Scientific
Research
Publishing

The Unexpected Existence of Coding and Non-Coding Fragments along the Eukaryotic Gene

Pietro Volpe[1,2]

[1]Department of Biology, University of Rome "Tor Vergata", Rome, Italy
[2]National Institute of Biostructures and Biosystems, Rome, Italy
Email: volpe@uniroma2.it

Abstract

The pathways leading to synthesis and post-synthetic modification of DNA employed methionine as donor of atoms: the carbon that came from its $-CH_3$ served for DNA replication and repair either in bacteria or humans; its entire $-CH_3$ served instead for building N^6-methyladenine and 5-methylcytosine on bacterial DNA and 5-methylcytosine alone on human DNA. In humans, although a slight extra-S asymmetric methylation appeared *de novo* yielding on parental DNA 5'-m^5CpC-3'/3'-GpG-5', 5'-m^5CpT-3'/3'-GpA-5' and 5'-m^5CpA-3'/3'-GpT-5' monomethylated dinucleotide pairs, a heavy symmetric methylation involved in S semiconservatively newly made DNA to guarantee genetic *maintenance* of $-CH_3$ in 5'-m^5CpG-3'/3'-Gpm^5C-5' dimethylated dinucleotide pairs. In this framework, an inverse correlation was found between bulk genomic DNA methylation occurring in S and bulk polyA-containing pre-mRNA transcription taking place in G_1 and G_2. Thus, *probes* of 1×10^6 Daltons (constructed using sheared by sonication newly made methylated DNA filaments) revealed a modular organization in genes: after the hypermethylated promoter, they exhibited an alternation of unmethylated coding and methylated uncoding sequences. This encouraged the search for a language that genes regulated by methylation should have in common. An initial deciphering of *restriction minimaps* with hypomethylatable exons *vs.* hypermethylatable promoters and introns was improved when the bisulfite technique allowed a direct sequencing of m^5C. In *lymphocytes*, where the transglutaminase gene is inactive, its promoter exhibited two fully methylated CpG-rich domains at 5' and one fully unmethylated CpG-rich domain at 3', including the site +1 and a 5'-UTR. At variance, in *HUVEC* cells, where the transglutaminase gene is active, in the first CpG-rich domain of promoter few doublets lost their $-CH_3$. Such an inverse correlation suggested new hypotheses especially in connection with *repair-modification*: UV radiation would cause de-methylation in given *loci* of a promoter by chance, whilst even a partial de-methylation in this promoter would be able to resume a previously silent pre-mRNA transcription.

Keywords

Coding *vs*. Non-Coding Pre-Messenger RNA Regions, Exons and Introns, Multigenic and Monogenic Transcriptional Units, Regulation of Gene Expression, Repair-Modification

1. Introduction

Initial contributions established that, while the *Sea urchin* DNA contains the sole 5-methylcytosine (m^5C) [1], the *E. coli* DNA contains N^6-methyladenine (m^6A) and m^5C [2] [3].

The finding of these two classes of post-synthetically modified nitrogen bases along the bacterial heteroduplex was accompanied by basic observations dealing with the *restriction-modification* enzyme system [4]-[6]: m^6A was shown to be due in part to the activity of the deoxyadenylate (dam) DNA methyltransferases (DNAmets) [7]-[9] and in part to the activity of type I, type II and type III methylating restriction-modification endonucleases involved in the digestion of the infecting m^6A-free phage DNA [10]-[13]; m^5C was found to be due in part to the activity of the deoxycytidylate (dcm) DNAmets [9] [14] [15] and in part to the methylase activity of the sole type II restriction-modification system involved in the digestion of the infecting m^5C-free phage DNA [11].

The presence of m^5C along the animal double helix was known earlier [16] [17], and a small amount of m^6A was soon found in DNA of protozoa [18] [19], fungi [20], algae [21] and higher vegetables [22]. The analysis of base composition in DNA of *Drosophila* was contradictory: one technique suggested an absolute absence of m^5C [23]; another showed that, in some of its sequences, C was still modified [24] [25]. Later a modest presence of m^5C in the genome of insects was correlated with lethality [26]-[28]. In any case, it was established that, if in animal genomes, 1 to 5 out of every 100 Cs are modified [29] [30], in plant genomes, this amount could increase up to 10 or even 30 per cent [31] [32].

Our laboratory at the International Institute of Genetics and Biophysics (IIGB) of the National Research Council (CNR), in Naples, got interested in the topic because of studies on structuristics of rare nucleotides found in ox brain [33] [34], *plus* and *minus* strand poliovirus RNAs synthesized at hourly intervals through the *HeLa* mitotic cycle [35] [36] and the nature of a high-molecular weight RNA extracted with hot phenol from nerve cells [37] [38] pertaining to the origin of mRNA. On the other hand, at variance with the polygenic structure of the *operon* regulating protein biosynthesis in bacteria [39], while tentatively an operon organization was thought to act in higher organisms too [40], the results collected by us at the Albert Einstein College of Medicine, in New York, on the amphi-directional regulation of a reversible reaction common to two inverted enzyme sequences [41] [42] accounted for a monogenic function of the transcriptional unit acting in mammals.

These points—apparently distant from each other—converged into an intriguing macromolecular manipulation, shedding light on the internal structure of the eukaryotic gene [46]. At the beginning of the 70s, in fact, while attention was paid to methylation of specific DNA sequences [47], a problematic discussion focused not only on the origin and size of the eukaryotic pre-mRNAs and mRNAs [37] [73] [74] but consequently on the design of the transcriptional unit in higher cells [75] [76]. The background experiment, performed to elucidate both these aspects, originated from previous knowledge regarding the timing of DNA methylation during the cellular cycle [47] [67] and from the non-random genetic scattering of m^5C along a semiconservatively replicating DNA chain [46]. This experiment showed a preferential methylation of gene promoter [46] (as reviewed in [88]) and of all regulatory and signal sequences that do not code for mRNAs [46] (as reviewed in [89]). In brief, in *HeLa* cells, if half of the population of hybrids between genomic DNA fragments of about 1×10^6 Daltons (used as "probes") and pre-processed high molecular weight mRNAs (purified from nuclei) contained a large number of m^5Cs, the whole population of hybrids between genomic DNA fragments of about 1×10^6 Daltons (also used as "probes") and processed low molecular weight mRNAs (purified from polysomes) contained few, if any, m^5Cs [46]. Since polysomal mRNAs were known to be much shorter than nuclear pre-mRNAs [75] [76] [111], these results suggested for the first time that, at variance with the structure of bacterial genes (in an operon the cistrons were known to be constituted of coding sequences [39]), the eukaryotic gene had to be thought of as a repetition, after the promoter, of intermittent "coding" and "uncoding" regions [46] [93]. For the sake of clarity, the DNA tracts containing methylated bases essentially were left out by mRNA [46], namely methyla-

tion did not significantly involve the coding regions: methylation preferentially involved the uncoding ones, complementary to parts of pre-mRNAs to be removed during processing [46]. Three years later, Mandel and Chambon supported this finding [116]. Their electron microscopic pictures showed how a cDNA intermittently excludes segments of the corresponding pre-mRNA from hybridization: the sequences of pre-mRNA, non-hybridized with cDNA [116], were nothing but the uncoding sequences [46] (Crick [87] and Gilbert [86] proposed to name them "introns"); the sequences of pre-mRNA, hybridized with cDNA [116], were nothing but the coding sequences [46] (Crick [87] and Gilbert [86] suggested to call them "exons").

The objective of this essay is to celebrate the eukaryotic gene fragmentation discovery and to emphasize how complex was the early 70s basic biochemical research promoting the modern Epigenetics and Genetic Engineering (as recognized in [150]-[160]). Our next investigation will concentrate on the physiological evolution of coding *vs.* non-coding pre-messenger RNA regions. The exon and intron methylation differences will be evaluated as a function of the vertebrate evolution, while the perturbed regulation of transcription will be investigated in mammals in relation to the radiation-induced neoplastic consequences of the "repair-modification" activity.

2. The Pathway Using Methionine to Synthesize DNA Remained Unchanged in Phylogeny, While the Pathway Using Methionine to Modify DNA Evolved from Bacteria to Humans

Let us begin with description of the biochemical reactions lying on the basis of DNA polymerization and modification. The study was initiated in *HeLa* cells [46] [47] and completed in *E. coli* [48] [49]. However, coherently with evolution, it seemed appropriate to first treat the results regarding prokaryotes and, then, those regarding eukaryotes.

2.1. Synthesis and Methylation of DNA in Prokaryotes

One of the main purposes of this part of analysis was to verify if *E. coli* would be able to employ Met as a sole common donor of atoms for DNA synthesis and methylation, as *HeLa* cells do [47] [50]; another purpose was to verify whether, on bacterial DNA, the timing for optimal yield of m^5C would coincide with that for optimal yield of m^6A [48] [49].

It emerged that, when percolated through reverse phase HPLC [51], control nitrogen bases, mixed in equimolar proportions, were eluted in the following time-order: C, m^5C, G, T, A and m^6A (**Figure 1(a)**). But there was a striking quantitative difference between the eluted unmethylated *vs.* methylated bases, when these were constituents of an *E. coli* DNA hydrolyzate: C, G, T, and A resulted in equimolar proportions, as expected; m^5C and m^6A were present in very small amounts (**Figure 1(b)**). **Figure 1(c)** illustrated the profile of an *E. coli* DNA hydrolyzate labelled with [^{14}C]methyl-L-methionine ([^{14}C]Met) in mid-culture growth cycle (CGC): C was the only base not incorporating the labelled carbon coming from the precursor; by contrast, m^5C was labelled to a notable extent comparable to that of m^6A. The radioactivity found in the two modified bases meant that post-synthetic DNA modification took place, as for *HeLa* cells [47] [50], by virtue of transfer, *via* S-adenosyl-L-methionine (AdoMet), of the whole radioactive –CH_3 from [^{14}C]Met to the DNA Cs and As. In addition, in **Figure 1(c)**, the radioactivity found in the A, G and T bases revealed that, once again, as for humans [47] [50], DNA synthesis emerged because there was, *via* C_1-chain oxidation, an insertion of the labelled carbon coming from the –CH_3 of [^{14}C]Met into the purine heterocycle of A and G and into the –CH_3 of T.

The quantitative analysis of the experiments performed through the CGC of *E. coli* (in the same conditions corresponding to those of **Figure 1(c)**), indicated that, while the curve of bacterial proliferation was S-shaped [48], the synthetic and methylase pathways of DNA developed in a differential fashion (**Figure 2(a)**).

i. The specific labelling accounting for synthesis of A, G and T appeared to be much lower than that accounting for construction of m^5C and m^6A (this meant that the methylase pathway evolved faster than the synthetic pathway).

ii. The specific labelling of T was always lower than that of G and A since, one thought, the exit of the C_1-chain, oriented towards synthesis of G and A, was highly facilitated in comparison with that oriented towards synthesis of T.

iii. The specific labelling of m^5C was much lower than that of m^6A, probably because the dam enzyme system worked faster than the dcm enzyme system.
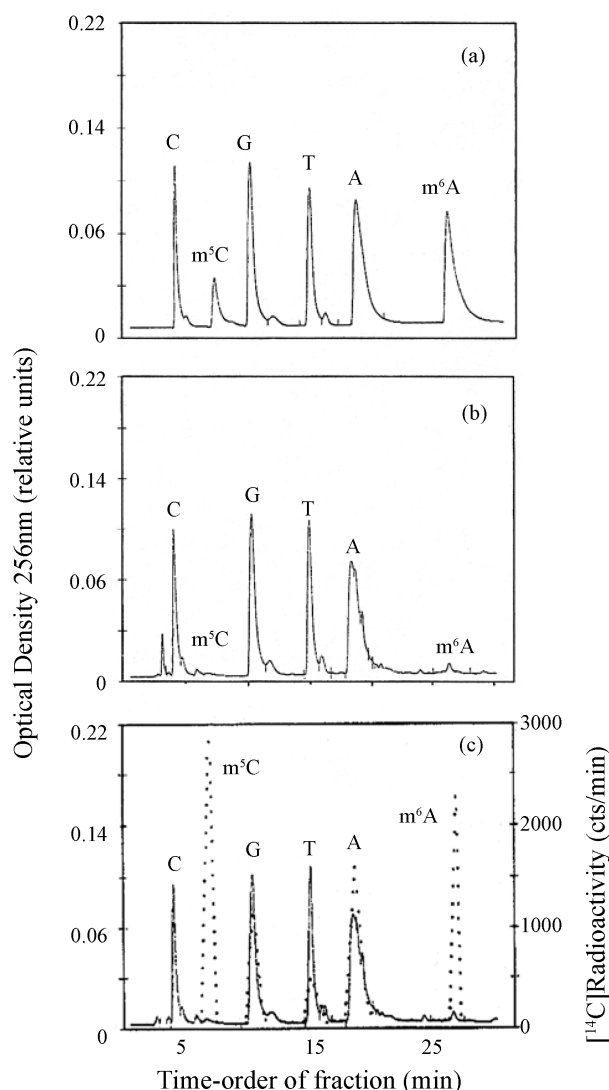
Figure 1. Chromatographic analysis of a bacterial DNA hydrolyzate. Cultures of *E. coli MRE* 600 were grown on Petri dishes by inoculating 0.5 - 1 ml of minimal medium (containing 1×10^{11} cells/ml) into 100 ml of Luria-Bertani medium. At intervals of time during their CGC (roughly lasting 10 hrs), DNA was labelled for 40 min at 37˚C with 2 microCi/ml [14C]Met [48]. The 14C-labelled DNA was extracted with minor modification of the method described by Marmur [145]. A 30-micrograms aliquot of purified DNA was introduced into a glass vial containing 300 microliters of 88% formic acid (75% final concentration). The vial was closed with flame and left at 160˚C for 1 hr [50]. The hydrolyzed residues were dried, suspended in 30 microliters of 20 mM trichloroacetic acid at pH 2.2 and run at 13,000 rev./min for 5 - 10 min in an Eppendorf centrifuge. A 20-microliters aliquot was percolated through a 131051 L ODS-55 Biorad Bio-Sil HPLC column (250×4 mm) in reverse phase condition [51]. Fractions of 0.4 ml were collected every 30 sec, while their content was resuspended in 50 microliters of water and tested for radioactivity in a Packard Tri-Carb TR beta-counter [48] [49]. (a) Free bases from a standard solution; (b) Bases present in 30 micrograms of an unlabelled DNA hydrolyzate extracted from a 2-ml suspension of $1 \times 10^8 - 1 \times 10^{10}$ bacteria maintained in minimal medium; (c) Bases present in 30 micrograms of a labelled DNA hydrolyzate obtained from a 2-ml suspension of $1 \times 10^8 - 1 \times 10^{10}$ bacteria maintained in Luria-Bertani medium. The left ordinate showed the concentration of bases detected in arbitrary units by a computerized instrument (continuous line); the right ordinate showed their radioactivity (dotted line). The analysis was repeated a dozen of times.

iv. In harmony with the development of the specific labelling of all nitrogen bases taken together and of the specific labelling of single A, G and T, the specific labelling of m⁶A gradually decreased during the CGC, when that of m⁵C rapidly decreased after the inoculum to rise again, with a sharp peak, in mid-CGC (an interpretation of these patterns was that the dam and dcm DNAmet activities are uncoupled) [48] [49].
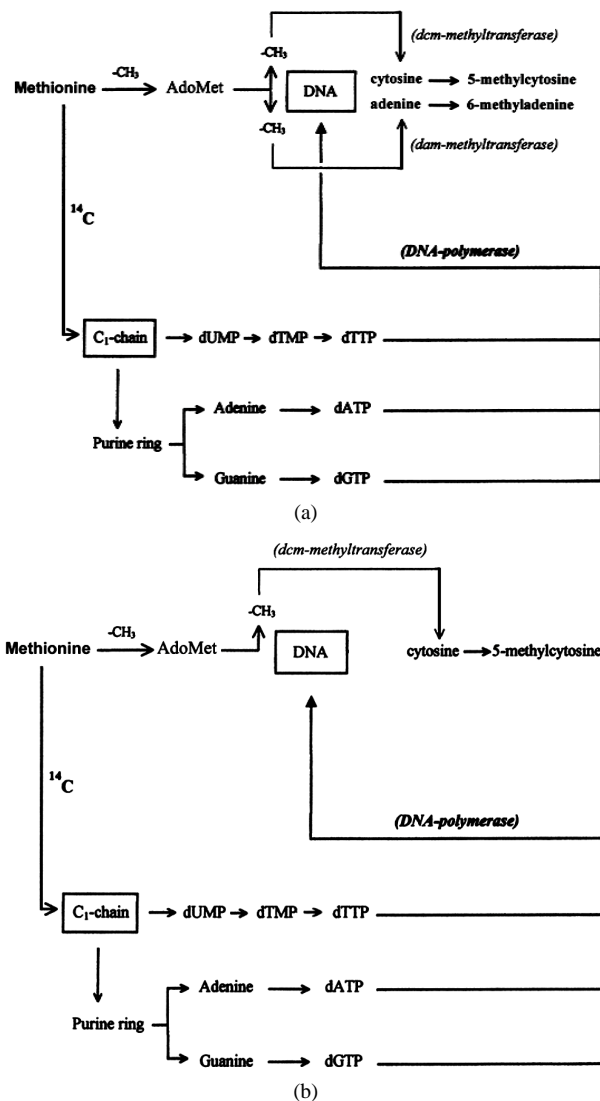
(a)

