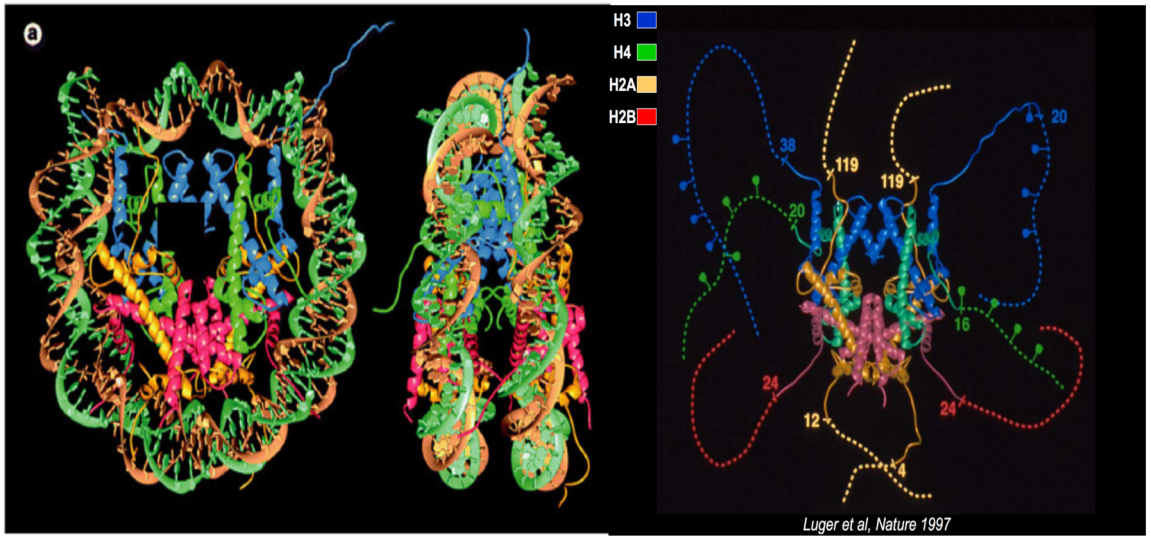
Organization of chromatin - Histones

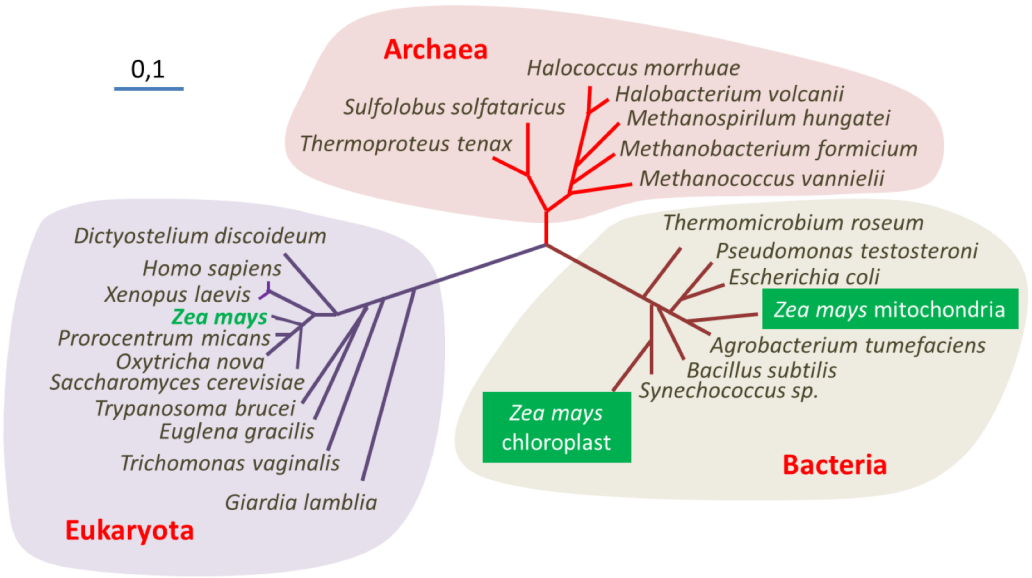
organism	genome size (base pairs)	protein coding genes	number of chromosomes
model organisms			
model bacteria *E. coli*	4.6 Mbp	4,300	1
budding yeast *S. cerevisiae*	12 Mbp	6,600	16
fission yeast *S. pombe*	13 Mbp	4,800	3
amoeba *D. discoideum*	34 Mbp	13,000	6
nematode *C. elegans*	100 Mbp	20,000	12 (2n)
fruit fly *D. melanogaster*	140 Mbp	14,000	8 (2n)
model plant *A. thaliana*	140 Mbp	27,000	10 (2n)
moss *P. patens*	510 Mbp	28,000	27
mouse *M. musculus*	2.8 Gbp	20,000	40 (2n)
human *H. sapiens*	3.2 Gbp	21,000	46 (2n)
viruses			
hepatitis D virus (smallest known animal RNA virus)	1.7 Kb	1	ssRNA
HIV-1	9.7 kbp	9	2 ssRNA (2n)
influenza A	14 kbp	11	8 ssRNA
bacteriophage λ	49 kbp	66	1 dsDNA
Pandoravirus salinus (largest known viral genome)	2.8 Mbp	2500	1 dsDNA
organelles			
mitochondria - *H. sapiens*	16.8 kbp	13 (+22 tRNA +2 rRNA)	1
mitochondria – *S. cerevisiae*	86 kbp	8	1
chloroplast – *A. thaliana*	150 kbp	100	1
bacteria			
C. ruddii (smallest genome of an endosymbiont bacteria)	160 kbp	182	1
M. genitalium (smallest genome of a free living bacteria)	580 kbp	470	1
H. pylori	1.7 Mbp	1,600	1
Cyanobacteria S. elongatus	2.7 Mbp	3,000	1
methicillin-resistant *S. aureus* (MRSA)	2.9 Mbp	2,700	1
B. subtilis	4.3 Mbp	4,100	1
S. cellulosum (largest known bacterial genome)	13 Mbp	9,400	1
archaea			
Nanoarchaeum equitans (smallest parasitic archaeal genome)	490 kbp	550	1
Thermoplasma acidophilum (flourishes in pH<1)	1.6 Mbp	1,500	1
Methanocaldococcus (Methanococcus) jannaschii (from ocean bottom hydrothermal vents; pressure >200 atm)	1.7 Mbp	1,700	1
Pyrococcus furiosus (optimal temp 100°C)	1.9 Mbp	2,000	1
eukaryotes - multicellular			
pufferfish *Fugu rubripes* (smallest known vertebrate genome)	400 Mbp	19,000	22
poplar *P. trichocarpa* (first tree genome sequenced)	500 Mbp	46,000	19
corn *Z. mays*	2.3 Gbp	33,000	20 (2n)
dog *C. familiaris*	2.4 Gbp	19,000	40
chimpanzee *P. troglodytes*	3.3 Gbp	19,000	48 (2n)
wheat *T. aestivum* (hexaploid)	16.8 Gbp	95,000	42 (2n=6x)
marbled lungfish *P. aethiopicus* (largest known animal genome)	130 Gbp	unknown	34 (2n)
herb plant *Paris japonica* (largest known genome)	150 Gbp	unknown	40 (2n)

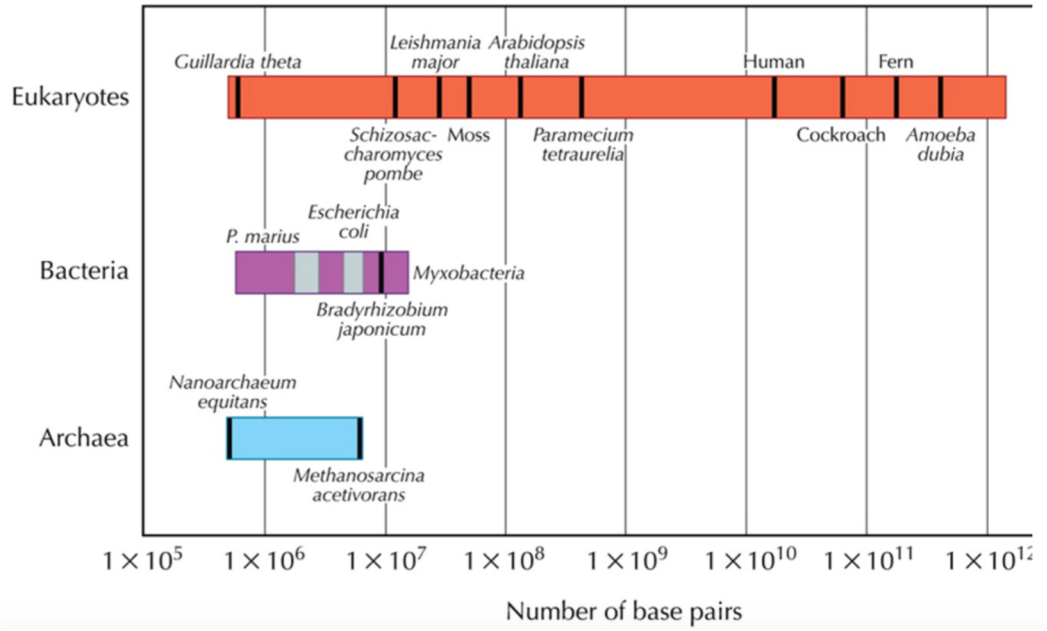
Genomes are organized



Luger et al., 1997

Genomes in the 3 domains of living organisms





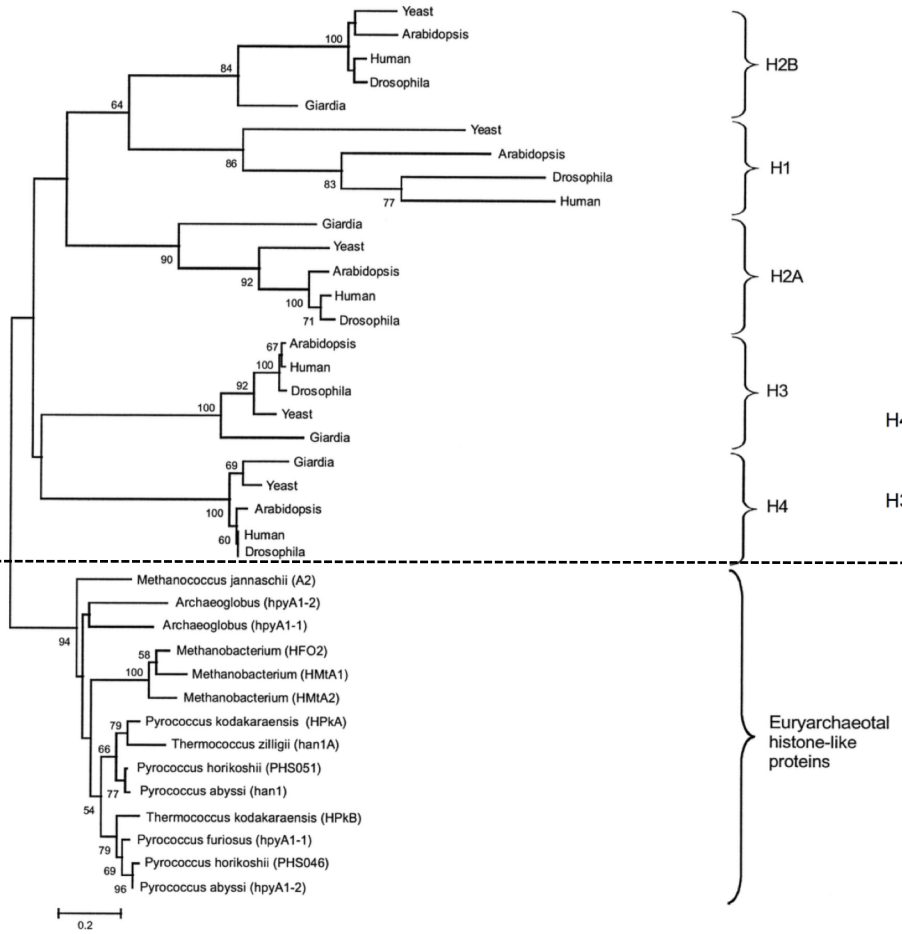
The unrooted phylogenetic tree of the three domains of living organisms, produced using a gene from the small ribosomal subunit

A phyllogenetic view on histone proteins

Eukaryotic and Archaeal histones have a common ancestor

Orgnization of bacterial genomes are unrelated to eukaryotes and archaee bacteria

Histones
are highly conserved
among eukaryotes

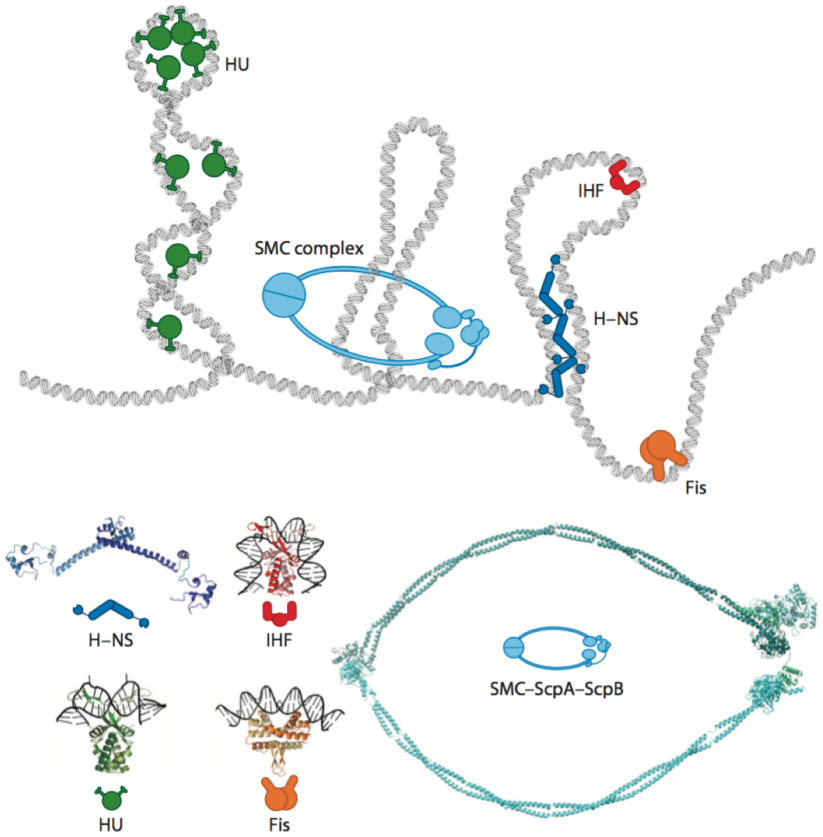


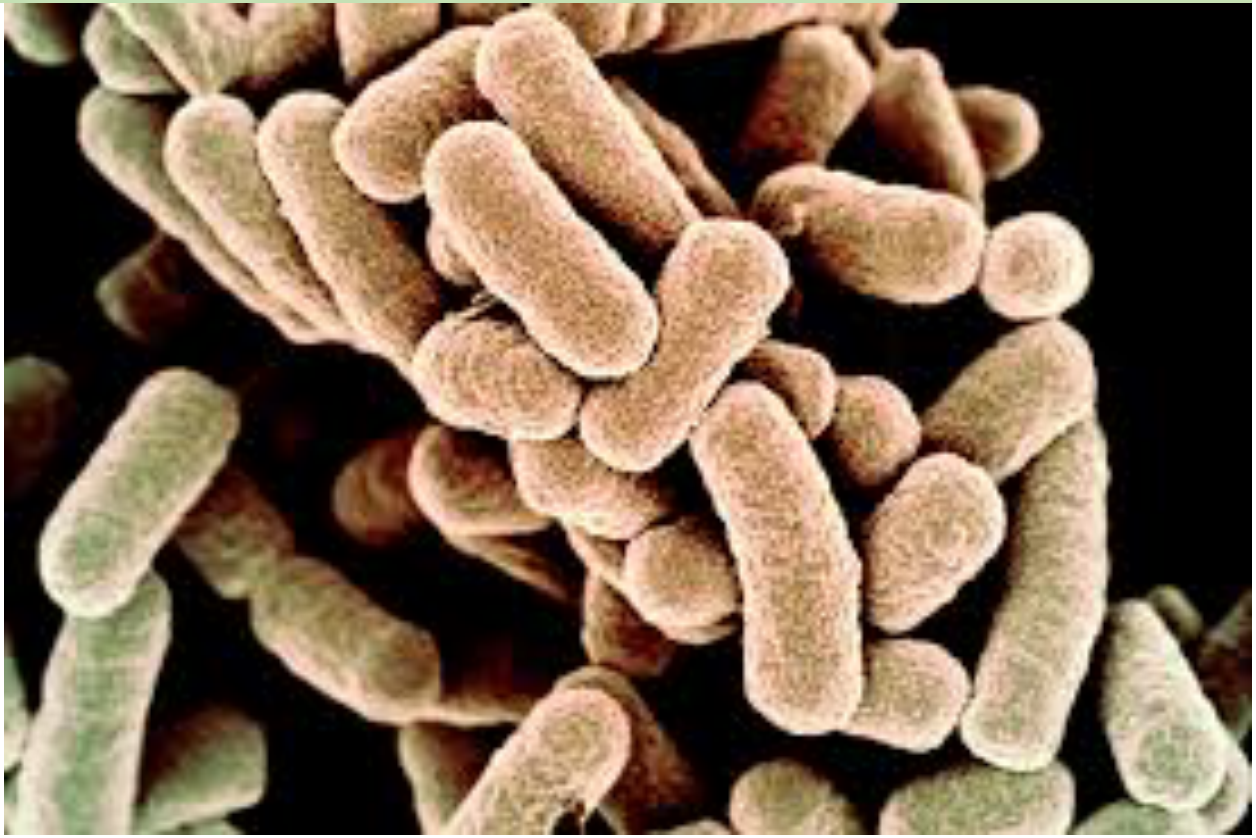
H4: very highly conserved between species
(98% identical between cows and peas)
~ 1% change in 600 million years (2 changes, Val to Ile, Lys to Arg)

