Biochemistry (Moscow), Vol. 70, No. 5, 2005, pp. 505-524. Translated from Biokhimiya, Vol. 70, No. 5, 2005, pp. 618-640.

On the Biological Significance of DNA Methylation

W. Doerfler

Institut für Genetik, Universität zu Köln, Zülpicherstr. 47, 50674 Köln, Germany; E-mail: walter.doerfler@uni-koeln.de
Institut für Klinische und Molekulare Virologie, Universität Erlangen, Schloßgarten 4, D-91054 Erlangen, Germany; E-mail: walter.doerfler@viro.med.uni-erlangen.de

Received December 20, 2004

Abstract—This chapter presents a personal account of the work on DNA methylation in viral and mammalian systems performed in the author's laboratory in the course of the past thirty years. The text does not attempt to give a complete and meticulous account of the many relevant and excellent reports published by many other laboratories, so it is not a review of the field in a conventional sense. The choice of viral model systems in molecular biology is well founded. Over many decades, viruses have proven their invaluable and pioneering role as tools in molecular genetics. When our interest turned to the demonstration of genome-wide patterns of DNA methylation, we focused mainly on the human genome. The following topics in DNA methylation will be treated in detail: (i) the *de novo* methylation of integrated foreign genomes; (ii) the long-term gene silencing effect of sequence-specific promoter methylation and its reversal; (iii) the properties and specificity of patterns of DNA methylation in the human genome and their possible relations to pathogenesis; (iv) the long-range global effects on cellular DNA methylation and transcriptional profiles as a consequence of foreign DNA insertion into an established genome; (v) the patterns of DNA methylation can be considered part of a cellular defense mechanism against foreign or repetitive DNA; what role has food-ingested DNA played in the elaboration of this mechanism?

Key words: DNA methylation, adenoviruses, cancer

The results of research on the biochemistry and biology of DNA methylation have grown into a sizable body of scientific information. A single article like this one can, of course, not even attempt to present an adequate overview of this rapidly developing field. This chapter has therefore been restricted to a synopsis of selected work performed in the author's laboratory between 1975 and 2004.

For a long time, many colleagues in molecular biology resisted recognizing the fact that the fifth nucleotide in DNA, 5-methyl-deoxycytidine (m^5C), exerts decisive functions in chromatin structure and in genetic control mechanisms. With m^5C, however, the arguments have finally become irrevocably strong enough not to be ignored. Nevertheless, text books still preach the existence of four, instead of five, nucleotides in DNA. Of course, it is good and essential scientific practice to cast most critical scrutiny on new claims and demand ample and definitive experimental proof. A large number of researchers have now provided this proof, and many of the findings will be summarized in this volume. My own group started contributing to the honing of problems related to DNA methylation in the mid-1970s, and this article presents a detailed summary of our results which have been adduced since then and stood the test of time. For further information, the reader can consult the references cited herein and previous reviews, which have been published as our work proceeded [1-7].

The discovery of m^5C [8] in eukaryotic, particularly in mammalian, DNA has provoked a challenging search for its functional significance. This search is by no means completed, and active investigations on numerous unsolved questions are still continuing. The modification of cytidine (C) to m^5C, apparently the only one among the nucleotides in mammalian DNA, is introduced postreplicationally by several DNA methyltransferases (DNMT) which are chosen depending on the functional context of their enzymatic activity: DNA can be methylated *de novo*, still a most enigmatic series of events, or a given pattern of DNA methylation in the genome can be maintained upon replication. In this latter mode of maintenance methylation, the parental DNA strand with the m^5C residue still in place can serve as the template to direct the DNMT to modify the newly synthesized DNA

0006-2979/05/7005-0505

complement. Although several DNMT have been well characterized, it is not clear whether any of them by itself suffices to facilitate either of the two modes of DNA methylation. In addition to the enzymatic activity proper, the function of these enzymes seems to depend critically on the conformation of the local chromatin segment in which the DNA is to be methylated. Since our understanding of chromatin structure is incomplete, we cannot expect to obtain a comprehensive description of the enzymatic activities of the DNMT. It appears more realistic to propose a complex interplay between DNA–chromatin structure and specific choices of enzymatic functions in which additional regulatory proteins have to participate.

In experimental terms, DNA methylation activities cannot be realistically assessed by relying on the measurement of enzymatic function using a naked DNA template, since the actually operational template for DNMT is a DNA–chromatin complex with site-specific, stochastically malleable functions which are targeted to individual loci in the genome. It will be some time before these processes can be elucidated or even mimicked by current technology. How can we approach a functional analysis of DNA methylation in eukaryotic, particularly in mammalian, systems? One important parameter in understanding this functional DNA modification is to realize that m^5C residues are not introduced randomly by a fortuitously acting enzymatic mechanism. In contrast, highly specific patterns in the distribution of m^5C residues exist all over the genome. These patterns appear to be different in each cell type and in each region of the genome. It will require a major effort to determine these patterns of DNA methylation in all parts of the mammalian, specifically in the human, genome.

In recognizing the very significant accomplishment of determining the nucleotide sequence of the human genome, I submit that the task has not been completed without the inclusion of the fifth nucleotide. Of course, it is technically impossible to differentiate between a C- and a m^5C-residue by the conventional sequencing reaction. The application of the bisulfite protocol of the genomic sequencing reaction [9, 10] is a demanding project, particularly when it has to be extended to many kilobases of DNA sequence. Nevertheless, this method is, at least for the time being, the only reliable procedure to ascertain levels and patterns of DNA methylation. By applying the bisulfite reaction, one can detect all m^5C residues in a sequence. The human epigenome project has just been initiated on an exploratory basis and will have to cope with the fact that patterns of DNA methylation can be different from cell type to cell type and, of course, in each segment of the genome [11]. In my laboratory, we have investigated methylation patterns in several areas of the human genome to obtain a first impression of the types of patterns. The structure of the genome inside its chromatin casing and its regulatory functions appear to depend on these patterns of DNA methylation. The function of the genome will not be understood before the completion of the analysis of these patterns. Hence, the study of more complex biomedical problems will undoubtedly escape a thoroughly informed experimental approach before this analysis has been finished. Presently available data imply that in 90 human genes in the major histocompatibility complex (MHC) of multiple tissues and individuals, the majority of regions were hypo- or hypermethylated. The patterns were tissue-specific, interindividually variable, and correlated with gene expression [12].

The following *Gedankenmodell* may aid the conceptual visualization of a more general function of patterns of DNA methylation across the entire genome. The model is based upon the notion that m^5C residues are modulators of DNA–protein interactions, as proposed earlier [2], and these modulators could facilitate and enhance or abrogate such interactions. The direction, in which these modulations work, depends on the type of protein and DNA sequences in functionally crucial interactions.

Imagine a bare wall represented here by the plain nucleotide sequence of A, C, G, and T residues onto which elaborate decorations have to be attached. Chromatin proteins then are the decorations, which eventually contribute to the chromatin structures and could be specific for different segments of the genome. Now, we insert into the blank wall the m^5C "pegs" to which proteins bind or are prohibited from binding. With this first and essential set of DNA–protein interactions, a central genome-associated scaffold will be generated which then will be able to inaugurate further protein and/or RNA assemblies until the final, yet enigmatic, chromatin structure has been established.

Local specificities in this structure will, of course, be determined by the site-specific pattern of DNA methylation, which thus assumes a functionally crucial role in this assembly process. There are several, but one particular, problem with this model: it is not apparent whether the generation of a given pattern of DNA methylation arises before or after chromatin formation. Possibly, both events are interdependent and develop concomitantly. Upon DNA replication, an established and inheritable pattern of DNA methylation is, of course, maintained by the array of m^5C residues, which are still preserved after DNA replication in the parental strand and which can serve as a template for the insertion of methyl groups in the newly synthesized DNA complement. In this way, patterns of DNA methylation are propagated and inherited. The methylation patterns in turn promote the site-specific chromatin structures.

A further tantalizing aspect arises from the fact that DNA methylation patterns are erased early in embryonic development and are thereupon re-imposed by an unknown mechanism of *de novo* DNA methylation, which cannot avail itself of the template pattern on the

complementary strand of DNA. Conversely, the fixation of *de novo* methylation patterns on integrated foreign DNA or in the course of embryonic development might be directed by local chromatin structures, which then would have to be "remembered" even in the absence of the fifth nucleotide. It is this crucial interdependence between methylation pattern and chromatin structure that we cannot yet satisfactorily explain. RNA could conceivably serve as a mediator for this functional gap in time and structure. This model is based on the fact that each individual segment of the genome is tightly associated with a given pattern of DNA methylation and, consequently, of chromatin structure. The same or a very similar site-specific pattern can also be conveyed to foreign DNA subsequent to its insertion into a specific segment of the mammalian genome.

In part, this model has been deduced from the observation that the site-specific re-integration of an unmethylated mouse gene, the B lymphocyte tyrosine kinase (*BLK*) gene, into the mouse genome by homologous recombination leads to the re-establishment of the original and authentic DNA methylation pattern in the integrate at its authentic site [13]. In contrast, when the *BLK* gene randomly hits host DNA sequences and recombines there by a non-homologous mechanism, patterns of DNA methylation are completely different from the authentic pattern in the *BLK* gene. For a working hypothesis, we assume that each genome segment is characterized by a "methylation memory". Its biochemical correlate is not known but must somehow be related to topical chromatin structure as well as local DNMT type, concentrations, and activities as well as auxiliary functions.

The most intensely studied function of DNA methylation in eukaryotic genomes is that of promoter activity and long-term gene silencing. Starting in the late 1970s, our laboratory has regularly contributed to the elaboration of this concept [1, 2]. In conjunction and – again in an interdependent mode – DNA (m^5C) and chromatin (histone acetylation and methylation) modifications collaborate in the long-term silencing of promoters and thus assume an essential function in regulating the activity of specific genome segments. In recent years, these mechanisms have been recognized to be of importance also for the understanding of more complex biomedical problems, in particular those which are related to genetic imprinting, embryonic and fetal development, genetic disease, and tumor biology. Here, we have another fine example of how basic research on fundamental mechanisms in molecular genetics can eventually help understand practical problems in biomedical research. Without turning to the study of simpler experimental systems, e.g. viral models, in the elucidation of promoter silencing by DNA methylation and related histone modifications, it would have been impossible to approach more complex problems in mammalian organisms or in plants.

THE *de novo* METHYLATION OF INTEGRATED FOREIGN DNA

Choice of experimental systems. There are many excellent examples to document that the study of viral systems has led to the discovery of fundamental mechanisms in prokaryotic and eukaryotic molecular genetics. Since in most instances, virus replication has to rely on the utilization of cellular mechanisms, it cannot be surprising that viruses have been efficiently exploited as "Trojan horses" inside the cellular milieus. In the 1970s, our laboratory was involved in detailed analyses on the mode of adenoviral DNA integration into the host genome in adenovirus-transformed cells and in adenovirus type 12 (Ad12)-induced hamster tumor cells. In the course of these studies, it became feasible to prove that integrated foreign (adenoviral) DNA was *de novo* methylated [14]. When we subsequently were able to document the first inverse relationship between genetic activity of integrated viral genes and the extent of their methylation [15], it was obvious that this experimental system could be applied to fundamental studies on the regulation of genetic activity and on the biological function of DNA methylation [2].

A second seminal observation, which emphasized the biological importance of DNA methylation, came from detailed investigations on patterns of DNA methylation in the 5′-upstream regions of two randomly chosen human genes, *TNFα* and *TNFβ*. The genomic sequencing method originally developed by Church and Gilbert [16], though difficult to use at the time, helped considerably in these studies. In the promoter and 5′-upstream regions of the *TNFα* and *TNFβ* genes, we found cell type-specific and interindividually highly conserved patterns in the distribution of m^5C residues, which agreed to the nucleotide site among individuals from different ethnic origins [17, 18].

Similar, though less precise, evidence came from large, randomly chosen segments of the human genome [19], many of them repetitive DNA sequences. These results implied that highly specific patterns of DNA methylation existed and most likely had to have a fundamentally important function. It was not easy in those days to convince others that a systematic endeavor to determine patterns of DNA methylation was not just a descriptive exercise but had to be initiated to learn about the wider gamut of possibilities with functional implications. Hopefully, the human epigenome project will help provide more ample evidence than the study of a single, pioneering, laboratory could possibly have adduced with limited means 15 years ago.

At this time, we also sought the collaboration of clinical researchers in order to extend the basic concepts derived from simpler experimental systems to more complex biomedical problems. In the course of these studies, it became even more obvious that the model developed

with the adenovirus system could reliably guide all our efforts. In collaboration with several groups, we determined patterns of DNA methylation in the Prader–Willi/Angelman regions of the human genome [20, 21], in the promoter regions of the *RET* protooncogene [22], of the *FMR1* gene [23, 24], and of several genes of the erythrocyte membrane [25, 26].

