

Methylation-specific PCR: four steps in primer design

Review Article

Radoslav S. Davidović*, Ana M. Božović, Vesna Lj. Mandušić, Milena M. Krajnović

*Institute of Nuclear Sciences "Vinča", University of Belgrade,
11001 Belgrade, Serbia*

Received 3 October 2013; Accepted 2 December 2013

Abstract: Methylation-specific PCR (MSP) is still the method of choice for a single gene methylation study. The proper design of the primer pairs is a prerequisite for obtaining reliable PCR results. Despite numerous protocols describing the rules for MSP primer design, none of them provide a comprehensive approach to the problem. Our aim was to depict a workflow for the primer design that is concise and easy to follow. In order to achieve this goal, adequate tools for promoter sequence retrieval, MSP primer design and subsequent in silico analysis are presented and discussed. Furthermore, a few instructive examples regarding a good versus a poor primer design are provided. Finally, primer design is demonstrated according to the proposed workflow. This article aims to provide researchers, interested in a single gene methylation studies, with useful information regarding successful primer design.

Keywords: *Methylation-specific PCR • Primer Design • Database • Software for MSP Primer Design • Transcriptional Start Site*

© Versita Sp. z o.o.

1. Introduction

Epigenetic regulation of gene expression has an essential impact on the growth and development of an organism [1]. This regulation occurs at many levels [1,2], of which methylation of cytosine at CpG dinucleotides is the most studied aspect of the phenomenon. It is estimated that approximately 20% of CpG's in the human genome are located inside so-called CpG islands [3]. A CpG island is a stretch of DNA with an overrepresentation of CpGs [4,5], spanning more than 200 bases with a G+C content accounting for at least 50%, and a ratio of observed versus statistically expected CpG frequencies of at least 0.6 [6]. Approximately 60% of human genes contain CpG islands in the promoter region [5], but the majority of these islands are unmethylated in most tissues. These two factors impact on the regulatory function of the islands. Methylation of cytosine inside CpG

islands has an inhibitory effect on transcriptional activity [4,5,7], suggesting that regulation of DNA methylation has significant importance in organism homeostasis. Therefore, unsurprisingly, much effort has been made to better understand DNA methylation pattern formation and its changes during physiological and pathological processes.

Aberrant methylation contributes to the development of numerous diseases and pathological processes such as carcinogenesis and imprinting disorders [8]. For example, hypermethylation of tumor suppressor gene promoters is a well characterized early event in tumorigenesis [9,10]; however, loss of imprinting, as a cause of particular syndromes, represents loss of methylation on one of the parental alleles [8]. These two events are sufficient to highlight the importance of DNA methylation studies. In order to assess the methylation status of the gene of interest, numerous methods are available. These methods rely on two strategies:

* E-mail: radoslav@vinca.rs

i) cleavage of DNA with methylation specific enzymes, and ii) bisulfite modification of genomic DNA (gDNA) [11,12]. Our focus is on the latter one.

Methylation-specific PCR (MSP) is the most convenient method for studying the methylation status of promoter regions of individual genes [13]. Although the method is considered obsolete by some authors, its cost effectiveness, sensitivity and rapid deployment in the laboratory make MSP the method of choice in single gene methylation studies. MSP is based on prior bisulfite modification of the DNA sample. After the treatment, unmethylated cytosines are converted to uracils, while 5-methylcytosines remain unaltered; thus as result, DNA strands are no longer complementary to each other (Figure 1). Further on, bisulfite modified DNA is subjected to PCR amplification using two primer pairs, of which one primer pair recognizes methylated, and another pair recognizes unmethylated alleles. As a result, well-optimized PCR reaction will provide detection of a single methylated allele among one thousand unmethylated ones. High sensitivity of the reaction enables potential application of MSP-based methods for diagnostic purposes [14,15]. However, optimization of MSP reaction can be quite challenging,

therefore, primer design is an absolutely essential step. There are numerous parameters that must be considered prior to and during the MSP primer design. Our aim is to provide researchers with a concise and easy to follow protocol for the MSP primer design.

2. Workflow and Databases

The whole process of MSP primer design can be divided into four steps: i) promoter sequence retrieval, ii) identifying the part of the promoter where the primers should be located, iii) selection of an appropriate software for primer design, and iv) selection of the primer pair with the best score according to in silico analysis (Figure 2). There are a considerable number of publications and internet resources that describe the general rules and software for MSP primer design. However, little attention has been paid to the first step, i.e. searching for the promoter sequence. Although databases like ENSEMBL (<http://www.ensembl.org/index.html>) and NCBI (<http://www.ncbi.nlm.nih.gov/>) contain information about gene and mRNA sequences, retrieval of the promoter sequence is not always simple and/or possible. On the

