

SURVEY AND SUMMARY

Presence and role of cytosine methylation in DNA viruses of animals

Karin Hoelzer¹, Laura A. Shackelton² and Colin R. Parrish^{1,*}

¹Baker Institute for Animal Health, Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853 and ²Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA

Received February 5, 2008; Revised February 29, 2008; Accepted March 4, 2008

ABSTRACT

Nucleotide composition varies greatly among DNA viruses of animals, yet the evolutionary pressures and biological mechanisms driving these patterns are unclear. One of the most striking discrepancies lies in the frequency of CpG (the dinucleotide CG, linked by a phosphate group), which is under-represented in most small DNA viruses (those with genomes below 10 kb) but not in larger DNA viruses. Cytosine methylation might be partially responsible, but research on this topic has focused on a few virus groups. For several viruses that integrate their genome into the host genome, the methylation status during this stage has been studied extensively, and the relationship between methylation and viral-induced tumor formation has been examined carefully. However, for actively replicating viruses—particularly small DNA viruses—the methylation status of CpG motifs is rarely known and the effects on the viral life cycle are obscure. In vertebrate host genomes, most cytosines at CpG sites are methylated, which in vertebrates acts to regulate gene expression and facilitates the recognition of unmethylated, potentially pathogen-associated DNA. Here we briefly introduce cytosine methylation before reviewing what is currently known about CpG methylation in DNA viruses.

INTRODUCTION

CpG underrepresentation in vertebrate genomes

The denotation ‘CpG’ is shorthand for the occurrence of a cytosine linked, through a phosphate bond, to a guanine. CpGs are underrepresented in most eukaryote genomes, but the frequency varies widely among species and is

negatively correlated with the presence and extent of cytosine methylation in the genome (1,2). In vertebrate genomes, CpGs are present at one-third to one-fourth of the expected frequency, yet the reasons are disputed (1–4). Cytosine and guanine tend to have higher stacking energies than adenine and thymine, so structural constraints may be important in CpG avoidance (5). In many species, the proportion of tRNAs containing CpG in their anticodons is lower than that of tRNAs with other dinucleotides; therefore, the transcription efficiency might be higher for codons not containing CpGs (6). Another explanation may lie in the fact that unmethylated CpGs can stimulate innate immune responses, potentially resulting in autoimmune reactions (7–9); therefore, large numbers of CpGs may be detrimental if not all are methylated. Finally, methylated cytosines have a tendency for spontaneous deamination, which may also account for the CpG depletion (10–12). While the deamination of unmethylated cytosines leads to uracil and can be corrected by cellular DNA repair machinery, the transition of methylated cytosines to thymines is irreversible, leading to elevated mutation frequencies in highly methylated genomes.

In general, vertebrate genomic CpGs are highly methylated. Sixty to 90% of genomic CpGs are thought to be in a methylated state (13,14), but both CpG frequency and methylation patterns can vary widely across a single vertebrate genome. Notably, some regions are CpG-enriched yet practically devoid of methylation. These sequence stretches, termed ‘CpG islands’, are >500 bp in length and comprise ~1% of total genomic DNA [e.g. the human genome contains more than 29 000 such islands, with estimates reaching as high as 45 000 islands per haploid genome (15,16)]. The islands are often associated with 5′ promoter regions of housekeeping genes [for reviews see, among others, (15,17,18)].

Vertebrate CpG methyltransferases

Vertebrate genomes are methylated by DNA methyltransferases (DNMTs) which convert cytosine to

*To whom correspondence should be addressed. Tel: +1 607 256 5649; Fax: +1 607 256 5608; Email: crp3@cornell.edu

Table 1. Overview of vertebrate DNA methyltransferases and their functions

DNMT	Functional role	Reference
De-novo DNMTs		
DNMT 3a	Embryonic development, methylation of CpG sites, meiosis induction in sperm(?)	(110–112)
DNMT 3b	Embryonic development, spermatogenesis(?)	(110,113)
DNMT 3L	Maternal genomic imprinting, silencing of retrotransposons in spermatogonial stem cells	(114)
Maintenance DNMTs		
DNMT 1	Cellular maintenance methylation, maintenance of imprinting, silencing of mobile elements during genomic demethylation, contribution to histone deacetylases	(115–118)
DNMT 2	Unclear. Methylation in <i>Drosophila melanogaster</i> (?)	(119)

The enzymatic functions attributed to known DNMTs are summarized here, and some key references are provided. Some simplifications were made for the purpose of clarity, and the reader is referred to specific reviews of DNMTs, indicated in the text, for more detail. Cases where the functional role has been proposed, but not yet established conclusively, are indicated by question marks.

5-methylcytosine (5Me-cytosine). DNMTs are functionally divided into ‘*de novo*’ (DNMT3a and DNMT3b) and ‘maintenance’ (DNMT1) methyltransferases (Table 1). Their catalytic domains appear highly conserved across species and *S*-adenosyl methionine (SAM) appears to function as the only methyl donor [reviews of mammalian DNA methyltransferases can be found in (19–22)]. DNMT expression levels in nontransformed tissues have rarely been studied explicitly. The available data indicates the translation of all three DNMTs in the majority of tissues, but translation levels seem to vary between tissues, cell differentiation levels, host developmental stages and potentially host species. Robertson *et al.* (23) analyzed DNMT mRNA levels in various human tissues and found DNMT1, 3a and 3b expressed in nearly all analyzed fetal and adult tissues. In adult tissues, DNMT1 appeared to be generally more highly expressed than DNMT3a and 3b, and DNMT1 mRNA was detected in all analyzed tissues except the small intestine. DNMT3a and 3b mRNAs were detected in large quantities in heart, skeletal muscle, thymus, kidney, liver and peripheral blood mononuclear cells, but were present at low levels in all other analyzed tissues. DNMT1 and 3a seemed to be expressed at high levels in all analyzed fetal tissues, while DNMT 3b was expressed at high levels only in fetal liver. The authors proposed fetal hepatic hematopoiesis as the reason for the observed high levels of DNMT3b in the fetal liver. In bovine fetuses, Golding *et al.* (24) found DNMT3b highly expressed in rumen, kidney, testes and lung while the highest mRNA levels were again detected in the liver. MessengerRNAs of all three DNMTs were again detected in all of the adult bovine tissues analyzed, but highest levels were detected in kidney, brain and testes. Mizuno *et al.* (25) detected mRNAs for all three DNMTs in human neutrophils, monocytes, T lymphocytes, bone marrow cells and CD34-positive immune cells, but mRNA levels varied between cell types. DNMT1 appeared to be expressed at high levels in all cell types except neutrophils, while DNMT3a mRNA levels in neutrophils and T lymphocytes appeared high. DNMT3b mRNA levels were low in differentiated cells such as neutrophils, but high in bone marrow and particularly CD34-positive cells.

Methylation patterns are generally stable, sustained by DNMT1 and are inherited by both daughter DNA molecules during mitosis. However, during meiosis both

parental genomes are demethylated prior to fertilization, and cellular methylation patterns become established *de novo* during embryogenesis. This process usually starts during implantation and is finished by the end of gastrulation (26). The development of site-specific methylation patterns is crucial for normal embryonic development (27–29) and is responsible for genomic imprinting and for X chromosome silencing (30,31).

Methylation-induced gene silencing

CpG methylation acts to suppress transcription in several ways (32). It can directly prevent the binding of transcription factors to the promoter regions of genes, most likely due to steric hindrance. Alternatively, a number of proteins, known as methyl-CpG binding (MeCP) proteins, can selectively bind methylated CpG sites. One of these, Kaiso, recognizes methylated CpG sites through a zinc-finger motif, while another subgroup shares a conserved methyl-CpG-binding domain (MBD). The MBD proteins MBD 1, 2 and MeCP2, as well as Kaiso, all function as transcriptional repressors. MBD 3, however, associates with the nucleosome remodeling and histone deacetylation (NuRD) complex, a co-repressor complex containing histone deacetylases. Deacetylation of histones increases their affinity for DNA, leading to the formation of inactive heterochromatin (30,33,34).

Methylation-induced silencing of foreign DNA

Methylation is involved in the inactivation of integrated foreign DNA, such as retrotransposons, proviral sequences and other transposable elements (TEs) (33). These ‘parasitic’ DNAs constitute a quantitatively significant portion of mammalian genomes, comprising, for example, ~40% of the human genome. They are notably GC rich and, within mammalian genomes, are hypermethylated at CpG motifs. In several studies, the inhibition of *de novo* or maintenance methylation was associated with increased transcription levels of TEs, indicating that effective TE silencing requires both *de novo* and maintenance methylation, likely catalyzed by DNMT3L and DNMT1, respectively (35,36).