The state of methylation in DNA viral genomes. *Many DNA virion genomes are unmethylated, others are methylated.* Modulation as a bidirectionally active parameter in DNA–protein interactions can be exemplified by the activity of the restriction endonucleases *Dpn*I and *Dpn*II. *Dpn*I cleaves the nucleotide sequences $G^{6m}ATC$ only when the A residue in the recognition sequence is methylated, whereas *Dpn*II is inhibited by a ^{6m}A residue in this sequence [27]. Hence a methylated nucleotide can both obstruct or facilitate and be required for the activity of a restriction endonuclease, i.e. for the interaction between a given nucleotide sequence and the protein which specifically recognizes this sequence. A similarly instructive example is not available for a m^5C containing recognition sequence of a restriction endonuclease.

Among the DNA containing viral genomes, examples of completely unmethylated as well as completely 5′-CG-3′ methylated virion DNA molecules exist. The encapsidated virion DNAs of the human adenoviruses [28] are unmethylated. In striking contrast, the double-stranded genome of frog virus 3 (FV3), an iridovirus, is completely methylated in all 5′-CG-3′ dinucleotides [29, 30]. As will be discussed later, the intracellular FV3 virion DNA becomes quickly remethylated after replication. Possibly, due to the specific nucleotide sequence of the FV3 genome, the viral and/or cellular proteins, which have to interact with this viral genome in the course of viral transcription and replication, are not inhibited by FV3 DNA methylation. Some of them may even require a methylated genome for full activity.

We have used several techniques, total hydrolysis of the virion DNA followed by bidirectional chromatography and electrophoresis [28], which allow the separation of C from m^5C residues, as well as genomic sequencing methods [31, 32], to demonstrate that the virion DNA as well as the free, i.e. not host cell genome integrated, intracellular adenovirus DNA in productively or abortively infected cells [33] remains unmethylated. In the latter study, restriction endonucleases were used to document the absence of m^5C residues at least in the *Hpa*II recognition sequences 5′-CCGG-3′. The intracellular genomes of the episomally persisting Epstein–Achong–Barr virus (EBV) have become methylated to a certain extent [34]. Similarly, the genome of another persisting virus, *herpes virus saimiri*, in lymphoid tumor cell lines has been shown to be extensively methylated [35]. The retroviral progenomes also become methylated [36, 37].

SYREC, an Ad12 recombinant genome, which carries unmethylated cellular DNA. When adenovirus type 12 (Ad12) was serially propagated on human cells in culture, a variant Ad12 genome arose which constituted a naturally generated recombinant between the left terminal 2081 nucleotides of Ad12 DNA and a large palindromic fragment of cellular DNA. This viral recombinant could be separated from the authentic Ad12 virions due to its lower buoyant density by equilibrium sedimentation in CsCl density gradients [38]. The existence of this symmetric recombinant (SYREC) proved that recombination could occur between viral and cellular DNAs in human cells, which were productively infected with human Ad12 [39].

The cellular DNA in this huge palindromic genome of some 34 kb with identical left ends of Ad12 on either terminus of the recombinant genome comprised cellular DNA sequences of both the unique and repetitive types. Interestingly, these cellular DNA sequences were completely unmethylated in the virion recombinant, but the same cellular DNA sequences were highly methylated in the human cellular genome from which they had been originally derived [38]. This finding demonstrates that free adenovirion DNA remains devoid of m^5C in the same human cell nucleus in which adenovirion DNA replicates and in which the methylation of cellular DNA is maintained in specific patterns. Apparently, the cellular DNMT fail to gain access to the free virion DNA, possibly because adenovirus DNA can avail itself of its own, specific, virion genome-encoded mechanism of DNA replication with the adenovirus terminal protein (TP), its viral DNA polymerase (pol), and the DNA binding protein (DBP). Alternatively, it is conceivable that free intranuclear adenovirus DNA becomes protected from *de novo* methylation by binding to specific proteins. Adenoviral DNA replication is at least partly independent of the cellular replication machinery, except for the requirement for nuclear factors I, II, and III which might not be linked to any of the cellular DNMT. On the other hand, the intracellularly located, episomal DNA of EBV must be tightly associated with the replication system for cellular DNA with which it replicates in synchrony. Thus, the EBV episomes might be in close contact with cellular DNMT and become methylated.

The adenovirus SYREC molecule and its ability to replicate in human cells in the presence of a helper adenovirus with an intact authentic viral genome has been the model for the construction of the gutless adenovirus vectors of the third generation [40]. These researchers have been able to separate the recombinant virus from its wild type precursor also by equilibrium sedimentation in CsCl density gradients.

Suppression of the frequency of 5′-CG-3′ dinucleotides in the genomes of the small eukaryotic viruses. The dinucleotide 5′-CG-3′ is statistically underrepresented in all but four of the small viruses with a genome size of <30 kb [41]. In the larger viral genomes, the abundance of this dinucleotide follows statistical expectations. The retrotransposons in eukaryotic genomes are also characterized

by low values of 5′-CG-3′ dinucleotides. There are several possible interpretations for these phenomena: methylation effects during the proviral states of some of these genomes which would lead to their silencing, and also dinucleotide stacking energies and mutation mechanisms or selection during evolution.

***De novo* methylation of foreign DNA, which was integrated into the mammalian genome.** *Studies on integrated Ad12 genomes in transformed or tumor cells.* In the course of investigations on the mode of Ad12 DNA integration in Ad12-transformed hamster cells by using restriction endonucleases, the *de novo* methylation of integrated foreign DNA was discovered. The generated fragments of cellular DNA were separated by electrophoresis on agarose gels and further analyzed by Southern blotting [42] and hybridization to ^{32}P-labeled Ad12 DNA or, more specifically, to the ^{32}P-labeled terminal fragments of Ad12 DNA. In this way, the terminal viral DNA fragments linked to the immediately abutting cellular DNA segments could be identified.

In an attempt to generate small junction fragments, which could be more easily analyzed, frequently cutting restrictases like *Hpa*II were employed. In these experiments, we discovered that the integrated form of Ad12 DNA was not effectively cleaved by *Hpa*II, whereas virion DNA, previously shown to be unmethylated [28], was readily cut. These data implied that the integrated Ad12 DNA had become *de novo* methylated upon integration into the established hamster genome [2, 14, 43].

This interpretation could be proven when the isoschizomeric restriction endonuclease pair *Hpa*II and *Msp*I became available. Both enzymes recognize the sequence 5′-CCGG-3′. *Msp*I cleaves irrespective of the presence of a m^{5}C residue in the 3′-located C-position in the recognition sequence, but *Hpa*II is capable of cleaving only the unmethylated sequence [44]. Along these lines, the integrated Ad12 DNA, like virion Ad12 DNA studied as a control, was completely cleaved by *Msp*I, whereas *Hpa*II could cleave only the virion DNA to completion. Integrated Ad12 DNA was cut incompletely by *Hpa*II and was thus recognized to be 5′-CCGG-3′ methylated in distinct patterns. Here, we were able to document one of the early examples for the notion that foreign DNA inserted into established mammalian genomes became heavily methylated [14, 15]. This now commonly reproduced finding was later extrapolated to numerous other eukaryotic genomes, including those of plants [45]. The human papilloma viruses 16 and 18 integrated into the genomes of cells from human cervical carcinomas are also methylated in functionally distinct patterns [46].

Site of initiation of de novo methylation – site of foreign DNA integration in the recipient genome. In later studies, we have demonstrated in numerous Ad12-transformed hamster cell lines and particularly in Ad12-induced hamster tumor cells that integrated Ad12 DNA is an excellent substrate for the action of cellular DNMT [47-49]. The patterns generated in different cell lines and tumors exhibited some similarities but did not appear to be identical. Extent and pattern of methylation of integrated foreign DNA were directed rather by the site of integration in the recipient genome than by the nucleotide sequence of the foreign DNA [50], although the latter could have some influence as well. In this context, the observation was of interest that a cloned E1 segment of adenovirus DNA genomically fixed by transfection into hamster cells and integrated at several different loci in the host hamster genome became methylated to different extents or could remain hypo- or unmethylated [50]. Hence, the site of insertion of foreign DNA had to be a strong determinant in its subsequent *de novo* methylation.

The sites of initiation of *de novo* methylation were determined by using the cloned *Hind*III DNA fragments of Ad12 DNA as hybridization probes on *Hpa*II- or *Msp*I-cleaved DNA from different Ad12-induced hamster tumors. These DNA fragments were separated by electrophoresis on agarose gels and subsequently analyzed by Southern blotting and hybridization to the ^{32}P-labeled Ad12 DNA fragments. The results of a large number of experiments [50] demonstrated that *de novo* methylation was initiated inside the integrated Ad12 DNA molecules in two paracentrally located regions of the Ad12 genomes. *De novo* methylation did not commence at or close to the termini of Ad12 DNA, which were linked to cellular DNA. The termini, in fact, remained hypomethylated, possibly because continued hypomethylation and expression of the gene products from the Ad12 regions E1 (left terminus) and E4 (right terminus) were likely selected for in Ad12-transformed cells and in Ad12-induced tumors.

We have investigated the site of initiation of *de novo* methylation in integrated Ad12 genomes also at the nucleotide level by applying the bisulfite genomic sequencing reaction. When CsCl-purified Ad12, 10^6 to 10^7 plaque forming units per animal, is injected intramuscularly into newborn hamsters (*Mesocricetus auratus*), numerous tumors of different sizes develop in the animals' peritoneal cavities. In these intraperitoneal tumors of different sizes, the paracentrally located regions of integrated Ad12 genomes were analyzed by the bisulfite protocol for the state of methylation.

Methylation levels did not exhibit an unequivocal relation to tumor size. Initiation of DNA methylation, moreover, was not emanating from a specific nucleotide or set of nucleotides in the region previously shown to be a site of initiation of *de novo* methylation. Initiation was rather regional and appeared to emerge from several sites within this region [50, 51]. Hence, *de novo* methylation seems to commence in a region and not at a single specific nucleotide. Of course, this experimental approach does not allow deriving a definite correlation between the time of foreign DNA integration and that of the initiation of DNA methylation.

In a related type of experiment, a plasmid construct, which contained the *E2A* late promoter of adenovirus type 2 (Ad2) and the prokaryotic gene for the chloramphenicol acetyltransferase (CAT) as a reporter, was transfected into hamster cells. The *Hpa*II-premethylated or unmethylated pAd2E2AL-CAT gene construct was genomically fixed in hamster cells by co-transfection with the unmethylated pSV2-neo plasmid. In this plasmid, the early *SV40* promoter controlled the neomycin phosphotransferase gene, which facilitated the selection of transgenic cells. Stability of methylation status and expression of the *CAT* gene were assessed in a number of clonal transgenic cell lines [52]. The foreign DNA was integrated frequently in multiple tandems of the transfected plasmid. Among 19 clonal cell lines, the unmethylated construct remained in that state; and in 18 of these lines, the *CAT* gene was continuously expressed. Among 14 cell lines transgenic for the premethylated construct, seven lines failed to express the test transgene, and the three 5′-CCGG-3′ sites in its late *E2A* promoter remained almost completely methylated. In five cell lines, the promoter remained partly methylated and the *CAT* gene was only weakly expressed. In two cell lines, the premethylated promoter lost the m^5C modification altogether, and the *CAT* gene was strongly expressed [52].

Factors determining de novo methylation: reinsertion of a mouse gene into its authentic position. We further pursued the general question of which factors would affect the *de novo* DNA methylation in mammalian genomes. The mouse B lymphocyte tyrosine kinase (*BLK*) gene was reintegrated by homologous recombination into the genome of mouse embryonal stem (ES) cells. Two different plasmid constructs containing that gene were used for these experiments. One construct also carried the weak *E2A* late promoter of Ad2 DNA in front of the luciferase gene. In the second, this gene was controlled by the strong early *SV40* promoter. Upon homing through homologous recombination to the authentic chromosome 14 or by heterologous recombination to many different loci in the mouse genome, methylation patterns in the integrates were assessed by restriction with the methylation-sensitive endonuclease *Hpa*II or the insensitive *Msp*I. The mouse *BLK* gene reinserted into the genome by homologous recombination had reestablished the identical methylation pattern characteristic for the authentic, non-manipulated mouse *BLK* gene [13]. The extent of *de novo* methylation in the DNA segments adjacent to the *BLK* gene in the integrated construct depended on the promoter present in the plasmid construct and on the location of the recombined construct in the ES genome. In homologously inserted DNA, which carried the weak Ad2 promoter, *de novo* methylation was extensive. Presence of the strong *SV40* promoter led to hypomethylation or no methylation at all. Even when the enhancer sequence was removed from the *SV40* promoter, hypermethylation was still observed. All randomly integrated constructs, independently of the type of promoter or enhancer included, were hypermethylated. We concluded [13] that an authentic mouse gene reinserted into its original genomic site was remethylated to the identical pattern as previously present on the target and on the allelic site; that heterologous recombination to randomly targeted loci did not confer the mouse *BLK* gene-specific methylation pattern, and that promoter strength in a construct was able to influence the pattern of methylation imposed *de novo* on the inserted construct after homologous recombination in the mouse genome.

