1. Introduction

1.1. DNA

Deoxyribonucleic acid (DNA) is a central molecule for life. The blueprint for every living being from the tiniest bacterium to the biggest redwood tree, from the simplest prokaryote to the most complex human neuron is written in the language of the four nucleobases dC, dG, dT and dA, which are the central building blocks of DNA. Although DNA was originally discovered by *Miescher* in 1869, it took another 50 years until its components, namely ribose, phosphate and the four DNA bases were deciphered.[1-3] Only since the ground-breaking work of *Avery* in 1944 and *Hershey* in 1952, it is proven that DNA actually is the carrier of heritable genetic information.[4-6] Shortly thereafter the structure of DNA was deciphered by *Watson* and *Crick* with major contributions by *Franklin*, *Gosling* and *Wilkins*.[7-10] DNA is a polymer with ribose units that are connected by phosphodiester groups. The nucleobases are attached to the sugars and form specific hydrogen bond mediated base pairs with each other: deoxycytidine (dC) pairs with deoxyguanosine (dG) and thymidine (dT) pairs with deoxyadenosin (dA) (*Figure 1*). These base pairing properties are the basis of DNA replication.[11] For the synthesis of proteins DNA is transcribed into messenger ribonucleic acid (mRNA), which is subsequently translated into proteins in the ribosomes. A triplet of three base pairs encodes one amino acid.[12]

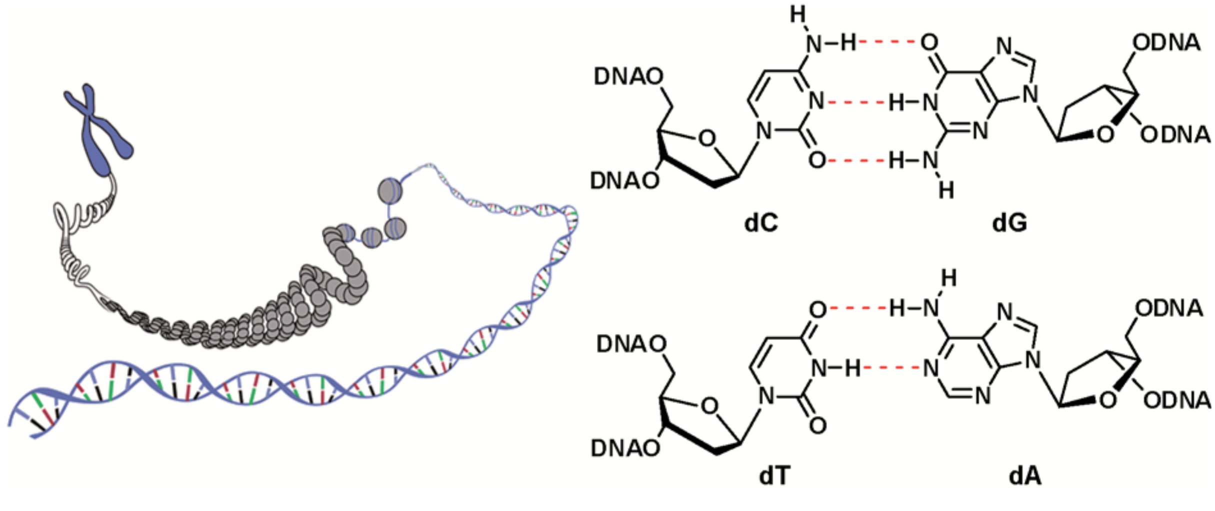

Figure 1: Specific hydrogen bond mediated base pairing in a DNA double strand. dC forms three hydrogen bonds with dG, dT forms two hydrogen bonds with dA. Red: hydrogen bonds.

Two pairing DNA strands adopt a double helical assembly with ten base pairs per turn. In natural B-DNA (further observed folds: A-DNA and Z-DNA) this gives rise to two areas where the nucleobases are exposed to the solvent: The broad major groove and the narrow minor grove (see *Figure 2*).[13] These areas allow specific interactions and are therefore the binding site for important proteins like transcription factors, but can also be exploited for targeting DNA with small molecules like for example dyes (Hoechst 33342 or 33258).[14]

