Introduction

In eukaryotic cells the genetic material is organized into a complex structure composed of DNA and proteins and localized in a specialized compartment, the nucleus. This structure, detected with basic dyes, was called chromatin (from the Greek "khroma" meaning coloured and "soma" meaning body) at the end of the 19th century (Flemming, 1882). Close to two meters of DNA in each cell must be assembled into a small nucleus of some \( \mu \) in diameter. Despite this enormous degree of compaction, DNA must be rapidly accessible to permit its interaction with protein machineries that regulate the functions of chromatin: replication, repair and recombination. The dynamic organization of chromatin structure thereby influences, potentially, all functions of the genome.

The fundamental unit of chromatin, termed the nucleosome, is composed of DNA and histone proteins. This structure provides the first level of compaction of DNA into the nucleus. Nucleosomes are regularly spaced along the genome to form a nucleofilament which can adopt higher levels of compaction (Figures 1 and 3), ultimately resulting in the highly condensed metaphase chromosome. The combined approaches of cell biology and genetic studies have led to the discovery that within an interphase nucleus chromatin is organized into functional territories (Cockell and Gasser, 1999).

Historically, based on microscopic observations, chromatin has been divided into two distinct domains, heterochromatin and euchromatin. Heterochromatin was defined as a structure that does not alter in its condensation throughout the cell cycle whereas euchromatin is decondensed during interphase (Heitz, 1928). Typically in a cell, heterochromatin is localized principally on the periphery of the nucleus and euchromatin in the interior of the nucleoplasm. We can distinguish constitutive heterochromatin, containing few genes and formed principally of repetitive sequences located in large regions coincident with centromeres and telomeres, from facultative heterochromatin composed of transcriptionally active regions that can adopt the structural and functional characteristics of heterochromatin, such as the inactive X chromosome of mammals (Lyon, 1999; Avner and Heard, 2001).

In this review we will define the components of chromatin and outline the different levels of its organization from the nucleosome to domains in the nucleus. We will discuss how variation in the basic constituents of chromatin can impact on its activity and how stimulatory factors play a critical role in imparting diversity to this dynamic structure. Finally we will summarize how chromatin influences the organization of the genome at the level of the nucleus.

The nucleosome

Historically, the periodic nature of chromatin was identified by biochemical and electron microscopic studies. The partial digestion of DNA assembled into chromatin, isolated from rat liver nuclei, generated fragments of 180-200 base pairs in length which were resolved by electrophoretic migration (Williamson, 1970; Hewish and Burgoyne, 1973). This regularity of chromatin structure was later
confirmed by electron microscope analysis that revealed chromatin as regularly spaced particles or "beads on a string" (Olins and Olins, 1974; Oudet et al., 1975). In parallel, chemical cross-linking analysis permitted the precise determination of the stoichiometry of DNA and histones in the nucleosome to be 1/1 based on their mass (Kornberg and Thomas, 1974). Together these observations led to the proposition that the nucleosome was the fundamental unit of chromatin. Pierre Chambon's laboratory was the first to use the term "nucleosome" (Germont et al., 1976). It is composed of a core particle and a linker region (or internucleosomal region) that joins adjacent core particles (Figure 1). The core particle is highly conserved between species and is composed of 146 base pairs of DNA wrapped 1.7 turns around a protein octamer of two each of the core histones H3, H4, H2A and H2B. The length of the linker region, however, varies between species and cell type. It is within this region that the variable linker histones are incorporated. Therefore, the total length of DNA in the nucleosome can vary with species from 160 to 241 base pairs (Compton et al., 1976; Morris, 1976; Noll, 1976; Spadafora et al., 1976; Thomas and Thompson, 1977).

The first crystal structure of the core particle was obtained at a resolution of 7A by diffraction of X rays (Finch et al., 1977). More recently the structure of the octamer was resolved at 3.1A (Arents and Moudrianakis, 1993). Finally, crystallization that enabled a resolution of 2.8A was obtained using a unique sequence of DNA and purified recombinant proteins (Luger et al., 1997). This analysis revealed, firstly, the distortion of the DNA wound around the histone octamer and, secondly, that the histone/DNA and histone/histone interactions through their "histone fold domain" formed a configuration reminiscent of a handshake. This structural information has facilitated experimental approaches used to study the functions of specific regions of histones, with the exception of their N-terminal tails that are not visualized in the crystal.