De novo DNA methylation – an ancient cellular defense mechanism? The *de novo* methylation of integrated foreign DNA is a phenomenon widely documented throughout the phyla of eukaryotic organisms, in mammals as well as in plants [45]. The nowadays, established genomes have evolved over many millennia. We tend to assume that the evolution of genomes is a continuous process which started right after the beginning of organismic life and probably before organisms arose and will continue as long as this biologic system can be maintained. Even under experimental conditions, many organisms have been shown to be capable of accepting and of accommodating foreign DNA, although with unpredictable – advantageous or catastrophic – consequences for the acceptor cell. Since all cells in culture or organisms in environment-challenged life are subject to stringent conditions of selection, even the cell which has been forced by its innate recombination mechanisms to tolerate the genomic insertion of foreign DNA can avail itself of an ancient defense mechanism against the genetic activity of foreign DNA which could carry active genes. Since promoter methylation has been identified as part of a mechanism for the long-term silencing of genes and DNA segments, the *de novo* methylation of integrated foreign DNA can be contemplated as such a defense mechanism or at least as an integral part of it [53, 54].

A large part of m^5C residues is found in the parasitic sequence elements of retrotransposons and endogenous retroviruses which constitute >35% of the human genome. Perhaps intragenomic parasites are recognized by their high copy number. Long-term inactivation by DNA methylation also entails the possibility of m^5C residues being deaminated to T's followed by permanent inactivation [55].

Are integrated foreign DNA sequences stabilized by hypermethylation? The Ad12-transformed cell line T637 was obtained by the *in vitro* transformation of BHK21 hamster cells with Ad12 [56] and carries about 15 copies of Ad12 DNA integrated at a single chromosomal locus [57-59]. Unintegrated, free Ad12 DNA could never be detected in any of the adenovirus-transformed cell lines or in Ad12-induced tumor cells. The integrated Ad12 DNA is methylated in functionally significant patterns [15, 51]. Upon continuous propagation in cell culture, a small number of morphological revertants arose from the

cell line T637 [60]. These revertants exhibit a fibroblastic phenotype in contrast to the more rounded, epithelioid appearance of the T637 cells. The revertants can be selected due to their resilience to spent, acidic medium in which T637 cells detach.

We have studied several of these revertants with respect to their content of residual viral DNA. In one of these revertants, TR12, only one copy of Ad12 DNA and an approximately 3.9 kb long fragment from the right terminus of Ad12 DNA persist in the integrated state (N. Hochstein, I. Muiznieks, and W. Doerfler, manuscript in preparation). Studies using methylation-sensitive restriction endonucleases revealed that the integrated Ad12 DNA in the revertant cell line TR12 was even more extensively methylated than in the cell line T637 from which TR12 originated [61]. Preliminary results adduced from experiments, in which the bisulfite genomic sequencing method was applied, confirm these data. We propose that hyper- or nearly completely 5′-CG-3′-methylated foreign DNA sequences would be more stably integrated than less completely methylated foreign genomes. Repetitive sequences, like endogenous retroviral retrotransposons, appear to be significantly, but not completely, methylated [62].

De novo methylation of Ad12 and of cellular DNA in hamster tumors. The insertion of foreign DNA into established mammalian genomes has consequences for the inserted foreign DNA and for the recipient host genome. We have chosen to study these events in Ad12-induced hamster tumor cells, a model in basic research as well as for its significance in viral oncology. When Ad12 is injected subcutaneously into newborn hamsters within 24 h after birth, undifferentiated tumors develop at the site of topical application of the virus within a few weeks in 70 to 90% of the animals, which survive injection. As mentioned above, upon the intramuscular injection of Ad12 into the gluteal region, numerous tumors are found intraperitoneally. Histologically, these tumors exhibit Homer–Wright rosette-like structures indicative of primitive neuroectodermal tumors (PNET).

The tumor cells express proteins, which are characteristic both for neuroepithelial as well as for mesenchymal cells [51, 63]. Each tumor cell carries multiple copies of integrated Ad12 DNA. Free viral DNA has never been found in any of these tumor cells.

The cells from an individual tumor carry the viral integrates, with few exceptions, all at a single chromosomal location which is identical in all cells of a given tumor. Among 60 different tumors investigated only one showed two different chromosomal loci to be occupied by Ad12 DNA when investigated by the fluorescent *in situ* hybridization (FISH) technique. When we compared the sites of Ad12 DNA integration in more than 100 different hamster tumors, the sites of viral DNA integration in the host cell genome were different from tumor to tumor [43, 47, 64]. We, therefore, favor a model of clonal origin of these Ad12-induced tumors. Aside from the *de novo* methylation of the viral integrates in all of the Ad12-induced tumors, there can be changes in the extent of DNA methylation also in the host genome [62]. Moreover, the transcription patterns of cellular genes in different tumors can be very similar but there are also differences. The patterns of transcription of the integrated Ad12 genes are strikingly similar in all Ad12-induced hamster tumors analyzed and resemble those patterns in the Ad12-transformed hamster cell line T637 [51, 63].

Loss of Ad12 genomes is compatible with maintenance of the oncogenic phenotype. Upon continuous cell culture, some of the Ad12-induced hamster tumor cells, which carried multiple copies of integrated and *de novo* methylated Ad12 genomes, lost the viral DNA sequences completely or almost completely. Surprisingly, these revertants devoid of Ad12 DNA and notably lacking its left terminus with the transforming E1 region of the viral genome retained their oncogenic phenotypes when reinjected into hamsters [48, 65]. We consider this result to be highly significant in that it demonstrates that Ad12 is capable of inducing tumors, which keep their oncogenic properties in animals, even when the E1 region is completely lost from these tumor cells.

Many researchers in the field of adenovirus tumor biology consider the E1 region of the adenoviral genome akin to an oncogene and its maintenance paramount in preserving the oncogenic potential of the adenovirus-induced tumor cells. However, this notion has been experimentally derived mainly by using rat embryo fibroblasts, which were transformed in culture through the transfection of adenovirus type 5 (Ad5) DNA fragments [66]. Obviously, tumor induction by Ad12 in animals can be a quite different process and can probably not appropriately be mimicked by transfection experiments in cell culture. Below I propose an alternate way of looking at the mechanism of viral oncogenesis. In this view, global changes in the cellular genome as a consequence of viral DNA integration are invoked as a mechanism of viral oncogenesis in addition to the expression of a single "oncogene" [4, 5, 7].

DNA methylation in non-CpG dinucleotides; hemimethylated DNA. The establishment of *de novo* patterns of DNA methylation in mammalian genomes is characterized by the gradual spreading of methylation, which has been documented to occur across multiple copies of integrated adenovirus genomes as well as, at the nucleotide level, in the integrated *E2A* promoter of Ad2 DNA [52, 67]. A few 5′-CG-3′ sequences can remain hemimethylated for several cell generations before they become totally methylated. Hemimethylation may be a transient phenomenon but could also persist for a certain period in specific segments of a transgene or in the genome in general. In the Ad2-transformed cell line HE2, the *E2A* promoter in the integrated Ad2 genome is heavily methylated not only in all 5′-CG-3′ dinucleotides

but also in some 5′-CA-3′ and 5′-CT-3′ dinucleotides [68]. Evidence for the occurrence of m^5C in non-5′-CG-3′ dinucleotides has also been presented by Woodcock et al. [69]. There is at present no plausible explanation whether and how this type of methylation can be maintained following DNA replication.

Initiation and spreading of de novo methylation. The mechanism of *de novo* methylation is not well understood. From the results of experiments performed with the adenovirus system, the following conclusions appear well founded:

– in a transgene the size of Ad12 DNA, 34,125 bp [70], *de novo* methylation starts in internal regions of the genome [49-51] and spreads from there across the transgene [67, 68]. In transformed or tumor cells, the early regions, particularly E1, can be partly spared from *de novo* methylation, because their gene products are required during the selection for the oncogenic phenotype;

– initiation of *de novo* methylation is regional and not confined to one or a few contiguous 5′-CG-3′ dinucleotides [51]. In Ad12-induced tumors, the extent of Ad12 transgene methylation is not only dependent on tumor size. Transcribed regions of the transgene are hypomethylated, inactive segments hypermethylated [15, 71, 72];

– spreading of DNA methylation is not entirely contiguous. For unknown reasons, certain 5′-CG-3′ dinucleotides can remain unmethylated (I. Hochstein, I. Muiznieks, and W. Doerfler, manuscript in preparation);

– the extent of *de novo* methylation and the speed of its spreading seem to be determined by the chromosomal site of transgene localization, to a lesser extent also by the nucleotide sequence of the transgene [13, 50];

– the reestablishment of the authentic pattern of DNA methylation in the correctly reinserted *BLK* gene in mouse embryonic stem cells [13] implies that each segment of the mammalian genome is capable of exerting a memory function which could be directly related to the mechanism of maintenance methylation. Specific chromatin structures at different sites of the genome may be decisive in directing the methylation reaction;

– in the same set of experiments [13], we noted that the same reintegrated *BLK* gene, when it was inserted at randomly selected sites in the genome, was methylated in completely different patterns. Moreover, the extent of the *de novo* methylation was influenced by the strength of the viral promoter in the construct integrated. The weaker *E2A* late promoter of Ad2 DNA led to a more extensive *de novo* methylation than the presence of the stronger early promoter of *SV40* DNA;

– when Ad12-transformed cells or Ad12-induced hamster tumor cells were continuously passaged in culture, revertants arose which had lost all or a part of the multiple copies of integrated Ad12 DNA. The then persisting Ad12 DNA genomes in the revertants seemed to be more completely methylated than the lost copies. Hence, the idea was put forward that the levels of DNA methylation of transgenes might be related to their stability of fixation in the host genome [61]. Moreover, continuous culture of different Ad12-induced tumor cells led sometimes to the selection of cell lines with very similar integration patterns [73]. It is conceivable that the integration sites in the thus selected cells were most compatible with the survival of these cells in culture.

INVERSE CORRELATIONS BETWEEN PROMOTER ACTIVITY AND METHYLATION

The field of DNA methylation in mammalian cells was pioneered by the discovery of inverse correlations between the extent of segmental DNA methylation and the genetic activity of these segments. In integrated Ad12 DNA in Ad12-transformed cells, the early viral genes are transcribed; the late Ad12 genes responsible for the synthesis of virion capsid proteins are permanently silenced. Hence, in Ad12-transformed cells or in Ad12-induced tumor cells, late viral gene products and mature virions are not synthesized [74]. Thus, Ad12-transformed cells provide a suitable tool to study the levels of DNA methylation in distinct sections of the viral genome and to document inverse correlations to gene transcription for the first time [15, 33, 75, 76]. A particularly clear example has been offered by the adenovirus type 2 (Ad2)-transformed cell lines HE1, HE2, and HE3 [77]. The E2A region of Ad2 DNA is not expressed in cell lines HE2 and HE3, whereas cell line HE1 does express the E2A region of the integrated Ad2 DNA [78]. Accordingly, the 5′-CCGG-3′ dinucleotides (*Hpa*II sites) in the promoter region of this gene in cell line HE1 are unmethylated and methylated in cell lines HE2 and HE3 [33].

At the time, one had to rely on the analyses of a limited number of 5′-CG-3′ dinucleotides whose state of methylation could be assessed only by the use of methylation-sensitive restriction endonucleases, in this example of *Hpa*II and *Msp*I (5′-CCGG-3′). These seminal observations started a burst of similar studies with numerous viral and cellular genes and confirmed almost without exceptions the initial observations which suggested that specific promoter methylation patterns are instrumental in the long-term silencing of eukaryotic genes [2, 72]. To this day however, we do not understand whether there have to be a specific number or pattern of m^5C residues in a promoter to assure its long-term silencing.

Productive versus abortive infection of cells with Ad12. Ad12 interacts with human cells in a productive infection leading to the synthesis of a large number of progeny virions. In contrast, the replication of Ad12 in hamster cells is completely blocked [79, 80], possibly because only few copies of Ad12 DNA are capable of

reaching the nucleus of the hamster cells [81]. However, the expression of the Ad12 genome in hamster cells appears to be suppressed in several steps of the normal replication cycle. Nevertheless, viral DNA can be traced in the nucleus of the abortively infected hamster cells and limited transcription of the early genes of non-integrated Ad12 has been documented [74, 81]. There is, however, no methylation of the intranuclear Ad12 DNA in productively or abortively infected cells [33]. Obviously, methylation of free viral DNA, even in the abortive system, does not serve to regulate or to inactivate the late Ad12 viral genes, which are not transcribed in abortively infected hamster cells.

The actively transcribed genome of frog virus 3 (FV3) is completely 5′-CG-3′ methylated. The hypermethylated state of the virion-encapsidated or of the intracellular FV3 genomes [29] in fish or mammalian cells has taught us that the biological significance of DNA methylation cannot be schematically interpreted and depends entirely on the biological system studied. While the inverse correlations described above hold true for most systems investigated so far, there are notable exceptions to this "rule" and, of course, no stringent dogmas in biology. It has been documented that the viral *L1140* gene is actively transcribed late after infection of fish cells with FV3, although it is methylated in all 5′-CG-3′ dinucleotides [82]. In FV3-infected fish or hamster cells, a transfected *L1140* promoter–indicator gene construct is active in the unmethylated or fully 5′-CG-3′-methylated form. When the same construct is methylated only in the 5′-CCGG-3′ (*Hpa*II) sequences, its activity is reduced. Compatibility between the methylation of an immediate-early FV3 promoter with its active transcription has also been reported [83]. These data confirm the special methylation requirements of this promoter in FV3 DNA. Special properties of the FV3 DNA–protein interactions may account for these unexpected activity patterns. It would be interesting to study in greater biochemical detail the transcription of FV3 genes and the proteins involved in their regulation.

