We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

6,600 Open access books available
177,000 International authors and editors
195M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter 1

Breaking the Silence: The Interplay Between Transcription Factors and DNA Methylation

Byron Baron

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/55021

1. Introduction

DNA methylation is best known for its role in gene silencing through a methyl group (CH$_3$) being added to the 5’ carbon of cytosine bases (giving 5-methylcytosine) in the promoters of genes leading to suppression of transcription [1]. However this is far from the whole story.

De novo methylation, which involves the addition of a methyl group to unmodified DNA, is described as an epigenetic change because it is a chemical modification to DNA not a change brought about by a DNA mutation. Unlike mutations, methylation changes are potentially reversible. Epigenetic changes also include changes to DNA-associated molecules such as histone modifications, chromatin-remodelling complexes and other small non-coding RNAs including miRNAs and siRNAs [2]. These changes have key roles in imprinting (gene-expression dependent on parental origin), X chromosome inactivation and heterochromatin formation among others [3-5].

DNA methylation leading to silencing is a very important survival mechanism used on repetitive sequences in the human genome, which come from DNA and RNA viruses or from mRNA and tRNA molecules that are able to replicate independently of the host genome. Such elements need to be controlled from spreading throughout the genome, by being silenced through CpG methylation, as they cause genetic instability and activation of oncogenes [6-10]. Such elements can be categorised into three groups: SINEs (Small Interspersed Nuclear Elements), LINEs (Long Interspersed Nuclear Elements) and LTRs (Long Terminal Repeats) [6,11-13]. Repetitive sequences are recognised by Lymphoid-Specific Helicase (LSH) also known as the ‘heterochromatin guardian’ [14,15], which additionally acts on single-copy genes [16].
DNA methylation generally occurs when a cytosine is adjacent 5’ to a guanine, called a CpG di-nucleotide. Such dinucleotides are spread all over the genome and over 70% of CpGs are methylated. Clusters of CpGs, called CpG Islands (CGI), consist of stretches of 200–4000bp that are 60 to 70% G/C rich, found in TATAless promoters and/or first exons of genes [17-19].

In the human genome almost 50% of transcription start sites (TSS) [20], and about 70% of all genes contain CGIs [21,22]. CGIs present in the promoters or first exons of ubiquitously expressed housekeeping and tightly regulated developmental genes are usually hypomethylated, irrespective of transcription activity [1,19,21,23-29] and become silenced when they are hypermethylated [20]. On the contrary, promoters of some tissue-specific genes, with low CpG density, are commonly methylated without loss of transcription activity [21,26,30]. Many active promoters were shown to contain a low percentage of methylation (4 - 7%) indicating that suppression through DNA hypermethylation is density-dependent [21]. The opposite was shown for the cAMP-responsive element (CRE)-binding sites, which are found in the promoters of numerous tissue-specific genes, including hormone-coding and viral genes [31]. Methylation of the CpG at the centre of the CRE sequence inhibits transcription, by inhibiting transcription factor binding, indicating that methylation at specific CpG sites can contribute to the regulation of gene expression [32].

Low-density gene body methylation has been observed in actively transcribed genes and is implicated in reducing ‘transcriptional noise’ – the inappropriate gene transcription from alternative start sites or in cells where it is meant to be silenced [33]. Moreover it is thought to inhibit antisense transcription, to direct RNA splicing and to have a role in replication timing [34-37]. Methylation is thought to play a role in transcriptional elongation, termination and splicing regulation due to higher CpG methylation in exons compared to introns [38,39] and the transcription start and termination regions lacking methylation [40,41].

CpG dinucleotides are not the only sequences that can be methylated, although non-CpG methylation was thought to be infrequent until the methylome of embryonic stem (ES) cells revealed that such non-CpG methylation, generally occurring in a CHG and CHH context, constitutes 25% of total methylated sites in the genome [40]. Non-CpG methylation was also reported in some genes from mouse ES cells [42,43]. The distribution of such non-CG methylation was high in gene bodies and low in promoters and regulatory sequences with almost complete loss during differentiation [40].

DNA methyltransferases (DNMTs) are enzymes that catalyse the addition of methyl groups to cytosine residues in DNA. Mammals have three important DNMTs: DNMT1 is responsible for the maintenance of existing methylation patterns following DNA replication, while DNMT3a and DNMT3b are de novo methyltransferases [1,44-46]. As a result of DNA replication, fully methylated DNA becomes hemi-methylated and DNMT1 binds hemi-methylated DNA to add a methyl group to the 5’ carbon of cytosines [2].