(b)

Figure 2. Pathways leading to synthesis and post-synthetic modification of DNA. In *E. coli* [48] [49], as in *HeLa* cells [47] [50], to accomplish replication, the DNApol system exploited the dATP, dGTP and dTTP resulting from the C_1-chain oxidation which used, *via* formic acid, the carbon coming from the $-CH_3$ of Met to build the purine heterocycle of A and G and the $-CH_3$ of T (this carbon did not enter the pyrimidine ring of C). (a) In bacteria, two parallel pathways led *via* AdoMet to post-synthetic DNA modification [48] [49]: on the one hand, the dam DNAmet system (including the methylation exerted by the type I, II and III RM endonucleases) constructed m^6A on new DNA circles; on the other, the dcm DNAmet system (including the methylation exerted by type II restriction-modification endonucleases) constructed m^5C on the same new DNA circles. (b) In Humans, Met also served as a common donor of atoms [46] [47] [138]: in the cytosol, the carbon coming from its $-CH_3$ through the C_1-chain entered the purine ring of A and G and the $-CH_3$ of T; after formation of the corresponding dATP, dGTP, and dTTP and, after release of a PP_i from each, in the nucleus the DNApol *alfa* [112] introduced, during S, the dAMP, dGMP, and dTMP into semiconservatively newly synthesized DNA chains, while Met catapulted *via* AdoMet its entire $-CH_3$ on given C residues located along these chains.

2.2. Synthesis and Methylation of DNA from the World of Bacteria to the World of Humans

The data reported above demonstrated that the two metabolic pathways, leading to synthesis and methylation of DNA, originated in the prokaryotic world by exploiting Met as donor of carbon atoms and $-CH_3$ groups (**Figure 2(a)**). This crucial property was preserved fairly well by the mammalian world (**Figure 3(a)** and **Figure 3(b)**), even if at the level of the DNA modifying enzyme bifurcation one of the two branches—the one yielding m^6A on DNA (**Figure 2(a)**)—was erased in *HeLa* cell DNA (**Figure 2(b)**).

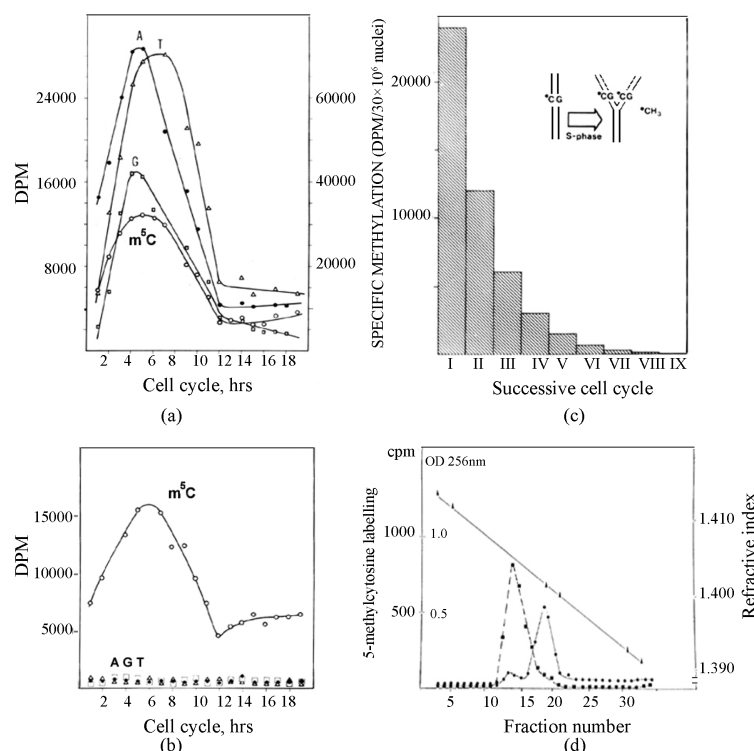**Figure 3.** Semiconservative transmission of m⁵C in human DNA. (a) Through an entire *HeLa* cell cycle [67], the genomic DNA (labelled with [¹⁴C]Met) was hydrolyzed to bases and chromatographed: radioactivity in A, G and T (right ordinate) showed synthesis, while radioactivity in m⁵C (left ordinate) showed methylation [46] [47]. (b) During an entire mitotic cycle, nuclei isolated from *HeLa* cells were labelled with [³H-methyl]AdoMet and their DNA was hydrolyzed to bases and chromatographed: radioactivity in A, G, and T, accounting for synthesis, was negligible, while radioactivity in m⁵C showed methylation [47]; (c) During S, *HeLa* cells were labelled with [¹⁴C]Met; then, after 14 hrs of growth in fresh medium (in the absence of radioisotope), their genomic DNA was hydrolyzed to bases and chromatographed: the measurement of radioactivity in m⁵C was repeated in the course of twelve CGCs. The insert suggested how, in the replication fork, methylation semiconservatively followed synthesis, since the labelled m⁵C per cell always decreased by half [109]. (d) Ultracentrifugation of genomic *HeLa* cell DNA in alkaline CsCl gradient [50]: the dashed line showed the m⁵C radioactivity (originating in [¹⁴C]Met) detected along the semiconservatively newly replicating chains, made heavier by a previous incorporation of BrdUrd; the solid line showed the OD at 256 nm of the separated lighter parental chains.

The abandonment of m⁶A by DNA of the majority of the highly evolved cells was a consequence of the substitution, in them, of the obsolete bacterial antiphage restriction-modification defense [6] with the innovative immune reaction. Actually, with the extinction of the system of dam and dcm DNAmets [9] and of the system of restriction-modification endonucleases [11]-[13], both characterizing the world of bacteria [52], molecular selection led in mammals to the loss of m⁶A not only in nuclear [46] but also in mitochondrial [53] [54] and chloroplastic [55] DNAs. In these different nucleic acids, the m⁵C base was conserved by a novel DNAmet family that likely evolved through an *alternate splicing* [56]-[58], as in mammals the fundamental defense against antigens was exerted by antibodies.

In such a structure, the question concerning the speed of reactions in the pathway leading to DNA synthesis *vs.* the speed of reactions in the pathway leading to post-synthetic DNA modification assumed great interest since, during the bacterial CGC, the rate of formation of T was always lower than that of A and G. Namely, the time required by the carbon, on the way from the –CH₃ of Met to the –CH₃ of T, was longer when compared to the time required for its insertion into the purine heterocycle of A and G [48] [49]. This differential rate was in harmony with the situation seen in *HeLa* cells: quantitatively, the biosynthetic rate of T was of the same order of that of A, even though a notable delay of its optimum with respect to the optimum of synthesis of A and G was observed (**Figure 3(a)**).

This kind of correspondence was expected, since either in bacteria (**Figure 1**) or mammals (**Figure 3(a)**) the process leading to synthesis of T had to join into a pyrimidine ring the carbamylphosphate and aspartate mole-

cules before accepting, at position 5 of the newly formed U, the $-CH_3$ constructed with the carbon coming from formic acid. This carbon entered directly the A and G purine heterocycle (**Figure 2(a)**).

Biosynthesis of A, G and T transpired in comparable loci (in liquid matrix of a bacterial cell or in cytosol of a mammalian cell), whereas post-synthetic modification of DNA occurred in different compartments: in *E. coli*, it yielded m^5C and m^6A in the same liquid matrix (**Figure 1(b)** and **Figure 1(c)**); in *HeLa* cells, it yielded the sole m^5C in the nucleus (**Figure 3(b)**). As follows, the longer time required by the nucleoside AdoMet to enter the nucleus would justify the advantage of the bacterial cell over the mammalian cell in exploiting the $-CH_3$ of Met: if in *E. coli* the specific building of A, G and T was much lower than that of m^5C and m^6A [48] [49], in *HeLa* cells the specific building of A, G and T turned out to be either similar or only a little lower than that of m^5C [47] [50].

Summarizing, the large percentage of DNA methylation, necessary for antiphage defense in bacteria, became unnecessary in mammals.

Was this a non-turning point in evolution?

To make a long reasoning short: the prokaryotic world used Met to promote both the pathways leading to synthesis and methylation of DNA (**Figure 1(b)** **Figure 1(c)** and **Figure 2(a)**); this capacity was inherited almost completely by the mammalian world, with the exception that, in *HeLa* cells, the reactions yielding m^6A on DNA were suppressed (**Figure 3(a)** and **Figure 3(b)**); the diversification of the speed of single reactions (in the pathways leading to synthesis and methylation of DNA) depended on the evolution of cellular organization, of course, since the eukaryotic nucleus concentrated in itself important genetic functions that were diffused earlier in the bacterial liquid matrix [47]-[49].

3. In Humans the Cell-Cycle Methodology Served to Reveal Two Basic Laws Regulating DNA Methylation

Notwithstanding the exhaustive investigation on the evolution of the pathways leading to DNA synthesis and methylation, there were doubts in elucidating the biocatalytic processes dealing with the complex phenomenon of DNA methylation in eukaryotes: the existence of DNA demethylases, deaminases and reaminases was debated [59]-[61]; the heterogeneity of the DNAmet system was revealed much later [56] [57] [62]-[64]. For this reason, in the early 70s, we preferred to investigate the timing of DNA methylation and consequently the precise macromolecular acceptors of the $-CH_3$ groups, namely the old or the new DNA chains: the experiments were made feasible by a method set up to determine the lengths of an entire mitotic cycle and its single G_1, S, G_2 and M phases [65]-[67].

3.1. DNA Methylation Followed DNA Synthesis

It was suggested that DNA methylation would continue for some hours after DNA synthesis is completed [68] [69]. In this research, by incubating synchronized *HeLa* cells with [^{14}C]Met used as common tracer for both DNA synthesis and methylation [47], it was possible to verify that DNA methylation accompanied DNA synthesis: the labelled carbon atom, coming from the $-CH_3$ of Met, did not enter the pyrimidine ring but pierced *via* C_1-chain the purine ring of A and G and the $-CH_3$ of T, while the entire $-CH_3$ of [^{14}C]Met was transferred *via* AdoMet to DNA C (**Figure 2(b)**). Truly, the maximal labelling emerging from the ^{14}C constituting the $-CH_3$ of [^{14}C]Met was always found in S in four hydrolyzed bases: in A, G and T, implying synthesis, and in m^5C, signifying methylation (**Figure 3(a)**).

3.2. Newly Replicating DNA Chains Were Semiconservatively Methylated

It was known that in isolated nuclei, in the absence of triphosphonucleosides (supplied by cytosol), DNA could not be synthesized. On this basis, by employing nuclei labelled with [^{14}C-methyl]AdoMet, one could verify whether or not the absence of DNA replication would influence DNA methylation. The result was unexpected: in nuclei isolated from *HeLa* cells during S, among other DNA bases, only C continued to be methylated in the absence of DNA synthesis (**Figure 3(b)**). This showed that the two pathways of DNA replication and DNA methylation were separated from one another (**Figure 2(b)**): in whole cells, DNA methylation followed DNA replication during S (**Figure 3(a)**); in isolated nuclei, DNA methylation proceeded during S in the absence of DNA synthesis (**Figure 3(b)**). By itself the occurrence of DNA methylation in S-phase nuclei was not an absolute demonstration that it involved new chains formed just before their isolation; but the correlations between *in vivo*

DNA synthesis and *in vivo* and *in vitro* DNA methylation strongly suggested this possibility [47]:

$$\left[\Delta DNA - CH_3 / \Delta t\right]_{in\ vivo} = K_1 \left[\Delta DNA\right]_{in\ vivo}$$

$$\left[\Delta DNA - CH_3 / \Delta t\right]_{in\ vitro} = K_2 \left[\Delta DNA\right]_{in\ vivo}$$

$$\left[\Delta DNA - CH_3 / \Delta t\right]_{in\ vitro} = K_3 \left[\Delta DNA - CH_3 / \Delta t\right]_{in\ vivo}$$

These equations established that the cells entering S would carry old chains inherited by parental cells, while targets of DNA met both *in vivo* and *in vitro* should be nothing but nascent chains: several of them still remained in the nuclei to be isolated [47]. A similar conclusion was reached thinking that, if old and new DNAs were methylated in isolated nuclei [70], new DNAs would be formed without m^5C in whole cells [71].

Whatever the case would be, the semiconservativity of DNA methylation (**Figure 3(c)**), revealed by treating synchronized cells with the [^{14}C]Met tracer [47], was confirmed not only by the analysis that used restriction endonucleases [72] but also by the direct separation in alkaline CsCl of heavy bromodeoxyuridine (BrdUrd)-containing new chains and light non-containing BrdUrd old chains (**Figure 3(d)**).

4. Hybridizations of Pre-Processed Messenger RNA or Processed Messenger RNA with DNA "Probes" (Constructed during the S Phase by Shearing Newly Methylated DNA Filaments) Revealed an Alternation of Coding and Non-Coding Regions in the Human Gene

The hypotheses regarding the organization of the eukaryotic transcriptional unit had some analogy with those concerning the role of DNA methylation in higher organisms. From this, one deduced that the easiest way to directly characterize the eukaryotic gene design would be to exploit pure biochemical tools.

4.1. From Finding of Pre-mRNA and mRNA to Construction of DNA Probes

It is worth re-visiting in some more details the basic dispute into which we get involved. Georgyev and Mantieva found the polyribonucleotide called *giant nuclear* RNA in mouse [73], as also evidenced by us in nerve cells [37] [38]. Penman, Vesco and Penman claimed that their polydisperse RNA had nothing to do with the giant nuclear RNA [74]. In return, Georgyev assumed that the polydisperse RNA would likely correspond to a product of maturation of the giant nuclear RNA [75].

Accepting this assumption, Darnell, Jelinek and Molloy considered the giant nuclear RNA as *pre-processed* mRNA (measuring the size of a gene) and the polydisperse RNA as *processed* mRNA (much shorter than a gene): an agreement stating that a transcriptase would copy the whole eukaryotic gene to form a large pre-processed pre-mRNA, namely the *heterogeneous nuclear* RNA [76].

This conclusion helped indeed to elucidate the structure of the monogenic transcriptional unit in mammals, since we judged such a monogenic unit as an extraordinary innovation exerted by evolution with respect to the complex bacterial polygenic transcriptional unit known as operon.

Therefore, the information reported above on semiconservative methylation of newly replicating DNA chains (**Figure 3(c)** and **Figure 3(d)**) served as a key for a crucial biochemical analysis [46]: segments of these chains, obtained through shearing, were employed as *probes*, labelled in m^5C, in order to look for particular sequences located inside the eukaryotic gene (**Figure 4(A)**).