Methylation and innate immunity

CpG methylation also plays an important role in immune surveillance and the detection of pathogens,

as unmethylated CpGs are a signature of bacterial and other pathogen-associated DNA. CpG-containing oligodeoxynucleotides (ODNs), even as short as 6 nt in length, can be potent stimulators of the vertebrate immune system, resulting in the activation of a wide variety of immune cells such as B cells, NK cells, monocytes/macrophages and dendritic cells (DCs). Stimulation with CpG ODNs directly activates plasmacytoid dendritic cells (PDCs) and B cells. The PDCs secrete large amounts of IFN- α and IL-12 which subsequently stimulate monocytes, myeloid DCs, T cells and type-1 NK cells (37–39). One pathway through which unmethylated CpGs trigger innate immune responses is by binding a member of the Toll-like receptor (TLR) family, TLR9 [see, as one example, (40)]. TLR9 is expressed at high levels on subsets of human and mouse DCs and gene homologs have been found in many other mammalian species. The degree to which particular CpG sequences stimulate immune responses depends upon their flanking nucleotide sequences, and motifs which confer maximal stimulation are at least partially species-specific (41). For example, ODNs containing the motif GACGTT very efficiently stimulate murine and rabbit B cells, but only weakly stimulate human B cells, which are most effectively stimulated by GTCGTT (42). Some similarities in recognition sequences among related species appear likely. For example, GTCGTT motifs are highly immune stimulatory in a variety of vertebrates including humans, dogs, cats, cows, chicken and nonhuman primates (41). ODNs that are thymine rich at the 3' end and contain a TpC dinucleotide at their 5' end appear to be generally potent immune stimulators, while those that contain CpG motifs toward the 3' end appear to be less immunogenic (41). Generalities, however, may be misleading in some cases, as resulting immune effects are thought to be highly cell-type and species specific (43–45).

CpG dinucleotides in vertebrate genomes are most commonly preceded by a C and followed by a G (46), a signature which does not appear to be immune stimulatory. Motifs such as TTCCGA, on the other hand, appear to be potent stimulators of the human immune system and may play a role in autoimmune diseases such as lupus erythematosus. Other CpG motifs, for example those containing poly(G) sequences, are thought to inhibit NF- κ B-induced immune stimulation (47). The relatively low occurrence of stimulatory motifs in mammalian DNA might explain the lower immunogenicity of unmethylated vertebrate DNA compared to bacterial DNA.

Stimulatory motifs in DNA virus genomes. The genomes of some, but not all, viruses are also lacking many such stimulatory sequences. For example, likely stimulatory sequences are highly underrepresented in the genomes of adenovirus serotypes 2 and 5, but not in serotype 12. Adenovirus type 2 and 5 can cause persistent infections, their DNA is not immune stimulatory and it might even suppress immune stimulation by DNA from other sources. Type 12, however, does not cause persistent infections, and its DNA appears highly immune stimulatory. Experimental evidence for a correlation between the stimulatory potential of the viral DNA and the number of supposedly stimulatory motifs has been presented (48).

Table 2 summarizes the frequencies of certain putatively stimulatory and nonstimulatory motifs in genomes from several members of different viral families. General conclusions are difficult to draw since the biological effects of individual motifs likely differ between virus families and host species. Moreover, the limited genome size and CpG underrepresentation complicates inference for small DNA viruses. However, certain patterns appear noticeable even without using formal statistical inference. The ratio of total stimulatory to total nonstimulatory motifs is higher than expected based on the viral nucleotide composition for Frog adenovirus, Frog virus 3 and herpes simplex virus type-1 (HSV1), while it is lower than expected for all other viruses for which this value can be calculated. Putatively stimulatory motifs are rarely present more frequently than expected based on the viral nucleotide composition, with the exception of HSV1 and Frog virus 3. Putatively nonstimulatory motifs, on the other hand, are present at higher frequency than expected in several cases, as seen for several adeno- and herpesviruses. The driving forces determining this nucleotide composition are difficult to disentangle. HSV1 induces mainly subclinical, persistent and recurrent infections but HSV1 DNA has been shown to efficiently stimulate TLR9 responses (49). The mammalian genomes examined contain a functional TLR9 ortholog, but chickens lack an orthologous TLR9 gene, and the same is likely true for most if not all avian species. Unfortunately, data on other avian species are thus far limited (50) and it is not known, in most cases, whether amphibian species have a functional TLR9. Several other factors likely influence presence and distribution of CpG-containing motifs, but many of these remain to be understood.

DNA VIRUSES AND CYTOSINE METHYLATION

To understand the impact of cytosine methylation on the viral life cycle and the evolution of base composition, the particularities of each virus will need to be considered. Differences will inevitably exist between actively replicating viral DNA and that which is integrated into the host genome. The type of viral persistence will also be of importance. The integration of adeno- or polyomavirus DNA into the host genome is usually a terminal process since the viruses cannot liberate their genomes and are therefore no longer infectious. The evolutionary roles of methylation in these cases will likely differ from those seen in other viruses, such as Herpesviruses, which can liberate their genome after periods of latency. But differences may also exist between large and small viruses—with many larger viruses encoding their own replication machinery and additional proteins which modify host cell processes and immune responses. The susceptibility of the viral genome to methylation and immune recognition will also be affected by other factors, such as the location of replication within the cell and the specific intracellular trafficking route (Figure 1).

Viruses that integrate into the host genome and the effects of methylation on their life cycles have long been of particular interest. Many studies have focused on

Table 2. Frequency of TLR9 stimulatory and non-stimulatory/inhibitory sequences

Virus	Host	Base frequencies					Stimulatory					Non-stimulatory					Ratio stimulatory- /non-stimulatory	Sequence length (nts)	Refseq number
		A (%)	C (%)	G (%)	T (%)	CpG	GpC	CpG/GpC	GACGTT	GTCGTT	GTCGCT	Total stimulatory	TTCCGA	GCGCGC	CCCCGG	Total non-stimulatory			
<i>Small ss DNA viruses</i>																			
FPV	Cat	1816 (35.44)	835 (16.3)	1036 (20.2)	1437 (28.04)	76 (169)	167 (169)	0.455 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	2 (0)	0 (0)	2 (1)	- (3)	5124	M38246 #
CPV	Dog	1923 (35.92)	839 (15.76)	1054 (19.8)	1507 (28.3)	74 (166)	161 (166)	0.46 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	1 (0)	0 (0)	1	- (3)	5323	NC_001539
AAV1	Human	1121 (23.76)	1393 (29.52)	1273 (26.98)	931 (19.73)	316 (376)	383 (276)	0.825 (1)	1 (1)	0 (1)	0 (1)	1 (3)	0 (1)	5 (2)	1 (3)	6 (6)	0.167 (0.5)	4718	NC_002077
AAV2	Human	1198 (25.6)	1262 (26.97)	1255 (28.82)	964 (20.6)	266 (364)	304 (364)	0.875 (1)	2 (1)	0 (1)	0 (1)	2 (3)	1 (1)	4 (2)	0 (2)	5 (5)	0.400 (0.6)	4679	NC_001401
AAV4	Human	1137 (23.85)	1405 (29.47)	1333 (27.96)	892 (18.7)	325 (392)	368 (392)	0.883 (1)	0 (1)	0 (1)	0 (1)	0 (3)	1 (1)	4 (3)	1 (3)	6 (7)	- (0.429)	4767	NC_001829
AAV5	Human	1167 (25.14)	1322 (28.48)	1238 (26.67)	915 (19.7)	287 (352)	320 (352)	0.897 (1)	0 (1)	0 (1)	0 (1)	0 (3)	1 (1)	1 (2)	1 (2)	3 (5)	- (0.6)	4642	NC_006152
<i>Small ds DNA viruses</i>																			
SV40	Human	1518 (28.95)	1100 (20.98)	1039 (19.82)	1586 (30.25)	27 (218)	266 (218)	0.102 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	0 (1)	0 (0)	0 (0)	- (3)	5243	NC_001669
JC polyoma-virus	human	1527 (29.77)	1035 (20.18)	1040 (20.27)	1528 (29.79)	16 (210)	255 210)	0.063 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	0 (0)	0 (0)	0 (1)	- (3)	5130	NC_001699
Murine polyoma-virus	Mouse	1396 (26.36)	1270 (23.98)	1233 (23.29)	1398 (26.4)	94 (295)	289 (295)	0.325 (1)	0 (1)	0 (1)	1 (1)	1 (3)	0 (1)	0 (1)	0 (1)	0 (3)	- (0.5)	5295	NC_001515
<i>Large ds DNA viruses</i>																			
HSV1	Human	24240 (15.92)	51458 (33.79)	52513 (34.49)	24050 (15.79)	17917 (17745)	16465 (17745)	1.088 (1)	31 (24)	37 (24)	49 (52)	117 (100)	24 (24)	236 (241)	297 (236)	557 (501)	0.21 (0.1996)	152261	NC_001806
Gallid herpes-virus 1	Chicken	38272 (25.74)	35807 (24.08)	35798 (24.08)	38810 (26.1)	8666 (8622)	9401 (8622)	0.922 (1)	25 (37)	22 (37)	31 (34)	78 (108)	45 (36)	74 (29)	40 (29)	159 (94)	0.49 (1.148)	148687	NC_006623
Human adeno-virus D	Human	7988 (22.76)	9877 (28.14)	9981 (28.44)	7250 (20.66)	2410 (2808)	3034 (2808)	0.794 (1)	1 (7)	2 (7)	8 (10)	11 (24)	2 (8)	36 (18)	22 (18)	60 (40)	0.183 (0.6)	35096	NC_002067
Simian adeno-virus 1	Non-human primate	7939 (23.04)	9677 (28.09)	9343 (27.11)	7491 (21.75)	2327 (2623)	2703 (2623)	0.861 (1)	3 (8)	7 (7)	10 (10)	20 (25)	6 (8)	11 (15)	19 (16)	36 (39)	0.556 (0.64)	34450	NC_006879
Fowl adeno-virus A	chicken	10284 (23.48)	12087 (27.59)	11699 (26.71)	9734 (22.22)	3332 (3228)	3109 (3228)	1.072 (1)	11 (10)	10 (10)	9 (12)	30 (32)	15 (11)	27 (18)	19 (18)	61 (47)	0.492 (0.68)	43804	NC_001720
Fowl adeno-virus D	Chicken	10485 (23.27)	12305 (27.3)	11931 (26.48)	10342 (22.95)	3857 (3258)	3178 (3258)	1.214 (1)	10 (11)	11 (10)	11 (12)	32 (33)	17 (11)	15 (17)	32 (18)	64 (46)	0.5 (0.72)	45063	NC_000899
Frog adeno-virus	Frog	7827 (29.9)	4693 (17.94)	5210 (19.91)	8433 (32.23)	509 (1011)	1017 (1011)	0.500 (1)	2 (6)	3 (7)	3 (4)	8 (17)	0 (6)	3 (1)	0 (1)	3 (8)	2.667 (2.1)	26163	NC_002501
Frog virus 3	Frog	24269 (22.92)	28815 (27.2)	29488 (27.84)	23331 (22.03)	5633 (8020)	6342 (8020)	0.888 (1)	31 (25)	14 (24)	11 (29)	56 (78)	10 (24)	7 (46)	25 (44)	42 (114)	1.333 (0.68)	105903	NC_005946