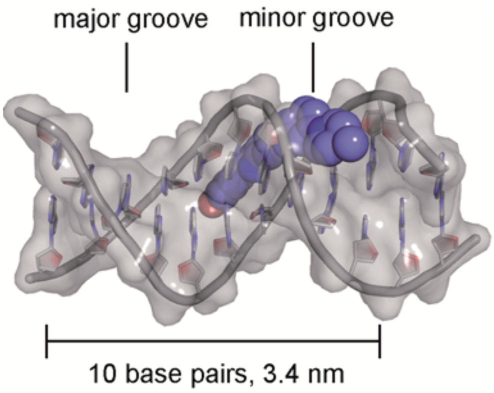

Figure 2: DNA in complex with the Hoechst 33258 dye, which selectively binds into the minor groove. PDB code: 264d

1.2. DNA within a Eukaryotic Cell

In eukaryotic cells, the DNA is located in the nucleus. Eukaryotic genomes consist of several independent strands of DNA, the chromosomes. In these, the DNA is closely associated with special DNA binding proteins. This protein-DNA assembly is called chromatin. The major functions of chromatin are packaging of the DNA in a smaller volume and the control of gene expression. The general organization of chromatin is shown in *Figure 3*.

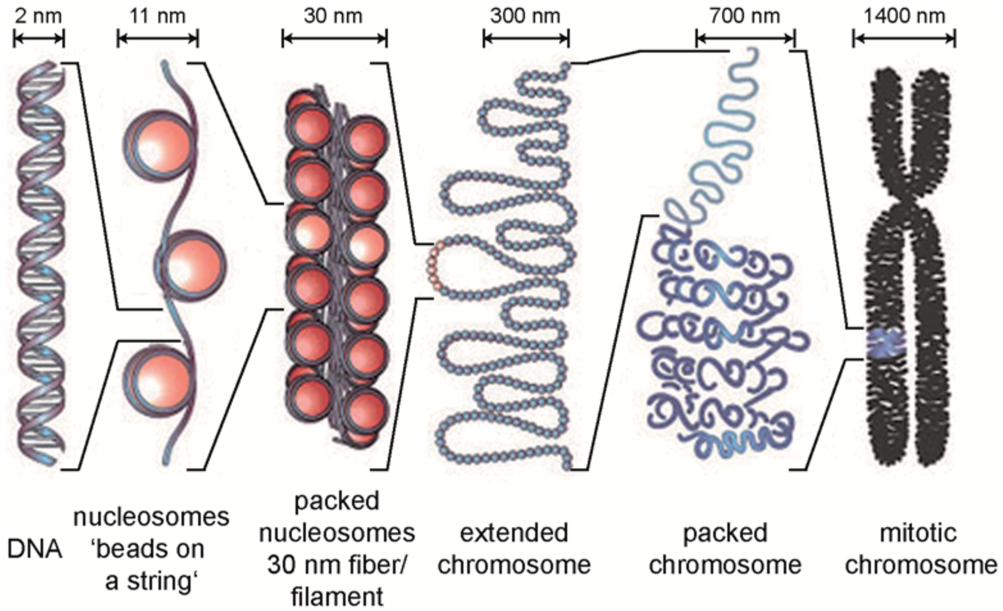

Figure 3: Organization of DNA in eukaryotic cells. DNA is wrapped around nucleosomes, which are further organized into the 30 nm fiber. The densest packing is observed in mitotic chromosomes. Based on Figure 1 in reference [15] with permission from Macmillan Publishers Ltd.

The DNA strand is wrapped around an assembly of eight basic proteins, the core histones (H2A, H2B, H3 and H4). This results in the formation of the nucleosome (*Figure 4*), which is the basic repeat unit of chromatin.[15-16] Major interactions occur between positively charged lysine residues and the negatively charged phosphodiester backbone of the DNA. Without linker histones (H1) the nucleosomes are only very loosely associated and form the 'beads on a string' structure which can be observed in an electron microscope.[16-17] The 'beads on a string' packaging is usually associated with active transcription of the corresponding stretch of DNA. The histones play an important role in the control of gene silencing and activation (see below). The linker histone H1 binds between the nucleosomes, which results in a coiling of the 'beads on a string' structure into a higher order assembly, called the 'filament' or '30 nm fiber'. The filament is highly dynamic and unfolds when RNA polymerase binds. This level of chromatin structure is considered the active state of chromatin, also called euchromatin. The 30 nm fiber is further organized in defined loops. It is believed that multiple folding topologies are possible.[18-19] In non-replicating cells, different structures result in regions where chromatin is densely packed and inactivated (heterochromatin) and others where the packaging is rather loose (euchromatin).[20] The most dense packaging of DNA is observed in mitotic chromosomes, which can be detected in a microscope even at low magnifications.