**Histone proteins**

**Core histones**
The core histones, H3, H4, H2A and H2B, are small, basic proteins highly conserved in evolution (Figure 2). The most conserved region of these histones is their central domain structurally composed of the "histone fold domain" consisting of three α-helices separated by two loop regions (Arents et al., 1991). In contrast, the N-terminal tails of each core histone are more variable and unstructured. The tails are particularly rich in lysine and arginine residues making them extremely basic. This region is the site of numerous post-translational modifications that are proposed to modify its charge and thereby alter DNA accessibility and protein/protein interactions with the nucleosome (Strahl and Allis, 2000).

It is significant to note that other proteins that interact with DNA also contain the "histone fold domain" (Baxevanis et al., 1995; Wolfe and Pruss, 1996). There is a web site that lists proteins containing a "histone fold domain" (http://genome.nhgri.nih.gov/histones/) (Baxevanis and Landsman, 1998; Sullivan et al., 2002).
Linker histones

Linker histones associate with the linker region of DNA between two nucleosome cores and, unlike the core histones, they are not well conserved between species (Richmond and Widom, 2000). In higher eukaryotes, they are composed of three domains: a globular, non-polar central domain essential for interactions with DNA and two non-structured N- and C-terminal tails that are highly basic and proposed to be the site of post translational modifications. The linker histones have a role in spacing nucleosomes and can modulate higher order compaction by providing an interaction region between adjacent nucleosomes. The precise role of the linker histones is still a controversial subject (Khochbin, 2001) and a variety of models have been proposed (Pruss et al., 1996; Thomas, 1999).

General steps in chromatin assembly

The assembly of DNA into chromatin involves a range of events, beginning with the formation of the basic unit, the nucleosome, and ultimately giving rise to a complex organization of specific domains within the nucleus. This step-wise assembly is described schematically in Figure 3. The first step is the deposition onto the DNA of a tetramer of newly synthesized (H3-H4)2 to form a sub-nucleosomal particle, which is followed by the addition of two H2A-H2B dimers (Senshu et al., 1978; Cremisi et al., 1977; Worcel et al., 1978). This produces a nucleosomal core particle consisting of 146 base pairs of DNA wound around a histone octamer. This core particle and the linker DNA together form the nucleosome (Kornberg and Thomas, 1974). Newly synthesized histones are specifically modified; the most conserved modification is the acetylation of histone H4 on lysine 5 and lysine 12 (Sobel et al., 1995). The next step is the maturation step that requires ATP to establish regular spacing of the nucleosome cores to form the nucleofilament. During this step the newly incorporated histones are de-acetylated. Next the incorporation of linker histones is accompanied by folding of the nucleofilament into the 30nm fibre, the structure of which remains to be elucidated. Two principal models exist: the solenoid model, an example of which is presented in Figure 3, and the zig zag (Woodcock and Dimitrov, 2001). Finally, further successive folding events lead to a high level of organization and specific domains in the nucleus.

At each of the steps described above, variation in the composition and activity of chromatin can be obtained by modifying its basic constituents and the activity of stimulatory factors implicated in the processes of its assembly and disassembly. Assembly begins with the incorporation of the H3/H4 tetramer (1), followed by the addition of two H2A-H2B dimers (2) to form a core particle. The newly synthesized histones utilized are specifically modified; typically, histone H4 is acetylated at Lys5 and Lys12 (H3-H4*). Maturation requires ATP to establish a regular spacing, and histones are de-acetylated (3). The incorporation of linker histones is accomplished by folding of the nucleofilament. Here the model presents a solenoid structure in which there are six nucleosomes per gyre (4). Further folding events lead ultimately to a defined domain organization within the nucleus (5) (for details see (Ridgway and Almouzni, 2001)).
Variation in basic constituents

