Review

Functional aspects of cytidine-guanosine dinucleotides and their locations in genes

Franz Varga1*, Heidrun Karlic2, Roman Thaler1 and Klaus Klaushofer1

1 Ludwig Boltzmann Institute of Osteology at the Hanusch Hospital of WGKK and AUVA Trauma Center Meidling, 1st Medical Department, Hanusch Hospital, Heinrich Collin Str. 30, A-1140 Vienna, Austria
2 Ludwig Boltzmann Institute for Leukemia Research and Hematology, Hanusch Hospital, Heinrich Collin Str. 30, A-1140 Vienna, Austria and Ludwig Boltzmann Cluster Oncology, Vienna, Austria

* Corresponding author
e-mail: franz.varga@osteologie.at

Abstract

Originally, the finding of a particular distribution of cytidine-guanosine dinucleotides (CpGs) in genomic DNA was considered to be an interesting structural feature of eukaryotic genome organization. Despite a global depletion of CpGs, genes are frequently associated with CpG clusters called CpG islands (CGIs). CGIs are prevalently unmethylated but often found methylated in pathologic situations. On the other hand, CpGs outside of CGIs are generally methylated and are found mainly in the heterochromatic fraction of the genome. Hypomethylation of those CpGs is associated with genomic instability in malignancy. Additionally, CpG-rich and CpG-poor regions, as well as CpG-shores, are defined. Usually, the methylation status inversely correlates with gene expression. Methylation of CpGs, as well as demethylation and generation of hydroxymethyl-cytosines, is strictly regulated during development and differentiation. This review deals with the relevance of the organizational features of CpGs and their relation to each other.

Keywords: DNA methylation; epigenetics; promoters.

Introduction

Epigenetics refers to meiotically and mitotically heritable changes in gene expression that are not coded in the DNA sequence itself. In eukaryotes, epigenetic processes mainly include DNA methylation and histone modifications (1). In the human genome, the cytosines of CpG dinucleotides are methylated at the 5 position. However, 5-methylcytosine is sensitive to deamination (2) yielding thymidine, which pairs with adenosine when DNA is replicated. In the course of evolution, this has led to a general depletion of CpG dinucleotides in the genome to about 20%–25% of the expected number, of which about 80% are still methylated. More than 90% of the methylated CpG dinucleotides are found in the heterochromatic fraction of the genome. The non-methylated CpG dinucleotides are arranged as CpG-islands (CGIs) in dense clusters located at promoters of about 60% of the genes; these include all housekeeping genes and about half of tissue-specific genes (3, 4). In addition, other types of CpG aggregates have been characterized, namely CpG-poor promoters, CpG-rich promoters and CpG-shores. CpG-poor promoters have been discussed in contrast to CGIs: the CpG frequency at such promoters is similar to the genome average, but their frequency at CpG island promoters is approximately 10-fold higher (5). Compared with the average of the genome, CpG-rich promoters have an increased number of CpGs. However, in such structures, the CpGs are considerably less densely clustered than in CGIs and are often localized near, or between, CGIs (6–11). CpG-shores are CpG-rich structures, which are localized up to 2000 bp distant to CGIs; they are strongly involved in gene regulation (12, 13). This review focuses on DNA-methylation, summarizes recent research on the dynamics of DNA-methylation and analyzes the arrangement and organization of CpG dinucleotides, as well as their functional relevance. In addition, the mechanistic aspects of how DNA methylation and demethylation can be achieved and regulated are covered.

The machinery involved in DNA methylation

In eukaryotes, the majority of CpG dinucleotides are chemically modified by covalent attachment of a methyl group to the C5 position of the cytosine ring. This modification is catalyzed by DNA (cytosine-5′)-methyltransferases (DNMTs) encoded by a highly conserved gene family. These enzymes transfer a methyl-group from S-adenosylmethionine (SAM) to a base of the DNA, in particular, to a cytosine at the 5 position, leaving S-adenosylhomocysteine (SAH), which is in equilibrium with homocysteine. The tetrahydrofolate/folate cycle regenerates SAM via methionine. Homocysteine is discharged either to methionine, catalyzed by vitamin B12-dependent methionine synthase or it can metabolize to cystathione, catalyzed by cystathione synthase using vitamin B6 as the cofactor. Mutations of enzymes from this pathway, or deficiency either in folate (vitamin B9) or vitamin B6 or B12, could cause the accumulation of homocysteine (Figure 1). This could result in high SAH levels that inhibit the methyl-donor activity of SAM, thus blocking DNMT-dependent DNA methylation. Moreover, elevated serum levels of homocysteine also represent a risk factor for several chronic disorders, such as cardiovascular disease, atherosclerosis, chronic renal failure, diabetes, or metabolic syndrome (14,
Figure 1 The THF/folate cycle regenerates methionine, which is necessary for methylation of the cytosines to 5-methylcytosine. The methyl group of SAM is transferred to the cytosines of the DNA yielding methylated DNA (DNAm). The biochemical reaction is catalyzed by DNMTs leaving SAH. SAH is regenerated by the folate metabolic pathway. The pathway depends on folate, vitamin B6 and B12. Depletion of the vitamins, mutations of methylene-THF reductase, or cystathione synthase could result in an increased homocysteine serum level and reduced availability of SAM. SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine; DNAm, methylated DNA; THF, tetrahydrofolate.

15). Furthermore, hyperhomocysteinemia is known to affect bone development and homeostasis (16, 17). This could be the result of reduced collagen quality caused by marred collagen cross-links. The latter is caused by a homocysteine-dependent inhibition of lysyl oxidase (Lox) (18), which is the collagen cross-linking enzyme and a tumor suppressor gene (19). However, homocysteine also down-regulates LOX-expression by increasing DNA-cytosine methylation (10).

As mentioned above, DNMTs are a family of genes conserved in eukaryotes. Homologous genes were found in algae, fungi (except yeasts) worms, insects, plants and animals (20). In mammals, three DNMT families (DNMT1, DNMT2, DNMT3), together with their isoforms, are found (21). Currently, four DNMT1 isoforms are described, which function as maintenance DNMTs propagating pre-existing methylation patterns of hemimethylated DNA (22) to newly synthesized daughter strands during the cell cycle; its expression level is critical for maintenance of imprints during early embryogenesis (23, 24).