4.2. Our So-called "Pioneering Experiment on Gene Fragmentation" Demonstrated the Existence in the Human Transcriptional Unit of Coding *vs.* Non-Coding Regions

Filaments of semiconservatively newly made methylated DNA, previously separated during the S-phase in alkaline CsCl gradients (**Figure 3(d)**), were shortened through sonication into fragments of about 1×10^6 Daltons (**Figure 4(A)**, left top) and hybridized with pre-mRNA (**Figure 4(Ab)**) extracted from *HeLa* cell nuclei [46] [66] [67] or with mRNA (**Figure 4(Ac)**) extracted from polysomes of these same cells [46] [77]: in the first case, the modified Cs were found in about 50 per cent of the hybrids "pre-mRNA/DNA probe" (signed with "plus" *vs.* "minus" in **Figure 4(A)**); in the second, the modified Cs were essentially absent in 100 percent of the hybrids "mRNA/DNA probe" (marked with "minus" in **Figure 4(A)**).
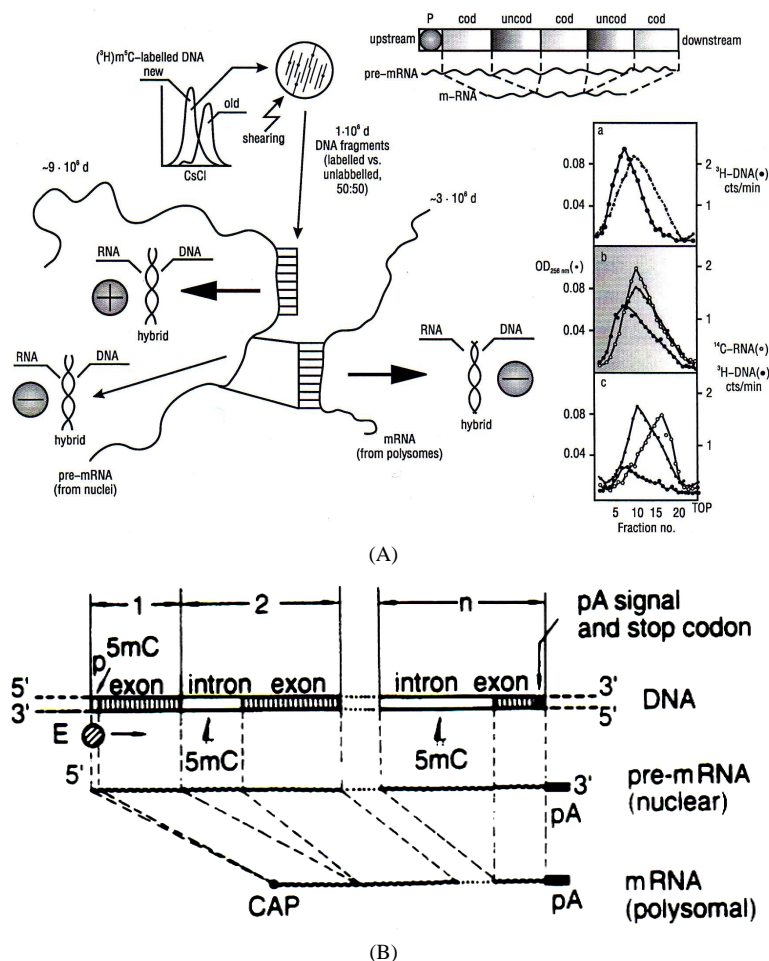
(A)

(B)

Figure 4. Gene design. (A) Employing *HeLa* cells, semiconservatively newly made DNA chains, separated through CsCl in mid S-phase (a), served to construct by shearing methylated probes of 1×10^6 Daltons; their association with larger polyA-pre-processed messenger RNA (from nuclei) yielded methylated (50%) and unmethylated (50%) DNA/RNA hybrids (b); the same probes, associated with shorter polyA-containing processed messenger RNA (from polysomes), essentially yielded 100% unmethylated DNA/RNA hybrids (c). Actually, after the promoter, an alternation of coding and uncoding regions appeared to take place, since a preferential methylation characterized the same promoter and the sequences that did not code for mRNAs [46] [138] [147]. (B) The monogenic transcriptional unit was thought to imply an adjacency of 1-, 2-, n-blocs, each containing an initial methylated part and a final unmethylated part [46] [93]. The promoter (p) would correspond to the methylated part of the first bloc, followed by the first unmethylated coding (exon) part. Each following bloc should contain a methylated non-coding (intron) part and an unmethylated coding (exon) part. The last bloc should carry, at the end of its coding part (exon), the polyA signal and the stop codon(s). The RNApol II would synthesize pre-mRNA along the whole gene, while the pre-mRNA tracts—complementary to the methylated non-coding DNA sequences (from 2- to n-blocs)–would remain out of processed mRNA. (CAP) 5'-Methylated mRNA end; (pA) 3'-post-transcriptional polyadenylated pre-mRNA end.

For clarity's sake, the result of the former experiment signified that the m^5Cs were located along the promoter and along the non-coding regions inside the gene (all uncoding sequences, including the promoter, being complementary to regions to be removed from the unprocessed pre-mRNA); by contrast, the result of the latter experiment suggested that the joined translatable sequences of processed mRNA were complementary to the unmethylated coding sequences scattered in the gene.

Such a design of human transcriptional unit-structured in alternating non-informative and informative sequences, starting from the promoter (**Figure 4(A)**, right top)—was illustrated at the 9*th Meeting of the Federation of the European Biochemical Societies* held in Budapest in 1974: the oral presentation was preceded by an invited paper that opened the ritual Special Issue of *FEBS-Letters* [46].

Three years later, in 1977, the novelty brought by this paper was confirmed in France by Breathnach, Mandel and Chambon [78]: in their electron microscopic observations, the ovalbumin mRNA repeatedly did not hybridize sequences of the ovalbumin gene, forming characteristic *loops* (these were nothing else but the uncoding regions found earlier by us); differently, the sequences of the ovalbumin gene, that did become hybridized to the ovalbumin mRNA, were nothing else but our coding regions.

Soon, a number of other laboratories, mostly in USA, observed that the eukaryotic gene looked *split*: the pre-mRNA coding sequences were interrupted by uncoding ones [79]-[82]; uncoding sequences were seen even in the *adenovirus* mRNA [83] [84].

Then, when an intervening sequence was also described in the mouse *beta*-globin gene [85], Gilbert [86] and Crick [87] proposed to call *exons* and *introns*, respectively, the coding and non-coding regions found in Naples [46].

Nevertheless, while two years earlier the proposal of these new appellations Maclean and Hilder [88], from United Kingdom, in the *International Review of Cytology*, attributed to us the priority of discovery of the "*preferential methylation of bases in the promoter proximal region of the transcriptional unit*", two years later the introduction of these new appellations, Tentravhi, Guntaka, Erlanger and Miller [89], from United States, in the *Proceedings of the National Academy of Sciences of USA*, attributed to us also the priority of discovery of the "*preferential methylation of the early replicating regulatory sequences that do not code for mRNAs*".

In sum, if in *HeLa* cells half of the population of hybrids between genomic DNA fragments (used as probes) and pre-processed mRNAs (purified from nuclei) contained a large number of m^{5}Cs (**Figure 4(Ab)**), the whole population of hybrids between genomic DNA fragments (used as probes) and processed mRNAs (purified from polysomes) contained few, if any, m^{5}Cs (**Figure 4(Ac)**). Since polysomal mRNAs were known to be much shorter than nuclear pre-mRNAs, these results suggested that, in comparison with the bacterial operon structure (containing *cistrons* essentially representing informative sequences), the eukaryotic gene had to be thought of as a repetition, after the promoter, of alternate *coding* and *uncoding* regions (**Figure 4(A)**, top right): methylation did not significantly involve the coding regions but involved the non-coding ones, complementary to parts of pre-mRNAs to be removed during processing [46].

5. Evolution Combined a Variable Number of *Coding-Uncoding Modules* in Eukaryotic Genes Depending on the Lengths of the Polypeptides to Be Expressed

Following the demonstration that the infection by a lytic virus, like *poliovirus*, was found to be rigorously cell-cycle dependent [35] [36], the observation of the design of the eukaryotic transcriptional unit [46] induced to investigate the site of recombination between the genome of an oncovirus, like *SV40*, and that of a host-cell, like *mouse fibroblasts* (the oncogenome integration needing at least an *uncoding-coding module* as specific target).

The working hypothesis was based on the assumption of two possible mechanisms of virus-induced cell transformation: one involving a complete loss of cell regulation and the other involving a partial loss of cell regulation. In the first instance, the virus DNA would be integrated between the uncoding and coding regions of the module while, in the second, the virus DNA would be integrated in the uncoding region only [90].

The genetic engineering experiments spoke in favor of the former mechanism, being facilitated during the S-phase [91] [92]. This mechanism—in addition to the information carried by the *SV*40 early gene—exploited the information carried by an adjacent short piece of host-cell DNA to yield the so-called *non-truncated*, *i.e.* entire (*fused*), T-antigen: the *truncated* T-antigen, encoded by the *SV40* early gene, did not contain the signals for polyadenylation and stop codons; these signals were brought by the joined host-cell DNA piece [93].

These new acquisitions led us to carefully consider the question, posed by Gilbert [86], on why the genes are broken down into pieces.

The origin and evolution of the uncoding-coding structures seemed problematic to him, not only because both the "exon theory of genes" and the "insertional theory of introns" debated the presence or absence of uncoding sequences in "primordial genes", but also because the statistical analysis assumed that an "exon shuffling" played an important role in the origin of either *ancient* or *modern* genes.

Despite the fact that we did agree with such a viewpoint, our attention was focused, rather, on the order of the two classes of regions in a module. Why, we questioned, an uncoding region should precede a coding one, and not *vice versa*?

Independently from the concept of *exon shuffling*, one would answer that transcription starts in any case from a promoter.

A *modular gene organization* should be in good harmony with this need: in the initial module, the promoter should play the role of an uncoding region, preceding the first exon; in the last module, after the intron, an exon should necessarily carry an end with signals for polyadenylation and stop codons (**Figure 4(B)**). Thus, inside the gene, the promoter and the various non-coding regions should exert a regulatory function, while the various coding regions should carry information for a given protein [46].

But were the $-CH_3$ groups (preferentially carried by the promoter and non-coding regions along the gene) correlable indeed with the genetic regulation of transcription?

A circumstantial attention to this correlation will be paid below.

Allow us emphasize the fact that the association of non-coding regions with methylation *vs*. the practical absence of methylation in coding regions (**Figure 4(B)**) represented ideal properties useful to evaluate the "age" of the regions themselves, *i.e.* if unmethylated coding sequences should be considered phylogenetically older while methylated non-coding sequences should be considered phylogenetically younger.

From the viewpoint of evolution, we assumed that the unmethylated coding regions probably became 'conservative' because the conventional excision-repair system always reconstructed them by maintaining any possibility of error at a minimum (e. g. the high conservation of genes coding for histones), as the continuous repetition of possible incomplete repair of the methylated uncoding sequences—through a long chain of cell cycles [90]—probably caused their wide variability.

On the basis of this variability selection would have operated by creating (through *ad hoc* transpositions) new transcriptional units containing non-coding regulatory and signal sequences [91] [93]-[96].

An insight of this kind might also explain why the amount of the uncoding portion in many eukaryotic genes was found to be so large when compared with that of their coding part, being the proportion dealing with the lengths of the proteins to be made, of course [53]. After all, the chain of biocatalytical reactions able to join in sequence various uncoding-coding modules to form the genes expressing these proteins had not yet been understood.

6. The Alternation of Coding and Non-Coding Regions in Eukaryotic Genes Was Convalidated by Their Restriction Minimaps Showing a Scattered Distribution of the Methylatable Sites

At the 7*th International Symposium on Macromolecules in the Functioning Cell*, held in Taormina in 1990, my talk entitled "*Is* 5-*methylcytosine a regulatory signal in eukaryotic gene expression*?" emphasized the concept of split eukaryotic gene [46] analyzing its methylatable sites recognized by restriction-modification endonucleases [97]. Then, the split design of the gene, observed in *HeLa* cells, that I extensively elucidated in research seminars given at the King College of London, in 1991, and at the Cancer Center of the University of Texas, Houston, in 1992, contributed to motivate at the "Incubator of Ideas" of Stony Brook, USA, in 1993, the project of a long-term biotechnological co-operation between the Italian CNR and the State of New York [98], while a review on "5-*methylcytosine in genes with methylation-dependent regulation*" opened an issue of *FEBS-Letters* [99].
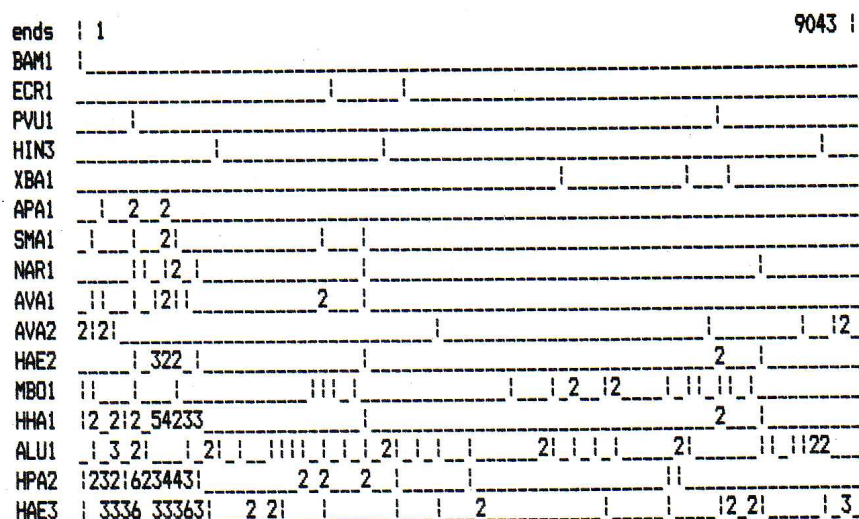
Complementing this, it is not superfluous to further make the subject clearer through a careful reading of *restriction minimaps* constructed for a large number of human genes regulated by m^5C: their genetically discontinuous "geometry" (**Figure 4(A)**) exhibited, after the promoter, a *sui generis* wavy chain of hypomethylatable coding and hypermethylatable uncoding regions (**Figure 4(B)** and **Figure 5**). Both in *housekeeping* (HK) and *tissuespecific* (TS) transcriptional units, they were roughly contained beneath the limits of a "scalene triangle".

For example, the *calcitonin* (CALC) transcriptional unit–representative of about thirty fully A, T, G, C-sequenced and m^5C-regulated TSs—was heavily methylatable upstream and slightly methylatable downstream (**Figure 5(b)**).
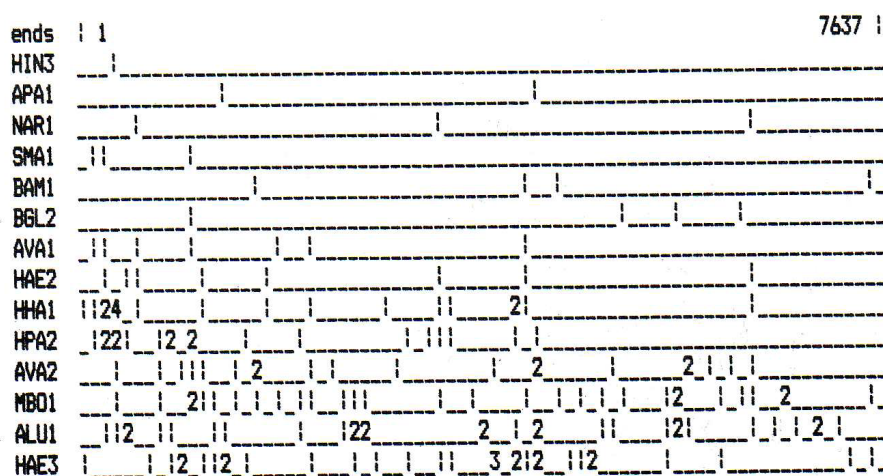
In such a framework, at least four pairs of dinucleotides behaved as targets recognizable by restriction-modification endonucleases [98] [99].