In the first steps of chromatin assembly, the elementary particle can assume variations at the level of DNA (for example by methylation) or at the histone level by differential post-translational modification and the incorporation of variant forms (for example CENP-A, a variant of H3). All of these variations are capable of introducing differences in the structure and activity of chromatin. The vast array of post-translational modifications of the histone tails summarized in Figure 2 (such as acetylation, phosphorylation, methylation, ubiquitination, polyADP-riboseylation), and their association with specific biological processes has led to a proposed hypothesis of a language, referred to as the "histone code", that marks genomic regions (Strahl and Allis, 2000). It must be emphasized that this code is a working hypothesis used to design experimental approaches that investigate the activity of chromatin. The code is "read" by other proteins or protein complexes that are capable of understanding and interpreting the profiles of specific modifications (Strahl and Allis, 2000; Jenuwein and Allis, 2001). The incorporation of histone variants such as H2A-X (Rogakou et al., 1999), CENP-A (Sullivan et al., 1994), macro H2A (Pehrson and Fried, 1992) and H2A.Z (Clarkson et al., 1999) may be important at specific domains of the genome. In this context, CENP-A, a variant of histone H3 is associated with silent centromeric regions (Sullivan et al., 1994) and macro H2A on the inactive X chromosome of female mammals (Mermoud et al., 1999). H2A-X is implicated in the formation of foci containing DNA repair factors in the regions of DNA double-strand breaks (Paule et al., 2000). Growing evidence exists that H2A.Z has a role in modifying chromatin structure to regulate transcription (Santisteban et al., 2000).

During the maturation step, incorporation of linker histones, non-histone chromatin associated proteins, called HMG (High Mobility Group), and other specific DNA-binding factors help to space and fold the nucleofilament. Therefore the early steps in assembly can have a great impact on the final characteristics of chromatin in specific nuclear domains (Marshall et al., 1997).

Stimulatory Assembly Factors

(for review see (Kaufman and Almouzni, 2000))

Histone interacting factors

Biochemical fractionation of extracts derived from cells or embryos permitted the identification of acidic factors that can form complexes with histones and enhance the process of histone deposition. They act as histone chaperones by facilitating the formation of nucleosome cores without being part of the final reaction product. These histone-interacting factors, also called chromatin-assembly factors, can bind preferentially to a subset of histone proteins. In Xenopus laevis, the proteins N1-N2 and nucleoplasmin are respectively associated with histones H3-H4 and histones H2A-H2B (Laskey et al., 1978). The situation is more complex for NAP-1 (Nucleosome Assembly Protein 1), depending on the system studied (Ito et al., 1997). Chromatin Assembly Factor-1 (CAF-1) interacts with newly synthesized acetylated histones H3 and H4 to preferentially assemble chromatin during DNA replication (Smith and Stillman, 1989; Kaufman et al., 1997).
CAF-1 is also capable of promoting the assembly of chromatin specifically coupled to the repair of DNA (Gaillard et al., 1996). The recent demonstration of the interaction of CAF-1 with the protein PCNA (Proliferating Cell Nuclear Antigen) established a molecular link between the assembly of chromatin and the processes of replication and repair of DNA (see for review Ridgway and Almouzni, 2000; Mello and Almouzni, 2001)). The assembly of specialized structures in centromeric regions, by deposition of variant histones such as CENP-A, or telomeres may be a result of the specificity and the diversity of as yet uncharacterised histone chaperones.

**Remodelling machines and histone-modifying enzymes**

Stimulatory factors also act during the chromatin maturation stage to organize and maintain a defined chromatin state. Their effects on chromatin can induce changes in conformation at the level of the nucleosome or more globally over large chromatin domains. These factors are of two types; one requiring energy in the form of ATP, generally referred to as chromatin remodelling machines, and the other that act as enzymes to post-translationally modify histones.

Chromatin remodelling machines are multi-protein complexes which have now been characterized from yeast, humans and *Xenopus laevis* and are summarized in Table 1. Complexes sharing the same ATPase subunit are classified in the same family. The ATPase subunit, mating-type switching/sucrose non-fermenting (swi2/snf2) defines the first family, the ATPase ISWI the second family, also termed ISWI, and the last family is called Mi2/NuRD (nucleosome remodeling histone deacetylase) complex after the name of the ATPase subunit Mi2. The activity of the ATPase permits the complex to modify nucleosomal structure, driven by the liberation of energy during the hydrolysis of ATP (Travers, 1999). The study of factors that stimulate the regular arrangement of nucleosomes during the assembly of chromatin led to the identification of several multi-protein complexes such as ATP-utilizing chromatin assembly and remodelling factor (ACF) (Ito et al., 1997; Ito et al., 1999), chromatin accessibility complex (CHRAC) (Varga-Weisz et al., 1997) and remodeling and spacing factor (RSF) (LeRoy et al., 1998). These complexes are capable of "sliding" nucleosomes along DNA *in vitro* (Langst et al., 1999; Ito et al., 1999). The common feature of these chromatin remodelling factors is their large size and multiple protein subunits including the ATPase, however, they display differences in abundance and activity. Remodels the Structure of Chromatin (RSC), for example, is ten times more abundant than SWI/SNF and contains 15 subunits in contrast to the 11 in SWI/SNF but it has six subunits in common with SWI/SNF including a homologous ATPase (reviewed in Bjorklund et al., 1999). In contrast to the SWI/SNF complex, all the subunits of RSC are essential for viability of yeast (Cao et al., 1997).