H3: also very conserved *(97% identical)*
~ 1% change in 300 million years

Histone-like proteins
In archaebacteria

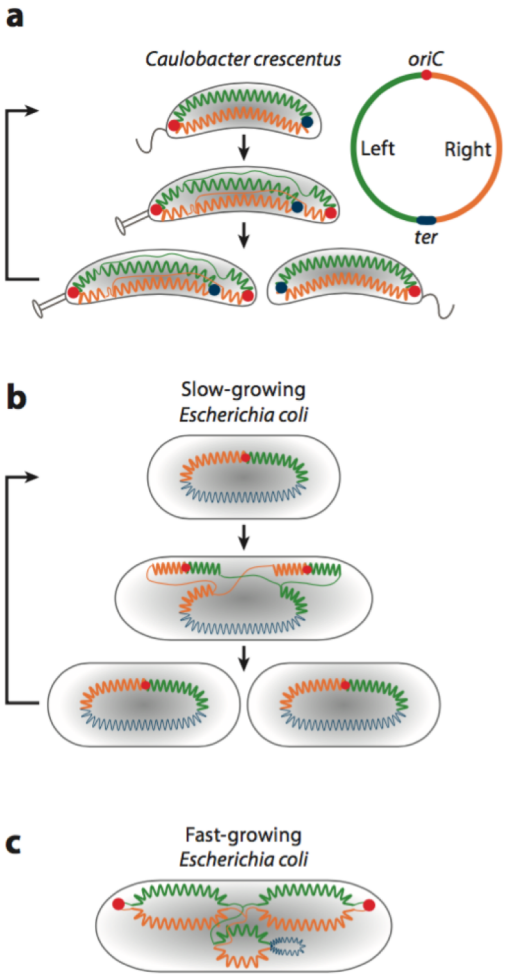
Genome organization in bacteria





Helical fold of DNA in bacteria

A helix-like conformation of DNA has been observed in replicating bacteria. The biological significance of a helical fold is unknown but may represent an energy-minimal configuration for fitting chromosomes within rod-shaped cells.



Caulobacter crescentus

- The spatial positions of loci within the cell recapitulated the genetic map, with the origin of replication (*oriC*) at one cell pole, the replication terminus (*ter*) at the opposite pole, and the left and right chromosomal arms likely running in parallel down the long axis of the cell, a pattern referred to as the ***ori-ter* configuration**.
- Found by DNA-FISH using specific DNA probes that map along the chromsom
- In replicating cells, one new copy of oriC is rapidly segregated to the opposite pole. As replication proceeds, the newly generated DNA moves to its respective position, again with loci arranged relative to the origin in a manner that reflects the genetic map

Escherichia coli

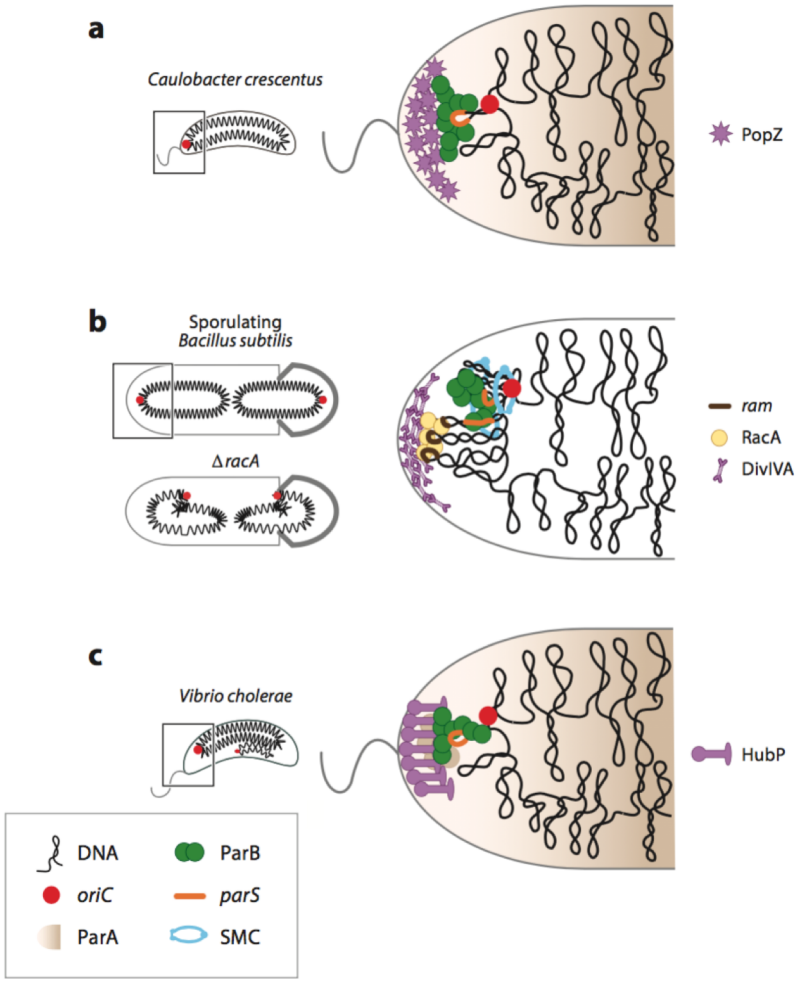
The origin resides near midcell, with the two chromosomal arms on opposite sides of the cell and the terminus typically near midcell, in a so-called left-*ori*-right configuration (Nielsen et al. 2006b, Wang et al. 2006). DNA replication and segregation of the origins to cell quarter positions regenerates a **left-*ori*-right organization** for each chromosome.

Chromosome consists of 4 macrodomanis – organisatorial units (found by DNA-FISH

By contrast, fast-growing *E. coli* cells adopt an ***ori-ter* configuration** of the chromosome with polarly localized origins

Anchoring of genomic DNA in bacteria

Polar anchoring of origins of chromosomes **in *ori-ter* configuration**, is thought to enforces the *ori-ter* pattern, and may help to ensure that each daughter cell inherits a full copy of the genome after DNA replication.



In *C. crescentus*, a *parS* site that is critical for chromosome segregation (discussed in the section titled The ParAB System for Origin Segregation) is located ~13 kb from the origin and is bound by ParB which also binds PopZ, a cytoplasmic protein that self-aggregates into a proteinaceous matrix at cell poles

A protein called RacA accumulates prior to sporulation and concentrates near the cell pole (Ben-Yehuda et al. 2003, Wu & Errington 2003). RacA binds 25 RacA-binding motif (*ram*) sites near *ori*C, helping tether *ori*-proximal regions of the chromosome to the pole. Polar localization of RacA requires a small peripheral membrane protein called DivIVA, which recognizes the concave curvature of the polar membrane

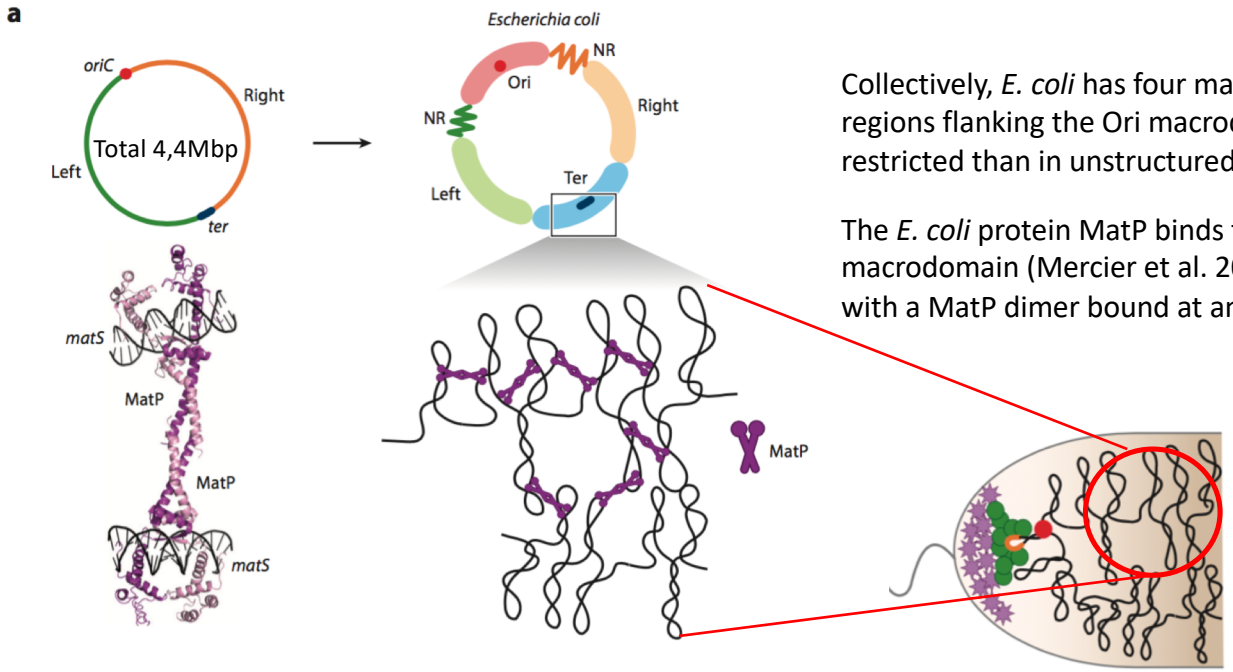
In *Vibrio cholerae*, a membrane-associated protein called HubP anchors the origin of the large chromosome, ChrI, to the pole. HubP interacts with ParAI, which likely interacts with ParBI, which in turn binds a *parS* site near the ChrI origin. HubP has a peptidoglycan-binding LysM domain, which is required for polar localization.

Note: Although PopZ, RacA, and HubP each anchor chromosomes to a cell pole, these proteins bear no sequence similarity, suggesting they arose independently

In *E. coli*, no polar anchoring complex has been identified. If one exists, it may function only during fast growth, when chromosomes exhibit an *ori-ter* pattern.

Macrodomains in bacteria

Bacterial chromosomes are further organized into Mb-sized domains called **macrodomains**, which were first suggested by FISH studies in *E. coli* that demonstrated certain loci frequently co-occupy the same restricted cytoplasmic space



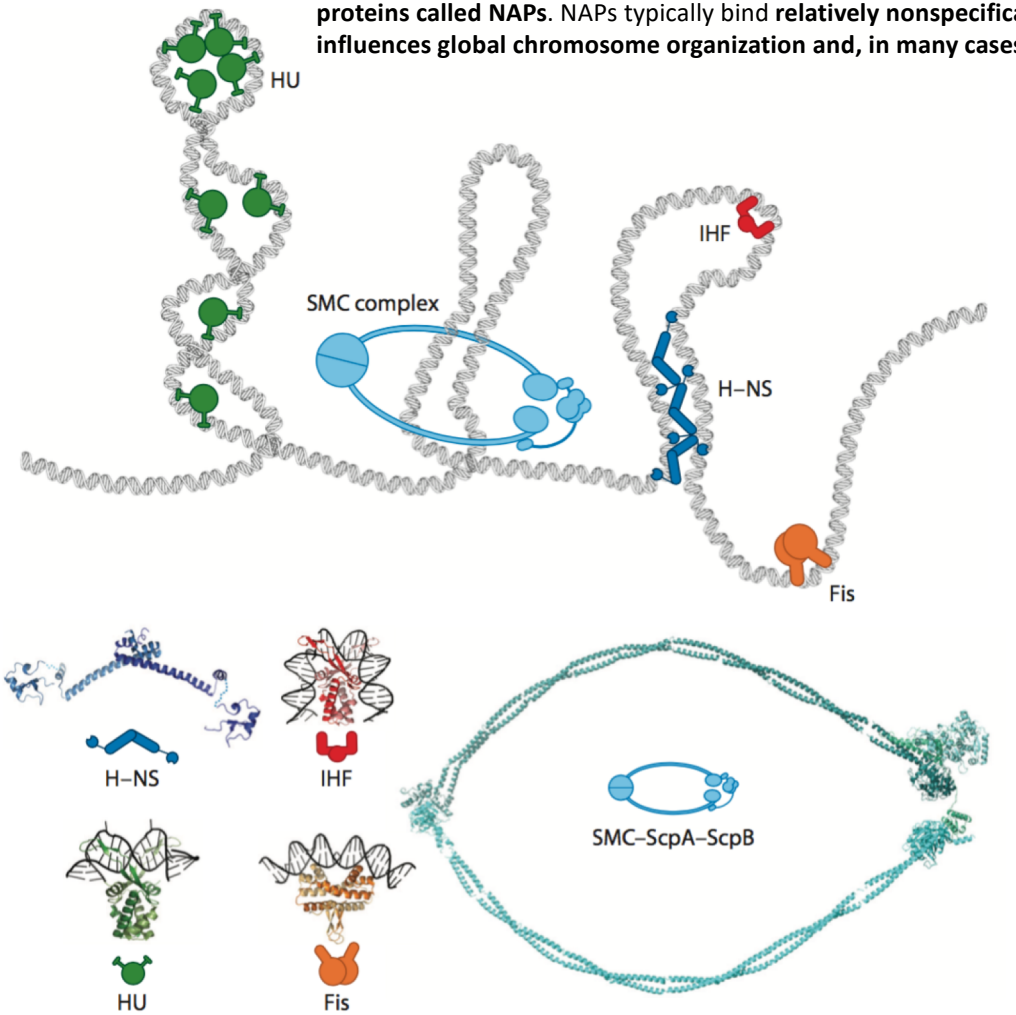
Collectively, *E. coli* has four macrodomains, Ori, Ter, Left, and Right, with two less-structured DNA regions flanking the Ori macrodomain (a). DNA movement within macrodomains is more restricted than in unstructured regions (NR)

The *E. coli* protein MatP binds to 13-bp *matS* sites present exclusively in the ~800-kb Ter macrodomain (Mercier et al. 2008). A MatP dimer bound to one *matS* site can form a tetramer with a MatP dimer bound at another site bringing different sites in the genome together.

Macrodomains and chromosomal interaction domains. (*a*) Macrodomain organization of the *Escherichia coli* chromosome, (*left*) or with the four macrodomains, Ori, Ter, Left, and Right, The crystal structure of two MatP dimers, each bound to a *matS* recognition site, is shown.

