Hairpin Bisulfite Sequencing: Synchronous Methylation Analysis on Complementary DNA Strands of Individual Chromosomes

Pascal Giehr and Jörn Walter

Abstract

The accurate and quantitative detection of 5-methylcytosine is of great importance in the field of epigenetics. The method of choice is usually bisulfite sequencing because of the high resolution and the possibility to combine it with next generation sequencing. Nevertheless, also this method has its limitations. Following the bisulfite treatment DNA strands are no longer complementary such that in a subsequent PCR amplification the DNA methylation patterns information of only one of the two DNA strands is preserved. Several years ago Hairpin Bisulfite sequencing was developed as a method to obtain the pattern information on complementary DNA strands. The method requires fragmentation (usually by enzymatic cleavage) of genomic DNA followed by a covalent linking of both DNA strands through ligation of a short DNA hairpin oligonucleotide to both strands. The ligated covalently linked dsDNA products are then subjected to a conventional bisulfite treatment during which all unmodified cytosines are converted to uracils. During the treatment the DNA is denatured forming noncomplementary ssDNA circles. These circles serve as a template for a locus specific PCR to amplify chromosomal patterns of the region of interest. As a result one ends up with a linearized product, which contains the methylation information of both complementary DNA strands.

Key words  Hairpin bisulfite sequencing, DNA strands, Chromosome, Restriction, Covalent linking, ssDNA circles

1 Introduction

Hairpin Bisulfite Sequencing (HBS) is a method to detect DNA methylation on both complementary DNA strands of individual DNA molecules [1]. HBS allows to discriminate if both strands are methylated or if a hemimethylation is present on only one of the two complementary DNA strands (upper or lower strand) or if both strands are symmetrically unmethylated. It also allows to discriminate a true non-CpG methylation from a genetic polymorphic (mutated) site. Hairpin bisulfite sequencing is more powerful and appropriate compared to conventional bisulfite sequencing
when one needs to detect the symmetry of DNA methylation patterns on both DNA strands, i.e., when analyzing active demethylation, de novo methylation or maintenance methylation events during cell replication or stages of reprogramming [2–4].

For the use of the HBS method the following general steps have to be considered. A standard HBS approach starts with the digestion of DNA by a defined restriction enzymes (usually four base cutter) that is not sensitive to DNA methylation, followed by covalent linking of the DNA fragments (upper and lower DNA strand) to a short hairpin DNA oligonucleotide using conventional ligation (Fig. 1). Restriction enzymes generating “sticky ends” should be preferred, since this will increase the efficiency of linker ligation. However, in our experience also the use of enzymes creating non-overhanging “blunt ends” is possible. The ligation is carried out using T4 DNA ligase. To ensure a high yield of DNA

---

**Fig. 1** Workflow of the Hairpin Bisulfite Sequencing protocol; genomic DNA is cut using a restriction enzyme, which is unaffected by DNA methylation. A complementary hairpin oligonucleotide is ligated to link upper and lower strand covalently together. The constructs are in the next step subject to bisulfite treatment resulting in single stranded circular DNA. After treatment the converted DNA serves as a template for a locus-specific PCR. PCR Products are then purified and sequenced. **Straight and dashed lines** indicate DNA strands, **red circles** illustrate CpG positions.
hairpin constructs the hairpin oligonucleotide is provided in excess to minimize the likelihood of religation of DNA fragments.

The overhang of the linker is designed with complementary overhangs. For example, the restriction enzyme MspI will leave a 5’ CG overhang, accordingly the hairpin linker also has to have a 5’ “CG” overhang (Fig. 2). Because of the enzymatic steps we recommend to use high quality (nondegraded) double stranded DNA (dsDNA). The circular constructs obtained after ligation are then subject to bisulfite treatment, in which all cytosines are converted to uracils. The steric closeness (intertwined ssDNA rings) of the complementary DNA strands favors a quicker renaturation to dsDNA. Hence the bisulfite conversion of hairpin constructs is more challenging than that of normal DNA. To avoid a reannealing we recommend using cycling bisulfite protocols with additional denaturation steps or alternatively higher incubation temperatures.