DNMT2 is a truncated protein showing no, or only weak, methyl-transferase activity. The specificity of this enzyme for methylating ‘non-CpG’ cytosines in Drosophila (25) is controversial (26). Recent results suggest that tRNA could be a substrate (27, 28). Although DNMT2 was studied in several model metazoans, its function is still not clear (29).

The DNMT3 family includes DNMT3A, DNMT3B, as well as many of the isoforms that emerge from alternative splicing. They act at different stages of animal development and also show tissue specific expression (21, 29). Both DNMT3A and DNMT3B are de novo DNMTs for the establishment of methylation patterns of originally unmethylated DNA (30, 31). Both genes are indispensable for embryogenesis, operating at different stages of embryogenesis. Thus, they may contribute distinct functions during early development and differentiation (32). The third member of this family, DNMT3L (DNA methyltransferase3-like), lacks catalytic activity and is unable to transfer methyl-groups to the DNA. However, it activates DNMT3A and DNMT3B to induce and improve their catalytic activity (33). Targeted disruption of Dnmt3L prevents methylation of sequences that are normally methylated on the maternal allele, while global genome methylation levels remain unaffected (34). Embryonic stem cells that lack Dnmt3L expression are essentially incapable of de novo methylation of newly integrated retroviral DNA and lose DNA methylation during culture time in vitro (35). DNMT3L mostly stimulates DNA methylation at poorly methylated sites, thus generating more uniform patterns. It modulates the intrinsic sequence preferences of DNMT3A and DNMT3B for distinct flanking sequences of target CpG sites (36). In human embryonic kidney cells (HEK 293) stably transfected with each of the 13 different DNMTs, as well as some of their isoforms, it has been shown that DNMTs have specific and overlapping target sites. Different isoforms (mainly the de novo methylating enzymes) differentially methylate repetitive DNA elements that are initially hypomethylated in these cells. Genes associated with histone (H3K4me3) modifications, which are transcriptionally active, are the preferred target sites of DNMT3A1, whereas the specific target sites of DNMT3B1 are associated with histone (H3K27me3) modifications that are transcriptionally inactive. In this model system, transfection of DNMT3L or DNMT1 has induced the methylation of a relatively small number of sites compared to the de novo DNMT3A and DNMT3B, suggesting cooperation with endogenous de novo methylases at specific CpG sites. These findings could explain specific gene methylation-patterns in tumors where increased expression of DNMT isoforms has been found (37). The specificity of DNA-methylation by different DNMTs is additionally steered by modification of histones (38), which may be controlled by Polycomb-group (PcG) proteins (39, 40) in cooperation with histone deacetylases and other repressors of transcription (41).
A key role in this process has been assigned to Hells, firstly described as lymphoid-specific helicase (LSH). It is required for genome-wide CpG methylation in plants and mammals (42–45). Hells has been described as an epigenetic guardian of the genome to protect against activation of retroviral transposable elements, thus stabilizing the genome (46, 47). In association with DNA-methyltransferases, DNMT3A and DNMT3B, Hells is involved in de novo CpG methylation (48). Careful evaluation of Hells-depleted mice revealed that these mice show premature aging with, for example graying and loss of hair, reduced skin fat deposition and osteoporosis (49, 50). Beyond global DNA CpG methylation of heterochromatic transposable elements, Hells is involved in the silencing of both stem cell genes (51) and developmentally regulated homeobox (Hox) genes by CpG methylation (52). Hells is co-regulated with DNMT1 when extracellular collagen matrix prevents up-regulation of the apoptosis mediator Fas (TNF receptor superfamily member 6) in pre-osteoblasts by promoter CpG methylation (11). For Fas promoter methylation, a plethora of genes are involved (41); in particular, the cooperation of histone methylases with Hells is required for developmentally programmed DNA methylation (53). A direct transcriptional regulation of DNMT1 by Friend leukemia integration 1 (Flt) via Janus kinase 2 (Jak2) was recently demonstrated by interleukin 6-IL6 mediated silencing of the collagen cross linker and tumor suppressor Lox by CpG-methylation; for this process co-regulation of DNMT3b and Hells was also found (10). These data suggest that HELLS, similarly to DNMT3L, supports catalytically active DNMTs during specific DNA CpG promoter methylation.

**Demethylation: physiologically, spontaneously or by drugs**

Until recently, CpG methylation has been considered as a stable epigenetic mark as active demethylation by enzymatic breakage of the C-C bonding is an unfavorable process, at least from the thermodynamic point of view. Removal of the methyl-groups from the cytosines occurs passively during replication when the methylation machinery is omitted (e.g., down-regulation of DNMTs), or by drug-mediated DNMT-inhibition and histone (re)acetylation (54, 55). This fact has provided the theoretical basis for treatment protocols in numerous malignancies (56–60). Passive demethylation leads to hypomethylation, as was found in tumorigenesis and aging. Mice expressing only 10% of the wild-type level of the DNMT1 showed tremendous global hypomethylation (61). Down-regulation of DNMT1 with concomitant demethylation of the Fas promoter was found in MC3T3-E1 cells when collagen was dispensed as a growth support (11). In addition to down-regulation of DNMT1, interaction of DNMT1 with RB1, HDAC1 and E2F1 has also been found, which is responsible for transcriptional repression. Moreover, this finding supports a link between transcriptional repression and sequence-specific DNA-demethylation (62).

However, reversibility is a basic requirement for a process that should function as a cellular signal, although, as mentioned above, C-C binding is chemically very stable. Such considerations led to a search for enzymes that execute active DNA-demethylation (63). Findings of a tight correlation between chromatin structure, gene expression and DNA-demethylation further supported a concept of active DNA-demethylation (64), a process which begins with the opening of the chromatin structure by histone acetylation. However, acetylation of histones is not sufficient for demethylation and active transcription of the respective genes by RNA-polymerase II is required for DNA-demethylation. The results of those experiments have suggested that, for this process, an active DNA-methylation is required (65). Furthermore, observations of paternal genome-wide demethylation in many mammals after fertilization suggested the existence of an active 5' -methyl-cytosine (5mC) demethylation process (2, 64). Early findings suggested an enzyme that directly removes methyl-groups by oxidizing it to methanol (66). Later, this demethylation pathway was confirmed by means of recombinant MBD2 (67). Since that time, several other mechanisms have been suggested for active demethylation. In plants, active demethylation uses specific DNA-glycosylases that remove 5mC as a free base and initiate a base excision repair (68). However in mammals, a homologous 5mC specific DNA glycosylase has yet to be identified.