SITE-SPECIFIC PROMOTER METHYLATION AND GENE SILENCING

The finding of inverse correlations between promoter activity and extent of DNA methylation led to the concept that sequence-specific promoter methylations exert a regulatory function on gene activity. In order to provide more direct evidence for this interpretation, we devised experiments in which a number of promoter–indicator gene constructs were tested for their genetic activities in the unmethylated or in the methylated state at 48 h after transfection into mammalian cells in culture. In general, these data corroborated the earlier interpretation of promoter inactivation by promoter methylation, although this experimental approach could, of course, not help decide whether in an intact mammalian genome promoter methylation was the cause or consequence of promoter inactivation. The former possibility, however, remains the more likely explanation.

The *E2A* promoter of Ad2 DNA. In a first set of experiments, the oocyte system from *Xenopus laevis* was adapted to test unmethylated or methylated promoter–gene constructs for genetic activities. The cloned *E2A* region of Ad2 DNA was then 5′-CCGG-3′-methylated with the *Hpa*II DNMT or was left unmethylated. Subsequently, either construct was microinjected into the nuclei of *Xenopus laevis* oocytes. The methylation status of these constructs was maintained in the oocyte nuclei. At 48 h after microinjection, the unmethylated construct was transcribed in the oocyte nuclei; the methylated construct was silenced [84]. Transcription was initiated at the authentic *E2A* late promoter of Ad2 DNA. Control constructs carrying the unmethylated histone H2 gene were actively transcribed when co-injected with the methylated and silenced *E2A* construct. Hence, there was no evidence for possible unspecific inhibitory effects exerted by the *in vitro* premethylated construct. Modification of the *E2A* construct by the *Bsu*RI (5′-GG*CC-3′) DNMT did not inactivate transcription [85]. These data provided direct evidence for the notion that 5′-CG-3′ sequence-specific promoter methylation was involved in the silencing of eukaryotic genes. The system was further refined by separating the promoter of the *E2A* gene from its body and preparing both DNA fragments in quantitative amounts. We then 5′-CCGG-3′ methylated either the promoter or the gene body part of the constructs. Subsequently, the methylated promoter was re-ligated to the unmethylated body of the *E2A* gene and, conversely, the unmethylated promoter was reattached to the methylated *E2A* gene sequence. Upon microinjection into the nuclei of *Xenopus laevis* oocytes, only the construct, in which the promoter had been methylated, was inactivated. The construct with an unmethylated promoter but a methylated gene body was actively transcribed [86]. We interpreted these data to demonstrate that sequence-specific promoter methylation led to gene inactivation.

The *E1A* promoter of Ad12 DNA. A similar set of experiments was performed with constructs that carried the chloramphenicol acetyltransferase (CAT) gene as an indicator for gene activity under the control of the *E1A* regulatory region of Ad12 DNA. Methylation of the two *Hpa*II (5′-CCGG-3′) or of the seven *Hha*I (5′-GCGC-3′) sequences in this promoter inactivated the *CAT* gene or severely decreased its activity at 48 h after the transfection of these constructs into mouse Ltk^- cells [87, 88]. Several additional sites in the promoter of the *E1A* gene of adenovirus type 12 were methylated, and the activity of the modified promoter was assessed with the *CAT* indicator gene. The C-residue methylation of two *Alu*I sites (5′-AGCT-3′) downstream from the TATA box had no effect

on promoter activity. However, when one *EcoR*I (5′-GAATTC-3′) sequence, 281 bp upstream, or one *Taq*I (5′-TCGA-3′) site downstream from the TATA signal in the promoter was deoxyadenosine methylated, the promoter became silent [89]. Deoxyadenosine methylation of an *Mbo*I (5′-GATC-3′) sequence downstream of the TATA signal had no effect. Apparently, methylated nucleotides introduced at highly specific promoter locations can play an important role in the down-regulation of the Ad12 *E1A* promoter at least in transfection experiments. Since N^6-mA is not known to occur in mammalian DNA, the effect of N^6-mA on promoter activity has been unexpected. In an extension of this experimental approach, additional viral and non-viral eukaryotic promoters were tested for their sensitivity towards 5′-CG-3′ or 5′-CCGG-3′ methylation. The *CAT* or luciferase gene was used as activity indicator 24 h after the transfection into different human cell lines (HeLa, PA-1, 293). The methylation of all 5′-CG-3′ sequences by the *Sss*I DNMT inactivated the *E2A* late promoter of Ad2 DNA, the human cytomegalovirus promoter, the *TNFα* promoter, the herpes simplex virus thymidine kinase promoter, and decreased the activity of the *SV40* early promoter [71]. In some experiments, *Hpa*II methylation just led to a decrease in genetic activity of some of these constructs.

The *L1140* promoter of frog virus FV3 DNA. The resistance of this promoter to complete 5′-CG-3′ methylation and its full activity in fish or mammalian cells in the completely methylated state has been described above [82, 83].

The *p10* promoter of the AcNPV insect virus. A construct, which contained the promoter of the *p10* gene of the insect virus *Autographa californica* nuclear polyhedrosis virus (AcNPV) and the *CAT* indicator gene, was active in AcNPV-infected *Spodoptera frugiperda* insect cells at 18 h after transfection of the construct. When the three 5′-CCGG-3′ (*Hpa*II) sites in the promoter and its downstream region were methylated, the *p10* promoter was silenced [90]. Although insect cells may contain only minor amounts of m^5C, the activity of an AcNPV insect virus promoter could be shown to be sensitive to sequence-specific methylation.

Human Alu sequences transcribed by RNA polymerase III. We have also demonstrated that the polymerase III transcription of Alu sequences associated with the human angiogenin, the tissue plasminogen activator (TPA), or the α_1-globin gene is inhibited by 5′-CG-3′ methylation of these sequences [91]. Their methylation also interferes with the binding of proteins to the B control region of these Alu sequences [92].

Bending of promoter DNA sequences due to methylation? The site-specific methylation in 5′-CCGG-3′ (*Hpa*II), 5′-CGCG-3′ (*Fnu*DII), or in 5′-CG-3′ (*Sss*I) sequences of the *E2A* promoter, the polymerase III-transcribed VAI RNA gene of Ad2 DNA or of the human angiogenin gene-associated Alu sequence can alter the electrophoretic mobility of these DNA sequences in non-denaturing polyacrylamide gels. This finding indicates that the bending of the tested sequences might be altered by DNA methylation [93].

Reversal of promoter inactivation by methylation. An adenovirus *E1A* gene product or the strong enhancer of human cytomegalovirus (HCMV) can overcome the transcription-inactivating effect of promoter methylation. The removal of the methyl group from m^5C in a methylated promoter in the absence of DNA replication seems to be a rare event. Hence, other mechanisms for transient reactivation of a permanently methylated promoter appear to be required. Of course, experimentally, the methylated *E2A* late promoter in the Ad2-transformed cell line HE3 can be demethylated and reactivated by growing the cells in culture in the presence of 50 μM 5-azacytidine (5-aza-C), an inhibitor of maintenance methylation [94]. This approach provides support of principle but does not adequately mimic the situation in a biological system.

In human 293 cells, which carry the left terminus of Ad5 DNA chromosomally integrated and express the E1 region of Ad5 constitutively, the inactivating effect of 5′-CCGG-3′ methylation of an *E2A* promoter construct of Ad2 DNA is released or markedly decreased [95]. We have also shown that the *E1A* gene encoding the 13S RNA and the 289 amino acid (aa) protein of Ad2, a well-known transactivator of genes [96, 97], is responsible for the reversal of the inactivating effect of *E2A* promoter methylation [98]. It is unknown, by which mechanism the 289 aa *E1A* function is capable of effecting this reactivation. The methylated *E2A* promoter did not lose its 5′-CCGG-3′ methyl groups in the reactivation process at 48 h after transfection. Moreover, the authentic *cap*-site of this promoter was used in the transcription following reactivation [94, 98]. Similarly, the 5′-CCGG-3′ methylated *E2A* promoter of Ad2 DNA was active when the strong immediate early enhancer of HCMV DNA was inserted into the promoter–indicator gene construct in a position either immediately antecedent to the promoter or several thousand nucleotides remote from it [99]. Transcription was initiated also at the authentic *cap*-site of the *E2A* gene, and 5′-CCGG-3′ methylation remained unaltered at least during the duration of the transient expression experiment.

Promoter methylation and protein binding. This topic has been extensively investigated in several laboratories. In the *E2A* promoter system of Ad2 DNA, the *in vitro* methylation of 5′-CCGG-3′ sequences at nucleotides +24, +6, and –215 relative to nucleotide +1, the site of transcriptional initiation, was demonstrated to lead to transcriptional inactivation in transient expression studies in *Xenopus laevis* oocytes [86, 95], in mammalian cells [95], after the genomic fixation of the promoter in mammalian cells [52], and in a cell-free transcription system using nuclear extracts from human HeLa cells [100].

DNA fragments, 50 or 73 bp in length, comprising the +24 and +6 5′-CCGG-3′ sequences of the *E2A* promoter of Ad2 DNA in the unmethylated, methylated, or hemimethylated state were incubated with partly purified nuclear extracts from human HeLa cells. Protein binding to these DNA preparations was assessed by electrophoretic mobility shift assays (EMSAs). The formation of one of the observed DNA–protein complexes in this system was compromised when the construct was methylated or hemimethylated [101]. The results of the necessary competition experiments confirmed the interpretation that specific promoter methylation interfered with the binding of nuclear proteins from human cells. There was evidence that the AP2 transcription factor was among the proteins sensitive to promoter methylation in this system [102].

PATTERNS OF DNA METHYLATION IN THE HUMAN GENOME

A more profound understanding of the multifaceted biological functions of DNA methylation in mammalian and other genomes will remain elusive unless we have at hand the complete nucleotide sequences of these genomes including the fifth nucleotide. Researchers interested in the function of m^5C have, therefore, been disappointed by the nevertheless admirable results of the human genome project. The human epigenome project has been initiated, and its results, once at least partly completed, will undoubtedly fill a serious gap in the anatomy of the human genome.

In the early 1990s, my laboratory has begun, as a pilot project as it were, to study DNA methylation patterns in various parts of the human genome. A part of these results has been recently summarized [5]. A more complete survey will be presented here. Our studies also had the aim to contribute to the understanding of epigenetic mechanisms and of human disease.

Interindividual concordance in human DNA methylation patterns. We have asked the question of how tightly preserved patterns of DNA methylation actually are in the promoter and 5′-upstream regions of a human gene among several individuals of different ethnic origins. The human, like many other eukaryotic genomes are characterized by the existence of complex patterns of DNA methylation which reflect in an unidentified way states of gene activities and inactivities and, equally important and related, of the chromosomal structure of the (human) genome. The 5′-upstream and promoter regions of the human genes for tumor necrosis factors TNFα and TNFβ were screened with the genomic sequencing technique [9, 10] for the presence of m^5C residues [17, 18].

Human DNA was derived from peripheral blood granulocytes, lymphocytes, or from sperm. The results indicated that patterns of DNA methylation at least in these genome segments were interindividually highly conserved. Thus, in the TNFα DNA from granulocytes of 15 individuals of African, Caucasian, or Chinese origin, the m^5C residues were consistently found in 5′-CG-3′ dinucleotide positions -IX, -X, -XI (upstream of the *cap*-site) and in position +XVI (downstream of it). Very different distributions of m^5C residues were observed in human cell lines HL60, Jurkat, and RPMI 1788. The *TNFα* gene is transcribed in human granulocytes. A very different result emerged for the promoter and upstream regions of the human gene for TNFβ, which is not transcribed in human granulocytes. All 13 5′-CG-3′ dinucleotides in this segment were methylated, two only hemimethylated. Again, this pattern held true in the granulocytes from nine different individuals. The same sequence was completely unmethylated in human lymphocytes from the same individuals, in sperm and in the human cell lines RPMI 1788 and HL60, but almost completely methylated in cell line Jurkat [17]. These data document that methylation patterns in human DNA can be very different in different cell lines, but can be interindividually highly concordant. Moreover, patterns of DNA methylation in specific genome segments can vary a great deal.

The patterns of DNA methylation in the human *TNFα* and *TNFβ* genes in granulocytes, monocytes and in several cases of acute (AML) or chronic myeloid leukemia (CML) were found to be very similar, except for one AML, in which the region in the *TNFα* gene was completely unmethylated, and several leukemia cases in which many sites in the *TNFβ* gene were only hemimethylated [18]. In T and B lymphocytes of many individuals and in a number of Hodgkin and non-Hodgkin lymphomas, both the *TNFα* and *TNFβ* genes were un- or hypomethylated. The DNA in HeLa cells in culture was completely methylated in the upstream and promoter regions of both genes [18]. If leukemia- or lymphoma-specific patterns should exist, they are very complex and not readily recognizable by this type of analysis. We have also compared methylation patterns by *Hpa*II (5′-CCGG-3′) and *Hha*I (5′-GCGC-3′) cleavages of human DNA from European and Japanese individuals across about 500 kb of randomly selected DNA sequences in the human genome and found complete interindividual congruence of patterns by this method of admittedly intermediate sensitivity [19].