The "histone code" hypothesis has been proposed to explain the diversity of chromatin activity in the nucleus. The unstructured N-terminal histone tails extend outside the nucleosome core and are the sites of action for enzymes that catalyze with high specificity their post-translational modification. The most well characterized of these modifications is the acetylation of lysine residues. Acetylation is the result of an equilibrium between two opposing activities: histone acetyl transferase (HAT) and histone deacetylase (HDAC) (for review see (Taddei and Almouzni, 1997)). An *in gel* electrophoretic protein separation method allowed the identification of the first protein with a histone acetyltransferase activity, HAT A, also called p55 in *Tetrahymena* (Brownell and Allis, 1995). The characterisation of specific inhibitors such as trapoxine resulted in the purification of the first histone deacetylase, human HDAC1 (Taunton et al., 1996). Numerous proteins that play a role in the regulation of transcription have intrinsic histone acetyltransferase activity such as GCN5, PCAF and TAFII250 (Brownell et al., 1996; Mizzen et al., 1996). Similarly, histone deacetylases have been described as components of multi-protein complexes associated with repressive chromatin. Also within these complexes are the Mi-2 family of remodeling factors (Knoepfler and Eisenman, 1999; Ng and Bird, 2000) providing a link between remodelling of nucleosomes and histone deacetylation during chromatin-mediated repression.

**Table 1:** Chromatin remodeling complexes divided into families based on the similarity of their ATPase subunit (see for review (Kingston and Narlikar, 1999; Fry and Peterson, 2001)).

<table>
<thead>
<tr>
<th>Family</th>
<th>Chromatin remodeling complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWI2/SNF2</td>
<td>SWI2/SNF, RSC, Brahma complex, BRM complex, BRG-1 complex</td>
</tr>
<tr>
<td>ISWI</td>
<td>ISWI, ISW2, INO80, NURF, CHRAC, ACF, RSF</td>
</tr>
<tr>
<td>Mi2/NuRD</td>
<td>NuRD, NRD, NuRD, Mi2, Mi2 complex</td>
</tr>
</tbody>
</table>

*Functional organization of the genome: chromatin* Ridgway P et al. 2002; 6(3) 245
Very recently it has been proposed that another modification, methylation of histones, plays a functionally important role (Jenuwein and Allis, 2001). The first histone-methyltransferase, called SUV39H1 in human, was only recently discovered (Rea et al., 2000). It specifically methylates histone H3 on lysine residue 9 and this methylation modifies the interaction of H3 with heterochromatin associated proteins (Bannister et al., 2001; Lachner et al., 2001). The two possible modifications (acylation and methylation) on the same residue (lysine 9) of the N-terminal tail of H3 is a perfect illustration of the "histone code" hypothesis in action. Indeed, acetylated lysine in H3 and H4 N-terminal tail selectively interact with chromodomain present in numerous proteins having intrinsic histone acetyltransferase activity. However, H3 methylated on lysine residue 9 interact specifically with the chromodomain of an heterochromatin associated protein HP1. Therefore, in addition to producing alterations in the overall charge of the histone tails, proposed to physically destabilize the nucleosome, modifications appear to impart specific protein:protein interactions with the histones. They are associated with different regions of the genome and are correlated with precise nuclear functions (Strahl and Allis, 2000).

Organization of the genome in the nucleus

The higher level of compaction of chromatin is not as well characterized. The nucleofilament is compacted to form the 30nm fibre that is organized into folds of 150 to 200 Kbp (250nm during interphase) to obtain a maximum level of compaction in the metaphase chromosome (850nm). At interphase the organization of the genome relies on the structure of chromosomes that have been characterized into different regions based on a specific banding pattern revealed by Giemsa staining (Comings, 1974; Belyaev et al., 1996). The principle bands are G and C bands that are late replicating in S phase and correspond to heterochromatin and the R bands that replicate earlier in S-phase and represent euchromatin. The R bands are enriched in acetylated histones and this modification is conserved through mitosis suggesting that histone acetylation may serve as a marker for the memory of domain organization through the cell cycle (Turner, 1998; Sadoni et al., 1999).