**Fig. 2** Workflow after bisulfite treatment; bisulfite treated circular hairpin constructs are amplified in two consecutive PCRs; the fusion primers used in the first amplification step carry on the 5’ end parts of the sequencing adapters which will become part of the PCR product. In the second PCR the rest of the adapter sequence is introduced to the amplicon. *Lines* indicate DNA strands, *red circles* illustrate CpG positions.
After bisulfite treatment the DNA molecules are present in form of single-stranded circular DNAs that contain uracil instead of unmethylated cytosines (Fig. 2). They serve in the next step as a template in a site specific PCR to amplify the region of interest. Here it is essential to utilize a polymerase that accepts uracils in the template. The product of this PCR holds now the methylation information of the upper as well as the lower strand. The generated amplicons can be treated like “normal” PCR products and can be sequenced using either Sanger or next generation sequencing (NGS). We have successfully combined hairpin bisulfite sequencing with the Roche FLX pyrosequencing system and the Illumina MiSeq system.

For the subsequent analysis we use the trimmed and quality-controlled FASTQ files and apply two bioinformatics tools develop in our lab. The first tool is the BiQ Analyzer (http://epigenetik.uni-saarland.de/de/software/) [5]. The program aligns the sequenced FASTQ reads to a reference HBS sequence (needs to be generated and provided). BiQ will provide an overview of CpG methylation and non-CpG methylation in the sequences. As outputs BiQ generates a tab separated table and different graphical representations, such as CpG methylation pattern map and quantitative pearl-necklace diagrams. The table output is then used in the next step by the Hairpinanalyzer script to back-fold the single strand information into a double stranded format. A more detailed description of the workflow is given in subsection 3.

The double stranded hairpin bisulfite sequencing output now allows to detect and quantitate the symmetry of CpG and CpNpG DNA methylation patterns on both DNA strands of one individual chromosome and to unambiguously identify nonsymmetrical cytosine methylation. We have applied the method to identify the massive occurrence of hemimethylated sites and general loss of methylation in Dnmt1KO ES-cells [2]. The hairpin bisulfite sequencing method provides matched stranded information, which directly shows the hemimethylated pattern formation.

The use of a hairpin linker also offers additional technical advantages. Since the added linker contains unmodified cytosines, it is possible to directly calculate the true conversion rate obtained during bisulfite treatment and at the same time can estimate the true amount of methylated non-CpG positions [3]. Further, in the loop sequences of the linker variable nucleotides can be introduced creating a barcoding for each DNA molecule. This allows to identify duplicated sequences generated during PCR and to exclude them from further analysis.

Despite of the many advantages of the Hairpin Bisulfite Sequencing method HBS also has some experimental limitations. In a locus-specific HBS analysis the obligate use of a suitable restriction enzyme (absence of recognitions site, distance to the analyzed region) can become a limitation. Moreover, the restriction
enzymes must not be affected by DNA methylation, which would otherwise lead to a massive underrepresentation of methylated sites. Further, the size of the region that can be analyzed is limited. Based on the fact, that the PCR product contains the information of both upper and lower strands the product will be double the size of the genomic region. However, this is only a small disadvantage, since modern sequencing techniques allow to analyze sequences with a size over 500 base pairs (bp) in length (particularly on a FLX or MiSeq sequencer).

2 Materials

2.1 Experimental Design

1. Identification of suitable restriction sites close to the region of interest using NEBcutter (http://tools.neb.com/NEBcutter2/) or WatCut restrictions analysis (http://watcut.uwaterloo.ca/template.php).

2. Design of the reference sequence (upper strand-linker-lower strand) for primer search and for the BiQ subsequent methylation analysis. The precise outline will be described in subsection 3.

3. Primers for PCR are either designed “by eye” or with the help of online tools like Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/).

4. Hairpin linkers are designed according to match the restriction enzyme ends and to containing indices/bar codes.

2.2 Sample Preparation

1. Qubit® BR assay kit for measurement of the DNA concentration (Thermo Fisher Scientific).

2. Restriction enzymes.


4. EZ DNA Methylation™ Kit; EZ DNA Methylation-Gold™ Kit (Zymo Research) or manual protocol.

5. HOT FIREPol® DNA Polymerase (Solis BioDyne) or Hot-StarTaq DNA polymerase (Qiagen) or similar.

6. Agencourt® AMPure® XP beads (Beckman Coulter).

7. Gel purification kit (e.g., AveGene or similar).

8. Primers and hairpin linkers.

9. dNTPs.