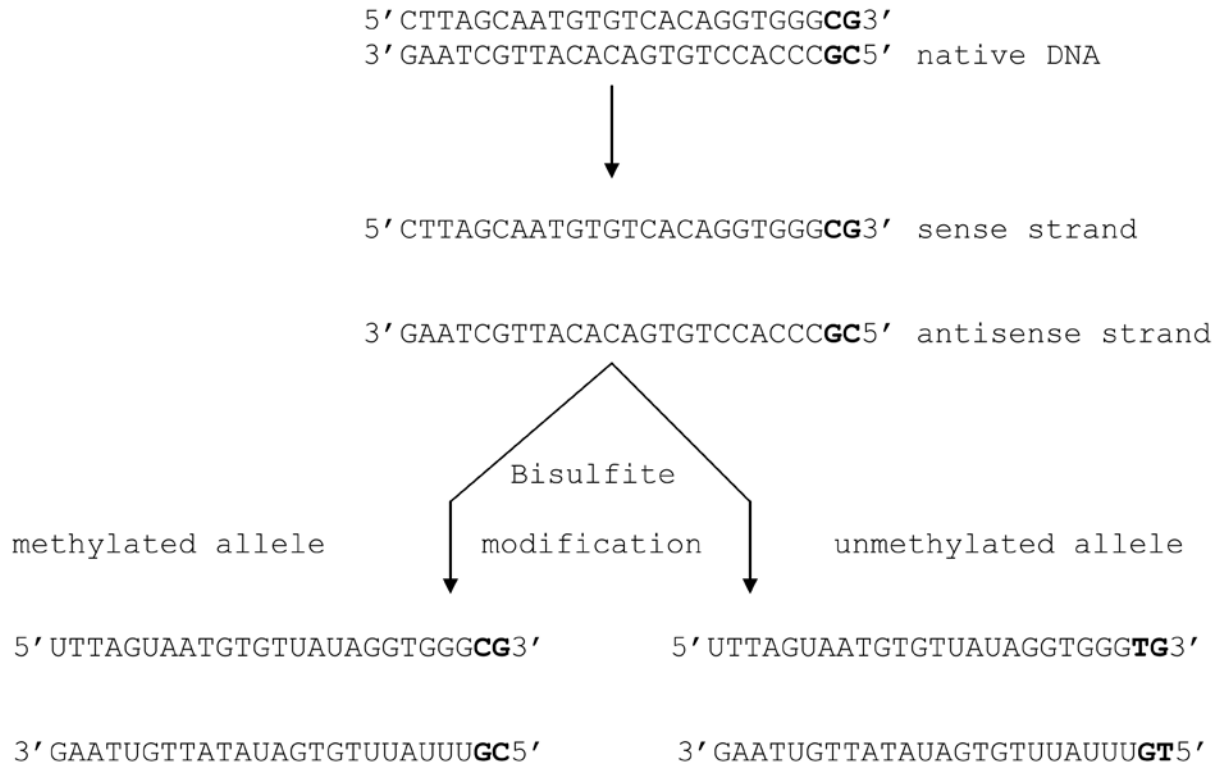


Figure 1. MSP is based on bisulfite modification of native DNA. Prior to the bisulfite treatment, DNA is denaturated. After denaturation, single stranded DNA is subject to the modification process. Methylated cytosines remain unaltered while unmethylated cytosines are converted to uracils. During PCR, uracils are replaced by thymines. As a consequence of bisulfite treatment, DNA strands are no longer complementary. CpG sites are denoted with bold font.

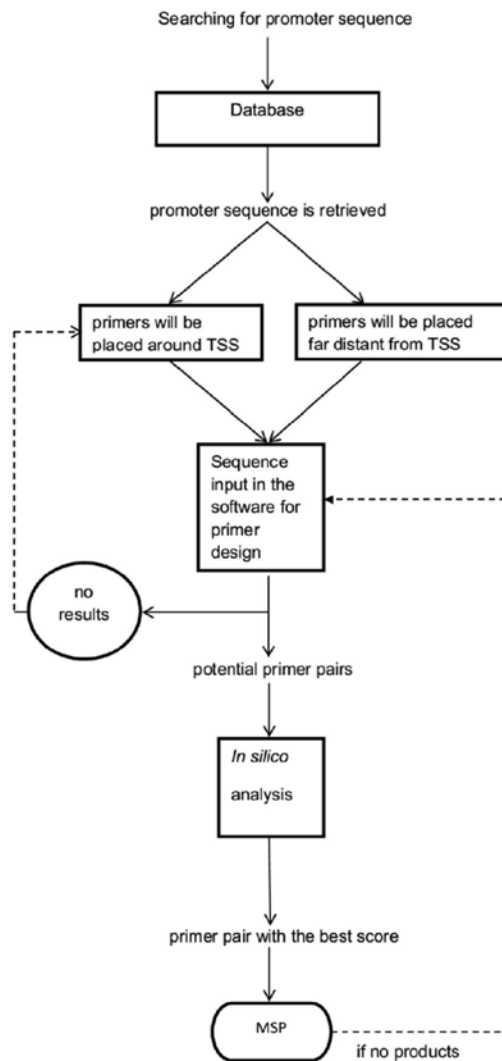


Figure 2. Four steps in the primer design presented as an algorithm. Promoter sequence retrieval is an essential step in the whole process. When the position of TSS is known then one should choose promoter region where primers will be located. A desirable sequence serves as a template for primer design with the help of appropriate software. As a result, several potential primer pairs are available to the user. In silico analysis helps us to choose the best one. However, the best primer pair does not have to perform well in the test tube. If this is the case, new primer pairs should be designed.

other hand, use of promoter prediction software is not recommended because of the necessity for the exact promoter location. Fortunately, substantial development of technology during the last decade has improved promoter characterization on the genomic level. Consequently, reliable databases containing promoter sequences with good coverage are now available. To this end, we will discuss two databases: i) DBTSS (Database of Transcriptional Start Sites), and ii) EPD (Eukaryotic Promoter Database). The reason we are

discussing these two particular databases is because we believe they are user-friendly and have excellent coverage. It is noteworthy that the primary intention of DBTSS and EPD is not to provide researchers with promoter sequences, but rather to facilitate functional studies.

DBTSS contains information about transcriptional start site (TSS) positions for 28 human and 4 mouse tissues/cell cultures [16]. Moreover, information about the TSS position for the same tissue in different conditions is provided (normoxia/hypoxia for example) as well. Clearly, the authors' intention was to make DBTSS a database that will provide a dynamic overview of transcriptional regulation. To create such a huge database, the authors used the so-called TSS-seq technique [17]. This technique is the combination of the oligo-capping method [18] followed by the use of an Illumina massively parallel sequencing platform.

Retrieval of the TSS position is straightforward. After selection of the species and cell/tissue type, one selects the gene of interest. Information about the precise gene chromosome position and annotation in Entrez and Ensemble databases, as well as the graphic presentation with information regarding the CpG island position and the presence of SNPs and the TSS position, is available. The user has an option to retrieve FASTA format of promoter sequence with the TSS position highlighted. The number of bases upstream and downstream of the TSS is selected by the user. The database is accessible at <http://dbtss.hgc.jp/>.

The EPD is the database founded more than 25 years ago [19]. During its long history, the database went through a significant number of changes. The most important change, however, is the way in which information about promoters is collected. Development of molecular biology methods had a great impact on the EPD organization. Instead of the time-consuming and tedious process of promoter sequence extraction from appropriate publications, methods for TSS identification on a genomic level are used. Today, the database contains, as stated by the authors, 25988 promoters for the human collection, 9773 promoters for the mouse collection, 11389 promoters for the *D. melanogaster* collection and 11719 promoters for the zebrafish collection. In order to get information about TSS, the user should select the gene as well as the appropriate species database. Users can also find references containing detailed explanation regarding database structure and organization, as well as the user manual and links to the other databases. It is important to emphasize that EPD takes into consideration only polymerase II binding sites. In addition, authors give the definition of the promoter as a TSS. Based on this, promoters are divided into

three classes: i) single TSS site, ii) multiple TSS sites, and iii) initiation region. The database website can be accessed at <http://epd.vital-it.ch/>.

It is important to highlight two facts regarding the presented databases: i) the detailed analysis of the database potential and performance are beyond the scope of the paper, and ii) position of the TSS for the same gene may vary between different databases. However, these differences are not significant and they do not affect later steps in the process.