Methylation patterns in genetically imprinted regions of the human genome. The Prader–Willi/Angelman region on chromosome 15q11-q13 of the human genome is genetically imprinted, i.e. on the maternally and on the paternally inherited chromosome different genes are activated and others silenced. The molecular mechanisms underlying genetic imprinting are not completely understood. However, there is much evidence that differences in the methylation patterns in imprinted regions on the two alleles play an important role in the imprinting phenomenon. As part of a study on DNA methylation pat-

terns in the human genome, we investigated all 5′-CG-3′ dinucleotides in the vicinity of exon 1 of the *SNRPN* and the *D15S63* loci on chromosome 15q. The SNRPN transcripts might be involved in imprint switching during gametogenesis. By using the bisulfite genomic sequencing technique, we looked at individual chromosomal PCR products from normal individuals, from Prader–Willi and from Angelman patients. In this region, non-5′-CG-3′ C-residues were never methylated. Around exon 1 of the *SNRPN* gene, >96% of the 23 5′-CG-3′-dinucleotides were methylated on the maternal chromosome, as apparent from the genomic sequencing data with DNA from Prader–Willi patients in whom this segment was deleted on the paternal chromosome. In contrast, the same region on the paternal chromosome was completely devoid of methylated 5′-CG-3′ dinucleotides [20]. Angelman syndrome patients carry a deletion of the region on the maternal chromosome. The methylation status in the *D15S63* locus, however, was quite different in that only two *Cfo*I/*Hha*I sites were methylated to >96% on the maternal chromosome. The remaining five 5′-CG-3′ dinucleotides in this segment were methylated only 45-70% on the maternal, and to only 5-14% on the paternal chromosome [20, 103]. In an extension of this study [21], it was demonstrated again by bisulfite genomic sequencing that the 16 5′-CG-3′ dinucleotides in the 1.15 kb ASSRO region on human chromosome 15q were methylated to 83 to 87% on both the maternal and paternal chromosomes in healthy individuals as well as in Prader–Willi and Angelman syndrome patients. There may be a low degree of mosaicism but there are no parent-of-origin-specific differences in the methylation patterns in this part of the genome [21]. These findings attest to the significant variability of the methylation patterns even in imprinted parts of the human genome.

Patterns of 5′-CG-3′ methylation in the promoter of the *FMR1* gene; relevance for the fragile X syndrome. In patients suffering from the fragile X (FRAXA) syndrome, a naturally occurring 5′-$(CGG)_n$-3′ repeat in the promoter and the 5′-untranslated regions (5′-UTR) of the *FMR1* gene on human chromosome Xq27.3 is expanded excessively. In normal individuals, the value *n* ranges between 6 and 40, in premutation females *n* assumes values between 40 and 200, while in affected individuals the repeat *n* lies between 200 and >2000. The expanded repeat is hypermethylated, perhaps because such expansions are recognized as foreign DNA and become subject to modification. The ensuing inactivation of the *FMR1* gene is the most likely cause for the disturbed embryonal and fetal development, which is the basis for the syndrome [104].

By applying the bisulfite genomic sequencing technique, we determined the methylation profiles in the promoter and 5′-UTR of the *FMR1* gene on single chromosomes of healthy individuals and of selected premutation carriers and FRAXA patients [24]. In the DNA from FRAXA patients, there is considerable variability in the lengths of the 5′-$(CGG)_n$-3′ repeats and in the levels of methylation in the repeats and the 5′-UTR regions in that all patients seem to be mosaics with respect to both parameters. In one patient with repeat lengths between $n = 15$ and >200, shorter repeats ($n = 20$ to 80) were methylated or unmethylated, longer repeats ($n = 100$ to 150) were often completely methylated. A particular repeat in this patient with $n = 160$, proved to be completely devoid of m^5C residues. This repeat mosaicism was observed in several FRAXA patients analyzed in our laboratory [24]. As expected for healthy females with one at least partially inactivated X chromosome, hypermethylated repeats and 5′-UTR sequences were found. We also demonstrated that the authentic *FMR1* promoter from healthy individuals was sensitive to methylation as shown by comparing the transient activities of constructs carrying the luciferase gene under the control of the unmethylated or the *Sss*I (5′-CG-3′) completely methylated *FMR1* promoter in human HeLa or 293 cells [24]. The methylated, inactive *FMR1* promoter regions do not bind to specific cellular proteins as determined by footprinting analyses, whereas active, unmethylated promoter regions do bind proteins [23].

The promoter and 5′-upstream region of the *RET* protooncogene, a gene involved in the causation of Hirschsprung disease. The *RET* (rearranged during transformation) protooncogene plays a role in the causation of some familial or sporadic cases of Hirschsprung disease which results from an impaired development of the neural crest-derived neurons of the enteric ganglia [105]. We investigated the level of DNA methylation in a DNA segment of about 1000 bp in the promoter and 5′-upstream region of this gene [22]. By again applying the bisulfite genomic sequencing technique, DNA from peripheral white blood cells (PWBCs) from healthy individuals or from Hirschsprung disease patients was used as well as DNA from different human tissues and from a human embryonic kidney cell line. In a DNA section starting about 790 bp upstream from the transcriptional start site, a few m^5C residues were found. However, in a 5′-CG-3′ rich 400 bp stretch in the *RET* gene promoter with 49 such dinucleotide pairs not a single m^5C residue was present, although the *RET* protooncogene was not transcribed in many human tissues. Weak transcriptional activity was detected in many neural crest-derived human tissues. Obviously, the *RET* gene promoter was not silenced by a long-term signal like promoter methylation, possibly because it had to be expressed occasionally, and its transcription was controlled by factors other than DNA methylation. In *in vitro* experiments, in which the transcriptional activity of the *RET* gene promoter was assessed in linkage to an indicator gene after transfection into human cells, the activity of this promoter was decreased by *Hpa*II (5′-CCGG-3′) methylation and abolished by *Sss*I (5′-CG-3′) methylation. Hence the

RET protooncogene promoter is sensitive to DNA methylation at least in transfection and transient transcription experiments [22].

Genes for proteins in the human erythrocyte membrane. Alterations in the structure of the erythrocyte membrane can be related to mutations in specific genes for proteins, which are essential elements of this membrane. These structural alterations of the erythrocyte membrane are responsible for hematologic diseases like hereditary elliptocytosis or hereditary spherocytosis [106]. By the bisulfite genomic sequencing procedure we determined patterns of methylation in the promoter and 5′-regions of the following human genes: the protein 4.2 (P4.2) gene (*ELB42*), the band 3 (B3) gene (*EPB3*), and the β-spectrin (β-Sp) gene (*SPTB*) [25, 26]. The promoter regions of the *EPB3* and *ELB42* genes were extensively methylated, whereas the promoters of the *SPTB* and the ankyrin genes were unmethylated. This finding again points to the interindividual conservation of certain patterns in the distribution of m^5C residues in the human genome. The human *SPTB* promoter conforms to expectations in that it is unmethylated and fully active throughout erythroid development. In contrast, high levels of promoter methylation correlate with promoter activity for the *EPB3* and *ELB42* genes during their sequential activation in erythrocyte differentiation (Remus et al., submitted). In this respect, the *EPB3* and *ELB42* genes may resemble the genes of frog virus 3. This analysis was extended to patients with red cell membrane diseases, such as complete P4.2 deficiency due to mutations in the *ELB42* gene, with hereditary spherocytosis with *EPB3* mutations, and to hereditary elliptocytosis with mutations in the *SPTB* gene. Patterns of methylation in these patients were in general very similar to those of normal individuals [26].

Promoter and exon 1 of the human gene for the interleukin 2-receptor α-chain (*IL-2Rα*). The *IL-2Rα* gene is expressed in stimulated, but not in resting human T lymphocytes and plays an important role in promoting the T cell-mediated immune response. The −300 to +300 promoter/exon 1 region of the *IL-2Rα* gene was analyzed by the genomic sequencing technique for its content of methylated 5′-CG-3′ dinucleotides. In the cell types investigated—sperm, placenta, granulocytes, T-CLL, B-CLL, Jurkat, KB cells—m^5C residues were not found in unstimulated or in stimulated lymphatic cells [107]. The 5′-CG-3′ sequence in position +198 was partly methylated. Even in cell types not relevant for the immune response, like in the human KB cell line, this regulatory region was consistently unmethylated. The promoter of a functionally essential human gene would not be long-term silenced by the methylation signal or else it could not be flexibly reactivated upon demand. Obviously, transient mechanisms of gene shut-off would be operationally preferred in these instances and thus remain independent of DNA methylation.

Human Alu sequences associated with specific genes. The human Alu sequences belong to the short interspersed repeat elements (SINE), comprise about 5% of the human genome and are about 300 bp long. The Alu elements might have been derived from reverse transcripts since they carry a 3′-dA-rich track. We analyzed the state of DNA methylation in the human Alu sequences associated with the genes for α_1-globin, the tissue plasminogen activator (TPA), the adrenocorticotropic hormone (ACTH), and for angiogenin. DNA was investigated from lymphocytes, granulocytes, brain, heart muscle, and sperm as well as from human HeLa and KB cells. Both methylation-sensitive restrictases and genomic sequencing techniques were employed. In primary human cells, these Alu elements were highly methylated, but there were distinct differences in specific Alu sequence elements. In the DNA from haploid spermatozoa, Alu elements were often hypomethylated. The *in vitro* transcription of Alu elements was inhibited by 5′-CG-3′ methylation of these sequences [91]. The patterns observed in these specific Alu elements were identical in different individuals. The high level of DNA methylation in the Alu sequences associated with specific genes was consistent with their transcriptional silencing.

Promoter of the polymerase I-transcribed human ribosomal genes. The 5′-CG-3′-rich promoter region of the DNA-dependent RNA polymerase I (rDNA) genes was analyzed for the presence of m^5C nucleotides by the genomic sequencing technique in DNA from primary human cells, from human tumor cells, and from human cell lines [108]. In none of the primary human cells or tumor cells was the rDNA promoter methylated. In contrast, in the DNA from the human cell lines HeLa (cervical cancer), KB (oral cancer), Jurkat (T cells), or CEM (T cells), the 5′-CG-3′ dinucleotides were methylated between 50% (KB) and 85% (Jurkat). Apparently, in the primary human cells, in granulocytes, T lymphocytes, spermatozoa, as well as in chronic T cell (T-CCL), myeloid (CML), or B cell (B-CLL) leukemia cells, which are all actively dividing, the essential rDNA genes need be transcribed actively and are not methylated [108]. In cell lines, rDNA genes are also actively expressed, and alternate mechanisms of overexpression must exist. Could some of the rDNA genes be over-amplified?

Randomly selected human genes in different Hodgkin's lymphoma and leukemia cell lines and in normal human lymphocytes. Several of the protooncogenes, parts of the *TNFα* and *TNFβ* genes, the insulin receptor and lamin C genes were investigated by using the methylation-sensitive restriction enzyme *Hpa*II and the control enzyme *Msp*I. There were regions completely devoid of methylation; others were completely or partly methylated. Various lymphoma and leukemia cell lines differed among each other in different regions of the genome and differed again from the patterns observed in normal primary human lymphocytes [109]. Obviously, there is great

variability, and no simple rules can be derived for the general characteristics of methylation patterns in leukemic versus normal human white cells.

The promoter of the human 5′-(CGG)$_n$-3′-binding protein (CGGBP1). From the nuclei of human HeLa cells, we isolated a 20 kD protein which binds specifically to 5′-(CGG)$_n$-3′ sequences [110, 111] and which might play a role in the control of promoters rich in CGG sequences like the *FMR1* promoter in human DNA [112]. The human gene for this protein, termed CGGBP1, was located to human chromosome 3p, and its promoter was characterized in detail [113]. In several different human cell types, this promoter was unmethylated. The complete *in vitro* premethylation of all 18 5′-CG-3′ dinucleotides in this promoter led to its inactivation upon transfection into human HeLa cells with the luciferase gene as activity indicator [113].

Towards a complete nucleotide sequence of the human genome [114]. Projects have been initiated to complete the human genomic DNA sequence by including the 5th nucleotide and to initiate a Human Epigenome Project [11, 12]. Obviously, this will be a very important, but at the same time demanding, task. Above, I have summarized our contributions, of course not towards the solution, but towards a more general appreciation of the problem that the mammalian genomes harbor functionally important patterns of DNA methylation. The distribution of m^5C residues in the human genome is thought to be essential for the understanding of chromatin structure and of the regulation of human gene expression in the many different cell types and during development.