10. 25 mM MgCl₂.

11. 1× TE buffer: 10 mM Tris–HCl, pH 7.4; 1 mM EDTA, pH 8.0.
2.3 Sequencing and Data Analysis

1. Genome Sequencer FLX 454 System (Roche) or MiSeq Desktop Sequencer (Illumina)

2. For Methylation analysis BiQ Analyzer HT (http://biq-analyzer-ht.bioinf.mpi-inf.mpg.de/) was used followed by the use of Hairpinanalyzer. The python script of the hairpin analyzer can be received from http://epigenetik.uni-saarland.de/de/software/).

3 Methods

3.1 Experimental Design

A proper planning and design of the experiment for Hairpin Bisulfite Sequencing includes three main steps.

3.1.1 Selecting Restriction Enzymes

The first step when designing the Hairpin Bisulfite experiment is the search for suitable restriction enzymes within the region of interest. When selecting enzymes a few things should be considered.

1. First, the restriction enzyme should not be affected by DNA methylation to ensure that unmethylated, hemimethylated, and fully methylated regions are equally represented in the later results.

2. The use of enzymes, which create sticky ends, should be preferred since the ligation will work more efficient. Nevertheless, the use of blunt end creating enzymes is possible.

3. There are several online tools, which are suitable for the search of restriction enzymes for example “NEBcutter” or “WatCut.” Both tools also provide the information on methylation sensitivity and sometimes even other DNA modifications like 5hmC, 5fC, and 5caC.

3.1.2 Hairpin Linker and Reference Sequence Design

The hairpin linker itself can be divided into three sections.

1. The first part is variable and depends on the used restriction enzyme (Fig. 3). For example MspI will create a 5'-CG overhang; therefore the linker must contain a 5'-CG overhang including a free phosphate group to allow ligation to the DNA.

2. The second part always consists of the same sequence and facilitates the formation of the hairpin structure (Fig. 3). The use of unmodified cytosine within this linker part allows later an exact and unbiased calculation of the conversion rate during the bisulfite treatment and permits a more accurate detection of non-CpG methylation.

3. The last part is forming the loop structure of the linker. It contains a unique sequence that cannot form any double strand structures (Fig. 3).
4. As shown in Fig. 3 the loop can obtain 6–8 variable nucleotide positions. This allows to create an individual barcode for each DNA molecule and to exclude duplicates of the PCR in further computational analysis. Using eight of these positions it is theoretically possible to distinguish between 6561 ($3^8$) sequences.

Fig. 3 Schematic illustration of the hairpin linker; $A$, variable, restriction enzyme dependent sequence; $B$, constant, hairpin formation facilitating sequence; $C$, variable loop sequence; as an example the structure and sequence of a hairpin oligo for MspI restriction is shown in a denatured, single strand and annealed, folded state.

Fig. 4 Example of the design of a hairpin construct; The left part next to the restriction site is removed. The linker sequence is pasted followed by the reverse complement sequence of the right part of the DNA sequence.

3.1.3 Reference Sequence and Primer Design

The next step in the experimental design is the construction of the reference sequence. This sequence is needed in order to design primers and also for later sequencing and methylation analysis.

The reference sequence can easily be designed with any software that can handle text files.

1. Download the genomic sequence of the region, mark the restriction site used, delete the sequence in front or after the restriction site, add the linker sequence and finally paste the reverse complement of the remaining sequence on the other site of the linker (Fig. 4).
2. Replace all cytosines outside a CpG context by T to obtain the bisulfite sequence of the hairpin construct.

3. The Primers can then be designed either manually or using software or online tools (see Note 1).

3.2 Experimental Procedure

Restriction, ligation and bisulfite treatment are done in the same reaction tube without any purification steps in between which minimizes the loss of DNA. The described protocol below is the standard protocol used in our lab, however depending on the amount of DNA or the restriction enzyme used in the reaction an optimization or adjustments might be necessary. As starting material, genomic DNA from phenol–chloroform or various extraction kits can be used. Due to the nature of the method it is important to work with high quality and high molecular weight dsDNA. The concentration of the DNA should be determined using the Qubit system which only detects dsDNA.

3.2.1 Restriction

1. Cleave 200–500 ng of genomic DNA with 10 units of a restriction enzyme in 1 x Reaction buffer in a total reaction volume of 17 μL.

2. Incubate the reaction for at least 3 h at the recommended temperature, followed by a heat inactivation (see Note 2).

3.2.2 Preparation of Hairpin Linker and Ligation

1. Dissolve the oligonucleotide, which will later form the hairpin linker, in 1 x TE, resulting in a 100 μM stock and store at −20 °C.

2. Before usage, form the oligo into the right structure. Heat 50 μL of the 100 μM solution to 98 °C for 15 min followed by cool down using the slowest cooling rate of a thermocycler until 20 °C is reached. In this form the hairpin linker is rather stable but should be stored at −20 °C.

3. Add 1 μL of a 100 μM hairpin linker solution to the reaction together with 200 U T4 DNA ligase and 2 μL of 10 mM ATP.

4. Incubate the reaction for at least 3 h at 16 °C. The ligation can also be performed overnight.

3.2.3 Bisulfite Treatment

As mentioned before, there are several bisulfite kits, which can be used for the conversion of hairpin constructs. Since hairpin DNA molecules tend to fold back rather fast and the conversion of cytosine can only occur on single stranded DNA, we recommend a protocol with higher incubation temperature or additional denaturation steps. Kits successfully used in our lab are listed above in subsection 2. When using a homemade protocol, the conversion rate can be verified by looking into the conversion of cytosines included in the hairpin linker. A manual protocol used in our lab has been described previously [2]. In addition, the hairpin protocol is also extendable to oxidative bisulfite sequencing (see also Chapter 34) or other “chemical” forms of sequencing [6].
3.2.4 PCR

Because of the bisulfite treatment it is essential to use a polymerase that recognizes uracil in the template strand for the PCR. The best results in our lab were achieved using the HotFirePol from SolisBioDyne or HotStarTaq from Qiagen. We recommend a total reaction volume 30 μL and to perform multiple PCRs in parallel to ensure a low number of duplicated reads.

1. Pipette the reaction mixture according to Table 1 using either enzyme.

2. Both enzymes share similar temperature characteristics during PCR. A typical cycler protocol for both enzymes is given in Table 2.

3. Purification of the PCR products is performed using 27 μL AMPureXP Beads (0.9×).

<table>
<thead>
<tr>
<th>HotFirePol PCR</th>
<th>HotStarTaq PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 μL 10× reaction buffer</td>
<td>3 μL 10× reaction buffer</td>
</tr>
<tr>
<td>3 μL 25 mM MgCl₂</td>
<td>1.2 μL 25 mM MgCl₂</td>
</tr>
<tr>
<td>2.4 μL 10 mM dNTPs</td>
<td>2.4 μL 10 mM dNTPs</td>
</tr>
<tr>
<td>0.5 μL 10 μM forward primer</td>
<td>0.5 μL 10 μM forward primer</td>
</tr>
<tr>
<td>0.5 μL 10 μM reverse primer</td>
<td>0.5 μL 10 μM reverse primer</td>
</tr>
<tr>
<td>0.5 μL HotFirePol DNA polymerase</td>
<td>0.3 μL HotStarTaq DNA polymerase</td>
</tr>
<tr>
<td>Add to 30 μL ddH₂O</td>
<td>Add to 30 μL ddH₂O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temperature profile of HotFire/HotStarTaq</th>
</tr>
</thead>
<tbody>
<tr>
<td>95 °C</td>
</tr>
<tr>
<td>95 °C</td>
</tr>
<tr>
<td>50–62 °C</td>
</tr>
<tr>
<td>72 °C</td>
</tr>
<tr>
<td>72 °C</td>
</tr>
<tr>
<td>4 °C</td>
</tr>
<tr>
<td>35–45 cycles</td>
</tr>
</tbody>
</table>
3.2.5 Sequencing

Hairpin bisulfite is compatible with both Sanger and next generation sequencing (NGS). To prepare the samples for NGS one has to introduce adapter on each side of the amplicon, which are compatible with the sequencing platform to be used. These adapter bind to the sequencing platform and are the start point of the sequencing process. In addition each adapter carries a unique sequence ID, which allows sequencing of multiple samples at the same time. The adapter sequence is introduced by the use of fusion primers in two consecutive PCRs (Fig. 2).