Nucleoid-Associated Proteins (NAP) in bacteria

The organization of bacterial chromosomes is profoundly influenced by DNA-binding proteins and in particular by **a heterogeneous class of abundant proteins called NAPs**. NAPs typically bind **relatively nonspecifically across bacterial genomes, wrapping, bending, or bridging DNA, ultimately influences global chromosome organization and, in many cases, transcriptional patterns.**



E. coli **H-NS** is a small (15.5 kDa) protein that can bridge DNA, **bringing loci** separated on the primary sequence level into close physical proximity**, constrains supercoiling**, can also **oligomerize and spread** along DNA to occlude binding sites for RNA polymerase or transcription activators (r**epressoion**)

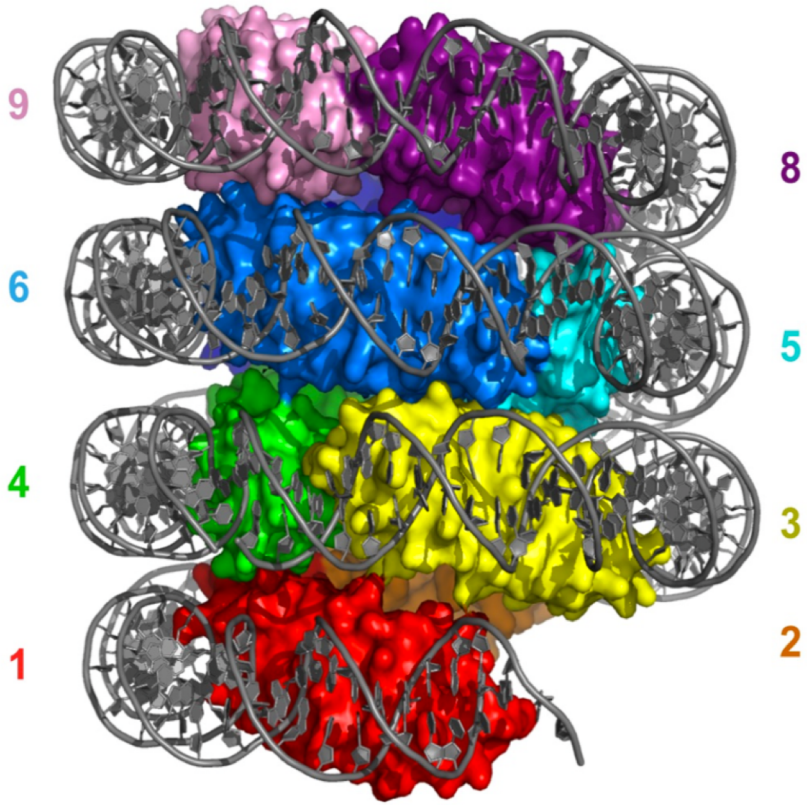
HU is another small (18 kDa), abundant (~30,000 copies/cell) NAP found in many bacteria that coats **and wraps chromosomal DNA around itself in a fashion grossly similar to that of histones**. Can coat 10% of DNA, aggregates, ha proposed role in DNA **compaction**, **promotes supercoiing**,

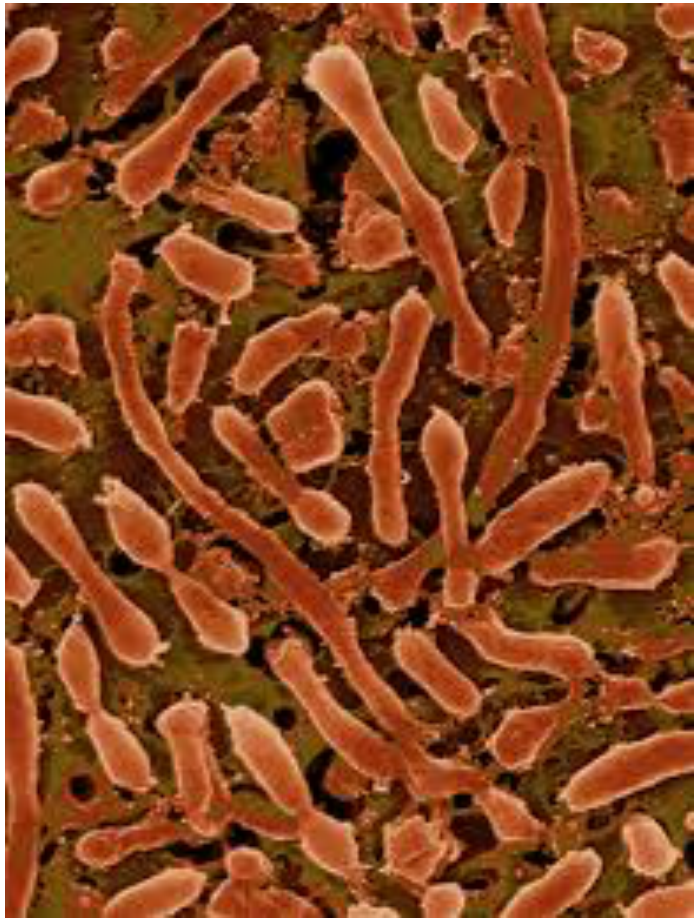
Integration host factor (**IHF**) and factor for inversion stimulation (**Fis**). Composed of two subunits, binds DNA more specifically and **introduces dramatic ~160◦ bends**. IHF alter DNA shape and facilitate **the formation of loops**, **frequently bringing RNA polymerase together with distant regulatory proteins**. IHF also impacts a range of other DNA-based processes, including **replication initiation and recombination**. **Fis** binds throughout the genome (Kahramanoglou et al. 2011), **impacting transcription, replication, and recombination. Modulates supercoiling**

SMC (125kDa) forms an extended, antiparallel coiled coil with a hinge domain at one end and an ATPase domain at the other. Homodimerization via the hinge domains creates a ring-like structure that **may encircle DNA**. SMC associates with two regulatory proteins, ScpA and ScpB, which likely modulate its ATPase activity, thereby affecting the opening and closing of the homodimeric ring. Mutations in SMC produce a range of chromosomal **defects in different bacteria, often including an increase in anucleate** cells. SMC likely contributes to both **chromosome segregation and chromosome compaction.**
HOMOLOG IN EUKARYOTES: CONDENSIN

Note: the function of these proteins was found by introducting mutations and study DNA topology/gene expression/recombination…

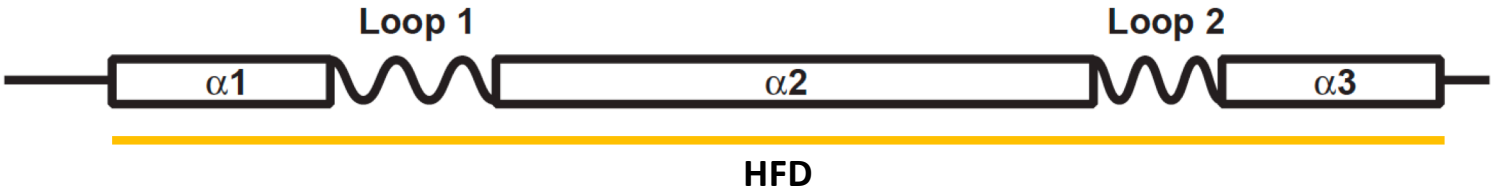
Genome organization in archaea bacteria

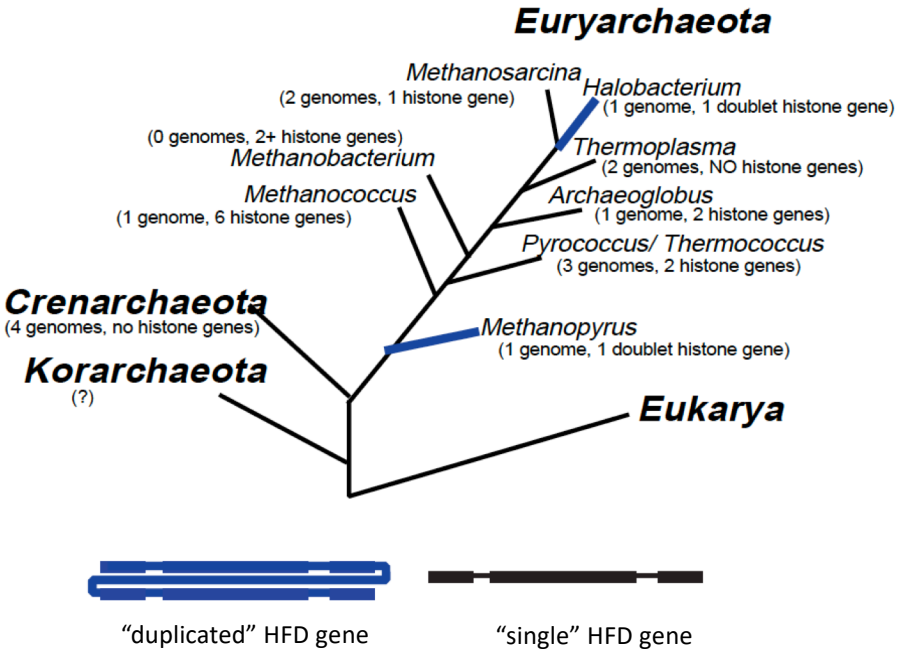




Halobacterium

The origin of histone proteins: Archaebacteria





"duplicated" HFD gene "single" HFD gene

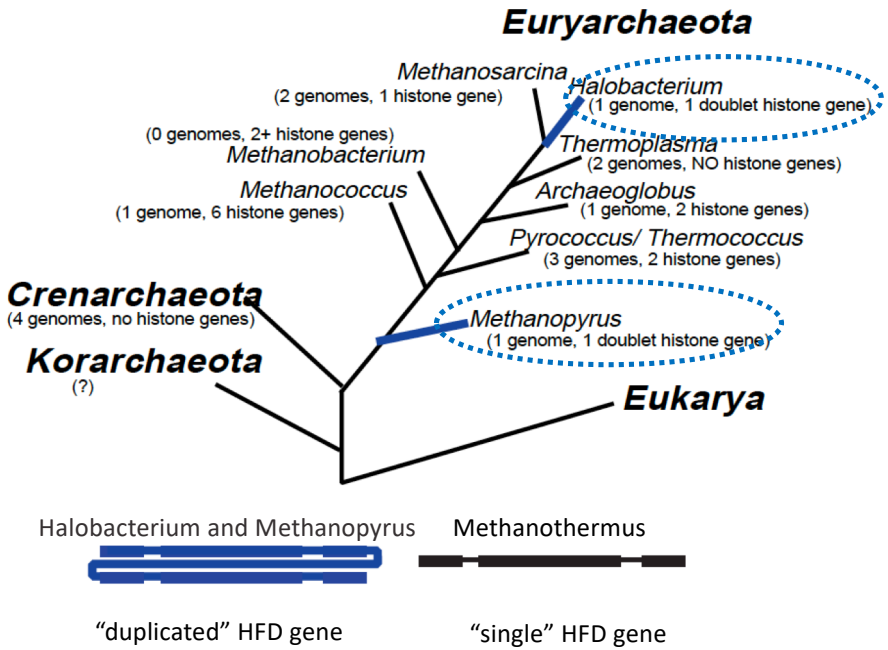
Histone proteins first evolved in Archaebacteria
→ Form «Archeal nucleosomes»

Histone fold (HFD) domain: dimerization and DNA binding

Most archaeal histones comprise a single histone fold domain (HFD), characterized by three α-helices, and two intervening loops, with no N-terminal or C-terminal 'tails'.

Malik 2013

The origin of histone proteins: Archaebacteria



The genome of Methanothermus fervidus encodes at least two distinct histones, **HMfA** (for histone M. fervidus A) and **HMfB**, which have been shown to compact DNA.

HMfA and HMfB form both, homodimers and heterodimers.

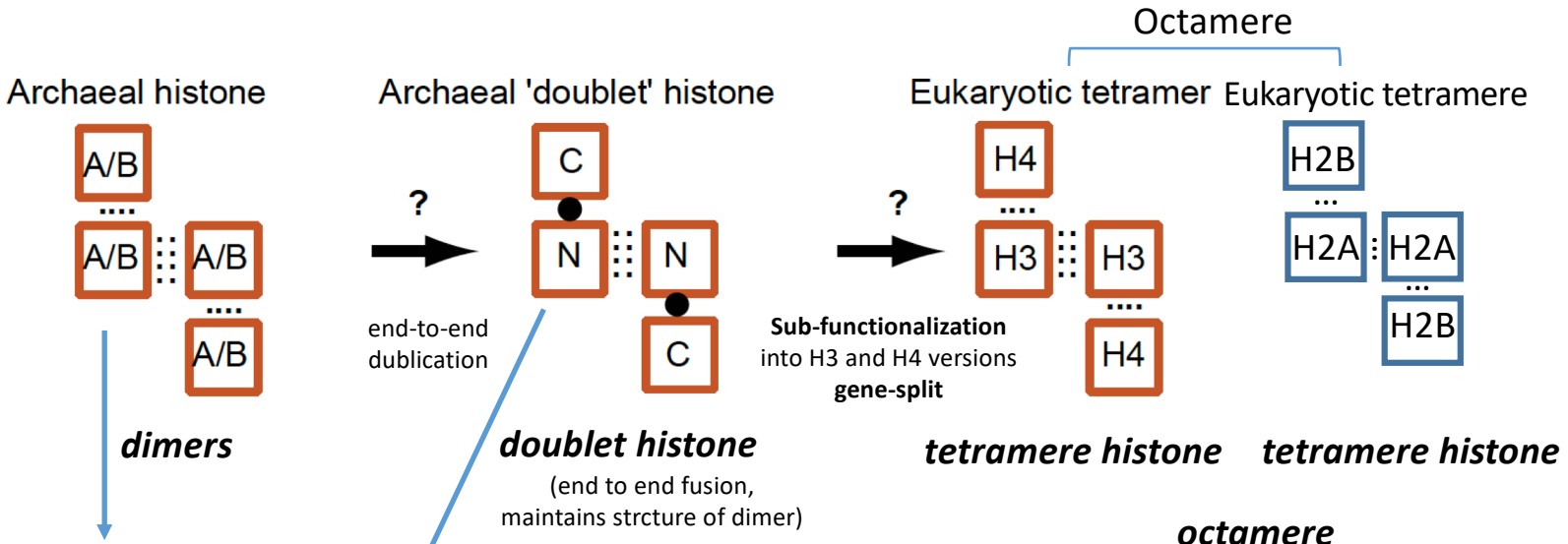
A tetramer of these histone proteins is able to protect 60 base pairs of packaged DNA from nuclease digestion, suggesting a **single wrap** of DNA around the tetramer

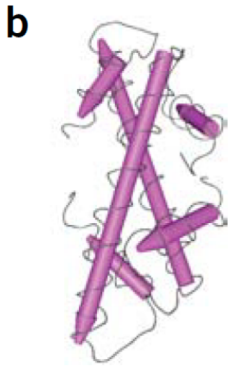
NOTE: Not all euryarchaeal lineages have a similar complement of two histones. The histone gene complement in completely sequenced euryarchaeal genomes varies from one to six genes

Halobacterium and Methanopyrus kandleri encode histones that are twice as long as typical archaeal histones and consist of an **end-to-end duplication** of the histone fold.
Evolutionary relevant: dublication requires less protein-protein interaction to form a «tetramere». N and C terminal portions can subfunctionalize

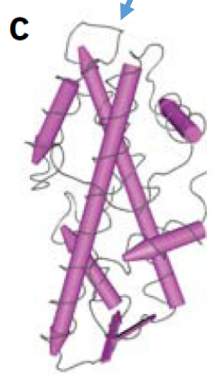
From Archaebacteria to Eukaryote histone tetramers

A model for the evolution of eukaryotic histones octamers containing 4 core histone components

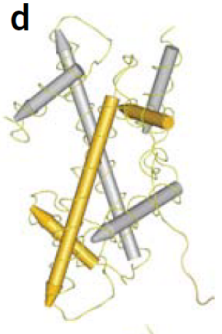




b) homodimer of HMfA from M. fervidus



c) doublet histone from M. kandleri – result of duplication of histone gene



d) H3–H4

e) H2A–H2B

The formation of the doublet histone directly resulted in an asymmetric (subfunctionalized N and C terminus) dimer that could have preceded the actual separation of the H3-H4 and H2A-H2B genes