From the available data, the following list of problems and desirable approaches can be compiled. Of course, we want to be cautious and cannot claim general validity of any of the presently plausible observations and conclusions:

– patterns of m^5C distribution across the human genome are highly specific. Each region, each promoter exhibits its own individual pattern. The patterns can be interindividually conserved at least in several regions of the human genome;

– it is likely that each human cell type could have a different pattern in each genome segment;

– long-term promoter inactivity is generally associated with hypermethylation of the promoter. Inactive promoters can, however, also be un- or hypomethylated, particularly when they have to be occasionally reactivated. The state of promoter methylation by itself cannot reveal the activity status of a promoter;

– promoter strength might affect the pattern imposed upon a particular promoter;

– there are distinct differences in the patterns of methylation between normal human lymphocytes and lymphoma or leukemia cell lines in different segments of the genome. It is at present not possible to derive functionally meaningful conclusions from these differences other than that there are extensive alterations. We pursue the possibility that the process of oncogenic transformation of a human cell is associated, possibly causally, with global changes in the genome organization which is also reflected in drastically altered methylation patterns. From our work on Ad12-induced hamster tumors and on Ad12 DNA- or bacteriophage λ DNA-transgenic hamster cells, we consider it likely that the insertion of foreign DNA is at least partly responsible for these alterations. Of course, it remains to be determined whether any of the global changes in methylation and transcription patterns demonstrated in tumor cells are the cause or the consequence of oncogenic transformation.

ALTERATIONS OF CELLULAR DNA METHYLATION UPON FOREIGN DNA INSERTION

In the Ad12-transformed hamster cell line T637 with 15 copies of viral DNA inserted at a single chromosomal site, extensive alterations, mainly increases, in the levels of DNA methylation in the *Hpa*II (5′-CCGG-3′) and *Hha*I (5′-GCGC-3′) sequences were apparent in the retrotransposon sequences of the about 900 copies of intracisternal A particle (IAPI) genomes [62]. The ~900 copies of IAPI sequences are a constitutive part of the hamster genome [115, 116]. In Ad12-transformed hamster cells, extensive changes in DNA methylation were also noted in the MHC class I and II, the Ig Cμ, the serine protease, and cytochrome P450 genes of the hamster cell genome. At least in the IAPI sequences, the increases in DNA methylation persisted in the revertant TR3 which had lost all 15 copies of Ad12 DNA [62]. Apparently, the alterations of the methylation patterns in the cellular genome are not dependent on the continued presence of the viral transgene DNA. The ~900 copies of IAPI DNA are distributed among many of the hamster chromosomes, often on their short arms [117, 118]. Since the increases in IAPI methylation were extensive, the effects of the Ad12 DNA insertion at a single chromosomal site had to transgress this site and lead to a disturbance in DNA *trans*-methylation patterns in the cellular genome even on different chromosomes.

Repetitive sequences in the mammalian genome appear to be particularly prone to respond with altered methylation patterns to perturbations in the genome caused by foreign DNA insertions. We surmise that the selection of the genes and DNA segments influenced in *trans* by the foreign DNA integration event might depend on the site of transgene insertion. The mechanism of this modulation of DNA methylation in the recipient genome remains unknown; it might be sought in the direct interaction of neighboring chromosomes. Soluble factors could obviously also have impacts. Further evidence in support of the contributions that foreign DNA insertions

rendered in altering DNA methylation patterns in the recipient genome came from experiments in which we generated clonal hamster BHK21 cell lines with multiple copies of integrated bacteriophage λ DNA. The integration phenomena of λ DNA resembled those of Ad12 DNA in that multiple copies of the phage DNA came to reside at a single site of the hamster chromosome and became progressively *de novo* methylated. However, in contrast to the integrated Ad12 DNA, the integrated λ DNA was not detectably transcribed. Alterations in cellular DNA methylation patterns were also observed in the IAPI sequences and could be unequivocally documented by the bisulfite genomic sequencing method with which 35 5′-CG-3′ dinucleotide positions were analyzed in a subsegment of the IAPI DNA region [119]. Even a transcriptionally inert transgene, like λ DNA, had led to alterations in the methylation profiles in the IAP transposons.

The question arose whether such differences in methylation patterns among the different copies of the ~900 IAPI equivalents might have preexisted in different BHK21 cell clones. We therefore examined >70 individual BHK21 cell clones for differences in methylation patterns in the investigated IAPI segment both by *Hpa*II and *Hha*I restriction patterns and by bisulfite genomic sequencing. Differences in patterns were not detectable [119]. Of course, we could not scrutinize thousands of individual cell clones for homogeneous methylation patterns. Nevertheless, on the basis of the available evidence, the preferred interpretation of a causative effect of foreign DNA integration on methylation patterns in *trans* will be pursued in future experimental projects.

We also entertained the possibility that the abortive infection of BHK21 cells with Ad12 [79, 120] with the transcription and expression of exclusively early Ad12 gene products [121] might have affected the stability of cellular DNA methylation patterns. At least on a time scale of days after Ad12 infection, changes in patterns of IAPI genome methylation could not be documented in BHK21 cells [119]. Productively Ad12-infected human or Ad2-infected hamster cells will also be examined for global changes in methylation patterns in the cellular genomes.

The method of methylation-sensitive representational difference analysis (MS-RDA) is based upon a subtractive hybridization protocol after selecting against DNA fragments, which were heavily methylated and, therefore, not cleaved by the *Hpa*II restriction endonuclease [122]. We applied this method to the investigation of transcripts from bacteriophage λ DNA-transgenic hamster cell lines in comparison to hamster cell lines devoid of integrated λ DNA [123]. By using the suppressive hybridization technique for the analysis of cDNA preparations from non-transgenic, Ad12 DNA-transgenic and λ DNA-transgenic hamster cells, several cellular genes were cloned which had altered transcriptional profiles in the transgenic as compared to the non-transgenic cells. Among individual non-transgenic hamster cell clones investigated as negative controls, no differences in cDNA isolates, and hence transcriptional profiles, were observed. We also studied these changes in one λ DNA-transgenic mouse strain: hypermethylation was found for the imprinted *IGF2R* gene for DNA from heart muscle. Two mouse lines transgenic for an Ad2 promoter–indicator gene construct showed hypomethylation in the *interleukin 10* and *IGF2R* genes. We concluded that in Ad12 DNA- or λ DNA-transgenic hamster cells or mice, cellular methylation and transcription patterns can be critically altered [122]. Detailed investigations on the heterogeneity of DNA transcription patterns in about 1170 genes among individual clones of BHK21 and T637 cells have revealed only minimal differences in five of these genes by DNA array analyses between the two cell lines and among different clones of each cell line (N. Hochstein and W. Doerfler, unpublished experiments).

Since the insertion of foreign DNA into established mammalian genomes has become a preferred regimen in experimental biology, e.g., in the generation of transgenic organisms, and increasingly also in gene therapy, I consider it an important problem to pursue the unanticipated, likely unwanted, effect of foreign DNA integration on the stability of the recipient genome. Alterations of patterns of DNA methylation might be merely one, but an experimentally recognizable, manifestation of this disturbance [7]. These problems may have considerable relevance for certain regimens in gene therapy in which the fixation of foreign DNA in an established human genome is considered. When retroviral gene transfer vectors were used to chromosomally fix the human adenine deaminase gene in children with hereditary immunodeficiency, rare T cell leukemias developed. In these cases, I consider the insertion of foreign DNA as one of the decisive factors explaining this unfortunate outcome of a well-intended medical procedure.

Towards a working hypothesis on viral oncogenesis. Viral oncogenesis is frequently accompanied by the integration of the viral genome into the genome of the transformed cell. Integration of viral DNA is a *conditio sine qua non* for transformation [124] in cells transformed by adenoviruses, by SV40, polyoma virus, by the papilloma viruses, and notably by retroviruses. Integration is, of course, an important mode of chromosomal fixation and continued expression of the viral genome in the transformed cell. In retroviral replication, proviral integration is an essential step in the viral life cycle. Conventionally, major attention has been directed towards the function of the expressed viral gene products to explain the mechanism of viral oncogenesis. Having identified the viral "culprit" does not exclude the possibility that the real action is somewhere else, namely in its direct effect on the recipient genome. For some time, we have pursued the possibility that the alterations of DNA methylation pat-

terns enacted in the wake of viral DNA insertion are a general phenomenon following the insertion of any foreign DNA [3-5, 7]. Altered methylation patterns then might be an indicator of more general perturbations in the cellular genome, which reach far beyond the immediate site of viral DNA integration. Furthermore, altered methylation patterns forebode changes in transcriptional patterns as well. Hence, upon foreign DNA insertion the recipient genome has undergone dramatic functional alterations, which might well be at the center of the oncogenic transformation process. Using the Ad12–hamster tumor system as a very efficient experimental model, we have only begun to document changes in cellular transcription patterns in Ad12-induced tumors [51].

STUDIES ON TRANSGENIC MICE: STABILITY OF PATTERNS OF DNA METHYLATION AND GENETIC BACKGROUND IN DIFFERENT STRAINS OF MICE

A construct consisting of the *E2A* late promoter of Ad2 DNA and the *CAT* indicator gene was integrated in the non-methylated or in the 5′-CCGG-3′ premethylated form into the genome of mice, and the state of methylation was analyzed by *Hpa*II cleavage of DNA from various organs of the transgenic animals [125]. In general, the transgenic construct remained stably integrated. In the founder animal, the non-methylated construct became *de novo* methylated at all or at most of the 5′-CCGG-3′ sites. Pre-imposed methylation patterns were stable for up to four generations beyond the founder animal. However, in the DNA from testes of two founder animals and two F_1-males, the premethylated, transgenic DNA was demethylated by an unknown mechanism. In all other organs, the transgenic DNA preserved the pre-imposed 5′-CCGG-3′ methylation patterns. Differences in these transmission modes were not seen depending on whether the transgene was inherited maternally or paternally [125]. There are studies to support the notion that genetic background in mice can have a decisive influence on the type of *de novo* methylation patterns imposed on a foreign DNA transgene and on their stability [126-128]. The molecular mechanisms involved in the "modifier gene" effects are not understood. We addressed this problem by introducing into the genomes of different mouse strains (DBA/2, 129/sv FVB/N or C57BL/6, CB20 or Balb/c) a construct, which consisted of the *E2A* late promoter of Ad2 DNA and the chloramphenicol acetyltransferase (CAT) gene as reporter. The patterns of *de novo* transgene methylation were transmitted to the offspring and remained stable for 11 backcross generations, regardless of the heterozygosity in the recipient mouse strain and the presence of presumptive modifier genes. In seven additional mouse strains carrying the same transgene in different chromosomal locations, strain-specific alterations of methylation patterns were not observed [129].

We also investigated the stability of DNA methylation patterns in the *Snurf/Snrpn* imprinted gene cluster in mouse embryonal stem cell lines cultured under different experimental conditions, like prolonged passaging, trypsinization, mechanical handling, single cell cloning, staurosporin-induced neurogenesis [130] or the insertion of foreign (viral) DNA into the ES cell genome. None of these *in vitro* manipulations affected the stability of the methylation patterns in the analyzed gene cluster [131]. Growth-related genes, *IGF2*, *H19*, *IGF2R*, *or GRB10*, are known to respond by altered imprint patterns. The analyzed neuronal gene cluster, however, exhibited stable patterns of DNA methylation under the experimental conditions chosen.

FATE OF FOOD-INGESTED FOREIGN DNA IN THE GASTROINTESTINAL TRACT OF MICE

The tempting interpretation that DNA methylation, particularly the *de novo* methylation of integrated foreign DNA, is part of an ancient cellular defense mechanism raises a number of questions. One of the obvious ones relates to the major origins of foreign DNA, e.g. in mammals. Virus infection as such a contingency has been extensively discussed in this chapter. Another apparent source of large amounts of foreign DNA all organisms are constantly exposed to is the DNA orally ingested with the food supply. We have therefore undertaken a study on the fate of food-ingested foreign DNA in mice as model organism. I will present a short summary of the major results my laboratory adduced in a project which we initiated in 1988.

In mammals, the gastrointestinal tract is the main portal of entry for foreign macromolecules, and its epithelial lining presents the immediate sites of contact with foreign DNA and proteins. In our investigations on the fate of foreign DNA in the digestive tract, we fed naked test DNAs of various derivations to laboratory mice at between two and six months of age [132-134]. The DNA of bacteriophage M13, the DNA of human adenovirus type 2, or the gene for the green fluorescent protein (GFP) from *Aequorea victoria* were administered as test DNAs in different experiments. None of these DNAs had homologies to bacterial or mouse DNA, except for perhaps very short stretches of DNA sequence, which were then excluded from being used for the detection of the foreign DNA in the mouse organism.