1.3. Epigenetics

In a multicellular organism, like for example humans, every cell contains the same genetic material. Cells do, however, perform vastly different functions. This is only possible, if different genes are active in different cell types. Every cell of the body furthermore originates from the same zygote. The genetic program therefore has to be modified during the differentiation of the cells. Together, these processes require precise control of gene expression which is established by both stable and dynamic gene silencing. The base sequence alone cannot control these processes, which are therefore *epi-genetic (*Greek: *epi-* over, above*)*. In a modern definition, the research field of epigenetics deals with "*the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states*".[21] The characteristic of these changes is that they are heritable during mitosis. In other words: *changes in the behavior of a cell persist through multiple cell generations without the change of the DNA sequence*. There is evidence for the heredity of epigenetic information during meiosis but the issue is still controversially discussed.[21-22] Chemically there are two main types of epigenetic signals: the covalent modification of DNA and the covalent modification of chromatin proteins. DNA modification is the more permanent of the two mechanisms and in terms of stability one can consider DNA modification as the epigenetic long term memory, while histone modifications are the epigenetic short term memory. Additionally, chromatin remodeling, the (active) repositioning of nucleosomes, is a third epigenetic mechanism. All epigenetic control mechanisms seem to influence each other, but the underlying mechanisms have not been fully understood, yet.[23-25]

Histone tails (N-termini) can be modified by methylation, acetylation, phosphorylation, ubiquitylation, sumoylation and ADP-ribosylation (*Figure 4*).[26-28] Methylation and acetylation are the best studied modifications and generally histone methylation seems to be a signal for gene inactivation (H3K9, H3K27), whereas acetylation is associated with transcriptional activation (H3K9, H3K14), but there are many exceptions (most notable: H3K4me3). Histone modifications are dynamic, and proteins that modify histones (like histone methyltranferases) as well as enzymes that remove histone modifications (like histone demethylases) are known. Despite intensive research in the field the histone code is far from being fully understood.[29]