Archaeal and eukaryotic histones

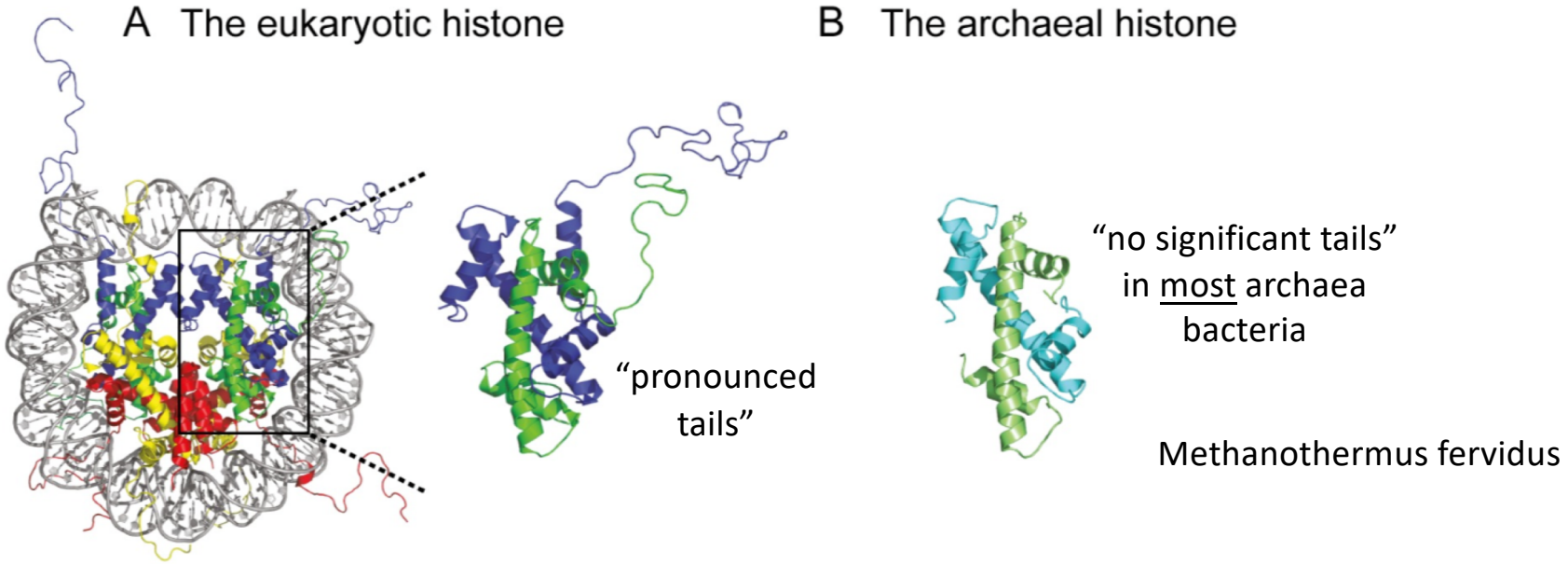


Fig 1. Eukaryotic and archaeal histones. (A) Eukaryotic nucleosome consisting of DNA wrapped around a core of a $(H3\text{-}H4)_2$ tetramer and two H2A-H2B dimers. Yellow, H2A; red, H2B; blue, H3; green, H4. (B) Archaeal histone homodimer of HMfB. HMfB, Histone B from *Methanothermus fervidus*.

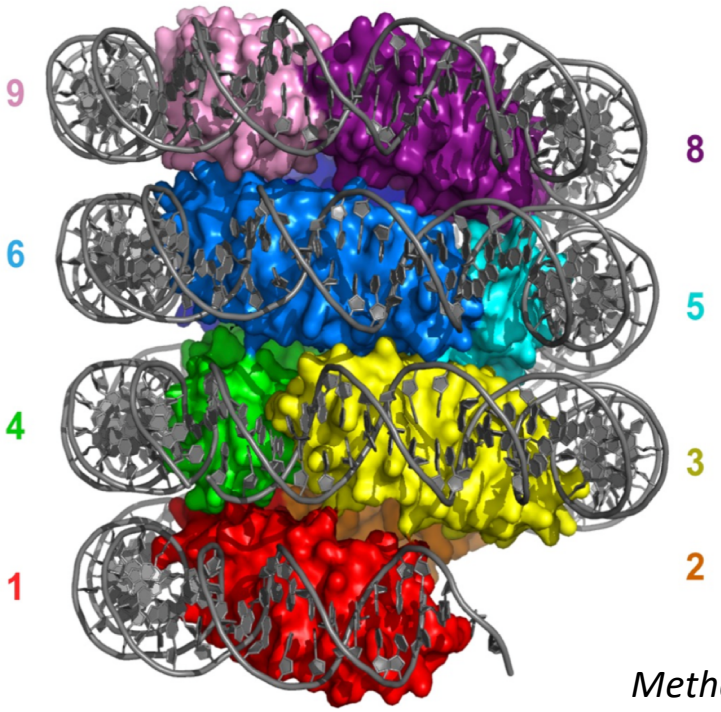
https://doi.org/10.1371/journal.pgen.1007582.g001

Although Archaea and Bacteria have common features, such as a circular genome and the absence of a nucleus, at the genetic level, Archaea seem to be more related to eukaryotes.

Amongst others, archaeal RNA polymerase, a key component of cellular life in all domains, is more similar to RNA polymerase from eukaryotes than bacterial RNA polymerase

The hypernucleosome in Archaea bacteria

Rod with 9 dimers



Histones (HMfb) assemble into an endless left-handed rod in vitro that bind DNA = “hypernucleosome”

DNA is wrapped around the protein rod

Histone–DNA complexes consist of discrete **multiples of a dimeric histone subunit that are not limited to dimers and tetramers** in vivo without obvious dependence on the DNA sequence

10-fold compaction

Methanothermus fervidus

Fig 2. Overview of the hypernucleosome structure. HMfB dimers stack to form a continuous, central protein core that wraps the DNA in a left-handed superhelix. Nine HMfB dimers are shown, each dimer in surface mode and in rainbow colors. Numbering indicates position of the nine histone dimers; note that dimer 5 and 6 occlude the view of dimer 7. DNA is in gray and shown as cartoon. *Image generated using PDB entry 5T5K* [64]. HMfB, Histone B from *Methanothermus fervidus*; PDB, Protein Data Bank.

https://doi.org/10.1371/journal.pgen.1007582.g002

Histone tails in Archaea bacteria hyper-nucleosomes

Most archaea bacteria do not show “no significant tails”

Exceptions exist: The tails of the two histones from *Heimdallarchaeota* and *Huberarchaea* are of roughly the same length and sequence composition as eukaryotic H4 tails

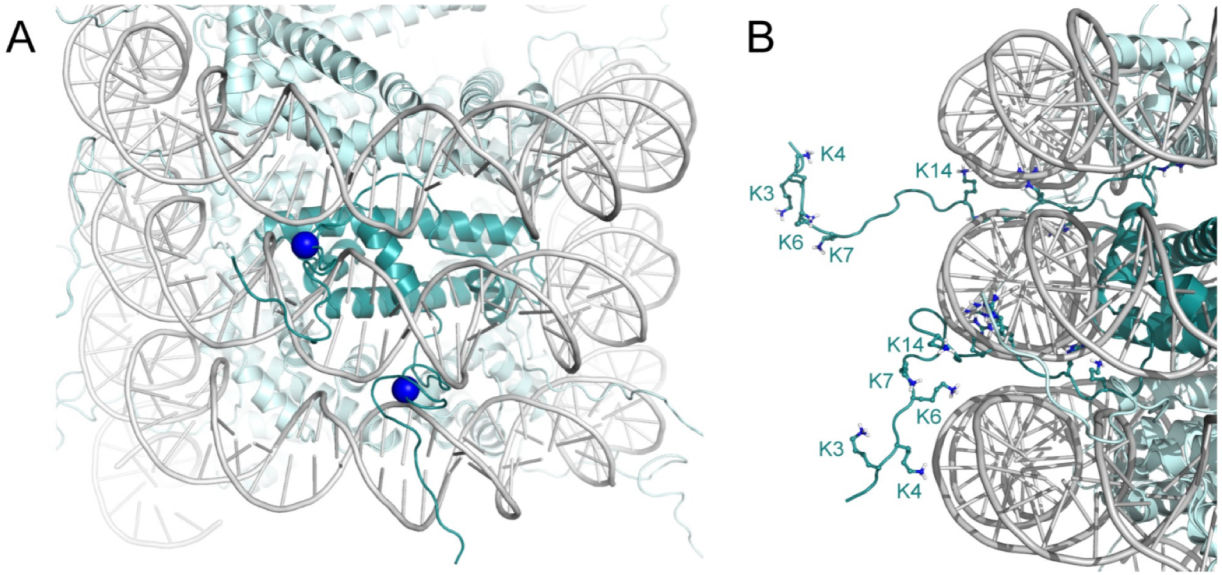


Fig 4. Model of a Heimdall LC_3 hypernucleosome with N-terminal tails. (A) View showing histone tails protruding through the DNA minor grooves. The R17 Cα-atom is shown as a blue sphere to mark the exit point of the tail. (B) Close up of the histone tails with lysine and arginine residues shown as sticks, and N-terminal lysines are labeled. Homodimers of Heimdall LC_3 histone HA are shown in teal; one dimer is highlighted in darker colors. Models are based on the structure of HMfB (PDB entry 5T5K); the tail in the top (bottom) of panel B is modeled in the H3 (H4) tail conformation (PDB entry 1KX5). HA, Histone A; HMfB, Histone B from *Methanothermus fervidus*; PDB, Protein Data Bank.

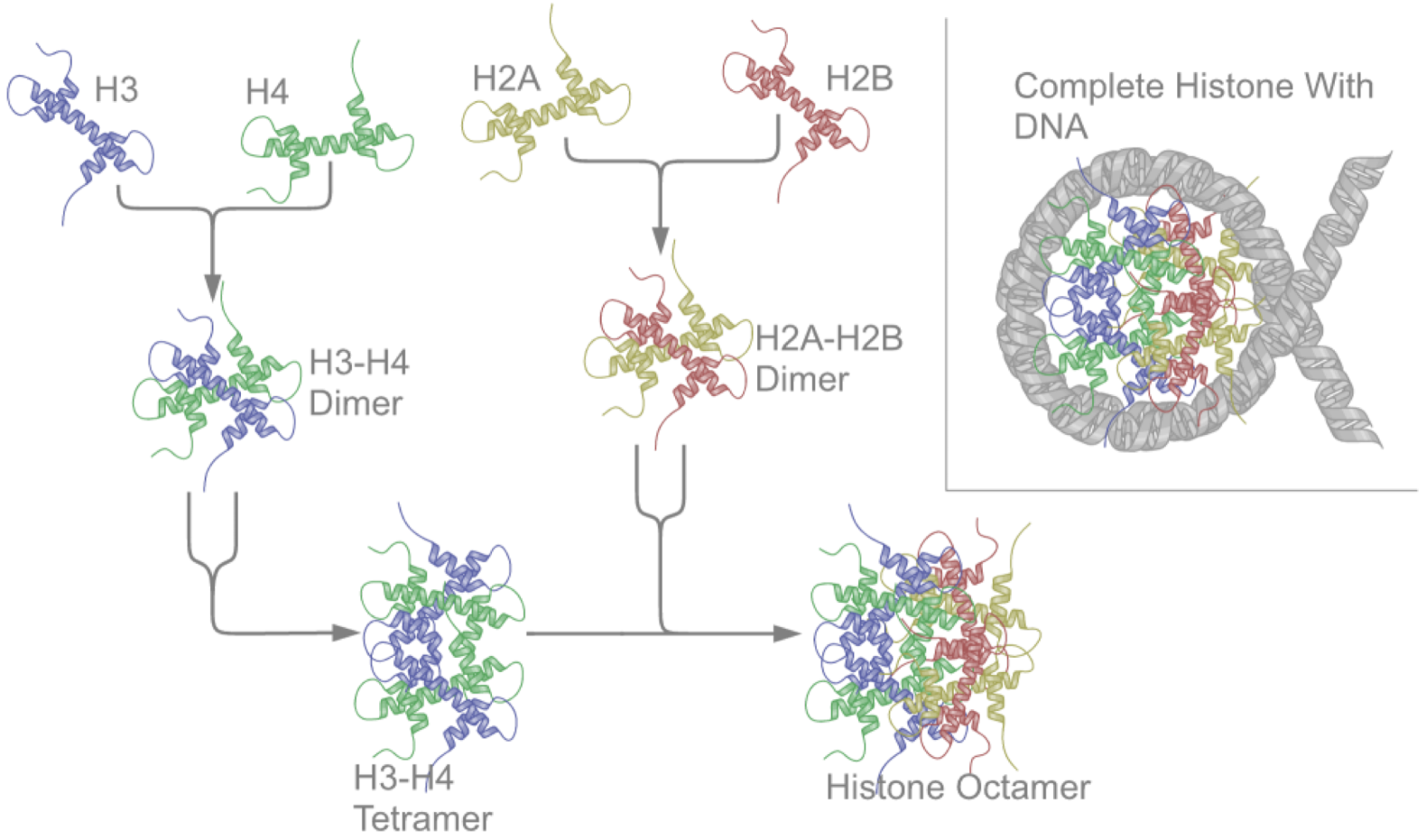
https://doi.org/10.1371/journal.pgen.1007582.g004

Three subsequent arginines (R17–R19) could facilitate passing of the tails through the DNA gyres.

The **tails exit the hypernucleosome through DNA minor grooves**, similar to eukaryotic histone tails, and might position their lysine side **chains to bind to the hypernucleosomal DNA or to other DNA close by**, facilitating (long-range) genomic interactions *in trans*.

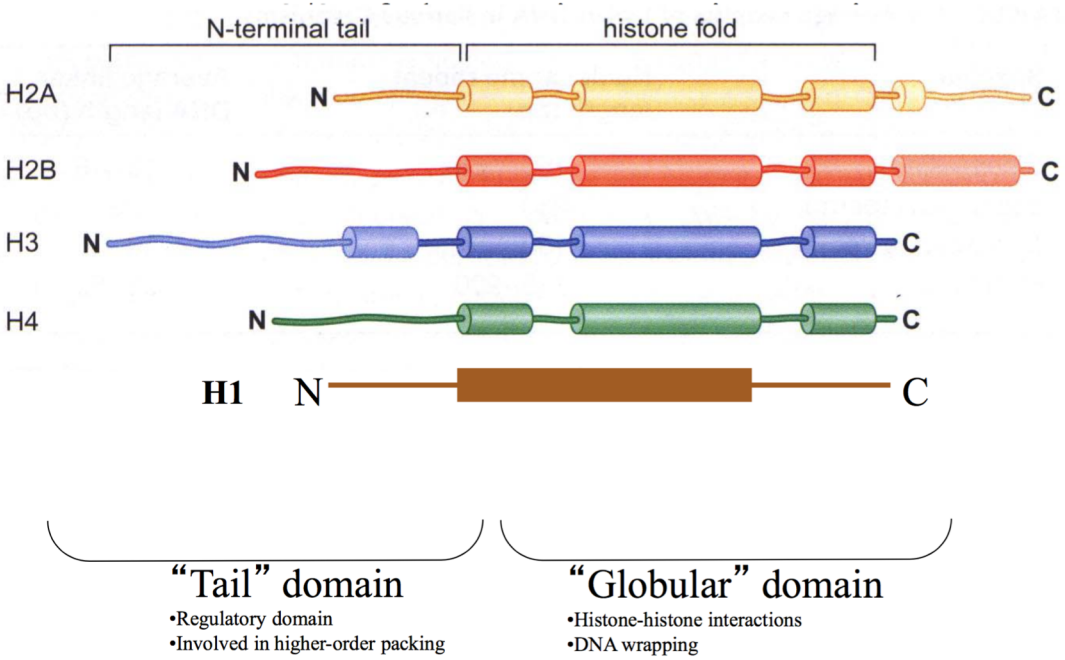
Like **the H4 tail that is subject to acetylation of lysines K5, K8, K12, and K16, lysines in the Heimdallarchaeal histone tail may well be subject to acetylation**. Archaeal genomes are known to **have several candidate lysine acetyltransferase and deacetylase enzymes**.