Overall, most DNA methylation changes can be observed invariantly in all tissues [47]. However, the small portion of tissue-specific methylation has a profound effect on cellular activity including cell differentiation, disease and cancer [48-53].
DNA methylation shows different effects on gene expression, brought about by an interplay of several different mechanisms, which can be grouped into three categories [2,54]: i. effects on direct transcription factor binding at CpG dinucleotides; ii. binding of specific methylation-recognition factors (such as MeCP1 and MeCP2) to methylated DNA; iii. changes in chromatin structure.

2. Methylation in development and aging

Key stages in development make use of methylation to switch on/off and regulate gene expression. DNA methylation was shown to be essential for embryonic development through homozygous deletion of the mouse Mtase gene which leads to embryonic lethality [52]. Germline cells show 4% less methylation in CGI promoters, including almost all CGI promoters of germline-specific genes, compared to somatic cells [21].

Immediately after fertilisation but before the first cell division, the paternal DNA undergoes active demethylation throughout the genome [55-58]. After the first cell cycle, the maternal DNA undergoes passive demethylation as a result of a lack of methylation maintenance after mitosis [56,59], and this genome-wide demethylation continues, except for the imprinted genes, until the formation of the blastocyst [60,61].

After implantation, the genome (except for CGIs) undergoes de novo methylation [54]. Active demethylation subsequently occurs during early embryogenesis [62] with tissue-specific genes undergoing demethylation in their respective tissues, creating a methylation pattern which is maintained in the adult, giving each cell type a unique epigenome. [54].

Somatic cells go through the process of aging as they divide and replicate. Aging is characterised by a genome-wide loss and a regional gain of DNA methylation [63]. CGI promoters present an increase in DNA methylation in normal tissues of older individuals at several sites throughout the genome [64,65]. This causes genomic instability and deregulation of tissue-specific and imprinted genes as well as silencing of tumour suppressor genes (controlling cell cycle, apoptosis or DNA repair) through hypermethylation of promoter CGIs [5,66].

The age-related change in methylation was shown in a genome-wide CGI methylation study comparing small intestine (and other tissues) from 3-month-old and 35-month-old mice, which presented linear age-related increased methylation in 21% and decreased methylation in 13% of tested CGIs with strong tissue-specificity [67]. Furthermore, human intestinal age-related aberrant methylation was shown to share similarities to mouse [67]. Although the majority of CGIs methylated in tumours are also methylated in a selection of normal tissues during aging, particular tumours exhibit methylation in specific promoters and are thus said to display a CpG island methylator phenotype (CIMP) [65].

Aging appears to exhibit common methylation features with carcinogenesis and in fact these processes share a large number of hypermethylated genes such as ER, IGF2, N33 and MyoD in colon cancer, NKX2-5 in prostate cancer and several Polycomb-group protein target genes, which suggests they probably have common epigenetic mechanisms driving them [68-70].
3. Methylation in carcinogenesis

DNA methylation can either affect key genes which act as a driving force in cancer formation or else be a downstream effect of cancer progression [71,72]. According to the widely accepted ‘two-hit’ hypothesis of carcinogenesis [73], loss of function of both alleles for a given gene, such as a tumour suppressor gene, is required for malignant transformation. The first hit is typically in the form of a mutation while the second hit tends to be due to aberrant methylation leading to gene suppression. While in familial cancers only one allele needs to be aberrantly methylated to result in carcinogenesis [74,75], both alleles have to be silenced by methylation in non-familial cancers [76,77]. Interestingly, cancer cells appear to use DNMT3b in addition to DNMT1 to maintain hypermethylation [78,79].

Hypermethylation and suppression of promoter CGIs through de novo methylation is well-documented for numerous cancer, affecting mostly general but occasionally tumour-specific genes [3,4,66,80,81]. A study of over 1000 CGIs from almost 100 human primary tumours deduced that on average 600 CGIs out of an estimated 45,000 spread throughout the genome were aberrantly methylated in cancers. It was shown that while some CGI methylation patterns were common to all test tumours, others were highly specific to a specific tumour-type, implying that the methylation of certain groups of CGIs may have implications in the formation, malignancy and progression of specific tumour types [82].

CGI shores (the 2kb region at the boundary of CGIs) are methylated in a tissue-specific manner to regulate gene expression but become hypermethylated in cancer [83-85]. Methylation boundaries flanking the CGIs in the E-cad and VHL tumour suppressor genes were found to be over-ridden by de novo methylation, resulting in transcription suppression and consequentially oncogenesis [86]. On the other hand, the location and function of non-CG methylation in cancer is still mostly unknown [87-88].

Aberrant methylation has been linked to cancer cell energetics. Most cancer cells exhibit the Warburg effect i.e. produce energy mainly through a high level of glycolysis followed by lactic acid fermentation in the cytosol even under aerobic conditions, rather than through a low level of glycolysis followed by oxidative phosphorylation in the mitochondria as is the case in normal cells [89].

In one study it was found that fructose-1,6-bisphosphatase-1 (FBP1), which reduces glycolysis, is down-regulated by the NF-κB pathway partly through hypermethylation of the FBP1 promoter [90]. It was proposed that NF-κB could interact with co-repressors such as Histone deacetylases 1 and 2 (HDAC1 and HDAC2) to suppress gene expression [91,92] and subsequently the HDACs could interact with DNMT1, which gives hypermethylation of the promoter resulting in gene silencing [93-96].

In another study it was proposed that environmental toxins bring about oxidative-stress which affects genome-wide methylation by activating the Ten-Eleven Translocation (TET) proteins (which convert methylcytosine to 5-hydroxymethylcytosine) and chromatin modifying proteins which interfere with oxidative phosphorylation [97].
4. Effect of CpG methylation on transcription factor binding

The methylation of CpGs in transcription factor binding sites in general leads to transcription suppression and gene silencing by directly inhibiting the binding of specific transcription factors. Transcription factors that have CpGs in their recognition sequences and are thus methylation-sensitive include AP-2 [98-100], Ah receptor [101], CREB/ATF [32,100,102], E2F [103], EBP-80 [104], ETS factors [105], MTF-1 [107], c-Myc, c-Myn [108-109], GABP [110], NF-κB [111-100], HiNF-P [112] and MSPF [113].

There are also some transcription factors that are not sensitive to methylation e.g. Sp1, CTF and YY1 [100]. Thus methylation does not hinder binding of gene-specific transcription factors, but rather prevents the binding of ubiquitous factors, and subsequently transcription, in cells where the gene should not be expressed [102].

A model of CpG de novo methylation through over-expression of DNMT1 revealed that despite the overall increase in CGI methylation, there was a differential response of specific sites. The vast majority of CGIs were resistant to de novo methylation, while seven novel sequence patterns proved to be particularly susceptible to aberrant methylation [114]. This essentially means that the sequence in itself plays a role in the methylation state of CGIs. The result of this study implies that specific CGI patterns have an intrinsic susceptibility to aberrant methylation, which means that the genes regulated by promoters containing such CGIs are more susceptible to de novo methylation and could lead to various cancers depending on the genes involved [114].

Various studies have identified three main groups of transcription factors as being important in human cancer: steroid receptors (e.g. oestrogen receptors in breast cancer and androgen receptors in prostate cancer), resident nuclear factors (always in the nucleus e.g. c-JUN) [115,116] and latent cytoplasmic factors (translocated from the cytoplasm to the nucleus after activation e.g. STAT proteins) [115].

Resident nuclear proteins are proteins ubiquitously present in the nucleus irrespective of cell type which include bZip proteins e.g. c-JUN, c-FOS, ATF5s, CREBs and CREMs, the cEBP family, the ETS proteins and the MAD-box family [117]. The different families vary greatly in overall structure and interaction profiles but have the common functional feature of promoting transcription by co-operating with other transcription factors through tandem recognition sequences in promoters as well as by interacting with co-activator proteins [116,118-124]. Resident nuclear transcription factors drive carcinogenesis by direct over-expression or as highly active fusion proteins e.g. MYC acting with MAX [125-127]. The two families of resident nuclear transcription factors that are most prominent in human cancers are the ETS family proteins and proteins composing the AP-1 transcription complexes. ETS family proteins are of particular interest because they promote transcription of a wide range of genes by providing a DNA-binding domain through fusion with other proteins or by mutation [123,128,129].

Latent cytoplasmic proteins are found in the cytoplasm of cells and rely on protein–protein interaction at the cell surface to produce a cascade which activates them as they are directed to the nucleus where they affect transcription by binding to activation sites in the promoters of indu-
cible genes and interacting with transcription initiation factors. They can be activated either directly by tyrosine or serine kinases at the cell surface or by complex processes which include kinases along the pathway [117]. STATs (signal transducers and activators of transcription) are activated by JAK (a tyrosine kinase family) which is activated by various receptors [130,131].

5. Protection mechanisms against methylation

It has been generally accepted that methylation-resistant CGIs are associated with broad expression or housekeeping genes while the majority of methylation-prone CGIs are associated with tissue-specific and thus restricted-expression genes [132]. Exceptions to this pattern have also been found, including WNT10B, NPTXR and POP3. Thus the hypothesis that active transcription has an indirect protective effect against aberrant methylation of CGIs [1,133] has been repeatedly proven to be valid though not absolute [114].

A number of mechanisms have been put forward to explain the relationship between aberrant de novo methylation and cancer. One hypothesis proposed that an initial random methylation event is selected for as proliferation progresses [80]. Another hypothesis proposed the recruitment of DNA methyltransferases to methylation-sensitive sequences by cis-acting factors [134,135], histone methyltransferases such as G9a [136,137], or EZH2 [138]. Yet another hypothesis proposed the loss of chromatin boundaries or the absence of ‘protective’ transcription factors, leading to the spread of DNA methylation in CGIs [139].

The most recent hypothesis proposes the protective character of co-operative binding of transcription factors in maintaining CGIs unmethylated [140]. CGIs showed an unexpected resistance to de novo methylation when DNMT1 was over-expressed. The general pattern that emerged was that most de novo methylated CGIs were characterised by an absence of tandem transcription factor binding sites and an absence of bound transcription factors. Thus protection from de novo methylation requires the presence of tandem transcription factor binding sites that are stably co-bound by at least one general transcription factor, with the second factor being either a general or a tissue-specific transcription factor. Among the most prominent transcription factors found to be linked with aberrant methylation were GABP, SP1, NFY, NRF1 and YY1 [140].

This study re-confirmed that methylation-resistant CGIs were bound by combinations of ubiquitous transcription factors which regulated genes of basic cellular functions, while methylation-prone CGIs were mostly associated with development, differentiation and cell communication, which are frequently regulated by tissue-specific transcription factors [140].

6. Specificity protein 1 (Sp1)

Sp1 is an Sp/KLF (Krüppel-like factor) family member containing a zinc-finger DNA-binding domain [141]. Many KLF proteins regulate cellular proliferation and differentiation.
Multiple Sp1 binding sites are found in the CGI-promoters of housekeeping genes [148,149] as well as CGIs downstream of the TSS [150]. Sp1 sites in gene promoters have been shown to protect CGIs from de novo methylation and maintain expression of downstream genes [151,153] e.g. Sp1-binding site protect the APRT gene from de novo methylation in humans and mice [154,155]. However, Sp1 binding is not methylation-sensitive [151,156,157] and resistance to de novo methylation by DNMT1 is not correlated to the frequency of Sp1 sites in CGIs [114].

Sp1 co-operates with the GABP complex to activate genes which include the folate receptor b [158], CD18 [159], utrophin [141,160], heparanase-1 [161], the pem pd homeobox gene [162], the mouse thymidylate synthase promoter [163] and mouse DNA polymerase alpha primase with E2F [164,165].

7. GA-Binding Protein (GABP)

GABP is a transcription factor composed of two distinct subunits: GABPα and GABPβ. GABPα, also known as Nuclear Respiratory Factor 2 (NRF-2) or Adenovirus E4 Transcription Factor 1 (E4TF1-60), is a member of the E26 Transformation-Specific (ETS) family of proteins [166-169]. However unlike other ETS factors GABPα forms an obligate heteromeric protein complex with GABPβ [170-172]. Together they generally form a heterotetramer consisting of 2α and 2β subunits [173,174] and the presence of sites for GABP binding containing 2 tandem ETS consensus motifs has been reported [175]. On the other hand, single GABP binding sites tend to combine with another site that recognises a different transcription factor e.g. NRF-1, Sp1 or YY1 [175]. GABP is able to recruit co-activators such as PCG1 and p300/CBP that create a chromatin environment favouring transcription [176,177].

GABPα (like all other ETS factors) binds to purine-rich sequences containing a 5'- GGAA/ T-3' core by means of a highly conserved DNA-binding domain made up of an 85 amino acid sequence rich in tryptophan which forms a winged-helix-turn-helix structure, characteristic of the ETS protein family near its carboxy terminal [166,167,170,172,178-181]. The domain through which GABPα binds to the ankyrin repeats of GABPβ is found just downstream of the DNA-binding domain [167,168]. GABPα also has another two domains, the helical bundle pointed (PNT) domain found in its mid-region, which consists of five α-helices [182,183] and the On-SighT (OST) domain near the amino-terminus (residues 35−121), which folds as a 5-stranded β-sheet crossed by a distorted helix and contains two predominant clusters of negatively-charged residues, which might be used to interact with positively-charged proteins [184].

The role of GABP is very versatile and its ability to co-operate with other transcription factors gives it a key role in transcription regulation. GABP and PU.1 compete for binding to the promoter of the b2-integrin gene, yet co-operate to increase gene transcription [185]. GABP also acts as a repressor of mouse ribosomal protein gene transcription [186], apparently by interfering with the formation of the transcriptional initiation complex [187].
GABP is a methylation-sensitive transcription factor [110] and its modulation is best seen in the transactivation of the Cyp2d-9 promoter for the male-specific steroid 16α-hydroxylase in mouse liver where GABP does not bind to the promoter when the CpG site at -97 is methylated [187]. Interestingly, CpG sites located at -93 and -85, outside of the GABP recognition sequence in the Thyroid Stimulating Hormone Receptor (TSHR) gene promoter when methylated, affect the binding of GABP to the promoter, leading to a reduction in basal transcription [187].

8. Therapeutic applications

As more such data is accumulated, it presents methylation as a very interesting and promising tumour-specific therapeutic target especially since the lack of methylation of CGIs in normal cells makes it a safe therapy. Demethylation is known to reactivate the expression of many genes silenced in cultured tumour cells [82]. While high doses of DNMT inhibitors can inhibit DNA synthesis and eventually lead to cell death by cytotoxicity, administration of low doses of these drugs over a prolonged period has a therapeutic effect [188-191]. In fact, the United States Food and Drug Administration has approved the DNMT inhibitors, 5-azacytidine and its derivative 5-aza-2′-deoxycytidine (decitabine), for therapy of patients with solid tumours, myelodysplastic syndrome (which can lead to the development of acute leukemia) and myelogenous leukemia [192].

5-azacitidine acts by becoming phosphorylated and being incorporated into RNA, where it suppresses RNA synthesis and produces a cytotoxic effect [3,193]. It is converted by ribonucleotide reductase to 5-aza-2′-deoxycytidine diphosphate and subsequently phosphorylated. The triphosphate form is then incorporated into DNA in place of cytosine. The substitution of the 5′ nitrogen atom in place of the carbon, traps the DNMTs on the substituted DNA strand and methylation is inhibited [194].

Several more stable analogues such as arabinofuranosyl-5-azacytosine [195], pseudo-isocytidine [196], 5-fluorocytidine [196], pyrimidone [197] and dihydro-5-azacytidine [198] have been tested, and others are undergoing clinical trials [199,200].

Targetting overactive transcription factors is another interesting tumour-specific therapeutic strategy. Many human cancers appear to have a small number of specific overactive transcription factors which are valid candidate targets to at least control further malignancy and metastasis. Such tumour-specific transcription factors are ideal targets because they are less numerous and more significant than other possible protein targets in the transcription activation pathway.

However it is not a simple task to target transcription factors in a controlled manner particularly if attempting to inhibit the interaction of DNA-binding proteins with their recognition sequences [201,202]. Inhibition of a DNA-binding transcription factor can alternatively be done in one of two ways: lowering the overall level of intracellular transcription factor through siRNA or directing methylation to the recognition sequence of the DNA-binding protein. Both options are extremely difficult to carry out in vivo even if their in vitro counterpart has proven to be successful.
9. Conclusion

Research into DNA methylation, particularly at CGIs has come a long way and it is now known that gene silencing, albeit essential, is not the only purpose of methylation processes. In particular, the interactions of transcription factors with promoters have been shown to modulate the function of genes through their methylation-sensitivity and may thus be regarded as viable targets for therapeutics. Unfortunately the biochemical mechanisms and principles required to successfully inhibit protein–protein interactions require further study and clarification [203-206]. Additionally, delivery systems for such cellular treatments also need further study and improvement. However as more focus is put on molecular medicine and with the shift towards personalised medicine, there will surely be significant advances in protein-targetting treatments.

Author details

Byron Baron¹,²

Address all correspondence to: angenlabs@gmail.com

1 Department of Anatomy and Cell Biology, Faculty of Medicine and Surgery, University of Malta, Msida, Malta

2 Department of Biochemistry and Functional Proteomics, Yamaguchi University Graduate School of Medicine, Ube-shi, Yamaguchi-ken, Japan

References