In the first PCR the primers consist at their 3’ end of the target specific sequence complementary to the bisulfite treated DNA and carry at their 5’ end the first part of the adapter sequence resulting in an amplification of the target sequence and the introduction of the first part of the adaptor (Table 1 for PCR conditions).

In the second PCR the primer will bind the adaptor part introduced during the first PCR step. These primers carry the sequence, which later bind to the sequencing platform and in addition carry a sample specific sequence ID. Table 3 provides all primer and adapter sequences needed for sequencing on the Illumina MiSeq platform.

The second amplification can be performed as a multiplex PCR where several amplicons of distinct genomic regions can be prepared for sequencing at the same time. For this the concentration of each amplicons must be adjusted to 5 nM and pooled into one reaction.

1. Pipette components and perform the second PCR amplification PCR as outlined in Table 4 (see Note 3).

2. After incubation, clean-up the reaction again using 55 μL AMPureXP beads (1.1×).

### Table 3

**Illumina adapter and primer sequences; i5/i7 = index; grey = flow cell binding sequence; Oligonucleotide sequences © 2016 Illumina, Inc. All rights reserved**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.PCR Forward</td>
<td>TCTTTCCCTACACGACGCTCTTCCGATCT, Amplicon Specific</td>
</tr>
<tr>
<td>1.PCR Reverse</td>
<td>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT, Amplicon Specific</td>
</tr>
<tr>
<td>2.PCR Forward</td>
<td>AATGATACGGCGACCCAGATCTACAC[i5]ACACTCTTTCCCTACACGACGCTCTTCCGATCT</td>
</tr>
<tr>
<td></td>
<td>(i5, 6bp index e.g.: CGTGAT)</td>
</tr>
<tr>
<td>2.PCR Forward</td>
<td>GATCGGAAGAGCACACGTCTGACTTCAGACGATCA[i7]ATCTCGTATGCGGTCTCCGATCT</td>
</tr>
<tr>
<td></td>
<td>(i7, 6bp index e.g.: AAGCTA)</td>
</tr>
</tbody>
</table>

582 Pascal Giehr and Jörn Walter
3. Adjust to a final concentration of 10 nM.
4. Prepare the final amplicon library by pooling all enrichment PCRs into one reaction.
5. Following the MiSeq preparation protocol from Illumina, dilute the library stepwise to a final concentration of 18 pM.

3.3 Data Analysis

To obtain the information of the symmetric DNA methylation two separate steps are necessary. First, the methylation information has to be obtained from the sequenced PCR product. Second, the methylation information has to be translated back to the DNA double strand. For this we developed in our lab two different bioinformatics tools.

3.3.1 Methylation Analysis

The methylation analysis is performed using the BiQ HT Analyzer [2]. The BiQ HT is a Java based program designed for locus-specific DNA methylation analysis of high-throughput bisulfite sequencing data (see Note 4). The program aligns the sequenced reads to a reference sequence. Thereby comparing all positions where a cytosine is expected it detects the methylation status of the sequenced loci. Both, reference sequence and sequencing data has to be provided in a FASTA file format. The program calculates different quality scores including alignment score, sequence identity, bisulfite conversion rate and number of missing sites. The different quality scores can also be used to filter the data, for example against low quality reads or low sequence identity. BiQ HT will automatically use the default settings for filtering, but each value can be adjusted manually by the user. Besides the analysis of CpGs, BiQ is also able to detect methylation in a non-CpG context (CpHpG; CpHpH). After analysis the data is presented and can be stored in different ways. A typical output consist of a tab stop
separated table which includes all quality and methylation values and different types of methylation diagrams such as pattern maps and pearl-necklace diagram.

For the Hairpin analysis, three independent analyzing steps with BiQ HT are necessary. The first step is the detection of CpG methylation. The reference sequence used in this step consists of the sequence of both DNA strands and the linker sequence in between.