Deamination of 5mec would generate G:T mismatch in DNA duplex, and a subsequent G:T to G:C mismatch repair is a suggested mechanism for active demethylation (69–74).

Nucleotide excision repair has been proposed as an alternative pathway for active DNA-demethylation. The key molecule in this mechanism is growth arrest and DNA-damage-inducible protein 45 a (Gadd45a), a nuclear protein involved in the maintenance of genomic stability, as well as DNA repair and suppression of cell growth. Gadd45a over-expression activates methylation-silenced reporter plasmids and promotes global DNA demethylation, while knockdown silences gene expression and leads to DNA hypermethylation (75). Targeted to rDNA, it triggers the demethylation of promoter-proximal DNA by recruiting the nucleotide excision repair machinery to remove methylated cytosines (76). Another member of this family, Gadd45b, is required for neuronal activity-induced DNA demethylation of specific promoters and the expression of the corresponding genes critical for adult neurogenesis, including brain-derived neurotrophic factor (BDNF) and fibroblast growth factor-1 (FGF1) (77). However, the involvement of Gadd45-family in active DNA-demethylation could not be substantiated (78) and an expected hypermethylation in Gadd45a knockout mice has not yet been found (79).

The detection of unexpected epigenetic modifications, such as 5-hydroxymethyl-cytosine (hmC), proposes a new understanding of gene regulation and DNA demethylation (80). Genes of the ten-eleven translocation (TET) oncogene family catalyze the conversion of 5mC to hmC in cultured cells (81, 82). Being similar to a part of the thymidine iso-rotate salvage pathway, hmC could easily be further oxidized via aldehyde to carboxy-cytosine that is decarboxylated to cytosine. Alternative mechanisms are associated with DNMTs or base excision repair (2). hmC is present in the genome of mouse embryonic stem cells and hmC levels...
involves genome-wide oxidation of 5mC by genome upon fertilization, it has been demonstrated that this of the TET family in early reprogramming of the paternal moter methylation (82). Regarding the involvement of genes that have the ability to integrate into the genome at different sites when activated by demethylation. Thus, they could also lead to passive demethylation during replication (83).

By means of knockdown experiments, it was demonstrated that Tet1 is involved in mouse embryonic stem cell maintenance. This is achieved by maintaining the expression of Nanog homeobox (Nanog), which correlates with its promoter methylation (82). Regarding the involvement of genes of the TET family in early reprogramming of the paternal genome upon fertilization, it has been demonstrated that this involves genome-wide oxidation of 5mC by Tet3 oxidase (84). Enhanced DNA strand breaks have been found in the paternal pronucleus which accumulates hmC, while the maternal genome is protected by the Polycomb gene named Developmental pluripotency-associated 3, Stella, PGC7 (Dppa3) (85). Knockdown experiments of the critical TET-enzyme confirmed the crucial role of this process (82, 86).

Mutations, deletions and chromosomal rearrangements involving the TET locus have been described in various myeloid malignancies (80, 81). Thus, defects of the above-mentioned mechanism may also contribute to the generation of cancer and leukemia (stem) cells.

Elp3 (elongator complex protein 3) contains a histone acetyltransferase-domain and a Radical SAM-domain that uses SAM to catalyze a variety of radical reactions (87). Knockdown of this gene by siRNA, or with similar results of Elp1 and Elp4, in oocytes before fertilization impaired paternal DNA demethylation in the zygote. The need for the presence of an intact Radical SAM-domain suggests an active demethylation with the involvement of the tetrahydrofolate/folate domain of 5-methylcytosine binding proteins localizes these high-methylated sequences into constitutive heterochromatin that is not transcribed (98). However, in mice, the absence of DNA methylation of normally densely methylated repetitive DNA sequences leaves heterochromatic foci visible by microscope, albeit with a somewhat altered composition (99).

Methylated CpGs are distributed in chromatin in a mosaic pattern

Most cytosines of DNA are methylated and distributed over the whole genome as CpGs. Cytosine methylation was found in phylogenetically distant plants and animals such as flowering plants and algae, as well as in mammals and the honeybee. Originally, non-CG methylations (CHG and CHH where H=A, C or T) were only found in plants, but recently such methylation patterns have also been found in animals at a reduced level. In general, methylation is highest in zebrafish and mice and lowest in the honeybee, while plants have intermediate methylation levels (88).

The vast majority of methylated CpGs is found in non-coding highly repetitive DNA, comprising long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs), the most abundant families accounting for 21% and 11%–13% of the genome fraction, respectively (89).

Both contain transposable elements that are DNA sequences that have the ability to integrate into the genome at different sites when activated by demethylation. Thus, they contribute to the genomic instability observed in cancer cells (90, 91).

Mutated mice expressing only 10% of wild type DNMT1, the enzyme responsible for the maintenance of DNA-methylation, are runt-like at birth and develop T-cell lymphomas (61). Moreover, these mice show a reduced overall methylation of a transposable endogenous retroviral element (intracisternal A particle) that is activated and translocated into genes (90). Genomic instability has also been reported in patients that have mutations in the DNMT3B gene leading to the immunodeficiency, centromeric region instability facial anomalies (ICF) syndrome. This disease shows instability of the centromers of chromosomes 1, 9 and 16, which is associated with abnormal hypomethylation of CpG sites in the pericentromeric satellite regions (92). Similarly, in Arabidopsis thaliana the role of DNA methylation in silencing transposable elements has been directly demonstrated. In DNA methylation-deficient mutants many transposable elements are mobilized and de-repressed (93).

