Identification of Potential Methylation Regions of the Smad4 mRNA and Determining Primer Sequences for MS-PCR with the 'Methprimer' Program

Metin Budak[1, 2, *], Ugur Ozkan[3], Mustafa Yildiz[1]

[1]Department of Biophysics, Faculty of Medicine, Trakya University, Edirne, Turkey

[2]Mirko Tos Ear and Hearing Research Center, Trakya University, Edirne, Turkey

[3]Department of Cardiology, Edirne Sultan 1. Murat State Hospital, Edirne, Turkey

Email address:

genomicdna2@yahoo.com (M. Budak)

[*]Corresponding author

Abstract: SMAD4 is a member of an intracellular signaling pathway protein family that is widely expressed in human tissues. This protein is responsible for carrying a chemical signal from the cell membrane to the nucleus. Since reduced SMAD4 expression leads to several tumors and neural disease, it is important to elucidate the mechanisms affecting the expression of this protein. Methylation is among the major factors that affect the expression of the SMAD4 gene. While methylation of the promoter and non-coding exons of SMAD4 gene appear to affect expression, there is no information regarding the other regions of this gene in this regard. Furthermore, cytosine methylation in mRNA is also important in gene activity. For this reason, the demonstration of possible cytosine methylation in mRNA of the SMAD4 gene may be important in understanding gene activity. In this study, we aimed to determine the potential methylation regions in the exons corresponding to SMAD4 protein generation which have not been investigated before. In order to do this, we used the MethPrimer program and identified 25 single CpG sequences and a double CpGpCpG across the exons as potential methylation regions. In addition, 5 pairs of methylated/unmethylated primer sequences were designed with the same program. The study results have shown the presence of potential methylation sequences that are candidates to affect SMAD4 gene expression.

Keywords: Smad4 Protein, Methylation, mRNA, PCR

1. Introduction

SMAD4 is a member of a gene family that produces signal transduction proteins. The SMAD4 protein carries a chemical signal from the cell surface to the nucleus. This signaling pathway works through the TGF-β pathway and may be influenced by cellular environment. Signal formation begins with the binding of the TGF-β protein to the relevant receptor on the cell surface, thereby activating the SMAD protein group. SMAD proteins create a protein complex by binding to SMAD4 and this complex moves towards the nucleus of the cell. Once inside the nucleus, it controls growth and activation of tumor suppressor genes [1, 2]. This gene is highly active and synthesized in several organs from human adrenal glands to the brain, colon, heart, kidneys and testes as well as the placenta [3, 4].

In living organisms, all molecular structures are determined by nucleotide sequences located in the nucleus and collectively referred to as genome. The genome expresses itself through RNA or protein generation and this is called gene expression. There is another mechanism which affects gene expression and has the ability to be inherited and transferred from cell to cell. This mechanism is known as the epigenetic code. During the formation of this code, the DNA sequence is not altered in any way,

and only methyl groups are added to some bases [5, 6]. The most common epigenetic modifications are the changes in histone proteins and DNA methylation. The most widely studied and the most well-established epigenetic mechanism is DNA methylation. It is an enzymatic change where cytosines are converted to 5'-methylcytosine. The cytosine-end methylation seen in mammalian genome often occurs at the 5'-CpG-3' dinucleotides which are also called CpG dinucleotides [2, 7]. In addition, this methylation can also be seen in mRNAs and gene expression can cause changes in the phenotype [8-10]. 'MethPrimer' is a web-based program which has been developed in 2002 by Li LC and Dahiya R (www.urogene/methprimer/ [11]). This program allows identifying the CpG islands and methylation regions in any gene or DNA sequence as well as determining bisulphite sequences for these regions and designing relevant primer sequences of SMAD4 for reverse transcriptase MS-PCR. Methylation-specific PCR is one of the most commonly utilized methods to identify methylation in DNA sequences. In this method, 2 pairs of specifically designed primers are employed for the region thought to be methylated. We used the 'MethPrimer' program in the present study. In the new gene sequence to be formed after DNA bisulphite modification; a pair is designed to bind if the region is methylated while the other pair is designed to bind if the region is unmethylated and the reaction occurs in two separate PCR tubes. Human SMAD4 gene is a considerably large gene containing non-coding exons and a major portion of the gene has not been investigated in terms of methylation.

The aim of the present study is to investigate the potential methylation regions in protein-coding exon regions which have not been previously shown in the literature by means of the MethPrimer program and to design appropriate primer sequences for these regions to be utilized in MS-PCR.