3. Where should primers be located?

After the promoter sequence retrieval, the next logical step is to define the part of promoter where primers should be located. Although this step may seem trivial, it could affect the results of the experiment. Routinely, the primers are placed in the vicinity of the TSS. The rationale behind this is that methylation in the region around the TSS must have a profound effect on the gene transcription. This solution frequently works in laboratory practice. However, promoter regions essential for transcriptional activity can be very distant from the TSS. Thus, it is advisable to search the literature data about the functional studies of the promoter of interest in order to find a sequence where methylation has biological meaning. Consequently, it is useful to know if there are sequences inside the CpG island to which transcription factors (TFs) bind. Methylation of Cs inside the CpG island can prevent binding of TFs to the promoter of the targeted gene, leading to decreased expression [20]. In this respect, interesting results were published by Zhu *et al.* [21]. These authors defined three so-called methylation “hot spots” in the ER β gene promoter ON. “Hot spots” are located in the consensus sequences containing transcription factor binding sites. These findings [23] were successfully used by Bozovic *et al* [22] to design primers for ER β methylation study (Figure 3A).

During the process of primer design, researchers sometimes lose sight of the context in which the target gene is placed, which may provide misleading information regarding the promoter methylation status. An example is depicted in Figure 4 where high sequence identity between part of the PTEN gene promoter and PTENP1 pseudogene is displayed. Zysman and Bapat [23] showed that some previously published results regarding the PTEN gene methylation status in various cancer types were false positive. As a matter of fact, primers used in these studies were located in the region having a high level of homology with PTEN

pseudogene (PTENP1). Indeed, Hesson and coworkers [24] confirmed the Zysman and Bapat conclusion. Additionally, they pointed out that assessment of PTEN methylation is further complicated by the fact that the PTEN CpG island is shared with KLLN gene. The results from these two publications suggest that carefully chosen primer sequences are a requirement for a successful methylation study. Furthermore, this example shows that primer design for methylation analysis of some genes is very demanding due to limitations imposed by the structural organization of the promoter and genome itself.

Based on the foregoing discussion, it is possible to draw a single conclusion that before primer placement, some strategic points have to be known. TSS is a strategic point which significantly influences the decision where primers should be located. As described above, DBTSS and EPD can provide information about the TSS position. Therefore, it is highly advisable to start the search for primer placement in the vicinity of the TSS. The region that spans 1000 bp upstream and up to 500 bp downstream from the TSS is a good starting point. However, the search for CpG islands may involve a considerably wider area (from -5 kb to +5 kb) [25]. Interestingly, methylation of the distal promoter region, but no core promoter, may play an important role in the target gene expression in some cases [26-28]. TF binding sites can also be strategic points that further specify the region where primers could be set. For example, four Sp1 transcription binding sites are located in the region close to the TSS of PADI2 gene [29], thus primers should be located as close as possible to these sites (Figure 3B).

Strategic points are important because they indicate where primers should be located, but some other factors have to be considered as well before making the final decision. These factors determine “restrictive points”, i.e., sequences where primers should not be located. For example, it is recommended to avoid sequences containing common SNPs [30]. Also, promoter regions having a high level of homology with other sequences in the genome should also be avoided.

4. Rules and software for primer design

There is an abundance of available free software for the MSP primer design on the internet (Table 1). Before presentation of the appropriate software, rules for the MSP primer design will be listed and explained. The general rules for PCR primer design [31,32] are applicable in this case, but some additional rules

emerged as a consequence of the bisulfite treatment. Initially, when primer design for bisulfite modified DNA is considered, the following two conditions must be remembered: i) primers must discriminate between

A.
 GACTGGGGCTCGGGGAGCGCCATCTGTGCGCCACTACTCTGTGGTGGACCA
 GGAGT**CGGTTC**CGAGGTGCTCCCACTTAGAGTCA**CGCGCGCTCGGGCGTTC**
 CTGAGAC**CGT**CGGGTCCCTGGCT**CGT**CAC**CGT**GGCTCAGGCACTACTCCCTC
 TACCCTCCTCT**CGT**TCTTAAAGGAAGAAGGGCTTACT**CGT**TAAG**CGT**TGTGAT
 CTT⁺TTCAGTTTCTCCAGCTGCTGGCTTTTGGACACCCTCC**CGC**AGGAGG
CAGTTGCAAG**CGCG**GAGGCTCGAGAAATAACTGCCTTTGAAACT**TCAGGGCG**
AAGAGCAGCGCGCAGCGCTGGGCGGGGAGGACCACCGAGCT**CGCAGCGG**
 CTCTGGGCTG**CGGGC**AGGCTGG**CGCCCG**AGCTGAGCT**CGCAGGAGTGC**
GCTCGCTTTCCTCAACAGTGG**CGCGGGCGCGCGC**CGGAGACCCCTAA
 TCGGGAAAAGCACGTGTCCGCATTATAGAGAAGCAAGG**CGGT**GTGTTTACTG
 CAAGGTAAG**CGCCCTCGCT**CGAGGTG

B.
 AAAGTGGTCCCTGATGACT**CGCCACGACCTCA****CGGGCGGGACGAGT****TAGGGG**
CGGGAC**AC**CGTTG**ACG**GACAGCCCG**CGACCCACTCGC**AGAG**ACTGGGGCGG**
CGCGCCCGCCACGGGCGCTGGATAAGGCT**CGCGCGGGCGTGGGGCG**⁺
 CAGGCTGCTGGAGAAGCGCACCTGCTGACAGTCTCC**CGCGCCCG**CGACCA
 CGGAGCGGGGACTCGGGCGGGAGGATGCT**CGCGAGCGGACCGT****CGGCT**
 GCAGTAC**CGGAGCGCGT**GGAGGCGGTGTACG

Figure 3. A. Sequence of ERβ gene promoter of 574 bp (-222 / +352). "Centers" of methylation are highlighted in gray. The designed primers for MSP are underlined; it can be seen that they are located in the second and the third methylation center. B. Promoter sequence of PADI2 gene. Sp1 binding sites are highlighted in gray. Hypothetical primer pair is designed in such a way that forward primer spans two Sp1 binding sites; transcription start site is designated as +1; start codon is italicized; CpG sites are bolded. The sequences displayed in the Figure are not bisulfite modified ones.

Table 1. Tools for MSP primer design and their features.

Tool	Description	Link
MSPprimer	Web-based tool for MSP, nested MSP and BSP primer design. Contains SDSS (Specificity-Determining SubSequence) algorithm that is responsible for higher primer specificity [50]. The program requires sequence of native DNA as input. User can adjust number of CpGs, Tm values, primer and product length. Also there is an option to set the position of TSS in the input sequence. The program can be used freely, but registration is required.	http://www.mspprimer.org/cgi-bin/design.cgi
MethMarker	Methmarker is an all-round program for methylation studies design. It supports primer design for the following methods: COBRA, bidulfite SnuPE, MSP, MethyLight, bisulfite pyrosequencing, MeDIP qPCR. An additional purpose of the tool is to support epigenetic marker optimization. MethMarker is free to download and use.	http://methmarker.mpi-inf.mpg.de/
Beacon designer	Software solution that can be used for designing primers and probes for quantitative MSP (MethyLight) studies. Provides a large number of options regarding primer design and analysis of primer properties. It is possible to choose which strand (sense/antisense) will serve as a template. Also, there is an option to set thermodynamic properties of the primers. The main disadvantage is that the software is shareware. After a free trial period expiration, only demo mode can be run.	http://www.premierbiosoft.com/molecular_beacons/
Primo MSP	Web-based tool for MSP and BSP primer design. Does not contain some important features such as adjustable number of non-CpG Cs and CpG island searches. It seems that the application is not optimized enough. Some results do not make sense (obtained product length is 1 bp, for example).	http://www.changbioscience.com/primo/primom.html

native and bisulfite modified DNA, and ii) primers must discriminate between methylated and unmethylated alleles. In order to meet these requirements, primers should contain: i) at least one CpG near its 3'end to

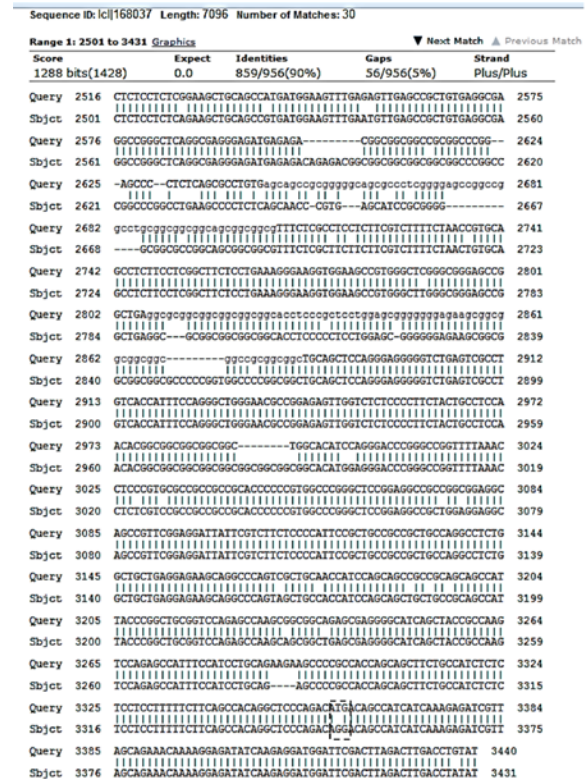


Figure 4. Alignment of PTEN (ENSG00000171862) gene and PTENP1 (ENSG00000237984) pseudogene using BLAST tool (<http://blast.ncbi.nlm.nih.gov/>). The upper sequence represents PTEN and the lower one represents pseudogene PTENP1. Translational start site of PTEN is framed by dashed lines. Transcriptional start site is located 1034 nucleotides upstream from the start codon and is not shown. The sequences show 90% identity.

discriminate between methylated and unmethylated sequence, and ii) non-CpG C's, the more the better, to favor amplification of the bisulfite modified sequence. Equally important, in order to assess methylation status of the gene, the primers for the methylated and unmethylated allele should contain the same CpGs in the sequence. This does not mean that the primers have to span the same stretch of DNA. On the contrary, it is preferred that the unmethylated primer is longer than the appropriate methylated counterpart. The bisulfite modified DNA contains long stretches of T's. Consequently, the unmethylated primer is AT rich and has low T_m. This can be compensated for by the longer sequence. In principle, MSP requires longer primers than standard PCR due to cytosines conversion. The optimal primer length for MSP ranges from 20 bp to 30 bp. Another consequence of the bisulfite treatment is fragmentation of the DNA sample. Accordingly, the resulting PCR product should not exceed 300 bp [33].

It is preferable to design the primer pair in which T_m's differ no more than 5°C, because otherwise efficacy of the reaction could be decreased. It is important to note that estimated T_m values can vary significantly between various programs for the same primer pair (Table 2). Calculation of T_m is mainly based on the nearest

neighbour method (NN), taking into account not only relative content of cytosine and guanine, but also the sequence of the primer [34,35]. The calculation of T_m by NN requires several thermodynamic parameters as an input. Differences among the calculated T_m values are mainly caused by using different thermodynamic tables [36]. Currently, there is no consensus about which thermodynamic table for calculation of T_m is the most accurate one. Description of other methods for primer T_m calculation can be found in [37,38].

From our point of view, listing the rules without providing adequate examples is not sufficient. Brandes and coworkers [39] showed that the role of ATM methylation is probably overestimated in breast and NSCLC samples. They found that results from previous publications were based on primers lacking specificity. Actually, the primers did not differentiate between native and bisulfite modified DNA due to the lack of non-CpG cytosines.

On this occasion, we will not debate the mathematical background of the programs or draw conclusions on the particular suitability of the program from these data. The approach we offer is not complex, yet it is equally successful and justifiable. It is based on the frequency of appearance in scientific publications, interface simplicity, and free access. According to these criteria, three programs will be reviewed: *Methprimer*, *Methylprimer express*, and *BiSearch*.

Methprimer is a web-based program freely accessible at <http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>. Besides MSP, there is an option for bisulfite sequencing PCR (BSP) primer design. A special benefit is that any format of DNA sequence is allowed for input. The user should provide the gDNA sequence because the software performs transformation of the sequence into the bisulfite modified one. The program offers numerous options regarding primer features and their location inside the input sequence (Table 3).

Table 2. Calculated T_m-s of two MSP primers [22] by three different free accessible online tools: Oligoanalyzer, Perlprimer and BiSearch. Parameters for T_m calculation of the primers are as follows: concentrations of monovalent cation (Na⁺), Mg²⁺ and primers were 50 mM, 1.5 mM and 200 nM, respectively.

Tool	Oligoanalyzer	Perlprimer	BiSearch
Forward primer T _m	60.1 °C	59.3°C	61.7°C
Reverse primer T _m	55.7 °C	54.4°C	60.4°C

Table 3. List of adjustable options of *Methprimer*, *BiSearch* and *Methylprimer express*. Given values refer to the default settings.

Parameter	<i>Methprimer</i>	<i>BiSearch</i>	<i>Methylprimer express</i>
Product length (bp)	100-300	up to 400	100-175
Primer T _m (°C)	50-60	45-70	56-64
Primer size (bp)	20-30	20-35	18-22
T _m difference between primer pairs (°C)	≤ 5	N/A	N/A
T _m difference between forward and reverse primer	N/A	< 8	< 8
Minimal CpG number per primer	1	1	2
Minimal non-CpG cytosines per primer	4	N/A	2
3' CpG constraint	yes	yes	yes
CpG island search is adjustable	yes	N/A	yes
Primer scoring	N/A	yes	N/A

The results are given in the form of graphic presentation and sequence alignment. However, *Methprimer* is not without drawbacks. For example, T_m overestimation of AT rich primers has been reported [39]. There is also the principal objection that the algorithm behind *Methprimer* is only a modification of the *Primer 3* software [40]. More information regarding the program can be found in [33].