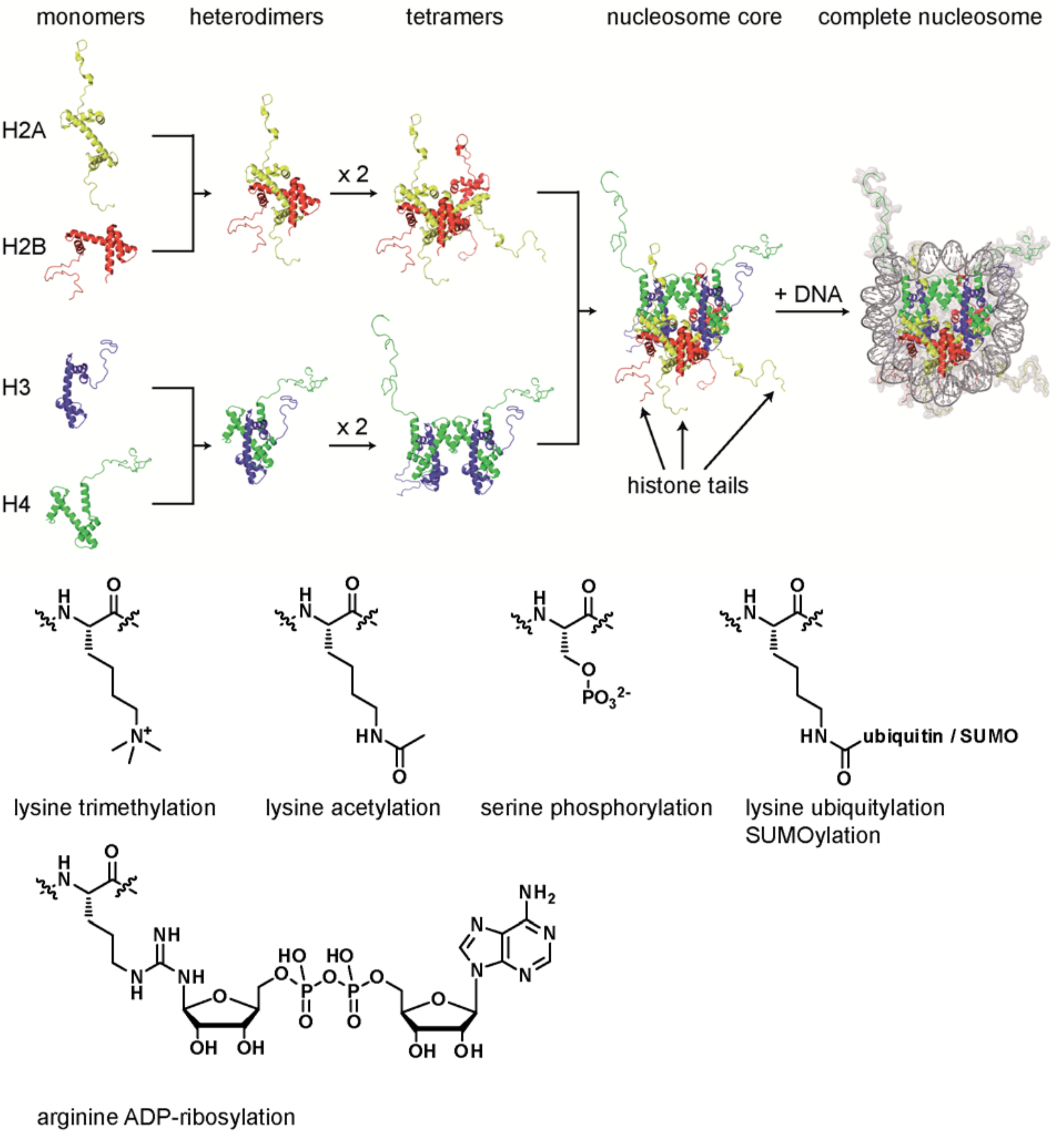

Figure 4: Composition of a nucleosome. 4 different histone proteins (two copies each) form the octameric nucleosome core, around a double strand with 147 base pairs is wrapped. The N-termini (tails) of the histones can be modified. A selection of possible modifications is shown. PDB code: 1kx5

1.4. DNA Methylation

When this thesis was started only one modification of DNA in mammals was known: dC can be post-replicatively derivatized with a methyl group to give 5-methyl-deoxycytidine ($^{5\text{-Me}}$dC).[30-31] The methyl group is located in the major groove and therefore influences DNA-protein contacts. In somatic cells methylation of cytosines nearly exclusively occurs in CpG dinucleotide sequences. Interestingly, this motif is underrepresented in the genome and only 1 % instead of the expected 4.4 % of all dinucleotides have this sequence. CpG units often cluster together, thereby forming the so called CpG islands.[32-34] These are predominately found in promoter regions, where they play an important role in the control of gene expression.[35] Methylation outside the CpG motif has been found in embryonic stem cells, but its function is currently unclear.[36] In mammals, 4 % to 6 % of all cytosines and consequently 70 % - 80 % of all CpG units are methylated.[30, 37]

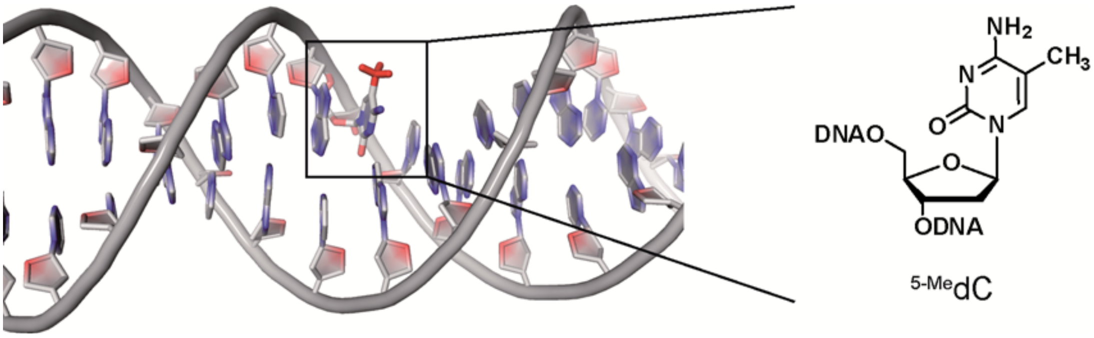

Figure 5: $^{5\text{-Me}}$dC is the most common DNA modification in eukaryotes. Its methyl group is directed into the major groove of DNA.