In later experiments, we fed leaves of the soybean plant to mice and followed the fate of the strictly plant-specific Rubisco (ribulose 1,5-diphosphate carboxylase) gene. During the passage through the gastrointestinal tract of mice, the bulk of the administered DNA is completely degraded. However, a few percent of the test

DNAs resist the digestive regimens of the gut and can be recovered for several hours after feeding in various parts of the intestinal tube as fragments between 1700 nucleotides (nt) (rare) and a few 100 nt. By applying a variety of techniques (Southern blotting, polymerase chain reaction (PCR), FISH, and rescue of the test DNA fragments by recloning and resequencing) the test DNAs could be followed to the wall of the intestinal tract, particularly the colon, to Peyer patches, peripheral white blood cells, and to cells of the liver and spleen [132-136]. When pregnant animals were test DNA fed, fragments of the test DNA could be traced by FISH and PCR to clusters of cells in various organs of the embryo, but never to all its cells. Moreover, when mice were fed daily and continuously for eight generations, transgenic animals were never observed. Hence, we assume that the germ line must be protected from the exposure to and the uptake of food-ingested foreign DNA. Moreover, we never obtained evidence for the test DNA being transcribed in any of the organ systems of the adult animals, which had been given test DNA [135]. The possible transcription of test DNA was assessed by RT-PCR, the most sensitive technique to detect trace amounts of specific transcripts. After feeding mice daily for one week, test DNA could be recloned, however extremely rarely, from the spleen of the animals. In a few of these clones, mouse specific DNA was found adjacent to the test DNA in the cloned DNA. Further proof will be required to investigate the possibility of whether foreign DNA could be integrated into the genome of defense cells in the recipient animals [133]. In a completely independent approach, we could demonstrate that the protein glutathione-S-transferase, a rather stable protein, survived in the stomach and small intestine of mice for up to 30 min after feeding [136].

Taken together, the results of this series of investigations indicate that foreign macromolecules, particularly the very stable DNA, can survive in the gastrointestinal tract at least transiently in small amounts and in fragmented form and can get access to various organ systems of the mouse. Even stable proteins survive only for a very short time in the gastrointestinal tube. We have not found any evidence for the entry of foreign DNA into the germ line, nor could we demonstrate transcription of foreign DNA in any of the organ systems tested. It is not known whether a tiny proportion of the thus persisting DNA may find entry into the genome of a rare defense cell and remain there with unknown functional consequences. These questions will be worth pursuing.

SYNOPSIS AND CONCLUSIONS

It appears that the following data and interpretations presented in this review have stood the test of time. Research, of course, is a never-ending enterprise, and conclusions have always to be considered subject to change as new data and concepts are being adduced. Here is a synopsis of concepts on the biological significance of m^5C in the genome the author feels reasonably certain about at the time of this writing (December 2004).

The virus particle (virion)-encapsidated genomes of most mammalian DNA viruses are not methylated. Likewise, cellular DNA haphazardly integrated into an adenoviral genome, which becomes virion enclosed, does not become methylated, irrespective of its methylation status in the genome it has been derived from. In contrast, the DNA of frog virus 3, an iridovirus, is extensively, probably completely, methylated.

The concept of sequence-specific promoter methylations being causally related to long-term gene silencing, which has been first deduced from work on adenovirus promoters, has proved to be generally applicable in most eukaryotic genomes. Frog virus 3 promoters are an interesting exception to this apparent rule. Concomitantly with promoter methylation, histone modifications and perhaps modifications of additional proteins involved in chromatin structure play a decisive role in the regulation of promoter activity. At this time, it seems undecided whether DNA or protein modifications initially orchestrate these regulatory processes. It is likely that a refined interplay between both biochemical mechanisms comes close to the correct answer.

Foreign DNA, which has become integrated into an established mammalian genome, becomes *de novo* methylated in distinct patterns. The sites of initiation of *de novo* methylation at least in integrated Ad12 genomes are located paracentrally in the transgenomes and not close to the junctions with cellular DNA. In integrated Ad12 genomes, this localization of methylation initiation sites might be influenced by the transcriptional activity of the terminally located E1 and E4 regions of the Ad12 genome in the transformed cell lines or in Ad12-induced tumors, which are selected for the genetic activity of these viral genome segments. In any event, subsequent to initiation, *de novo* methylation extends continuously across the transgenomes in a spreading reaction. Initiation seems regional and does not emanate from a specific 5′-CG-3′ dinucleotide.

It is likely that hypermethylated or rather completely methylated transgenomes are more stably integrated than less completely methylated foreign DNA molecules. At the immediate sites of foreign DNA insertion, the patterns of cellular DNA methylation can be altered extensively.

Alterations of cellular DNA methylation patterns are, however, not restricted to the cellular junction sites [137, 138] but involve remote areas of the recipient genomes, even on different chromosomes. This *trans* effect is most striking in retrotransposon sequences, like the endogenous IAP DNA sequences in hamster cells, but can affect genuine cellular sequences as well. These remote perturbations of methylation patterns are not only

observed after the integration of Ad12 DNA, which is partly transcriptionally active, but also after insertion of transcriptionally inactive bacteriophage λ genomes. Possibly, ancient retrotransposons might be more responsive to local alterations of chromatin structure due to foreign DNA insertions into the recipient genome. There is evidence that in addition to alterations of methylation patterns in *trans*, the insertion of foreign DNA could also alter transcriptional patterns in the recipient genomes.

Many of the notions summarized here hold true not only for mammalian organisms but also for other eukaryotic genomes, particularly for those of plants.

In mammalian genomes, distinct patterns in the distribution of m^5C residues exist which, at least in humans, can be interindividually preserved in several (many?) genome segments. These patterns are specific for each genome segment and can be different from cell type to cell type. These observations constitute a major challenge to the so-called Human Epigenome Project.

The biological importance of these patterns, which have obviously been conserved over long periods of time, has not been clarified. Long-term gene silencing and chromatin structure as well as the defense against foreign retrotransposons may be factors of significance in explaining the nature of these patterns of genome methylation.

Many members of my laboratories in Koeln (1972 to 2002) and in Erlangen (2002 to the present) have essentially contributed to the data summarized in this article. Their work has been acknowledged in the references cited herein. I am indebted to Petra Böhm, Koeln, for expert editorial work.

During different times, our research on DNA methylation has been made possible by grants and/or support from the following organizations: Deutsche Forschungsgemeinschaft, Bonn – SFBs 74 and 274; Center for Molecular Medicine in Cologne (CMMC, TV13); Federal Ministry for Research and Technology (BMFT), Bonn; Amaxa GmbH, Köln; Bayerisches Staatsministerium für Landschaftsentwicklung und Umwelt-gestaltung, München; Alexander von Humboldt-Stiftung, Bonn; Thyssen Stiftung, Köln; Sander Stiftung, München; Fonds der Chemischen Industrie, Frankfurt; European Union, Bruxelles; Universität zu Köln; Institut für Klinische und Molekulare Virologie, Erlangen.

It is a great pleasure to contribute this chapter to a volume of *Biochemistry (Moscow)* published in honor of Professor Boris F. Vanyushin on the occasion of a special birthday. Over a period of more than 20 years, I have valued our personal and scientific friendship, which started at a time when scientists of our countries had to devise special tools to be able to meet.

Accept my cordial greetings, Boris Fedorovich!

REFERENCES

1. Doerfler, W. (1981) *J. Gen. Virol.*, **57**, 1-20.
2. Doerfler, W. (1983) *Annu. Rev. Biochem.*, **52**, 93-124.
3. Doerfler, W. (1995) *Adv Cancer Res.*, **66**, 313-344.
4. Doerfler, W. (1996) *BBA Rev. Cancer*, **1288**, F79-F99.
5. Doerfler, W. (2000) *Foreign DNA in Mammalian Systems*, Wiley-VCH, Weinheim-New York-Chichester-Brisbane-Singapore-Toronto.
6. Doerfler, W., Weisshaar, B., Hoeveler, A., Knebel, D., Müller, U., Dobrzanski, P., Lichtenberg, U., Achten, S., and Herrman, R. (1988) *Gene*, **74**, 129-133.
7. Doerfler, W., Hohlweg, U., Müller, K., Remus, R., Heller, H., and Hertz, J. (2001) *Ann. New York Acad. Sci.*, **945**, 276-288.
8. Hotchkiss, R. (1948) *J. Biol. Chem.*, **175**, 315-332.
9. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1827-1831.
10. Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994) *Nucleic Acids Res.*, **22**, 2990-2997.
11. Beck, S., and Olek, A. (eds.) (2003) *The Epigenome*, Wiley-VCH, Weinheim.
12. Rakyan, V. K., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., Fischer, J., Gut, I. G., Berlin, K., and Beck, S. (2004) *PloS. Biol.*, **2**, e405.
13. Hertz, J., Schell, G., and Doerfler, W. (1999) *J. Biol. Chem.*, **274**, 24232-24240.
14. Sutter, D., Westphal, M., and Doerfler, W. (1978) *Cell*, **14**, 569-585.
15. Sutter, D., and Doerfler, W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 253-256.
16. Church, G. M., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991-1995.
17. Kochanek, S., Toth, M., Dehmel, A., Renz, D., and Doerfler, W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8830-8834.
18. Kochanek, S., Radbruch, A., Tesch, H., Renz, D., and Doerfler, W. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 5759-5763.
19. Behn-Krappa, A, Hölker, I., Sandaradura de Silva, U., and Doerfler, W. (1991) *Genomics*, **11**, 1-7.
20. Zeschnigk, M., Schmitz, B., Dittrich, B., Buiting, K., Horsthemke, B., and Doerfler, W. (1997) *Hum. Molec. Genet.*, **6**, 387-395.
21. Schumacher, A., Buiting, K., Zeschnigk, M., Doerfler, W., and Horsthemke, B. (1998) *Nature Genet.*, **19**, 324-325.
22. Munnes, M., Patrone, G., Schmitz, B., Romeo, G., and Doerfler, W. (1998) *Oncogene*, **17**, 2573-2584.
23. Schwemmle, S., de Graaff, E., Deissler, H., Gläser, D., Wöhrle, D., Kennerknecht, I., Just, W., Oostra, B. A., Doerfler, W., Vogel, W., and Steinbach, P. (1997) *Amer. J. Hum. Genet.*, **60**, 1354-1362.
24. Genc, B., Müller-Hartmann, H., Zeschnigk, M., Deissler, H., Schmitz, B., Majewski, F., von Gontard, A., and Doerfler, W. (2000) *Nucleic Acids Res.*, **28**, 2141-2152.
25. Remus, R., Zeschnigk, M., Zuther, I., Kanzaki, A., Wada, H., Yawata, A., Muiznieks, I., Schmitz, B., Schell, G., Yawata, Y., and Doerfler, W. (2001) *Gene Funct. Dis.*, **2**, 171-184.
26. Remus, R., Kanzaki, A., Yawata, A., Nakanishi, H., Wada, H., Zeschnigk, M., Zuther, I., Schmitz, B., Doerfler, W., and Yawata, Y. (2005) *Int. J. Hematol.*, in press.

27. McClelland, M., and Nelson, M. (1988) *Gene*, **74**, 291-304.
28. Günthert, U., Schweiger, M., Stupp, M., and Doerfler, W. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 3923-3927.
29. Willis, D. B., and Granoff, A. (1980) *Virology*, **107**, 250-257.
30. Schetter, C., Grünemann, B., Hölker, I., and Doerfler, W. (1993) *J. Virol.*, **67**, 6973-6978.
31. Wienhues, U., and Doerfler, W. (1985) *J. Virol.*, **56**, 320-324.
32. Kämmer, C., and Doerfler, W. (1995) *FEBS Lett.*, **362**, 301-305.
33. Vardimon, L., Neumann, R., Kuhlmann, I., Sutter, D., and Doerfler, W. (1980) *Nucleic Acids Res.*, **8**, 2461-2473.
34. Ernberg, I., Falk, K., Minarovits, J., Busson, P., Tursz, T., Masucci, M. G., and Klein, G. (1989) *J. Gen. Virol.*, **70**, 2989-3002.
35. Desrosiers, R. C., Mulder, C., and Fleckenstein, B. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3839-3843.
36. Conklin, K. F., Coffin, J. M., Robinson, H. L., Groudine, M., and Eisenman, R. (1982) *Mol. Cell. Biol.*, **2**, 638-652.
37. Jähner, D., Stuhlman, H., Stewart, C. L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982) *Nature*, **298**, 623-628.
38. Deuring, R., Klotz, G., and Doerfler, W. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3142-3146.
39. Deuring, R., and Doerfler, W. (1983) *Gene*, **26**, 283-289.
40. Kochanek, S., Clemens, P. R., Mitani, K., Chen, H. H., Chan, S., and Caskey, C. T. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 5731-5736.
41. Karlin, S., Doerfler, W., and Cardon, L. R. (1994) *J. Virol.*, **68**, 2889-2897.
42. Southern, E. M. (1975) *J. Mol. Biol.*, **98**, 503-517.
43. Doerfler, W., Gahlmann, R., Stabel, S., Deuring, R., Lichtenberg, U., Schulz, M., Eick, D., and Leisten, R. (1983) *Curr. Top. Microbiol. Immunol.*, **109**, 193-228.
44. Waalwijk, C., and Flavell, R. A. (1978) *Nucleic Acids Res.*, **5**, 3231-3233.
45. Meyer, P. (ed.) (1995) *Curr. Top. Microbiol. Immunol.*, **197**.
46. Badal, S., Badal, V., Calleja-Maicas, I. E., Kalantari, M., Chuang, L. S., Li, B. F., and Bernard, H. U. (2004) *Virology*, **324**, 483-492.
47. Kuhlmann, I., and Doerfler, W. (1982) *Virology*, **118**, 169-180.
48. Kuhlmann, I., Achten, S., Rudolph, R., and Doerfler, W. (1982) *EMBO J.*, **1**, 169-180.
49. Orend, G., Kuhlmann, I., and Doerfler, W. (1991) *J. Virol.*, **65**, 4301-4308.
50. Orend, G., Knoblauch, M., Kämmer, C., Tjia, S. T., Schmitz, B., Linkwitz, A., Meyer zu Altenschildesche, G., Maas, J., and Doerfler, W. (1995) *J. Virol.*, **69**, 1226-1242.
51. Hohlweg, U., Hösel, M., Dorn, A., Webb, D., Hilger-Eversheim, K., Remus, R., Schmitz, B., Mende, Y., Buettner, R., Schramme, A., Corzilius, L., Niemann, A., and Doerfler, W. (2003) *Virus Res.*, **98**, 45-56.
52. Müller, U., and Doerfler, W. (1987) *J. Virol.*, **61**, 3710-3720.
53. Doerfler, W. (1991) *Biol. Chem. Hoppe-Seyler*, **372**, 557-564.
54. Yoder, A., Walsh, C. P., and Bestor, T. H. (1997) *Trends Genet.*, **13**, 335-340.
55. Bestor, T. H. (1998) *Novartis Found. Symp.*, **241**, 187-199.
56. Strohl, W. A., Rabson, A. S., and Rouse, H. (1967) *Science*, **156**, 1631-1633.
57. Stabel, S., Doerfler, W., and Friis, R. R. (1980) *J. Virol.*, **36**, 22-40.
58. Knoblauch, M., Schröer, J., Schmitz, B., and Doerfler, W. (1996) *J. Virol.*, **70**, 3788-3796.
59. Schröer, J., Hölker, I., and Doerfler, W. (1997) *J. Virol.*, **71**, 7923-7932.
60. Groneberg, J., Sutter, D., Soboll, H., and Doerfler, W. (1978) *J. Gen. Virol.*, **40**, 635-645.
61. Orend, G., Knoblauch, M., and Doerfler, W. (1995) *Virus Res.*, **38**, 261-267.
62. Heller, H., Kämmer, C., Wilgenbus, P., and Doerfler, W. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5515-5519.
63. Hohlweg, U., Hösel, M., Dorn, A., Webb, D., Remus, R., Buettner, R., and Doerfler, W. (2004) *Curr. Top. Microbiol. Immunol.*, **273**, 215-244.
64. Hilger-Eversheim, K., and Doerfler, W. (1997) *Cancer Res.*, **57**, 3001-3009.
65. Pfeffer, A., Schubbert, R., Orend, G., Hilger-Eversheim, K., and Doerfler, W. (1999) *Virus Res.*, **59**, 113-127.
66. Zantema, A., and van der Eb, A. J. (1995) *Curr. Top. Microbiol. Immunol.*, **199/III**, 1-23.
67. Toth, M., Lichtenberg, U., and Doerfler, W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 3728-3732.
68. Toth, M., Müller, U., and Doerfler, W. (1990) *J. Mol. Biol.*, **214**, 673-683.
69. Woodcock, D. M., Crowther, P. J., Jefferson, S., and Diver, W. P. (1988) *Gene*, **74**, 151-152.
70. Sprengel, J., Schmitz, B., Heuss-Neitzel, D., Zock, C., and Doerfler, W. (1994) *J. Virol.*, **68**, 379-389.
71. Muiznieks, I., and Doerfler, W. (1994) *FEBS Lett.*, **344**, 251-254.
72. Munnes, M., and Doerfler, W. (1997) in *Encyclopedia of Human Biology* (Dulbecco, R., ed.) 3rd Edn., Academic Press, San Diego-New York-Boston-London-Sydney-Tokyo-Toronto, pp. 435-446.
73. Orend, G., Linkwitz, A., and Doerfler, W. (1994) *J. Virol.*, **68**, 187-194.
74. Ortin, J., Scheidtmann, K. H., Greenberg, R., Westphal, M., and Doerfler, W. (1976) *J. Virol.*, **20**, 355-372.
75. Sutter, D., and Doerfler, W. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, **44**, 565-568.
76. Vardimon, L., Kuhlmlann, I., Doerfler, W., and Cedar, H. (1981) *Eur. J. Cell Biol.*, **25**, 13-15.
77. Cook, J. L., and Lewis, A. M., Jr. (1979) *Cancer Res.*, **39**, 1455-1461.
78. Johansson, K., Persson, H., Lewis, A. M., Pettersson, U., Tibbetts, C., and Philipson, L. (1978) *J. Virol.*, **27**, 628-639.
79. Doerfler, W. (1969) *Virology*, **38**, 587-606.
80. Doerfler, W. (1991) *Adv. Virus Res.*, **39**, 89-128.
81. Webb, D., Dorn, A., Hösel, M., Schmitz, B., Auerochs, S., and Doerfler, W. (2005), under revision.
82. Munnes, M., Schetter, C., Hölker, I., and Doerfler, W. (1995) *J. Virol.*, **69**, 2240-2247.
83. Thompson, J. P., Granoff, A., and Willis, D. B. (1988) *J. Virol.*, **62**, 4680-4685.
84. Vardimon, L., Kressmann, A., Cedar, H., Maechler, M., and Doerfler, W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1073-1077.
85. Vardimon, L., Günthert, U., and Doerfler, W. (1982) *Mol. Cell. Biol.*, **2**, 1574-1580.

86. Langner, K. D., Vardimon, L., Renz, D., and Doerfler, W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2950-2954.
87. Kruczek, I., and Doerfler, W. (1982) *EMBO J.*, **1**, 409-414.
88. Kruczek, I., and Doerfler, W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7586-7590.
89. Knebel, D., and Doerfler, W. (1986) *J. Mol. Biol.*, **189**, 371-375.
90. Knebel, D., Lübbert, H., and Doerfler, W. (1985) *EMBO J.*, **4**, 1301-1306.
91. Kochanek, S., Renz, D., and Doerfler, W. (1993) *EMBO J.*, **12**, 1141-1151.
92. Kochanek, S., Renz, D., and Doerfler, W. (1995) *FEBS Lett.*, **360**, 115-120.
93. Muiznieks, I., and Doerfler, W. (1994) *Nucleic Acids Res.*, **22**, 2568-2575.
94. Knust, B., Brüggemann, U., and Doerfler, W. (1989) *J. Virol.*, **63**, 3519-3524.
95. Langner, K. D., Weyer, U., and Doerfler, W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1598-1602.
96. Flint, J., and Shenk, T. (1989) *Annu. Rev. Genet.*, **23**, 141-161.
97. Nevins, J. R. (1995) *Curr. Top. Microbiol. Immunol.*, **199/III**, 25-32.
98. Weisshaar, B., Langner, K. D., Jüttermann, R., Müller, U., Zock, U., Klimkait, T., and Doerfler, W. (1988) *J. Mol. Biol.*, **202**, 255-270.
99. Knebel-Mörsdorf, D., Achten, S., Langner, K. D., Rüger, R., Fleckenstein, B., and Doerfler, W. (1988) *Virology*, **166**, 166-174.
100. Dobrzanski, P., Hoeveler, A., and Doerfler, W. (1988) *J. Virol.*, **62**, 3941-3946.
101. Hermann, R., Hoeveler, A., and Doerfler, W. (1989) *J. Mol. Biol.*, **210**, 411-415.
102. Hermann, R., and Doerfler, W. (1991) *FEBS Lett.*, **281**, 191-195.
103. Zeschnigk, M., Lich, C., Buiting, K., Doerfler, W., and Horsthemke, B. (1997) *Eur. J. Human Genet.*, **5**, 94-98.
104. Sutherland, G. R., Gecz, J., and Mulley, J. C. (2002) in *Principles and Practice of Medical Genetics* (Rimoin, D. L., Connor, J. M., Pyeritz, R. E., and Korf, B. R., eds.) 4th Edn., Churchill Livingstone, London-Edinburgh-New York-Philadelphia-St. Louis-Sydney-Toronto, pp. 2801-2826.
105. Passarge, E. (2002) in *Principles and Practice of Medical Genetics* (Rimoin, D. L., Connor, J. M., Pyeritz, R. E., and Korf, B. R., eds.) 4th Edn., Churchill Livingstone, London-Edinburgh-New York-Philadelphia-St. Louis-Sydney-Toronto, pp. 1747-1759.
106. Yawata, Y. (2003) *Cell Membrane*, Wiley-VCH, Weinheim.
107. Behn-Krappa, A., and Doerfler, W. (1993) *Human Molec. Genet.*, **2**, 993-999.
108. Kochanek, S., Hosokawa, K., Schiedner, G., Renz, D., and Doerfler, W. (1996) *FEBS Lett.*, **388**, 192-194.
109. Achten, A., Behn-Krappa, A., Jücker, M., Sprengel, J., Hölker, I., Schmitz, B., Tesch, H., Diehl, V., and Doerfler, W. (1991) *Cancer Res.*, **51**, 3702-3709.
110. Deissler, H., Behn-Krappa, A., and Doerfler, W. (1996) *J. Biol. Chem.*, **271**, 4327-4334.
111. Deissler, H., Wilm, M., Genc, B., Schmitz, B., Ternes, T., Naumann, F., Mann, M., and Doerfler, W. (1997) *J. Biol. Chem.*, **272**, 16761-16768.
112. Müller-Hartmann, H., Deissler, H., Naumann, F., Schmitz, B., and Doerfler, W. (2000) *J. Biol. Chem.*, **275**, 6447-6452.
113. Naumann, F., Remus, R., Schmitz, B., and Doerfler, W. (2004) *Genomics*, **83**, 108-120.
114. Doerfler, W. (2003) in *The Epigenome* (Beck, S., and Ole, A., eds.) Wiley-VCH, Weinheim-New York-Chichester-Brisbane-Singapore-Toronto, pp. 23-38.
115. Lueders, K. K., and Kuff, E. L. (1981) *Nucleic Acids Res.*, **9**, 5917-5930.
116. Ono, M., and Ohishi, H. (1983) *Nucleic Acids Res.*, **11**, 7169-7179.
117. Kuff, E. L., Fewell, J. E., Lueders, K. K., DiPaolo, J. A., Amsbaugh, S. C., and Popescu, N. C. (1986) *Chromosoma*, **93**, 213-219.
118. Meyer zu Altenschildesche, G., Heller, H., Wilgenbus, P., Tjia, S. T., and Doerfler, W. (1996) *Chromosoma*, **104**, 341-344.
119. Remus, R., Kämmer, C., Heller, H., Schmitz, B., Schell, G., and Doerfler, W. (1999) *J. Virol.*, **73**, 1010-1022.
120. Schiedner, G., Schmitz, B., and Doerfler, W. (1994) *J. Virol.*, **68**, 5476-5482.
121. Hösel, M., Webb, D., Schröer, J., Schmitz, B., and Doerfler, W. (2001) *J. Virol.*, **75**, 16041-16053.
122. Ushijima, T., Morimura, K., Hosoya, Y., Okonogi, H., Tatematsu, M., Sugimura, T., and Nagao, M. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 2103-2105.
123. Müller, K., Heller, H., and Doerfler, W. (2001) *J. Biol. Chem.*, **276**, 14271-14278.
124. Doerfler, W. (1968) *Proc. Natl. Acad. Sci. USA*, **60**, 636-643.
125. Lettmann, C., Schmitz, B., and Doerfler, W. (1991) *Nucleic Acids Res.*, **19**, 7131-7137.
126. Sapienza, C., Paquette, J., Tran, T. H., and Peterson, A. (1989) *Development*, **107**, 165-168.
127. Reik, W., Howlett, S. K., and Surani, M. A. (1990) *Development Supp.*, 99-106.
128. Engler, P., Haasch, D., Pinkert, C. A., Doglio, L., Glymour, M., Brinster, R., and Storb, U. (1991) *Cell*, **65**, 939-947.
129. Schumacher, A., Koetsier, P. A., Hertz, J. M., and Doerfler, W. (2000) *J. Biol. Chem.*, **275**, 37915-37921.
130. Schumacher, A., Arnhold, S., Addicks, K., and Doerfler, W. (2003) *Mol. Cell. Neurosci.*, **23**, 669-680.
131. Schumacher, A., and Doerfler, W. (2004) *Nucleic Acids Res.*, **32**, 1566-1576.
132. Schubbert, R., Lettmann, C., and Doerfler, W. (1994) *Mol. Gen. Genet.*, **242**, 495-504.
133. Schubbert, R., Renz, D., Schmitz, B., and Doerfler, W. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 961-966.
134. Schubbert, R., Hohlweg, U., and Doerfler, W. (1998) *Mol. Gen. Genet.*, **259**, 569-576.
135. Hohlweg, U., and Doerfler, W. (2001) *Mol. Genet. Genomics*, **265**, 225-233.
136. Palka-Santini, M., Schwarz-Herzke, B., Brondke, H., Renz, D., Schmitz, B., Auerochs, S., and Doerfler, W. (2003) *Mol. Genet. Genomics*, **270**, 201-215.
137. Lichtenberg, U., Zock, C., and Doerfler, W. (1987) *J. Virol.*, **61**, 2719-2726.
138. Lichtenberg, U., Zock, C., and Doerfler, W. (1988) *Virus Res.*, **11**, 335-342.