Methylprimer express is a software solution for the MSP primer design developed by *Applied Biosystems*. The program is now available by request from local *Applied Biosystems* office. *Methylprimer express* can be used for MSP and BSP primer searches. Defining the TSS base or origin of translation inside the input sequence is the logical option. If the user decides to use this option, then the primers will be located as close as possible to the marked base. The user can adjust primer characteristics, as well as the CpG island search. Output results are displayed in graphic form and as base sequences. Below the resulting primer sequences T_m value, the number of CpG's and non-CpG cytosines are displayed. Also, the region where primer modification is possible is indicated. However, there are several shortcomings that should be mentioned: T_m overestimation, and sometimes unmethylated primers are not long enough and have low T_m values.

BiSearch is a web-based tool freely accessible at <http://bisearch.enzim.hu/>. In contrast to the previous programs, *BiSearch* has the ability to carry out a similarity search with designed primers on selected native or even bisulfite modified genomes [40]. This option is intended for checking primers for specificity. A further advantage of *BiSearch* is the very fast server [41]. The software proposes only significantly different primer pairs, which enables setting up alternative PCR programs for the same sample [40]. An interesting option might be to choose which strand, plus or minus, will be template for the MSP primer design. The user should provide native (not bisulfite modified) DNA sequence in plain text format.

A practical consideration would be how to set the parameter values in order to obtain reliable primers for the methylation study. The optimal choice would eliminate the main drawback of the method: the possibility that obtained results are false positives [30]. False positive results are the consequence of incomplete bisulfite modification, or low ability of the methylated primer pair to differentiate between methylated and unmethylated alleles. In order to prevent amplification of unconverted DNA, at least one primer should contain several non-CpG cytosines located close to the 3'-end [42]. Mismatches at the 5'-end cannot guarantee primer specificity. Also, a low annealing temperature would stabilize binding of mismatched primers in 30

the PCR reaction [43]. In order to investigate, *in silico* test was performed. Five promoter sequences (region 1000 bp upstream and 200 bp downstream from TSS for following genes: p14, p16, DAPK, MGMT, RASSF1A) were extracted using DBTSS, and were used as the templates for primer design. Initially, default settings of the programs described above were used. These results showed that: i) some primers designed by *Methprimer* have low ability to differentiate between methylated and unmethylated alleles (Figure 5A), which is in line with the study done by Brandes *et al.* [39], ii) *BiSearch* designed primers should be used only in the case of p14 and MGMT, and iii) primers provided by *Methylprimer express* have short sequences (18-22 nt). Therefore, it is not surprising that primer pairs specific for unmethylated alleles have low T_m . For example, the predicted T_m of the MGMT unmethylated forward primer is only 49.89°C (data not shown). Based on the foregoing discussion, it is not the best option to use default settings (Table 3) for primer design, regardless of the program.

A.
 TTAGTTAGAGGGTAGTTAGIAATGTGTIAAGGTGGGGCGTTCCGCGTTTCGGCGGACGT
 ATTGGTTTTTCGGTCGGCGTGGGTGTGGGGCGAGTGGGTGTGTCCGGGTGTGCCGCG
 GTAGAGCGCGTTAGCGAGTTCGGAGCGCGGAGTTGGGAGGAGTAGCGAGCGTCGCGCT
 AGAATTCGTAGCGTCGGTTTGGTAGGGTAGTTCGGAGGTGGGTGGTCCGCTCGTAAAGT
 CGTTTGAAGGGTTHAATTGGTCGTTTTTCGGTCGTTTTTCGTTTAAAGGCCGTAAGGAGTC
 B.
 TGGGGGAGGCGTTGAAGTCGGGGTTCGTTTTGTGGTTTCGTTCCGGTTCGGCTTTGTTAG
 CGTTAAAGTACGGAAGTACGGGTTAATCGGGTATGTCGGGGGAGTTTGTAGTTATTGA
 GTTCCGGGAGTTGGTATTCGTTGGCGCGTTGGGAAGGGTTCGATTCGGTTGGAGCGTG
 C.
 TATTTTTAGTCGTAAAGGTGATTTCGATTTCGGTGGGTTTTTATAAATAGGAAGAATA
 GTTTTGTTTTTTTATGATTAAGAAGAAGTTATATTTTTTTATGATTAATAATTTTCG
 ATTTAATTTGGTAGTTACGAAGGTTGTATCCGGAGGAAGGAACGGGGCGGGGG

Figure 5. A. MSP primer pair designed for assessment of DAPK promoter methylation status using *Methprimer* default settings. The forward primer contains a single CpG and four non-CpG Cs. A single C even at the very 3'-end cannot guarantee primer specificity, i.e., the forward primer has low ability to discriminate between methylated and unmethylated alleles. Four non-CpGs are poorly positioned within the sequence. None of them are positioned at 3'-end. The reverse primer has two CpGs and non-CpG cytosines are positioned well. However, 3 CpG (forward + reverse) are insufficient for reliable assessment of the promoter methylation status. B. MSP primer pair design by *Methylprimer express* using custom settings. Promoter sequence of RASSF1A gene served as a template for the primer design. C. Methylated primer pair designed by *BiSearch* using p16 promoter sequence as a template. Reverse primer does not have any non-CpG cytosines, while forward primer contains only two. This primer pair would be prone to false positive results. Thymines denoted with lower case in the primer sequences mark position of non-CpG cytosines in the native DNA.

However, *Methprimer* and *Methylprimer express* (Figure 5B) gave satisfactory results when the primer search is properly customised. It turned out that *BiSearch* had difficulties designing specific primers in some cases (Figure 5C). Interestingly, *BiSearch* does not have an option to set the number of non-CpG cytosines. Also, there are no options regarding CpG island searches.

For designing specific primers, some principal recommendations should be followed. It is advisable to set the number of CpGs from two to four per primer, and the more CpGs located at the very 3'-end of the primer used, the better. On the other hand, the number of non-CpGs cytosines should be set to four or more, but it is better that these cytosines are located as close as possible to the 3'-end of primer. T_m should be set between 60°C and 70°C, but this is not a strict criteria. According to our laboratory findings, the best results are achieved by using primer pairs which give < 150 bp amplicons. This may stem from the use of formalin fixed paraffin embedded samples in our studies. However, there is a potential advantage to design primer pairs resulting in amplicons not greater than 150 bp. Such primers could be used in the quantitative assessment of methylation status if an appropriate probe is designed. Researchers are encouraged to test several combinations of parameters, and to compare given primers according to the aforementioned criteria.

