

GENOME STRUCTURE

Course: Molecular Biology (BIOL 333)

Instructor: Dr. Mahmoud A. Srouf

Textbook:

Watson J, et al. (2014). Molecular Biology of the Gene, 7th ed. Chapter 8

Introduction

- ❑ Inside the cell (prokaryote or eukaryote) each DNA molecule and its associated protein is called a chromosome.
- ❑ Half of the molecular mass of an eukaryotic chromosome is protein.
- ❑ In eukaryotic cells, a given region of DNA with its associated proteins is called chromatin,
- ❑ Chromatin = DNA + proteins (histone and non-histone proteins)

- Packaging of DNA into chromosomes achieves several functions including”
 - ▣ The chromosome is a compact form of DNA that fits inside cells
 - ▣ Protects the DNA from damage
 - ▣ Only DNA packaged into a chromosome can be transmitted efficiently to both daughter cells when a cell divides.
 - ▣ The chromosome confers an overall organization to each molecule of DNA

How large is the human genome?

- Human haploid genome is $\sim 3 \times 10^9$ bp
- The average thickness of each bp is 0.34 nm.
- If the DNA molecules in a haploid set of chromosomes were laid out end to end, the total length of DNA would be 10^9 nm, or 1 m!
- For a diploid cell (as human cells typically are), this length is doubled to 2 m.
- The diameter of a typical human cell nucleus is only 10-15 μm , thus it is obvious that the DNA must be compacted by many orders of magnitude to fit in such a small space.
- **How is this achieved?**

Nucleus vs. a tennis ball??

- If the 2m DNA is transformed to 2 km thread (or 4 km when we consider it as a duplex), and the diameter of nucleus is transformed from 10 μm to 10 cm>
- Can you fit this 2 km duplex thread into a tennis ball???

Chromosome diversity: linear vs. circular

TABLE 8-1 Variation in Chromosome Makeup in Different Organisms

Species	Number of Chromosomes	Chromosome Copy Number	Form of Chromosome(s)	Genome Size (Mb)
Prokaryotes				
<i>Mycoplasma genitalium</i>	1	1	Circular	0.58
<i>Escherichia coli</i> K-12	1	1	Circular	4.6
<i>Agrobacterium tumefaciens</i>	4	1	3 circular, 1 linear	5.67
<i>Sinorhizobium meliloti</i>	3	1	Circular	6.7
Eukaryotes				
<i>Saccharomyces cerevisiae</i> (budding yeast)	16	1 or 2	Linear	12.1
<i>Schizosaccharomyces pombe</i> (fission yeast)	3	1 or 2	Linear	12.5
<i>Caenorhabditis elegans</i> (roundworm)	6	2	Linear	97
<i>Arabidopsis thaliana</i> (weed)	5	2	Linear	125
<i>Drosophila melanogaster</i> (fruit fly)	4	2	Linear	180
<i>Tetrahymena thermophilus</i> (protozoa)				
Micronucleus	5	2	Linear	125
Macronucleus	225	10–10,000	Linear	
<i>Fugu rubripes</i> (fish)	22	2	Linear	393
<i>Mus musculus</i> (mouse)	19+ X and Y	2	Linear	2600
<i>Homo sapiens</i>	22+ X and Y	2	Linear	3200

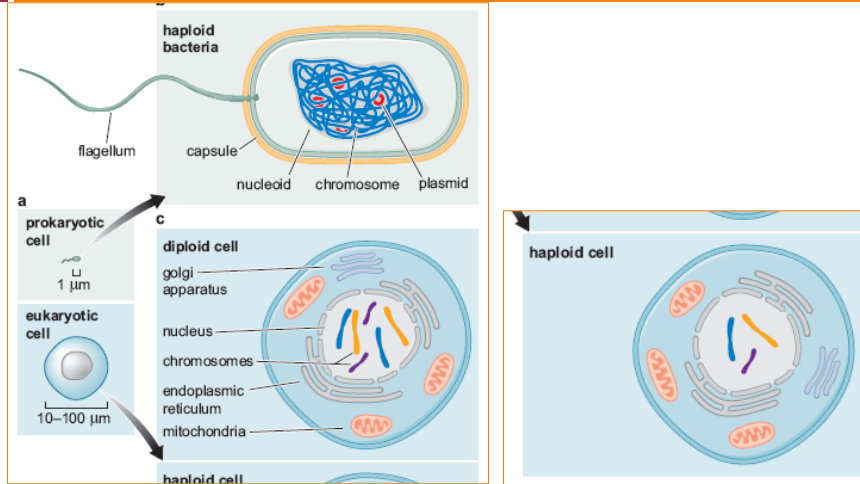
Every cell maintains a characteristic number of chromosome