2. Materials and Methods

Human SMAD4 gene contains non-coding exons and 11 protein-coding exons [12]. Exact product of these sequences constitutes the SMAD4 protein. DNA sequence of this region has been obtained from www.ensembl.org and exon onset-termination nucleotide sequences have been checked from the http://www.rcsb.org/pdb/gene/SMAD4?chromosome=chr18&range=48573417&v=hg19database. Subsequently, the sequence in question was uploaded to the MethPrimer program and the program was run to identify CpG islands, methylation regions and primers.

2.1. Bisulfite Modification

mRNA sequence was modified to target the methylation specific primers. The aim is to convert the cytosines in the entire mRNA sequence to uracil while cytosines are converted to thymine in the CpG sequences with methylation potential. Primers were designed assuming that CpGs cannot be turned at the same locus. In this way, RNA for that region was considered to be two alleles for CpG and C5mpG [13]. Bisulfite modification with Sodium bisulfite is considered the "gold standard" for analyzing polinucleotid such as DNA or RNA methylation and can be followed by downstream applications such as PCR or sequencing. In this method, sodium bisulfite deaminates un-methylated cytosines to uracil and leaves methylated cytosines untouched (in other words that is methylated cytosines are resistant to conversion induced by sodium bisulfite). So un-methylated cytosine residues are deaminated to uracil and methylated cytosine (5-mC) residues remain unaffected, enabling for PCR amplification to recognize uracil residues as thymine and 5-mC or 5-hmC as cytosine residues before cDNA synthesis with revers transcription reactions [11].

2.2. Mertprimer and General Parameters for Primer Design

Designing primers in the Methprimer program Tm: (Primer melting temperature). Tm: between 50 and 60 °C, size: 20-30 bp (For bisulfite PCR, primer size is usually bigger than that for regular PCR). Primer GC%: 10-80 (GC percentage of the primer), Primer poly X: 5 (The maximum allowable number of consecutive mononucleotides in a primer sequence except 'T') Primer poly T; 8 (The maximum allowable number of consecutive 'T's (or 'A's in a reverse primer) in a primer sequence. The 'T's here includes both original 'T's and non-CpG 'C's which have been converted to 'T'.) and product size: adjust 100-400 bp.

Parameters for MSP primers; 3'CpG constraint: 3, CpG in primer: 1, and Max Tm difference: 5.

3. Results

A total of 25 single CpGs and a CpGpCpG sequence were identified as potential methylation regions in the DNA sequence consisting of 1659 nucleotides investigated after running the MethPrimer program in Figure 4. A region suitable for MS-PCR was detected and the program generated 5 pairs of methylated and unmethylated primer sequences for this region. Among these, while forward sequences are specific to the CGCG region, reverse sequences are specific to two single potential CpG sequences and two individual CpG potential methylation sequences are located between the forward and reverse sequences respectively in Figures 1, 2 and 3.
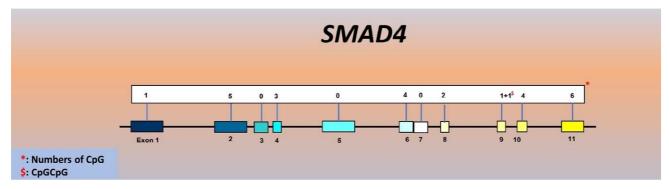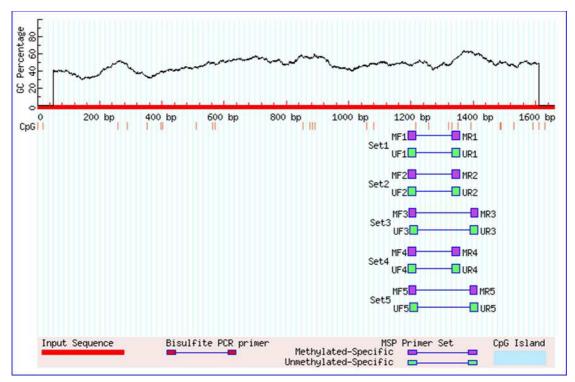
Figure 1. Exons distribution of mCpG base pairs.