The methyl-CpG-binding protein 2 (MeCP2) is a multifunctional protein, which interacts with promoter DNA, either dependently (94, 95) or independently on methylated CpGs (96). In addition to its activity as transcriptional repressor, MeCP2 is also involved in chromatin compaction and heterochromatin formation which is altered in Rett-syndrome (97).

Recently, increased genomic instability has also been found in 5-methylcytosine binding protein 2 (Mecp2)-depleted mice (a murine Rett-syndrome model). The binding domain of 5-methylcytosine binding proteins localizes these highly methylated sequences into constitutive heterochromatin that is not transcribed (98). However, in mice, the absence of DNA methylation of normally densely methylated repetitive DNA sequences leaves heterochromatic foci visible by microscope, albeit with a somewhat altered composition (99).

Methylated CpGs also appear in the coding region of genes (exons, gene-body methylation) and between genes (intergenic regions) and are highly conserved, thus indicating that they were already present in the last common ancestor (20). In plants, where other types of methylation (mCHG and mCHH, where H=A, C or T) exist, gene-body methylation occurs exclusively at CG sites. While methylation in promoters represses transcription (55, 100, 101), body-methylated genes are usually transcribed at moderate-to-high levels (102). High gene-body methylation with low promoter methylation has also been found in an unbiased choice of highly-expressed target genes in human B-lymphocytes, fibroblasts and induced pluripotent stem cells (103, 104). The loss of body methylation of such genes does not lead to increased expression. In contrast, gene-body methylation shows rather moderate up-regulation of transcription compared with body-unmethylated genes (102, 103).

Using high definition profiling of mammalian DNA methylation, dermal fibroblast cells were compared to an invasive breast tumor cell line. Methylated CpGs were enriched in exons compared with introns and the trend was quite pronounced in the cancer cell line with higher overall methylation levels (105).
As already found in plants, non-CG methylation (CHG and CHH) has also been recently found in humans (106), particularly in embryonic stem cells (107). This was supported by results from single base resolution analyses of methylomes from human embryonic stem cells (108). In contrast to parallel analyzed fetal lung fibroblasts, which do not show significant non-CG methylation, in H1 embryonic stem cells almost 25% of all methylated cytosines comprise mCHG and mCHH. Not surprisingly, treatment of embryonic stem cells with BMP4 induced a loss of non-CG methylation at several loci examined, whereas methylation in the CG context was maintained. This indicates that the pervasive non-CG methylation is lost upon differentiation. Moreover, when fetal lung fibroblasts are induced to acquire stem cell characteristics, the analysis of those loci revealed non-CG methylation, suggesting that non-CG methylation is characteristic of stem cell state (108).

As already mentioned above, substantial gene body methylation is an evolutionarily conserved feature of eukaryotic genomes. Interestingly, gene body methylation has been found in plants and animals irrespective of phylegenetic distance. Although the function of gene body methylation remains unknown, it has been proposed to suppress spurious transcription from cryptic promoters that might otherwise interfere with gene regulation (109).

Methylation status of CGIs in normal, aged and tumor cells

As already mentioned, most CpGs are methylated and appear in the heterochromatic fraction of the genome, whereas the non-methylated CpG dinucleotides are arranged as CGIs in dense clusters at the promoter of genes. However, the number of CGIs depends on the definition of a CGI, which has changed during the last years. In the beginning, to predict a CGI, an algorithm was used that was based on sequences longer than 200 bp with a GC content higher than 50% and a ratio of observed-to-expected CpGs of 0.6 (110). However, this criterion resulted in a vast excess of false positive CGIs (111). The increase in the minimum length over which the base compositional (0.65) and CpG frequency (55%) criteria must apply to 400–500 bp eliminates most false positives and has become accepted as standard (112, 113). About 37% of the CGIs are found in promoters of the genes and approximately 70% of known genes have a CGI within -2000 bp to +1000 bp distant to the transcription start sites (TSS) (91, 114).

Physiologically, most CGIs of genes remain unmethylated throughout development, regardless of expression state but a few are methylated during development; this correlates with transcriptional silencing. A unique feature concerns the X-chromosomes in females where CGI hypermethylation, CGI hypomethylation and gene body methylation can be observed. In mammals, during early embryogenesis of females, one of the two active X-chromosomes must be transcriptionally silenced to achieve dosage compensation compared with XY-males. After randomly choosing one of the two X-chromosomes, inactivation starts, spreads along the chromosome and results in heritable repression of most genes on the now inactivated X (115, 116). The inactivated X (Xi) behaves like late-replicating chromatin and shows CGI methylation that correlates with lack of gene expression as demonstrated by hypoxanthine phosphoribosyl-transferase (HPRT) (117). As expected, the HPRT on Xi is hypermethylated at the 5’ promoter region relative to the gene on the activated chromosome (Xa). The middle region of this gene on Xi, however, is consistently hypomethylated compared with that on the Xa (118). These findings were later supported by an allele-specific analysis of more than 1000 loci along the human X chromosome. The alleles on Xi were less than half methylated than that on the Xa, where the methylation was found at multiple neighboring CpGs on the gene bodies; before X-inactivation, both alleles were gene-body hypermethylated (119). These findings agree with gene-body methylation patterns found in plants and honey bees, indicating phylogenetic conservation of this process (20, 120).

This may be involved in the regulation of gene expression and/or splicing (121–123) as found for osteocalcin as a consequence of treatment with a tyrosine kinase-targeting drug in pro-myelocytic leukemia cells (124).