The nucleoside $^{5\text{-Me}}$dC is formed post-replicatively by addition of a methyl group to the C(5) position of dC.[30] This reaction is catalyzed by the enzyme family of the methyltransferases.[38-39] In humans, three different active DNA-methyltransferases have been detected: Dnmt1, Dnmt3A and Dnmt3B. Based on structural similarities it was believed that Dnmt2 is also a DNA methylating enzyme, but recently it has been shown that it does not modify DNA but tRNA.[40] Dnmt1 methylates the unmethylated strand of a hemimethylated duplex, as it occurs after replication of the genome. The enzyme is associated with the replication fork and is responsible for copying the 'methylome' after cell division. It is therefore referred to as the maintenance methyltransferase.[41-42] The crystal structure of Dnmt1has recently been solved.[43] As shown in *Figure 6A* a CXXC domain (red) and an autoinhibitory loop (yellow) prevent unmethylated DNA to enter the active site, which accounts for its specificity for hemimethylated DNA. Dnmt3A and Dnmt3B, in contrast, are also capable of methylating unmodified DNA and are therefore considered to be the *de novo* methyltransferases.

All methyltransferases share the same mechanism, which requires *S*-adenosylmethionine (SAM) as a cofactor. In order to render the cytosine accessible, the enzyme flips it out of the duplex, as proven by the crystal structure of the bacterial methyltransferase M.Hha1 (*Figure 6B*).[44] In the first step of the catalytic cycle a conserved cystein residue attacks the electrophilic C(6) position of the cytosine (*Figure 6C*). This reaction is facilitated by protonation of N(3) by a conserved glutamate or aspartate. In the next step the C(4) - C(5) double bond attacks the methyl group of SAM and displaces the positively charged sulfur atom in an S_N-type reaction. Based on the geometry observed in the crystal structure the methyl group resides in an *anti* position relative to the cystein. In the last step of the catalytic cycle the nucleobase rearomatizes via a *syn*-elimination of cystein.