- Prokaryotic cells typically have only one complete copy of their chromosome(s) that is packaged into a structure called the nucleoid
- unlike chromosomal DNA, plasmids are often present in many complete copies per cell
- The majority of eukaryotic cells are diploid; that is, they contain two copies of each chromosome.
- The two copies of a given chromosome are called homologs—one being derived from each parent

Every cell maintains a characteristic number of chromosome

- A subset of eukaryotic cells are either haploid or polyploid
- In extreme cases, there can be hundreds or even thousands of copies of each chromosome
- For example, megakaryocytes are specialized polyploid cells (about 28 copies of each chromosome) that produce thousands of platelets,

Comparison of typical prokaryotic and eukaryotic cells



Genome size is related to the complexity of the organism

TABLE 8-2 Comparison of the Gene Density in Different Organisms' Genomes

Species	Genome Size (Mb)	Approximate Number of Genes	Gene Density (genes/Mb)
Prokaryotes (bacteria)			
<i>Mycoplasma genitalium</i>	0.58	500	860
<i>Streptococcus pneumoniae</i>	2.2	2300	1060
<i>Escherichia coli</i> K-12	4.6	4400	950
<i>Agrobacterium tumefaciens</i>	5.7	5400	960
<i>Sinorhizobium meliloti</i>	6.7	6200	930
Eukaryotes (animals)			
Fungi			
<i>Saccharomyces cerevisiae</i>	12	5800	480
<i>Schizosaccharomyces pombe</i>	12	4900	410
Protozoa			
<i>Tetrahymena thermophila</i>	125	27,000	220
Invertebrates			
<i>Caenorhabditis elegans</i>	103	20,000	190
<i>Drosophila melanogaster</i>	180	14,700	82
<i>Ciona intestinalis</i>	160	16,000	100
<i>Locusta migratoria</i>	5000	nd	nd
Vertebrates			

Table 8-2/cont'd

TABLE 8-2 Comparison of the Gene Density in Different Organisms' Genomes

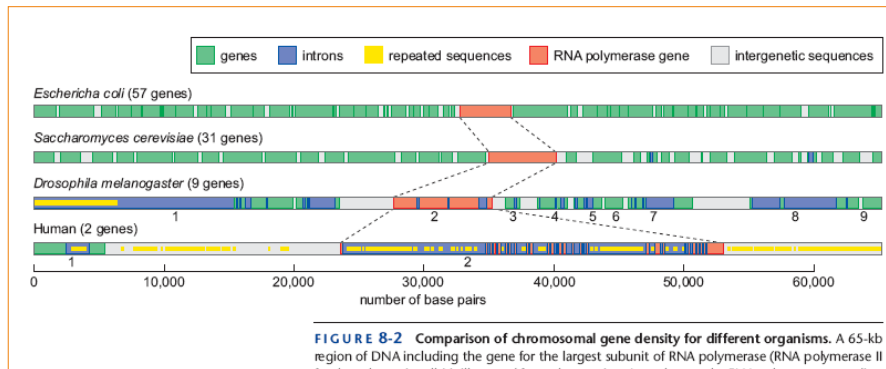
Species	Genome Size (Mb)	Approximate Number of Genes	Gene Density (genes/Mb)
Vertebrates			
<i>Fugu rubripes</i> (pufferfish)	393	22,000	56
<i>Homo sapiens</i>	3200	20,000	6.25
<i>Mus musculus</i> (mouse)	2600	22,000	8.5
Plants			
<i>Arabidopsis thaliana</i>	120	26,500	220
<i>Oryza sativa</i> (rice)	430	~45,000	~100
<i>Zea mays</i> (corn)	2200	>45,000	>20
<i>Triticum aestivum</i> (wheat)	16,000	nd	nd
<i>Fritillaria assyriaca</i> (tulip)	~120,000	nd	nd

nd, Not determined.

Genome size and biological complexity

- Although there is a rough correlation between genome size and organism complexity, this relationship is far from perfect.
- For example: the rice genome is about 40 times smaller than that of wheat examples point out that the number of genes, rather than genome size, is more closely related to organism complexity.

The E. coli Genome Is Composed Almost Entirely of Genes



More Complex Organisms Have Decreased Gene Density

- ❑ What explains the dramatically different genome sizes of organisms of apparently similar complexity (such as the fruit fly and locust)?
- ❑ The differences are largely related to gene density
- ❑ There is a roughly inverse correlation between organism complexity and gene density—the less complex the organism, the higher the gene density
- ❑ For example, the highest gene densities are found for viruses that, in some instances, use both strands of the DNA to encode overlapping genes.

Fruit fly vs. Locust



Genes Make Up Only a Small Proportion of the Eukaryotic Chromosomal DNA

- Two factors contribute to the decreased gene density observed in eukaryotic cells:
 - ▣ increases in gene size and increases in the DNA between genes, called intergenic sequences
 - ▣ protein-encoding genes in eukaryotes frequently have discontinuous protein-coding regions
- For example, the average transcribed region of a human gene is 27 kb (this should not be confused with the gene density), whereas the average protein-coding region of a human gene is 1.3 kb.

Schematic of RNA splicing to remove non-coding introns to produce messenger RNA (mRNA).

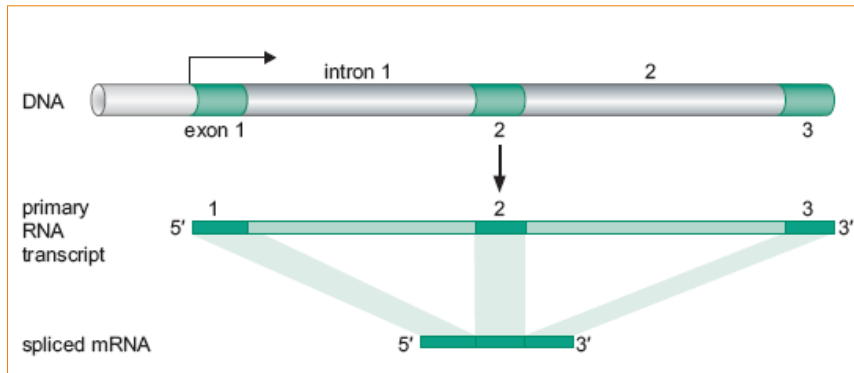
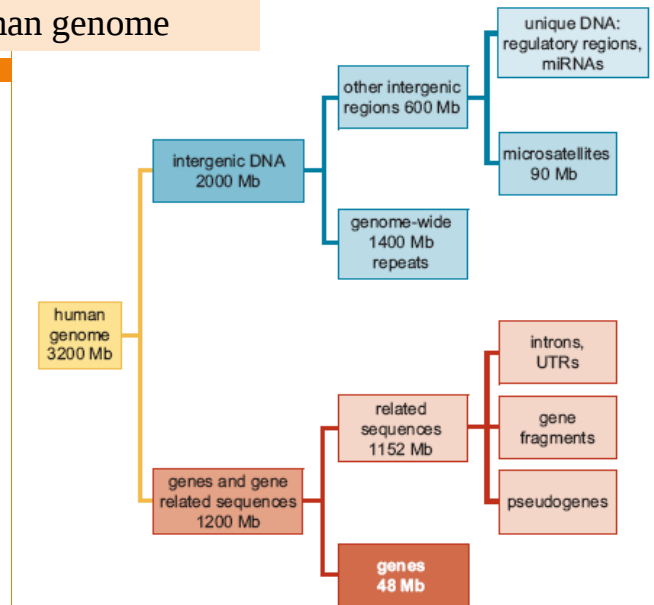


TABLE 8-3 Contribution of Introns and Repeated Sequences to Different Genomes

Species	Gene Density (genes/Mb)	Average Number of Introns per Gene	% of Repetitive DNA
Prokaryotes (bacteria)			
<i>Escherichia coli</i> K-12	950	0	<1
Eukaryotes (animals)			
Fungi			
<i>Saccharomyces cerevisiae</i>	480	0.04	3.4
Invertebrates			
<i>Caenorhabditis elegans</i>	190	5	6.3
<i>Drosophila melanogaster</i>	82	3	12
Vertebrates			
<i>Fugu rubripes</i>	56	5	2.7
<i>Homo sapiens</i>	6.25	6	46
Plants			
<i>Arabidopsis thaliana</i>	220	3	nd
<i>Oryza sativa</i> (rice)	~100	nd	42

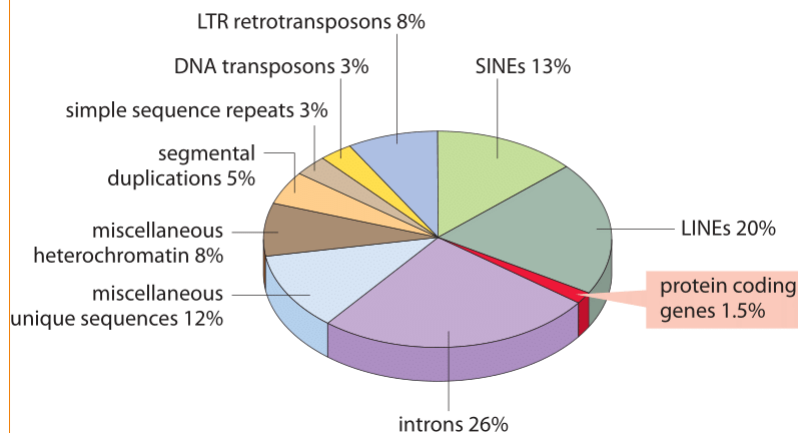
nd, Not determined.

Organization and content of the human genome



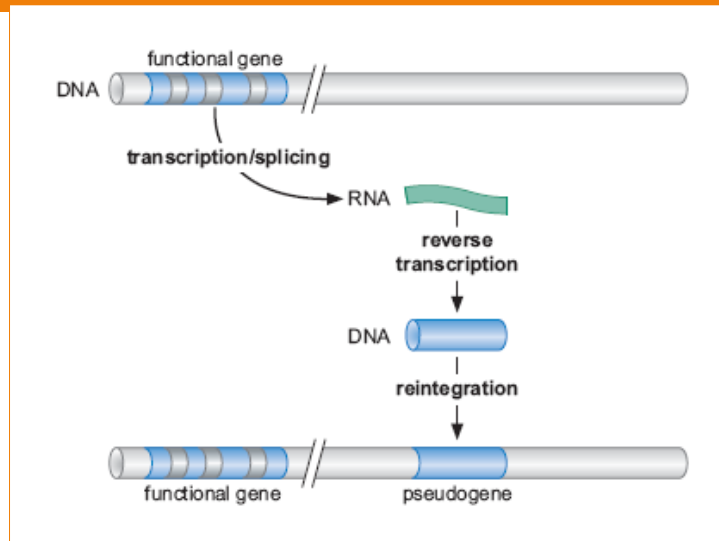
The different sequence components making up the human genome.

main components of the human genome



Non-LTR retrotransposons include SINES (~300 bp long) & LINEs (~6kb long)

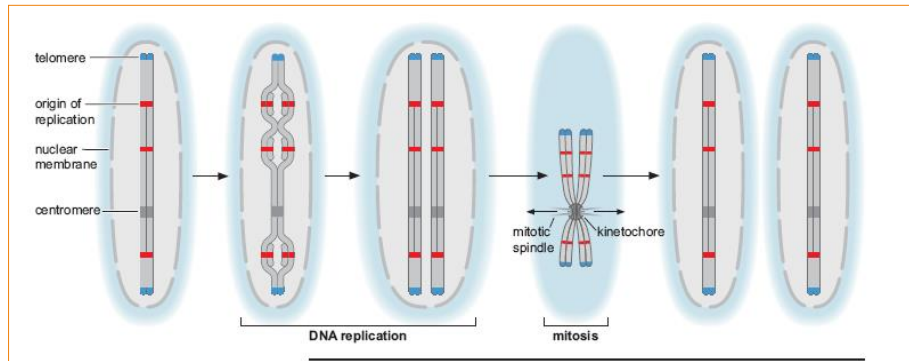
Processed pseudogenes arise from integration of reverse-transcribed messenger RNAs.



The Majority of Human Intergenic Sequences Are Composed of Repetitive DNA

- Almost half of the human genome is composed of DNA sequences that are repeated many times in the genome. There are two general classes of repeated DNA:
 - ▣ Microsatellite DNA is composed of very short (<13 bp), tandemly repeated sequences. The most common are dinucleotide repeats (e.g., CACACACACACACA) & represent nearly 3% of the human genome.
 - ▣ Genome-wide repeats (transposable elements) are much larger than their microsatellite counterparts. Each genome-wide repeat unit is >100 bp in length and many are >1 kb.

Eukaryotic Chromosomes Require Centromeres, Telomeres, and Origins of Replication to Be Maintained during Cell Division



Chromosome elements

- ❑ **Origins of replication** are the sites at which the DNA replication machinery assembles and replication is initiated.
 - ▣ Eukaryotic chromosomes, have ori's every 30–40 kb
 - ▣ Prokaryotic chromosomes typically have only a single ori
 - ▣ Ori's are found in non-coding regions
- ❑ **Centromeres** are required for the correct segregation of the chromosomes after DNA replication

Chromosome elements (cont'd)

- Telomeres are located at the two ends of a linear chromosome, they perform 2 functions:
 - ▣ Telomeric proteins distinguish the natural ends of the chromosome from sites of chromosome breakage and other DNA breaks in the cell
 - ▣ Telomeres act as specialized origins of replication that allow the cell to replicate the ends of the chromosomes

Nucleosomes Are the Building Blocks of Chromosomes

- The majority of the DNA in eukaryotic cells is packaged into nucleosomes. Each nucleosome is composed of a core of eight histone proteins and the DNA wrapped around them
- the core DNA, is wound about 1.65 times around the histone octamer
- The core DNA is 147-bp long. The length of core is the same in all eukaryotic cells. In contrast, the length of the linker DNA between nucleosomes is variable. Typically, this distance is 20–60 bp, and each eukaryote has a characteristic average linker DNA length

- In any cell, there are stretches of DNA that are not packaged into nucleosomes. Typically, these are regions of DNA engaged in gene expression, replication, or recombination

Nucleosomes Are the Building Blocks of Chromosomes

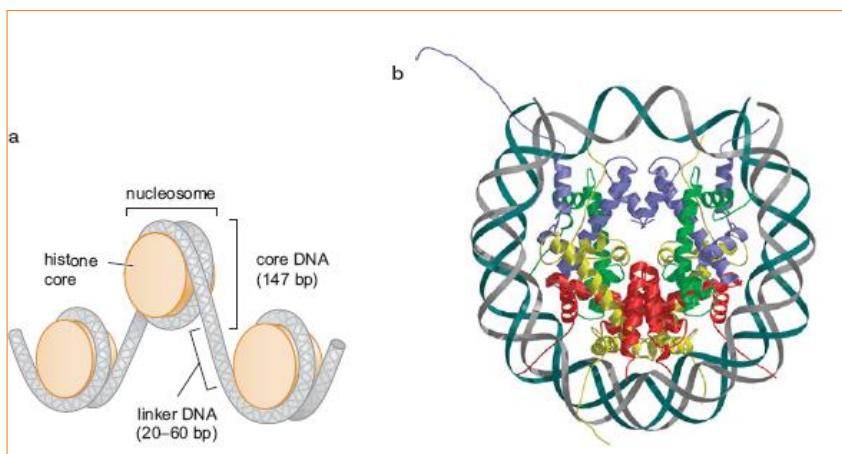


TABLE 8-4 Average Lengths of Linker DNA in Various Organisms

Species	Nucleosome Repeat Length (bp)	Average Linker DNA Length (bp)
<i>Saccharomyces cerevisiae</i>	160–165	13–18
Sea urchin (sperm)	~260	~110
<i>Drosophila melanogaster</i>	~180	~33
Human	185–200	38–53

Histones Are Small, Positively Charged Proteins

TABLE 8-5 General Properties of the Histones

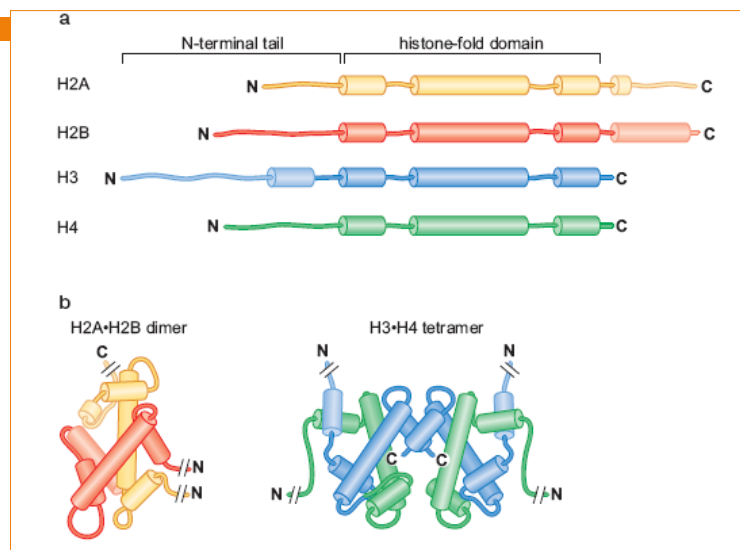
Histone Type	Histone	Molecular Weight (M_r)	Lysine and Arginine (%)
Core histones	H2A	14,000	20
	H2B	13,900	22
	H3	15,400	23
	H4	11,400	24
Linker histone	H1	20,800	32

Histones have a high content of positively charged amino acids. At least 20% of the residues in each histone are either lysine or arginine.

Histone proteins

- A conserved region found in every core histone, called the **histone-fold domain**, mediates the assembly of these histone-only intermediates
- The core histones each have an amino-terminal extension, called a **tail** because it lacks a defined structure and is accessible within the intact nucleosome.
- **Tails** are the sites of extensive posttranslational modifications that alter the function of individual nucleosomes. These modifications include phosphorylation, acetylation, and methylation on serine, lysine, and arginine residues

Core histones share a common structural fold



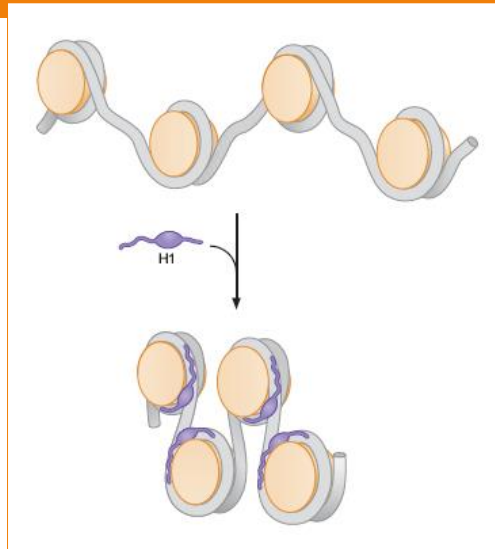
Many DNA Sequence–Independent Contacts Mediate the Interaction between the Core Histones and DNA

