Introduction
Sequencing of sodium-bisulfite modified genomic DNA originally introduced by M. Frommer (Frommer et al., 1992) is a widely used "gold standard" method for DNA-methylation analysis. Since this method relies on a harsh chemical treatment of DNA it causes a lot of DNA damage and hence a dramatic loss of high quality DNA for PCR amplification and further analysis. In the meantime several commercial kits are available for this procedure which work reasonably well when starting with large amounts of DNA.

Here we describe a protocol for small numbers of cells and little DNA which requires some specific handling. The protocol is based on a strategy originally introduced by our lab (Hajkova et al., 2002) using agarose embedded DNA. This physical trapping helps to avoid DNA loss during the various incubation steps while maintaining a good bisulphite conversion rate. We will introduce two alternative procedures to perform bisulphite treatment of agarose embedded small DNA aliquots or cells and guide through some generally critical points in the bisulphite reaction and primer design. We also include tips for the process of data processing after sequencing which is facilitated by a new and very useful software tool (BiQ Analyzer). This tool allows rapid and reproducible processing and evaluation of bisulphite sequencing data. It generates standardized table output formats allowing direct database integration.

Procedure
Two alternative protocols (A and B) for agarose embedded DNA of cell preparation

In the following we will describe two protocols for bisulphite treatment of small aliquots of mammalian cells/DNA. Both procedures only differ in the first step, i.e. the preparation of the DNA for bisulphite treatment. We routinely use protocol A for collected mammalian cells in the range >100 and B for very scarce material (such as zygotes/early mouse embryos), i.e. few to 100 cells.

Protocol A: Preparation of genomic DNA from samples with more than 100 cells per sample

1. Mix the collected mammalian cells (>100) with 50 µl of solution A (25 mM EDTA, 75 mM NaCl), 50 µl of solution B (10 mM EDTA, 10 mM Tri-HCl, pH 8.0, 1% SDS), add 2 µl of Proteinase K (20 mg/ml) and 5 µg of yeast t-RNA (use 1.5 ml Eppendorf tubes)
2. Incubate 1h at 37°C
3. Add 50 µl of phenol and 50 µl of chloroform/isoamylalcohol (24:1)
4. Mix the samples gently by inverting the tubes
5. Centrifuge the samples at 13,000 rpm for 10-15 min
6. Transfer the supernatant into fresh tubes and add 100 µl of chloroform/isoamylalcohol (24:1)
7. Mix the samples gently by inverting the tubes
8. Centrifuge 5 min at 13,000 rpm
9. Transfer the supernatant into fresh tubes and add another 5 µg of t-RNA and 1/10 volume of 3M sodium acetate (pH5.5), then add 350 µl of 100% ethanol
10. Precipitate the DNA at -20°C over night or longer
11. Centrifuge the samples for 30 min at 13,000 rpm at 4°C
12. Remove the ethanol and let the pellets dry at room temperature
13. Dissolve the pellet in 4 µl ddH₂O over night (do not dissolve the pellets in larger volumes since the DNA is later embedded into one single agarose bead – highest yield; dissolving the pellet over night is necessary to bring the DNA entirely into solution)
14. Add 0.4 µl of freshly prepared 2 M NaOH and 10 µl of molten 2% LMP agarose (comment 1)
15. Overlay with 200-500 µl of heavy mineral oil (white, Aldrich)
16. Boil for 10 min in a water bath
17. Cool down on ice until the bead is solid
18. Proceed with bisulphite treatment

Protocol B: Bisulfite treatment in agarose embedded cells

This Protocol should be used as an alternative if you want to avoid precipitation steps, e.g. when working with very low amounts of cells/DNA; few to 100 cells or equivalent DNA content (up to 300 pg).

1. Add 15 µl of 2% LMP agarose to the sample and overlay with 100-300 µl of heavy mineral oil (note: total volume of the Cell suspension should not exceed 5µl)
2. Boil the samples for 10 min in a water bath to destroy the cells
3. Cool the samples for at least 10 min on ice until the bead is solid
4. Remove the oil (be careful in case if the bead is still not solid)
5. Add 400 µl of solution A (25mM EDTA, 75 mM NaCl) and 50 µl of Proteinase K (20mg/ml). Centrifuge briefly to make sure that there is no residual oil between the bead and the solution
6. Incubate at 50°C overnight
7. Carefully remove the solution as complete as possible
8. Wash the samples 2x with 400 µl of 1xTE at RT
9. Rinse the bead with 400 µl of 0.3M NaOH
10. Wash the sample 2x with 400 µl of 0.3M NaOH
11. Wash the samples 2x with 1xTE
12. Remove the TE entirely and store the bead at 4°C (no longer than a couple of days)
13. Proceed with bisulphite treatment