GpC and CpG contents, as well as individual nucleotide and motif frequencies, were determined from the publicly available reference sequences (RefSeq) in Genbank (MatLab script available from the authors upon request). Sequence length, frequencies and percentages of individual nucleotides, GpC and CpG content, as well as the frequencies of potentially stimulating and non-stimulating motifs, are provided.

Potentially stimulatory sequences are those described by Rankin *et al.* (120); values represent the total number of the respective motif present in the genome. The number in parentheses indicates the number of motifs expected in the sequence, given the individual nucleotide composition. This was calculated using the following formula: $E(UVWXYZ) = 10^{-12} * \text{sequence length} * (\%U * \%V * \%W * \%X * \%Y * \%Z)$, where $E(UVWXYZ)$ = expected value of the motif $UVWXYZ$; U, V, W, X, Y, Z = nucleotides of which the motif consists, and $\%U$ = % of viral sequence consisting of base U. Expected values were rounded to the nearest integer. Likely non-stimulatory and potentially inhibitory sequences are those described by Krieg; values again represent the total number of the respective motif present in the genome, and expected values were calculated as describe above. 'Total stimulatory' or 'total non-stimulatory' values represents the sum of all putatively stimulatory or non-stimulatory motifs in the sequence. 'Ratio of stimulatory/non-stimulatory sequences' represent the fraction of the total number of stimulatory divided by the total number of non-stimulatory sequences. 'CpG' or 'GpC' represents the total number of the respective dinucleotide CpG or GpC in the viral DNA sequence. 'CpG/GpC' represents fraction of the number of CpG motifs divided by the number of GpC motifs in the viral DNA. 'RefSeq' number represent GeneBank accession number. '#' no RefSeq FPV sequence available.

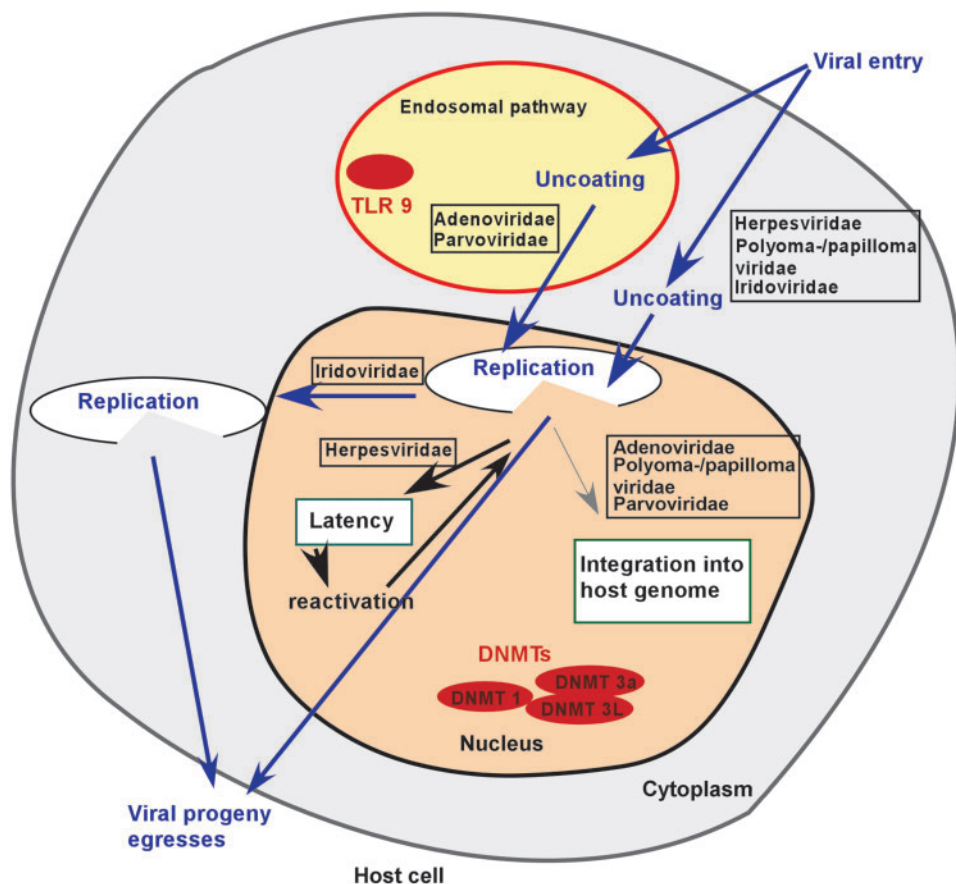


Figure 1. Comparison of DNA virus infection pathways. Major intracellular trafficking routes and characteristics of replication are shown for the DNA virus families discussed. The importance of methylation and immune recognition is indicated. Simplifications and generalizations were made for the purpose of clarity.

adenoviruses, but comprehensive knowledge of the methylation status of other viruses, such as polyomaviruses and herpesviruses, has also been obtained (Table 3).

Early studies of viral methylation focused on DNA from polyomavirus-infected cells (51). The early methods (e.g. quantification of methyl- H^3 incorporation into the genome) used to measure methylation were limited in their sensitivity and did not show the genomic distribution of methylated sites. Technical advances later made it possible to study cytosine methylation in a more detailed and site-specific manner. However, the various roles of methylation during persistent viral infections, in the silencing of viral genomes, and in immune evasion and tumor formation are still being uncovered.

The impact of methylation during active viral replication is generally incompletely understood, and for many small DNA viruses it is currently not known whether the viral genome is methylated during active replication.

Adenovirus DNA methylation: summarizing technical advances in viral methylation studies

Early studies of adenovirus methylation relied on chromatographic and radioactive techniques to distinguish methylated from unmethylated bases. The techniques were

complicated by limited sensitivity, the difficulty of isolating pure viral fractions and the inability to discriminate between host and integrated viral DNA. The investigators nevertheless carefully compared actively replicating and integrated viral DNA and their conclusions have stood the test of time. Genomic DNA from adenovirus type 2 and 12 infected cells was heavily methylated, more than cellular DNA from uninfected control cells (52). Whether the integration of viral DNA altered methylation patterns in the host genome or whether the observed differences were due exclusively to the integrated viral DNA was not determined. However, actively replicating adenovirus type 2 and 12 DNA appeared to have few or no methylated bases (52).

