BIOINFORMATICS

Vol. 18 no. 11 2002
Pages 1427–1431

MethPrimer: designing primers for methylation PCRs

Long-Cheng Li and Rajvir Dahiya

Department of Urology, Veterans Affairs Medical Center, and University of California San Francisco, San Francisco, CA 94121, USA

Received on February 26, 2002; revised on April 25, 2002; accepted on May 1, 2002

ABSTRACT

Motivation: DNA methylation is an epigenetic mechanism of gene regulation. Bisulfite- conversion-based PCR methods, such as bisulfite sequencing PCR (BSP) and methylation specific PCR (MSP), remain the most commonly used techniques for methylation mapping. Existing primer design programs developed for standard PCR cannot handle primer design for bisulfite-conversion-based PCRs due to changes in DNA sequence context caused by bisulfite treatment and many special constraints both on the primers and the region to be amplified for such experiments. Therefore, the present study was designed to develop a program for such applications.

Results: MethPrimer, based on Primer3, is a program for designing PCR primers for methylation mapping. It first takes a DNA sequence as its input and searches the sequence for potential CpG islands. Primers are then picked around the predicted CpG islands or around regions specified by users. MethPrimer can design primers for BSP and MSP. Results of primer selection are delivered through a web browser in text and in graphic view.

Availability: MethPrimer is freely accessible at the following Web address http://itsa.ucsf.edu/~urolab/methprimer

Contact: longli@itsa.ucsf.edu; urologylab@aol.com

INTRODUCTION

Methylation of cytosine at CpG dinucleotides is a common feature of many higher eukaryotic genomes, and is most likely to be restricted to CpG dinucleotides, where both cytosine residues on the opposite strands are methylated (Paulsen and Ferguson-Smith, 2001). Compared with other dinucleotides, CpG dinucleotides are underrepresented in vertebrate DNAs except in clusters known as CpG islands. CpG islands are usually hypomethylated and often linked to promoter regions of genes (Antequera and Bird, 1993). The accepted definition of CpG island is regions of DNA greater than 200 bp, with a guanine/cytosine content above 0.5 and an observed or an expected presence of CpG above 0.6 (Gardiner-Garden and Frommer, 1987). Approximately 60% of all human genes contain at their 5′ ends CpG islands (Larsen *et al.*, 1992). Cytosine methylation is recognized as an important mechanism of epigenetic regulation of genomic function and plays important roles in diverse biological processes including embryogenesis (Monk *et al.*, 1987), genomic imprinting (Singer-Sam and Riggs, 1993), X-chromosome inactivation (Li *et al.*, 1993), and cancer (Baylin *et al.*, 1998; Li *et al.*, 2000a,b, 2001; Nojima *et al.*, 2001). Mapping of methylation patterns in CpG islands has become an important tool for understanding both normal and pathologic gene expression events.

Numerous techniques have been invented for the mapping of cytosine methylation, among them, bisulfite-conversion-based methods are probably the most widely used in recent years because they permit the rapid identification of methylated cytosine (5mC) in any sequence context. The bisulfite reaction was first described in early 1970s (Hayatsu *et al.*, 1970; Shapiro and Weisgras, 1970) and was used by Frommer *et al.* (1992) in 1992 to distinguish between cytosine and 5mC in DNA. In this reaction, DNA is first treated with sodium bisulfite to convert cytosine residues to uracil in single-stranded DNA, under conditions whereby 5mC remains essentially non-reactive. The DNA sequence under investigation is then amplified by PCR with primers specific for bisulfite-modified DNA (Clark *et al.*, 1994). Since the first description of bisulfite reaction in the application of studying 5mC, many methods based on the same principle have been developed including bisulfite-sequencing PCR (BSP; Clark *et al.*, 1994), methylation-specific PCR (MSP; Herman *et al.*, 1996), a quantitative method called COBRA (Xiong and Laird, 1997), and methylation-sensitive single nucleotide primer extension (MS-SNuPE; Gonzalgo and Jones, 1997), with the first two being the most commonly used (Figure 1). All methods share the same procedure of modifying DNA with sodium bisulfite as the first step and subsequently PCR amplification with primers specific for modified DNA.

For the successful implementation of these methods, the most critical step is the design of primers for the

To whom correspondence should be addressed.