Figure 2. Potential mCpG of SMAD4 mRNA exons and MS-PCR primer regions.

```
1187  TGATGTTTGGGTCAGGTGCCTTAGTGACCACGCGGTCTTTGTACAGAGTTACTACTTAGA
      ||||||||||||:|||||::||||||||:::|++++||:|||||:|||||||:||:|||||
1187  TGATGTTTGGGTTAGGTGTTTTAGTGATTACGCGGTTTTTGTATAGAGTTATTATTTAGA
          M>>>>>>>>>>>>>>>>>>>>>>>>
          U>>>>>>>>>>>>>>>>>>>>>>>>


1247  CAGAGAAGCTGGGCGTGCACCTGGAGATGCTGTTCATAAGATCTACCCAAGTGCATATAT
      :|||||||:||||++||:|::||||||||:||||:|||||||:||:::|||||:||||||
1247  TAGAGAAGTTGGGCGTGTATTTGGAGATGTTGTTTATAAGATTTATTTAAGTGTATATAT


1307  AAAGGTCTTTGATTTGCGTCAGTGTCATCGACAGATGCAGCAGCAGGCGGCTACTGCACA
      ||||||:||||||||||++|:|||||:||++|:|||||:||:||:|||++:||:||:|:|
1307  AAAGGTTTTTGATTTGCGTTAGTGTTATCGATAGATGTAGTAGTAGGCGGTTATTGTATA
                    M<<<<<<<<<<<<<<<<<<<<<<<<<
                    U<<<<<<<<<<<<<<<<<<<<<<<<<
```
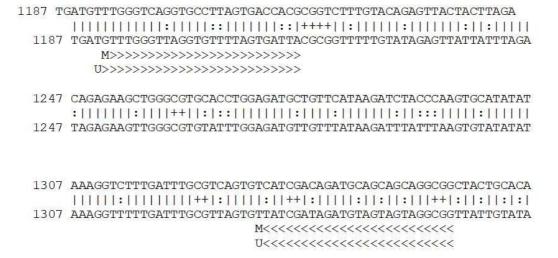
Figure 3. Upper row: Original sequence Lower row: Bisulfite modified sequence for display, assume all CpG sites are methylated ++ unmethylated CpG sites,:::: Unmethlated-CpG 'C' converted to 'T', M >>>>> Left or forward methylated-specific primer, M <<<<<< Right or reverse methylated-specific primer, U >>>>>> Left unmethylated-specific primer, U <<<<<< Right unmethylated-specific primer.

Primer	Start	Size	Tm	GC%	'C's	Sequence
1 Left M primer	1191	25	57.76	56.00	5	TTGGGTTAGGTGTTTTAGTGATTAC
Right M primer	1355	25	58.77	64.00	5	ACCGCCTACTACTACATCTATCGAT
Product size: 165, Tm: 67.0						
Left U primer	1190	26	58.51	53.85	5	TTTGGGTTAGGTGTTTTAGTGATTAT
Right U primer	1355	25	52.45	64.00	5	ACCACCTACTACTACATCTATCAAT
Product size: 166, Tm: 66.7						
2 Left M primer	1191	25	57.76	56.00	5	TTGGGTTAGGTGTTTTAGTGATTAC
Right M primer	1355	25	58.77	64.00	5	ACCGCCTACTACTACATCTATCGAT
Product size: 165, Tm: 67.0						
Left U primer	1190	26	58.51	53.85	5	TTTGGGTTAGGTGTTTTAGTGATTAT
Right U primer	1356	26	53.91	61.54	5	AACCACCTACTACTACATCTATCAAT
Product size: 167, Tm: 66.6						
3 Left M primer	1194	24	58.29	62.50	5	GGTTAGGTGTTTTAGTGATTACGC
Right M primer	1413	24	59.58	66.67	7	ACCAAAAATATTTCCTACCACGAC
Product size: 220, Tm: 66.5						
Left U primer	1195	25	56.13	64.00	5	GTTAGGTGTTTTAGTGATTATGTGG
Right U primer	1414	25	58.76	68.00	8	AACCAAAAATATTTCCTACCACAAC
Product size: 220, Tm: 66.3						
4 Left M primer	1190	26	58.93	53.85	5	TTTGGGTTAGGTGTTTTAGTGATTAC
Right M primer	1355	25	58.77	64.00	5	ACCGCCTACTACTACATCTATCGAT
Product size: 166, Tm: 66.9						
Left U primer	1190	26	58.51	53.85	5	TTTGGGTTAGGTGTTTTAGTGATTAT
Right U primer	1355	25	52.45	64.00	5	ACCACCTACTACTACATCTATCAAT
Product size: 166, Tm: 66.7						
5 Left M primer	1194	24	58.29	62.50	5	GGTTAGGTGTTTTAGTGATTACGC
Right M primer	1412	23	58.85	65.22	6	CCAAAAATATTTCCTACCACGAC
Product size: 219, Tm: 66.5						
Left U primer	1195	25	56.13	64.00	5	GTTAGGTGTTTTAGTGATTATGTGG
Right U primer	1414	25	58.76	68.00	8	AACCAAAAATATTTCCTACCACAAC
Product size: 220, Tm: 66.3						

Figure 4. Possible Primer Sequences for Reverse Transcriptase MS-PCR.

