ABSTRACT

We report the profiling of the 5-methyl cytosine distribution within single genomic-sized DNA molecules at a gene-relevant resolution. This method linearizes and stretches DNA molecules by confinement to channels with a dimension of about 250×200 nm². The methylation state is detected using fluorescently labeled methyl-CpG binding domain proteins (MBD), with high signal contrast and low background. DNA barcodes consisting of methylated and non-methylated segments are generated, with both short and long concatemers demonstrating spatially resolved MBD binding. The resolution of the technique is better than 10 kbp, and single-molecule read-lengths exceeding 140 kbp have been achieved.

KEYWORDS: DNA, Nanochannels, Epigenetics, Cytosine Methylation

INTRODUCTION

Epigenetic regulation is the inheritable modification of gene activity without influencing the underlying DNA sequence. DNA 5-cytosine methylation is one of the most widely-studied mechanisms influencing epigenetic gene regulation, and is generally thought to suppress gene expression. A CpG pattern in which cytosines on both strands carry this modification can be maintained through DNA replication, and thus cell division. The methylation of CpG islands contributes to various biological processes such as parental genomic imprinting, X-chromosomal inactivation, cellular differentiation, aging, and cancer. The ability to detect the hypo- or hyper-methylation state of the CpG sites is useful in predicting gene transcription, which can have profound consequences for human health.

Traditional ensemble methods for methylation profiling include methylation-specific PCR, combined bisulfite restriction analysis (COBRA), methylation-sensitive single-nucleotide primer extension (Ms-SNuPE), methylated DNA immunoprecipitation, hybridization arrays, restriction landmark genome scanning, next-generation sequencing after bisulfite conversion. These methods require a multitude of cells for analysis, with the disadvantage that rare cells are poorly represented or have to be specifically enriched in order to make meaningful statements. That becomes particularly important when the epigenetic state of small sub-populations such as cancer stem cells is sought.

In the field of protein biology, single-molecule investigations have yielded important insights into the intricate relationships of rare states in a population. Single-molecule methylation detection has been recently demonstrated through nanopore technology, and single-molecule real-time (SMRT) DNA sequencing. In principle, these methods require no front-end amplification and sample preparation of the DNA. Bisulfite conversion and sequencing schemes can also yield single-molecule data, but cannot generally guarantee that single molecules were targeted and requires intensive preparation.

In our work, methylation patterns are detected through binding of a fluorophore-tagged methyl-CpG-binding domain (MBD) protein fragment to the interrogated dsDNA segment, as shown in Fig. 1. The binding pattern along the DNA is detected by fluorescence microscopy. In order to achieve single-gene relevant resolution, DNA is stretched by confinement to a quasi one-dimensional nanochannel (Fig. 1). The technique thus is conceptually similar to fluorescence in-situ hybridization on molecules that were elongated molecules (fiber-FISH), which are arrested in their extended configuration through a technique such as molecular combing.

Nanochannel stretching itself is an emergent technique that has been used to map the length of DNA fragments, image the binding of GFP-fusion transcription factors, observe real-time ordered restriction mapping, and perform single-molecule melting temperature mapping. While nanochannel mapping appears similar to stretching through molecular combing, it differs in that the stretched state represents the equilibrium configuration of the molecule, such that the molecule can fluctuate around that equilibrium configuration. In contrast, fiber-FISH and molecular combing lock the nucleic acid in a single non-equilibrium configuration.

In order to provide a robust testing vehicle with minimal biological complexity, we used λ-phage DNA concatemers as a model system for genomic DNA (Fig. 1). By concatenating fully CpG-methylated and non-methylated strands, we created a predictable barcode that enables us to judge both the detection efficacy and the mechanical properties of the probe-substrate complex. We have shown that MBD binds specifically to methylated DNA, and that the spatial location of binding sites within the molecules can easily be mapped. However, we also noticed that binding of MBD to methylated DNA segments leads to a contraction that is dependent on the MBD quality. We believe that the success in using this artificial barcode sample implies that the technique can be extended to real genomic DNA molecules. Our results imply detection resolution on the order of 10 kbp, which is roughly the size of an average human gene.
FIG. 1. a) Schematic of possible outcomes of DNA concatemer formation with 5-cytosine methylated (5mC) and non-methylated segments. (b) Schematic of Alexa568-MBD to DNA concatemer. The entire molecule is stained using the green stain YOYO-1, and Alexa568-MBD binds to methylated stretches. c) Schematic of a device with two microchannel feeds (top and bottom) that are bridged by a nanochannel (inflowing arrows) containing an Alexa568-MBD labeled DNA concatemer. A shallow central shunt channel (outflowing arrows) allows the use of pressure-driven flow.