© Oxford University Press 2002

Geneoscopy Exhibit 1059, Page 1

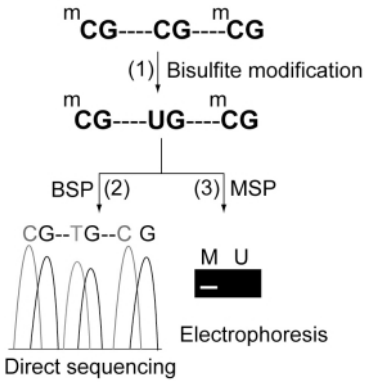

Fig. 1. Schematic representation of bisulfite methylation mapping techniques. (1) DNA is modified with sodium bisulfite to convert unmethylated Cs to Us, and subjected to PCR amplification. (2) BSP is carried out using primers containing no CpG sites and the resulted PCR products are sequenced. (3) For MSP, two sets of primers are used to amplify methylated and unmethylated allele respectively. MSP products are separated by gel electrophoresis. m: 5mC; M: methylated; U: unmethylated.

modified DNA (Clark *et al.*, 1994). Existing programs for standard PCR cannot handle primer design for bisulfite-conversion-based PCRs, because primers for such PCRs must be picked from the bisulfite-modified DNA sequence and extra constraints are needed for primer selection in addition to those required for standard PCR, such as no CpG site should occur in the primer sequence for BSP, while for MSP, at least one CpG site should appear in the primer sequence. These extra constraints make it impossible for existing programs to pick primers even if input DNA sequences are edited before fed to a primer design program.

In this paper, we introduce a new program called MethPrimer that was based on the well-known primer design program Primer3 (Rozen and Skaletsky, 2000) and was developed specifically for bisulfite-conversion-based PCRs. MethPrimer integrates CpG island prediction into primer design and is able to design primers for BSP and MSP. It will also be able to pick primers for COBRA and MS-SnuPE with minor modifications to the program.

SYSTEM AND METHODS

Sequence conversion

To design primers for the bisulfite-modified DNA, users are required to input an original sequence and no sequence conversion and editing is required. The program will generate two versions of modified sequences internally, one is the bisulfite-modified and methylated sequence in which all 'C's except 5mC are converted to 'T's; the other is the bisulfite-modified and unmethylated sequence in which all 'C's as well as 5mCs are converted to 'T's.

CpG island prediction

CpG islands are predicted using a simple sliding window algorithm. The algorithm slides across the sequence at a specified shift value examining the GC content and the ratio observed/expected (Obs/Exp) in a windows size defined by users (Gardiner-Garden and Frommer, 1987). By default a CpG island is defined as a DNA stretch at least 200 bp-long with a GC content 50% and an Obs/Exp ratio of CpG dinucleotides 0.6 (Gardiner-Garden and Frommer, 1987). If users choose to pick primers using the predicted CpG island as target region, the following rules are applied.

(1) If more than one island is found, any of the predicted islands will be the target region for amplification.

(2) If a CpG island size is smaller than minimal product size, the primer pair should span the whole island.

(3) If a CpG island size is greater than maximal product size, the primer pair should be within the island.

(4) If a CpG island size is between the minimal and maximal product size, primer pair should cover at least two thirds of the island.

General primer selection criteria

For standard PCR, the important parameters to be considered when selecting primers are the ability of the primers to form a stable duplex with the specific site on the target DNA and no duplex formation with another primer molecule or no hybridization at any other target site (Rychlik, 2000). The same is true of bisulfite-conversion-based PCRs. To assess these parameters along with other common criteria for standard PCR, MethPrimer uses Primer3's algorithms and code to compute self annealing, self-end annealing, pair complementarity, GC content, and melting temperature ((T_m); Rozen and Skaletsky, 2000). Additional constraints are also applied, such as T_m difference between left and right primer, the maximal allowable length of a mononucleotide repeat in primer sequence. In the last case, 'T' repeat is treated differentially as other nucleotides, since all non-5mCs are treated as 'T' during primer selection. By default, a maximal number of eight consecutive 'T's is allowed, while five is the default value for other nucleotides.

Primer selection criteria specific for bisulfite-conversion-based pcrs

Despite variations among bisulfite-conversion-based methylation mapping methods, they all share the same procedure of modifying DNA with sodium bisulfite as the first step and subsequently PCR amplification with primers specific for modified DNA. Incomplete bisulfite modification of

Geneoscopy Exhibit 1059, Page 2

DNA is sometimes a concern (Clark *et al.*, 1994) and results in high representation of methylation levels in samples studied. Thus, selective amplification of only completely modified DNA is important. To bias bisulfite-modified DNA against unmodified or incompletely modified DNA, primers should be picked from a region that have enough number of non-CpG 'C's in the original sequence (Herman *et al.*, 1996). Primers with more 'C's will be preferred by receiving higher weighing scores.

Primer selection for BSP

For BSP, DNA is first modified by treatment with sodium bisulfite to convert all 'C's to uracil residues except 5mCs, then PCR is performed to amplify the bisulfite-modified DNA. The resulted PCR product can be used in the following three ways: (1) cloning and subsequent sequencing to study the methylation status of individual molecules; (2) direct sequencing to examine strand-specific methylation for the population of molecules; (3) digestion with methylation-sensitive restriction enzyme to examine methylation of a specific CpG site. Besides meeting those common criteria as stated above, the following constraints are enforced.

(1) Primers should not contain any CpG sites within their sequence to avoid discrimination against methylated or unmethylated DNA.

(2) A good primer pair should span a maximal number of CpG sites. The more CpG sites a pair spans, the higher weighing scores it will be assigned. This parameter forces the program to return primers spanning CpG-rich regions in which users are most interested.

Primer selection for MSP

For MSP experimentation, two pairs of primers are needed, one of them is specific for modified and methylated DNA (M pair), and the other for modified and unmethylated DNA (U pair). For each sample to be studied, two PCRs are performed with each pair of primers. Amplification with M pair indicates methylation of CpG site(s) within the primer sequences, U pair no methylation, and both pairs partial methylation. To satisfy such requirements, the following constraints are applied to primer selection for MSP.

(1) For maximal discrimination between methylated and unmethylated allele, primers should contain at least one CpG site at the most 3′-end. Users can specify the maximal distance from the 'C' in the CpG dinucleotide to the most 3′-end of the primer. By default this value is set to 3, which means among the last three bases in the primer at least one of them should be a CpG 'C'.

(2) Other than CpG site(s) at the most 3′ end, more CpG sites in primer sequence are preferred.

(3) Primers in M pair and U pair should contain the same CpG sites within their sequence. For example, if the sequence for forward primer in M pair is ATTTAGTTTCGTTTAAGGTTCGA, the forward primer in U pair must also contain the two CpG sites (underlined) as in M pair, but with 5mC replaced by 'T'. This constraint is necessary because nearby CpG sites are not always equally methylated (Li *et al.*, 2000a,b). If two pairs of primers do not anneal to the same CpG sites, PCR results from the primers may not truly reflect the DNA methylation status of the sample studied. However, primers in M pair and U pair may not span exactly the same sequence and may vary in start position or length. Usually primers in U pair are longer than those in M pair. This is due to the effect of the T_m constraint and the constraint stated below. Since replacement of the 5mCs by 'T' in U pair results in lower GC content, thus lower T_m, compared with that of M, provided that M pair and U pair shared the same sequences, thus, for primers in U pair to achieve optimal T_m and to match the T_m value of primers in M pair, longer primers for U pair will be selected automatically by the program.

(4) Two sets of primers should preferably have similar product T_m values. This constraint will produce primers with them two PCR reactions can be carried out in a single PCR machine using the same cycling conditions. By default, the difference is set to 5° C.

Visualization of primer selection results

To visually display results for CpG island prediction and primer selection, the Perl GD module (http://stein.cshl.org/WWW/software/GD) is used to generate a PNG image for each input sequence. An image map is also generated and embedded into HTML code using a Perl script for each image to display text explanation as tool tips for each element in the image.

Implementation

MethPrimer was developed on a Linux platform and written mainly in C language. The web interface to MethPrimer was written in Perl. MethPrimer is hosted on a Linux server and is accessible via a web browser from any computer that has access to the server via Internet or intranet. Performance assessment was carried out on a PC DELL, 400 MHz workstation running Mandrake Linux 8.2.

Figure 2 shows a typical session of primer searching for MSP. We used the human estrogen receptor (ER) β promoter sequence (GenBank accession no. AF191544)