- The association of DNA with the nucleosome is mediated by a large number (about 40) of hydrogen bonds between the histones and the DNA

Higher-order chromatin structure

- **Heterochromatin and Euchromatin:**
- Heterochromatic regions of chrom had very limited gene expression
- Euchromatic regions showed higher levels of gene expression

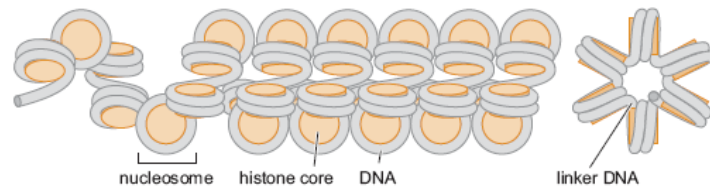
Nucleosome assembly: Beads-on-string



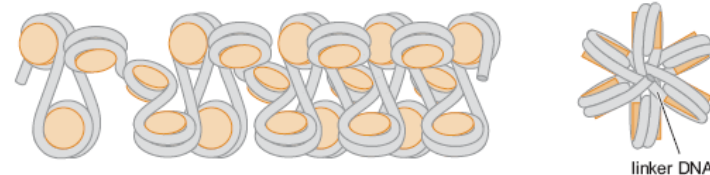
HistoneH1 induces tighter DNA wrapping around the nucleosome

Two models for the 30-nm chromatin fiber

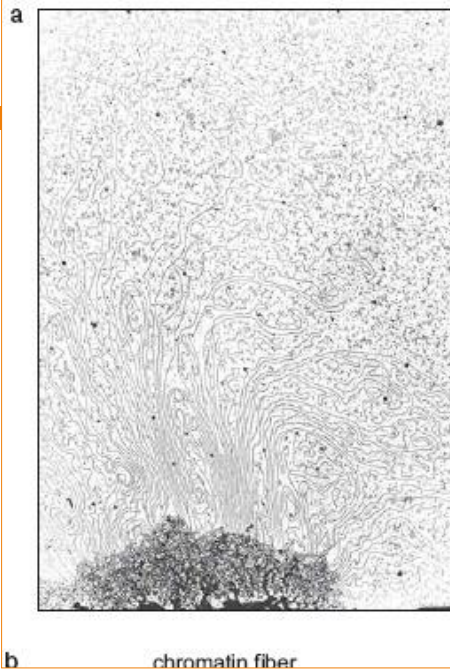
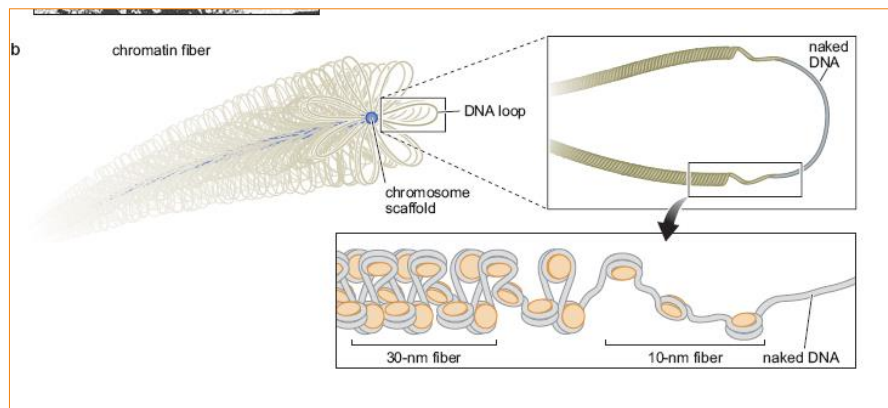
a solenoid



b zigzag



Further Compaction of DNA Involves Large Loops of Nucleosomal DNA



Higher-order
structure of
chromatin

Nucleosome assembly

- Nucleosomes Are Assembled Immediately after DNA Replication

Chromatin assembly factors facilitate the assembly of nucleosomes