4. Discussion

The human genome contains plenty of methylated cytosine residues which play crucial roles in cellular development and differentiation. Regional DNA methylations begin to take shape in embryonic stages and regulate several reactions that are effective throughout human life. SMAD4 is an intracellular signaling protein involved in cell growth and differentiation, and the expression control of this protein has been shown be highly important in certain stages, with SMAD4 methylations also playing a role in this control [1]. Promoter methylations in the SMAD4 gene has been shown to be decreased and associated with SMAD4 expression in prostate cancers while this was not observed in colon cancer. Promoter hypermethylation has also been shown to reduce SMAD4 expression, thereby leading to tumor formation in gastric tumors [14] Furthermore, some studies have demonstrated an association between the methylation of exon 1 and transcriptional silencing of SMAD4 [15]. In gliomas, decreased SMAD4 expression has been demonstrated although the underlying reason remains unclear. However, we believe that it may affect the neurological system during development of the organism as a high level of expression is seen in neurological structures such as the brain [4].

In addition, highly dynamic modifications in mRNA can affect the posttranscriptional regulations of the RNA and the function of the gene or genes. For the last 20 years, covalent modifications of mRNA have been shown to be quite frequent and are particularly effective in RNA splicing. The most commonly studied of these covalent modifications are N6-methyladenosine (m6A), N6,20-O-dimethyladenosine (m6Am) N1-methyladenosine (m1A) and 5-methylcytosine m5C. There are more than 100 covalent base modifications that are known and have in all types of RNA (mRNA, rRNA, tRNA and snRNA). These modifications are effective in brain tissues that carry out highly complex functions [16, 17]. 5-methylcytosine (5mC) is a nucleotide covalent transformation that can be seen in both DNA and different cellular RNAs. 5-methylcytosine has been a highly studied modification in recent years and the causes and consequences of such changes in RNAs have not been fully elucidated [18]. For the detection of 5mC, one of the methods that usually separates from unmethylated cytonzines is; thin layer chromatography (TLC), high-performance liquid chromatography (HPLC) and mass spectrometry. The most widespread use is by expensive methods such as two-dimensional TLC [19]. Recently, developed DNA bisulfite modification systems, which are relatively more flexible than the above methods, such changes in DNA can also be studied by PCR-based methods. Studies on this subject increase that such methods in RNAs are applied. RNA bisulfite modifications are covalently effective application of 5-methylcytosine that does not alter 5-methylcytosine on the RNA molecule (regardless of the type of RNA), while converting Citozines to Thymines without changing the A and U and T bases [20, 21]. As a result of this application there are two different case series based on 5-methylcytosine and thymine for the same locus in the medium. We believe

that these PCR-based methods will be applicable in SMAD4 mRNA with primer regions and sequences we have defined with appropriate PCR conditions to be applied. For this reason, we find it highly important to identify the potential methylation regions in the SMAD4 gene.

5. Conclusions

To the best of our knowledge, the methylation regions in RNA for SMAD4. We have identified in the active exons of SMAD4 have not been investigated previously. Therefore, we believe that our study is important in terms of showing the methylation changes in exon regions corresponding to protein products of the SMAD4 gene by means of computer-assisted databases. Smad4 protein may be a molecular target for novel drug therapies due to detection of new potential changes in mRNAs of this gene due to the presence of a signaling molecule between cell surface and nucleus and its function as a tumor suppressor in cancer cells. Furthermore, because of physiological role of Smad4 in body systems, the role of balance between atrophy and hypertrophy may also have potential to help in development of new mechanisms to increase expression of smad4, even in future organ repairs. This may help elucidating the factors that affect SMAD4 gene expression. Furthermore, we believe that our study may contribute to wet lab studies by providing ready-to-use primer sequences.

Acknowledgements

References

[1]   S. Roth, P. Laiho, R. Salovaara, V. Launonen, L. A. Aaltonen, (2000). No SMAD4 hypermethylation in colorectal cancer, British journal of cancer 83 (8); 1015.