Geneoscopy Exhibit 1059, Page 3

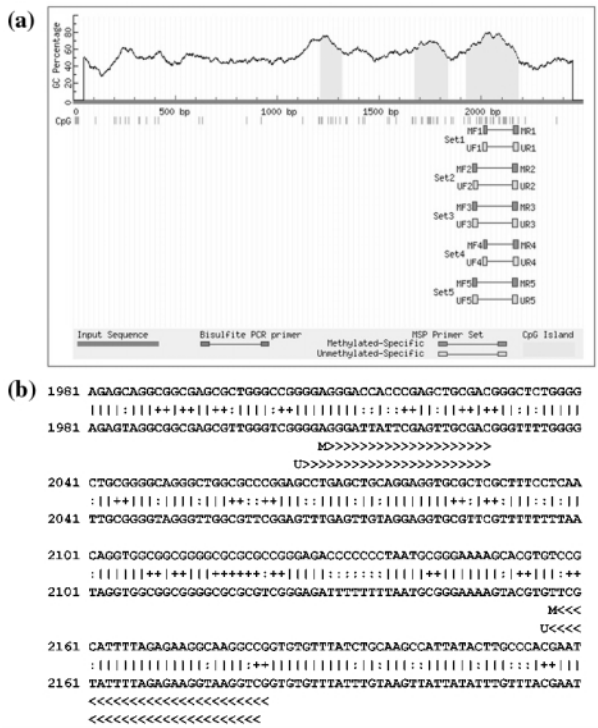

Fig. 2. Visualization of primer selection results for MSP. A 1655-bp estrogen receptor β promoter sequence (GenBank Accession: AF191544) was input to the program using the web interface. CpG island prediction for primer selection was used as an input parameter. All other parameters were default values. (a) Graphic view showing primers and sequence features such as GC percent, CpG islands, and CpG site. (b) Text view showing sequence alignment and location of primers.

as the test sequence which is known to contain CpG islands in its sequence (Li *et al.*, 2000a,b). Five sets of MSP primers were returned, each of them consists of two pairs of primers; one was for the amplification of methylated template (methylation-specific), the other unmethylated template (unmethylation-specific). In this example, the program was forced to pick primers around predicted CpG island regions. All other parameters used were default values. As expected, all pairs either spanned or overlapped a CpG island. In the text view (Figure 2(b)), the original sequence and the bisulfite-modified sequence are aligned together. 5mCs and non-5mCs are denoted differently to aid easy visualization.

RESULTS AND DISCUSSION

To estimate the performance and scalability of MethPrimer, we tested the program with random DNA sequences ranging from 10 to 100 kb in length for tasks of designing BSP and MSP primers. The execution time

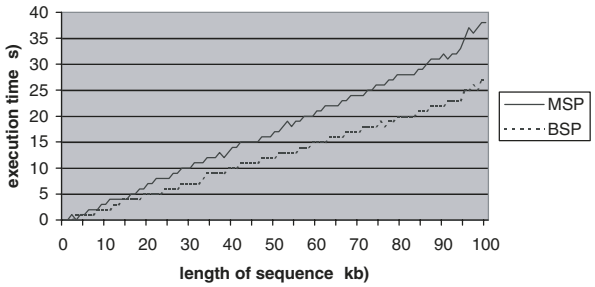

Fig. 3. Performance of MethPrimer. Both BSP and MSP primer designing tasks were assessed with random DNA sequences of varying length as input sequences. A DELL 400 MHz workstation was used for the assessment. The calculation included CpG island prediction.

is linear with respect to the length of the query sequence (Figure 3). It takes more time to process sequences with the same length for MSP than for BSP. Considering the extra calculations needed for M pair and U pair matching in designing MSP primers, this increase is reasonable.

Primer design is crucial for successful PCR amplification of bisulfite-modified DNA. Bisulfite reaction not only causes the expected conversion of cytosines to uracils, resulting in universal low GC content and long stretches of Ts in the sequence, but also causes undesired DNA strand breakage. Loss of DNA during the subsequent purification step is another concern especially when studying microdissected DNA samples. All these factors pose challenges to downstream PCR applications and should be taken into consideration when designing primers for such PCRs.

Usually, a product size greater than 300 bp will be difficult to amplify from bisulfite-modified DNA template (unpublished observations); hence, we set the default product size range as 100–300 bp with 200 bp as the optimal product size. Another rule that differs from standard PCR is primer length. Bisulfite-conversion-based PCRs generally require longer primers. Primers with a length of approximately 30 bp usually yield successful results (Clark *et al.*, 1994). The reason is that bisulfite modification decreases considerably GC content of DNA templates and produces long stretches of 'T's in the sequence, making it difficult to pick primers with acceptable T_m values or stability. On the other hand, in order to discriminate modified DNA and unmodified or incompletely modified DNA, enough number of 'C's is required in primers, which makes the job of picking stable primer more demanding. Thus, to achieve better duplex stability, choosing longer primer is necessary as T_m of DNA also depends on its length (Rychlik, 2000). In practice, size of primers for such PCRs usually ranges from 20 to 30 bp (Herman *et al.*, 1996; Graff *et al.*, 1997;

Geneoscopy Exhibit 1059, Page 4

Li *et al.*, 2000a,b). In MethPrimer, 20–30 bp is set as the default range of primer size with 25 bp being the optimal size.