In exons, two of them were monomethylatable:

5'-Tpm⁵C-3'   5'-Cpm⁵C-3'
3'-ApG-5'   3'-GpG-5'

```
ends  | 1                                                                    9043 |
BAM1  |_____
ECR1  _____|_____|_____
PVU1  ____|_____|_____
HIN3  _____|_____|_____|___
XBA1  _____|_____|__|_____
APA1  __|_2_2_____
SMA1  _|___|_2|_____|___|_____
NAR1  _____||_|2_|_____|_____
AVA1  ||__|_|2||_____2___|_____
AVA2  2|2|_____|_____|_____|_|2_
HAE2  _____|_322_|_____|_____2__|_____
MBO1  ||__|__|__|_____|||_|_____|__|_2_|2___|_||_||_|_____
HHA1  |2_2|2_54233_____|_____2__|_____
ALU1  _|_3_2|___|_2|_|_||||_|_|_|_2|_|_|_|____2|_|_|_|____2|_____||_|122__
HPA2  |232|623443|_____2_2__2_|_____|_____||_____
HAE3  |_3336_33363|__2_2|___|_____|__|_|__2_____|_____|___|2_2|____|_3_
```
(a)

```
ends  | 1                                                                    7637 |
HIN3  ___|_____
APA1  _____|_____|_____
NAR1  ____|_____|_____|_____
SMA1  _||_____|_____
BAM1  _____|_____|_|_____|_
BGL2  _____|_____|___|____|_____
AVA1  _||_|___|____|_____|_|_____
HAE2  _|_||____|____|_____|_____|_____
HHA1  ||24_|____|___|___|___|____|___||___2|_____|_____
HPA2  _|22|__|2_2____|___|_____|_|||____||_____
AVA2  ___|__|_|||__|_2____|_|_____|_____|__2_____|____2_|_|_|_____
MBO1  ___|__|_2||_|_|_|_|_||_|||_____|_|_____|_|_|_|_|_|2__|_||_2_____|_
ALU1  _||2_||__||_____|__|22_____2_|_2____||___|2|____|_|_|2_|___
HAE3  |____|_|2_||2_|_____|___|_|_|_||__3_2|2_||2_____|____|_____||_|
```
(b)

Figure 5. Scattered methylatable cites along the HK and TS genes in mammals. The restriction minimaps were drawn using the Mapdrow DNAstar Program. At the 5'-end of a gene, the promoter region showed the highest number of dcm-sites, accounting for a significant hypermethylation [98] [99] [146]. Going from 5' to 3', while the alternate coding (exon) sequences carried a relatively low number of dcm-sites, the alternate uncoding (intron) sequences carried a relatively high number of dcm-sites. However, this level steadily decreased going to the 3'-end. (a) The HK *ODC* gene contained 9.043 nucleotide pairs; (b) The TS *CALC* gene contained 7.637 nucleotide pairs. (Ends) dcm-inhibited restriction endonucleases; (1) single dcm-sites; (2-6) multiple dcm-sites.

In promoter and introns, the other two were dimethylatable:

5'-m^{5}CpG-3'   5'-Gpm^{5}C-3'

3'-Gpm^{5}C-5'   3'-m^{5}CpG-5'
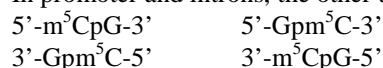
At any rate, the average quantitative distribution of these pairs–which corresponded to those isolated from the DNAse I digests [100]-[102]—was similar in HKs (**Figure 5(a)**) and TSs (**Figure 5(b)**), despite the fact that in the case of TSs (going in the direction from 5' to 3') the monomethylatable dinucleotide pairs proportionally increased while the dimethylatable dinucleotide pairs proportionally decreased [98] [99].

Thanks to identification of all these methylatable "words", the promising 1993 seemed to anticipate, therefore, that the biomolecular studies were moving into a challenging new cycle, whilst new resources such as *restriction gene minimaps* were enabling genome-wide investigations that could potentially identify most common genetic determinants of human health, disease or drug design.

7. The "Repair-Modification" System Played a Role in Either Gene Evolution or Neoplastic Cell Transformation

If we look back for a moment at the investigations regarding the cell-cycle dependence of DNA methylation [47] *vs.* the non-random scattering of m^{5}C along semiconservatively newly replicating DNA chains [46], we can appreciate how they not only paved the way to the discovery of the hypermethylation of promoter and all regulatory and signal sequences that, in the eukaryotic gene, did not code for mRNAs [46] but also considered the corollary experiments which, albeit at the beginning oriented towards other goals (particularly towards deciphering of the –CH$_3$ language), at the end helped to shed some new light on the fine control of post-synthetic DNA modification.

7.1. A Differential Methylation of Euchromatic *vs.* Heterochromatic DNAs Took Place

Evidence indicated that the S-phase should be subdivided into two parts with respect to the characteristics of DNA, since early replicating euchromatic DNA tended to be CG-rich, while late replicating heterochromatic DNA tended to be AT-rich [103].

Furthermore, the DNA fraction extracted from *Chinese hamster* cells in early S was methylated to a greater extent when compared to the DNA fraction extracted from the same cells in late S [104].

This difference was reproduced in *HeLa* cells [46]. On its basis, we supposed that, while in newly replicating DNA chains there would be sequences not uniformly methylated in the course of S, the CG-rich sequences would be preferentially methylated [46].

7.2. A Biological Clock Controlled Methylation of Specific Replicating Sequences

Since DNA did not appear to be uniformly methylated in S, investigations continued to focus on nitrogen bases specifically adjacent to m^{5}C [46] rather than considering the rough subdivision of hypermethylated euchromatic *vs.* hypomethylated heterochromatic replicons [104].

With this purpose in mind, methylated DNA coming from *HeLa* cells was fractioned in alkaline Ag$^+$/Cs$_2$SO$_4$ gradients: a small heavy CG-rich fraction and a large light AT-rich fraction were obtained (the first—containing genes for rRNA—was mainly expressed in early S and, the second, was mainly expressed in late S); as for the m^{5}C concentration, if in early S it was found to increase on the heavier peak, in late S it was found to increase on the lighter peak [105]-[107].

All this suggested that, at the different stages of S, one dealt with CG- or AT-rich regions probably polymerized in correspondence of specific hypo- or hypermethylated templates: diversely, it seemed that the genes would be methylated according to a given order and intensity along the newly replicating chains [105].

This idea was definitely demonstrated when we observed that the foldback DNAs were subjected to hypermethylation: if in *HeLa* cells—through the S-phase—a methylation wave chronologically involved palindromic, highly repeated and moderately repeated sequences, the unique sequences were characterized by a minimal late methylation [108].

7.3. Both CG- and AT-Rich Sequences Were Targets for Methylation

A circumstantial analysis of the two genomic DNA fractions, separated in alkaline Ag$^+$/Cs$_2$SO$_4$ gradients during S, confirmed that m^{5}C had a differential distribution in them: the highest concentration of methylated sequences was found in the denser side of the heavy fraction; in the light DNA the highest concentration of methylated sequences was found in the lighter side [105]-[107].

This implied two generalizations: hypermethylation on CG-rich sequences would offer no surprise because of its rather statistical character (the CG-rich rRNA genes would be hypermethylated according to the size of their CG targets); hypermethylation on AT-rich sequences would seem to work, instead, against the principle of CG target size, since a decrease in C residues would correspond to an increase in their methylation. Hence, when the extra-S time was characterized by a minimal, probably *de novo* DNA methylation, S exhibited maintenance DNA methylation [47].

For this reason, one assumed, the mechanisms discussed above should function in S [90].

But it was obvious that, with consecutive cell cycles (**Figure 6(a)**), the methylated sequences could not increase without limits [109], also because changes in the amount of methylated DNA during development [100]
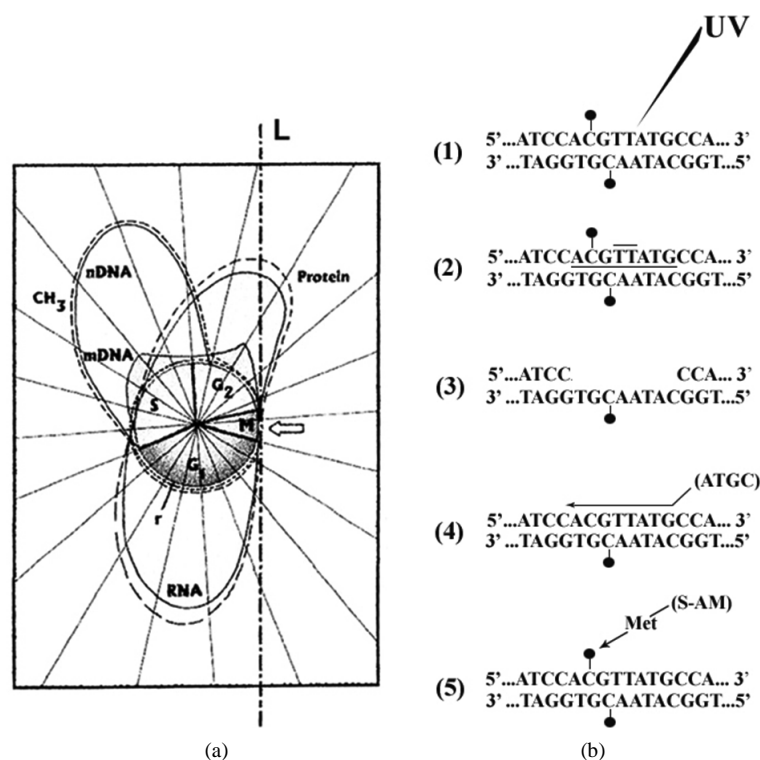
Figure 6. Inverse correlation between DNA methylation and gene expression as a function of cell cycle or repair-modification. (a) DNA methylation [47] *vs*. biosynthesis of DNA [101], RNA [110], and protein [111]: the hours were signed by points where spokes crossed the circular "abscissa"; the ordinates were represented by the same spokes; the white arrow, against M, referred to a line of cell-cycle asymmetry L [65] [67]. Following appropriate labelling, at hourly intervals, macromolecules were extracted, purified and analyzed on gradients [50] [101] [110] [111]. Replicative synthesis of genomic DNA (—) and its methylation (--) showed apogees in S, as in **Figures 3(a)-(b)**, while repair synthesis of genomic DNA (--) was rather constant around the cycle [50] [90] [138]. Replication of mtDNA (—) showed maxima in S and G_2, whereas its methylation also occurred in S [54] [101]. Differently, transcription of RNA (—) and its translation into protein (−−) exhibited almost overlapped apogees in G_1 and G_2 [90] [138]. In M, all macromolecular biosyntheses were rather repressed [35] [36] [90] [138]. (b) Repair-modification mechanism [53] [95] [138] [142]-[144] [148]: UV at random would induce the building of TT-dimers in vicinity of symmetrically dimethylated 5'-m^{5}CpG-3'/3'-Gpm^{5}C-5' dinucleotide pairs, for instance (2); in this case, after digestion of damaged regions (3), which previously included a m^{5}CpG dinucleotide (1), the cell cycle-independent excision-repair would replace the m^{5}C nucleotide with a simple C nucleotide, since repair DNApol would not find any methylated dCTP in the soluble pool of triphosphonucleosides; the resultant hemi-methylated CpG/Gpm^{5}C dinucleotide pairs, namely the incompletely reconstructed RPs (4), should gain the capability of resuming transcription; in S, the cell cycle-dependent DNAmet would add to C the lost –CH$_3$ (5), providing completely reconstructed RPs able to silence the transcription again.

and in differentiated tissues [29] [30] were not conspicuous.

So, whatever the mechanism of DNA methylation was in S, statistical or specific, it was reasonable to imagine that, mostly during the extra-S time, each small wave of *de novo* DNA methylation should be followed by a corresponding small wave of DNA de-methylating repair [53].

7.4. After a Gene Radiodamage, the Complete Rebuilding of Unmethylated Coding Regions Was Guaranteed by "Excision-Repair", While the Complete Rebuilding of Methylated Uncoding Regions Was Guaranteed by "Repair-Modification"

In *HeLa* cells, replication of genomic DNA [65] [67] [101] was followed by its methylation [47] in S, whereas transcription [110] and translation [111] mainly resulted in G_1 and G_2. This was the first concrete suggestion as to a macroscopic inverse correlation observable between gene methylation and expression (**Figure 6(a)**). As for the enzyme components: DNA polymerase (DNApol) *alpha*, responsible for replication of genomic DNA, was induced in S; DNApol *gamma*, responsible for replication of mtDNA, was made in S and G_2; DNApol *beta*, in-

volved in rebuilding, was produced at a rather low level through the whole mitotic cycle with some increase in early S (the increased DNApol *beta* repair activity in early S signifying an effort by the cell to highly reconstruct the genomic DNA just before its replication) [121]. Moreover, in conformity with the sharp peak of maintenance genomic DNA methylation in S (**Figure 6(a)**), the DNAmet system [63] was induced in S as well (**Figure 3(b)**), while a moderate presence of DNAmet characterized the rest of the cell cycle mainly accounting for *de novo* genomic DNA methylation (**Figure 3(b)**). During S, a *coupling*—definable as *repair-modification*—took place between the activities of DNApol *beta* (involved in *excision-repair*) and DNAmet (making *re-methylation*), whereas during the extra-S time an *uncoupling* occurred between these two enzyme systems [53] [64].

Thus, using synchronized *HeLa* cells [65] [67], we took a look at the differential *reconstruction* of the UV-damaged unmethylated coding and methylated uncoding gene sequences (**Figure 6(b)**): in a *repair patch* (RP) located along given promoters or introns (1), it could happen that a symmetrically dimethylated 5'-m^{5}CG-3'/3'-Gm^{5}C-5' dinucleotide pair would flank a UV-induced TT-dimer (2); after digestion of the damaged nucleotide sequence (3), the *excision-repair* (being the methylated dCTP absent in the soluble pool of triphosphonucleosides) would replace—in the "incompletely" reconstructed sequence—the previous m^{5}C nucleotide with a normal unmethylated cytosine nucleotide (4). Even if the existence of DNA de-methylating proteins should not be excluded in general [113], this "non-enzymatic de-methylation" certainly would occur through the whole cell cycle. Anyhow, during S—*via* AdoMet—DNAmet would add to the incompletely reconstructed (unmethylated) DNA sequence (4) the lost –CH$_3$ group, rebuilding the previous symmetrically di-methylated 5'-m^{5}CG-3'/3'-Gm^{5}C-5' dinucleotide pair (5). This "complete" rebuilding of a previously methylated RP would require, in any situation, a "re-methylation" catalyzed by DNAmet [53] [95] [114], as demonstrated by the fact that the labelled carbon atom coming from the –CH$_3$ of [^{14}C]Met (**Figure 3** and **Figure 2(b)**) was found in the T, G, A and m^{5}C nitrogen bases carried by the re-methylated RPs (**Figure 7(c)**) distributed along the light parental strands (**Figure 7(b)**). Yet, to what kind of regions—along the eukaryotic transcription unit—did such RPs belong?

Sedimentation of the semiconservatively newly replicating chains, made heavier in alkaline CsCl by BrdUrd incorporation (**Figure 7(a)**), was confronted with that of the labelled RPs appearing in the light parental chains (**Figure 7(b)**). These light parental chains, after an extensive shearing by sonication [46], were hybridized with polyA-containing pre-mRNA extracted from nuclei (**Figure 7(c)**) or with polyA-containing mRNA extracted
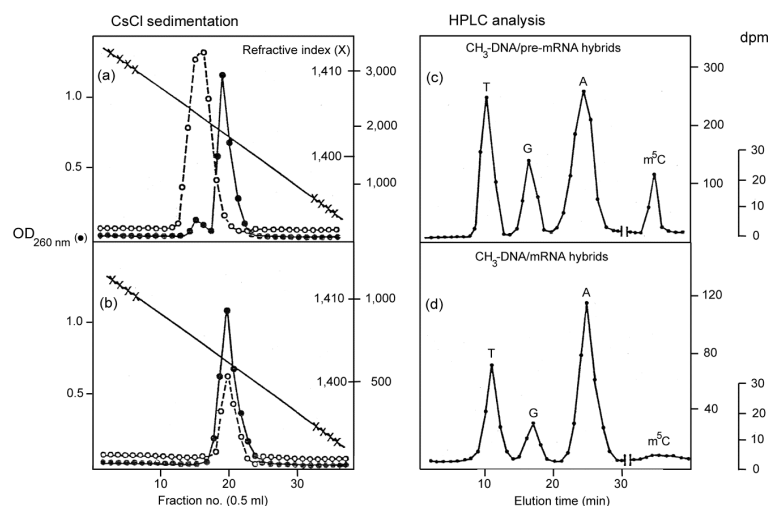

Figure 7. Differential repair-modification of introns and exons following a DNA radiodamage [96]. (a) *HeLa* cells were labelled, in mid S-phase, for 1.5 hrs with 80 Ci/mmol [^{14}C]Met in the presence of 10 mM BrdUrd; (b) *HeLa* cells were UV-irradiated for 45 sec and labelled with 80.0 Ci/mmol [^{14}C]Met in the presence of 10 mM BrdUrd plus 5 mM hydroxyurea. In both cases, the CsCl gradients, brought to pH 8.5, were run in the rotor SW55Ti for 60 hrs at 32,000 rev./min. The [^{14}C]-repaired parental DNA (b), sheared by sonication to fragments of about 1×10^6 Daltons, was hybridized with the [^{3}H]-labelled pre-processed polyA-containing mRNA fraction, purified from nuclei (c), or with the [^{3}H]-labelled processed polyA-containing mRNA fraction, purified from polysomes (d). Both pre-processed and processed messenger RNA fractions were labelled in unsynchronized cells with [^{3}H]uridine for 30 min in the presence of 0.04 micrograms/ml Actinomycin D and purified through oligo-dT cellulose. The hybrids were hydrolyzed to bases and chromatographed by HPLC. The reduced scale at the right-hand side (^{14}C-label) concerned m^{5}C.

from polysomes (**Figure 7(d)**). In the first combination, the hybrids "DNA/pre-mRNA" showed an appreciable amount of m^{5}Cs; in the second, the hybrids "DNA/mRNA" did not appear to be methylated. These results signified that re-methylation involved the uncoding parts of the repaired genes and that the coding parts of these genes only underwent a rebuilding of their primary base sequence, without modified bases.

In summary, we considered one of the possible mechanisms—*repair-modification*—through which the presence of m^{5}C in eukaryotic genes might have exerted a key role in the evolution of their unmethylated coding *vs*. methylated uncoding regions: after radioinduced damage, the RPs formed in exons (where the modified bases were practically absent) would be completely reconstructed by the sole excision-repair, whereas the RPs formed in promoters and introns (where the modified bases were normally present) would be incompletely reconstructed by excision-repair. Only later, in S, the unmethylated RPs would become completely reconstructed by re-methylation.

We assumed, therefore, that the repair-modification machinery exerts a continuous differential pressure on evolution of given gene pieces, when methylated to a different extent (**Figure 7(c)** and **Figure 7(d)**).

In particular, according to Darwinian concepts, *repair-modification* would lead to a very high variability of *introns* (if hypermethylated), whereas on the basis of this wide variability selection would favor their transposition from one gene to another. In them, the conservative nature of *exons* (if unmethylated) would be ensured by *excision-repair*.

Besides, it is worth stressing the possible influence that the incomplete repair of RPs (see the de-methylation shown in **Figure 6(b4)**) could exert on resumption of transcription by virtue of the cell cycle-dependent inverse correlation seen between DNA methylation and RNA synthesis (**Figure 6(a)**).

In this case, in harmony with the *azacytidine effect* [94] observed in *HeLa* cells (**Figure 8**), certain previously silent (**Figure 6(b1)**) and then re-activated (**Figure 6(b4)**) genes would lead even to neoplasy.

After all, the idea suggesting that the radio-induced DNA de-methylation could be considered as a feasible molecular basis for cancerogenesis [53] conformed with the known fact denouncing that at least 30 per cent of tumors are really induced in humans by radiations [148] [149].
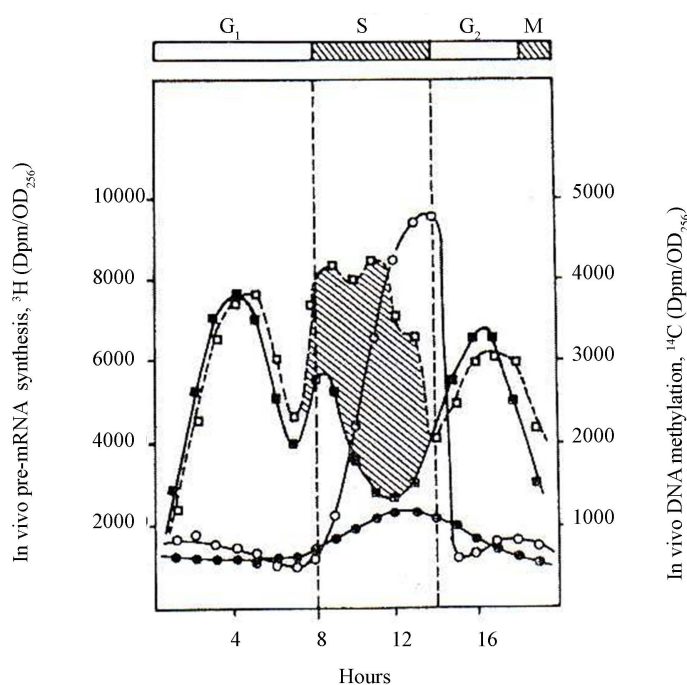
Figure 8. Inverse correlation in S phase between inhibition of DNA methylation by 5-azacytidine and polyA-containing pre-mRNA transcription. The experiments were performed in synchronized *HeLa* cells, as for **Figure 3(a)** and **Figure 6(a)**: following the inhibition of genomic DNA methylation by 5-azacytidine [94], some resumption of transcription was evidenced (dashed area). The left ordinate showed the rates of pre-mRNA synthesis in control (black squares) and 5-azacytidine-treated (white squares) cells labelled with [^{3}H]uridine; the right ordinate showed the rates of genomic DNA methylation in control (open circles) and 5-azacytidine-treated (closed circles) cells labelled with [^{14}C]Met.

8. The Arduous Biochemical Approach to Decipher the Language of Gene Methylation in Eukaryotes Was Crowned with Success by Its Direct Bisulfite Reading

The finding of an alternation, after the promoter, of coding and uncoding regions in the eukaryotic transcriptional unit and the consequent demonstration that these regions are differentially methylated, or re-methylated, inevitably stimulated curiosity to decipher the language marked by their m^5Cs.

8.1. At Variance with the Systems of Regulation of the Operon Activity in Bacteria, an Inverse Correlation between Gene Methylation and Transcription Was Shown in Humans

In bacterial DNA the m^6A and m^5C modified bases were related not only to the restriction-modification anti-phage defense [6]: complementing the induction or repression operon functions [39], they were further associated with an auxiliary mechanism capable of partially regulating transcription in the same phage-infected microorganisms [115]. This "embryonic" stage of control evolved in the major part of eukaryotes: in them the higher concentration of m^5C, present in the promoter and introns, was suggested to be responsible for modulation of gene expression [46], as supported by others [60] [116] [117] and reviewed [89] [118]. Such an involvement found its first clear demonstration in the fact that the maximal rate of DNA methylation [47] [119] followed the maximal rate of DNA replication [101] during S, while the maximal rates of transcription [110] and translation [77] [111] appeared during G_1 and G_2 (**Figure 6(a)**). Further results supporting the idea about the role of m^5C in regulation of gene expression were soon detected [120]-[123], afresh to those on the methylation-dependent regulation of integrated viral genes [92]-[109] [117] [124] [125]. Moreover, *maintenance* DNAmet [47] [72] was shown to recognize hemimethylated 5'-CpG-3'/3'-Gpm^5C-5' dinucleotide pairs to yield dimethylated 5'-m^5CpG-3'/3'-Gpm^5C-5' dinucleotide pairs (**Figure 3(c)**), whilst an inverse correlation between methylation and expression was established for a large number of sequenced HK and TS genes [98] [99]. In fact, an inverse correlation between gene methylation and transcription did occur in animals signifying, at the same time, that it could be caused most likely by the higher concentration of m^5C in promoter and intron sequences: one theory justifies the other, and *vice versa*.

8.2. A Study Focused on Isolation of Specific Targets for Methylation in Native Human DNA

After establishing that in eukaryotes the $-CH_3$ groups are preferentially attached to the CpG dinucleotides [106] [107] [126], as in bacteria [127], clearcut evidence about the existence of two classes of targets for DNAmet emerged, as mentioned already, from analysis of *HeLa* cell double-helical DNA in alkaline Ag^+/Cs_2SO_4 [106].

This type of gradient separated a heavier fraction (representing about 20% of the total) from a lighter one (representing about 80% of the total) [128].

However, the consequent ultracentrifugation of these fractions in CsCl [106] further showed that the heavier, banding at 1.715 g/cm^3, contained 53% CG (10% of the total CG), whereas the lighter, banding at 1.703 g/cm^3, contained 40% CG (32% of the total CG).

Generalizing, four possible triplets as sites for DNAmet recognition were suggested: GCG and CCG, in CG-rich sequences; ACA and ACG, in AT-rich sequences [107].

The evidence that the specific methylation of CG-rich sequences (as those for rRNAs) was maximal in early S and that the specific methylation of AT-rich sequences was maximal in late S not only demonstrated that the DNA sequences were replicated and methylated following an order during S [107] [108] but also pointed that the DNAmet activity exerted the role of *maintenance* modification along CG-rich sequences and the role of *de novo* modification along AT-rich sequences [107].

This important deduction was in accordance with the fact that the extra-S methylation (about 10% of the total) involved some AT-rich sequences either along the old or the new chains, when the S phase methylation (about 90% of the total) almost exclusively involved newly replicating chains [47] [50].

In S, the *maintenance* DNAmet function [63] followed at a distance of about 30 min [47] the DNApol *alpha* activity [112].

The reasoning made sense because, in S, the *maintenance* modification—involving CpG targets in newly rep-

licating chains—would be induced by the –CH₃ of the m⁵C present in the complementary strand as a genetically encoded signal.

Otherwise, a CpG dinucleotide (**Figure 3(c)**) would be methylated on the new chain in the presence of a complementary already methylated GpC dinucleotide on the old chain [47] [72]; differently, during the extra-S time, a *de novo* modification occurring on an ApC dinucleotide, for instance, could not be induced by a signal coming from the complementary TpG dinucleotide since this cannot be methylated at all.

8.3. Both *De Novo* Monomethylated and *Maintenance* Dimethylated Dinucleotide Pairs Were Found in Humans

For the sake of clarity, the above-established sedimentation of CG- and AT-rich sequences in alkaline Ag^+/Cs_2SO_4 suggested the existence of at least four targets for methylation [106]:

GCG, CCG, ACA, and ACG.

In one of the trinucleotides there was a methylatable CpA; in three of them, the methylatable CpG was repeated. For this reason, a direct isolation of such restricted targets was required (**Figure 9**).
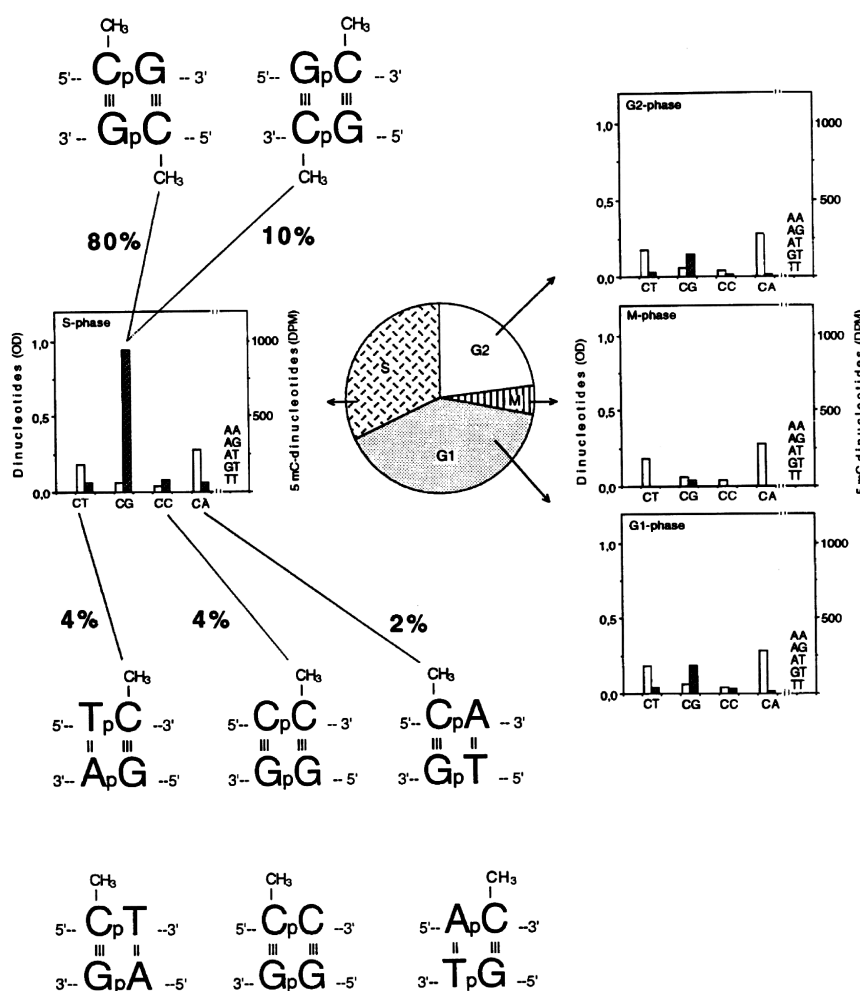
Figure 9. Methylated "words" in eukaryotic genomic DNA. Experiments with synchronized *HeLa* cells [65] [67] demonstrated that, in double helix (characterized by antiparallelism of complementary chains) two symmetrically dimethylated palindromic dinucleotide pairs and six asymmetrically monomethylated non-palindromic dinucleotide pairs were detected [102] [138]. DNA was labelled in isolated nuclei with [¹⁴C-methyl]AdoMet, as for **Figure 3(b)**: after extraction, in the middle of each cell-cycle phase (G₁, S, G₂ and M), it was digested with pancreatic DNAse I [129] to chromatografically separate various dinucleotides [102]. The white columns show the dinucleotide molar concentration, expressed in OD at 256 nm; the black columns, expressed in DPM, show the radioactivity of their m⁵C.

This was achieved by exploiting DNAase I digests [126] [129] of DNA samples methylated in isolated nuclei of previously synchronized cells with [^{3}H-methyl]AdoMet [47] [109]. Chromatographic analysis of the material contained in these digests [102] yielded a number of obviously unmethylated dinucleotides

ApA, ApG, ApT, CpT, and TpT

and, for the rest, four dinucleotides methylated to a different extent:

m^{5}CpT, Cpm^{5}C, m^{5}CpA, and m^{5}CpG.

Their methylation level was cell-cycle dependent, in agreement with [47]: in M, few CpGs were methylated; in G$_1$ and G$_2$, methylation of CpGs slightly increased, in comparison with that seen in M, and there appeared some methylation on CpTs, CpCs, and CpAs; in S, again, a methylation of four dinucleotides

m^{5}CpTs, Cpm^{5}Cs, m^{5}CpAs, and m^{5}CpGs

was detected. In this phase, methylation of CpGs became intensive (90% of the total), accounting for its *maintenance* character along newly made chains (**Figure 3(c)**). The occurrence of methylation on CpTs, CpCs, and CpAs accounted instead for a *de novo* phenomenon, which characterized especially the extra-S part of the interphase [47]. Considering the two possible directions of the separated dinucleotides, due to the double helix anti-parallelism (the 5'-3' and 3'-5' directions could not be distinguished chromatographically), the analysis actually regarded the identification of eight *words* written in the genome in terms of dinucleotide pairs:

5'-m^{5}CpG-3'	5'-Tpm^{5}C-3'	5'-Cpm^{5}C-3'	5'-m^{5}CpA-3'
5'-Gpm^{5}C-3'	3'-ApG-5'	3'-GpG-5'	3'-GpT-5'
+	+	+	+
3'-m^{5}CpG-5'	5'-m^{5}CpT-3'	5'-m^{5}CpC-3'	5'-Apm^{5}C-3'
3'-Gpm^{5}C-5'	3'-GpA-5'	3'-GpG-5'	3'-TpG-5'

Briefly, if the CpG target only allowed a final dimethylation of dinucleotide pairs, as expected from semiconservative methylation [47] [72] [102] [130], the other combinations always led to a monomethylation of dinucleotide pairs [102]. Still, what could be demonstrated about the diversity of the methylation code at the level of the promoter and at that of the intervening uncoding sequences?

8.4. In the Promoter of the Transglutaminase Gene the Bisulfite Sequencing of 5-Methylcytosine Showed Two Fully Methylated CpG-Rich Domains Close to the 5'-End and One Fully Unmethylated CpG-Rich Domain Close to the 3'-End

Using the model of human *transglutaminase* (hTGc) gene [131] and the method based on bisulfite conversion of C to T residues along a DNA filament [132], the following analysis showed that m^{5}C, behaving as bisulfite-independent base [133], could be sequenced directly despite its low molar proportion [134].

