



Biology of Chromatin

Contents

- 1.1 Introduction: Epigenetic Regulation in the Context of the Genome – 2**
 - 1.1.1 Background: Gene Expression and Chromatin – 2
 - 1.1.2 Discovery of the Nucleosomal Structure of the Genome – 3
- 1.2 The Structure of the Nucleosome – 3**
 - 1.2.1 Histone Variants – 4
- 1.3 Histone Modifications – 7**
 - 1.3.1 Nomenclature for Histone Modifications – 7
 - 1.3.2 Combinatorial Modifications at Pericentric Heterochromatin – 8
 - 1.3.3 Histone Modifications at High Resolution – 9
 - 1.3.4 Chromatin Modifications Associated with Transcription Units – 12
 - 1.3.5 A Concept of Writers, Readers, and Erasers of Histone Modifications – 13
- 1.4 DNA Modifications – 14**
 - 1.4.1 DNA Cytosine Methylation – 15
 - 1.4.2 DNA Cytosine Hydroxymethylation – 20
 - 1.4.3 Interaction of DNA and Histone Modifications – 22
- 1.5 Chromatin Organization and Compartmentalization in the Cell Nucleus – 22**
 - 1.5.1 Replication of Pericentric Heterochromatin Domains – 24
 - 1.5.2 Topologically Associating Domains – 24
 - 1.5.3 Structural Maintenance of Chromosomes Complexes – 26
- References – 28**

What You Will Learn in This Chapter

This chapter provides an introduction to chromatin. We will examine the organization of the genome into a nucleosomal structure. DNA is wrapped around a globular complex of 8 core histone proteins, two of each histone H2A, H2B, H3, and H4. This nucleosomal arrangement is the context in which information can be established along the sequence of the DNA for regulating different aspects of the chromosome, including transcription, DNA replication and repair processes, recombination, kinetochore function, and telomere function. Posttranslational modifications of histone proteins and modifications of DNA bases underlie chromatin-based epigenetic regulation. Enzymes that catalyze histone modifications are considered writers. Conceptually, erasers remove these modifications, and readers are proteins binding these modifications and can target specific functions. On a larger scale, the 3-dimensional (3D) organization of chromatin in the nucleus also contributes to gene regulation. Whereas chromosomes are condensed during mitosis and segregated during cell division, they occupy discrete volumes called chromosome territories during interphase. Looping or folding of DNA can bring regulatory elements including enhancers close to gene promoters. Recent techniques facilitate understanding of 3D contacts at high resolution. Lastly, chromatin is dynamic and changes in histone occupancy, histone modifications, and accessibility of DNA contribute to epigenetic regulation.

1.1 Introduction: Epigenetic Regulation in the Context of the Genome

1.1.1 Background: Gene Expression and Chromatin

All organisms inherit traits from their parents, which are encoded in the succession of four bases in nucleic acids. All eukaryotic organisms possess deoxyribonucleic acid (DNA)-based genomes, whereby DNA comprises of antiparallel strands wound in a right-handed double helix. Although the sequence of bases as well as the 3D structure of the DNA helix contribute to the expression of traits, it is thought that the DNA sequence facilitates trait generation by the regulated expression of genes. Gene expression is not the only function of the genome, but replication and faithful inheritance of the genome to descendant somatic cells and the next generation, and evolutionary modifications of the genome are principal functions of the heritable material. Importantly, a genome is not sufficient for generating an organism; this requires a suitable reader that can be represented by a cell or, in the case of higher organisms, an oocyte or zygote that must be from the same species as the genome. The reader implements molecular processes in the cell's nucleus that lead to the production of biomolecules – ribonucleic acids and other biosynthetically active molecules – that replenish the cells and maintain organismal tissues. Through feedback by transcription factors that are encoded in the genome and typically bind to specific DNA sequences, complex regulatory circuits are established. In addition, the DNA itself is organized in a chromatin fiber that facilitates the imposition of information along the DNA sequence. Chromatin also supports the transduction of this epigenetic information into regulatory processes that, in turn, affect transcription, replication, repair, and, in specialized cases, even changes of the DNA sequence. Reciprocal feedback of transcription, combinatorial activity of regulators, genome size, and

temporal dynamics contribute to the complexity of gene regulatory networks. On the one hand, this has led to the evolution of developmental programs for complex body plans and, on the other hand, often makes the understanding of individual processes difficult. Therefore, exploration of mechanism of chromatin regulation relies on well suited model systems that disambiguate the function of the components involved. Genomic imprinting and X chromosome inactivation are phenomena where expressed and repressed copies of individual genes are present in the same nucleoplasm and facilitate the study of chromatin-based regulation of transcription factor activity. In addition, ingenious approaches have been designed to analyze chromatin-based heritable expression states controlled by *Polycomb* (PcG) and *Trithorax* group (TrxG) complexes. This chapter contains an introduction to the organization and function of chromatin and provides a basis for understanding its role in regulating the cellular function of DNA sequences.

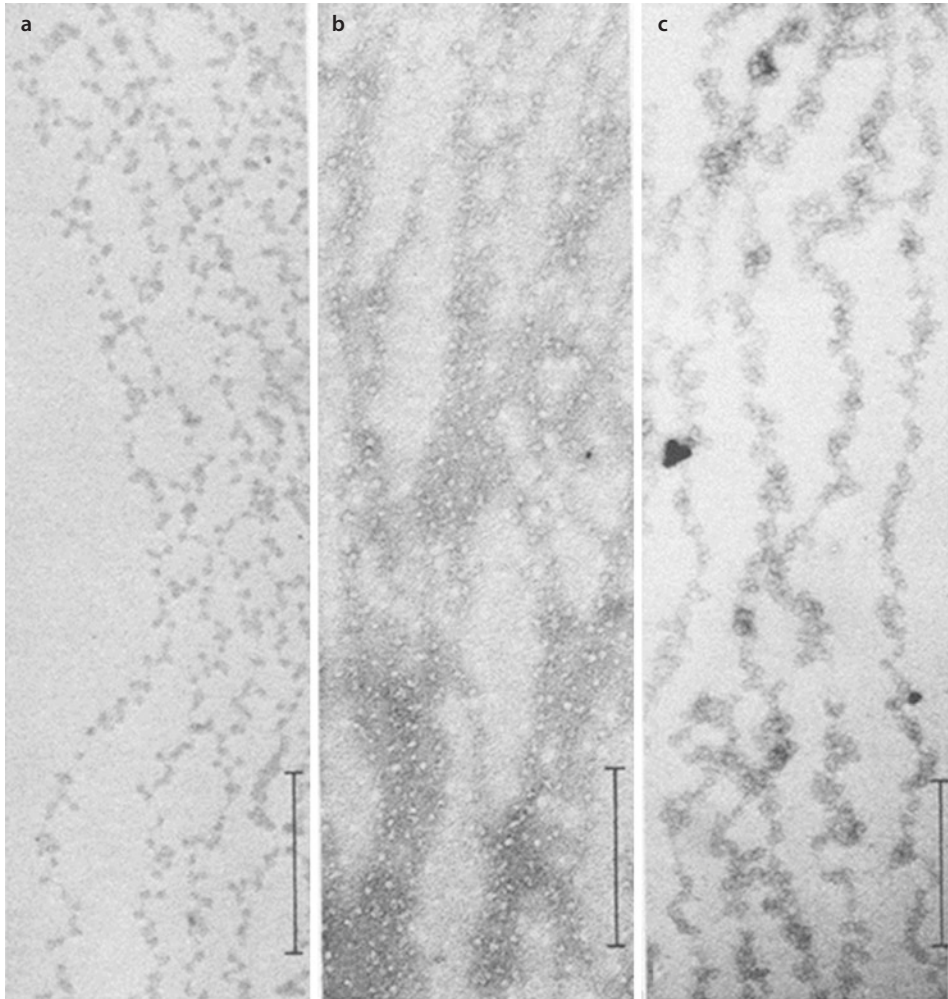
1.1.2 Discovery of the Nucleosomal Structure of the Genome

Linear arrays of spherical particles of about 70 Ångström in diameter were initially observed by electron microscopy of chromatin released from animal cells (■ Fig. 1.1) (Olins and Olins 1974). The regular spacing of the spherical units and the fact that similar arrangements were found in many eukaryotes suggested a basic form of organization of the genome. Consistent with a regular structure, experiments using limited digestion of chromatin with nucleases and gel electrophoresis analysis revealed DNA fragments at regular intervals at multiples of 150 base pairs (bp). This is consistent with protection of around 150 bp of DNA from nuclease digestion, whereby nuclease cleavage occurs on the stretch of DNA that lies between spherical particles. Based on these studies, it has become clear that an understanding of the genome needs to consider the molecular components of nucleosomal structure.

1.2 The Structure of the Nucleosome

A nucleosome represents a single repeat unit for organizing the majority of the DNA of a cell.¹ A nucleosome consists of 8 histone proteins and 146 bp of DNA, which is wrapped around them in two left-handed turns² (■ Fig. 1.2a–c). Histone proteins form an octamer complex that comprises positively charged surfaces formed by basic amino acid side chains, which interact with the negatively charged phosphate groups of the DNA backbone. The octamer is assembled from two of each histone H3/H4 and histone H2A/H2B dimers (■ Fig. 1.2d). The histone H3/H4 dimers occupy the core of the nucleosome and the H2A/H2B dimers are more loosely associated

-
- 1 An exception to nucleosomal organization of eukaryotic genomes is found in male gametes of animals, where histones are largely replaced with protamine facilitating a distinct and highly compacted genome configuration.
 - 2 Although it is clear that the large majority of DNA in cells is arranged in left-handed turns around the histone octamer, it has been suggested that right-handed DNA looping can be accommodated by the nucleosomal structure and has experimentally been produced *in vitro*.



■ **Fig. 1.1** Electronmicrographs of eukaryotic chromatin. Rat thymus chromatin **a** positive and **b** negative staining, and **c** chicken erythrocyte chromatin. (From Olins and Olins (1974))

(■ **Fig. 1.3**). Histone proteins form a globular domain with a characteristic alpha-helical arrangement called the histone fold (Luger et al. 1997). Flexible N-terminal regions of histone proteins, the so-called histone tails, associate more loosely with the nucleosome and are accessible for posttranslational modifications.

1.2.1 Histone Variants

Although the nucleosomal structure appears homogenous, variation in histone composition can introduce different functionalities. Different variants of histones can be incorporated, whereby variations of histone H3 and histone H2A are common, and histones H2B and H4 appear to be predominantly canonical.

1.2 · The Structure of the Nucleosome

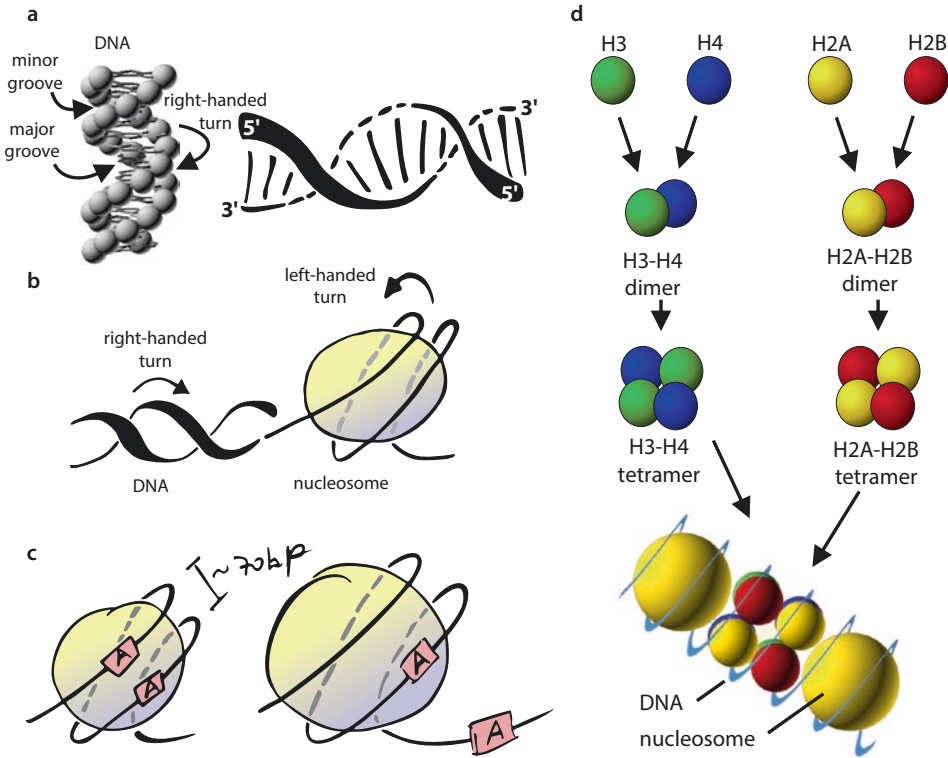


Fig. 1.2 Scheme representing biologically relevant aspects of the nucleosome structure. **a** DNA forms a right-handed double helix with two strands aligned in opposite (antiparallel) orientation. The helix forms a major and minor groove, whereby DNA bases are more accessible from the major groove. **b** Winding of the DNA in left-handed turns over the histone octamer. **c** At a distance of approximately 70 bp, the two DNA loops face the same surface of the nucleosome. Two hypothetical factor binding-sites are indicated by the boxed “A”. As one turn of the DNA is completed every 10 bp, both “A” boxes can be facing the surface through the major groove, depending on the position of the nucleosome. Nucleosome positioning along the DNA thereby allows for changes in binding-site geometry that can have a regulatory function. **d** A model for assembly of nucleosomes from histone dimers. DNA assembled into a nucleosomal structure is depicted at the bottom

Histone variants are encoded by a different set of genes and show slight variations in the amino acid sequence. In mammals, histone H3.1 and H3.2 are deposited during DNA replication in S-phase, whereas histone H3.3 is involved in histone exchange at active transcription units and at pericentric and telomeric heterochromatin. Incorporation of histone H3.3 outside of S-phase can be explained as it is the only histone H3 gene whose expression is not restricted to S-phase. Histones are very basic proteins and are associated with chaperones, or loaders, in the cell when they are not in a chromatin context (see book ► Chap. 2 of Paro). Different histone loaders are involved in deposition of histone H3 variants. During S-phase, chromatin assembly factor 1 (CAF-1) incorporates histones H3.1 and H3.2 into DNA, whereas *histone cell cycle regulator* (HIRA) has a role in incorporating H3.3 into transcribed

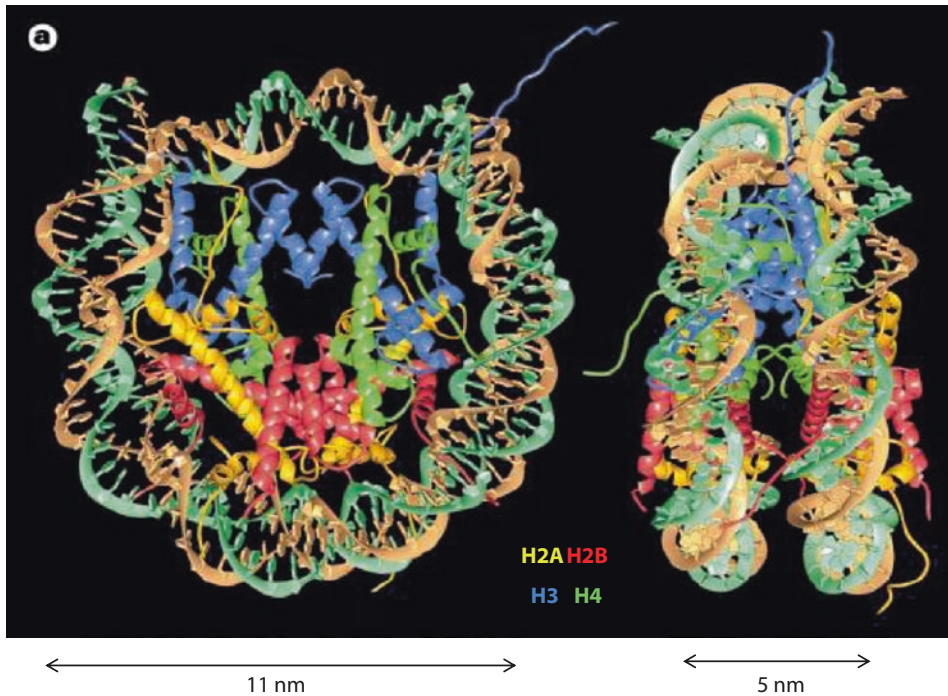


Fig. 1.3 A model based on X-ray crystallographic structure determination reveals details of a nucleosome. DNA is wrapped around a protein core of 8 histone proteins (histone H3/H4 dimers, green and blue; histone H2A/H2B dimers, red and yellow). Globular domains of histone proteins show helical folds and long, unstructured N-terminal peptides are pasted into the model as protruding from the core. (From Luger et al. (1997))

genes. At pericentric³ and telomeric heterochromatin, ATRX/DAXX is required for histone H3.3 incorporation. At the centromere of mammalian cells, canonical histone H3 is replaced by CENP-A, which contains a number of amino acid differences and adopts a more compact structure. It is thought that CENP-A contributes to the mechanical rigidity of centromeric chromatin that forms the basis of the kinetochore, where spindle microtubules attach for chromosome segregation in mitosis. The specific loader JHURP is involved in CENP-A incorporation into centromeres.

Among histone H2A genes in mammals, the incorporation of the H2A.z variant is restricted to promoters of active genes. This is somehow surprising as H2A.z appears to represent the ancestral form of H2A and is the mammalian H2A gene that is most similar to the H2A gene of *Saccharomyces cerevisiae*. Therefore, the canonical H2A, which is found in the majority of mammalian chromatin, has a slightly different amino acid sequence to the ancestral histone. Additional variants of histone H2A have been correlated with gene activity. Whereas H2A.B is incorporated into chromatin over the transcription unit of active genes, macroH2A accumulates in

³ Pericentric chromatin lies next to the centromere and is normally in a repressed heterochromatic configuration.

silent chromatin. MacroH2A is a vertebrate-specific histone H2A variant that contains a large C-terminal extension, which is commonly referred to as a macro-domain. MacroH2A is enriched at the inactive X chromosome of female mammals, consistent with its association with transcriptionally repressed chromatin (see book ► Chap. 4 of Wutz). Through variation of the composition of histone proteins, the function of nucleosomes can be changed, which affects the function of the 146 bp long DNA associated with the nucleosome.

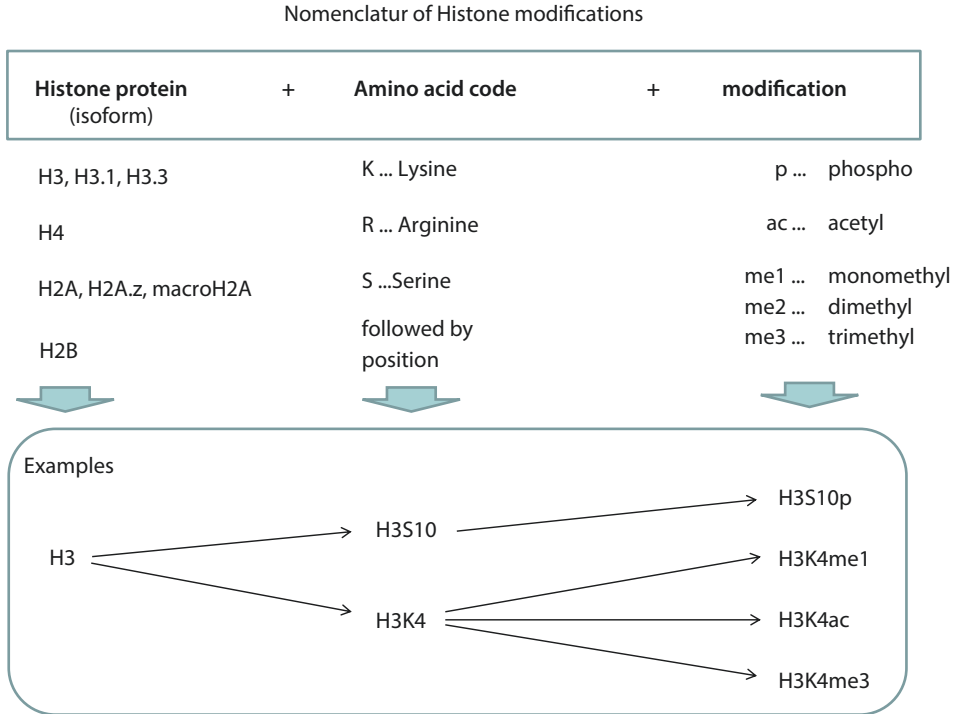
1.3 Histone Modifications

Changes in histone composition are not the only way by which information can be added to nucleosomes. Histone proteins are the subject to posttranslational modifications. In particular, the unstructured N-termini of histones can be modified in a variety of ways, whereby the chemical spectrum of modifications and the combinatorial complexity is high. Phosphorylation of serine, acetylation and methylation of lysine, and methylation of arginine residues are the most prominent histone modifications. Notably, multiple methyl-groups can be added to lysines and arginines, thereby increasing the complexity. Transfer of a ubiquitin to lysines in histone H2A and H2B has also been described. Improvements of analytical techniques have revealed an increasing number of histone modifications and it is likely that this trend will continue in the future. Posttranslational modifications of histones and their functions are incompletely understood at present. However, the development of specific antisera to detect modified histones has provided key insights.

1.3.1 Nomenclature for Histone Modifications

The chemical diversity of modifications and the large number of acceptor sites on histone proteins makes it difficult to describe the modification state of nucleosomes. To facilitate the precise and comprehensive documentation of experimental results and further theoretical elaborations, the scientific community has adopted guidelines for a systematic nomenclature. The class of the histone protein is prefixed to the single character code and position⁴ of the amino acid carrying the modification, followed by an abbreviation of the chemical nature of the modification (► Fig. 1.4). For example, the short form for histone H3 carrying di-methylation of lysine 27 is H3K27me₂. Similarly, H2AK119ub identifies histone H2A carrying a mono-ubiquitin modification on lysine 119. If the precise histone subtype is known this can also be incorporated, for example H3.3K4ac stands for histone H3.3 that is acetylated on lysine at position 4. This versatile nomenclature can easily be extended for multiple modifications: H3K9me₃S10p specifies a doubly modified histone H3 with tri-methylation on lysine 9 and phosphorylation on serine 10.

4 Numbering of amino acids starts at the biochemical N-terminus, which is different from the methionine encoded by the start ATG of the histone gene due to the activity of methionine aminopetidase.



■ Fig. 1.4 A schematic of short-form nomenclature for posttranslational histone modifications

1.3.2 Combinatorial Modifications at Pericentric Heterochromatin

Histone modifications can be detected by suitable antibodies either using biochemical methods or by staining techniques followed by microscopy. The latter has been used in combination with fluorescence-labeled secondary antibody detection to analyze different histone modification states within the nucleus. Microscopy techniques for dual- or multi-color labeling have further facilitated the observation of colocalization of specific protein complexes with histone modifications. These approaches have been important to characterize different types of chromatin in the nucleus. When combined with fluorescent probes that hybridize with specific DNA sequences, an assessment of the genomic context can be made, albeit at a modest resolution of several thousand bp and, hence, tens or hundreds of nucleosomes.

The pericentric repeats in mouse cells are observed as DNA-dense clusters that are brightly stained with DNA binding dyes such as DAPI.⁵ Association of H3K9me3 with this DAPI-dense region suggested a correlation of a specific histone methylation mark with pericentric heterochromatin. Binding of H3K9me3 by the chromo-domain of heterochromatin protein 1 (HP1) leads to HP1 accumulation at the pericentric

⁵ DAPI (4',6-diamidino-2-phenylindole) is a fluorescent dye that binds to DNA, allowing the detection of the relative density of DNA in the nucleus by fluorescence microscopy.

heterochromatin. The assay of co-localization with pericentric heterochromatin has been exploited to identify additional components. The curious and instructive anecdote of the human autoimmune serum MCA1 provides a concise understanding. Initial interest in the MCA1 antiserum was raised when it seemed to recognize centromeres in mitosis and, thus, suggested specificity for a potential component of the kinetochore. Since MCA1 did not identify any known kinetochore-associated proteins, mass spectrometry was performed, leading to identification of none but histone proteins. Finally, the mystery was resolved when the specificity of MCA1 was investigated on a panel of doubly modified histone H3 peptides (Hirota et al. 2005). It turned out that MCA1 specifically recognizes H3K9me3S10p, which occurs exclusively at pericentric heterochromatin in mitosis, when serine 10 is phosphorylated by cell cycle kinases, including polo-like kinase (PLK). The physiological effect of serine 10 phosphorylation is to displace HP1 from pericentric heterochromatin in mitosis. This shows that the interaction between HP1 and histone H3 is specific for the H3K9me3 state and is disrupted by doubly modified H3K9me3S10p. This result is important for two reasons. Firstly, it clearly demonstrates that combinations of histone modifications can act in cellular regulation. Secondly, in hindsight, it became clear that antisera for specific histone modifications are sensitive to modifications on neighboring amino acids. This has implications for the interpretation of results obtained with such immunoreagents as a failure to detect a signal does not always correlate with the absence of the modification but can also be caused by interference with a neighboring modification. Although it is assumed that in the majority of cases interference from neighboring modifications can be neglected, there is no systematic study to ascertain the validity of this assumption.

Research on pericentric heterochromatin has also led to the identification of histone methyltransferases. Suppressor of position variegation 3-9 homologues 1 and 2 (Suvar3-9h1 and Suvar3-9h2) associate with pericentric chromatin and catalyze H3K9me3 (■ Fig. 1.5). H3K9me3 acts as a binding site for HP1 and mediates the recruitment of the additional histone methyltransferases Suvar4-20h1 and Suvar4-20h2 which, in turn, catalyze H4K20me3 at pericentric regions (Schotta et al. 2004). This example shows that modifications from histone H3 can be propagated onto histone H4 within the same chromatin context. The observation that multiple histones and modifications contribute to pericentric chromatin suggests a mechanism for the stability of heterochromatin.

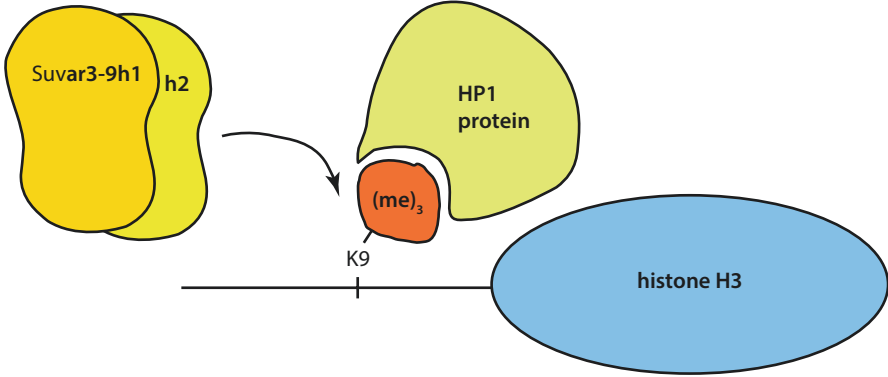
1.3.3 Histone Modifications at High Resolution

Practically, it is of interest to investigate which modifications are present on nucleosomes at a given gene promoter. Knowing of the activity of genes, this can further identify correlations with histone modifications that are generally associated with active or repressed chromatin. For this analysis, two prerequisites need to be fulfilled. Firstly, suitable and specific reagents are required to detect a particular modification of a histone. This has been facilitated by the development of a large array of antibodies that specifically recognize histone modifications. Secondly, the methodology needs to establish a link between modified histones and the DNA sequence, with which these are associated. Chromatin immunoprecipitation (ChIP) does exactly allow for such an analysis (► Method Box 1.1). Chromatin enriched for a defined

Combinatorial interaction of histone modifications

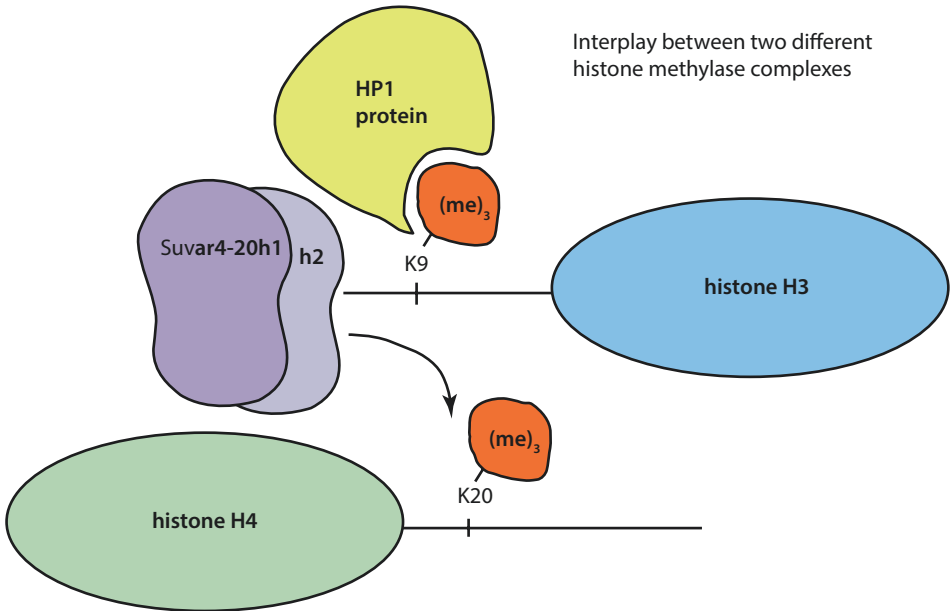
1

The HMTs Suvar3-9h1 and Suvar3-9h2 catalyse tri-methylation of histone H3 lysine 9 in pericentric heterochromatin



tri-methylation of histone H3 lysine 9 is a signal for recruiting HP1 through its chromo-domain in pericentric heterochromatin

Interplay between two different histone methylase complexes



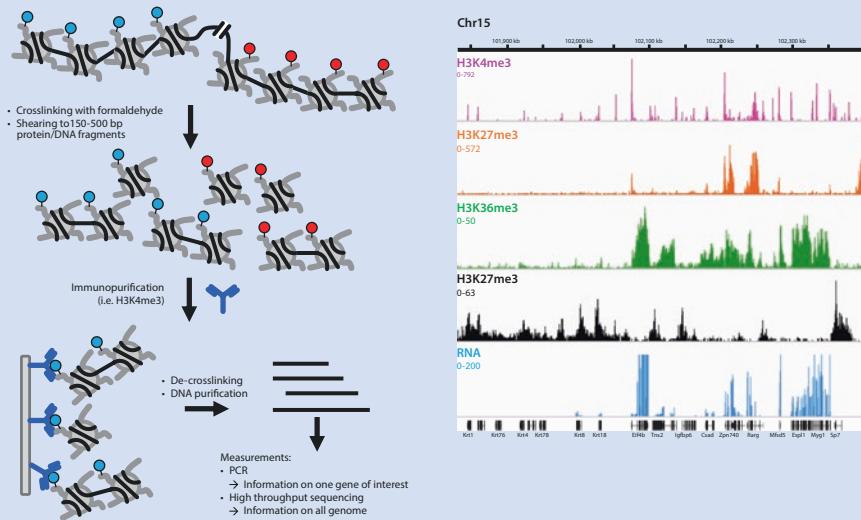
The HMTs Suvar4-20h1 and Suvar4-20h2 are recruited and catalyse tri-methylation of histone H4 lysine 20

Fig. 1.5 A schematic of histone marks at pericentric heterochromatin. The histone methyltransferases Suv3-9h1 and Suv3-9h2 establish H3K9me3. The tri-methylated lysine 9 serves as a binding signal for HP1 to histone H3. HP1 binding and recruitment of the histone-methyltransferases Suv4-20h1 and Suv4-20h2 leads to establishment of H4K20me1 (Schotta et al. 2004)

histone modification is isolated through immunoprecipitation using antibodies that recognize specific histone modifications. The DNA can then be purified and analyzed by a variety of methods including hybridization to microarrays (ChIP-array), adapter ligation and sequencing (ChIPseq), PCR with gene-specific primers, or by hybridization on dot blots. The latter has originally been used to investigate highly repeated parts of the genome, such as the pericentric repeats, whilst the former can provide genome-wide maps of histone modifications at nucleosomal resolution.

Method Box 1.1: Chromatin Immunoprecipitation

Histone modifications, and transcription factor binding, can be analysed at high resolution by chromatin immunoprecipitation (ChIP). For ChIP analysis, the DNA is cross-linked to histone proteins using formaldehyde, when chromatin is in its native form within cells. Subsequently, cells are lysed and chromatin is sheared either by ultrasound treatment or partial digestion with nucleases. Chromatin that is associated with a specific histone modification or protein is enriched by immunoprecipitation with a specific antibody. After purification, the chromatin associated DNA can be analysed by dot blot, gene specific PCR, and next generation sequencing (Box Fig. 1.1).

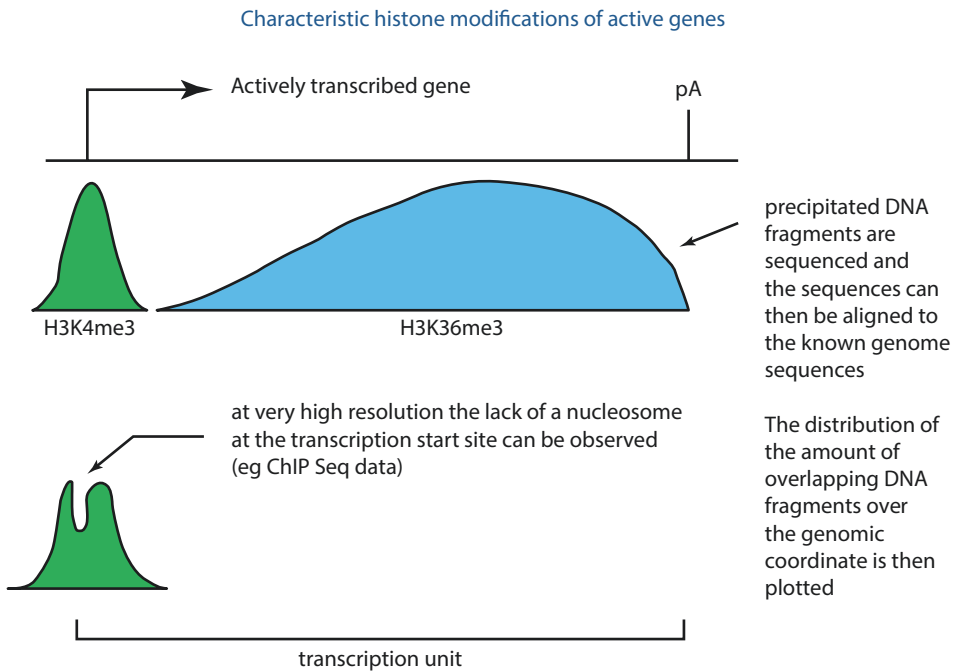


Box Fig. 1.1 Analysis of chromatin modifications by ChIP. A schematic shows an overview of the ChIP analysis (left) and an example view from a genome browser of a histone modification map obtained by next generation sequencing (right)

1.3.4 Chromatin Modifications Associated with Transcription Units

From analyses in multiple cell types, correlation of gene activity and histone modifications along transcription units has been performed. Some general rules can be deduced. Gene promoters are frequently marked by H3K4me3 and carry acetylated lysines on histone H4 and H3. The gene body of active genes is often marked by H3K36me3. The occurrence of H3K4me3 at the promoter and H3K36me3 over the gene body is a reliable indicator for transcription and has been used to identify new genes (■ Fig. 1.6) whose transcripts would have been difficult to detect due to their low abundance (Guttman et al. 2009).

In contrast, genes that are characterized by H3K9me2 or H3K9me3 tend to be transcriptionally inactive. Similarly, H3K27me3 and H2AK119ub are associated with the activity of PcG complexes that are well-known repressors of developmentally regulated genes (see book ► Chap. 3 of Paro). In progenitor cells and embryonic stem cells (see book ► Chap. 7 of Paro), these marks can also co-occur with the active mark H3K4me3. This observation has led to the discovery of bivalent chromatin at promoters with apparently active H3K4me3 and repressive H3K27me3 modifications. It appears that this configuration is resolved when cells differentiate and yields active H4K4me3 or repressed H3K27me3 configurations in separate cell



■ **Fig. 1.6** Illustration of ChIPseq analysis of the transcription unit of a hypothetical active gene. A peak of H4K4me3 enrichment overlaps the gene promoter whereas the transcription unit shows increasing H3K36me3 towards the 3' end of the gene. If data is obtained at very high resolution, a gap in the H4K4me3 peak can be observed at the position of the transcription start site (TSS). At TSS one nucleosome is displaced and the DNA is bound by general transcription factors

lineages. Therefore, PcG complexes might pre-mark certain developmentally regulated genes before an expression state is committed, which is consistent with the developmental plasticity of progenitor cells (see book ► Chaps. 3 and 7 of Paro).

1.3.5 A Concept of Writers, Readers, and Erasers of Histone Modifications

Chemical reactions that lead to posttranslational modifications of proteins use co-factors from metabolic pathways. Acetyl-Co-enzyme A, S-adenosyl-methionine, and ATP are used for the acetylation, methylation, and phosphorylation of histones, respectively (see book ► Chap. 9 of Santoro). The corresponding enzymatic activities are referred to as histone acetyltransferases (HATs), histone methylases (HMTs), and histone kinases. Analogously, histone ubiquitin-transferases catalyze mono-ubiquitylation. Posttranslational modifications of proteins are frequent and can have different functions. Not all modifications do possess an apparent physiological function and might be observed, to varying extents, as likely bystander reactions that the cell could not prevent. In a number of cases, highly specific modifications have been selected during evolution as signals that can have profound effects on the function of chromatin. Several histone modifications have been implicated in the transcription of genes. Other modifications act to establish a heritable signal for repression. Highly specific functions have been identified for particular methylation states of lysines in the N-termini of histone H3, whereas acetylation can be less specific. Acetylation of lysines of histones is part of the chromatin assembly process and is thought to remove the positive charge from lysines, thereby preventing strong interactions with negatively charged phosphate groups of the DNA. Acetylation of lysines of histone H3 and H4 is also associated with active promoters and could similarly loosen the association of histones with DNA to facilitate accessibility. These observations suggest a charge effect of acetylation and biophysical mechanisms involved in regulating local chromatin accessibility. However, this is not the entire story.

Bromodomain⁶-containing proteins can bind to acetylated lysines and attract other regulators. Proteins that specifically bind modified histones are considered readers of histone modifications. This group of proteins is particularly important for recognizing specific methylations states of lysines in the N-terminus of histone H3. HP1 protein can associate with H3K9me3 through a chromodomain. A single interaction provides only little binding energy given the small size of the tri-methyl-modification. However, HP1 can bind through a cooperative binding mechanism that enhances the binding of several HP1 proteins to longer stretches of H3K9me3 modified chromatin. This explains the enrichment of HP1 on the pericentric regions of mouse chromosomes, where H3K9me3 is abundant. Similar to HP1, the Polycomb (Pc) protein contains a chromodomain which is specific for H3K27me3 (see book ► Chap. 3 of Paro). The relevance of chromodomains in cells has been confirmed by replacing the chromodomain of Pc by the one from HP1, whereby the protein was

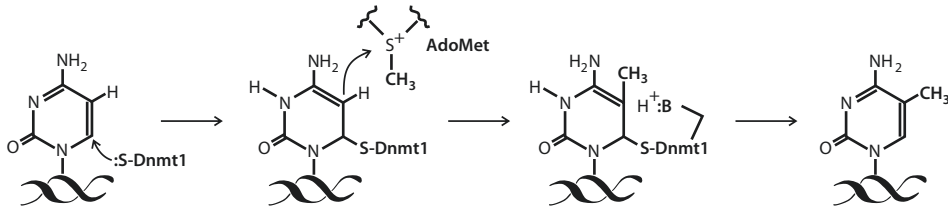
6 Bromo- and chromodomains are defined based on sequence homology in a number of chromatin-associated proteins. They recognize posttranslational modifications and sequence variation of histones in the context of the nucleosome.

redirected to pericentric chromatin. Similar protein domains exist for reading phosphorylated serines on histones. Importantly, 14-3-3 proteins have high specificity for the doubly modified histone H3S10pK14ac. This specificity plays a role in gene activation. A switch from an inactive H3K9me3 and HP1-bound state to an active H3K9me3S10pK14ac state has been described during the activation of the cell cycle inhibitor p21, which is an important tumor suppressor gene (see ► Chap. 8 of Santoro). Histone H3 serine 10 phosphorylation by ERK kinase activity interferes with HP1 binding to H3K9me3 and simultaneously facilitates the association of 14-3-3 proteins if lysine 14 is acetylated. This mechanism demonstrates how cell signaling, in combination with combinatorial histone modifications, can contribute to complex gene regulatory mechanisms.

A last aspect of histone modifications is their stability. Depending on their chemical nature, histone modifications possess different lifetimes. Whereas phosphorylation is readily reversible through phosphatases, tri-methylated lysine modifications can persist for extended periods. Histone deacetylases (HDACs) and histone demethylases (HDMs, or KDMs for lysine demethylases) remove acetyl- and methyl-groups from lysines, respectively. In addition, deubiquitinating enzymes contribute to the turnover of ubiquitin moieties. To date, mechanisms for the enzymatic removal of all histone modifications have been described, suggesting that posttranslational modifications of chromatin are dynamic and actively regulated by the cell. Proteins or complexes that remove histone modifications are considered erasers of epigenetic information. Whereas some histone modifications have been selected by evolution for a regulatory function other may be less important. Examples for the function of histone modifications will be discussed in their physiological context in the following chapters of this book.

1.4 DNA Modifications

Modifications of histones illustrate how information can be added to the genome without changing the sequence of the DNA. This is exactly how we defined “epigenetics” in the opening of the book. Enzymatic activities for establishing and removing modifications also illustrate how this information can be dynamically regulated during development or in response to external stimuli. However, we also had one expectation that is less easily explained: How can the potentially complex patterns of histone modifications be transmitted through cell division? Upon replication of the DNA, twice the number of nucleosomes will need to be assembled and this necessitates that half of the histones are freshly produced whereas the other half keeps the previously established modification patterns (see book ► Chap. 2 of Paro). How can the information be reestablished on the new histones? Or is it not restored but lost? There are good indications that epigenetic modifications are maintained but the mechanisms appear to be complex and are poorly understood. To understand the problem of epigenetic heritability, we turn to a much simpler system where maintenance is mediated by a concise mechanism.



■ **Fig. 1.7** Schematic of the catalytic mechanism of DNA methylation. DNMTs attach to the 6 position of the pyrimidine ring of the cytosine and enter a covalent intermediate. Methylation of the 5 position induces a shift of electrons and releases the enzyme

1.4.1 DNA Cytosine Methylation

The cytosine base of DNA in animal and plant genomes can be chemically modified to 5-methyl-cytosine (5mC). This modification is observed in the majority of animals but is conspicuously absent from some popular laboratory model organisms including the nematode *Caenorhabditis elegans*, the fly *Drosophila melanogaster*, and the yeasts *S. cerevisiae* and *Schizosaccharomyces pombe*. Although 5mC is not ubiquitous, it is widely distributed among animals⁷ and plants. In particular, DNA methylation is essential for mammalian development and has been extensively studied for its role in silencing tumor suppressor genes in certain types of human cancer (see book ► Chap. 8 by Santoro).

5mC is catalyzed by the activity of DNA methyltransferases (DNMTs) that use S-adenosyl-methionine (SAM) as a methyl-donor. The catalytic center of mammalian DNMTs resembles the one of DNA methylases that are components of bacterial restriction systems. The reaction involves an attack at the 6 position of the cytosine ring by the thiol group of a cysteine, leading to the formation of a covalent bond between the cytosine and the DNMT. The reaction mechanism of DNMTs thereby comprises an intermediate that links DNMTs temporarily to DNA (■ Fig. 1.7). This intermediate is subsequently resolved by transfer of a methyl-group from SAM to the 5 position of the cytosine ring, abstraction of a proton, and release of the DNMT enzyme. After release, DNMT enzymes are available for another reaction cycle. Derivatization of the ring system of a DNA base requires that the base is accessible. From structural analysis of bacterial DNA methylases, it has been proposed that the cytosine is flipped out from the DNA double helix by rotation of the phospho-deoxyribose backbone. In this way, the base can be inserted into a deep pocket where SAM and catalytic residues are in close contact to the 5 and 6 positions of the cytosine ring.

A variety of methods has been developed to analyze DNA methylation at specific genes and genome-wide (► Method Box 1.2). 5mC in mammalian genomes is preferentially found in the context of CG dinucleotides, but not on cytosines in other sequence contexts like CC, CA, and CT. The reason for this becomes clear when replication of the DNA and the maintenance of methylation patterns are considered. CG dinucleotides represent a symmetric configuration on the antiparallel DNA strand as C pairs with G and DNA strands have opposite polarity in the double helix.

⁷ DNA methylation has been observed in honey bees and ants suggesting that it is present in insects. The DNA methylation system was likely lost during the evolution of *D. melanogaster*.

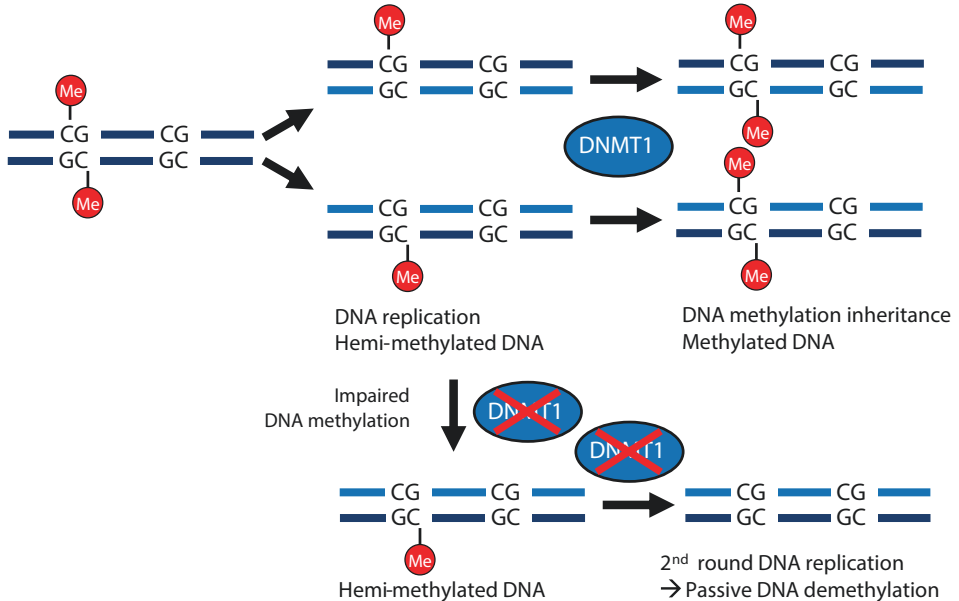
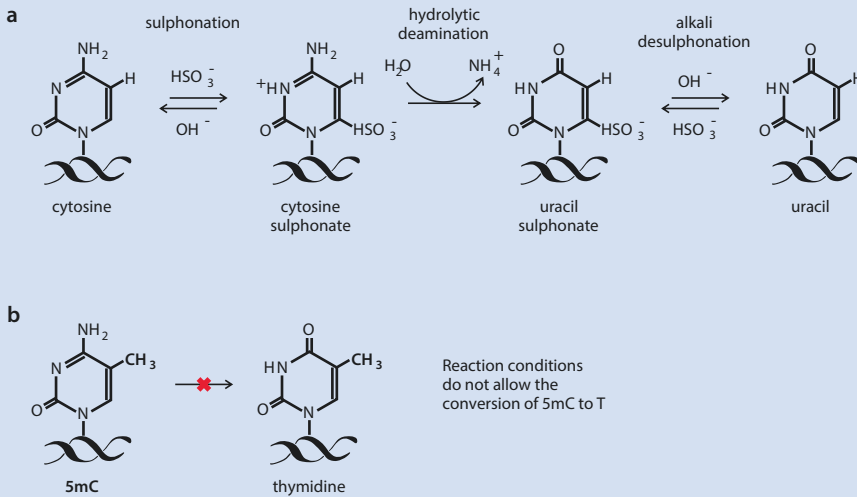


Fig. 1.8 A model for maintaining CpG methylation patterns in mammalian genomes. After a single round of DNA replication, hemi-methylated DNA is generated. Restoration of DNA methylation on the newly synthesized DNA strand leads to a heritable pattern of DNA methylation. DNMT1 is an enzyme that is targeted to hemi-methylated substrates and has maintenance DNMT activity. The original 5mC is not lost but unmethylated C is present in newly synthesized DNA. Hence, if CpG methylation patterns are not restored after 2 rounds of DNA replication, fully unmethylated (newly synthesized) DNA is expected (bottom). This leads to a loss of epigenetic information by passive, replication-dependent demethylation

If DNA with symmetrically methylated CG dinucleotides on both strands is replicated, two copies result, each having a methylated and an unmethylated (the newly synthesized) strand. Maintenance DNMT enzymes are recruited to hemi-methylated DNA to restore a fully symmetric methylation pattern (Fig. 1.8). This mechanism requires the recognition of hemi-methylated CG dinucleotides and the activation of methylation activity to restore the pattern on the newly synthesized strand for a faithful maintenance of the epigenetic marks. This conceptually simple mechanism can explain the heritability of this particular epigenetic information through cell division. We will see shortly that the molecular details are not quite as simple but the overall concept is clear and easy to comprehend.

Method Box 1.2: Analysis of DNA Modifications

DNA methylation can be detected by different methods. Bacterial restriction endonucleases that are inhibited by cytosine methylation can be used to digest genomic DNA. Electrophoretic size separation and hybridization with complementary probes (Southern blot) reveals unmethylated DNA at a lower molecular weight than the methylated DNA, which remains undigested. PCR with gene-specific primers can



■ **Box Fig. 1.2** Analysis of 5mC by bisulfite conversion. **a** The reaction mechanism for the conversion of cytosine to uracil in the presence of bisulfite. **b** The conversion of 5mC to thymine is blocked under suitably chosen reaction conditions

also be used to detect methylated DNA. When the endonuclease cuts the unmethylated template, no PCR product is obtained. Suitable controls are required as interpretation relies on the completeness of the enzymatic reaction. By using restriction enzymes that recognize the same sequence but are not inhibited by cytosine methylation, the digestibility of the DNA can be confirmed. *HpaII* and *MspI* recognize CCGG, but only *HpaII* is inhibited by C5mCGG.

5mC can also be detected by antibodies. Similar to ChIP, methylated DNA is precipitated using 5mC-specific antibodies and can then be studied by a range of techniques including dot blots, PCR, and sequencing. Similar antisera have been developed for 5hmC. These antisera have further been used for immunofluorescence staining to observe the distribution of DNA modifications in tissues and cells by microscopy techniques.

Bisulfite conversion allows to examine DNA methylation at single base resolution. Heating DNA in the presence of bisulfite leads to deamination of cytosine to uracil (■ Box Fig. 1.2). Under suitable reaction conditions, the conversion of 5mC is not observed and, thus, cytosine remains unchanged. Sequencing of the converted DNA shows unmethylated cytosines as thymines, so that methylation can be identified by comparison to a reference sequence.

Recently, nanopore sequencing and mass spectrometry have also been applied to detect and quantify DNA modifications.

What would happen if methylation does not take place, possibly because the DNMT enzyme is inactive or unavailable? Well, after another round of cell division, genomes would emerge that have completely lost DNA methylation from both strands and, hence, the corresponding epigenetic information (■ Fig. 1.8).

We would expect this to happen to half of all cells that have descended from the last ancestor with fully symmetric methylation. Further divisions would lead to increased loss as methylation patterns would not be restored and the only remaining hemi-methylated DNA strands are the ones inherited from the last ancestral cell that was fully methylated.

In mammalian genomes, CG nucleotides are observed in less than the expected frequency.⁸ As a consequence, long stretches of DNA contain 5mC at a relatively low density. To explain this observation, evolutionary erosion of methylated CG dinucleotides has been suggested. Deamination of cytosine leads to uracil that is recognized as an illicit base in DNA and is efficiently replaced. The deamination product of 5mC is thymine, which is a valid base in DNA and might not be repaired and, thus, can become fixed. If C to T mutations occur in the germline they can accumulate over evolutionary time and could explain the overall deficiency of CG dinucleotides in the genome of species with DNA methylation. This view is further supported by the observation that the unmethylated genome of *D. melanogaster* does not show a deficiency in CG dinucleotides.

Also in mammalian genomes, there are regions with the expected number of CG dinucleotides. These regions appear as islands of about 500–1000 bp of locally elevated CG density compared to the majority of the genome. These CpG islands⁹ have been associated with gene promoters and are normally devoid of any 5mC. However, methylation of cytosines in CpG islands does occur in specific physiological contexts. CpG islands are methylated on the inactive X chromosome in female mammalian cells as a consequence of transcriptional repression (see book ► Chap. 4 of Wutz). In addition, methylation of CpG island promoters of tumor suppressor genes is important in human tumors (see book ► Chap. 8 of Santoro). It has been shown that loss of DNA methylation from the promoters of tumor suppressor genes can lead to their reactivation and cause cell cycle arrest and death of tumor cells. Therefore, DNMTs have been pursued as potential targets for treating human tumors (see book ► Chap. 8 by Santoro).

The physiological function of DNA methylation has been explored by reverse genetic analysis in mice. Mice possess three genes with catalytic DNMT activity. DNMT1 is considered the maintenance DNMT and catalyzes most of the replication-coupled cytosine methylation. This is the enzyme that we invoked in the simple mechanism for inheriting DNA methylation patterns earlier in this chapter. In addition, there are two enzymes, DNMT3A and DNMT3B, that are considered *de novo* DNMTs. These DNMTs are thought to newly establish DNA methylation patterns and are targeted by other factors to chromatin. All DNMTs possess characteristic sequence homology in their catalytic domains. In mice, DNMT1 and DNMT3B are essential for embryo development. Mutations of DNMT3B are also associated with the immunodeficiency, chromosomal instability, and facial abnormalities (ICF) syndrome in humans, suggesting a contribution of DNMT3 to centromere function and

8 In a stretch of mammalian DNA, fewer CGs occur than any of the other possible dinucleotide combinations GC, AT, TA, AC, CA, AG, GA, CT, TC, GT, TG, AA, CC, GG, TT. In plants, CNG appears the sequence motif instead of CG. Like CG, CNG is symmetric on both DNA strands.

9 The term CpG islands, the “p” representing the phosphor-ribose linkage, is often used instead of CG islands to indicate that CpG dinucleotides and not CG base pairs are meant.

Table 1.1 Components of the DNA modification systems that are mutated in human disease

Gene	Disease
DNMT3B	ICF syndrome
MeCP2	RETT syndrome
DNMT3A	Haematopoietic malignancies
TET2	Haematopoietic malignancies

genomic stability. Part of the defects caused by loss of DNMT1 are derepression of retrovirus-like genomic elements, so called intra-cisternal A-particles (IAPs). Loss of DNMT1 also causes cell death of differentiated cells but might be tolerated in transformed cells to some extent. Surprisingly, early embryonic cells appear to tolerate a loss of DNA methylation. Studies have shown that DNA methylation is critical for gene regulation. In particular, gene regulation by genomic imprinting is intimately linked to cytosine methylation (see book ► Chap. 5 of Grossniklaus).

Analyses of the effects of mutations on DNA methylation led to the identification of factors that are involved in the mammalian DNA methylation system. These include structural maintenance of chromosomes hinge domain 1 (SmcHD1), which contributes to DNA methylation and gene repression at promoters on the inactive X chromosome, the non-catalytic DNMT3L that acts together with DNMT3A in establishing DNA methylation patterns in the germline, and UHRF1 (also called NP95). UHRF1 is required for association of DNMT1 with the replication machinery. UHRF1 contains a domain with specificity for hemi-methylated DNA. Further studies have led to the discovery that UHRF1 possesses catalytic activity and mediates ubiquitinylation of histone H3 lysines 23 and 18. H3K23ub and/or H3K18ub recruit DNMT1 to DNA where it changes the hemi-methylated to a fully methylated pattern. It is interesting to note that histone modifications are an integral part of the maintenance pathway of DNA methylation patterns. We will see more overlap and crosstalk between DNA and histone modifications in the chapters throughout this book.

The methyl-group of 5mC is located in the major groove of the DNA double helix where it can be accessed by readers of DNA methylation. Protein domains have been identified that mediate specific binding to methylated or unmethylated DNA. UHRF1 contains such a domain for recognizing hemi-methylated DNA. A family of DNA-binding proteins contains a methyl-DNA binding (MDB) domain. In this family of proteins, methyl-cytosine binding protein 2 (MeCP2) specifically recognizes 5mC. The function of MeCP2 is not entirely clear but it appears to affect gene expression in a subtle way. Mutations affecting MeCP2 have been shown to cause RETT-syndrome in humans (► Table 1.1). RETT syndrome is named after its discoverer and characterizes a neurodevelopmental disorder. It mainly affects girls who initially appear to develop normally but, at about 1 year of age, regress and display neurologic symptoms that overlap with symptoms of autism. The mutation is in a heterozygous state and *MeCP2* is located on the X chromosome. Due to random

X inactivation, only about half of the patients' cells lack *MeCP2* expression whereas the other half is phenotypically normal. It has been suggested that the presence of MeCP2-deficient cells might have a dominant effect disrupting brain functions, which is likely due to subtle changes in gene expression in neurons and glia cells. A mouse model has been established that recapitulates some neurological defects to study the disease mechanism. Experiments that allow the restoration of *MeCP2* function after symptoms have developed, suggest that neurologic phenotypes can be reversed (Guy et al. 2007). This finding has spurred efforts to reactivate the intact copy of *MeCP2* in human patients where an intact copy of the gene resides on the inactive X chromosome. If these approaches were successful and *MeCP2* could be reactivated, possibly by removing DNA methylation, a treatment of this devastating disease could be found. However, caution is needed as the mouse model on which these hopes were based does not fully recapitulate all phenotypes that are observed in humans. Mice that lack MeCP2 completely are viable and show neurologic symptoms, whereas absence of MeCP2 in all cells is lethal in humans. Therefore, MeCP2 mutations are generally not observed in males.

A second way how DNA methylation can affect gene expression is by preventing the recognition of binding sites by transcription factors. One might think of a steric hindrance by methyl groups that stick out into the major groove of a DNA segment containing a binding motif. The transcription factor CCCTC-binding factor (CTCF) recognizes a motif containing cytosines that can be methylated. CTCF plays a role in chromatin organization and has been associated with the formation of chromatin boundaries and insulator function (see ► Sect. 1.7.3 of this Chap., and book ► Chap. 5 of Grossniklaus). CTCF binding is blocked by DNA methylation, which can have profound consequences for the genomic region.

Studies of DNA methylation in mouse development have revealed that early embryos lose most of their DNA methylation at the blastocyst stage (see book ► Chap. 5 of Grossniklaus). Initially, the genomes of the gametes are characterized by substantial methylation, which is removed during the cleavage stages of preimplantation mouse development. Both active and passive (through DNA replication) demethylation mechanisms have been considered. It is interesting to note that the precise mechanism of enzymatic removal of 5mC from DNA has not yet been established. It is thought that DNA repair pathways play a role as it appears impossible to chemically remove the methyl-group from 5mC without opening the pyrimidine ring system. This eraser mechanism is therefore different from mechanisms of erasing histone modifications and, thus, DNA methylation is considered a relatively stable epigenetic mark.

1.4.2 DNA Cytosine Hydroxymethylation

A fundamental advance in understanding DNA methylation in animal genomes has come from the discovery of a family of enzymes that convert 5mC in DNA. The ten eleven translocation (TET) family of proteins has three members in mice, TET1, TET2, and TET3. These are iron- and α -ketoglutarate-dependent dioxygenases that use molecular oxygen to oxidize the methyl-group of 5mC (■ Fig. 1.9). The first

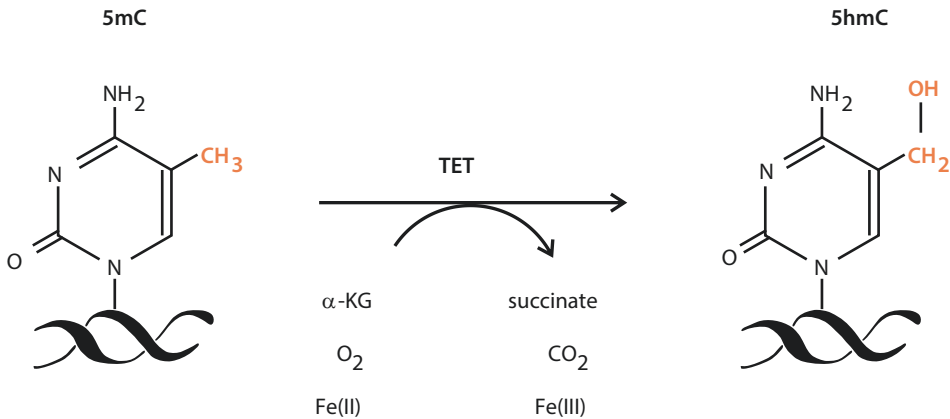


Fig. 1.9 Conversion of 5mC to 5hmC by TET dioxygenases. TET enzymes use oxygen to hydroxylate the methyl-group of 5mC. The reaction requires iron in oxidation state 2 [Fe(II)] and leads to the decarboxylation of α -ketoglutarate to succinate

product of this oxidation is 5-hydroxy-methyl-cytosine (5hmC). Further oxidation converts 5hmC to 5-formyl-cytosine (5fC), and 5-carboxy-cytosine (5caC). It has been suggested that recognition and removal of these oxidation products by the base excision DNA repair machinery could be a mechanism for active demethylation of DNA. Additional roles for different cytosine modifications have been proposed in gene regulation, organization of nucleosomes, and the regulation of histone modifications. It is important to recognize that the substrate for generating any of these modifications is 5mC and, therefore, DNMTs are required. The dependence of some modifications on a preceding reaction is interesting as a more complex regulation of epigenetic information might be established similar to an “AND” gate in Boolean logic. However, whether this aspect is used by cells is presently unclear. Also, constraints for the system of DNA modifications come from the observation that DNA methylation is dispensable in some organisms (*D. melanogaster* and *C. elegans*), suggesting that evolutionary important processes are not strongly dependent on it.

TET1 was initially discovered at the breakpoint of chromosomal translocations associated with leukemia and gave the family its name: ten-eleven-translocation. Mutation of *Tet2* in mice leads to aberrant blood cell differentiation and leukemia-like phenotypes (see book ► Chap. 8 of Santoro). These observations suggest a role of TET dioxygenases in the regulation of gene expression during cell differentiation. The function of *Tet3* has been associated with the demethylation of the genome in mouse preimplantation embryos. TET3 is present in the oocyte as a maternally supplied protein. After fertilization TET3 converts 5mC on the paternal, i.e., sperm-derived, genome to 5hmC. In contrast, the maternal genome is protected and maintains 5mC in the zygote. This leads to a differential marking of maternal and paternal chromosomes by 5mC and 5hmC, respectively. Although, the differences are removed through the subsequent passive demethylation process during embryonic cleavage, it is important to note that parent-of-origin marking is initially genome-wide. This is surprising considering that only a small number of imprinted genes maintain their parent-of-origin information throughout development.

1.4.3 Interaction of DNA and Histone Modifications

At first sight, the large number of chromatin modifications might appear daunting to comprehend. Conceptually, it can be helpful to see different modifications in the context of their biological functions. Thereby, the same chromatin modifications might be established by different enzymes and contribute to different molecular pathways. Biochemical studies have identified domains that recognize specific chromatin modifications and, thus, allow the prediction of successive modifications if binding domains are part of enzymes. A number of rules can be considered to predict and describe the makeup and function of different states of chromatin.

CG-rich DNA appears to be recognized by the PcG complexes that establish H3K27me3 and H2AK119ub. Cytosine methylation of CpG islands prevents recruitment, suggesting an antagonism between DNA methylation and PcG proteins, at least in CG-rich DNA. H3K27me3 has been shown to coexist with H3K4me3 that marks promoters. Such H3K27me3 and H3K4me3 doubly marked chromatin is considered to carry a “bivalent” mark, indicating that it has an ambivalent state between active and inactive. Yet, active promoters are acetylated on H3K27 and H3K27ac and H3K27me3 cannot coexist on the very same lysine. This logic suggests a potential step for activation by blocking PcG activity through H3K27ac (see book ► Chap. 3 of Paro). Notably, H3K9me3 does not appear to occur together with H3K4me3, such that a strong anti-correlation is observed. At pericentric heterochromatin, H3K9me3 is paired with H4K20me3 as we have seen earlier in this chapter. In addition, it is correlated with H3K4me1. Preventing methylation of histone H3 lysine 4 causes DNA *de novo* methylation. It has been shown that all mammalian DNMTs bind to H3K4 when it is unmodified. Binding leads to structural changes that activate *de novo* DNMT activity. In addition, DNMT3A and DNMT3B bind to H3K36me2 and H3K36me3 that are associated with the gene body of active genes. It is thought that DNA methylation within the transcription unit helps to prevent spurious transcription initiation, thereby reducing transcriptional noise. Undoubtedly, future studies have to extend our understanding of functional pathways in chromatin but keeping a few general rules in mind is helpful to evaluate epigenomic data and to decipher its biological relevance.

1.5 Chromatin Organization and Compartmentalization in the Cell Nucleus

We have now taken a look at chromatin and identified molecular characteristics that can carry epigenetic information along the DNA sequence. This has led to the view of a highly functional structure of chromatin. We end this chapter with a brief analysis of how chromatin is arranged in the cell nucleus. This aspect becomes important when one considers the large number of nucleosomes encompassing an animal genome. In the case of the diploid human genome, over 40 million nucleosomes will need to be assembled. This has implications of how genes can be identified and how transcription factors can bind to their recognition sites. Transport is not the only issue. The RNA polymerase II complex that transcribes most messenger RNAs exerts force on the DNA fiber. Transcribing in the order of 10,000 genes in a cell would lead to excessive movement within the nucleus. Observations from microscopy suggest

1.5 · Chromatin Organization and Compartmentalization in the Cell Nucleus

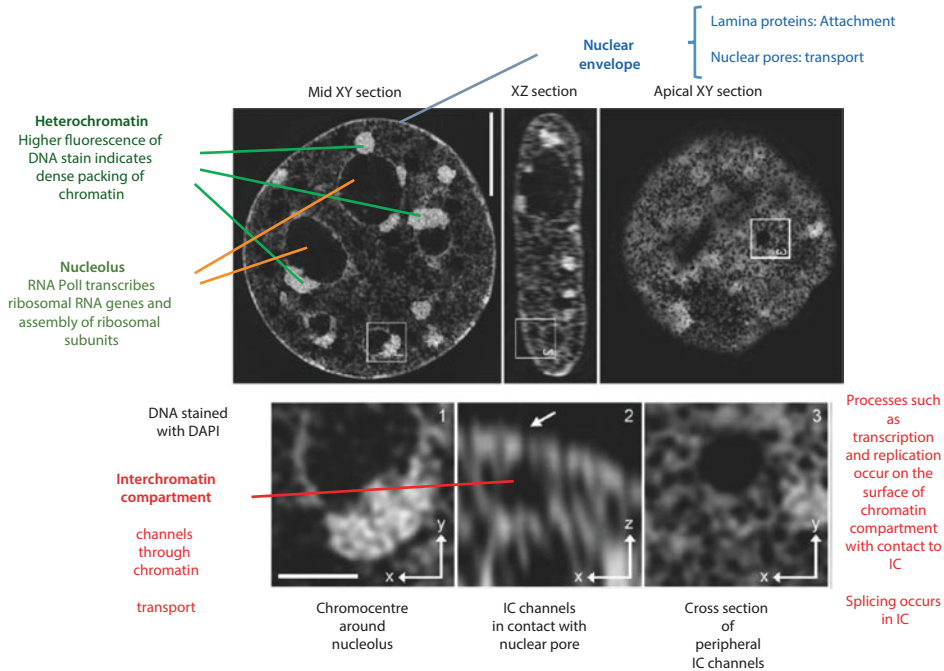


Fig. 1.10 Organization of the cell nucleus. Super-resolution microscopy images of the cell nucleus have been annotated with prominent structures and compartments. (Adapted from Markaki et al. (2012))

that chromatin maintains a relatively stable distribution. Discrete foci are observed from pericentric heterochromatin, with similar chromatin appearing to cluster together. Also, chromosomes appear to occupy discrete volumes within the cell nucleus, referred to as chromosome territories (CTs). Although different CTs are in close contact, there is very little mixing of the majority of the DNA from different chromosomes. These observations constitute evidence that the chromatin fiber is constrained and mechanisms exist that compartmentalize the nucleus.

Microscopy images of DNA-stained nuclei show the nucleoli as prominent structures (Fig. 1.10). These are the locations where ribosomal RNA is transcribed by RNA polymerase I and the assembly of ribosomal subunits takes place. Nucleoli contain RNA and ribosomes at different stages of assembly. As they contain little DNA, they stain weakly with DNA dyes and appear dark under the microscope. Nucleoli are bordered by a rim of heterochromatin called perinucleolar heterochromatin. Throughout the nucleus, DNA staining is distributed with different brightness corresponding to nucleosome density. Notably, small holes appear throughout the DNA staining. These are channels within chromatin, which are thought to allow genes to be accessed by factors and transcripts to be transported (Markaki et al. 2012). Connections of these channels to nuclear pores have been observed in super-resolution images, illustrating how mRNAs might be exported from the interior of the nucleus. Channels and gaps within chromatin are referred to as interchromatin compartment (IC). Collectively, these observations suggest that processes might take place on the surface of chromatin with factor access and transport being dependent on the IC.

1.5.1 Replication of Pericentric Heterochromatin Domains

The use of 3D fluorescent *in situ* hybridization (FISH) with complementary DNA probes makes it possible to correlate nuclear sub-domains with repetitive sequences. Fluorescently labeled DNA probes can be used to detect minor and major satellite repeat DNA of the pericentric regions in fixed nuclei. These preparations can be further stained using antibodies for histone modifications. From such experiments, it could be concluded that the pericentric repeats are organized into chromatin domains that are stained homogeneously by H3K9me3 antisera. Thus, microscopy showed that pericentric heterochromatin is indeed marked with certain histone modifications and associates with specific proteins as we discussed earlier in this chapter. Microscopy can also be useful to gain insights into how heterochromatin behaves during DNA replication. Newly synthesized DNA can be marked by incorporation of modified nucleotides such as bromo-deoxyuridine (BrdU), which can be visualized by specific antisera. Replication of pericentric repeats appears to take place at the outer border of the heterochromatin domain. The replicated DNA then associates again with the domain of heterochromatin. As we have seen earlier, pericentric heterochromatin recruits enzymes that catalyze histone modifications. It is, therefore, reasonable to predict that newly deposited histones on the replicated DNA will be targeted by histone modifying activities that are resident in pericentric heterochromatin. Restoration of the characteristic chromatin marks is accomplished during S-phase and chromosomes with faithfully marked pericentric repeats can then be segregated in mitosis. Therefore, self-association of pericentric heterochromatin can be regarded as a mechanistic basis for the heritability of histone modifications. Although this is a plausible model, it is likely that the situation is more complex. Firstly, pericentric repeats are also subject to DNA methylation and, therefore, additional interactions might have to be taken into account to understand the mechanism. Secondly, it is presently not clear how self-association of pericentric chromatin is mediated and how it can be segregated from neighboring chromatin. Could it be an intrinsic property of chromatin modifications? Are specific H3K9me3 binding proteins mediating self-association by cooperative binding mechanisms? Potentially, additional factors could be involved in this process (see book ► Chap. 3 of Paro). Anchoring of chromatin at non-chromatin structures has been proposed, including a nuclear matrix or nuclear scaffold. However, the components of these conceptually plausible structures remain to be revealed.

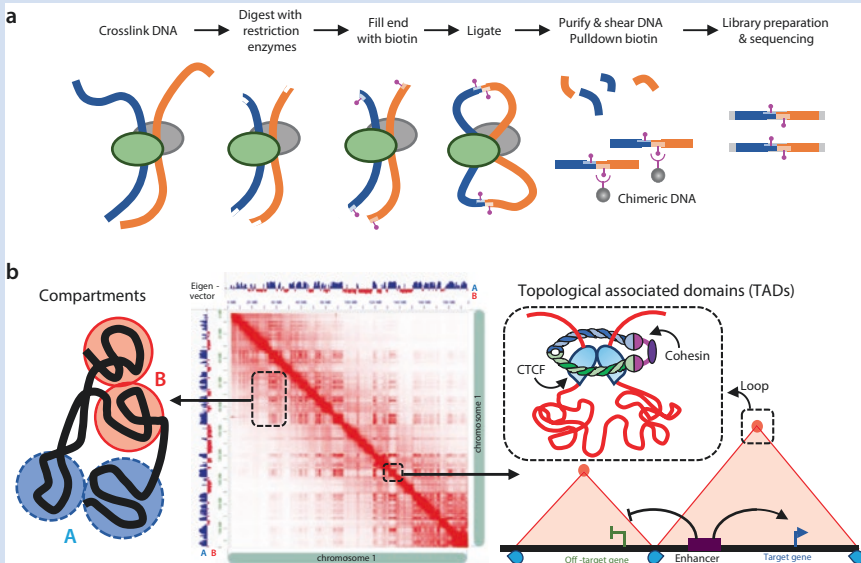
1.5.2 Topologically Associating Domains

Even in images from super-resolution microscopy, which can resolve details of 30–100 nm, a single voxel¹⁰ represents between 30 and 100 nucleosomes. Microscopy analysis is, therefore, on a different scale than our molecular understanding of reg-

10 A voxel represents the smallest volume that can be resolved in a 3D image. It is defined in analogy to a pixel, which represents a dot of a 2D picture.

ulatory elements. A typical 1 kb long promoter would be assembled into 7 nucleosomes. To bridge this gap, new methods have been developed. Chromatin conformation capture techniques (3C to HiC) aim to identify DNA fragments that are in close proximity of each other in the nucleus but might be far apart in the linear DNA sequence. The method preserves DNA contacts by cross-linking chromatin when it is in its native state in the cell (► Method Box 1.3). Thereafter, chromatin is digested with restriction enzymes. Ligation of the ends of the cut DNA in very dilute conditions yields chimeric DNA fragments. DNA fragments that are frequently in close proximity in the nucleus have a higher probability to be cross-linked in the same chromatin particle and, hence, show an elevated number of chimeric fragments. Sequence determination of the ligation products and plotting the number of chimeric DNA fragments against their distance in the linear DNA sequence yields a contact frequency estimate. Using this approach, the *in vivo* 3D topology of chromosomes has been investigated. From these studies it has become apparent that chromosomes are made up of domains of interacting chromatin that are partitioned by boundaries between them. Intervals with high internal contact frequencies are termed topologically associating domains (TADs, ► Method Box 1.3). Although TADs might share regulatory elements and be co-regulated, it is important to realize that 3C and related techniques usually investigate a large number – several thousands to millions – of cells and, therefore, only averages are observed. It is not always clear whether, in a given cell at a given time, all contacts take place. 3C techniques have been very successful in identifying gene regulatory elements such as enhancers that are located at a distance from their promoters but show a high contact frequency. Chromatin conformation data is, therefore, useful to determine regulatory interactions and possibly to identify genes that are coregulated by certain *cis*-regulatory elements. Recently, chromatin conformation capture-based techniques have been extended to single cell analysis. This has further contributed to our understanding of chromatin organization. From a study that used haploid G1 phase cells to make unambiguous assignments of DNA sequence contacts, it has become clear that a general subnuclear arrangement is adopted in all cells (Stevens et al. 2017). A-compartments predominantly consist of sequences that are correlated with active chromatin marks. These can be distinguished from B-compartments that contain largely inactive chromatin and frequently locate to the nuclear periphery. The precise winding and geometry of the chromatin fiber is different in each cell. Therefore, chromatin in the nucleus is stochastically distributed at the smallest scale but becomes more deterministic as the scale increases. A potential mechanism for such a topology has been proposed as local self-organization of chromatin, whereby locally acting elements would affect chromatin structure in their proximity. The combined activity of many elements would then result in the overall observed topology.

Method Box 1.3: Chromatin Conformation Capture (Box Fig. 1.3)



Box Fig. 1.3 Chromatin conformation capture. **a** The 3D organization of chromatin is preserved by crosslinking protein complexes and DNA with formaldehyde. The DNA is then digested with nucleases into short fragments. Subsequent ligation of DNA ends from fragments that are crosslinked facilitates the capture of interactions between sequences that are distant in the linear sequence of the genome but have been in close proximity in the nucleus. Sequencing of the ligation products and analysis of the frequency of chimeric DNA can be used to calculate the frequency of interaction between distant genomic regions. **b** Organization of chromatin compartments in the nucleus (left) can be deduced from interaction frequencies. A matrix of color-coded interaction frequencies (middle) can visualize the interaction of sequences in the genome. The diagonal line that runs from top left to right bottom represents interactions of regions that are in close proximity in the DNA sequence and are, thus, closely linked by the DNA backbone. Topological associated domains (TADs) (right) are confined by borders that limit interactions. High frequencies of interactions are observed within TADs and little interactions are observed outside. Cohesin complexes have been associated with TAD borders. Schematic representation of A and B compartments is also shown

1.5.3 Structural Maintenance of Chromosomes Complexes

The association of enhancers with gene promoters and the organization of TADs raises the expectation for mechanisms that facilitate the winding and bending of the chromatin fiber to achieve the observed topology. This aspect has been a long-standing question, which appeared difficult to answer for many years. Recent evidence illustrates a surprising and very powerful mechanism involving protein complexes of the structural maintenance of chromosomes (SMC) family. SMC proteins associate with chromatin and have important roles in chromosome condensation and segregation during mitosis. Two distinct SMC complexes have been identified, referred to as cohesin and condensin complexes in anticipation of their function in mediating sister chromatid cohesion and condensation of mitotic chromosomes, respectively. SMC proteins

contain ATPase head domains that can hydrolyze ATP to exert motor function. Reconstituted recombinant condensin complexes (containing SMC1 and SMC3) have the ability to induce changes in the topology of fluorescently marked DNA molecules. This observation has been made in sophisticated microscopy-based biochemical assays (Ganji et al. 2018). DNA loops were formed when DNA fragments were incubated with condensin complexes and ATP. A similar function has subsequently been observed for a recombinant human cohesin complex (containing SMC2 and SMC4), when additional cohesin-associated factors were present (Kim et al. 2019). These experiments suggest the SMC family as important regulators of chromatin topology.

Cohesins have a role in the cell cycle where they mediate the cohesion of sister chromatids after replication of the chromosomal DNA until segregation at mitosis. In addition, cohesins have been found to be involved in gene regulation. Cohesins are loaded on chromatin in interphase and likely can slide along the chromatin fiber, mediating loop formation. Some chromatin loops correlate with the presence of CTCF sites at the loop base (► Method Box 1.3). CTCF binds sequences that are asymmetric and, therefore, have an orientation. CTCF sites at the base of chromatin loops are oriented towards each other, suggesting that CTCF might act in a directional manner to block cohesin sliding. A number of TAD boundaries can be correlated with CTCF binding sites. In addition, other molecules can also contribute to the localization of cohesins and possibly other SMC complexes. Therefore, the recent discovery of the DNA motor function of SMC complexes suggests a fundamental mechanism for establishing DNA topology at a larger scale that will certainly continue to attract attention in the future (Banigan and Mirny 2020).

Take-Home Message

- Chromatin organizes the genome of eukaryotic cells into a nucleosomal structure that facilitates the imposition of epigenetic information and assists in regulating the DNA sequence.
- The nucleosomal structure can be modified through histone variants, posttranslational histone modifications, and DNA modifications that act combinatorically and synergistically.
- Specific enzymes that establish histone modifications can be considered writers of epigenetic information. Conversely, readers are proteins with affinity for histone marks that can recruit protein complexes with additional functions to chromatin. Erasers are enzymes that remove chromatin marks and, thus, ensure that epigenetic information is reversible.
- Folding of the linear DNA in the nucleus facilitates the compartmentalization of chromatin into domains with different functional properties.
- Chromosomes are not randomly distributed but are organized into discrete volumes called chromosome territories, whereby separation of active A and repressed B compartments can be observed.
- Modifications of chromatin structure and chromatin topology both contribute to establishing information along the sequence of the DNA that regulates the function of distinct genomic regions.
- The understanding of nuclear pathways for chromatin regulation is a basis for the consideration of epigenetic mechanisms that will further be discussed in following chapters.

References

- Banigan EJ, Mirny LA (2020) Loop extrusion: theory meets single-molecule experiments. *Curr Opin Cell Biol* 64:124–138. <https://doi.org/10.1016/j.ceb.2020.04.011>
- Ganji M, Shaltiel IA, Bisht S, Kim E, Kalichava A, Haering CH, Dekker C (2018) Real-time imaging of DNA loop extrusion by condensin. *Science* 360(6384):102–105. <https://doi.org/10.1126/science.aar7831>
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235):223–227. <https://doi.org/10.1038/nature07672>
- Guy J, Gan J, Selfridge J, Cobb S, Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315(5815):1143–1147. <https://doi.org/10.1126/science.1138389>
- Hirota T, Lipp JJ, Toh BH, Peters JM (2005) Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* 438(7071):1176–1180. <https://doi.org/10.1038/nature04254>
- Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H (2019) Human cohesin compacts DNA by loop extrusion. *Science* 366(6471):1345–1349. <https://doi.org/10.1126/science.aaz4475>
- Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648):251–260. <https://doi.org/10.1038/38444>
- Markaki Y, Smeets D, Fiedler S, Schmid VJ, Schermelleh L, Cremer T, Cremer M (2012) The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. *BioEssays* 34(5):412–426. <https://doi.org/10.1002/bies.201100176>
- Olins AL, Olins DE (1974) Spheroid chromatin units (v bodies). *Science* 183(4122):330–332. <https://doi.org/10.1126/science.183.4122.330>
- Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* 18(11):1251–1262. <https://doi.org/10.1101/gad.300704>
- Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, Cramard J, Faure AJ, Ralser M, Blanco E, Morey L, Sanso M, Palayret MGS, Lehner B, Di Croce L, Wutz A, Hendrich B, Klenerman D, Laue ED (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544(7648):59–64. <https://doi.org/10.1038/nature21429>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