EXPERIMENTAL

MBD expressed from a vector, pET6HMBD (a gift from Sally H. Cross), was purified using Ni-NTA agarose beads followed by FPLC. MBD was then labeled using Alexa Fluor 568 according to standard protocols. Methylation of λ-DNA was performed using CpG Methyltransferase (M.SssI) from New England Biolabs (NEB) according to their standard protocol. Non-methylated and methylated λ-DNA were assembled into barcodes at a 1:1 ratio, by annealing and then slow cooling. A mixture of concatemers was obtained, with shorter length concatemers occurring at higher frequency. Alexa568-labeled MBD was then applied to the resulting mixture. In the final imaging solution, filtered 10 kDa MW PVP is used at 1% by weight, and the concentrations of DTT and PMSF were adjusted to 5 mM and 0.5 mM respectively.

Integrated nano/microfluidic channels were fabricated in fused silica using methods described elsewhere. The device layout is illustrated in Fig. 1, following a design principle demonstrated by Reisner et al., in which DNA is localized in the field of view through the combination of nanogrooves and a thin shunt layer that allows liquid to escape but traps DNA within the grooves. Note that the actual device is an array of such channels. The effective channel cross-section was 250×200 nm², and the shunt channel was 50 nm deep.

DNA was driven through both micro- and nanofluidic channels using total pressures of about 30 psi. Once molecules had been localized, the pressure was removed to recover an equilibrium configuration that is independent of liquid flow. After observation the channels were flushed using one-sided pressure application. Molecules were observed using an inverted fluorescence microscope. Simultaneous dual channel imaging of the green and red images is enabled by an image splitter and alternating illumination by two lasers.

RESULTS AND DISCUSSION

We tested false positive and false negative rates by incubating Alexa568-MBD with fully methylated and non-methylated DNA samples containing mostly λ-DNA monomers and halfmers. By counting each molecule as one event, we quantified the false positive rate as less than 1%, and the false negative rate as less than 0.6%. We then introduced DNA into nanochannel devices, as described in the methods section. We observe binding patterns as predicted by the random design of the DNA substrate. Fig. 2 shows a collection of observed fluorescence patterns from molecules incorporating MBD-labeled stretches. MBD binding occurred in continuous stretches, and not as dots along a line. Each λ-monomer has about 3000 CpG sites. Fig. 2(a) shows a λ-DNA trimer where the central monomer is methylated. A trimer with a terminal methylated monomer is shown in Fig. 2(b). A heterodimer is shown in Fig. 2(c). In all panels we note that both DNA and MBD are co-localized and stretched out. However, we typically observed that the length of a MBD-conjugated methylated λ-monomer is roughly a quarter to a third of that of a fully extended lambda monomer. This shortening upon MBD binding is attributable to the same MBD self-interaction that we reported for our controls. Tests on pure methylated DNA showed that methylation does not influence the mechanical properties in channels, at least on the scale seen when bound to MBD. Self-interaction also appeared to increase with the age of the Alexa568-MBD construct. A considerable fraction of molecules showed patterns indicating that the concatenation was terminated by half-mers, which is expected. For instance, Fig. 2(d) shows a methylated monomer flanked by two non-methylated half-mers. In some cases even shorter fragments are present, such as in Fig. 2(e), although a small probability exists that MBD is mechanically stripped from one end upon insertion into the nanochannel.
FIG. 2. (a) Fluorescence images of concatenated methylated and non-methylated λ-DNA labeled with Alexa568MBD (red) and YOYO-1 (green), stretched out in nanochannels. Within each panel colors are split for clarity; (left) YOYO-1 only (DNA), (center) composite, (right) Alexa568 only (Alexa568-MBD). Schematic drawings in each panel illustrate the spatial position of the Alexa Fluor 568 MBD and the length of the λ-DNA. The scale bar in panel (b) is 5 microns.

We can estimate the resolution of our technique from the extension of partially stretched MBD domains. We have observed that MBD-bound methylated λ-DNA monomers of 48.5 kbp length consistently stretch to a length of 1.3 microns. Since the resolution of our technique is approximately given by the diffraction limit, that would correspond to a resolution of about 10 kbp, which is comparable to the size of a human gene. Hence we believe that our technique is applicable to gene specific mapping of hyper- and hypo-methylated sites on genomic DNA.

CONCLUSION
We present a technique for the mapping of 5-methyl cytosine modification of CpG clusters in genomic length DNA with a resolution of about 10 kbp. We have demonstrated low false positive and negative rates, and have shown methylation patterns consistent with a prepared barcode pattern. We believe that the technique will be capable to derive gene-relevant data from single molecules.

ACKNOWLEDGEMENTS
We acknowledge funding from the National Institutes of Health (R21CA132075, R21HD065222). A portion of this research was conducted at the Center for Nanophase Materials Sciences, which is sponsored at Oak Ridge National Laboratory by the Scientific User Facilities Division, U.S. Department of Energy. This work was performed in part at the Cornell NanoScale Facility, a member of the NNIN, which is supported by the National Science Foundation (Grant ECS-0335765).

REFERENCES

CONTACT
*R. Riehn, tel: +1-919-513-0841; rriehn@ncsu.edu

1676