Agarose embedded bisulfite modification

Agarose beads formed according to protocols A or B are used for the following procedure:

1. Prepare bisulfite solutions freshly as follows: mix 1.9 g sodium metabisulfite (Merck no. 1.06528.0500) with 2.5 ml ddH₂O and add 750 µl of freshly prepared 2 M NaOH, dissolve the powder by strong vortexing (it takes time, if the powder does not dissolve it can be done at 50°C). Dissolve 55 mg of hydroquinon (Sigma) in 500 µl ddH₂O at 50°C. Mix both solutions when chemicals are dissolved completely. (comment 2)
2. Add 400 µl of freshly prepared bisulfite solution to each bead (as prepared according to the protocol A or B. Overlay with mineral oil (following protocol B). (When adding the solution to beads from protocol A make sure the beads are in contact with the solution and there is no residual oil between the bead and the solution; brief spin down helps)
3. Transfer the samples to 50°C and incubate for 3.5 h
4. Cool the samples briefly on ice (makes the bead better visible and “harder” for solution exchange)
5. Remove the bisulfite solution and rinse the bead with 400 µl of 1x TE at RT (put 400 µl of 1x TE onto the bead and remove the TE immediately to avoid starting diffusion)
6. Wash the samples 2x with 400 µl 1xTE for 15 min at RT
7. Rinse the bead with 400 µl of 0.3M NaOH at RT
8. Wash the sample 2x with 400 µl of 0.3M NaOH at RT
9. Wash the samples 2x with 1xTE at RT
10. Remove the TE entirely and store the bead at 4°C (can be stored for 2 weeks max.)
11. Proceed with Bisulphite PCR amplification (Bis-PCR)

Bisulphite PCR amplification, cloning and sequencing
For Bisulphite PCR-amplification we regularly use the HotFirePol from Solis BioDyne, Tartu, Estonia. It combines the characteristics of a Hot Start Taq with a good performance und a high cost-efficiency. We experienced that some amplicons may work better with HotStarTaq from Qiagen (has to be tested).

Our standard Bis-PCR protocol runs as follows:

Melt the agarose bead at 70-80°C and transfer 3 µl of the molten agarose into the PCR tube containing the following ingredients (kept at RT or on ice):

- 5 µl 10xbuffer (Solis BioDyne)
- 5 µl MgCl₂ (25 mM)
- 4 µl dNTPs (2.5 mM each)
- 1 µ l forward primer (10 µM)
- 1 µl reverse primer (10 µM)
- 0.6 µl HotFirePol (5 U/µl)
- 30.4 µl ddH₂O

Transfer into a preheated (95°C) PCR block and start the following program:

PCR program:

- 97°C  15 min (required for HotFirePol)
- 40 cycles of:
  - 95°C  1 min
  - 52°C  1 min
  - 72°C  1min30sec

Load 5 µl on a 1.2% agarose gel and purify the PCR product by gel extraction (using JetsorbTM from Genomed) or precipitation with SureClean from Bioline, store rest at -20°C. (comment 3)

Cloning and “colony” PCR

We usually use 3 µl of the purified PCR product for pGEM-T T/A cloning, following the recommendations of the manufacturer, and plate on LB agar containing 100 mg/ml Amp, 0.04% X-GAL, 0.2 mM IPTG. Colonies are picked after overnight incubation with a sterile tooth pick. Briefly dip the tooth pick into a PCR-tube containing 30 ml of the following Colony-PCR Mix :

- 3 µl 10x Standard PCR buffer (containing 20mM MgCl)
- 2.4 µl dNTPs (2.5 mM each)
- 1 µl colony PCR forward primer (10 µM)
- 1 µl colony PCR reverse primer (10 µM)
- 0.5 µl Taq polymerase ( U/µl)
- 22.1 µl ddH₂O

For amplification we use the Colony-PCR program:

- 95°C  5 min
- 30 cycles of:
  - 95°C  30 sec
  - 65°C  30 sec
  - 72°C  1min
- 72°C  5 min

Comments to cloning and colony PCR: We use pGEM-T specific “colony PCR” forward (5´ gct att acg cca gct ggc gaa agg ggg atg tg 3´) and reverse (5´ ccc cag gct tta cac ttt atg ctt ccc gct cg 3´) primers which are located in the vector approx. 150 bp upstream and downstream of the insert hence enlarging the insert by 300bp. The “colony PCR fragments” include the universal M13 forward and reverse primer sequences which make them amendable for Sanger sequencing.

For cloning we routinely use the pGEM-T cloning kit from Promega, but PCR products may behave different with
respect to cloning efficiency, so that it might be occasionally useful test other vector systems. We even experienced
that some fragments are not clonable at all into a particular vector. We do not find much of a difference with respect
to the bacterial strain used (such as Sure, Top10). Different ligation times and temperatures may also influence the
cloning efficiency. Best results we obtained incubating the ligation mix at 16°C over night.

The particular sequence composition of bisulphite converted DNA often results in repeated structures and extended
homopolymer stretches. This might influence the orientation of the cloned fragment and occasionally the
maintenance of the cloned fragments in bacteria (instabilities). We recommend to pick clones from fresh overnight
plates. (comment 4)

Sequencing of bisulphite clones

We perform the sequencing following standard Sanger-procedures (Sanger et al., 1977) on a Beckmann-Coulter
sequencer using the DTCS QuickStart Sequencing kit. In principle all standard sequencing platforms can be used
such as ABI 3100, 3730, Amersham Megabase etc. For sequencing we usually prepare ("purify") the colony-PCR
templates by PEG precipitation (protocol available on request) or use gel purification. For a standard sequencing
reaction about 100-200ng of template is needed. We mostly use standard sequencing primers (M13 forward: 5´ gtt
ttc cca gtc acg acg 3´, M13 reverse: agg aaa cag cta tga cca t 3´) for the reaction but occasionally perform reactions
with template specific primers.

Author Notes

Primer design

One of the most critical issues for successful bisulphite experiments is to carefully select the regions suitable for
bisulphite analysis and the design of accurately positioned and matching primers. Our experience is that there is no
good specific prediction program for the design of bisulphite PCR primers available. Hand made selections (with
some "experience") work best. (comment 5) Selected primers should however be examined with software such as
Oligo 6.0; Vector NTI etc. manually for similarities in melting temperatures, secondary structures and primer dimer
formation. We usually vary the oligo length to adjust annealing performance and avoid secondary structure effects.
Primers work best between 25-28 bp in length. Both primers should have an optimal Tm at around 65°C which
means an effective annealing temperature of 52°C in the PCR reaction. (comment 6) As a threshold we recommend
the primers do not differ more than 1-2°C in their Tm's. The melting temperature of the PCR product should also not
differ too much from the melting temperature of the primers.

Primers should be designed within originally cytosine rich but CpG depleted regions. One should try to include at
least 4-5 cytosines preferably at the 3’ end of the primer which will be converted to T's. This enhances the
“specificity” to truly amplify bisulphite converted DNA sequences and to “eliminate” not fully converted material. CpG
positions within primers should be avoided - our experience with the use of "wobble" primers (to cover CpG
positions with either C or T) were not very positive, i.e. we observed preferential amplification of methylated
templates.

In older protocols (Hajkova et al., 2002) we recommended nested PCR approaches to enhance sensitivity and
specificity. The current protocol yields better DNA quality and we mostly find the nested approach to be
unnecessary. Avoiding nested amplifications has advantages because it reduces the risk of cross contamination
and of preferential primary template amplification (amplification bias). According to our experience in almost all
cases one round of PCR (max. 40 cycles) is sufficient, even if small amount of material is used (analysis of single
copy genes from 20 cells, i.e. 60 pg genomic DNA). (comment 7) (figure 1)

BiQ-Analysis of Bisulphite-sequence data (clones)

The conventional procedure to analyze bisulfite-modified DNA is to amplify the region of interest, to clone the PCR
product and analyze single clones by Sanger-sequencing. By this high resolution single chromosome profiles are
obtained. A big issue in this process is the evaluation (quality) and quantification of sequencing profiles. To facilitate
the analysis we have developed the BiQ Analyzer program (Bock et al., 2005, which runs on PC and Mac's and is
freely available at http://biq-analyzer.bioinf.mpi-sb.mpg.de/). The program uses ClustalW to automatically align the
single amplicon sequences to the original genomic sequence. It calculates sequence identity, conversion rate and
methylation degree for each clone and provides a colour coded alignment (see figure 2a). Sequences below 80%
sequence identity to the original sequence and below 90% conversion rate are marked and can be excluded from
analysis. The data are exported in excel type tables and automatically provide a graphical representation ("Lollipop"
graphs) (figure 2b) and a bar diagram calculating the methylation degree for every CpG position in all clones are
given as output files (figure 2c). Whole methylation of the amplicon is calculated by the following formula: no. of
methylated sites/ no. of all analyzed sites. BiQ Analyzer saves time and provides standardized data and output formats. BiQ still requires some manually quality control of the processing - the program sometimes does not exclude "bad" sequence traces that show very low sequence quality.

The statistical significance of the bisulphite results is very much dependent on the number of sequences analyzed. This number can vary from experiment to experiment depending on the methylation variation observed. We recommend analyzing at least some 15-20 clones for each amplicon, ideally derived from two independent bisulphite treatments to obtain a first reasonable estimate. Ideally the data obtained by bisulphite clone sequencing should be confirmed by a different method since cloning bias (i.e. preferred cloning of unmethylated or methylated sequences) and preferential amplification may play a role (see also Clark et al., 2006).