However, in cancers and leukemias, many CGI-associated genes, especially tumor suppressor genes, are hypermethylated – this means that more than 70% of the CGPs of CGIs are methylated and inactivated. On the other hand, a global demethylation process is also observed, resulting in a hypomethylated genome with a CpG methylation level lower than 30% (91, 125–128). Demethylation was observed in repetitive elements (89, 129), including LINE-1 (130, 131), moderately repetitive sequences such as human endogenous pro-virus K (132), and centromeric sequences (133, 134). Recently, the Alu-repeat containing SINEs (135) has been found hypermethylated in tissue-dependent and differentially-methylated regions of embryonic stem cells (136) and in cancer-associated differentially-methylated regions (137). In addition, there was increased transcription of SINE and reverse transcriptase-associated LINE elements in hypoxia (138) and in genomic regions which are involved in epigenetic dysregulation of cancer-associated genes (137). The activated transcription and transposition of those retroviral and transposable elements may contribute to genomic instability and chromosomal rearrangements, and tumor progression (61, 90, 99, 132, 133, 139). However, as a consequence, global hypomethylation also leads to decreased methylation of genes, especially of their promoters. Examples include the loss of imprinting of the IGF2 gene in colorectal cancer, the gamma-globin gene in breast and colon adenocarcinomas, MN, carboxic anhydrase IX (CA9) in renal cell carcinomas and v- myc myelocytomatosis viral oncogene homolog (avian) (MYC) in colorectal cancers and hepatocellular carcinomas (129). Recently, some gene families have been detected as hypomethylated in diverse cancers. The B melanoma antigen (BAGE) sequence family consists of at least 15 loci that map to the juxtacentromeric (centromeric-adjacent) heterochromatic regions of autosomes. Juxtacentromeric regions are hypermethylated in normal cells and hypomethylated in...
tumors (127). Some of these genes are selectively transcribed in the testis and were found hypomethylated and therefore transcribed in a large number of tumors (140). Another gene family that is hypomethylated and expressed in a large number of tumors is melanoma antigen family (MAGE). This cancer/testis antigen family, which is physiologically only expressed in testis, is distributed over the X-chromosome (141–143). Interestingly, class I-part of this family is regulated epigenetically by KIT, which is the cellular homolog of the feline sarcoma viral oncogene, v-kit. Those genes are hypomethylated and highly expressed in the human malignant mast cell line (HMC1). They are down regulated by the inhibition of the mutated and, therefore, constitutively activated KIT by the tyrosine-kinase inhibitor, Imatinib. Analysis of the methylation status of their promoters indicates epigenetic control of aberrant gene expression in tumors by tyrosine kinases (144).

Hypomethylation or mutation of one of two related cellular oncogenes, c-Ha-ras, v-Ha-ras Harvey rat sarcoma viral oncogene homolog (HRAS) or c-Ki-ras, v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS) in primary human carcinomas results in decreased expression of lysyl oxidase (LOX). This gene was first identified by its ability to revert the Ras-transformed phenotype of tumor cell lines and was, therefore, denominated as ras-reversion gene (rrg). The protein comprises two parts: when processed by bone morphogenetic protein-1 (BMP1), the C-terminal peptide works as the oxidizing enzyme while the N-terminal part functions as a tumor-suppressor (145). It was demonstrated in human osteosarcoma cell lines that Suramin, a drug with anti-cancer activity, induces differentiation of these cells possibly by up-regulation of LOX (146). Stable expression of the tumor-suppressor part and, to a lesser extent, of the complete protein can reverse the HRAS(ras)- and ERBB2 (Her-2/neu)-transformed phenotype in mammary epithelial cells and suppress tumor formation (147). The tumor suppressor activity of the LOX-propeptide was also demonstrated in lung and pancreatic cancer cells (148). Recent research suggests that reversal of the cellular tumor phenotype is mediated by the whole pro-enzyme (145). Although the gene is often mutated in lung and pancreatic cancer (148), hypermethylation of CpGs in the LOX promoter also causes down-regulation of its expression in many other tumors; oncogene-driven hypermethylation of CGIs of tumor suppressor genes is common in cancers and leukemias. In human hepatocellular cancer the hypomethylation rate of MYC and HRAS was found to amount to 30% and 61%, respectively (149). In hepatocellular cancer, promoter hypermethylation affects genes of the cell cycle inhibitor CDKN2B (p15, p15INK4b, cyclin-dependent kinase inhibitor 2B), CDKN2B (p16, p16INK4A, cyclin-dependent kinase inhibitor 2A) and genes involved in cell adhesion (CDH1, CDH13, CDH15 corresponding to E-, H- and M-Cadherin), DNA repair and apoptosis and like caspase-8 (CASP8, apoptosis-related cysteine peptidase) (150). Hepatitis C virus (HCV) is a common cause of chronic liver disease. A recent study evaluated the methylation status of multiple tumor suppressor genes in aging liver, chronic hepatitis and hepatocellular carcinoma. In this study, a group of genes, including tumor suppressor genes, were defined where methylation level was increasing with age. This group of genes was even more methylated in aged tissue when infected with HCV and in tumor tissue, even compared to tissue infected with hepatitis B virus (HBV). Two further groups of genes were identified whose methylation status sequentially progresses with advancing stages of chronic viral infection and tumor development (151).

A general change in methylation status has also been suggested to be involved in the aging process as aging is one of the highest risk factors for cancer. Similarly, during aging, global hypomethylation of DNA (89, 152, 153), as well as hypermethylation of promoters of tumor suppressor genes, is found (151, 154–156). During replication, passive methylation by DNMTs preserves the methylation status of the DNA. Failure during this operation, either for stochastic or for systematic reasons, may be responsible for global hypomethylation (157) preferentially of heterochromatic DNA (5, 158, 159). Loss of methylation at those sites could induce genomic instability. Changed expression or activity of DNMTs, especially of the de novo methylases DNMT3B, could result in aberrant methylation of specific promoters of (tumor-suppressor promoter) genes (10, 156, 160–164).

CpG-islands, CpG-rich, CpG-poor regions and CpG-shores

Since the discovery of CpG-clusters in gene-promoters and their denomination as CGIs, the organization of the CGIs, the density, the string and the search for their relevance in the genome have attracted much attention. This interest has increased with the finding that CGIs strings are involved in the regulation of gene expression, and their possible methylation in development, age and tumor. Many algorithms have been developed to predict the number of CGIs in the genome and to correlate it with the number of putative genes; such estimations have revealed that about 60% of genes are associated with CGI, including all housekeeping genes and half of the tissue-specific genes. It has been generally assumed that CGI methylation should be developmentally-dependent and/or tissue-specific and should correlate with gene expression. Therefore, functionally-relevant changes in DNA methylation should occur in CGIs localized either near, or in the direct neighborhood, of gene-promoters. It has been verified for 6%–8% of 17 000 genes that were stably silenced by CGI-methylation during development (112).