Heterodimerization + tetramerization = eukaryotic octamere formation



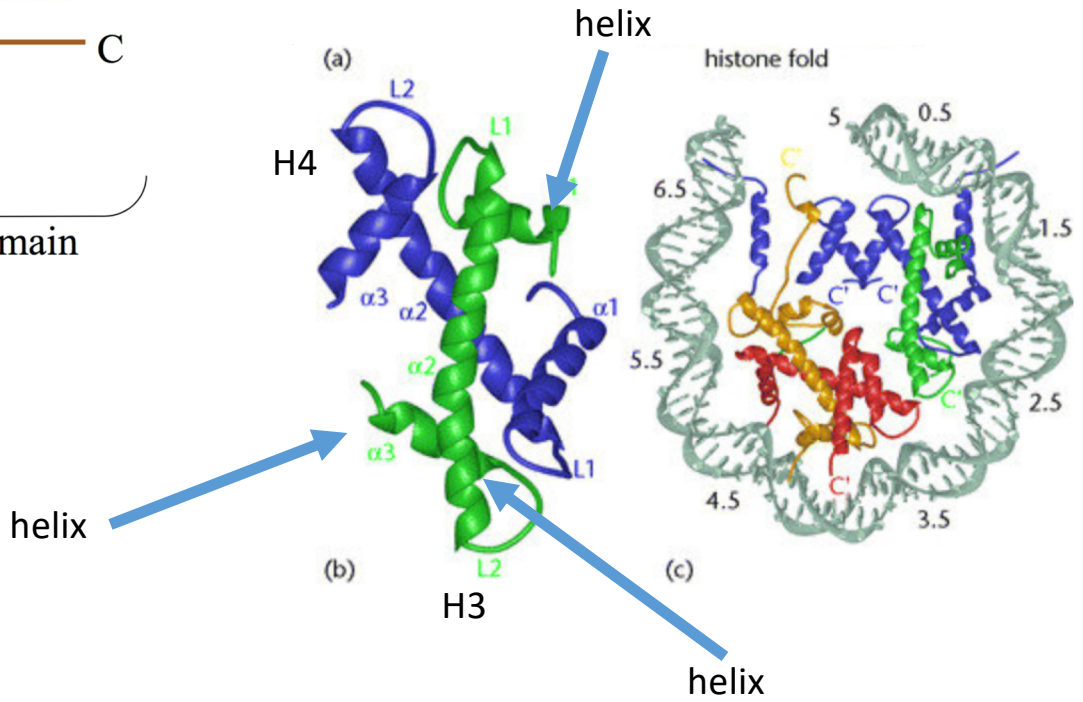
Nucleosome
2 wraps of DNA around Octamere:

Eukaryotic histones have key structural domains



The histone fold averages about 70 amino acids and consists of three alpha helices connected by two short, unstructured loops

'helix turn helix turn helix' motif (DNA-binding protein and protein interaction)



Properties of histone proteins; Interaction Histones - DNA

Types and Properties of Histones

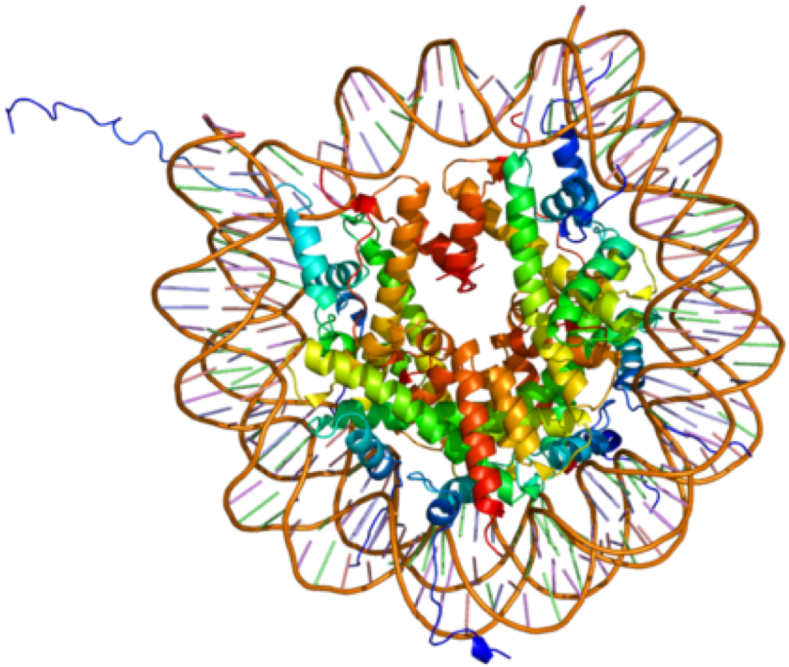
Histone	Molecular weight	Number of amino acid residues	Content of basic amino acids (% of total)	
			Lys	Arg
H1*	21,130	223	29.5	1.3
H2A*	13,960	129	10.9	9.3
H2B*	13,774	125	16.0	6.4
H3	15,273	135	9.6	13.3
H4	11,236	102	10.8	13.7

kDa

- **Helix-dipoles form alpha-helixes in H2B, H3, and H4 cause a net positive charge to accumulate at the point of interaction with negatively charged phosphate groups on DNA**
- Hydrogen bonds between the DNA backbone and the amide group on the main chain of histone proteins
- Non-polar interactions between the histone and deoxyribose sugars on DNA
- Salt bridges and hydrogen bonds between side chains of basic amino acids (especially lysine and arginine) and phosphate oxygens on DNA
- Non-specific **minor groove** insertions of the H3 and H2B N-terminal tails into two minor grooves each on the DNA molecule

Histone proteins are basic

- They contain many **positively-charged amino acids Lysine and arginine**
- These bind with the phosphates along the DNA backbone



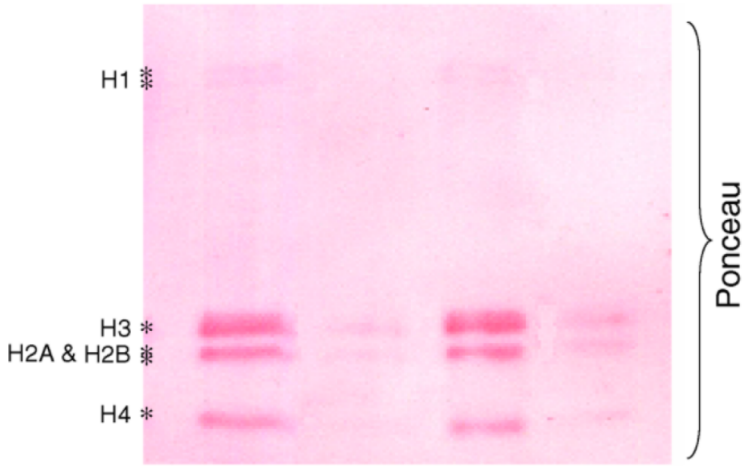
Properties of histone proteins; Interaction Histones - DNA

Meeting histones in lab → western blotting

Types and Properties of Histones

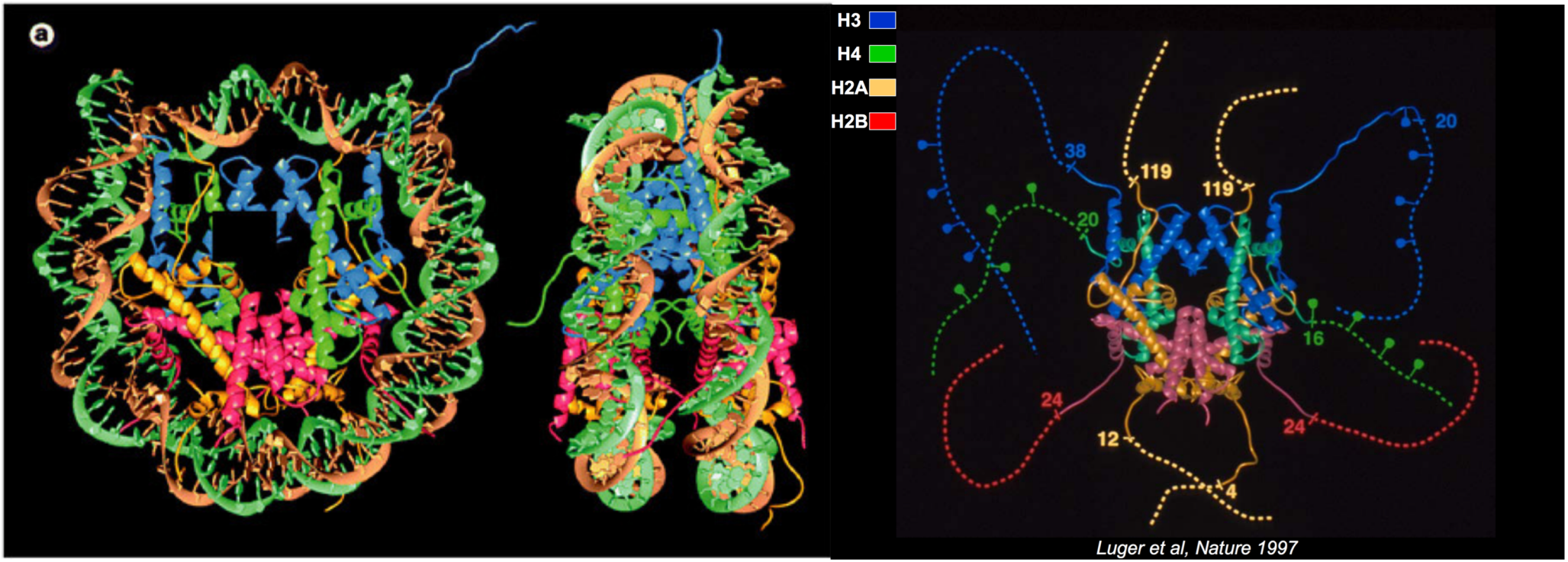
Histone	Molecular weight	Number of amino acid residues	Content of basic amino acids (% of total)	
			Lys	Arg
H1*	21,130	223	29.5	1.3
H2A*	13,960	129	10.9	9.3
H2B*	13,774	125	16.0	6.4
H3	15,273	135	9.6	13.3
H4	11,236	102	10.8	13.7

kDa



High % PAA gel; proteins transferred to western blotting membrane;
Stained with ponceau (labels proteins on membrane

The structure of the eukaryotic nucleosome



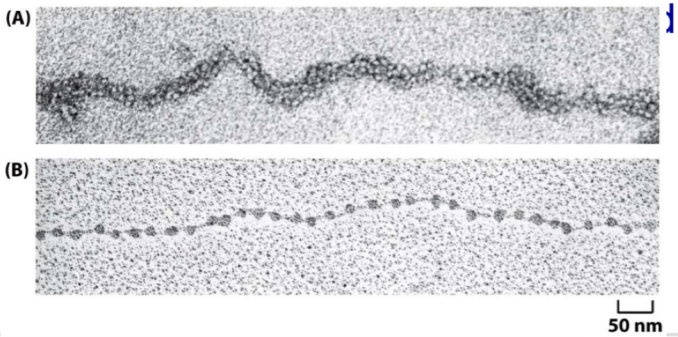
Luger et al., 1997

DNA Packaging by histones

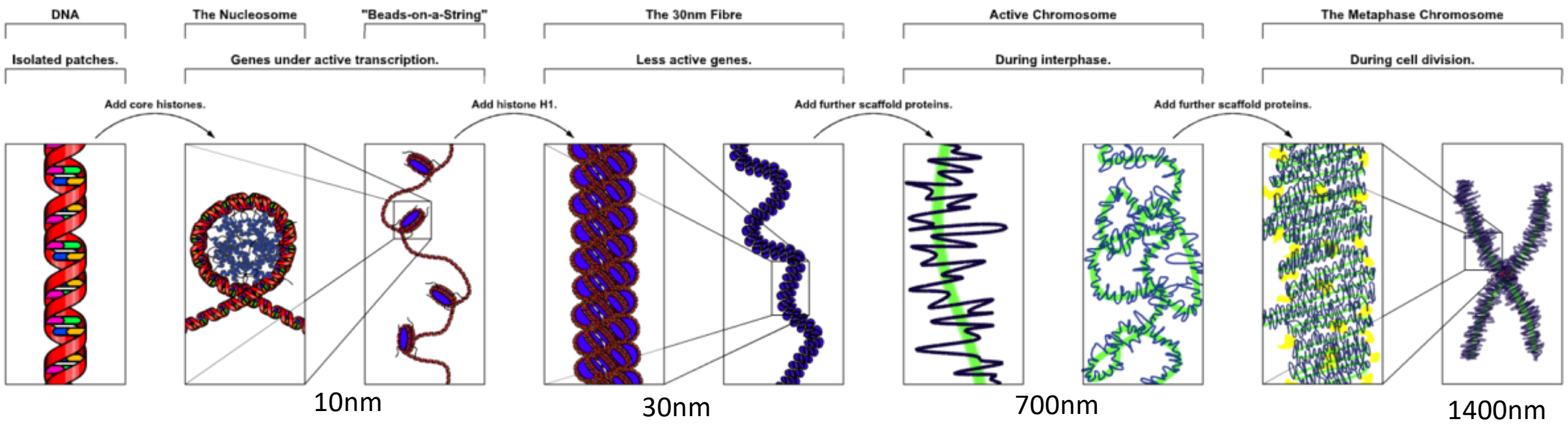
- Long sequence of DNA must be stored within the geometry of a nucleus

  - Example: human chromosome 22, 48 million bp
  - Extends to length of ~1.5 cm
  - Measures 2 μm in mitosis
  - Packaging ratio on the level of 10^4 in mitosis
  - Packaging ratio ~500 in interphase

- Packaged DNA must provide [illegible] for gene expression.

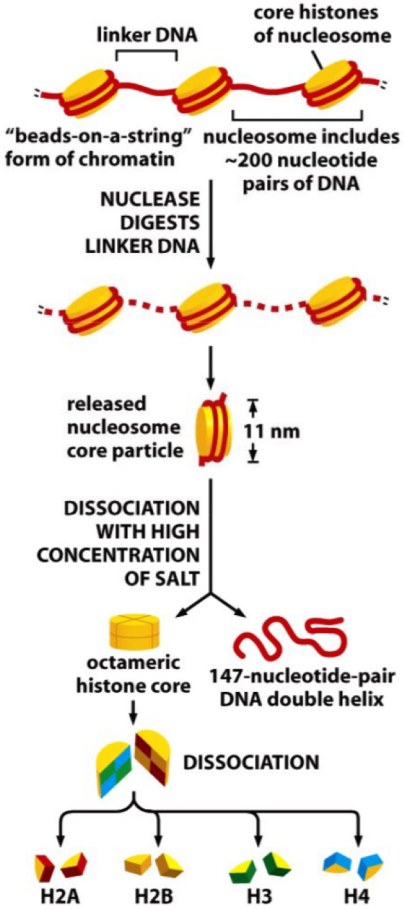


The major structures in DNA compaction



The 10nm fibre – Nucleosomes

- DNA is coiled around a protein core to form nucleosomes
- ~7 folds in packaging.
- Histone H2A, H2B, H3, H4 with 147 bp DNA.
- Nucleosomes repeat at every 200 bp. So ~30 million nucleosomes in a human cell.
- Total mass of histones approximately equal to that of DNA.



Nucleosomes are repeating units

Figure 19.7 Micrococcal nuclease digests chromatin in nuclei into a multimeric series of DNA bands that can be separated by gel electrophoresis. Photograph kindly provided by Markus Noll.

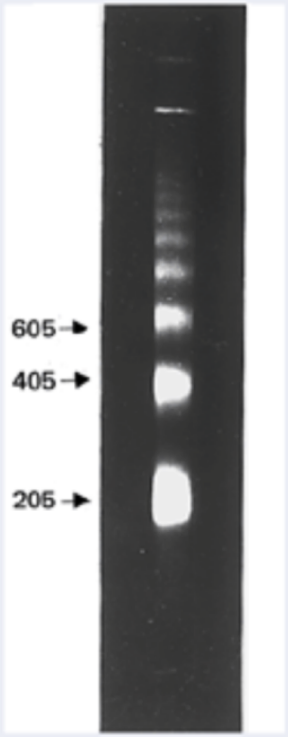
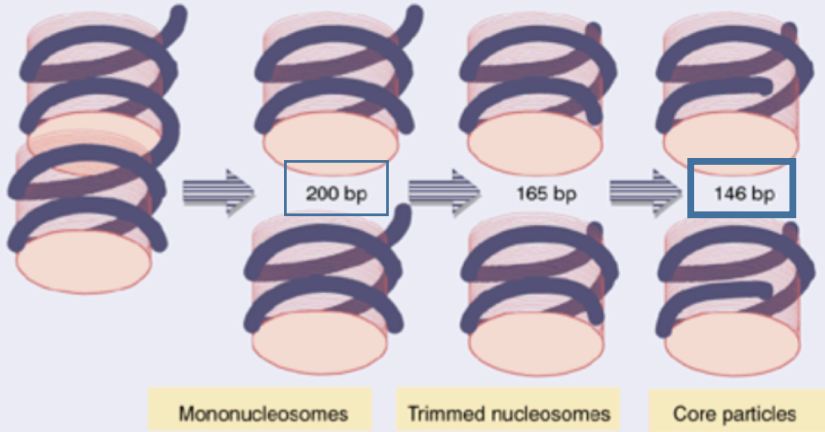
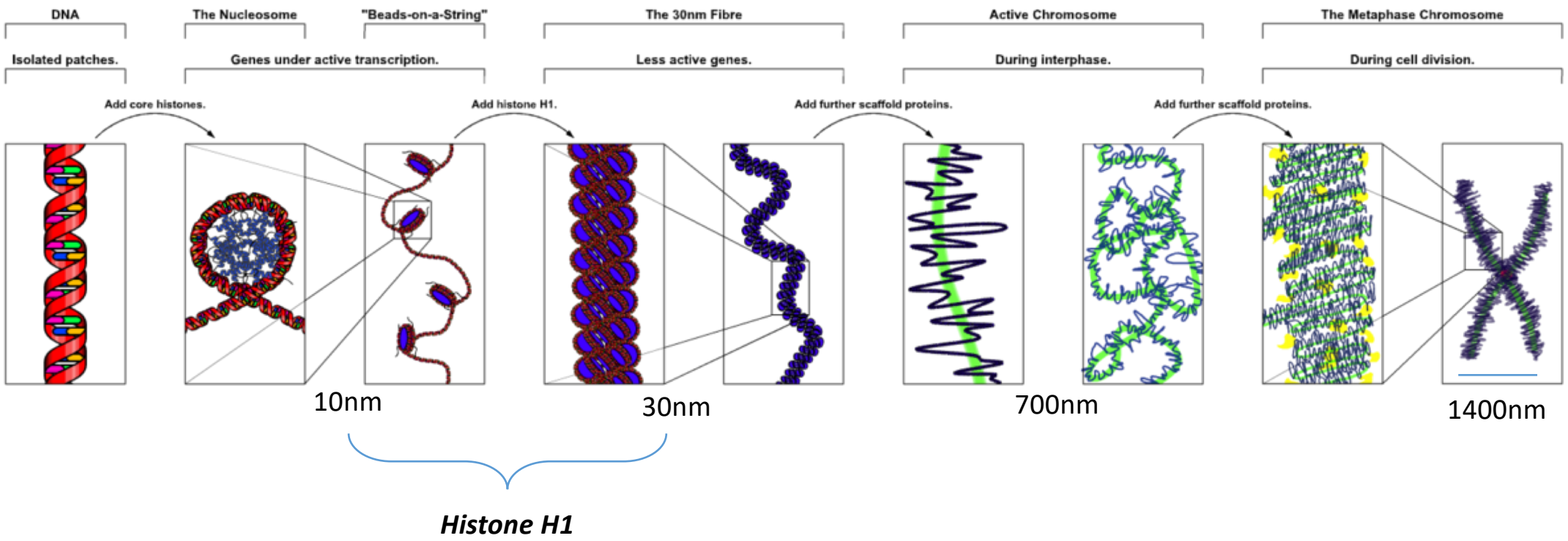


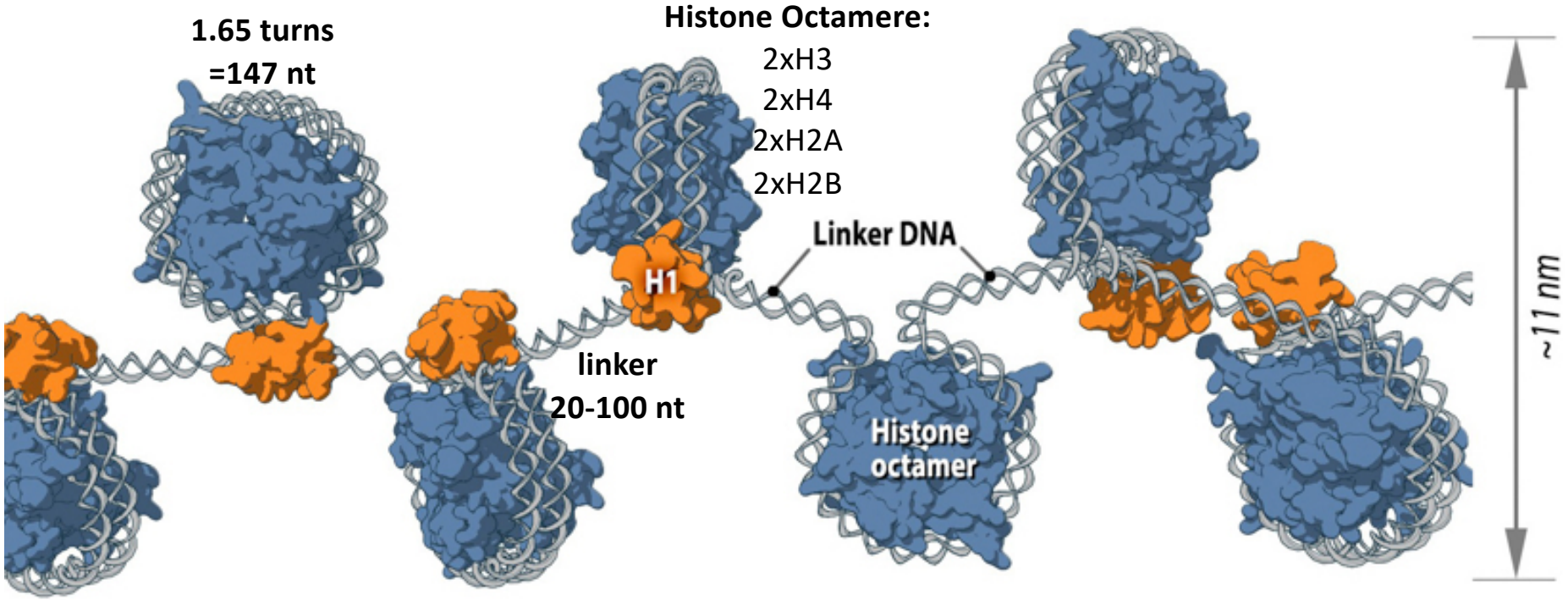
Figure 19.10 Micrococcal nuclease initially cleaves between nucleosomes. Mononucleosomes typically have ~200 bp DNA. End-trimming reduces the length of DNA first to ~165 bp, and then generates core particles with 146 bp.



The major structures in DNA compaction



Histone H1 compacts chromatin to the 30nm fibre



7-fold shortening of DNA

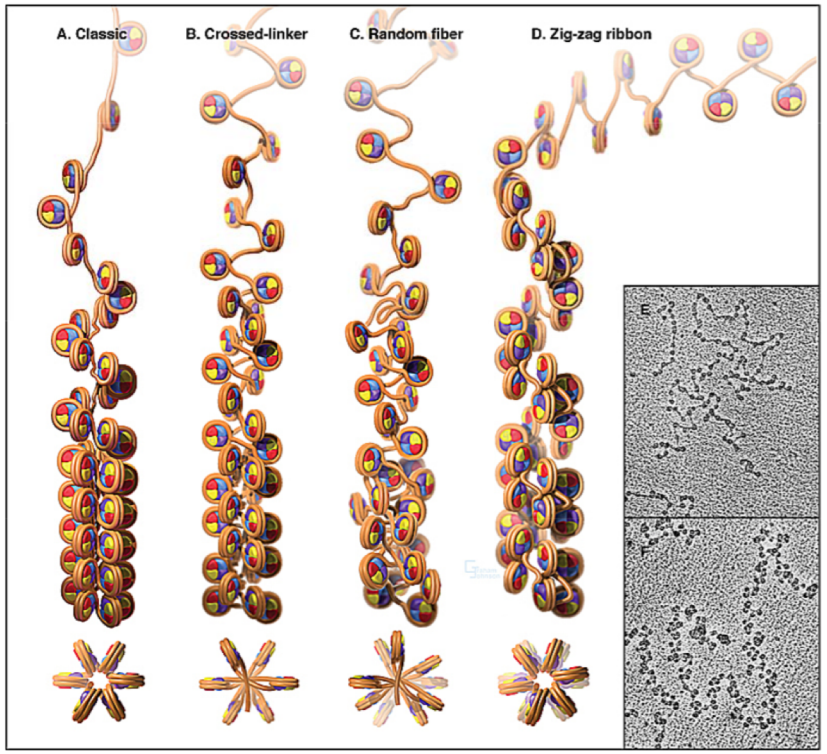
Nucleosome core particle: 147nt DNA + 2xH3, 2xH4, 2xH2A, 2xH2B

Histone Octamerne: 2xH3, 2xH4, 2xH2A, 2xH2B

Histone H1 compacts chromatin to the 30nm fibre

- Nucleosomes are further packaged into 30-nm fibers.
- The precise structure of the 30-nm fiber is not yet known.
- Chromatin structure beyond nucleosomes is poorly understood.
- ~40 folds in packaging.

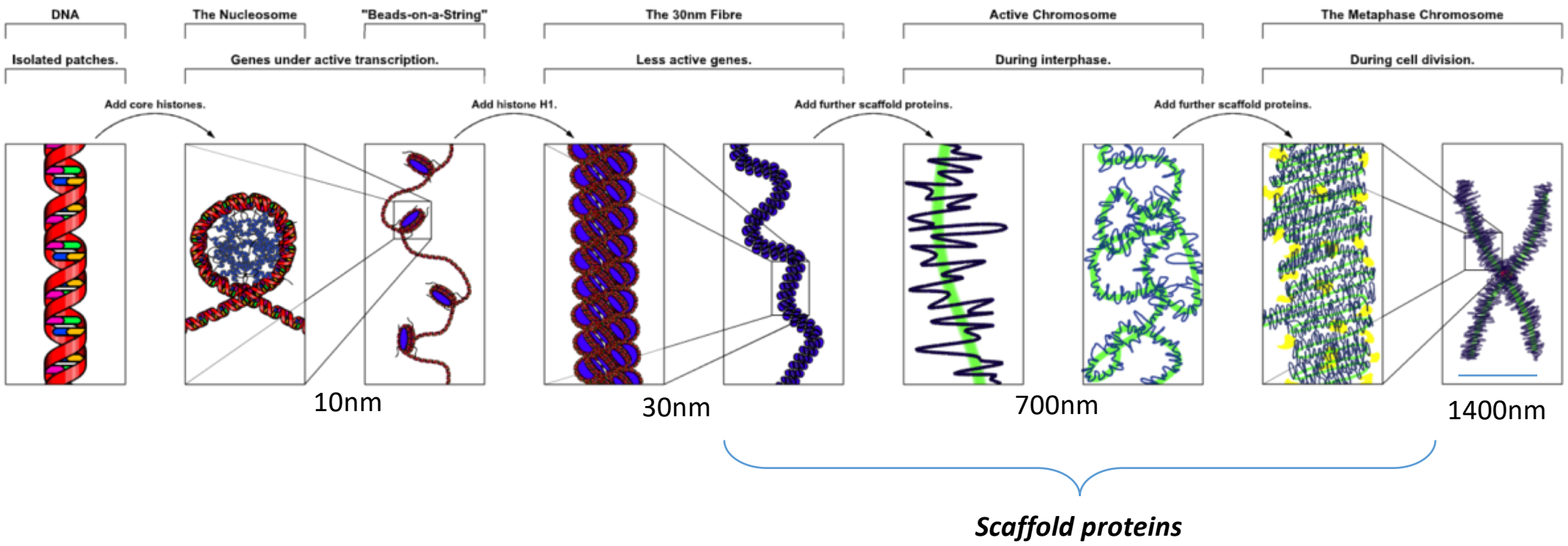
The structure of the 30-nm chromatin fiber remains unclear



Different models of the 30-nm fiber

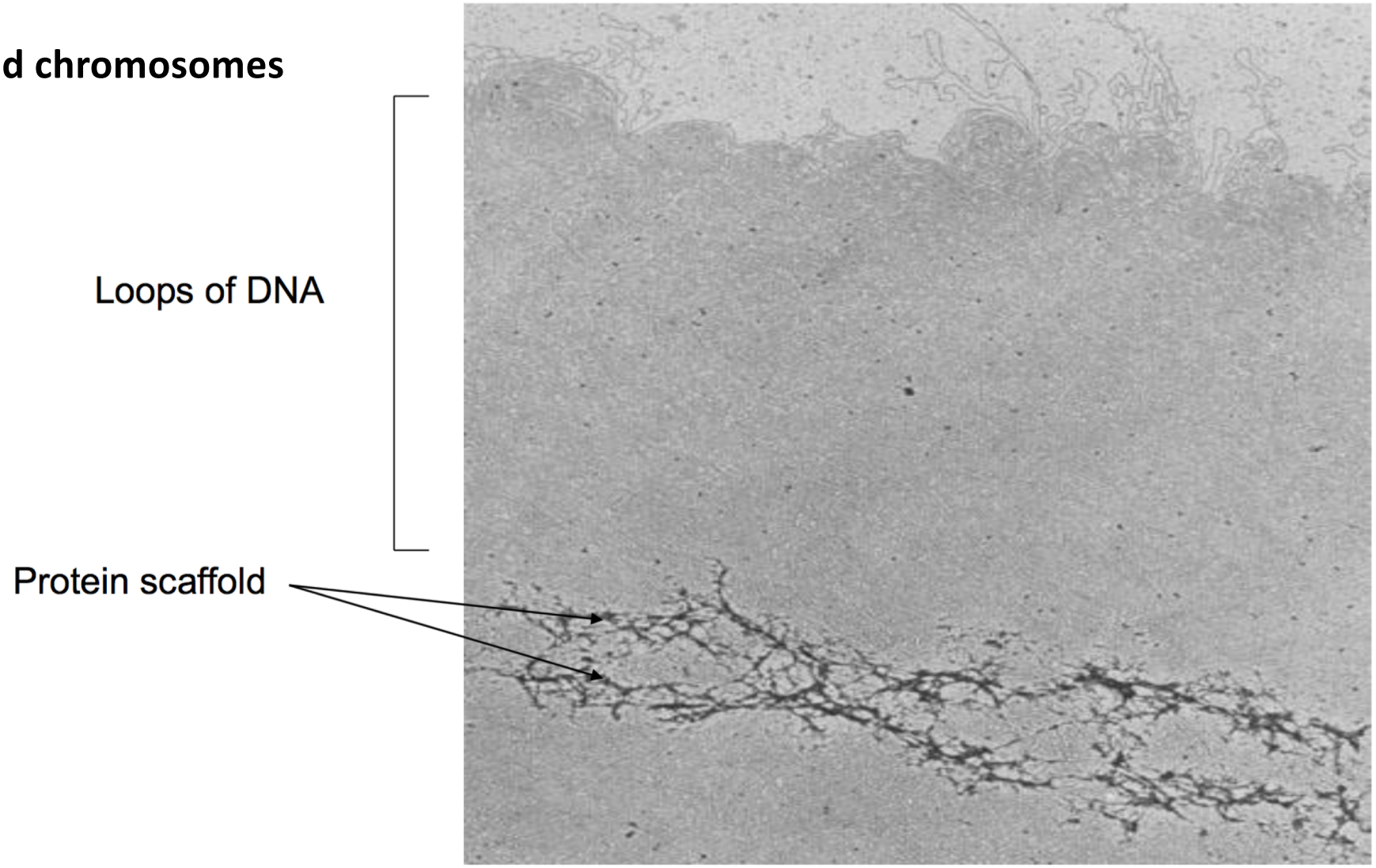
30-nm fiber most likely represents a delicate balance of different configurations that can be modified by a number of factors, including but not limited to, deposition of linker histones and high mobility group (HMG) proteins, incorporation of variant histones, modifications of histone tails, activity of chromatin remodeling factors, presence of phased nucleosome arrays