The localization of chromosomes in the interphase nucleus by Fluorescence In Situ Hybridization (FISH) reveals that each chromosome occupies a defined space (Lamond and Earnshaw, 1998). This observation is in accordance with the notion of chromosomal territories proposed, in 1885 by Rabl, for the organization of chromosomes in plants. The Rabl configuration, in which the telomeres are attached to the nuclear envelope beside the nucleus with the centromeres on the other side, has been described in a number of cell types (Marshall et al., 1997). However, in mammals, this configuration does not exist and the organization of the chromosomes in the nucleus varies as a function of cell type (He and Brinkley, 1996). During interphase, regions that correspond to the bands of metaphase chromosomes are located in the nucleus based on the timing of their replication. On the nuclear periphery are the later replicating regions, corresponding to G and C bands and the transcriptionally silent telomeres, while gene rich regions are preferentially localized more internally. Therefore, although each chromosome occupies a different territory, distinct parts of chromosomes can unite to form functional domains (Cockell and Gasser, 1999; Croft et al., 1999). The localization of coincident and non-coincident regions by FISH suggests that genes tend to be localized at the surface of chromosome territories. In the model proposed by Cremer, based on the localization of some genes, transcripts are released into interchromosomal channels, transferred to sites for processing, then exported to the cytoplasm after maturation (Cremer et al., 1993) (and see for review (Cremer and Cremer, 2001)).


<table>
<thead>
<tr>
<th>HMG motif</th>
<th>Functional motif</th>
<th>Root symbol</th>
<th>Newname (canonical HMGs)</th>
<th>Old name (canonical HMGs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMG-box proteins</td>
<td>HMG-box</td>
<td>HMG-GB</td>
<td>HMG-E1,2,..,n</td>
<td>HMG-1/HMG-2</td>
</tr>
<tr>
<td>NED proteins</td>
<td>NED</td>
<td>HMGN</td>
<td>HMG-N1,2,..,n</td>
<td>HMG-14/HMG-17</td>
</tr>
<tr>
<td>ATH proteins</td>
<td>ATH</td>
<td>HMG-A</td>
<td>HMG-A1,2,..,n</td>
<td>HMG-1/HMG-C</td>
</tr>
</tbody>
</table>

Abbreviations: ATH: AT-hook; HMG: high mobility group; NED: nucleosome binding domain
Several studies have led to the proposal that the nucleus is organized into domains (van Holde and Zlatanova, 1995; Lamond and Earnshaw, 1998; Bridger and Bickmore, 1998). The localization of DNA in these domains is perhaps, in part, a consequence of the activities of chromatin. Targeting proteins might help to bring specialized proteins to specific domains in the nucleus (Sutherland et al., 2001). In a hypothetical model, the proteins associated with heterochromatin (for example HP1, Polycomb, Sir3p/Sir4p and ATRX), transcription factors (such as Ikaros) and assembly factors (such as CAF-I) may all be involved in the for establishment and maintenance of nuclear domains.

References


Heitz E. Das heterochromatin der Moose Jb Wist Botan. 1928; 69: 728.


Hewish DR, Burgoyne LA. The calcium dependent endonuclease activity of isolated nuclear preparations. Relationships between its occurrence and the occurrence of other classes of enzymes found in nuclear preparations. Biochern Biophys Res Commun. 1973 May 15;52(2):475-81


Senshu T, Fukuda M, Ohashi M. Preferential association of newly synthesized H3 and H4 histones with newly replicated DNA. J Biochem. 1978 Oct;84(4):398-8


Pehrson JR, Fried VA. MacroH2A, a core histone containing a large nonhistone region. Science. 1992 Sep 4;257(5075):1398-400

Arents G, Moudrianakis EN, Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. Proc Natl Acad Sci U S A. 1993 Nov 15;90(22):10489-93


Atlas Genet Cytogenet Oncol Haematol. 2002;6(3) 248


Jenuwein T, Allis CD. Translating the histone code. Science. 2001 Aug 10;293(5532):1074-80


This article should be referenced as such: