Analytical Biochemistry 377 (2008) 46–54

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Single-molecule polymerase chain reaction reduces bias: Application to DNA methylation analysis by bisulfite sequencing

Aparna Chhibber, Benjamin G. Schroeder *

Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA

ARTICLE INFO

Article history:
Received 10 November 2007
Available online 4 March 2008

Keywords:
DNA methylation
5-Methylcytosine
Bisulfite
Single-molecule PCR
Artifacts
DNA sequencing

ABSTRACT

The treatment of DNA with bisulfite, which converts C to U but leaves 5-methyl-C unchanged, forms the basis of many analytical techniques for DNA methylation analysis. Many techniques exist for measuring the methylation state of a single CpG but, for analysis of an entire region, cloning and sequencing remains the gold standard. However, biases in polymerase chain reaction (PCR) amplification and in cloning can skew the results. We hypothesized that single-molecule PCR (smPCR) amplification would eliminate the PCR amplification bias because competition between templates that amplify at different efficiencies no longer exists. The amplified products can be sequenced directly, thus eliminating cloning bias. We demonstrated this accurate and unbiased approach by analyzing a sample that was expected to contain a 50:50 ratio of methylated to unmethylated molecules: a region of the X-linked *FMR1* gene from a human female cell line. We compared traditional cloning and sequencing to smPCR and sequencing. Sequencing smPCR products gave an expected methylated to unmethylated ratio of 48:52, whereas conventional cloning and sequencing gave a biased ratio of 72:28. Our results show that smPCR sequencing can eliminate both PCR and cloning bias and represents an attractive approach to bisulfite sequencing.

DNA methylation, in which cytosine in a 5′-CpG-3′ sequence context is methylated at the 5 position, is associated with normal processes such as epigenetic reprogramming [1], X chromosome inactivation [2], environmental exposure and aging [3,4], and disease processes [5,6] including cancer [7,8], mental retardation, and autoimmune disease [9].

Advances in the bisulfite conversion of DNA [10,11] have led to increased use of this method for methylation analysis. Once normal C is converted to U, analysis essentially becomes a problem of identifying and measuring sequence differences. Many techniques have been developed, spanning the spectrum from those that analyze a single nucleotide difference to those that measure global changes in methylation [8,12]. One very popular technique is bisulfite sequencing, first described by Frommer et al. [13] in 1992. In contrast to single-nucleotide-polymorphism-based techniques, sequencing a CpG-containing region measures the methylation state of all the CpGs within the region and is useful for detailed analysis and initial surveys of regions suspected to be under methylation regulation. Another advantage of bisulfite sequencing is the built-in quality control for bisulfite conversion. Because methylation happens only at CpG, all CpH (CpA, CpC, and CpT) cytosines should be converted to T in the resulting bisulfite sequence. The detection of CpH is evidence of incomplete bisulfite conversion.

There are two methods commonly used when performing bisulfite sequencing, each with benefits and drawbacks. One method is to directly sequence the amplicon. The other method is to PCR amplify a region of interest, clone it, and then pick several clones for sequencing.

Although direct amplicon sequencing is faster and requires less sequencing, there are several issues to consider. Sequence data generated from fluorescently labeled dideoxy terminators requires extensive normalization and processing to quantitatively detect 20% differences [14]. Pyrosequencing is more quantitative but is limited to at most 15 CpGs in an amplicon of no larger than 100 bp [15]. Furthermore, any approach that analyzes a pool of molecules, regardless of its quantitative accuracy, cannot provide information about the methylation haplotype, that is, the methylation status of each CpG within an individual DNA molecule. Finally, when analyzing PCR amplicons derived from a mixed pool of template molecules, care must be taken to measure PCR bias, which can often be severe [16–18]. If PCR bias is observed, steps must be taken to correct for it or to change reaction conditions to try to reduce it. PCR bias is a result of the difference in amplification efficiencies of sequences within a PCR. Sequences that amplify more efficiently become overrepresented in the final amplicon population. Several approaches for the reduction of bisulfite-specific PCR bias have been reported. Voss et al. [17]

* Corresponding author. Fax: +1(650)-554-2660.
E-mail addresses: schroebg@appliedbiosystems.com, bschroeder100@gmail.com (B.G. Schroeder).

0003-2697/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2008.02.026

reported that bias in favor of unmethylated targets could be reduced by the addition of betaine to the PCR. Wojdacz and Hansen [19] found that the amplification bias in favor of unmethylated targets could be reversed by redesigning the primers to include one or two CpGs. Shen et al. [20] reported that the amount of PCR bias observed can change depending on the annealing temperature of the PCR.

Unlike direct PCR sequencing, PCR amplification followed by cloning and sequencing has the advantage of producing methylation haplotypes. However, the process still relies on PCR from a pool of molecules and can be subject to PCR bias. One additional source of bias associated with cloning and sequencing is the cloning step itself [21], where the ratio of methylated to unmethylated PCR products is not faithfully maintained through the ligation, transformation, and bacterial growth steps.

We sought to circumvent these problems by using single-molecule PCR (smPCR)[1] bisulfite sequencing. smPCR has been employed for the analysis of single-copy genes in sperm (see [22] and references therein), for digital PCR [23], for analysis of somatic mutations (reviewed in [24]), for massively parallel sequencing (reviewed in [25]), and for screening protein expression libraries (see [26] and references therein).

Here we demonstrate that single-molecule PCR bisulfite sequencing provides all the benefits of bisulfite cloning and sequencing (quantitation, methylation haplotypes, and internal QC) while eliminating the problems of PCR and cloning bias. The result is detailed, unbiased, molecule by molecule methylation mapping which can provide quantitative descriptions of the methylation state of the region of interest and the methylation status of each individual CpG within the region.

Materials and methods

Bisulfite conversion

Female human cell line genomic DNA from CEPH individual 1347-02 (Applied Biosystems 403062) or in vitro universally methylated genomic DNA (Chemicon S7821) was converted using the methylSEQr Bisulfite Conversion Kit (Applied Biosystems 4379580) according to the manufacturer's instructions. Genomic DNA (300 ng) was converted and isolated in a total volume of 50 μl for a concentration of 6 ng/μl.

PCR, cloning, and sequencing

The *FMR1* region was amplified in a 50-μl PCR containing 1X Gold Buffer, 200 nM each dNTP, 2 mM $MgCl_2$, 0.1 unit AmpliTaq Gold polymerase (all from Applied Biosystems), 0.5 mg/ml bovine serum albumin (Sigma B8667), 0.5% (v/v) glycerol (Shelton Scientific IB15760), 250 nM each primer (forward: 5′-GTGTAAAACGACGGCCAGTTGAGTGTATTTTTGTAGAAATGGG; reverse: 5′- GCAGGAAACAGCTATGACCTCTCTCTTCAAATAACCTAAAAAC; underlined portions correspond to sequencing primer tails), and 30 ng (approximately 10,000 copies) bisulfite-converted genomic DNA. The reaction was thermally cycled in a GeneAmp PCR System 9700 (Applied Biosystems) as follows: 5 min at 95 °C (to activate the hot-start polymerase), 5 cycles of 95 °C/30 s, 60 °C/2 min, 72 °C/3 min; 30 cycles of 95 °C/30 s, 65 °C/1 min, 72 °C/3 min; hold at 72 °C/7 min; hold at 4 °C. The control was an identical reaction without genomic DNA. An aliquot of each reaction was analyzed by agarose gel electrophoresis to confirm the presence of the expected 312-bp amplicon and the absence of any product in the control reaction. One microliter of PCR product was cloned into pCR4-TOPO using the TOPO TA Cloning Kit (Invitrogen) according to the manufacturers instructions, and a 2-μl aliquot of this reaction was used to transform One Shot TOP 10 chemically competent cells (Invitrogen). After overnight growth on LB agar plates containing 30 μg/ml kanamycin, isolated colonies were inoculated into 2.3 ml of LB containing 100 μg/ml ampicillin for overnight growth at 37 °C with shaking. Plasmids were purified with the QIAprep Spin Miniprep Kit (Qiagen). Sequencing reactions contained 2 μl plasmid DNA, 4 μl BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems), and 320 nM primer 5′- ATTAATGCAGCTGGCACGAC in a 10-μl volume. Reactions were thermally cycled as follows: 1 min at 96 °C, 25 cycles of 96 °C/10 s, 50 °C/4 min, then purified with the BigDye XTerminator Purification Kit (Applied Biosystems) according to the manufacturer's instructions. Purified sequencing products were analyzed on a 3130*xl* Genetic Analyzer with a 50-cm array and POP-7 polymer (all from Applied Biosystems). Data were analyzed with Sequencing Analysis 5.2 with the KB Basecaller and SeqScape 2.5 (all from Applied Biosystems).

Primer screening

Twelve primer sets previously described [18] were screened for their propensity to generate template-independent amplification (primer dimer). Duplicate 10-μl no-template reactions containing 1X SYBR Green Master Mix (Applied Biosystems 4309155) and 250 nM each primer were thermally cycled as follows: 10 min at 95 °C (to activate the hot-start polymerase), 5 cycles of 95 °C/15 s, 60 °C/2 min, 72 °C/3 min; 40 cycles of 95 °C/15 s, 65 °C/1 min, 72 °C/3 min. Reactions were monitored on a 7900HT Real-Time PCR instrument (Applied Biosystems).

Single molecule PCR

SYBR Green reaction mixtures were prepared with a series of 10-fold serial dilutions of bisulfite-converted genomic DNA template. For each template dilution, and a no-template control mixture, a 5-μl aliquot was transferred into each of 12 wells for real-time PCR analysis as described above. Negative wells were counted and used to calculate the copies/μl concentration of the bisulfite-converted genomic DNA stock as follows.

The distribution of molecules into wells is modeled as a Poisson distribution, where the probability of getting k molecules in a well is $P(k) = e^{-\lambda}\,\lambda^k / (k!)$, where e is the base of the natural logarithm, $k!$ is the factorial of k, and λ is equal to the expected number of occurrences per well, in other words, the average number of molecules per 5-μl aliquot. From this relation, it follows that the probability of getting 0 molecules in a well is $P(0) = e^{-\lambda}$, the probability of getting one molecule in a well is $P(1) = e^{-\lambda}\lambda$, and the probability of getting two or more molecules in a well is $P(>1) = 1-P(0)-P(1)$. A positive amplification can occur as a result of one or more than one molecule present in the well, whereas a negative amplification indicates that no molecules were present or that the PCR failed, but this can be ruled out if higher concentrations in the dilution series generated amplification. The fraction of wells that do not amplify should therefore equal $P(0)$, and thus the average number of molecules per 5-μl aliquot is $\lambda = -\ln(P(0))$. Based on the dilution used, the stock concentration can then be calculated.

For single-molecule methylation analysis of *FMR1*, a 1X SYBR Green reaction containing 250 nM each primer and 0.01 copies/μl bisulfite-converted genomic DNA was prepared and 20 μl per well distributed to 336 wells of a 384-well plate. The remaining 48 wells were filled with 20 μl each of the corresponding no-template control reaction. The plate was thermally cycled as above. For analysis without a real-time PCR instrument, after thermal cycling the

[1] *Abbreviations used:* smPCR, single-molecule PCR; NTC, no-template control.