1. Replace in the linker sequence all cytosines by thymines because the linker will be analyzed later in an independent step. Depending on the type of loci, adjust the filter sequence identity. When analyzing repetitive elements for example, a lower sequence identity (80%) should be chosen because of the variability of the sequence of those elements.

2. Calculate the non-CpG methylation. Here it can be advantageous to replace all CpGs in the sequencing data by NpN to avoid confounding by possible mutations or sequencing failures, which create new CpG positions and lead to wrong estimations of non-CpG methylation. Again this is especially important when analyzing repetitive elements, due to their sequence variability. Note that also all CpGs in the reference should be replaced by NpN to allow an accurate alignment of the sequenced reads.

3. Estimate the conversion rate using the unconverted linker sequence as the reference sequence. Like in the non-CpG analysis the method to analyze Cs has to be chosen. When using wobble-position, within the linker loop, lower the sequence identity to 70 or even 60%, otherwise most of the sequences will be filtered out.

For each of the three analysis steps a separate folder has to be created in which the results are stored. The information in the different folders is then used to reconstruct the double strand information. Only the CpG folder is needed to reconstruct the double strand. However, conversion rate and non-CpG positions will then not be analyzed (see Note 5).

3.3.2 Reconstructing the Double Strand

The four folders containing the information about CpG methylation, non-CpG methylation, conversion rate and optionally SNPs are then used to reconstruct the double strand and calculate the amount of both strands methylated, only upper strand methylated, only lower strand methylated or both strands unmethylated.

For this purpose we have developed in our group a script we called Hairpinanalyzer. The program is based on python and is able to translate the single strand information, created by the BiQ HT,
into a double strand output. The Hairpin analyzer has no graphical interface and has to be run via command line and configuration of the source code. Like BiQ the Hairpin analyzer creates an output in form of a tab separated table and a methylation pattern map. An example of a pattern map created by the Hairpin analyzer is given in Fig. 5. Each column of the map represents a CpG position whereas each row indicates a sequenced read. The different methylation statistics include, both strands methylated, left (of the linker) strand methylated, right strand methylated, both strands unmethylated and mutated or not detectable. The colors can be chosen manually. The tab separated table contains information about number of reads, number of analyzed CpGs, Methylation status, Mutations or Sequencing errors as well as the information about SNPs and non-CpG methylation (see Note 6).

Fig. 5 Example of a methylation pattern map created by the Hairpin analyzer; each column represents one CpG dyad (1–5) and one row a sequenced read; red = fully methylated CpG dyads; light and dark green = hemimethylated CpG dyads; blue unmethylated CpG dyads; white bars indicate mutated or not analyzable CpGs; the bar on the left site shows a summary of the methylation over all CpG positions analyzed (M)
4 Notes

1. Even though both strands are no longer perfectly complementary after bisulfite treatment due to the conversion of cytosine; there are still regions that are complementary to some extent. This makes the primer design sometimes difficult because of dimer formations. To avoid this, it is sometimes necessary to use relatively short primers. This on the other hand will result in relatively low annealing temperatures. In our experience primers with a size of 23 bases and an annealing temperature over 50 °C are working fine. However, if possible a higher temperature should always be preferred, since it will increase the specificity of the PCR.

2. The restrictions conditions described in this protocol were suitable for all restriction enzymes used in our lab so far (BsaWI, DdeI, Eco47I, MspI, and TaqI). However, depending on the amount of DNA and the used restriction enzyme it might be necessary to change the parameters to obtain optimal reaction conditions.

3. When using next generation sequencing, the fusion primers of the PCR have to be adjusted to the sequencing system. In our lab we have successfully used FLX as well as Illumina systems for the sequencing of Hairpin Bisulfite amplicons.

4. The BQ Analyzer HT can be downloaded from http://biq-analyzer-ht.bioinf.mpi-inf.mpg.de. There is also a Java Web Start version available as well as detailed documentation.

5. It is also possible to include a fourth folder that contains the information about SNPs and the barcode of the linker. For that a fourth analysis with BiQ has to be performed. By repeating the analysis for CpG including the option output SNPs.

6. The Hairpin Analyzer was programmed by Mathias Bader in our lab. Unfortunately this script is not available online, but can be obtained upon request from the authors.

References