The major structures in DNA compaction

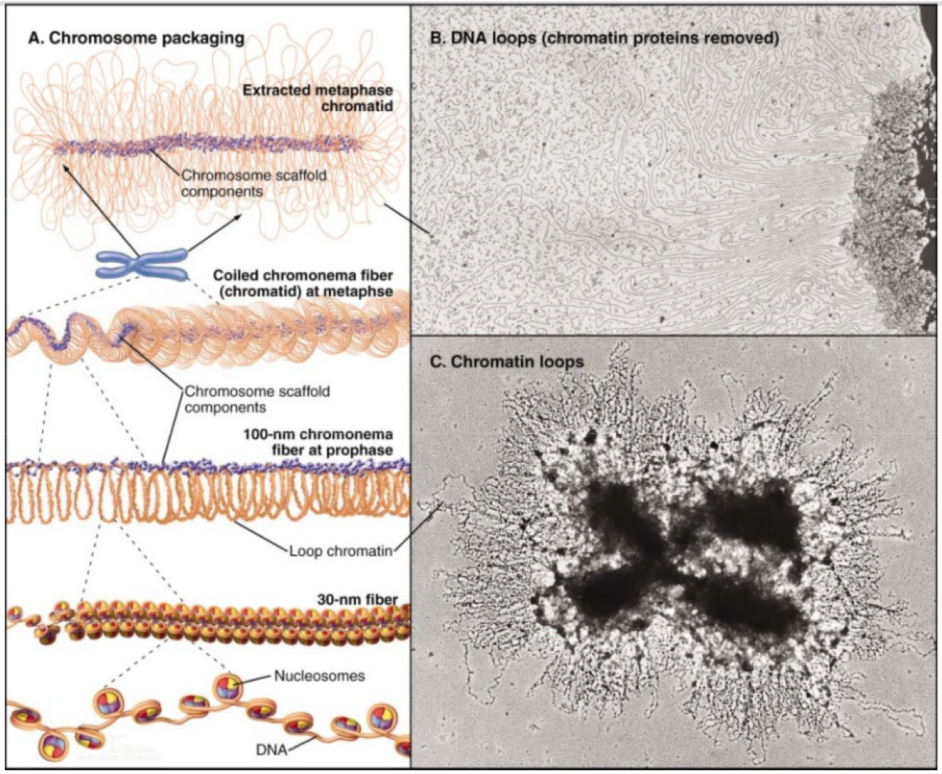


A protein scaffold is required to further compact the genome

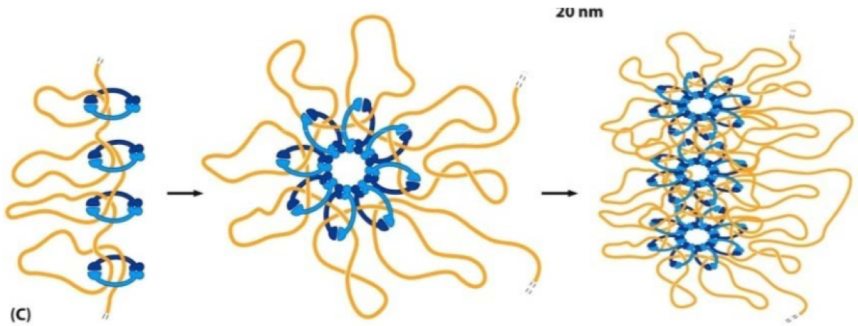
Histone depleted chromosomes



A protein scaffold is required to further compact the genome



Model of mitotic chromosomes

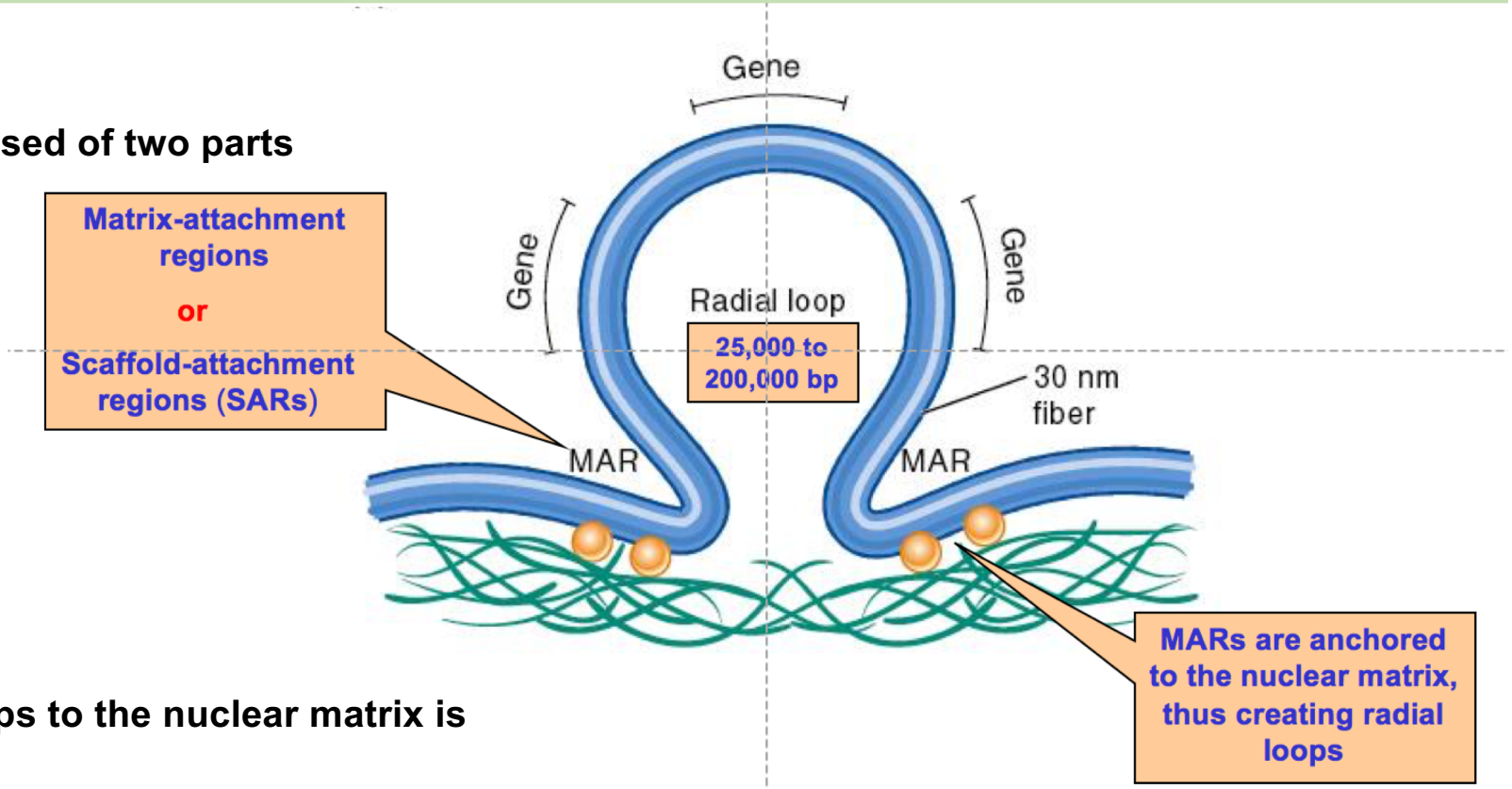


The axially-positioned chromosome scaffold of both chromatids mainly comprises non-histone proteins: so-called scaffold proteins, including **condensin**, **topoisomerase IIα** (Topo IIα) and kinesin family member 4

The nuclear mtatrix

The nuclear matrix is composed of two parts

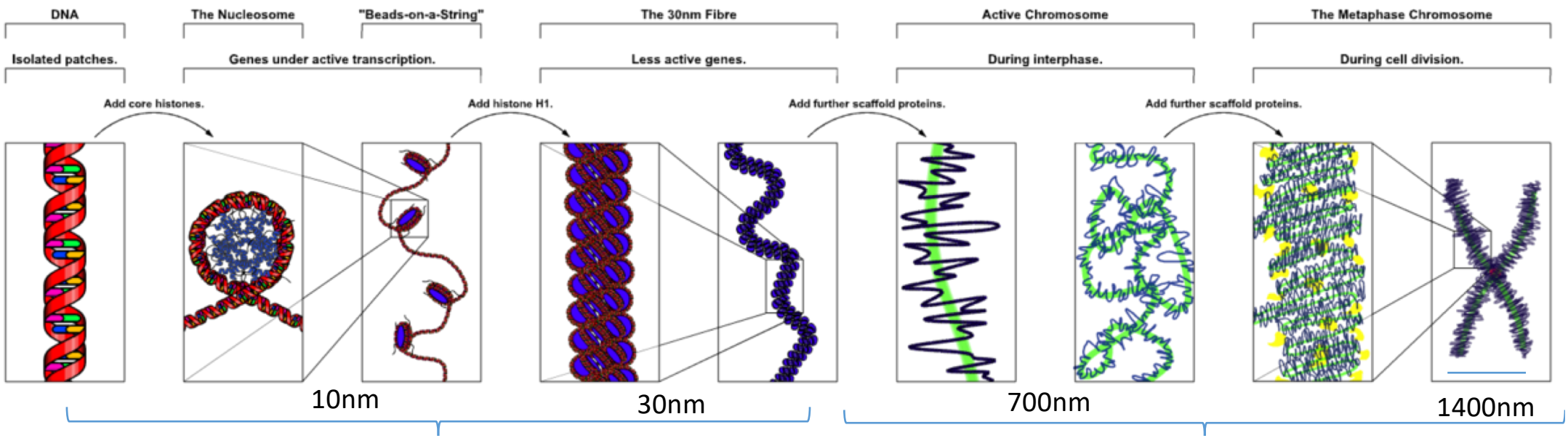
Nuclear lamina
Internal matrix proteins



The attachment of radial loops to the nuclear matrix is important in two ways

1. It plays a role in gene regulation
2. It serves to organize the chromosomes within the nucleus
   Each chromosome in the nucleus is located in a discrete and non-overlapping chromosome territory

The major structures in DNA compaction



Transcription **No transcription**

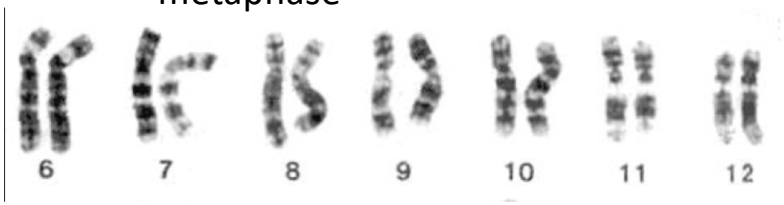
Transitions between the 30-nm and 10-nm fibers are thought to be essential for the control of chromatin transcriptional status.

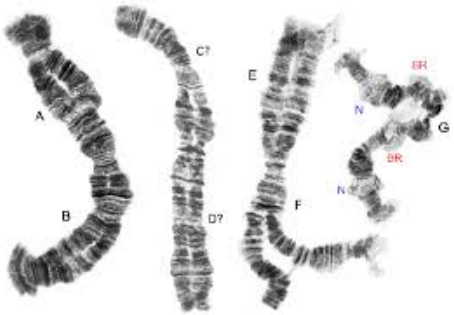
Euchromatin **Facultative heterochromatin** **Consitutive heterochromatin**

Chromatin comes in different flavors

Different types of chromatin

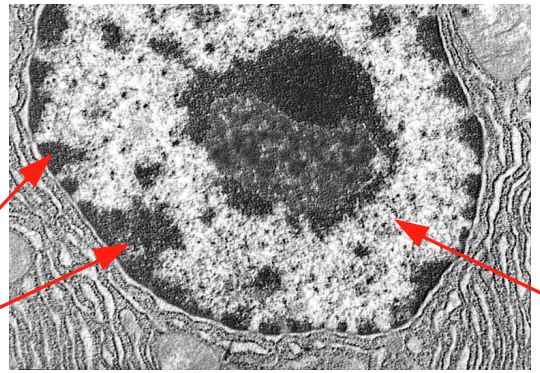
metaphase





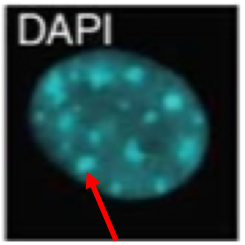
Giemsa

Interphase



heterochromatin

euchromatin
(and facultative
heterochromatin)



Chromocenter
(aggregates of centromeres
= constitutive heterochromatin)

Constitutive heterochromatin:

- constitute ~ 10% of nuclear DNA; telomeres, centromeres, and a considerable fraction of repetitive sequences
- highly compacted, replicates late in S phase, (transcriptionally inert)

Euchromatin + facultative heterochromatin:

- constitute ~ 90% of nuclear DNA
- less condensed, rich in genes, replicates early in S phase
  however,
- only small fraction of euchromatin is transcriptionally active
- the rest is transcriptionally inactive/silenced (but can be activated in certain tissues or developmental stages) → these inactive regions are also known as "facultative heterochromatin"

Histone gene clusters in vertebrates

Histone genes are organized in histone gene clusters

Human:

Chr.6: **major histone gene locus: HIST1 cluster: 45 core histone genes ; 6 Histone H1 genes**

Chr.1**: minor histone gene locus HIST2 cluster: 6 histone genes**

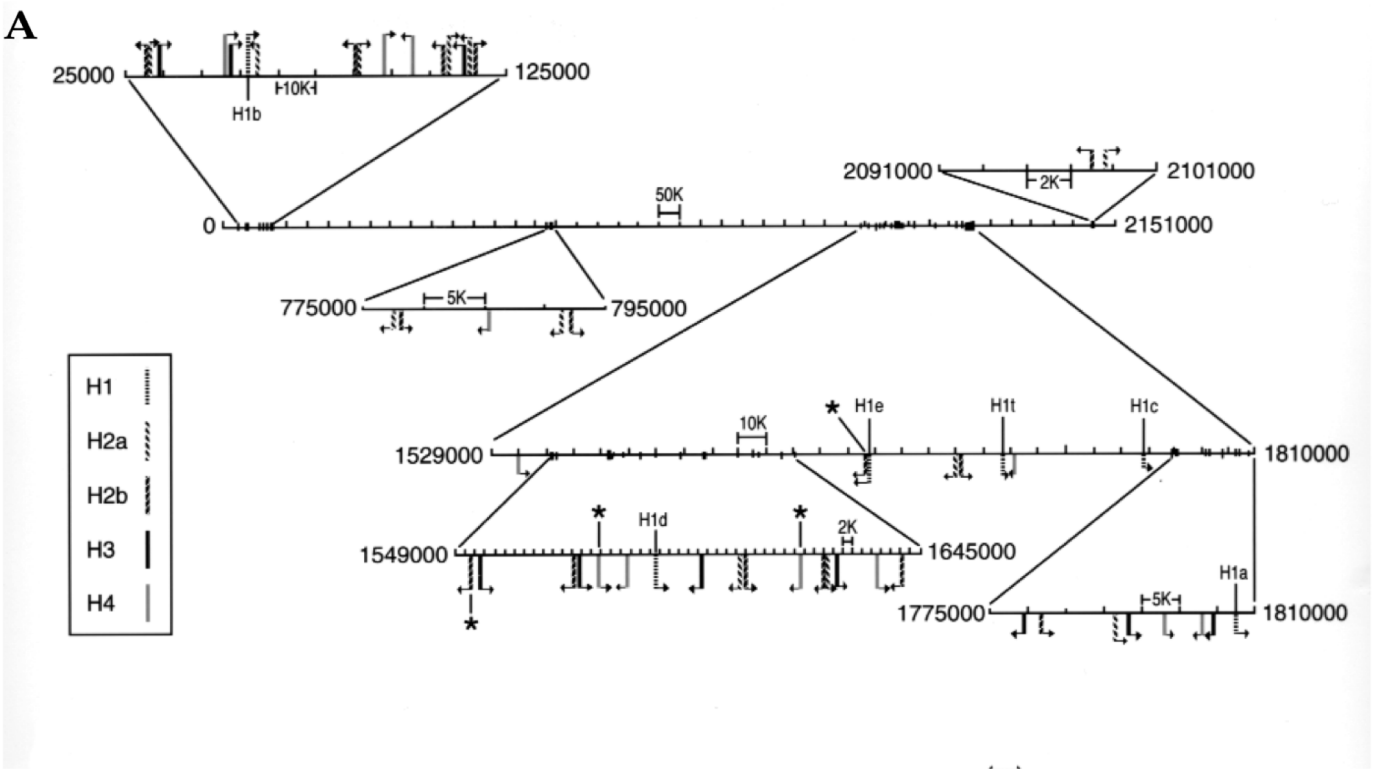
Chr.1**: minor histone gene locus HIST3 cluster: 3 histone genes**

Yeast:

2 copies for each core histone

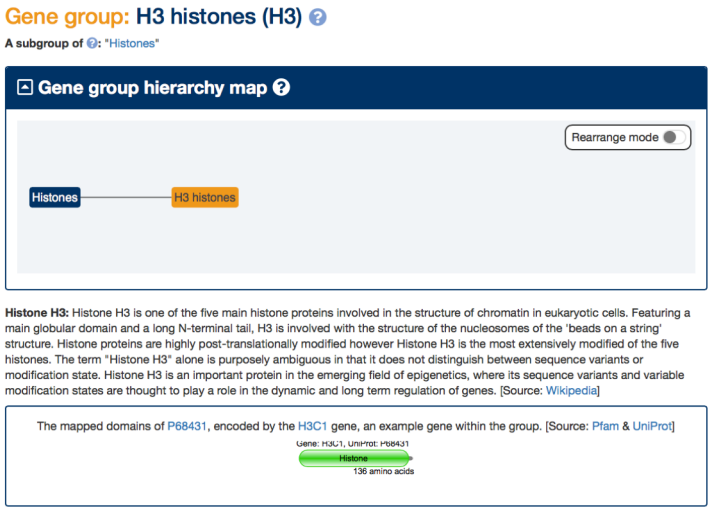
Sear Urchin:

More than copies for each core histone



The human and mouse major histone gene cluster. (A) The histone gene cluster on human chromosome 6p21–p22 is shown. The position and direction of transcription of the 55 histon genes in this region are indicated, with the genes for the five histone pro- teins indicated in the box. Only “real” genes are shown (defined as genes that contain the expected 3 end of histone mRNA). The portion of chromosome 6 is going (left to right) from the cen- tromere to telomere. HISTH4A is the first H4 gene starting from the right and the same is true for the other core histone genes. The numbers are nucleotides from the arbitrary start of the cluster at 0. The regions where there are tightly grouped clusters of histone genes have been expanded. The scale of each section is indicated in kilobases (kb). The position of each of the histone H1 histone gene is indi- cated with the nomenclature H1a–e, H1t, and the symbols for the core his- tone genes are in the inset. The aster- isks indicate the position of genes pres- ent in human and not in mouse.