**Brief overview of other methods for sequence analysis**

Besides labor and cost intensive single clone analysis by Sanger sequencing it is also possible to analyze the average methylation of CpG positions within the PCR amplicon by "direct" sequencing of the bisulphite-PCR product. Here pyrosequencing has been shown to be very efficient and quantitative but reaching only few CpG positions (Tost et al., 2007). Chromatograms from direct sequencing by the Sanger-method are rather difficult to quantitate without sophisticated software, i.e. such an analysis requires software such as ESME (Lewin et al., 2004; see also [http://www.epigenome.org/](http://www.epigenome.org/)), which are not publicly available. If only single or few CpG position shall be interrogated SNuPE based methods such as SIRPH (El-Maarri et al., 2002, El-Maarri, 2004) or real time PCR approaches such as MethyLight (Trinh et al., 2001) can be recommended since they are less laborious and once established very reliable and quantitative.

**SIRPH as an alternative to sequencing**

In case the exact CpG positions to be quantified for methylation are known in a given fragment, SIRPH (and other SNuPE-based methods) offer(s) a cheaper and less laborious possibility to obtain average methylation values. We are using this method routinely for large scale sample comparison. Briefly, a primer is placed next to the CpG position to be interrogated. Extension by ddCTP (or ddGTP on the other strand) indicated methylation, extension by dTTP an unmethylated state. The ratio of incorporated nucleotides (extended primers) can be examined simultaneously on a DHPLC WAVE™ system from Transgenomic (using DNASep column). T and C extended primers are simply separated according to their hydrophobicity and can be distinguished from unextended primers by length. By measuring the height or area of the obtained peaks the ratio C/C+T is calculated. Further information on the method is available on request.

Although SIRPH is a simple, quick and cost-efficient method it can only analyze one or two CpG positions simultaneously. Therefore it is not suitable for the analysis of complex methylation patterns. Moreover to really quantify the methylation for a given position requires the establishment of a reference curve – since the C and T incorporations at a specific position may vary form position to position. Additionally, conversion rate (i.e. the bisulphite quality check) will have to be tested using an additional primer or partial sequencing of clones.

**Reviewer Comments**

Reviewed by: Jörg Tost Centre National de Génotypage, CEA / Institut de Génomique, 91057 Evry Cedex, France.

1. Supplier should be specified. We are using SeaPlaque™ low melting agarose from FMC.
2. We have experienced no difference in performance if the solution stays slightly clouded and contains some undissolved crystals.
3. A simple and quick alternative for purification consists of a centrifugation step of the PCR product through a Multiscreen plate (Millipore) filled with P100 Bio-Gel (Bio-Rad) acrylamide suspension.
4. This is an important point and we have frequently experienced problems with instability of inserts. Shorter incubation times and even lower concentrations of antibiotics might help to overcome this problem. In some cases a slightly lower incubation temperature 36°C might have beneficial effects.
5. We have good success rates using freely available design software such as MethPrimer, which is based on the widely used Primer3 algorithm or MethylPrimer express from Applied Biosystems.
6. We use an annealing temperature of 60°C, which in our hand reflects quite well the real temperature when using the Qiagen HotStar Taq.
7. We do also not recommend any nested PCR approaches.

**Figures**
Fig. 1 An example of proper (green) and improper (red) primer positioning. Highlighted in green is the best position to place the primer. This region contains originally 7 cytosines where at least four of them are located at the 3’ part of the primer. The region highlighted in red shows a position where in total 5 originally cytosines are included, but they are all located in the 5’ part of the primer. In this case it might happen that not fully converted sequences are amplified as well.

Fig. 2a Example of a BiQ generated ClustLW alignment representing the methylation of 7 CpG positions in single clones; unconverted Cs are highlighted in green.

Fig. 2b “Lollipop” scheme representing unmethylated (white circles) and methylated (black circles) CpG positions; this diagram is generated based on the ClustLW alignment including the relative CpG distances.
Fig. 2c Bar diagram showing the numbers of methylated and unmethylated CpG positions; the unmethylated portion is shown in blue, the methylated in yellow; not present designates CpGs that were not analysable due to sequencing errors; the distance between the single CpG positions is given in bp.

<table>
<thead>
<tr>
<th>CpG position</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmethylated</td>
<td>10</td>
<td>5</td>
<td>14</td>
<td>3</td>
<td>14</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Methylated</td>
<td>9</td>
<td>14</td>
<td>5</td>
<td>16</td>
<td>5</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Not present</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

References