5. *In silico* analysis of the primers suitability

In most cases, the software used for primer design does not provide much information about primer hairpin structures, primer-dimer formation and 3'-end primer stability. These parameters significantly influence PCR efficacy and yield [31,44]. Formation of primer-dimers and hairpin structures does not impact effective primer concentration, but decreases amplification efficiency, as well. On the other hand, 3'-end primer stability affects its priming specificity [45]. Stable 3'-end priming is desirable to a certain extent; overly stable 3'-end causes non-specific amplification.

Hairpin structures are formed because of the primer self-complementarity. In principle, 3'-end hairpin structures have a more substantial effect on PCR amplification [45]. Actually, Taq polymerase may extend the primer at its 3'-end using the primer 5'-end as a template. Consequently, the structure is stabilized and production of non-specific products is favored. In the study dealing with hairpin structures' effect on PCR, Singh *et al.* reported that the stem length of four or more

nucleotide hairpins significantly influenced amplification efficiency [44].

Primer-dimers are formed by intermolecular interactions between two of the same primers (self-dimers), or a sense and antisense primer within complementary sequences (cross-dimers). Primer-dimers formed as a consequence of 3'-ends complementarity have a more striking effect than internal primer-dimers. During PCR, Taq polymerase carries out extension of 3'-ends of the complementary primers, and a primer-dimer is formed. Thus, competition between primer-dimer and template sequence is established. The consequence is poor amplification.

In order to minimize primer secondary structure formation, the following should be avoided: sequence homology between primers, palindrome sequences, and runs of three or more G's or C's at the primer 3'-end [46].

The degree of stability of these structures is described by dG. dG stands for the Gibbs free energy, a measurement which can be defined as the amount of work that can be extracted from a process operating at a constant pressure. Simply stated, it is a measure of the spontaneity of the reaction. If $dG < 0$, the reaction is favorable and *vice versa*. Information about acceptable dG values for hairpin structures, primer dimers, and 3'-end primer stability can be found at http://www.premierbiosoft.com/tech_notes/. It is important to note that the software calculates dG at 25°C or 37°C in the default settings. Stability of the secondary structure that could be produced during PCR is not relevant at these temperatures. Thus, it is recommended to enter the annealing temperature.

In silico analysis of the primers' suitability involves the use of appropriate software that predicts primer pair behaviour in the test tube. In this way, it is more likely to recognize and discard the primer pair based on its potentially insufficient specificity, or stable hairpin structures. The process could be divided into two steps: i) calculation and prediction of primer thermodynamic characteristics, and ii) checking primer pair specificity by performing a similarity search using an appropriate genome. There are many web-based tools suitable for the primer analysis. Some applications and programs that could help estimating potential formation and stability of hairpin and primer-dimer structures are shown in Table 4.

General features of tools for *in silico* analysis of the primers' suitability will be described using the example of *NetPrimer* (<http://www.premierbiosoft.com/>). It is a web-based application routinely used in our laboratory for checking primer properties. The program allows the user to set values for nucleic acid concentration, ion

Table 4. List of the software and web applications that may be used for checking primer thermodynamic properties.

Tools	Description	Reference
Primerlist	Java application; simple interface; calculation of T _m , T _a and possible primer secondary structures.	http://primerdigital.com/tools/
PerlPrimer	Program written in Perl. It is free for download and use. Enables primer design for standard PCR, qPCR, BSP and sequencing. Provides user with relevant information regarding designed primers (T _m , secondary structures). Pre-designed primers can be analyzed. The software calculates primer T _m , GC content and dG values for primer-dimers and hairpin structures.	http://perlprimer.sourceforge.net/
Primer Premier	Requires download and installation process. Primer premier is a commercial software. Demo version is free for use and enables pre-designed primer analysis. The software provides users with detailed information regarding primer features. User can change parameters that are used for calculating T _m and dG.	http://www.premierbiosoft.com/primerdesign/index.html
Autodimer	Designed for rapid screening of preselected PCR primer pairs for potential cross-reactivity. Convenient tool for checking primer pairs designed for multiplex PCR. User can adjust temperature for dG calculation, total strand concentrations and minimum score requirement. User should provide primer sequences in FASTA format.	http://www.cstl.nist.gov/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm
Oligoanalyzer	A web based tool for preselected primer checking. Provides information about T _m , primer dimers and hairpins. User can adjust primer, dNTPs, Na ⁺ and Mg ²⁺ concentration.	http://eu.idtdna.com/analyzer/applications/oligoanalyzer/

concentrations (Na⁺, Mg²⁺) and temperatures that will be used in calculations. *NetPrimer* displays information such as: secondary structures stability, T_m value, stability of primer 5' and 3'-ends, and primer molecular weight, etc. In order to estimate primer efficiency, *NetPrimer* calculates its rating. This calculation is based on self-dimer and hairpin stability according to the following formula:

$$\text{Rating} = 100 + (\text{dG (Self-Dimer)} \times 1.8 + \text{dG (Hairpin)} \times 1.4).$$

Therefore, if dG (Self-Dimer) = dG (Hairpin) = 0, then the primer rating is 100. During the rating calculation, the software takes into consideration the most stable structures.

A good practice would be to “calibrate” *NetPrimer* and other similar applications. In other words, a researcher is advised to check for primer pairs which are known for providing good results (experimentally verified), for secondary structure and the 3'-end stability. The rationale behind this is that the software does not include all variables that affect PCR outcome. For example, there is no an option to set the concentration of additives that could be used in the PCR mix. Also, there are differences regarding prediction of potential secondary structures formation among the various software options. Consequently, dG values calculated by *NetPrimer* could be greater than the proposed values, but this does not mean that the examined primer pair will have poor performances *in vitro*.

Tools such as *NetPrimer* do not provide any information about priming specificity. In order to check

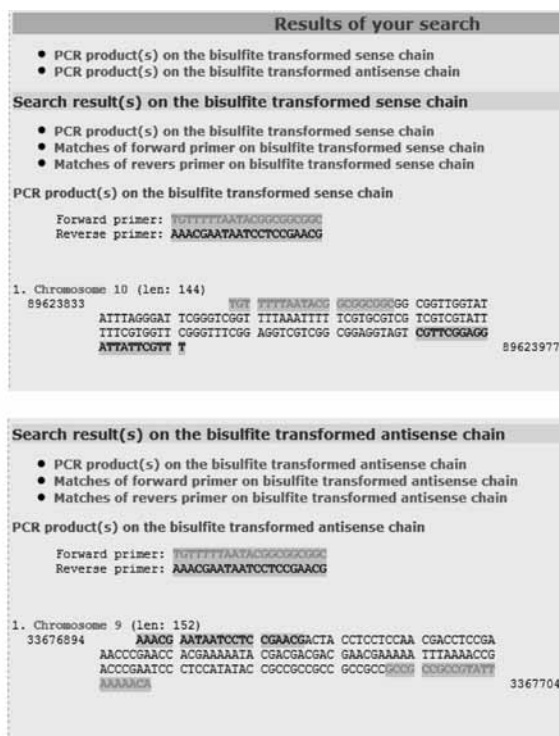


Figure 6. The similarity search done for a primer pair chosen from [23] on bisulfite treated methylated genome (<http://bisearch.enzim.hu/?m=genompsearch>). The results show that the primer pair binds to the target sequence (chromosome 10), but binds to the sequence at chromosome 9, as well.