The hTGc gene was chosen principally because it was among those that are regulated by methylation [135]. In its 1665 bp long promoter [131], three CpG-rich domains were characterized (**Figure 10**): the first domain (330 bp), close to 5', contained 12 CpGs corresponding to an average frequency of 3.63% with respect to the total number of nucleotides; the second (227 bp), roughly located in the middle of the promoter, contained eight CpGs corresponding to an average frequency of 3.52%; the third (264 bp), close to 3' (including 70 bp of the 73 bp long 5'-UTR), contained 31 CpGs corresponding to an average frequency of 11.74% [136].

In leukocytes and lymphocytes, where hTGc is silent [135], out of the three CpG-rich domains, only the first two were found to be methylated, while the third, on the 3'-side, did not present any m^{5}Cs: the 11 CpGs of domain 1 and the 7 CpGs of domain 2 were methylated 100%; the 31 CpGs of domain 3 were on the contrary unmethylated 100% (in this domain the 5'-UTR, also resulting 100% unmethylated, was almost entirely included) [136].

The lack of m^{5}C in domain 3 was expected, in agreement with the conventional definition of unmethylated CpG-rich "island" (where the intrafilament CpG/GpC ratio has to be higher than 0.6); but the occurrence of methylation in all CpGs of domains 1 and 2 was unexpected. Domain 1 contained 79 Cs and 11 CpGs out of 330 bases; domain 2 contained 44 Cs and 7 CpGs out of 227 bases; domain 3 contained 116 Cs and 31 CpGs out of 264 bases. Compared with the average 5%-methylation of human genomic DNA [137], methylation of domains 1 and 2 corresponded to 15.18% and 18.18%, respectively [136]. This confirmed the idea that the promoter of an inactive gene should be characterized by hypermethylation [46] [98] [99].

Certainly, *HUVEC* cells joined their hTGc gene activity with a loss of m^{5}C along the hTGc promoter domain 1, at least at the −1380, −1349, −1338, and −1320 CpG sites [138]. This is an appropriate example of direct
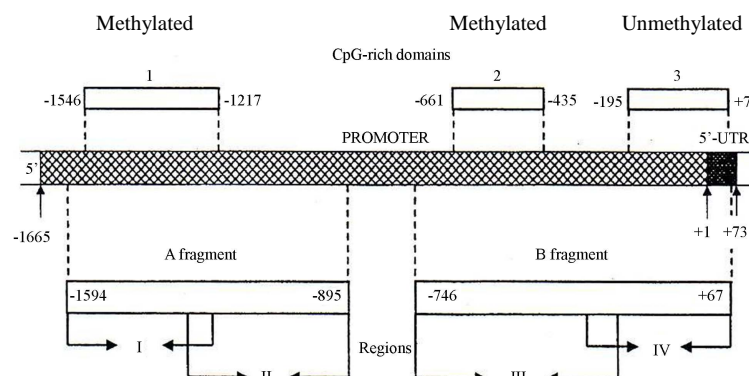
Figure 10. Scattered methylation along the repressed hTGc gene promoter in lymphocytes. The sequenced bisulfite-converted fragments A and B were investigated [136]: all CpGs found in domain 1 were methylated; all CpGs found in domain 2 also were methylated; in contrast, all CpGs found in domain 3 were unmethylated (starting from site +1, the fragment B, with its domain 3, overlapped a 5'-UTR, also unmethylated). In *HUVEC* cells, carrying an active hTGC gene, several CpGs of domain 1 were unmethylated [136] [138].

base-sequence analysis showing the inverse correlation between the activity of a gene and the methylation of its promoter (of course, in harmony with this correlation, the on-going research is aimed at clarifying the regulatory role of m^{5}Cs along the non-coding sequences).

9. Interpretations of the Mechanisms Switching-on and -off Transcription in Higher Cells: Future Perspectives

From the picture described herewith what hypothesis could one deduce on the role of methylation in regulating gene activity?

Supporting the message coming from **Figure 8** and **Figure 10** and from the long list of genes regulated by methylation [98] [99], the analysis based on the use of methylation-sensitive restriction endonucleases further showed that in human lymphocytes and monocytes an inverse correlation between the hTGc gene expression and methylation of its promoter took place [135].

Furthermore, the study based on the direct sequencing of m^{5}C showed that in the same cell species, in the absence of hTGc activity, the hTGc promoter was hypermethylated in its CpG-rich domains 1 and 2 and unmethylated in its CpG-rich domain 3 [136].

This neat division of hTGc promoter in two parts—one methylated and the other not—deserved particular attention in correlating DNA heterochromatization with A or B m^{5}CpG-binding proteins [123] [139]-[141] [139]-[141]: one could assume that they might transform domains 1 and 2 into heterochromatic structures and that, in turn, these structures would be sufficient to prevent transcription (methylation would be relevant to the repressor complex able to bind the promoter 5'-end).

Then, one could ask why, in hTGc promoter, the unmethylated domain 3, remaining in normal relaxed state, would be unable to interfere with the basal transcription mechanism at the promoter 3'-end [138].

Combined with the inactivity of the hTGc gene in leukocytes [136], the postulated heterochromatization of domains 1 and 2, in its promoter, could acquire great interest if considered within the framework of the repair-modification scheme (**Figure 6(b)**). Causing an *at random* DNA de-methylation [95] [142], it would lead, in general, to switch-on of previously silent genes, as in the case of genes for *alpha* and *beta* chains of hemoglobin in *Friend erythroleukemia* cells [99] [143].

In this situation, the switch-on and -off of transcription—implying conformational changes of the two hemoglobin gene promoters—would correspond, first, to their de-methylation and, second, to their re-methylation [99] [143] [144].

In other words, following the damages caused to double helix by ionizing radiations [95] [142], excision-repair would be sufficient to guarantee a complete reconstruction of previously unmethylated regions; nevertheless, to complete the reconstruction of previously methylated regions, a coupling would arise [53] between excision-repair (re-establishing the basic code in A, G, T, and C) and DNAmet (re-establishing the position of m^{5}C among A, G, T, and C). Actually, a coupling (properly meaning *repair-modification*) took place in S [64]: a spe-

cific DNApol, crucial for excision-repair, was active during the whole cell cycle [112]; DNAmet was highly active in S and almost inactive during the major part of the extra-S time [63]. Once de-methylated through excision-repair, given genes, silent when methylated, could be expressed if their transcription was inversely correlated with their methylation.

10. Conclusion

In conclusion, our basic hypothesis was that m^5C might function as a signal for genetic regulation of transcription through the hypermethylated promoter and the intermittent methylated non-coding regions. After a possible association with m^5C-binding proteins, they would physically interfere with the slip of RNA polymerase (RNApol) on the transcriptional unit. Probably, the block of transcription starts from the methylated promoter (initially recognized by a type A m^5C-binding protein able to compete with RNApol) and then is potentiated, inside the gene, by an intermittent association of uncoding methylated regions with a type B m^5C-binding protein.

Acknowledgements

References

[1] Scarano, E., Iaccarino, M., Grippo, P. and Winckelmans, D. (1965) On Methylation of DNA during Development of the Sea Urchin Embryo. *Journal of Molecular Biology*, **14**, 603-607. http://dx.doi.org/10.1016/S0022-2836(65)80211-X

[2] Vanyushin, B.F., Belozersky, A.N., Kokurina, N.A. and Kadirova, D.X. (1968) 5-Methylcytosine and 6-Methylaminopurine Inbacterial DNA. *Nature*, **218**, 1066-1067. http://dx.doi.org/10.1038/2181066a0

[3] Lark, C. (1968) Studies on the *in Vivo* Methylation of DNA in *Escherichia coli* 15T. *Journal of Molecular Biology*, **31**, 389-399. http://dx.doi.org/10.1016/0022-2836(68)90416-6

[4] Arber, W. (1965) Host Specificity of DNA Produced by *Escherichia coli*. V. The Role of Methionine in the Production of Host Specificity. *Journal of Molecular Biology*, **11**, 247-256. http://dx.doi.org/10.1016/S0022-2836(65)80055-9

[5] Arber, W. and Linn, S. (1969) DNA Modification and Restriction. *Annual Review of Biochemistry*, **38**, 467-500. http://dx.doi.org/10.1146/annurev.bi.38.070169.002343

[6] Arber, W. (1974) DNA Modification and Restriction. *Progress in Nucleic Acids Research and Molecular Biology*, **14**, 1-37. http://dx.doi.org/10.1016/S0079-6603(08)60204-4

[7] Hattman, S., Brooks, J.E. and Masurekar, M. (1978) Sequence Specificity of the PI Modification Methylase (M·*Eco P*1) and DNA Methylase (M·*Eco dam*) Controlled by *Escherichia coli dam* Gene. *Journal of Molecular Biology*, **126**, 367-380. http://dx.doi.org/10.1016/0022-2836(78)90046-3

[8] Messer, W. and Noyer-Weidnetr, M. (1988) Timing and Targeting: The Biological Functions of *dam* Methylation in *E. coli*. *Cell*, **54**, 735-737. http://dx.doi.org/10.1016/S0092-8674(88)90911-7

[9] Palmer, B.R. and Marinus, M.G. (1994) The *dam* and *dcm* Strains of *Escherichia coli*. *Gene*, **143**, 1-12. http://dx.doi.org/10.1016/0378-1119(94)90597-5

[10] Marinus, M.G. and Morris, N.R. (1974) Biological Function for 6-Methyladenine Residues in the DNA of *Escherichia coli* K12. *Journal of Molecular Biology*, **85**, 309-322. http://dx.doi.org/10.1016/0022-2836(74)90366-0

[11] Raleigh, E.A. and Wilson, G. (1986) *Escherichia coli* K-12 Restricts DNA-Containing 5-Methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 9070-9074. http://dx.doi.org/10.1073/pnas.83.23.9070

[12] Heitman, J. and Model, P. (1987) Site-Specific Methylases Induce the SOS DNA Repair Response in *Escherichia coli*. *Journal of Bacteriology*, **169**, 3243-3250.

[13] Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Wastaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J. and Hanahan, D. (1988) *McrA* and *McrB* Restriction Phenotypes of Some *E. coli* Strains and Implications for Gene Cloning. *Nucleic Acids Research*, **16**, 1563-1575. http://dx.doi.org/10.1093/nar/16.4.1563

[14] Marinus, M.G. and Morris, N.R. (1973) Isolation of Deoxyribonucleic Acid Methylase Mutants of *Escherichia coli K*12. *Journal of Bacteriology*, **114**, 1143-1150.

[15] May, M. and Hattman, S. (1975) Analysis of Bacteriophage Deoxyribonucleic Acid Sequences Methylated by Host- and R-Factor-Controlled Enzymes. *Journal of Bacteriology*, **123**, 768-770.

[16] Wyatt, G.R. (1951) The Purine and Pyrimidine Composition of Deoxypentose Nucleic Acids. *Biochemical Journal*, **48**, 584-590.

[17] Vanyushin, B.F., Tkacheva, S. and Belozersky, A.N. (1970) Rare Bases in Animal DNA. *Nature*, **225**, 948-994. http://dx.doi.org/10.1038/225948a0

[18] Gorovsky, M.A., Hattman, S. and Plager, G.L. (1973) [^{6}N]methyl-adenine in the Nuclear DNA of a Eucaryote, *Tetrahymena pyriformis*. *Journal of Cell Biology*, **56**, 697-701. http://dx.doi.org/10.1083/jcb.56.3.697

[19] Cummings, D.J., Tait, A. and Goddard, J.M. (1974) Methylated Bases in DNA from *Paramecium aurelia*. *Biochimica et Biophysica Acta*, **374**, 1-11. http://dx.doi.org/10.1016/0005-2787(74)90194-4

[20] Buryanov, Y.I. and Skryabin, G.K. (1970) Detection of 6-Methylaminopurine in DNA of the Fungus *Mucor hiemalis*. *Doklady Akademii Nauk SSSR*, **195**, 728-730.

[21] Pakhomova, M.V., Zaitseva, G.N. and Belozersky, A.N. (1968) The Presence of 5-Methylcytosine and 6-Methylaminopurine in the Composition of DNA in Some Algae. *Doklady Akademii Nauk SSSR*, **182**, 712-715.

[22] Buryanov, Y.I., Eroshina, N.V., Vagabova, L.M. and Ilyin, A.V. (1972) On the Detection of 6-Methylaminopurine in DNA of Higher Plant Pollen. *Doklady Akademii Nauk SSSR*, **206**, 992-994.

[23] Urieli-Shove, S., Gruenbaum, Y., Sedat, J. and Razin, A. (1982) The Absence of Detectable Methylated Bases in *Drosophila melanogaster* DNA. *FEBS Letters*, **146**, 148-152. http://dx.doi.org/10.1016/0014-5793(82)80723-0

[24] Adams, R.L.P., MacKay, E.L., Graig, L.M. and Burdon, R.H. (1979) Methylation of Mosquito DNA. *Biochimica et Biophysica Acta*, **563**, 72-81. http://dx.doi.org/10.1016/0005-2787(79)90008-X

[25] Achwal, C.W., Ganguly, P. and Chandra, H.S. (1984) Estimation of the Amount of 5-Methylcytosine in *Drosophila melanogaster* DNA by Amplified ELISA and Photoacoustic Spectroscopy. *EMBO Journal*, **3**, 263-266.

[26] Lyko, F., Ramsahoye, B.H., Kashevsky, H., Tudor, M., Mastrangelo, M.A., Orr-Weaver, T.L. and Jaenish, R. (1999) Mammalian (Cytosine-5) Methyltransferases Cause Genomic DNA Methylation and Lethality in *Drosophila*. *Nature Genetics*, **23**, 363-366. http://dx.doi.org/10.1038/15551

[27] Lyko, F., Ramsahoye, B.H. and Jaenisch, R. (2000) DNA Methylation in *Drosophila melanogaster*. *Nature*, **408**, 538-540. http://dx.doi.org/10.1038/35046205

[28] Gowher, H., Leismann, O. and Jeftsch, A. (2000) DNA of *Drosophila melanogaster* Contains 5-Methylcytosine. *EMBO Journal*, **19**, 6918-6923. http://dx.doi.org/10.1093/emboj/19.24.6918

[29] Vanyushin, B.F., Mazin, A.L., Vasilyev, V.K. and Belozersky, A.N. (1973) The Content of 5-Methylcytosine in Animal DNA: The Species and Tissue Specificity. *Biochimica et Biophysica Acta*, **299**, 397-403. http://dx.doi.org/10.1016/0005-2787(73)90264-5

[30] Vanyushin, B.F., Nemirovsky, L.E., Klimenko, V.V., Vasilyev, V.K. and Belozersky, A.N. (1973) The 5-Methylcytosine in DNA of Rats: Tissue and Age Specificity and the Changes Induced by Hydrocortisone and Other Agents.

Gerontology, **19**, 138-152. http://dx.doi.org/10.1159/000211967

[31]   Vanyushin, B.F. and Belozersky, A.N. (1959) Base Composition of Deoxyribonucleic Acids of Higher Plants. *Doklady Akademii Nauk SSSR*, **129**, 944-946.

[32]   Vanyushin, B.F. (1984) Replicative DNA Methylation in Animals and Higher Plants. *Current Topics in Microbiology and Immunology*, **108**, 99-114. http://dx.doi.org/10.1007/978-3-642-69370-0_7

[33]   Giuditta, A. and Volpe, P. (1967) Studies on an Unknown Compound from Ox Brain. I. Isolation and Analysis. *Journal of Neurochemistry*, **14**, 626-636. http://dx.doi.org/10.1111/j.1471-4159.1967.tb09566.x

[34]   Volpe, P. and Giuditta, A. (1967) Studies on an Unknown Compound from Ox Brain: II. Separation into Several Components. *Journal of Neurochemistry*, **14**, 637-647. http://dx.doi.org/10.1111/j.1471-4159.1967.tb09567.x

[35]   Eremenko, T., Benedetto, A. and Volpe, P. (1972) Poliovirus Replication during the *HeLa* Cell Life Cycle. *Nature New Biology*, **237**, 114-116. http://dx.doi.org/10.1038/newbio237114a0

[36]   Eremenko, T., Benedetto, A. and Volpe, P. (1972) Virus Infection as a Function of the Host Cell Cycle: Replication of Poliovirus RNA. *Journal of General Virology*, **16**, 61-68. http://dx.doi.org/10.1099/0022-1317-16-1-61

[37]   Volpe, P. and Giuditta, A. (1967) Kinetics of RNA Labelling in Fractions Enriched with Neuroglia and Neurons. *Nature*, **216**, 154-155. http://dx.doi.org/10.1038/216154a0

[38]   Volpe, P. and Giuditta, A. (1967) Biosynthesis of RNA in Neuron- and Glia-Enriched Fractions. *Brain Research*, **6**, 228-240. http://dx.doi.org/10.1016/0006-8993(67)90193-X

[39]   Jacob, F. and Monod, J. (1961) Genetic Regulatory Mechanisms in the Synthesis of Proteins. *Journal of Molecular Biology*, **3**, 318-356. http://dx.doi.org/10.1016/S0022-2836(61)80072-7

[40]   Tomkins, G.M. and Ames, B.N. (1966) The Operon Concept in Bacteria and in Higher Organisms. *National Cancer Institute Monographs*, **27**, 221-234.