MethPrimer does not implement checks for repetitive elements or vector sequences in input sequences before picking primers. For methylation mapping, researchers usually focus on well-characterized sequences or well-defined regions in a sequence. Most of such sequences or regions are promoter sequences. Therefore, contamination by vector sequences is not a concern; on the other hand, because input sequences are converted internally by the program to mimic bisulfite modification, repetitive elements may be no longer repetitive, thus having no impact on primer quality even the primer is located at a region that was repetitive before sequence conversion.

In summary, MethPrimer was developed specifically for designing primers for a number of bisulfite-conversion-based methylation PCRs including BSP and MSP. Further development is undergoing to make it supportive of primer design for other bisulfite-conversion-based PCRs such as COBRA and MS-SNuPE.

REFERENCES

Antequera,F. and Bird,A. (1993) CpG islands. *Exs*, **64**, 169–185.

Baylin,S.B., Herman,J.G., Graff,J.R., Vertino,P.M. and Issa,J.P. (1998) Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer. Res.*, **72**, 141–196.

Clark,S.J., Harrison,J., Paul,C.L. and Frommer,M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic. Acids Res.*, **22**, 2990–2997.

Frommer,M., McDonald,L.E., Millar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L. and Paul,C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad. Sci. USA*, **89**, 1827–1831.

Gardiner-Garden,M. and Frommer,M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.*, **196**, 261–282.

Gonzalgo,M.L. and Jones,P.A. (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.*, **25**, 2529–2531.

Graff,J.R., Herman,J.G., Myohanen,S., Baylin,S.B. and Vertino,P.M. (1997) Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation. *J. Biol. Chem.*, **272**, 22322–22329.

Hayatsu,H., Wataya,Y., Kai,K. and Iida,S. (1970) Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*, **9**, 2858–2865.

Herman,J.G., Graff,J.R., Myohanen,S., Nelkin,B.D. and Baylin,S.B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad. Sci. USA*, **93**, 9821–9826.

Larsen,F., Gundersen,G., Lopez,R. and Prydz,H. (1992) CpG islands as gene markers in the human genome. *Genomics*, **13**, 1095–1107.

Li,E., Beard,C. and Jaenisch,R. (1993) Role for DNA methylation in genomic imprinting. *Nature*, **366**, 362–365.

Li,L.C., Chui,R., Nakajima,K., Oh,B.R., Au,H.C. and Dahiya,R. (2000a) Frequent methylation of estrogen receptor in prostate cancer: correlation with tumor progression. *Cancer Res.*, **60**, 702–706.

Li,L.C., Yeh,C.C., Nojima,D. and Dahiya,R. (2000b) Cloning and characterization of human estrogen receptor beta promoter. *Biochem. Biophys. Res. Commun.*, **275**, 682–689.

Li,L.C., Zhao,H., Nakajima,K., Oh,B.R., Filho,L.A., Carroll,P. and Dahiya,R. (2001) Methylation of the E-cadherin gene promoter correlates with progression of prostate cancer. *J. Urol.*, **166**, 705–709.

Monk,M., Boubelik,M. and Lehnert,S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, **99**, 371–382.

Nojima,D., Nakajima,K., Li,L.C., Franks,J., Ribeiro-Filho,L., Ishii,N. and Dahiya,R. (2001) CpG methylation of promoter region inactivates E-cadherin gene in renal cell carcinoma. *Mol. Carcinog.*, **32**, 19–27.

Paulsen,M. and Ferguson-Smith,A.C. (2001) DNA methylation in genomic imprinting, development, and disease. *J. Pathol.*, **195**, 97–110.

Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.

Rychlik,W. (2000) Primer selection and design for polymerase chain reaction. In Rapley,R. (ed.), *The Nucleic Acid Protocols Handbook*. Humana Press, Totowa, NJ, pp. 581–588.

Shapiro,R. and Weisgras,J.M. (1970) Bisulfite-catalyzed transamination of cytosine and cytidine. *Biochem. Biophys. Res. Commun.*, **40**, 839–843.

Singer-Sam,J. and Riggs,A.D. (1993) X chromosome inactivation and DNA methylation. *Exs*, **64**, 358–384.

Xiong,Z. and Laird,P.W. (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic. Acids Res.*, **25**, 2532–2534.

Geneoscopy Exhibit 1059, Page 5