More sensitive analysis was allowed by pairs of methylation-sensitive and insensitive restriction endonucleases. In particular, the isoschizomers HpaII and MspI have been used frequently in studies of DNA methylation. Both enzymes recognize the motif CCGG, but the catalytic activity of HpaII is inhibited by CpG methylation while the activity of MspI is not affected. The new technique verified the previous results with actively replicating adenovirus type 2 and 12 DNA showing no detectable methylated CCGG sites and the restriction patterns were consistent among viral DNAs from

Table 3. Overview of GC content, CpG frequency and methylation status of small and large DNA viruses

Virus	Genome size ^a (kb)	GC frequency ^a	CpG content ^a (pCPG)	Methylation status during active replication		Methylation status during latency			Host species	Effect on host methylation
				Replicating	Reference ^b	Integrated	Episomal	Reference ^b		
Large dsDNA viruses										
Adenoviridae	28–45	0.3–0.65	0.5–1.13	Un/hypomethylated	(121)	Methylated	–	(121)	Mammal, bird	DNMT upregulation
Alpha-herpesvirinae	130–150	0.4–0.71	0.9–1.17	Un/hypomethylated	(67)	–	Un/hypo-methylated	(89)	Mammal, bird	DNMT upregulation
Beta-herpesvirinae	140–240	0.4–0.67	1.0–1.25	Unknown	N/A	–	Unknown	N/A	Mammal	Unknown
Gamma-herpesvirinae (122)	110–185	0.3–0.61	0.3–0.66	Un/hypomethylated	(66)	–	Methylated	(65)	Mammal	DNMT upregulation
Ranid herpesvirus ^c	220–230	0.5–0.55	0.8–0.95	Methylated	(70)	–	Unknown	N/A	Amphibian	Viral 5-cytosine methyltransferases?
Iridoviridae	140–383	0.2–0.56	0.5–0.84	Methylated	(74)	–	–	–	Amphibian, fish	Viral 5-cytosine methyltransferases?
Poxviridae	130–375	0.2–0.64	0.8–1.23	Unknown	N/A	–	–	–	Mammal bird invertebrate	Unknown
Small dsDNA viruses										
Papilloma-viridae	7–8	0.4–0.54	0.1–0.57	Partially methylated	(60)	Methylated	–	(123)	Mammal	DNMT upregulation
Polyoma-viridae	5	0.4–0.48	0.05–0.78	Un/hypomethylated	(51)	Methylated	–	(51)	Mammal, bird	DNMT upregulation
Small ssDNA viruses										
Autonomous Parvoviridae	4–6	0.3–0.5	0.3–0.71	Unknown	N/A	–	–	–	Mammal	Unknown
Dependo-virinae	4–6	0.4–0.58	0.6–1.03	Unknown	N/A	–	–	–	Mammal	Unknown
Circoviridae	2	0.5–0.57	0.4–0.87	Unknown	N/A	–	–	–	Mammal, bird	Unknown
Anellovirus	4	0.5	0.67	Unknown	N/A	–	–	–	Human	Unknown

The GC content, CpG content and, where known, methylation status is shown for viral families/subfamilies. Where applicable, a distinction is made between active replication and latency. If latent, the state of the genome (i.e. integrated or episomal) is specified. Preferred host species are indicated along with any known effect the virus has on host cell methylation. Where relevant, representative references are provided.

Dash indicates this form is not known to occur for this virus.

^aGC and CpG contents as determined by authors [here or in (83)] according to methods described therein; GC content represents the relative frequency of G and C in the sequence. CpG content represents the observed divided by the expected frequency of the dinucleotide CpG in the sequence.

^bReference refers to a representative study, which focuses on one member of the virus family.

^cUnpublished results, analysis done by authors as described above (83).

different cells (53). Inference from restriction enzyme analyses, however, remained limited to those sites recognizable by restriction enzyme pairs. Several subsequent studies focused on adenovirus DNA and the absence of methylated sites could later be conclusively established through bisulfate sequencing (51–53). Bisulfate converts unmethylated cytosines to thymines and thus allows the sensitive detection of specific methylated sites by sequencing the genome prior to and after treatment. The absence of methylated sites in the replicating adenovirus 2 genome has been shown using this technique (54–56).

Many studies of adenovirus DNA have focused on integrated viral genomes or genomic segments, largely because of their oncogenic potential. Integrated adenoviral DNA is known to be hypermethylated, but the regulatory events which determine the methylation status of replicating or integrated sequences are not well understood. A recombinant between adenovirus 12 and host cell DNA (referred to as SYREC) is unmethylated over the whole genome during active replication including the host-derived sequence, which is methylated in the chromosome. This led to speculations about the existence of adenovirus-encoded de-methylation proteins, or a role for the virus-encoded replication machinery in methylation avoidance (57). However, no specific mechanisms have been defined, and the speed of viral replication or compartmentalization of this process within the host cell has also been proposed as factors contributing to the lack of methylation.

Methylation status of polyoma- and papillomaviruses

Studies of papovavirus DNA methylation also focused on integrated sequences, mostly of human papillomavirus strains and SV40 polyomavirus. Polyomavirus and papillomavirus genomes are CpG-depleted (Table 3) and any CpG motifs present are clustered within certain regions of the viral genome (58), complicating the analysis of methylation patterns. The early studies, however, concluded that there was little methylation in replicating polyomaviruses (51), with the possible exception of a CpG site near the early promoter (59). Several other studies concurred and found several human papillomavirus strains to be hypomethylated during active replication, again with methylated sites clustering in specific regions of the viral genome. The integrated viral DNA, on the contrary, appeared heavily methylated. However, distinguishing between actively replicating and integrated DNA appears difficult in some cases (60–62).

Methylation status of herpesvirus genomes

Herpesviruses establish latent infections without integrating into the host genome, and reactivate to re-establish active replication. During latency, the circularized viral genome remains quiescent in an episomal state, and can be replicated by the host cell replication machinery. Active (e.g. lytic) viral replication differs in several respects. Production of viral progeny inevitably results in cell lysis, utilizes the viral-encoded replication machinery and involves a separate origin of replication [e.g. in the case

of Epstein–Barr Virus (EBV) *oriLyt* functions as the origin of replication in the lytic cycle while *oriP* is responsible for replication of episomal EBV DNA (63)]. The methylation status during lytic DNA replication is likely different from that during latency, but most studies have focused on latent infections and the differences have rarely been studied explicitly.

Differences might also exist among herpesvirus sub-families, as CpG motifs are highly underrepresented in gammaherpesvirus genomes, but not in alpha- or beta-herpesvirus genomes (64). The reasons for these differences remain elusive. Gammaherpesviruses cause predominantly persistent, lymphoproliferative diseases. Studies of some gammaherpesviruses, such as EBV and Kaposi's Sarcoma-associated herpesvirus (KSHV), indicate that methylation serves an intricate regulatory role during the viral life cycle (see below). The EBV genome appears to be hypomethylated or unmethylated during lytic infection, but highly methylated during latency, and appears to become demethylated during reactivation (65). There is some evidence for methylation during active replication of another gammaherpesvirus, herpesvirus saimiri, since Kaschka-Dierich *et al.* (66) found both the linear and circular viral DNA to be heavily methylated. The alphaherpesvirus herpes simplex virus (HSV) appears unmethylated during active replication and latency (67–69) while two frog herpesviruses (ranid herpesvirus 1 and 2) appear to be methylated during replication. The ranid herpesvirus genomes also contain putative DNA cytosine-5 methyltransferases sequences located in ORFs 86 and 120 (70).

Methylation plays a pivotal role in the life cycle of EBV (71,72), where methylation of specific genes appears to be involved in the transition from lytic to latent infection. Most EBV nuclear antigens (EBNAs) are expressed only for short periods during the lytic infection, after which their transcription is silenced through hypermethylation of the viral promoters Cp and Wp. Production of the indispensable EBNA1 is then achieved by expression from an alternative 'TATA-less' promoter Qp. Reinitiation of lytic infection can occur spontaneously or may be triggered by events such as immunoglobulin crosslinking of the host cell and cell proliferation. Reactivation is mediated through the expression of the immediate early genes Zta and Rta, which are regulated by the promoters Zp and Rp, the latter of which is hypermethylated during latency. While Rp hypermethylation is a major regulator of latent infection, Zp hypermethylation is dispensable. EBV therefore seems to use methylation-induced gene silencing as an immune evasion strategy. While a role for methylation in the EBV lifecycle is widely accepted, its importance for other herpesviruses is still disputed.

Methylation status of iridiviruses and ascoviruses

The genomes of iridoviruses appear to be heavily methylated during active replication and methylation might again have a regulatory role. Iridoviruses are large DNA viruses that infect fish, amphibians and reptiles. Initial stages of replication occur in the nucleus, after which the nascent viral genomes are transported into

the cytoplasm where the second stage of replication occurs (73). Willis and Granoff (74) analyzed the iridovirus frog virus 3 (FV3) by radioactive labeling/restriction enzyme digestion and found the viral DNA to be heavily methylated, with an estimated 20% of cytosines methylated. Time-course experiments later indicated changing methylation patterns over the course of infection. Willis *et al.* (75) analyzed nuclear and cytoplasmic extracts collected at various times postinfection (p.i.) and found methylated viral DNA after ~6–7h p.i. This DNA appeared to be located in the cytoplasm, while nuclear DNA, collected at earlier time points, appeared to be hypomethylated. The authors also provided evidence against demethylation of parental viral genomes upon host cell infection. Further, Schetter *et al.* (76) showed that nascent FV3 DNA in the nucleus is unmethylated or hypomethylated at early times p.i, but later becomes hypermethylated (after ~6h p.i. in their assay). Cytoplasmic viral DNA, on the other hand, appears to be methylated at all times. The authors reported DNA methyltransferase activity in nuclear extracts of FV3 infected cells, beginning at a time p.i. which coincided with the appearance of methylated viral genomes. This activity was absent from cytoplasmic extracts, indicating FV3 might regulate methylation of its genome. It was later discovered that the FV3 genome encodes a putative DNA cytosine-5 methyltransferase sequence, which bears similarities to other known methyltransferases (77). The putative methyltransferase, however, is more homologous to prokaryotic enzymes than to eukaryotic methyltransferases. While transcription (from an early promoter) and translation of this FV3 protein have been verified, direct evidence for its methylation activity is still limited.