[2]   L. P. Cacheaux, S. Ivens, Y. David, A. J. Lakhter, G. Bar-Klein, M. Shapira, U. Heinemann, A. Friedman, D. Kaufer, (2009). Transcriptome profiling reveals TGF-β signaling involvement in epileptogenesis, Journal of Neuroscience 29 (28); 8927-8935.

[3]   O. Legendre, A. Sookdeo, D. A. Foster, (2014). BxPC3 pancreatic cancer cells express a truncated Smad4 protein upon PI3K and mTOR inhibition, Oncology letters 7 (4); 1165-1168.

[4]   L. Fagerberg, B. M. Hallström, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M. Habuka, S. Tahmasebpoor, A. Danielsson, K. Edlund, (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics, Molecular & Cellular Proteomics 13 (2); 397-406.

[5]   R. Jaenisch, A. Bird, (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals, Nature genetics 33: 245.

[6]   B. M. Turner, (2000). Histone acetylation and an epigenetic code, Bioessays 22 (9); 836-845.

[7]   B. H. Ramsahoye, D. Biniszkiewicz, F. Lyko, V. Clark, A. P. Bird, R. Jaenisch, (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a, Proceedings of the National Academy of Sciences 97 (10); 5237-5242.

[8]   T. Pollex, K. Hanna, M. Schaefer, (2010). Detection of cytosine methylation in RNA using bisulfite sequencing, Cold Spring Harbor Protocols 2010 (10); pdb. prot5505.

[9]   M. Schaefer, T. Pollex, K. Hanna, F. Lyko, (2008). RNA cytosine methylation analysis by bisulfite sequencing, Nucleic acids research 37 (2); e12-e12.

[10]   F. Tuorto, R. Liebers, T. Musch, M. Schaefer, S. Hofmann, S. Kellner, M. Frye, M. Helm, G. Stoecklin, F. Lyko, (2012). RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis, Nature Structural and Molecular Biology 19 (9); 900.

[11]   L. -C. Li, R. Dahiya, (2002). MethPrimer: designing primers for methylation PCRs, Bioinformatics 18 (11); 1427-1431.

[12]   S. A. Hahn, M. Schutte, A. S. Hoque, C. A. Moskaluk, L. T. Da Costa, E. Rozenblum, C. L. Weinstein, A. Fischer, C. J. Yeo, R. H. Hruban, (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21. 1, Science 271 (5247); 350-353.

[13]   M. Rolando, L. Gomez‐Valero, C. Buchrieser, (2015). Bacterial remodelling of the host epigenome: functional role and evolution of effectors methylating host histones, Cellular microbiology 17 (8); 1098-1107.

[14]   L. -H. Wang, S. -H. Kim, J. H. Lee, Y. -L. Choi, Y. C. Kim, T. S. Park, Y. -C. Hong, C. -F. Wu, Y. K. Shin, (2007). Inactivation of SMAD4 tumor suppressor gene during gastric carcinoma progression, Clinical cancer research 13 (1); 102-110.

[15]   F. Brenet, M. Moh, P. Funk, E. Feierstein, A. J. Viale, N. D. Socci, J. M. Scandura, (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing, PloS one 6 (1) e14524.

[16]   F. Yang, H. Jin, B. Que, Y. Chao, H. Zhang, X. Ying, Z. Zhou, Z. Yuan, J. Su, B. Wu, (2019). Dynamic m 6 A mRNA methylation reveals the role of METTL3-m 6 A-CDCP1 signaling axis in chemical carcinogenesis, Oncogene 38 (24); 4755.

[17]   M. Engel, A. Chen, (2018). The emerging role of mRNA methylation in normal and pathological behavior, Genes, Brain and Behavior 17 (3); e12428.

[18]   N. A. Gkatza, C. Castro, R. F. Harvey, M. Heiß, M. C. Popis, S. Blanco, S. Bornelöv, A. A. Sajini, J. G. Gleeson, J. L. Griffin, (2019). Cytosine-5 RNA methylation links protein synthesis to cell metabolism, PLoS biology 17 (6); e3000297.

[19]   Y. Motorin, F. Lyko, M. Helm, (2009). 5-methylcytosine in RNA: detection, enzymatic formation and biological functions, Nucleic acids research 38 (5); 1415-1430.

[20]   M. Budak, (2018). Logic of Epigenetics and Investigation of Potential Gene Regions, Epigenetics, Intech Open.

[21]   O. Yalcin, M. Budak, (2017). Un-methylation of the survivin gene has no effect on immunohistochemical expression of survivin protein in lung cancer patients with squamous cell carcinoma, Bratislavske lekarske listy 118 (3); 160-163.