if a primer pair binds only to the region of interest, it is advisable to do a similarity search on bisulfite-modified

genome using *BiSearch* (Figure 6). This option is known as *ePCR* (electronic PCR). There are three options: i) similarity search on native genome, ii) similarity search on bisulfite-modified unmethylated genome, and iii) similarity search on bisulfite-modified methylated genome. Users can choose between the following genomes: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* and *Pan troglodytes*. It should be noted that the default settings allow for displaying unspecific primer binding only if a mismatch is at 5'-end. This makes sense considering the significance of primer 3'-end for PCR specificity. However, it is recommended that the software be adjusted to accept sequences as 'similar' if a mismatch is at their 3'-end. A single mismatch even at the very 3'-end cannot guarantee specificity of the reaction [47-49].

The whole process of the primer design is summarized and demonstrated with an example depicted in Figure 7.

RASSF1A (NM_007182)

The figure illustrates the process of designing primers for a hypothetical methylation study of the RASSF1A gene (human). The process is shown in four main steps:

- Gene Identification:** The user searches for the RASSF1A gene (NM_007182) in NCBI Entrez Gene. The screenshot shows the gene structure and the option to "Retrieve Promoter Sequence".
- Promoter Sequence Retrieval:** The user uses DBTSS to retrieve the promoter sequence. The screenshot shows the "Retrieve Promoter Sequence" interface with the following parameters:
 - Upstream: 1000 bp (0-10000)
 - Downstream: 200 bp (0-10000)
 - TSS position (designated as 0): NM_007182 (555 56370267) → 56370267
- Primer Design:** The user uses Methprimer to design primers. The screenshot shows the "General Parameters for Primer Selection" and "Parameters for MSP primers" sections.

General Parameters for Primer Selection			
Sequence name (optional):			
Target (optional):	"start_size", such as (660, 30)		
Excluded Regions (optional):	"start_size", such as (160, 50 1100, 50)		
Number of output pairs (optional):	5		
Product Size:	Min: 100	Opt: 150	Max: 200
Primer Tm:	Min: 50	Opt: 65	Max: 70
Primer Size:	Min: 20	Opt: 25	Max: 30
Product CpGs:	4	Primer Poly N:	3
Primer non-CpG C's:	4	Primer Poly T:	8

Parameters for MSP primers	
CpG constraint:	1
CpG in primer:	2
Max Tm difference:	5
- Primer Analysis:** The user uses NetPrimer to analyze the resulting primer pairs. The screenshot shows the "NetPrimer" interface with the following parameters:
 - Gene: RASSF1A
 - Description: Reverse primer
 - Sequence: 5' ACCCGTACTTCGCTTAACTTTAAACG 3' Length: 25 bp

The resulting primer pairs are analyzed, and the best primer pair specific for the methylated allele is shown in the figure:

5'-CGAGAGCGCGtttAGtttCGttTTC-3'
 5'-aCCCGTACTTCGCTTaaCTTTaaCG-3'

Figure 7. Designing primers for hypothetical methylation study of RASSF1A gene (human). The first, RefSeq ID (NM_007182), was retrieved from NCBI (http://www.ncbi.nlm.nih.gov/nucore/NM_007182.4). In order to retrieve RASSF1A promoter sequence, DBTSS was used. RASSF1A ID was entered in the appropriate field which allowed access to information about RASSF1A promoter (see the main text). Using the default setting, DBTSS displays promoter sequence 1000 bp upstream and 200 bp downstream from TSS. This option was selected. Obtained sequence was copied and pasted in the appropriate field of *Methprimer*. *Methprimer* option was configured as the program screenshot display. The resulting primer pairs were analysed by *NetPrimer*. The best primer pair specific for methylated allele is shown in the Figure.

Acknowledgements

This work was supported by the Grant 173049 from the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

- [1] Kiefer J. C., Epigenetics in development, *Dev Dyn*, 2007, 236, 1144-1156
- [2] Delcuve G. P., Rastegar M., Davie J. R., Epigenetic control, *J Cell Physiol*, 2009, 219, 243-250
- [3] Wan J., Oliver V. F., Zhu H., Zack D. J., Qian J., Merbs S. L., Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs, *Nucleic acids research*, 2013, 10.1093/nar/gkt652,
- [4] Deaton A. M., Bird A., CpG islands and the regulation of transcription, *Genes Dev*, 2011, 25, 1010-1022
- [5] Santos K. F., Mazzola, T. F., Carvalho, H. F., The prima donna of epigenetics: the regulation of gene expression by DNA methylation, *Brazilian Journal of Medical and Biological Research*, 2005, 38, 1531-1541
- [6] Portela A., Esteller M., Epigenetic modifications and human disease, *Nature biotechnology*, 2010, 28, 1057-1068
- [7] Esteller M., CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future, *Oncogene*, 2002, 21, 5427-5440
- [8] Egger G. L., G.; Aparicio, A.; Jones, P.A., Epigenetics in human disease and prospects for epigenetic therapy, *Nature*, 2004, 429, 457-463
- [9] Das P. M., Singal R., DNA methylation and cancer, *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 2004, 22, 4632-4642
- [10] Kanai Y., Hirohashi S., Alterations of DNA methylation associated with abnormalities of DNA methyltransferases in human cancers during transition from a precancerous to a malignant state, *Carcinogenesis*, 2007, 28, 2434-2442
- [11] Kondo Y., Issa J. P., DNA methylation profiling in cancer, *Expert Rev Mol Med*, 2010, 12, 23
- [12] Mansego M. L., Milagro F. I., Campion J., Martinez J. A., Techniques of DNA methylation analysis with nutritional applications, *Journal of nutrigenetics and nutrigenomics*, 2013, 6, 83-96
- [13] Herman J. G., Graff J. R., Myohanen S., Nelkin B. D., Baylin S. B., Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands, *Proceedings of the National Academy of Sciences of the United States of America*, 1996, 93, 9821-9826
- [14] Zhang Z., Sun D., Hutajulu S. H., Nawaz I., Nguyen Van D., Huang G., et al., Development of a non-invasive method, multiplex methylation specific PCR (MMSP), for early diagnosis of nasopharyngeal carcinoma, *PloS one*, 2012, 7, 45908
- [15] Delpu Y., Cordelier P., Cho W. C., Torrisani J., DNA methylation and cancer diagnosis, *Int J Mol Sci*, 2013, 14, 15029-15058
- [16] Yamashita R., Sugano S., Suzuki Y., Nakai K., DBTSS: DataBase of Transcriptional Start Sites progress report in 2012, *Nucleic acids research*, 2012, 40, 150-154
- [17] Tsuchihara K., Suzuki Y., Wakaguri H., Irie T., Tanimoto K., Hashimoto S., et al., Massive transcriptional start site analysis of human genes in hypoxia cells, *Nucleic acids research*, 2009, 37, 2249-2263
- [18] Maruyama K., Sugano S., Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides, *Gene*, 1994, 138, 171-174
- [19] Dreos R., Ambrosini G., Cavin Perier R., Bucher P., EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era, *Nucleic acids research*, 2012, 41, 157-164
- [20] Brait M., Sidransky D., Cancer epigenetics: above and beyond, *Toxicology mechanisms and methods*, 2011, 21, 275-288
- [21] Zhu X., Leav I., Leung Y. K., Wu M., Liu Q., Gao Y., et al., Dynamic regulation of estrogen receptor-beta expression by DNA methylation during prostate cancer development and metastasis, *The American journal of pathology*, 2004, 164, 2003-2012