Another iridovirus, fish lymphocystis disease virus (FLDV), was also found to have a highly methylated genome, even though its GC content is considerably lower than that of FV3 (78,79). FLDV also contains a putative 5-cytosine methyltransferase (53.3% identical to that of FV3) (80) but evidence for its function is, thus far, scarce. Besides the frog herpesviruses and iridoviruses, ascoviruses, which infect insects and are closely related to the iridoviruses, are also heavily methylated during replication (81).

Methylation status of small DNA viruses

In contrast to most large DNA viruses, CpG is underrepresented in the majority of small DNA viruses (64,82,83). It is possible that cytosine methylation is partly responsible for this underrepresentation. Small DNA viruses encode only a minimal number of proteins and use the host cell replication machinery for replication. Low CpG frequencies may have been selected to avoid methylation by host methyltransferases, to maximize translation efficiencies, or as a means of reducing CpG mediated immune responses. However, apart from the small polyomaviruses and papillomaviruses described above, there appears to have been few close examinations of the methylation status of small DNA viruses (particularly those that do not integrate), such as autonomous parvoviruses, circoviruses and anelloviruses. The roles and

effects of methylation might be different from those seen in integrating and/or large DNA viruses, but further research is needed to understand the specific effects of methylation on small, nonintegrating DNA viruses.

Adeno-associated virus (AAV) integrates into the host genome, and has been examined primarily in the context of guanine methylation. Productive infection with an AAV normally requires co-infection with a helper virus (usually an adenovirus or, less frequently, a herpesvirus). In the absence of a helper virus, AAVs may remain in the cell as persistent dsDNA, or they may be stably integrated into the host genome. Upon infection by a helpervirus, the integrated AAV DNA can be rescued and undergo active replication. The CpG contents of AAVs more closely resemble those of their helper viruses than those of the closely related autonomous parvoviruses, which show strong CpG suppression. The mechanisms of integration have been extensively studied in AAV serotype 2 (AAV2), which predominantly integrates in a specific region of chromosome 19 (locus 19q13.3t). Integration is directed by AAV-encoded Rep proteins, which likely recognize guanine residues within GCTC repeating motifs (located in the AAV terminal hairpin). *In vitro* methylation of these guanine residues has been shown to effectively inhibit DNA–protein interactions and likely interferes with integration *in vivo* (84). Similar results were obtained when examining the integration of AAV4 (which utilizes a slightly different recognition motif) into African green monkey cells (85). Nevertheless, it is not known what role, if any, cytosine-5 methylation plays in the natural AAV life cycle, if AAV methylation influences integration, or if guanines or cytosines are methylated during active replication.

METHYLATION AND INTEGRATED OR LATENT VIRAL DNA

Methylation of integrated viral DNA has been studied extensively and the topic is extensively covered in the scientific literature. Areas of interest have included changes in methylation patterns during integration, the effects of viral methylation on host cells and the correlation between virally induced methylation and tumor formation. Excellent reviews have been published on these topics and the reader is referred to these articles, indicated in the following discussion, for additional detail. Here we briefly summarize what is known about methylation during integration and latency and the regulatory elements involved, contrasting the level of understanding to what is known about methylation during active replication.

As already mentioned above, most DNA viruses such as adenoviruses (57,59,86–88) and polyomaviruses (83) appear specifically methylated during latency or when in an integrated form. In contrast, the alphaherpesvirus HSV does not appear to be methylated during latency (89), but the reasons for this difference are unclear. The processes by which integrating viruses are *de novo* methylated are only partly understood, but it is clear that the extent and pattern of *de novo* methylation depend upon the time after

integration and site in the host genome (90), the integrated sequences (90–94), and possibly the cell type (57,90). *De novo* methylation tends to be initiated within confined genomic regions and then spreads throughout the viral genome (90,93,95,87). For example, in a study examining the methylation pattern of adenovirus 12 over the course of integration into hamster cells, methylation began in the center of the viral DNA (between map units 30 and 75) and subsequently spread outward in both directions. Some regions of the viral genome remained unmethylated throughout the study, namely the right terminal repeat and regions near the left terminus (95). It is not known which factors initiate methylation of the genome, nor the determinants of the sequential methylation pattern.

Methylation patterns appear to be site specific. In many viruses there seems to be an inverse correlation between viral gene expression and degree of methylation, with late viral genes being generally more susceptible to methylation effects than early genes (53,91). Promoter regions of integrated viruses are frequently hypomethylated. For example, promoter regions in adenovirus 12 are hypomethylated at CCGG sites while the viral genome is otherwise heavily methylated (57,59,96). Methylation of viral genes, particularly at promoter sequences, appears to be a reversible event (96) but the mechanism by which viral genomes are demethylated remains to be determined. Demethylation has also been shown to occur after *in vitro* methylation of several viral DNAs. For example, SV40 DNA recovered 20 h after microinjection into cells retained the artificial methylation pattern under nonpermissive culture conditions, while allowing early gene transcription. The DNA, however, became demethylated after viral replication (97). Loss of methylation has also been observed in adenoviruses, in the case of the SYRECs mentioned above (86,98) which are methylated when covalently linked to the host genome, but unmethylated during active replication. They require the presence of fully functional adenovirus helper for efficient lytic replication, which led some to postulate a role for the viral transcription machinery in the loss of methylation. Unfortunately, experimental data in support of this hypothesis are so far scarce.

Infection with integrating (and even some nonintegrating) viruses can induce alterations in the methylation patterns of host DNA, a potential factor in the development of malignant tumors. For example, herpesviruses KSHV and EBV, polyomaviruses BK and SV40, papillomavirus HPV, several adenoviruses and hepatitis B virus encode proteins that activate/upregulate DNMT1, 3a, or 3b. This results in the hypermethylation and, therefore, the downregulation of a number of cellular genes. One of the cellular genes frequently hypermethylated in these infections is the cell cycle regulating tumor suppressor gene p16INK4a, which is commonly hypermethylated in many types of cancer (99). Differentiation of host cell tumors and the level of genome methylation of integrated viruses generally appear to be positively correlated, as seen for various papillomaviruses such as Shope papillomavirus (100) and human papillomavirus (HPV) (101,102), the herpesvirus EBV (72,103) and various adenoviruses (96).

METHYLATION AND VIRAL GENE EXPRESSION

For iridoviruses, gene expression at functional levels can be maintained despite genome hypermethylation. Methylated sites, however, are not distributed randomly across the genome and methylation levels differ among CpG motifs. FLDV, for example, has a considerably higher level of methylation at CCGG than CGCG motifs (79). The physiological significance of this, however, is still disputed. *In vitro* studies of the related FV3 showed that specific methylation (using HpaII methyltransferase) of CCGG motifs in a late promoter, L1140, abolished its function. However, the methylation of all CpG sites by the indiscriminate SssI methyltransferase or the specific methylation of GCGC motifs by HhaI methyltransferase had no dramatic effect. This appears to be a rare example where, counterintuitively, complete methylation of a late promoter does not abolish its function (methylation frequently abolishes the function of late viral promoters, while early promoters tend to be less susceptible to the repressive effect of cytosine methylation). The underlying mechanism has not been elucidated, but steric hindrance in asymmetrically methylated DNA might play a role (104). It seems possible that methylation of CCGG motifs results in complex secondary structures in the promoter region, whereas indiscriminate CpG methylation or methylation of GCGC motifs balances these secondary structures. A similar phenomenon was seen in adenovirus type 2, where methylation of CCGG sites within the (late) E2a promoter inhibited the transcription of viral DNA, while methylation of GCGC motifs had no repressive effect (105,106). The transcription activity of the polyomavirus SV40 was found not drastically decreased when *in vitro* methylated with indiscriminate DNMTs from rat liver (107,108). In contrast, upon specific methylation of HpaII sites (CCGG-specific), transcription activity was reduced markedly (59). However, the interpretation of these latter results is complicated by the possibility of incomplete methylation by the rat liver extract, as rat liver DNA is a competitive inhibitor of adenovirus methylation (55) and low-level contaminants might likewise have inhibited complete methylation of SV40.