[41]   Volpe, P. and Strecker, H.J. (1968) Amphi-Directional Control of a Reversible Reaction Common to Two Enzyme Sequences. *Biochemical and Biophysical Research Communications*, **32**, 240-245. http://dx.doi.org/10.1016/0006-291X(68)90375-6

[42]   Volpe, P., Sawamura, R. and Strecker, H.J. (1969) Control of Ornithine-Delta-Ransaminase in Rat Liver and Kidney. *Journal of Biological Chemistry*, **244**, 719-726.

[43]   Volpe, P. and Eremenko, T. (1971) Sructural Enzyme Heterogeneity as Expressed by Differential Regulation during the Phases of the Cell Cycle. *Advances in Cytopharmacology*, **1**, 257-261.

[44]   Volpe, P. and Eremenko, T. (1971) Biochemical Differentiation and Cancerogenesis. In: Neskovic, B., Ed., *Proceedings of the International Symposium on Experimental Oncology*, Ognjon Prica Publishers, Zagreb, 200-203.

[45]   Volpe, P., Menna, T. and Pagano, G. (1974) Ornithine-Delta-Transaminase Heterogeneity and Regulation: Sequential Expression of the "Liver" and "Kidney" Enzyme Forms during the *HeLa* Cell Cycle. *European Journal of Biochemistry*, **44**, 455-458. http://dx.doi.org/10.1111/j.1432-1033.1974.tb03503.x

[46]   Volpe, P. and Eremenko, T. (1974) Preferential Methylation of Regulatory Gene Sequences in *HeLa* Cells. *FEBS Letters*, **44**, 121-126.

[47]   Geraci, D., Eremenko, T., Cocchiara, R., Granieri, A., Scarano, E. and Volpe, P. (1974) Correlation between Synthesis and Methylation of DNA in *HeLa* Cells. *Biochemical and Biophysical Research Communications*, **57**, 353-361. http://dx.doi.org/10.1016/0006-291X(74)90937-1

[48]   Duranti, T., La Teana, A., Cacciamani, T. and Volpe, P. (2006) The Prokaryotic Origin of the Pathways for Synthesis and Post-Synthetic Modification of DNA. *RNA Biology*, **3**, 49-53. http://dx.doi.org/10.4161/rna.3.1.2794

[49]   Duranti, T., La Teana, A., Cacciamani, T. and Volpe, P. (2006) Evolution of the Pathways Leading to Synthesis and Modification of DNA. *Journal of Biological Research*, **81**, 47-50.

[50]   Eremenko, T., Palitti, F., Morelli, F., Whitehead, E.P. and Volpe, P. (1985) Hypomethylation of Repair Patches in *HeLa* Cells. *Molecular Biology Reports*. **10**, 177-182. http://dx.doi.org/10.1007/BF00778526

[51]   Gehrke, C.W., Kuo, K.C., Davis, G.E., Suits, R.D., Waalkesb, R.D. and Borek, E. (1978) Quantitative High-Performance Liquid Chromatography of Nucleosides in Biological Materials. *Journal of Chromatography*, **150**, 455-476. http://dx.doi.org/10.1016/S0021-9673(00)88205-9

[52]   Marinus, M.G. (1987) DNA Methylation in *Escherichia coli. Annual Review of Biochemistry*, **21**, 113-131. http://dx.doi.org/10.1146/annurev.genet.21.1.113

[53]   Volpe, P. and Eremenko, T. (1989) Repair-Modification and Evolution of the Eukaryotic Genome Organization. *Cell Biophysics*, **15**, 41-60. http://dx.doi.org/10.1007/BF02991578

[54]   Volpe, P. (2007) Post-Synthetic Methylation of Macromolecules in Synchronized Mammalian Cells. *Progress in DNA Methylation Research*, **8**, 203-224.

[55] Feng, T.Y. and Chang, K.S. (1984) Persistence of Maternal Inheritance in *Chlamydomonas* Despite Hypomethylation of Chloroplast DNA Induced by Inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 3438-3442. http://dx.doi.org/10.1073/pnas.81.11.3438

[56] Volpe, P. and Cascio, O. (1993) Detection of a Mammalian DNA Methylase Protein Family. *Rendiconti Lincei*, **4**, 345-357. http://dx.doi.org/10.1007/BF03001194

[57] Volpe, P. and Cascio, O. (1994) Evolution of the DNA Methylase Protein Family in Vertebrates. *Rendiconti Lincei*, **5**, 79-87. http://dx.doi.org/10.1007/BF03001439

[58] Franchina, M., Hooper, J. and Kay, P.H. (2001) Five Novel Alternatively Spliced Transcripts of DNA (Cytosine-5) Methyltransferase 2 in Human Peripheral Blood Leukocytes. *International Journal of Biochemistry and Cell Biology*, **33**, 1104-1115. http://dx.doi.org/10.1016/S1357-2725(01)00074-7

[59] Scarano, E., Iaccarino, M., Grippo, P. and Parisi, E. (1967) Heterogeneity of Thymine Methyl Group Origin in DNA Pyrimidine Isostichs of Developing Sea Urchin Embryos. *Proceedings of the National Academy of Sciences of the United States of America*, **57**, 1394-1400. http://dx.doi.org/10.1073/pnas.57.5.1394

[60] Holliday, R. and Pugh, J.E. (1975) DNA Modification Mechanisms and Gene Activity during Development. *Science*, **187**, 226-232. http://dx.doi.org/10.1126/science.1111098

[61] Borek, E. (1975) Model for Differentiation Based on DNA Modifying Enzymes. *Science*, **190**, 591-593. http://dx.doi.org/10.1126/science.1188359

[62] Bestor, T.H. and Ingram, V.M. (1983) Two Species of DNA Methyltransferase from Murine Erythroleukemia Cells: Purification, Sequence Specificity, and Mode of Interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 5559-5563. http://dx.doi.org/10.1073/pnas.80.18.5559

[63] Delfini, C., Crema, A.L., Alfani, E., Lo Presti, E., Eremenko, T. and Volpe, P. (1987) DNA-Methylases Separated through the *HeLa* Cell Cycle Methodology Show Allosteric Properties. *FEBS Letters*, **210**, 17-21.

[64] Eremenko, T., Delfini, C., Crema, A.L., Alfani, E. and Volpe, P. (1988) Uncoupling of the DNA Polymerase and Methylase Systems Leads to Hypomethylation of Repair Patches. *Macromolecules in the Functioning Cell*, **5**, 37-41.

[65] Volpe, P. and Eremenko, T. (1970) A Method to Measure the Length of Each Phase of the Cell Cycle in Spinner Cultures. *Experimental Cell Research*, **60**, 456-458. http://dx.doi.org/10.1016/0014-4827(70)90542-2

[66] Volpe, P. and Eremenko, T. (1970) Quantitative Studies on Cell Proteins in Suspension Cultures. *European Journal of Biochemistry*, **12**, 195-200. http://dx.doi.org/10.1111/j.1432-1033.1970.tb00837.x

[67] Volpe, P. and Eremenko, T. (1973) A Method for Measuring Cell Cycle Phases in Suspension Cultures. *Methods in Cell Biology*, **6**, 113-126. http://dx.doi.org/10.1016/S0091-679X(08)60049-7

[68] Kappler, J.A. (1970) The Kinetics of DNA Methylation in Cultures of a Mouse Adrenal Cell Line. *Journal of Cell Physiology*, **75**, 21-34. http://dx.doi.org/10.1002/jcp.1040750104

[69] Evans, H.H., Evans, T.E. and Littman, S. (1973) Methylation of Parental and Progeny DNA Strands in *Physarum polycephalum*. *Journal of Molecular Biology*, **74**, 563-574. http://dx.doi.org/10.1016/0022-2836(73)90047-8

[70] Adams, R.L.P. and Hogarth, C. (1973) DNA Methylation in Isolated Nuclei: Old and New DNAs Are Methylated. *Biochimica et Biophysica Acta*, **331**, 214-220. http://dx.doi.org/10.1016/0005-2787(73)90434-6

[71] Adams, R.L.P. (1974) Newly Synthesized DNA Is Not Methylated. *Biochimica et Biophysica Acta*, **335**, 365-373. http://dx.doi.org/10.1016/0005-2787(74)90159-2

[72] Bird, A.P. (1978) Use of Restriction Enzymes to Study Eucaryotic DNA Methylation: The Symmetry of Methylated Sites Supports Semi-Conservative Copying of the Methylation Pattern. *Journal of Molecular Biology*, **118**, 49-60. http://dx.doi.org/10.1016/0022-2836(78)90243-7

[73] Georgyev, G.P. and Mantieva, V.I. (1962) The Isolation of DNA-Like RNA and Ribosomal RNA from the Nucleolo-Chromosomal Apparatus of Mammalian Cells. *Biochimica et Biophysica Acta*, **61**, 153-162.

[74] Penman, S., Vesco, C. and Penman, M. (1968) Localization and Kinetics of Formation of Nuclear Heterodisperse RNA, Cytoplasmic Heterodisperse RNA and Polyribosome Associated Messenger RNA in *HeLa* Cells. *Journal of Molecular Biology*, **34**, 49-62. http://dx.doi.org/10.1016/0022-2836(68)90234-9

[75] Georgyev, G.P. (1969) On the Structural Organization of Operon and on the Regulation of RNA Synthesis in Animal Cells. *Journal of Theoretical Biology*, **25**, 227-231. http://dx.doi.org/10.1016/S0022-5193(69)80034-2

[76] Darnell, J.E., Jelinek, W.R. and Molloy, G.R. (1973) Biogenesis of mRNA: Genetic Regulation in Mammalian Cells. *Science*, **181**, 1215-1221. http://dx.doi.org/10.1126/science.181.4106.1215

[77] Volpe, P. and Eremenko, T. (1973) Polysomes during the *HeLa* Cell Life Cycle. 9*th International Congress of Biochemistry*, Stockholm, Abstracts, 195.

[78] Breathnach, R., Mandel, J.L. and Chambon, P. (1977) Ovalbumin Gene Is Split in Chicken DNA. *Nature*, **270**, 314-

319. http://dx.doi.org/10.1038/270314a0

[79] Brack, C. and Tonegawa, S. (1977) Variable and Constant Parts of the Immunoglobulin Light Chain Gene of a Mouse Myeloma Cell Are 1250 Nontranslated Bases Apart. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5652-5656. http://dx.doi.org/10.1073/pnas.74.12.5652

[80] Doel, M.T., Houghton, M., Cook, E.A. and Carey, N.H. (1977) The Presence of Ovalbumin mRNA Coding Sequences in Multiple Restriction Fragments of Chicken DNA. *Nucleic Acids Research*, **4**, 3701-3713. http://dx.doi.org/10.1093/nar/4.11.3701

[81] Jeffreys, A.J. and Flavell, R.A. (1977) A Rabbit Beta-Globin Gene Contains a Large Insert in the Coding Sequence. *Cell*, **12**, 1097-1108. http://dx.doi.org/10.1016/0092-8674(77)90172-6

[82] Leder, P., Tilghman, S.M., Tiemeier, D.C., Polsky, F.I., Seidman, J.G., Edgell, M.H., Enquist, L.W., Leder, A. and Norman, B. (1977) Cloning of Mouse Globin and Surrounding Gene Sequences in *Bacteriophage lambda*. *Cold Spring Harbor Symposium on Quantitative Biology*, **42**, 915-920. http://dx.doi.org/10.1101/SQB.1978.042.01.093

[83] Berget, S.M., Moore, C. and Sharp, P.A. (1977) Spliced Segments at the 5' Terminus of Adenovirus 2 Late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 3171-3175. http://dx.doi.org/10.1073/pnas.74.8.3171

[84] Chow, L.T.R., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) An Amazing Sequence Arrangement at the 5' Ends of *Adenovirus* 2 Messenger RNA. *Cell*, **12**, 1-8. http://dx.doi.org/10.1016/0092-8674(77)90180-5

[85] Tilghman, S.M., Tiemeier, D.C., Seldman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V. and Leder, P. (1978) Intervening Sequence of DNA Identified in the Structural Portion of a Mouse Beta-Globin Gene. *Proceedings of the National Academy of Sciences of the United States of America*, **75**, 725-729. http://dx.doi.org/10.1073/pnas.75.2.725

[86] Gilbert, W. (1978) Why Genes in Pieces. *Nature*, **271**, 501. http://dx.doi.org/10.1038/271501a0

[87] Crick, F.H.C. (1979) Split Genes and RNA Splicing. *Science*, **204**, 264-271. http://dx.doi.org/10.1126/science.373120

[88] Maclean, M. and Hilder, V.A. (1977) Mechanisms of Chromatin Activation and Repression. *International Review of Cytology*, **48**, 54-97. http://dx.doi.org/10.1016/S0074-7696(08)61742-8

[89] Tentravahi, V., Guntaka, R.V., Erlanger, B.F. and Miller, O.J. (1981) Amplified Ribosomal RNA Genes in a Rat Hepathoma Cell Line Are Enriched in 5-Methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 489-493. http://dx.doi.org/10.1073/pnas.78.1.489

[90] Volpe, P. (1976) The Gene Expression during the Cell Life Cycle. *Horizons in Biochemistry and Biophysics*, **2**, 285-340.

[91] Eremenko, T., Menna, T. and Volpe, P. (1979) Differential Gene Expression during the Cell Life Cycle. *Macromolecules in the Functioning Cell*, **1**, 79-108. http://dx.doi.org/10.1007/978-1-4684-3465-1_7

[92] Eremenko, T. and Volpe, P. (1984) Virus Genome Integration as a Function of the Host Cell Genome Replicative Cycle. *FEBS Letters*, **169**, 211-214.

[93] Volpe, P. and Eremenko, T. (1984) Position of the Integration Site of SV40 Virus Genome at the Level of the Transcriptional Unit. *Voprosy Virusologii*, **5**, 610-619.

[94] Volpe, P. and Eremenko, T. (1984) Resumption of Transcription Following Inhibition of DNA Methylation. *Macromolecules in the Functioning Cell*, **3**, 59-69.

[95] Volpe, P. and Eremenko, T. (1985) The Gene Repair-Modification System in Eukaryotes. In: Ovcinnikov, Y.A., Ed., *Proceedings of the* 16*th FEBS Congress*, Vol. 3, Science Press, Utrecht, 123-129.

[96] Volpe, P. and Eremenko, T. (1986) Differential Evolution of Coding and Non-Coding Eukaryotic DNAs through Repair-Modification. *Macromolecules in the Functioning Cell*, **4**, 8-20.

[97] Eremenko, T., Esposito, C., Iacovacci, P., Benedetto, A., Elia, G., Delfini, C., Alfani, E., Cascio, O., Sarpietro, M.G., Fazzio, A. and Volpe, P. (1990) Is 5-Methylcytosine a Regulatory Signal in Eukaryotic Gene Expression? 7*th International Symposium on Macromolecules in the Functioning Cell*, Taormina, Abstracts, 22.

[98] Volpe, P., Esposito, C., Iacovacci, P., Butler, R.H. and Eremenko, T. (1993) Language of Genes with Inverse Correlation between Methylation and Transcription. *Macromolecules in the Functioning Cell*, **7**, 59-71.

[99] Volpe, P., Iacovacci, P., Butler, R.H. and Eremenko, T. (1993) 5-Methylcytosine in Genes with Methylation-Dependent Regulation. *FEBS Letters*, **329**, 233-237. http://dx.doi.org/10.1016/0014-5793(93)80228-M

[100] Scarano, E. (1969) Enzymatic Modifications of DNA and Embryonic Differentiation. *Annals in Embryology and Morphology*, Suppl. 1, 51-61.