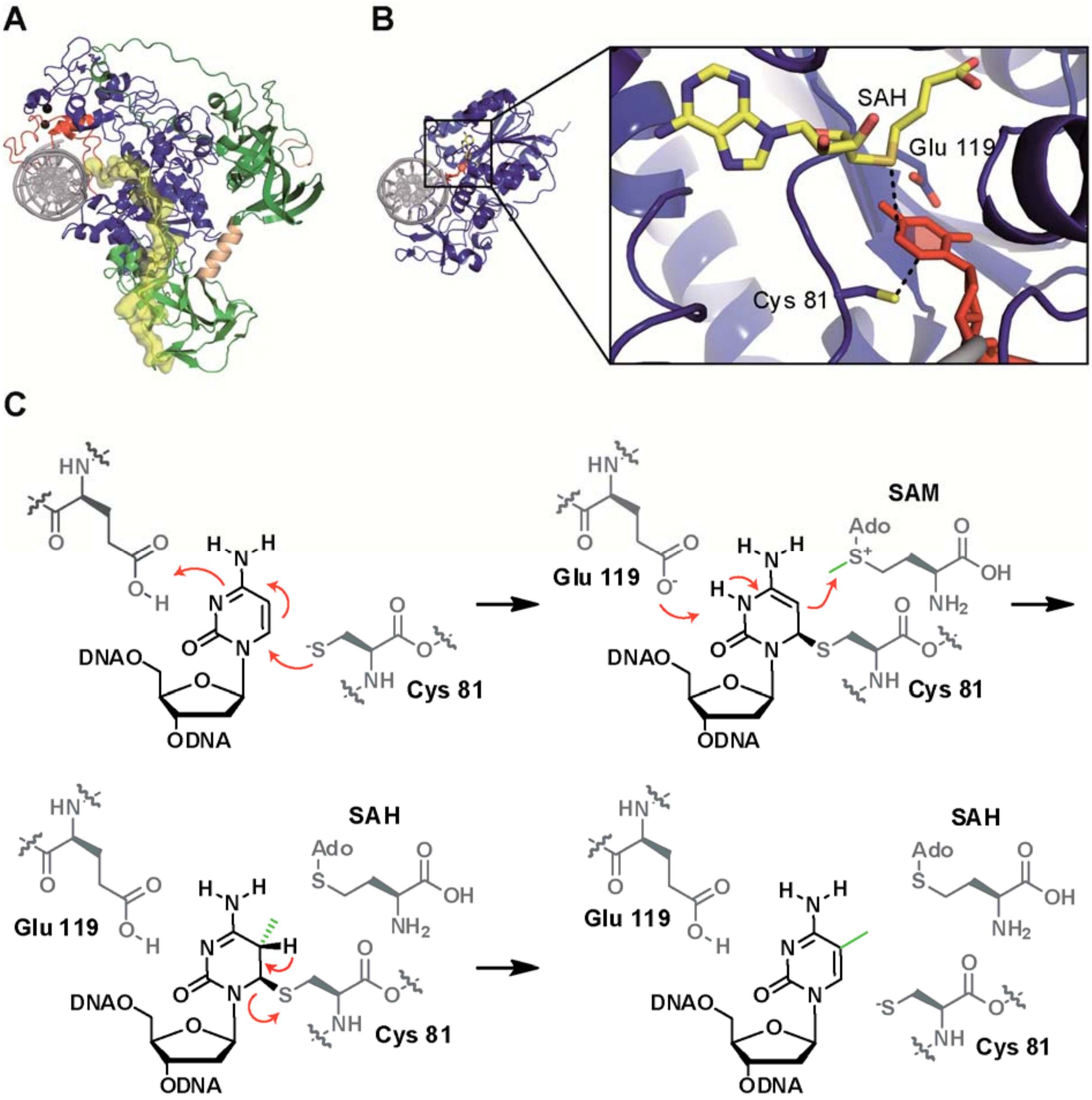

Figure 6: A) Crystal structure of mouse Dnmt1. Blue: catalytic subunit; red: CXXC domain; green bromo-domains; yellow: autoinhibitory loop; PDB code: 3PT6. B) Crystal structure of M.Hha1 methyltransferase. Blue: catalytic domain, red: dC, yellow: SAH; PDB code: 3MHT. C) Catalytic mechanism of methyltransferases. Example: M.Hha1, SAM: *S*-adenosylmethionine, SAH: *S*-adenosylhomocystein.

1.4.1. Consequences of DNA Methylation

Methylation of cytosines in promoter sequences is strongly associated with transcriptional repression. Most probably, methylation is a secondary mechanism that permanently silences already repressed genes.[45] The presence of the methyl group in the major groove alters the interaction patterns of the genome with the proteome.[31] This is often mediated by proteins that contain a special methyl-CpG binding domain (MBD).[46] The most prominent member of this family is methyl-CpG binding protein 2 (MeCP2). The MBD-proteins can either simply block the DNA or recruit other proteins like histone modifying enzymes, which provides a connection between the different epigenetic signals.[24, 47] Inversely, histone modifications can also target methyltransferases to DNA and therewith induce DNA methylation.[48-49] Furthermore, $^{5\text{-Me}}$dC can directly modulate the binding of transcription factors or chromatin modifying enzymes to DNA, which leads to the inhibition of transcription.[50-51] Through silencing of genes DNA methylation has important functions in X-chromosome inactivation,[31] imprinted genes,[52] embryonic development[45] and the occurrence of certain diseases.[24]

1.4.2. The Role of $^{5\text{-Me}}$dC in Cellular Differentiation

As stated before, all cells of a multicellular organism originate from the same cell, the zygote. The zygote and the early blastomers are totipotent - they can differentiate to any tissue of the embryo and the placenta. In the blastocyst there are already two different pluripotent cell types: first the inner cell mass, which forms the three germ cell layers of the embryo, and second the trophectoderm which forms the placenta. Embryonic stem cells are inner-cell mass explants and retain pluripotency when they are cultured under suitable conditions.[45]

DNA-methylation is crucial for mammalian development. Dnmt1 knockout mice die until embryonic day 9, Dnmt3a and Dnmt3b knockout mice are viable but show developmental defects and die latest 1 month after birth. During differentiation the epigenetic landscape changes dramatically. Directly after fertilization the epigenetic marks of the paternal genome rapidly change. First it is loaded on histones, and the protamines that packed the DNA in the sperm cell are replaced. Furthermore, the signal for $^{5\text{-Me}}$dC decreases, which for a long time was believed to be the result of global demethylation.[53-54] Recent findings, however, suggest that this demethylation actually is a hydroxymethylation.[55-58] During early cell divisions also the maternal genome is passively demethylated, which leads to a hypomethylation of the inner cell mass of the blastocyst. Cultured embryonic stem cells surprisingly show the normal extent of methylation.[59] Although their global amount of $^{5\text{-Me}}$dC is the same as in somatic cells, the methylation pattern in the genome of ES cells differs dramatically.[60] Not surprisingly, the promoters of the key pluripotency factors like *nanog* and *oct4* are unmethylated in the pluripotent state and thereby expressed. Promoters of genes, that are associated with tissue specific genes are mostly methylated.[61] During differentiation genes like *oct4* are repressed and especially key promoters are methylated in a tissue specific manner.[45] Methylation is probably used to suppress alternative differentiation pathways by