In vitro methylation studies involving human papillomavirus type 16 (HPV16) showed that indiscriminate SssI methylation significantly decreased the transcriptional activity of the long control region (LCR), which is devoid of ORFs but contains several *cis*-acting regulatory elements. Enhancers located in the LCR are responsive to both cellular and viral factors, such as the viral E2 gene products. Transcriptional silencing was measured using an LCR-containing reporter plasmid and inhibition was again believed to be steric, specifically due to interference with the binding of transcription regulator E2. Notably, hypomethylation of HPV16 E2-binding sites has been observed in highly differentiated, but not in undifferentiated, tumor cells (101). However, direct comparison of the HPV16 and SV40 studies is difficult. The effects of indiscriminate methylation on early and late viral gene products in the genomic context are hard to infer from this study.

The time of transcription appears to impact the effects of methylation in several viruses. For example, the *in vitro* HpaII (CCGG-specific) methylation of SV40 early genes (i.e. those coding for the large T-antigen) does not lead to transcriptional repression, and the *in vitro* methylated sites are subsequently lost during active replication (108). Late viral gene expression, however, can be efficiently inhibited by *in vitro* methylation of HpaII sites within the 5' part of this region (109). Site-specific methylation might be involved in the transition to latency in some cases (as in the case of EBV). On the other hand, methylation does not seem to be involved in the switch from early to late gene expression in HSV (89). Likewise, the major late promoter of adenovirus 2, which induces the switch from early to late gene expression, is unmethylated in nonintegrated viruses (56), suggesting methylation does not play a major role in the regulation of early versus late gene expression for this virus.

CONCLUSIONS

A number of studies show that CpG methylation can greatly affect the life cycles of DNA viruses, but its exact role in natural infections remains unclear. Those effects likely differ between different viruses and are dependent upon many factors, such as the stage in the viral life cycle, the host species and infected tissue, the flanking nucleotide motifs, the genes in question and the genomic location. Where it has been studied, major differences have been seen for some viruses between the actively replicating and the latent or integrated viral DNA; the former is often unmethylated or hypomethylated, while the later often shows specific and regulated methylation. Differences may also exist between large and small DNA viruses, potentially because of their rate and site of replication and differences in their ability to manipulate the host responses to unmethylated DNA. Thus, it is necessary to not only examine individual viruses, but to examine their dynamic methylation status and the effects of any viral methylation on both the virus and the host cell. The various ways in which mutational bias, codon usage and host selections (e.g. through TLR9-dependent responses) influence CpG frequency and methylation status need to be defined. It will be especially important to determine the unique roles and regulators of methylation, and their divergent effects on the genome evolution of small and large DNA viruses during active or latent infection, and in packaged, integrated or episomal DNA. A deeper knowledge of the relationships among the virus, the host cell and methylation (and de-methylation) machinery will provide essential insights into determinants of methylation status, gene expression, replicative behavior, as well as activation of pattern recognition receptors and immune responses.

ACKNOWLEDGEMENTS

Funding to pay the Open Access publication charges for this article was provided by National Institutes of Health Grant R01 GM080533 to ECH.

REFERENCES

- Bird, A.P. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.*, **8**, 1499–1504.
- Schorderet, D.F. and Gartler, S.M. (1992) Analysis of CpG suppression in methylated and nonmethylated species. *Proc. Natl Acad. Sci. USA*, **89**, 957–961.
- Swartz, M.N., Trautner, T.A. and Kornberg, A. (1962) Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J. Biol. Chem.*, **237**, 1961–1967.
- Burge, C., Campbell, A.M. and Karlin, S. (1992) Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl Acad. Sci. USA*, **89**, 1358–1362.
- Sponer, J., Gabb, H.A., Leszczynski, J. and Hobza, P. (1997) Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophys. J.*, **73**, 76–87.
- Andersson, S.G. and Kurland, C.G. (1990) Codon preferences in free-living microorganisms. *Microbiol. Rev.*, **54**, 198–210.
- Krieg, A.M. (2002) A role for toll in autoimmunity. *Nat. Immunol.*, **3**, 423–424.
- Bachmann, M.F. and Kopf, M. (2001) On the role of the innate immunity in autoimmune disease. *J. Exp. Med.*, **193**, F47–F50.
- Janeway, C.A., Jr. and Medzhitov, R. (2002) Innate immune recognition. *Ann. Rev. Immunol.*, **20**, 197–216.
- Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, **274**, 775–780.
- Sved, J. and Bird, A. (1990) The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc. Natl Acad. Sci. USA*, **87**, 4692–4696.
- Fryxell, K.J. and Zuckerkandl, E. (2000) Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.*, **17**, 1371–1383.
- Kass, S.U., Pruss, D. and Wolffe, A.P. (1997) How does DNA methylation repress transcription? *Trends Genet.*, **13**, 444–449.
- Ng, H.H. and Bird, A. (1999) DNA methylation and chromatin modification. *Curr. Opin. Genet. Dev.*, **9**, 158–163.
- Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA*, **90**, 11995–11999.
- McClelland, M. and Ivarie, R. (1982) Asymmetrical distribution of CpG in an 'average' mammalian gene. *Nucleic Acids Res.*, **10**, 7865–7877.
- Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA*, **99**, 3740–3745.
- Caiafa, P. and Zampieri, M. (2005) DNA methylation and chromatin structure: the puzzling CpG islands. *J. Cell. Biochem.*, **94**, 257–265.
- Pradhan, S. and Esteve, P.O. (2003) Mammalian DNA (cytosine-5) methyltransferases and their expression. *Clin. Immunol.*, **109**, 6–16.
- Hermann, A., Gowher, H. and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell. Mol. Life Sci.*, **61**, 2571–2587.
- Bird, A.P. and Wolffe, A.P. (1999) Methylation-induced repression—belts, braces, and chromatin. *Cell*, **99**, 451–454.
- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A. and Jones, P.A. (1999) The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res.*, **27**, 2291–2298.
- Golding, M.C. and Westhusin, M.E. (2003) Analysis of DNA (cytosine 5) methyltransferase mRNA sequence and expression in bovine preimplantation embryos, fetal and adult tissues. *Gene Expression Patterns*, **3**, 551–558.
- Mizuno, S., Chijiwa, T., Okamura, T., Akashi, K., Fukumaki, Y., Niho, Y. and Sasaki, H. (2001) Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia. *Blood*, **97**, 1172–1179.
- Salozhin, S.V., Prokhorchuk, E.B. and Georgiev, G.P. (2005) Methylation of DNA—one of the major epigenetic markers. *Biochemistry*, **70**, 525–532.