- [22] Bozovic A., Markicevic M., Dimitrijevic B., Jovanovic Cupic S., Krajnovic M., Lukic S., et al., Potential clinical significance of ERbeta ON promoter methylation in sporadic breast cancer, *Med Oncol*, 2013, 30, 642
- [23] Zysman M. C., WB. , Bapat B., Considerations When Analyzing the Methylation Status of PTEN Tumor Suppressor Gene, *American Journal of Pathology*, 2002, 60, 795-800
- [24] Hesson L. B., Packham D., Pontzer E., Funchain P., Eng C., Ward R. L., A reinvestigation of somatic hypermethylation at the PTEN CpG island in cancer cell lines, *Biol Proced Online*, 2012, 14, 5
- [25] Cho Y. G., Chang X., Park I. S., Yamashita K., Shao C., Ha P. K., et al., Promoter methylation of leukemia inhibitory factor receptor gene in colorectal carcinoma, *International journal of oncology*, 2011, 39, 337-344
- [26] Fasan A., Alpermann T., Haferlach C., Grossmann V., Roller A., Kohlmann A., et al., Frequency and prognostic impact of CEBPA proximal, distal and core promoter methylation in normal karyotype AML: a study on 623 cases, *PloS one*, 2013, 8, e54365
- [27] Tada Y., Brena R. M., Hackanson B., Morrison C., Otterson G. A., Plass C., Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein alpha activity in lung cancer, *Journal of the National Cancer Institute*, 2006, 98, 396-406
- [28] Lin T. C., Hou H. A., Chou W. C., Ou D. L., Yu S. L., Tien H. F., et al., CEBPA methylation as a prognostic biomarker in patients with de novo acute myeloid leukemia, *Leukemia*, 2011, 25, 32-40
- [29] Dong S., Kojima T., Shiraiwa M., Mechin M. C., Chavanas S., Serre G., et al., Regulation of the expression of peptidylarginine deiminase type II gene (PADI2) in human keratinocytes involves Sp1 and Sp3 transcription factors, *The Journal of investigative dermatology*, 2005, 124, 1026-1033
- [30] Hernandez H. G., Tse M. Y., Pang S. C., Arboleda H., Forero D. A., Optimizing methodologies for PCR-based DNA methylation analysis, *Biotechniques*, 2013, 55, 181-197
- [31] Dieffenbach C. W., Lowe T. M., Dveksler G. S., General concepts for PCR primer design, *PCR Methods Appl*, 1993, 3, 30-37
- [32] Chuang L. Y., Cheng Y. H., Yang C. H., Specific primer design for the polymerase chain reaction, *Biotechnol Lett*, 2013, 10.1007/s10529-013-1249-8,
- [33] Li L. C., Dahiya, R. , MethPrimer designing primers for methylation PCRs, *Bioinformatics*, 2002, 18, 1427-1431
- [34] Borer P. N., Dengler B., Tinoco I., Jr., Uhlenbeck O. C., Stability of ribonucleic acid double-stranded helices, *J Mol Biol*, 1974, 86, 843-853
- [35] SantaLucia J., Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc Natl Acad Sci U S A*, 1998, 95, 1460-1465
- [36] Panjkovich A., Melo F., Comparison of different melting temperature calculation methods for short DNA sequences, *Bioinformatics*, 2005, 21, 711-722
- [37] Marmur J., Doty P., Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature, *J Mol Biol*, 1962, 5, 109-118
- [38] Wallace R. B., Shaffer J., Murphy R. F., Bonner J., Hirose T., Itakura K., Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch, *Nucleic acids research*, 1979, 6, 3543-3557
- [39] Brandes J. C., Carraway H., Herman J. G., Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter, *Oncogene*, 2007, 26, 6229-6237
- [40] Tusnady G. E., Simon I., Varadi A., Aranyi T., BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes, *Nucleic acids research*, 2005, 33, 9
- [41] Aranyi T., Varadi A., Simon I., Tusnady G. E., The BiSearch web server, *BMC Bioinformatics*, 2006, 7, 431
- [42] Kristensen L. S., Raynor M. P., Candiloro I., Dobrovic A., Methylation profiling of normal individuals reveals mosaic promoter methylation of cancer-associated genes, *Oncotarget*, 2012, 3, 450-461
- [43] Rand K., Qu W., Ho T., Clark S. J., Molloy P., Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives, *Methods*, 2002, 27, 114-120
- [44] Singh V. K., Govindarajan, R., Naik, S., Kumar, A., The Effect of Hairpin Structure on PCR Amplification Efficiency, *Molecular Biology Today*, 2000, 1, 67-69
- [45] Rychlik W., Selection of primers for polymerase chain reaction, *Mol Biotechnol*, 1995, 3, 129-134

- [46] Innis M. A., Gelfand, D. H., Optimization of PCRs, in PCR Protocols (Innis, M A, Gelfand, D H, Sninsky, J J, and White, T J, eds), Academic, New York, 3-12, 1990,
- [47] Simsek M., Adnan H., Effect of single mismatches at 3'-end of primers on polymerase chain reaction, Journal for scientific research Medical sciences / Sultan Qaboos University, 2000, 2, 11-14
- [48] Kwok S., Kellogg D. E., McKinney N., Spasic D., Goda L., Levenson C., et al., Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies, Nucleic acids research, 1990, 18, 999-1005
- [49] Huang M. M., Arnheim N., Goodman M. F., Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR, Nucleic acids research, 1992, 20, 4567-4573
- [50] Miura F., Uematsu C., Sakaki Y., Ito T., A novel strategy to design highly specific PCR primers based on the stability and uniqueness of 3'-end subsequences, Bioinformatics, 2005, 21, 4363-4370