However, a comprehensive colon cancer methylome analysis suggested the existence of CpG-shores, which are CpG-rich regions about 2000 bp distant to CGIs and are differentially methylated in tissues and tumors. Moreover, their methylation status strongly correlates with gene expression and is tissue-specifically conserved in mice. It should be emphasized that altered methylation in cancer occurs neither primarily in promoters nor in CGIs (13). The methylation status of those CpG-shores also distinguishes tissue- and cancer-specific cells from artificially-induced human pluripotent stem cells (iPS), human embryonic stem cells (hESCs) and fibroblasts. iPS derived from fibroblasts differ in more
than 4400 regions, where about 2000 are localized within a 2 kb promoter region and have either been hyper- or hypo-methylated. Interestingly, iPS and hESC have a very similar methylome, although 71 regions are different and 51 show hyper- and 20 hypomethylation. Global gene expression studies have revealed a strong inverse correlation between differential gene expression and methylation around 1 kb to the transcription start. Differential DNA-methylation correlated more significantly to transcriptional activity and was associated with CpG-rich sequences.

Interestingly, methylation of a single CpG in the first exon correlated more strongly to transcriptional activity and was associated with CpG-islands (12).

Variably-methylated regions around the TSS with a low CpG ratio and distant to CGIs have been described as transcription factor binding sites (165). Promoters that are poor in CpGs compared with CGIs, led to their denomination as CpG-poor promoters. To increase confusion, some promoters with an increased number of CpGs compared with the surrounding DNA-sequence were also labeled as CpG-rich, although they would resemble CpG-poor promoters (5).

Using the term CpG-rich for areas around the transcription start of a gene with dense accumulation of CpGs that do not meet the stringent definition of a CGI (GC content 0.65, CpG frequency 55% for 400–500 bp, see above) hypermethylation, as well as hypomethylation, will strictly correlate with decreased or increased expression of a gene.

Collagen, type II, α1 (COL2A1), which is associated with a CGI, is expressed in nearly all chondrocytic cell types, while collagen, type X, α1 (COL10A1) that has only few CpGs in its promoter, is only found in hypertrophic articular chondrocytes in close vicinity to subchondral bone. As a model system, adipocytic mesenchymal stem cells (MSC) can be induced to differentiate into chondrocyte-like cells expressing both genes. The CGI of COL2A1 was neither methylated in uninduced nor in induced MSC, while two CpGs of the promoter of COL10A1 changed their methylation status during differentiation (166). Again this result suggests that predominantly non-island CpGs are involved in cytosine-methylation-dependent regulation of expression.

In neuronal cells, the expression level of tyrosine hydroxylase (TH) correlates with the overall methylation level of a CpG-rich sequence. Treatment of non-expressing cells with 5-azacytidine, a DNMT-inhibitor, increases TH expression. Interestingly, methylation of a single CpG in the first exon of the gene correlates well with expression or silencing, respectively. The sequence around this CpG is known for binding the CpG binding protein and repressor of transcription KAIso, thus explaining expression according to the methylation status of this CpG (167).

A subpopulation of the mouse stromal cell line ST2 supports osteoclastogenesis by expressing tumor necrosis factor ligand superfamily, member 11, RANKL (Tnfsf11), while others do not. Mapping of the CpG methylation pattern has revealed that one specific CpG locus was methylated; this stromal subpopulation does not express this gene and does not support osteoclastogenesis. This suggests that the methylation status of a single CpG, three bases upstream of the TATA-box, controls Tnfsf11 expression (168). The structure of this gene reveals a small CGI closely following the transcription start and spans the first exon; the mentioned CpG, however, is localized in a CpG-rich region (Figure 2).

FAS promoter hypermethylation is well known in malignant cells and shows a complex pattern of CpG methylation. Methylation of specific CpGs in promoter- and enhancer-regions regulates its expression and its sensitivity to tumor protein p53 (TP53) in colon carcinoma (169) or to nuclear factor NF-kappa-B transcription factor family (170). In Sezary syndrome, three CpGs are critical for FAS transcription (171). Analysis of the of the human FAS promoter region with a window size of 200 [MethPrimer program (172)] clearly shows two CGIs, the first starts with the TSS, while the other begins with the first exon and reaches into the first intron (Figure 3A). However, increasing the stringency by increasing the window size to 400 caused the first CGI to disappear. In the now CGI-free but CpG-rich region, the critical CpGs are localized (Figure 3A). In contrast, the murine FAS promoter shows no CGI; even when analyzed from -2.000 to 1.000 flanking the TSS (Figure 3B) and also when extended to -5.000 bp upstream; a CpG-rich region is solely found at the TSS. This region is methylated 'pathologically' in Kras- transformed NIH3T3 mouse cells where it prevents these cells from apoptosis (41) and is methylated 'physiologically' in MC3T3-E1 mouse osteoblasts, where CpG demethylation is prevented by interaction with the extracellular matrix guarding the cells from apoptosis by anoikis (11).

Figure 2 Schematic representation of the Tnfsf11 gene starting at -2000 to +1000 bp estimated from the transcription start site (TSS). Analyzing the methylation status around the TSS (TATA-box) revealed that methylation of only one CpG (black triangle) abrogates the expression of Tnfsf11 in osteoblasts. This CpG was never methylated in the osteoblasts of BALB/c mice, but is methylated in other tissues that do not express Tnfsf11. The methylation status of CpG indicated by the gray triangle did not influence the expression of this gene in osteoblasts and was found to be methylated in all tissues investigated (168). The results indicate that methylation of individual nucleotides in CpG-rich regions at the TSS is critical for regulation of gene expression. The black line indicates the 5’ non-coding region. The 5’ untranslated region is indicated by the gray light box and the translated region of the first exon is indicated by the dark gray box. The start of the light gray box is the TSS. The black line after the first exon indicates a part of the first intron. The CGI was calculated using the MethPrimer software (172). The applied settings define a CGI when a DNA stretch of 200 bp (window size) has a GC-content >50% and the observed-to-expected ratio of C/G >0.6 (110).
As already highlighted, the N-terminal propeptide of LOX functions as a tumor suppressor in humans and mice and is, therefore, often inactivated in tumors by promoter CpG methylation (145). Both genes have comparable methylomes with CpG islands beginning around the TSS (Figure 4). However, by increasing the observed-to-expected ratio of C/G to >0.7, the CGI of the mouse Lox gene was shortened and released at an area around the TSS. Detailed analysis of the released CpG string around the TSS of the mouse Lox promoter has revealed a CpG-rich region which is methylated by constitutively active Kras in a tumorigenic situation (41), but also in non-tumorigenic cells when IL6 acts on MC3T3-E1 mouse osteoblasts to methylate CpGs in the promoter, blocking its expression and thus modulating collagen cross-linking (10).

All these findings reveal CpG methylation-dependent gene expression in CpG-rich regions of promoters of those genes when also hidden by a CGI. Changes in methylation status, often of a single, or of several, neighboring CpG, will prevent, or force, the binding of transcriptional activators (169, 170), repressors or methyl-CpG-binding proteins regulating gene transcription, or possibly splicing.

Single CpG-methylation seems to be physiologically regulated during differentiation or, possibly, somatically when only the expression status of a gene must be changed. Whether tumorigenic or other pathologic methylation processes affect the methylation status of different CpGs is a matter of future investigation; however, it is obvious that detailed analysis of the methylation status of every single CpG will be important in understanding epigenetic gene regulation.

Concluding remarks

To date, following the history of CpG-methylation that began with the discovery of CpG clusters in gene-promoters, CGIs have been most attractive subjects for research, while single CpGs in the 5′-regulatory region of promoters have barely been noticed. Only recently, with the characterization of CGI-shores, CpG-poor or CpG-rich regions, single CpGs have become more attractive. CGIs, when methylated, often are subject to heterochromatization. As a consequence, the corresponding genes turn inaccessible and transcriptionally inactivated (119, 173, 174). Recent data, however, suggest that CpG-rich regions are reversibly methylated during cellular development and differentiation; single CpGs, or only some neighboring ones, are frequently affected.

From the available information, we suggest two structural elements with different functions: CGIs and less dense CpG structures that resemble possible CpG-shores, CpG-rich or CpG-poor regions. Methylation of CGIs could be important for heterochromatization and could, therefore, contribute to the silencing of extensive chromatin segments. The individual CpGs, however, are necessary for stable regulation of genes which can also be transmitted to the daughter cells during mitosis. Neither the density nor the number of CpGs is critical as every CpG can be important. The location and the methylation status of the CpGs in the promoter is essential. This status determines whether activators or repressors bind to this promoter region and whether gene expression is stimulated or attenuated.

The probable dynamics of this process is underlined by the recent discovery of active demethylation of CpGs. Such
considerations support new concepts for biomolecular analysis of methylated CpGs at the single cytosine level. Such analyses will reveal differences in DNA methylation in ‘physiologic’ processes during development and differentiation and in ‘pathologic’ situations in cancer, as well as in other chronic diseases which affect CpG-rich regions embedded in CGIs. This implicates the rising relevance of the analysis of the methylation apparatus. The specificity of DNMT-isoforms and their accessory proteins for CpG-containing sequences will gain much more interest. Only then will it be possible to understand the methylation process and the complex action of epigenetic drugs (175). The development and improvement of drugs specifically influencing the methylation process will help to modulate physiologic development and to prevent pathologic processes in organisms.

Acknowledgements

Supported by Medical Scientific Fund of the Mayor of the City of Vienna (Medizinisch-Wissenschaftlicher Fonds des Bürgermeisters der Bundeshauptstadt Wien). The authors thank E. Schiemel, K. Schiemel and M. Adams for carefully reading the manuscript.

References

11. Thaler R, Karlic H, Spitzer S, Klaushofer K, Varga F. Extra-
cellular matrix suppresses expression of the apoptosis mediator 
Fas by epigenetic DNA methylation. Apoptosis 2010; 15: 728–
37.

Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger 
T, Daley GQ, Feinberg AP. Differential methylation of 
tumor- and cancer-specific CpG island shores distinguishes human 
induced pluripotent stem cells, embryonic stem cells and fibro-

Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, 
Pothas JB, Sabunciyan S, Feinberg AP. The human colon cancer 
metastome shows similar hypo- and hypermethylation at 
41: 178–86.


15. Iannucci CV, Capoccia D, Calabria M, Leonetti F. Metabolic 
syndrome and adipose tissue: new clinical aspects and thera-

16. Raisz LG. Homocysteine and osteoporotic fractures—culprit or 

High levels of homocysteine inhibit lysyl oxidase (LOX) and 
downregulate LOX expression in vascular endothelial cells. 
Atherosclerosis 2004; 177: 1–8.

18. Liu G, Nellaiappan K, Kagan HM. Irreversible inhibition of 
lysyl oxidase by homocysteine thiolactone and its selenium 
and oxygen analogues. Implications for homocystinuria. J Biol 

19. Kagan HM, Li W. Lysyl oxidase: properties, specificity, and 
biological roles inside and outside of the cell. J Cell Biochem 

20. Jeltsch A. Molecular biology. Phylogeny of methylomes. Sci-

21. Lan J, Hua S, He X, Zhang Y. DNA methyltransferases and 
methyl-binding proteins of mammals. Acta Biochim Biophys 
Sin (Shanghai) 2010; 42: 243–52.

22. Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of 
the murine DNA methyltransferase Dnmt1 is controlled by 
interaction of the catalytic domain with the N-terminal part of 
the enzyme leading to an allosteric activation of the enzyme 
after binding to methylated DNA. J Mol Biol 2001; 309: 
1189–99.

23. Branco MR, Oda M, Reik W. Safeguarding parental identity: 
Dnmt1 maintains imprints during epigenetic reprogramming 
71.

H, Asano T, Kurihara H. Maintenance of genomic methylation 
patterns during preimplantation development requires the 
somatic form of DNA methyltransferase 1. Dev Biol 2008; 313: 
335–46.

25. Phalke S, Nickel O, Walluscheck D, Hertig F, Onorati MC, 
Reuter G. Retrotransposon silencing and telomere integrity in 
somatic cells of Drosophila depends on the cytosome-5 methyl-

26. Schaefer M, Lyko F. Lack of evidence for DNA methylation of 
Invader® retroelements in Drosophila and implications for 
Dnmt2-mediated epigenetic regulation. Nat Genet 2010; 42: 
920–1.

27. Schaefer M, Pollex T, Hanna K, Tuerto F, Meusburger M, Helm 
M, Lyko F. RNA methylation by Dnmt2 protects transfer RNAs 

28. Tovy A, Hofmann B, Helm M, Ankri S. In vitro RNA meth-
ylation assay with the Entamoeba histolytica DNA and tRNA 

L, Xu Z. Molecular and enzymatic profiles of mammalian DNA 
methyltransferases: structures and targets for drugs. Curr Med 

30. Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and 
maintenance of genomic methylation patterns in mouse embryo-
onic stem cells by Dnmt3a and Dnmt3b. Molecular and cellular 

31. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases 
Dnmt3a and Dnmt3b are essential for de novo methylation and 

32. Watanabe D, Suetake I, Tada T, Tajima S. Stage- and cell-
specific expression of Dnmt3a and Dnmt3b during embryogen-

33. Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. 
DNMT3L stimulates the DNA methylation activity of Dnmt3a 
and Dnmt3b through a direct interaction. J Biol Chem 2004; 
279: 27816–23.

34. Bourc’his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L 
and the establishment of maternal genomic imprints. Science 

35. Ooi SK, Wolf D, Hartung O, Agarwal S, Daley GQ, Goff SP, 
Bestor TH. Dynamic instability of genomic methylation patterns 

36. Wenzel RL, Kureta MS, Moarefi AH, Gordon CA, Ginn PA, 
Chedin F. DNMT3L modulates significant and distinct flanking 
sequence preference for DNA methylation by DNMT3A and 

37. Choi SH, Heo K, Byun HM, An W, Lu W, Yang AS. Identifi-
cation of preferential target sites for human DNA methyltrans-

38. Black JC, Whitestine JR. Chromatin landscape: methylation 

39. Cedar H, Bergman Y. Linking DNA methylation and histone 
methylation: patterns and paradigms. Nat Rev Genet 2009; 10: 
295–304.

40. Sawarkar R, Paro R. Interpretation of developmental signaling 
at chromatin: the Polycomb perspective. Dev Cell 2010; 19: 
651–61.

41. Gazin C, Wajapeyee N, Gobeil S, Virbasius CM, Green MR. 
An elaborate pathway required for Ras-mediated epigenetic 

42. Bourc’his D, Bestor TH. Helicase homologues maintain cyto-
sine methylation in plants and mammals. Bioessays 2002; 24: 
297–9.

43. Dennis K, Fan T, Geiman T, Yan Q, Muegge K. Lsh, a member 
of the SNF2 family, is required for genome-wide methylation. 

44. Jeddoh LA, Stokes TL, Richards EJ. Maintenance of genomic 
methylation requires a SWI2/SNF2-like protein. Nat Genet 

45. Suzuki T, Farrar JE, Yegnasubramanian S, Zahed M, Suzuki N, 
Arcelli RJ. Stable knockdown of PASG enhances DNA deme-
thylation but does not accelerate cellular senescence in TIG-7 
human fibroblasts. Epigenetics: official journal of the DNA 

46. Fan T, Yan Q, Huang J, Austin S, Cho E, Ferris D, Muegge K. 
Lsh-deficient murine embryonal fibroblasts show reduced pro-
liferation with signs of abnormal mitosis. Cancer Res 2003; 63: 
4677–83.


166. Thorogood NP, Brown CJ. Active chromatin marks are retained on X chromosomes lacking gene or repeat silencing despite XIST/Xist expression in somatic cell hybrids. PLoS One 2010; 5: e10787.

Received April 11, 2011; accepted July 12, 2011
Franz Varga completed his PhD in Molecular Biology at the Institute of Molecular Biology at the University of Vienna in 1986. In the same year he became assistant professor at the Department of Genetics and Cytogenetics at the Institute of Botany at the aforementioned university. Since 1992 he has worked as a senior scientist at the Ludwig Boltzmann Institute of Osteology at Hanusch-Hospital in Vienna. His laboratory work focuses on the epigenetics of osteoblast differentiation, and how epigenetically active drugs, steroid hormones and the extracellular matrix influence that process.

Heidrun Karlic studied biology, biochemistry and philosophy. In 1980, she received her PhD at the University of Vienna. Since 2000, she is an associate professor for Experimental Hematology at the Medical University of Vienna (Department of Internal Medicine). From 1981 to 1987, she worked as an assistant professor at the Medical University of Vienna at the Departments for Molecular Biology and Gynecology. Since 1987, she has been employed as a senior scientist at the Ludwig Boltzmann Institute for Leukemia Research in Vienna (part of Ludwig Boltzmann Cluster Oncology since 2008). Her research interests include epigenetics and analyses of gene expression pathways with a special focus on myeloid leukemias and myelodysplastic syndromes.

Roman Thaler completed his undergraduate degree in Nutritional Sciences at the Department of Nutritional Sciences at the University of Vienna. He completed his PhD at the Ludwig Boltzmann Institute of Osteology in the Hanusch Hospital under the supervision of Dr. Franz Varga and Professor Klaus Klaushofer. For his PhD thesis, investigating effects of extracellular matrix and homocysteine on epigenetics of bone cells, he received a university. He has received awards from the Austrian Society of Endocrinology and Metabolism and the European Calcified Tissue Society. He currently holds a postdoctoral position at the Ludwig Boltzmann Institute of Osteology.

Klaus Klaushofer was appointed Associate Professor of the Medical University Vienna in 1989 and Professor in 1995. He specializes in Internal Medicine and has additional degrees in Rheumatology, Endocrinology and Metabolism. Moreover, he is Head of the Ludwig Boltzmann Institute of Osteology (since 1991). He was Head of the 4th Medical Department from 2001 until 2010. Presently he is Head of the 1st Medical Department (since 2010) and Medical Director of Hanusch Hospital (since 2001).