[101] Volpe, P. and Eremenko, T. (1973) Nuclear and Cytoplasmic DNA Synthesis during the Mitotic Cycle of *HeLa* Cells. *European Journal of Biochemistry*, **32**, 227-232. http://dx.doi.org/10.1111/j.1432-1033.1973.tb02600.x

[102] Volpe, P. and Eremenko, T. (1978) A Language of DNA Modification during Replication. 14*th International Congress*

of Genetics, Moscow, Abstracts, 1, 194.

[103] Lima de Faria, A. (1969) Handbook of Molecular Cytology. Wiley, New York, 277-283.

[104] Comings, D.E. (1972) Methylation of Euchromatic and Heterochromatic DNA. *Experimental Cell Research*, **74**, 383-391. http://dx.doi.org/10.1016/0014-4827(72)90391-6

[105] Volpe, P., Granieri, A. and Eremenko, T. (1975) Negli eucarioti i geni si metilano ordinatamente durante la replica. 12*th Meeting of the Italian Society of Biophysics and Molecular Biology*, Pavia, Abstracts, 49.

[106] Eremenko, T., Granieri, A. and Volpe, P. (1978) Organization, Replication and Modification of the Human Genome: I. Differential Methylation of Two Classes of *HeLa* Nuclear DNA Separated on Ag^+/Cs_2SO_4 Gradients. *Molecular Biology Reports*, **4**, 163-170. http://dx.doi.org/10.1007/BF00777518

[107] Eremenko, T., Granieri, A. and Volpe, P. (1978) Organization, Replication and Modification of the Human Genome: II. Temporal Order of Synthesis and Methylation of Two Classes of *HeLa* nDNA Separated in Ag^+/Cs_2SO_4 Gradients. *Molecular Biology Reports*, **4**, 237-240. http://dx.doi.org/10.1007/BF00777561

[108] Eremenko, T., Timofeeva, M.Y. and Volpe, P. (1980) Organization, Replication and Modification of the Human Genome: Synthesis and Methylation of Palindromic, Repeated and Unique *Hela* nDNA Sequences during the S-Phase. *Molecular Biology Reports*, **6**, 131-136. http://dx.doi.org/10.1007/BF00775405

[109] Volpe, P. and Eremenko, T. (1982) Cell and Virus Macromolecular Modification during the Cell Life Cycle. *Macromolecules in the Functioning Cell*, **2**, 179-195.

[110] Volpe, P., Menna, T. and Eremenko, T. (1976) RNA Transcription and Processing as a Function of the Cell Cycle. *Bulletin of Molecular Biology and Medicine*, **1**, 18-28.

[111] Eremenko, T. and Volpe, P. (1975) Polysome Translational State during the Cell Cycle. *European Journal of Biochemistry*, **52**, 203-210. http://dx.doi.org/10.1111/j.1432-1033.1975.tb03988.x

[112] Delfini, C., Alfani, E., De Venezia, V., Oberholtzer, G., Tomasello, C., Eremenko, T. and Volpe, P. (1985) Cell-Cycle Dependence and Properties of the *HeLa* Cell DNA Polymerase System. *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 2220-2224. http://dx.doi.org/10.1073/pnas.82.8.2220

[113] Battacharya, S.K., Ramchandani, S., Cervoni, N. and Szyf, M. (1999) A Mammalian Protein with Specific Demethylase Activity for mCpG DNA. *Nature*, **397**, 579-583. http://dx.doi.org/10.1038/17533

[114] Volpe, P. and Eremenko, T. (1994) Complete and Incomplete Repair of Radiodamaged Gene Sequences. *UNESCO Technical Reports*, **19**, 27-43.

[115] Kruger, D., Schroeder, C., Reuter, M., Bogdarina, I., Buryanov, Y. and Bickle, T. (1985) DNA Methylation of Bacterial Viruses T3 and T7 by Different DNA Methylases in *Escherichia coli* K12 Cells. *European Journal of Biochemistry*, **150**, 323-330. http://dx.doi.org/10.1111/j.1432-1033.1985.tb09024.x

[116] Mandel, J.L. and Chambon, P. (1979) DNA Methylation: Organ Specific Variations in the Methylation Pattern within and Oround Ovalbumin and Other Chicken Genes. *Nucleic Acids Research*, **7**, 2081-2092. http://dx.doi.org/10.1093/nar/7.8.2081

[117] Sutter, D. and Doerfler, W. (1980) Methylation of Integrated Adenovirus Type 12 DNA Sequences in Trasformed Cells Is Inversely Correlated with Viral Gene Expression. *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 253-256. http://dx.doi.org/10.1073/pnas.77.1.253

[118] Liau, M.C., Chang, C.F., Saunders, G.F. and Tsai, Y.H. (1981) S-Adenosylhomocysteine Hydrolases as the Primary Target Enzymes in Androgen Regulation of Methylated Complexes. *Archives of Biochemistry and Biophysics*, **208**, 261-272. http://dx.doi.org/10.1016/0003-9861(81)90148-X

[119] Geraci, D., Eremenko, T., Granieri, A., Scarano, E. and Volpe, P. (1973) The *in Vivo* and *in Vitro* Methylation of DNA in Synchronized *HeLa* Cells. 9*th International Congress of Biochemistry*, Stockholm, Abstracts, 3m64, 190.

[120] Razin, A. and Riggs, A.D. (1980) DNA Methylation and Gene Function. *Science*, **210**, 604-610. http://dx.doi.org/10.1126/science.6254144

[121] Eden, S. and Cedar, H. (1994) Role of DNA Methylation in the Regulation of Transcription. *Current Opinions in Genetics and Development*, **4**, 255-259. http://dx.doi.org/10.1016/S0959-437X(05)80052-8

[122] Razin, A. (1998) CpG Methylation, Chromatin Structure and Gene Silencing—A Three-Way Connection. *EMBO Journal*, **17**, 4905-4908. http://dx.doi.org/10.1093/emboj/17.17.4905

[123] Bird, A.P. and Wolffe, A.P. (1999) Methylation-Induced Repression: Belts, Braces and Chromatin. *Cell*, **99**, 451-454. http://dx.doi.org/10.1016/S0092-8674(00)81532-9

[124] Volpe, P., Menna, T. and Eremenko, T. (1978) Supermethylation of Eukaryotic nDNA after Oncovirus Genome Integration. 11*th European Tumor Virus Meeting*, Balatonfured, Abstracts, 4, 194.

[125] Eremenko, T. and Volpe, P. (1984) *In Vitro* Methylation of Total and Foldback DNAs in Normal and Virus-Trans-

formed Cells. *FEBS Letters*, **173**, 233-237.

[126] Grippo, P., Iaccarino, M., Parisi, E. and Scarano, E. (1968) Methylation of DNA in Developing Sea Urchin Embryos. *Journal of Molecular Biology*, **36**, 195-208. http://dx.doi.org/10.1016/0022-2836(68)90375-6

[127] Doskocil, J. and Sormova, Z. (1965) The Occurrence of 5-Methylcytosine in Bacterial Deoxyribonucleic Acid. *Biochimica et Biophysica Acta*, **95**, 513-515. http://dx.doi.org/10.1016/0005-2787(65)90199-1

[128] Eremenko, T., Granieri, A., Scarano, E. and Volpe, P. (1975) Fractionation of Methylated Nuclear DNA in *HeLa* Cells. 10*th FEBS Meeting*, Paris, Abstracts, 130.

[129] Augusti-Tocco, G., Carestia, C., Grippo, P., Parisi, E. and Scarano, E. (1968) Thin-Layer Electrophoresis of Nucleic Acid Derivatives on Microcrystalline Cellulose. *Biochimca et Biophysica Acta*, **155**, 8-18. http://dx.doi.org/10.1016/0005-2787(68)90329-8

[130] Eremenko, T. and Volpe, P. (1976) Replica semiconservativa del DNA metilato. 13*th Meeting of the Italian Society of Biophysics and Molecular Biology*, Albano Laziale, Abstracts, 23.

[131] Lu, S., Saydak, M., Gentile, V., Steins, J.P. and Davies, J.A. (1995) Isolation and Characterization of the Human Tissue Transglutaminase Gene Promoter. *Journal of Biological Chemistry*, **270**, 9748-9756. http://dx.doi.org/10.1074/jbc.270.17.9748

[132] Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) High Sensitivity Mapping of Methylated Cytosines. *Nucleic Acids Research*, **22**, 2990-2997. http://dx.doi.org/10.1093/nar/22.15.2990

[133] Frommer, M., McDonald, L.E., Millar, D.S., Collins, C., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (1992) A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 1827-1831. http://dx.doi.org/10.1073/pnas.89.5.1827

[134] Clark, S.J., Harrison, J. and Frommer, M. (1995) CpNpG Methylation in Mammalian Cells. *Nature Genetics*, **10**, 20-27. http://dx.doi.org/10.1038/ng0595-20

[135] Lu, S. and Davies, P.J.A. (1997) Regulation of the Expression of the Tissue Transglutaminase Gene by DNA Methylation. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 4692-4697. http://dx.doi.org/10.1073/pnas.94.9.4692

[136] Cacciamani, T., Virgili, S., Centurelli, M., Bertoli, E., Eremenko, T. and Volpe, P. (2002) Specific Methylation of the CpG-Rich Domains in the Promoter of the Human Tissue Transglutaminase Gene. *Gene*, **297**, 103-112. http://dx.doi.org/10.1016/S0378-1119(02)00874-0

[137] Razin, A. and Cedar, H. (1977) Distribution of 5-Methylcytosine in Chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 2725-2728. http://dx.doi.org/10.1073/pnas.74.7.2725

[138] Volpe, P. (2005) The Language of Methylation in Genomics of Eukaryotes. *Biochemistry* (*Moscow*), **70**, 584-595. http://dx.doi.org/10.1007/s10541-005-0152-x

[139] Meehan, R., Antequera, F., Lewis, J., McLeod, D., McKay, S., Kleiner, E. and Bird, A.P. (1990) A Nuclear Protein That Binds Preferentially to Methylated DNA *in Vitro* May Play a Role in the Inaccessibility of Methylated CpGs in Mammalian Nuclei. *Philosophical Transactions of the Royal Society of London*, **B326**, 199-205. http://dx.doi.org/10.1098/rstb.1990.0004

[140] Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A.P. (1998) Transcriptional Repression by the Methyl-CpG-Binding Protein MeCP2 Involves a Histone Deacetylase Complex. *Nature*, **393**, 386-389. http://dx.doi.org/10.1038/30764

[141] Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F. and Wolffe, A.P. (1999) Mi-2 Complex Couples DNA Methylation to Chromatin Remodelling and Histone Deacetylation. *Nature Genetics*, **23**, 62-66. http://dx.doi.org/10.1038/12664

[142] Volpe, P. and Eremenko, T. (1995) Repair-Modification of Radiodamaged Genes. *Radiation Protection Dosimetry*, **62**, 19-22.

[143] Volpe, P., Parasassi, T., Sapora, O., Ravagnan, G. and Eremenko, T. (1999) Influence of the Low Doses of Radiation on the DNA Double Helix, Gene Expression and Membranal State. *International Journal of Radiation Medicine*, **1**, 78-89.

[144] Volpe, P. (2001) The Gene "Repair-Modification" Theory in Biophysics of Radiations. *International Journal of Radiation Medicine*, **3**, 123-137.

[145] Marmur, J. (1961) A Procedure for the Isolation of Deoxiribonucleic Acid from Microorganisms. *Journal of Molecular Biology*, **3**, 208-218. http://dx.doi.org/10.1016/S0022-2836(61)80047-8

[146] Volpe, P., Iacovacci, P., Esposito, C. and Eremenko, T. (1992) A Common Language in Eukaryotic Genes with Methylation-Dependent Regulation. *Rendiconti Lincei*, **3**, 383-394.

[147] Volpe, P. (1996) Dall'origine del codice all'evoluzione del gene eucariotico (*Lectio Doctoralis* in Occasion of the *Laurea Honoris Causa* in *Biological Sciences with Biotechnological Specialization*). Università degli Studi, Ancona, 1-35.

[148] Volpe, P. (1999) Introduzione alla Biofisica delle Radiazioni. UNESCO Publishers, Venice, 1-256.

[149] Volpe, P. (2014) A Theoretical Overview of Bioresponse to Magnetic Fields on the Earth's Surface. *International Journal of Geosciences*, **5**, 1149-1162. http://dx.doi.org/10.4236/ijg.2014.510097

[150] Kruger, D.H., Shroeder, C., Reuter, M., Bogdarina, I.G., Buryanov, Y.I. and Bickle, T.A. (1985) DNA Methylation of Bacterial Viruses T3 and T7 by Different DNA Methylases in *Escherichia coli* K12 Cells. *European Journal of Biochemistry*, **150**, 323-330. http://dx.doi.org/10.1111/j.1432-1033.1985.tb09024.x

[151] Gorelick, R. (2003) Transposable Elements Suppress Recombination in All Meiotic Eukaryotes, Including Automictic Ancient Asexuals: A Reply to Schön and Martens. *Journal of Natural History*, **37**, 903-909. http://dx.doi.org/10.1080/0022293021000007705

[152] Gorelick, R. (2005) Theory for Why Dioecious Plants Have Equal Length Sex Chromosomes. *American Journal of Botany*, **92**, 979-984. http://dx.doi.org/10.3732/ajb.92.6.979

[153] Lingua, G., Trotta, A., Prigione, V., Ugoccioni, R. and Berta, G. (2005) Nucleus Size in the Host Cells of an Arbuscular Mycorrhizal System: A Mathematical Approach to Estimate the Role of Ploidy and Chromatin Condensation. *Karyologia*, **58**, 112-121.

[154] Zemskov, E.A., Janiak, A., Hang, J., Waghray, A. and Belkin, A.M. (2006) The Role of Tissue Transglutaminase in Cell-Matrix Interactions. *Frontiers in Bioscience*, **11**, 1057-1076. http://dx.doi.org/10.2741/1863

[155] Lai, T.S., Liu, Y., Li, W. and Greenberg, C.S. (2007) Identification of Two GTP-Independent Alternatively Spliced Forms of Tissue Transglutaminase in Human Leukocytes, Vascular Smooth Muscle, and Endothelial Cells. *FASEB Journal*, **21**, 4131-4143. http://dx.doi.org/10.1096/fj.06-7598com

[156] Malousi, A., Maglaveras, N. and Kouidou, S. (2008) Intronic CpG Content and Alternative Splicing in Human Genes Containing a Single Cassette Exon. *Epigenetics*, **3**, 69-73. http://dx.doi.org/10.4161/epi.3.2.6066

[157] Łukasik, M., Karmalska, J., Szutowski, M.M. and Łukaszkiewicz, J. (2009) Wplyw metylacji DNA na funkcjonowanie genomu. *Biuletyn Wydziału Farmaceutycznego Warszawskiego Uniwersytetu Medycznego*, **2**, 13-18.

[158] Shleikin, A.G. and Danilov, N.P. (2011) Evolutionary-Biological Peculiarities of Transglutaminase. Structure, Physiological Functions, Application. *Journal of Evolutionary Biochemistry and Physiology*, **47**, 1-14. http://dx.doi.org/10.1134/S0022093011010014

[159] Lytvynenko, O.O. and Bugaicov, S.G. (2013) Ionizing Radiation in the Process of Initiation and Development of Cancer Pathology. *Archiv Klinichnoj ta Eksperimentalnoj Medizini*, **22**, 120-123.

[160] Lytvynenko, O. and Bugaitsov, S. (2013) Ionizing Radiation in Origin and Development Oncology Pathology. *Actual Problems of Transport Medicine*, **4**, 18-23.

Abbreviations

N^6-methyladenine (m^6A), 5-methylcytosine (m^5C), deoxyadenylate (dam), deoxycytidylate (dcm), methionine (Met), [^{14}C]methyl-L-methionine ([^{14}C]Met), S-adenosyl-L-methionine (AdoMet), bromo-deoxyuridine (BrdUrid), DNA polymerase (DNApol), RNA polymerase (RNApol), methyltransferase (met), transglutaminase (hTGc), housekeeping (HK) and tissuespecific (TS) genes, calcitonin (CALC), mitochondrial DNA (mtDNA), culture growth cycle (CGC), Simian virus 40 (SV40), ultraviolet (UV), repair patch (RP).