27. Reik,W., Dean,W. and Walter,J. (2001) Epigenetic reprogramming in mammalian development. *Science*, **293**, 1089–1093.
28. Fulka,H., Mrazek,M., Tepla,O. and Fulka,J., Jr. (2004) DNA methylation pattern in human zygotes and developing embryos. *Reproduction*, **128**, 703–708.
29. Aranyi,T. and Paldi,A. (2006) The constant variation: DNA methylation changes during preimplantation development. *FEBS Lett.*, **580**, 6521–6526.
30. Jaenisch,R. and Bird,A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33** (Suppl.), 245–254.
31. Goll,M.G. and Bestor,T.H. (2005) Eukaryotic cytosine methyltransferases. *Ann. Rev. Biochem.*, **74**, 481–514.
32. Attwood,J.T., Yung,R.L. and Richardson,B.C. (2002) DNA methylation and the regulation of gene transcription. *Cell. Mol. Life Sci.*, **59**, 241–257.
33. Singal,R. and Ginder,G.D. (1999) DNA methylation. *Blood*, **93**, 4059–4070.
34. Hendrich,B. and Tweedie,S. (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.*, **19**, 269–277.
35. Slotkin,R.K. and Martienssen,R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.*, **8**, 272–285.
36. Robertson,K.D. and Wolffe,A.P. (2000) DNA methylation in health and disease. *Nat. Rev. Genet.*, **1**, 11–19.
37. Rothenfusser,S., Hornung,V., Krug,A., Towarowski,A., Krieg,A.M., Endres,S. and Hartmann,G. (2001) Distinct CpG oligonucleotide sequences activate human gamma delta T cells via interferon-alpha/-beta. *Eur. J. Immunol.*, **31**, 3525–3534.
38. Kojima,Y., Xin,K.-Q., Ooki,T., Hamajima,K., Oikawa,T., Shinoda,K., Ozaki,T., Hoshino,Y., Jounai,N., Nakazawa,M. *et al.* (2002) Adjuvant effect of multi-CpG motifs on an HIV-1 DNA vaccine. *Vaccine*, **20**, 2857–2865.
39. Krieg,A.M., Yi,A.K., Matson,S., Waldschmidt,T.J., Bishop,G.A., Teasdale,R., Koretzky,G.A. and Klinman,D.M. (1995) CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature*, **374**, 546–549.
40. Krug,A., Towarowski,A., Britsch,S., Rothenfusser,S., Hornung,V., Bals,R., Giese,T., Engelmann,H., Endres,S., Krieg,A.M. *et al.* (2001) Toll-like receptor expression reveals CpG DNA as a unique microbial stimulus for plasmacytoid dendritic cells which synergizes with CD40 ligand to induce high amounts of IL-12. *Eur. J. Immunol.*, **31**, 3026–3037.
41. Krieg,A.M. (2002) CpG motifs in bacterial DNA and their immune effects. *Ann. Rev. Immunol.*, **20**, 709–760.
42. Hartmann,G. and Krieg,A.M. (2000) Mechanism and function of a newly identified CpG DNA motif in human primary B cells. *J. Immunol.*, **164**, 944–953.
43. Mutwiri,G., Pontarollo,R., Babiuk,S., Griebel,P., van Drunen Littel-van den Hurk,S., Mena,A., Tsang,C., Alcon,V., Nichani,A., Ioannou,X. *et al.* (2003) Biological activity of immunostimulatory CpG DNA motifs in domestic animals. *Vet. Immunol. Immunopathol.*, **91**, 89–103.
44. Iho,S., Yamamoto,T., Takahashi,T. and Yamamoto,S. (1999) Oligodeoxynucleotides containing palindrome sequences with internal 5'-CpG-3' act directly on human NK and activated T cells to induce IFN-gamma production in vitro. *J. Immunol.*, **163**, 3642–3652.
45. Kanellos,T.S., Sylvester,I.D., Butler,V.L., Ambali,A.G., Partidos,C.D., Hamblin,A.S. and Russell,P.H. (1999) Mammalian granulocyte-macrophage colony-stimulating factor and some CpG motifs have an effect on the immunogenicity of DNA and subunit vaccines in fish. *Immunology*, **96**, 507–510.
46. Han,J., Hsu,C., Zhu,Z., Longshore,J.W. and Finley,W.H. (1994) Over-representation of the disease associated (CAG) and (CGG) repeats in the human genome. *Nucleic Acids Res.*, **22**, 1735–1740.
47. Lenert,P., Stunz,L., Yi,A.K., Krieg,A.M. and Ashman,R.F. (2001) CpG stimulation of primary mouse B cells is blocked by inhibitory oligodeoxyribonucleotides at a site proximal to NF-kappaB activation. *Antisense Nucleic Acid Drug Dev.*, **11**, 247–256.
48. Krieg,A.M., Wu,T., Weeratna,R., Efler,S.M., Love-Homan,L., Yang,L., Yi,A.K., Short,D. and Davis,H.L. (1998) Sequence motifs in adenoviral DNA block immune activation by stimulatory CpG motifs. *Proc. Natl Acad. Sci. USA*, **95**, 12631–12636.
49. Krug,A., Luker,G.D., Barchet,W., Leib,D.A., Akira,S. and Colonna,M. (2004) Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9. *Blood*, **103**, 1433–1437.
50. Roach,J.C., Glusman,G., Rowen,L., Kaur,A., Purcell,M.K., Smith,K.D., Hood,L.E. and Aderem,A. (2005) The evolution of vertebrate Toll-like receptors. *Proc. Natl Acad. Sci. USA*, **102**, 9577–9582.
51. Winocour,E., Kaye,A.M. and Stollar,V. (1965) Synthesis and transmethylation of DNA in polyoma-infected cultures. *Virology*, **27**, 156–169.
52. Gunthert,U., Schweiger,M., Stupp,M. and Doerfler,W. (1976) DNA methylation in adenovirus, adenovirus-transformed cells, and host cells. *Proc. Natl Acad. Sci. USA*, **73**, 3923–3927.
53. Vardimon,L., Neumann,R., Kuhlmann,I., Sutter,D. and Doerfler,W. (1980) DNA methylation and viral gene expression in adenovirus-transformed and -infected cells. *Nucleic Acids Res.*, **8**, 2461–2473.
54. Wienhues,U. and Doerfler,W. (1985) Lack of evidence for methylation of parental and newly synthesized adenovirus type 2 DNA in productive infections. *J. Virol.*, **56**, 320–324.
55. von Acken,U., Simon,D., Grunert,F., Doring,H.P. and Kroger,H. (1979) Methylation of viral DNA in vivo and in vitro. *Virology*, **99**, 152–157.
56. Kammer,C. and Doerfler,W. (1995) Genomic sequencing reveals absence of DNA methylation in the major late promoter of adenovirus type 2 DNA in the virion and in productively infected cells. *FEBS Lett.*, **362**, 301–305.
57. Doerfler,W. (2005) On the biological significance of DNA methylation. *Biochemistry*, **70**, 505–524.
58. Soeda,E., Arrand,J.R., Smolar,N. and Griffin,B.E. (1979) Sequence from early region of polyoma virus DNA containing viral replication origin and encoding small, middle and (part of) large T antigens. *Cell*, **17**, 357–370.
59. Doerfler,W. (1981) DNA methylation—a regulatory signal in eukaryotic gene expression. *J. Gen. Virol.*, **57**, 1–20.
60. Danos,O., Katinka,M. and Yaniv,M. (1980) Molecular cloning, refined physical map and heterogeneity of methylation sites of papilloma virus type 1a DNA. *Eur. J. Biochem./FEBS*, **109**, 457–461.
61. Danos,O., Katinka,M. and Yaniv,M. (1982) Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. *EMBO J.*, **1**, 231–236.
62. Burnett,T.S. and Sleeman,J.P. (1984) Uneven distribution of methylation sites within the human papillomavirus 1a genome: possible relevance to viral gene expression. *Nucleic Acids Res.*, **12**, 8847–8860.
63. Hammerschmidt,W. and Sugden,B. (1988) Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein–Barr virus. *Cell*, **55**, 427–433.
64. Takacs,M., Segesdi,J., Balog,K., Mezei,M., Toth,G. and Minarovits,J. (2001) Relative deficiency in CpG dinucleotides is a widespread but not unique feature of Gammaherpesvirinae genomes. *Acta Microbiol. Immunol. Hungarica*, **48**, 349–357.
65. Szyf,M., Eliasson,L., Mann,V., Klein,G. and Razin,A. (1985) Cellular and viral DNA hypomethylation associated with induction of Epstein–Barr virus lytic cycle. *Proc. Natl Acad. Sci. USA*, **82**, 8090–8094.
66. Kaschka-Dierich,C., Bauer,I., Fleckenstein,B. and Desrosiers,R.C. (1981) Episomal and nonepisomal herpesvirus DNA in lymphoid tumor cell lines. *Haematol. Blood Transfusion*, **26**, 197–203.
67. Low,M., Hay,J. and Keir,H.M. (1969) DNA of herpes simplex virus is not a substrate for methylation in vivo. *J. Mol. Biol.*, **46**, 205–207.
68. Lundberg,P., Welander,P., Han,X. and Cantin,E. (2003) Herpes simplex virus type 1 DNA is immunostimulatory in vitro and in vivo. *J. Virol.*, **77**, 11158–11169.
69. Kubat,N.J., Tran,R.K., McAnay,P. and Bloom,D.C. (2004) Specific histone tail modification and NOT DNA methylation is a determinant of herpes simplex virus type 1 latent gene expression. *J. Virol.*, **78**, 1139–1149.

70. Davison, A.J., Cunningham, C., Sauerbier, W. and McKinnell, R.G. (2006) Genome sequences of two frog herpesviruses. *J. Gen. Virol.*, **87**, 3509–3514.
71. Ambinder, R.F., Robertson, K.D. and Tao, Q. (1999) DNA methylation and the Epstein–Barr virus. *Semin. Cancer Biol.*, **9**, 369–375.
72. Tao, Q. and Robertson, K.D. (2003) Stealth technology: how Epstein–Barr virus utilizes DNA methylation to cloak itself from immune detection. *Clin. Immunol.*, **109**, 53–63.
73. Chinchar, V.G. (2002) Ranaviruses (family Iridoviridae): emerging cold-blooded killers. *Arch. Virol.*, **147**, 447–470.
74. Willis, D.B. and Granoff, A. (1980) Frog virus 3 DNA is heavily methylated at CpG sequences. *Virology*, **107**, 250–257.
75. Willis, D.B., Goorha, R. and Granoff, A. (1984) DNA methyltransferase induced by frog virus 3. *J. Virol.*, **49**, 86–91.
76. Schetter, C., Grunemann, B., Holker, I. and Doerfler, W. (1993) Patterns of frog virus 3 DNA methylation and DNA methyltransferase activity in nuclei of infected cells. *J. Virol.*, **67**, 6973–6978.
77. Kaur, K., Rohozinski, J. and Goorha, R. (1995) Identification and characterization of the frog virus 3 DNA methyltransferase gene. *J. Gen. Virol.*, **76**(Pt 8), 1937–1943.
78. Darai, G., Anders, K., Koch, H.G., Delius, H., Gelderblom, H., Samalecos, C. and Flugel, R.M. (1983) Analysis of the genome of fish lymphocystis disease virus isolated directly from epidermal tumours of pleuronectes. *Virology*, **126**, 466–479.
79. Wagner, H., Simon, D., Werner, E., Gelderblom, H., Darai, C. and Flugel, R.M. (1985) Methylation pattern of fish lymphocystis disease virus DNA. *J. Virol.*, **53**, 1005–1007.
80. Tidona, C.A., Schnitzler, P., Kehm, R. and Darai, G. (1996) Identification of the gene encoding the DNA (cytosine-5) methyltransferase of lymphocystis disease virus. *Virus Genes*, **12**, 219–229.
81. Bigot, Y., Stasiak, K., Rouleux-Bonnin, F. and Federici, B.A. (2000) Characterization of repetitive DNA regions and methylated DNA in ascovirus genomes. *J. Gen. Virol.*, **81**, 3073–3082.
82. Karlin, S., Ladunga, I. and Blaisdell, B.E. (1994) Heterogeneity of genomes: measures and values. *Proc. Natl Acad. Sci. USA*, **91**, 12837–12841.
83. Shackelton, L.A., Parrish, C.R. and Holmes, E.C. (2006) Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.*, **62**, 551–563.
84. Weitzman, M.D., Kyostio, S.R., Kotin, R.M. and Owens, R.A. (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl Acad. Sci. USA*, **91**, 5808–5812.
85. Amis, T.J., McCarty, D.M., Skulimowski, A. and Samulski, R.J. (2003) Identification and characterization of an adeno-associated virus integration site in CV-1 cells from the African green monkey. *J. Virol.*, **77**, 1904–1915.
86. Doerfler, W., Kruczek, I., Eick, D., Vardimon, L. and Kron, B. (1983) DNA methylation and gene activity: the adenovirus system as a model. *Cold Spring Harbor Symp. Quant. Biol.*, **47**(Pt 2), 593–603.
87. Doerfler, W. (1993) Adenoviral DNA integration and changes in DNA methylation patterns: a different view of insertional mutagenesis. *Prog. in Nucleic Acid Res. Mol. Biol.*, **46**, 1–36.
88. Doerfler, W. (1996) A new concept in (adenoviral) oncogenesis: integration of foreign DNA and its consequences. *Biochim. Biophys. Acta*, **1288**, F79–F99.
89. Kubat, N.J., Tran, R.K., McAnany, P. and Bloom, D.C. (2004) Specific histone tail modification and not DNA methylation is a determinant of herpes simplex virus type 1 latent gene expression. *J. Virol.*, **78**, 1139–1149.
90. Orend, G., Knoblauch, M., Kammer, C., Tjia, S.T., Schmitz, B., Linkwitz, A., Meyer, G., Maas, J. and Doerfler, W. (1995) The initiation of de novo methylation of foreign DNA integrated into a mammalian genome is not exclusively targeted by nucleotide sequence. *J. Virol.*, **69**, 1226–1242.
91. Sutter, D. and Doerfler, W. (1980) Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. *Proc. Natl Acad. Sci. USA*, **77**, 253–256.
92. Kruczek, I. and Doerfler, W. (1982) The unmethylated state of the promoter/leader and 5'-regions of integrated adenovirus genes correlates with gene expression. *EMBO J.*, **1**, 409–414.
93. Kuhlmann, I. and Doerfler, W. (1983) Loss of viral genomes from hamster tumor cells and nonrandom alterations in patterns of methylation of integrated adenovirus type 12 DNA. *J. Virol.*, **47**, 631–636.
94. Hertz, J.M., Schell, G. and Doerfler, W. (1999) Factors affecting de novo methylation of foreign DNA in mouse embryonic stem cells. *J. Biol. Chem.*, **274**, 24232–24240.
95. Orend, G., Kuhlmann, I. and Doerfler, W. (1991) Spreading of DNA methylation across integrated foreign (adenovirus type 12) genomes in mammalian cells. *J. Virol.*, **65**, 4301–4308.
96. Doerfler, W. (1991) Patterns of DNA methylation—evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. A proposal. *Biol. Chem. Hoppe-Seyler*, **372**, 557–564.
97. Graessmann, M., Graessmann, A., Wagner, H., Werner, E. and Simon, D. (1983) Complete DNA methylation does not prevent polyoma and simian virus 40 virus early gene expression. *Proc. Natl Acad. Sci. USA*, **80**, 6470–6474.
98. Dearing, R., Klotz, G. and Doerfler, W. (1981) An unusual symmetric recombinant between adenovirus type 12 DNA and human cell DNA. *Proc. Natl Acad. Sci. USA*, **78**, 3142–3146.
99. Flanagan, J.M. (2007) Host epigenetic modifications by oncogenic viruses. *Br. J. Cancer*, **96**, 183–188.
100. Sugawara, K., Fujinaga, K., Yamashita, T. and Ito, Y. (1983) Integration and methylation of Shope papilloma virus DNA in the transplantable Vx2 and Vx7 rabbit carcinomas. *Virology*, **131**, 88–99.
101. Kim, K., Garner-Hamrick, P.A., Fisher, C., Lee, D. and Lambert, P.F. (2003) Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *J. Virol.*, **77**, 12450–12459.
102. Badal, S., Badal, V., Calleja-Macias, I.E., Kalantari, M., Chuang, L.S., Li, B.F. and Bernard, H.U. (2004) The human papillomavirus-18 genome is efficiently targeted by cellular DNA methylation. *Virology*, **324**, 483–492.
103. Verma, M. (2003) Viral genes and methylation. *Ann. NY Acad. Sci.*, **983**, 170–180.
104. Munnes, M., Schetter, C., Holker, I. and Doerfler, W. (1995) A fully 5'-CG-3' but not a 5'-CCGG-3' methylated late frog virus 3 promoter retains activity. *J. Virol.*, **69**, 2240–2247.
105. Vardimon, L., Kuhlmann, I., Doerfler, W. and Cedar, H. (1981) Methylation of adenovirus genes in transformed cells and in vitro: influence on the regulation of gene expression? *Eur. J. Cell Biol.*, **25**, 13–15.
106. Vardimon, L., Gunthert, U. and Doerfler, W. (1982) In vitro methylation of the BsuRI (5'-GGCC-3') sites in the E2a region of adenovirus type 2 DNA does not affect expression in *Xenopus laevis* oocytes. *Mol. Cell. Biol.*, **2**, 1574–1580.
107. Simon, D., Grunert, F., von Acken, U., Doring, H.P. and Kroger, H. (1978) DNA-methylase from regenerating rat liver: purification and characterisation. *Nucleic Acids Res.*, **5**, 2153–2167.
108. Subramanian, K.N. (1982) Effect of in vitro methylation at CpG sites on gene expression in a genome functioning autonomously in a vertebrate host. *Nucleic Acids Res.*, **10**, 3475–3486.
109. Fradin, A., Manley, J.L. and Prives, C.L. (1982) Methylation of simian virus 40 Hpa II site affects late, but not early, viral gene expression. *Proc. Natl Acad. Sci. USA*, **79**, 5142–5146.
110. Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
111. Oka, M., Rodic, N., Graddy, J., Chang, L.-J. and Terada, N. (2006) CpG sites preferentially methylated by Dnmt3a in vivo. *J. Biol. Chem.*, **281**, 9901–9908.
112. Yaman, R. and Grandjean, V. (2006) Timing of entry of meiosis depends on a mark generated by DNA methyltransferase 3a in testis. *Mol. Reprod. Dev.*, **73**, 390–397.
113. Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W.W., Okumura, K. and Li, E. (1999) Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, **236**, 87–95.
114. Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. and Bestor, T.H. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science*, **294**, 2536–2539.
115. Bestor, T.H. and Ingram, V.M. (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl Acad. Sci. USA*, **80**, 5559–5563.

116. Li,E., Beard,C. and Jaenisch,R. (1993) Role for DNA methylation in genomic imprinting. *Nature*, **366**, 362–365.
117. Yoder,J.A., Walsh,C.P. and Bestor,T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.
118. Jones,P.L., Veenstra,G.J., Wade,P.A., Vermaak,D., Kass,S.U., Landsberger,N., Strouboulis,J. and Wolffe,A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, **19**, 187–191.
119. Kunert,N., Marhold,J., Stanke,J., Stach,D. and Lyko,F. (2003) A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development*, **130**, 5083–5090.
120. Rankin,R., Pontarollo,R., Ioannou,X., Krieg,A.M., Hecker,R., Babiuk,L.A. and van Drunen Littel-van den Hurk,S. (2001) CpG motif identification for veterinary and laboratory species demonstrates that sequence recognition is highly conserved. *Antisense Nucleic Acid Drug Development*, **11**, 333–340.
121. Gunthert,U., Schweiger,M., Stupp,M. and Doerfler,W. (1976) DNA methylation in adenovirus, adenovirus-transformed cells, and host cells. *Proc. Natl Acad. Sci. USA*, **73**, 3923–3927.
122. Perlmann,C., Saemundsen,A.K. and Klein,G. (1982) A fraction of Epstein–Barr virus virion DNA is methylated in and around the EcoRI-J fragment. *Virology*, **123**, 217–221.
123. Badal,V., Chuang,L.S., Tan,E.H., Badal,S., Villa,L.L., Wheeler,C.M., Li,B.F. and Bernard,H.U. (2003) CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression. *J. Virol.*, **77**, 6227–6234.