Histone gene clusters in vertebrates



HGNC ID (gene)	Approved symbol	Approved name	Previous symbols	Aliases	Chromosome
HGNC:4766	H3C1	H3 clustered histone 1	H3FA, HIST1H3A	H3/A	6p22.2
HGNC:4776	H3C2	H3 clustered histone 2	H3FL, HIST1H3B	H3/l	6p22.2
HGNC:4768	H3C3	H3 clustered histone 3	H3FC, HIST1H3C	H3/c, H3.1	6p22.2
HGNC:4767	H3C4	H3 clustered histone 4	H3FB, HIST1H3D	H3/b	6p22.2
HGNC:54427	H3C5P	H3 clustered histone 5, pseudogene			6p22.2
HGNC:4769	H3C6	H3 clustered histone 6	H3FD, HIST1H3E	H3/d, H3.1	6p22.2
HGNC:4773	H3C7	H3 clustered histone 7	H3FI, HIST1H3F	H3/i	6p22.2
HGNC:4772	H3C8	H3 clustered histone 8	H3FH, HIST1H3G	H3/h	6p22.2
HGNC:18982	H3C9P	H3 clustered histone 9, pseudogene	HIST1H3PS1	dJ45P21.6, H3F3AP1, p36	6p22.2
HGNC:4775	H3C10	H3 clustered histone 10	H3FK, HIST1H3H	H3/k, H3F1K	6p22.1
HGNC:4771	H3C11	H3 clustered histone 11	H3FF, HIST1H3I	H3/f, H3.f	6p22.1
HGNC:4774	H3C12	H3 clustered histone 12	H3FJ, HIST1H3J	H3/j	6p22.1
HGNC:25311	H3C13	H3 clustered histone 13	HIST2H3D		1q21.2
HGNC:20503	H3C14	H3 clustered histone 14	H3F2, H3FM, HIST2H3C	MGC9629, H3/m, H3, H3.2, H3/M	1q21.2
HGNC:20505	H3C15	H3 clustered histone 15	HIST2H3A	H3/n, H3/o	1q21.2
HGNC:43735	H3Y1	H3.Y histone 1		H3.Y, H3.Y.1	5p15.1
HGNC:43734	H3Y2	H3.Y histone 2		H3.X, H3.Y.2	5p15.1
HGNC:1851	CENPA	centromere protein A		CENP-A, CenH3	2p23.3
HGNC:32060	H3-2	H3.2 histone (putative)	HIST2H3PS2	p06	1q21.1
HGNC:4764	H3-3A	H3.3 histone A	H3F3, H3F3A	H3.3A	1q42.12
HGNC:4765	H3-3B	H3.3 histone B	H3F3B	H3.3B	17q25.1
HGNC:4778	H3-4	H3.4 histone	H3FT, HIST3H3	H3t, H3/g, H3.4	1q42.13
HGNC:33164	H3-5	H3.5 histone	H3F3C	H3.5	12p11.21

Canonical H3

Histone H3 variants with special function

In vertebrates, there are a total of 10–20 genes encoding each of the core histone proteins.

Each of these genes encodes a unique mRNA, with a **distinct 5 - and 3 -UTR**, as well as **nucleotide changes** in the coding region.

All the histone **H4 genes encode the same protein**, **histone H3 genes** encodes variants with aa changes, there **10–12 different H2a and H2b** proteins are known

https://www.genenames.org/cgi-bin/genefamilies/set/864

Coordinated control of histone gene expression and incorporation into DNA

1. Histone and cell cycle regulation:

Maintaining a stable and balanced histone pool is of vital importance for appropriate gene regulation, cell cycle progression and genome stability. Excess of free histones is lethal → precise control of histone gene expression required → cell cycle

2. Controlled production of nucelosomes:

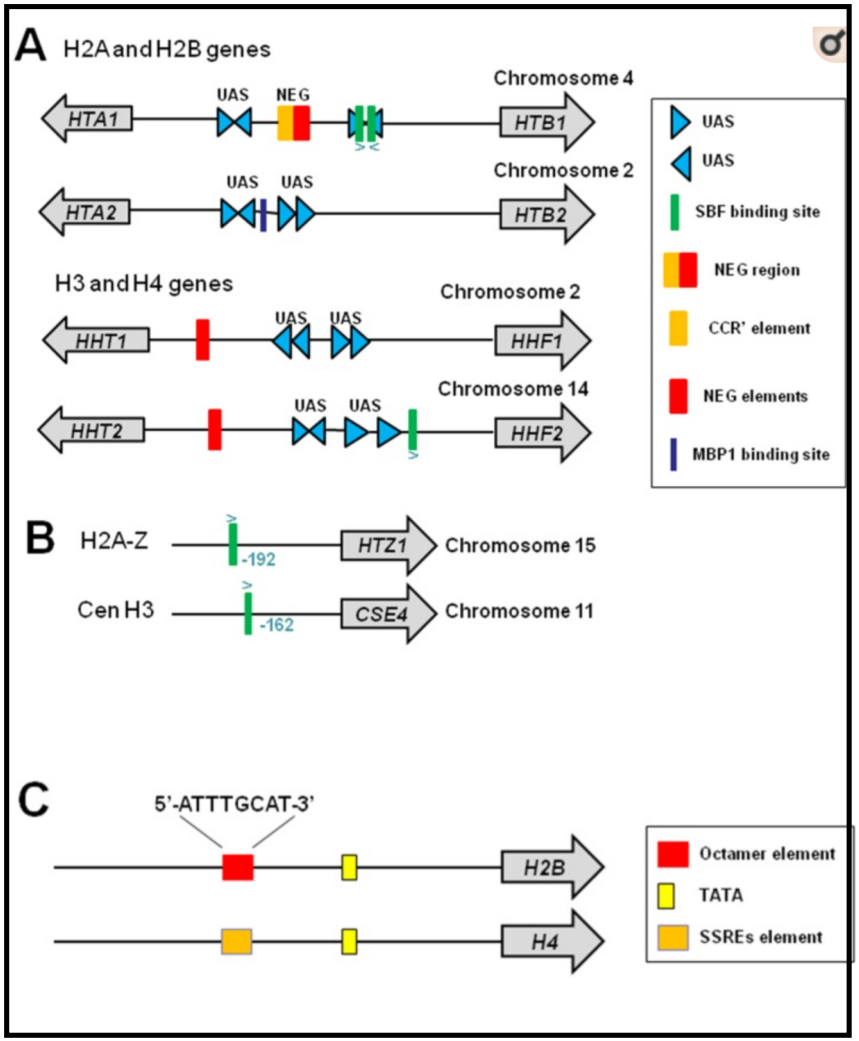
A. REPLICATION DEPENDENT REGULATION: **Canonical histone proteins**: Histone mRNA levels increase ca 15-fold increase in the level of histone mRNAs during S phase. At the end of S phase or DNA synthesis is interrupted, cells turned off histone transcription and histone mRNA levels declined rapidly. Replication dependent histone chaperons are linked to DNA polymerase and generate nucleosomes in S-Phase

B. REPLICATION INDEPENDENT REGULATION: histone variants: Nucleosome need also be formed in G1, G2 phase when nucleosome arrays get disturbed (trasncription, DNA damage, etc…). The second class of histones is composed of **histone variants that are expressed at a relatively low level throughout the cell cycle, and are therefore regulated in a DNA replication-independent manner. Histone variants have specialized functions!!**

Coordinated control of histone gene expression

Structure of yeast canonical histone genes

The core histone gene promoters contain specialized DNA elements that enable cis-regulation of histone gene expression: UAS (upstream activating sequence) and NEG (negative regulation of expression) Divergent arrangement of histone promoters allows coordinated gene expression to get equal amount of all four core histones.



*HTA1-HTB1*and *HTA2-HTB2*:
encode H2A-H2B pairs

HHT1-HHF1 and *HHT2-HHF2*: encode H3-H4 pairs;

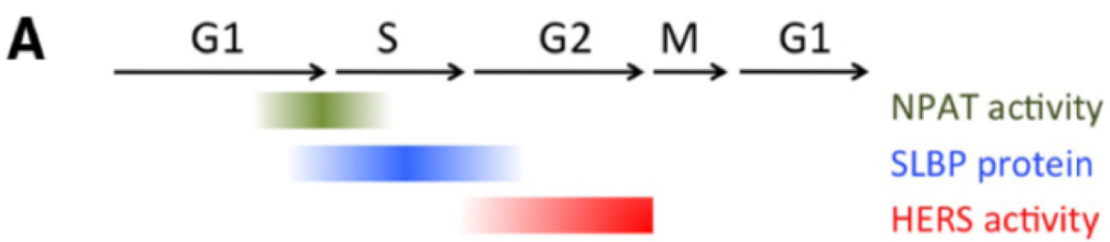
Structure of yeast histone variants H2AZ and CenH3.

Structure of mammalian canonical histone genes

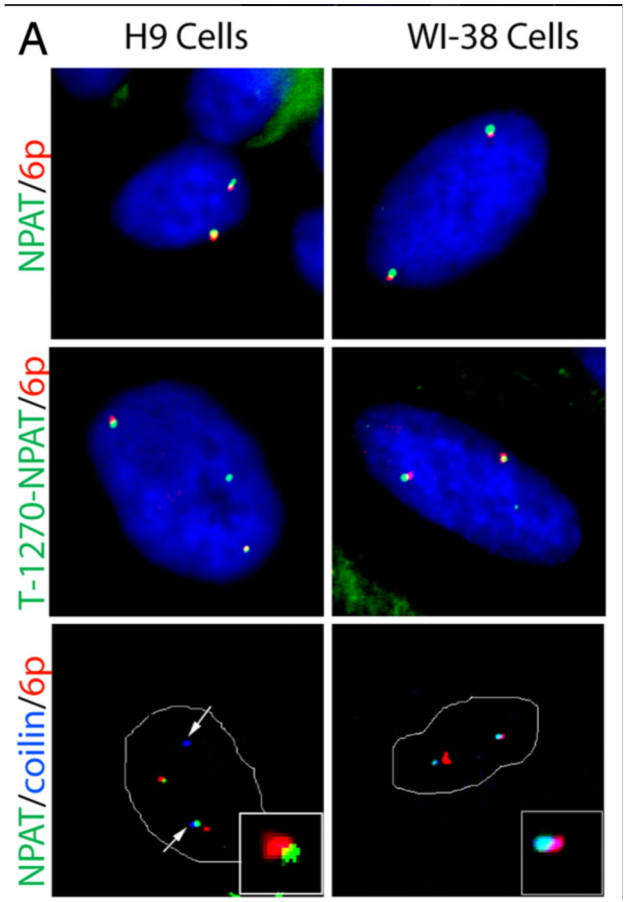
Histone promoters also contain specialized cis elements that required for histone gene expression. Histone H2B promoter contains an octamer element (5'-ATTTGCAT-3'), which is bound by transcription activator Oct-1 (octamer-binding factor 1). Histone H4 promoter contains subtype-specific regulatory elements (SSREs)

CONTROL OF HISTONE EXPRESSION IN S-PHASE

Histone synthesis is limited to S-Phase



The transcription of histone gene takes place in a subnuclear organelle termed the **histone locus body (HLB)**, containing factors required for the processing of histone pre-mRNAs which have an **unusual mRNA structure, with a 3'UTR that forms a stem-loop structure instead of a polyA tail** (White et al., 2007; Nizami et al., 2010).



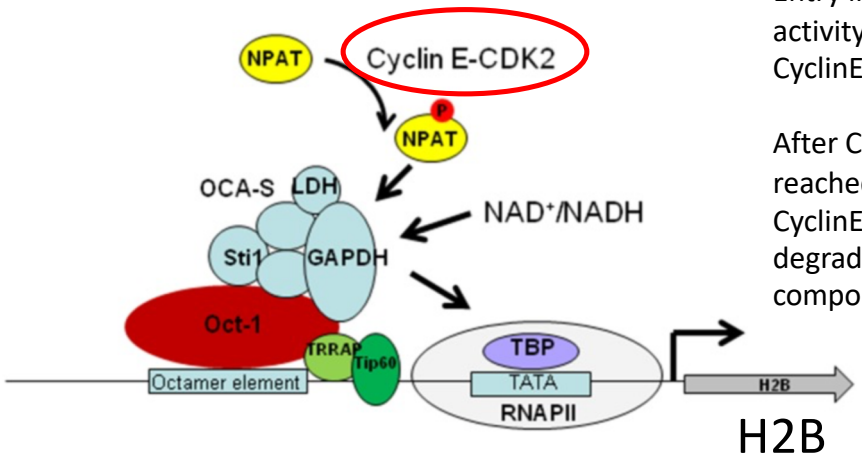
Human cells in S-phase
IF staining for NPAT combined with
DNA FISH for the HIST1 cluster on Chr.6

CONTROL OF HISTONE EXPRESSION IN S-PHASE

NPAT – TRANSCRIPTIONAL ACTIVATION

Histone synthesis is limited to S-Phase

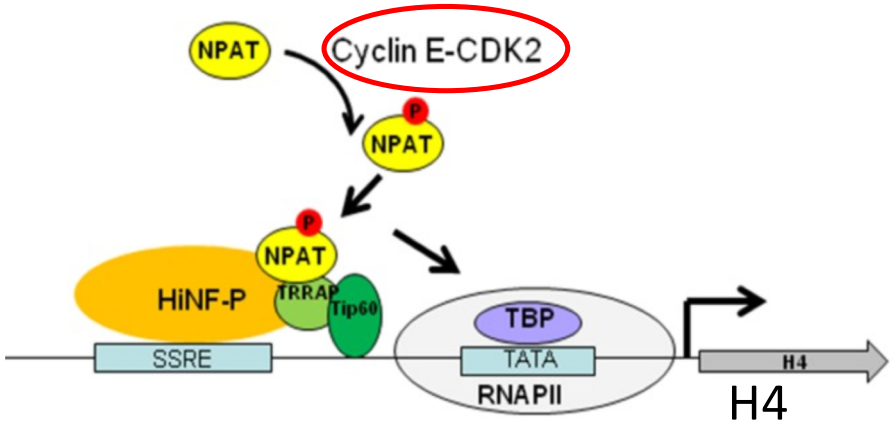
Activation of histone H2B. Oct-1 binds to octamer elements in H2B promoter. During S phase, activated cyclin E/CDK2 complex phosphorylates NPAT. In combination with NPAT, Oct-1 recruits OCA-S to H2B promoter to activates the expression of H2B.



CDK2 and the cell cycle:
Entry into S-phase is triggered by the activity of the G1-S Cyclin complex, CyclinE/Cdk2.

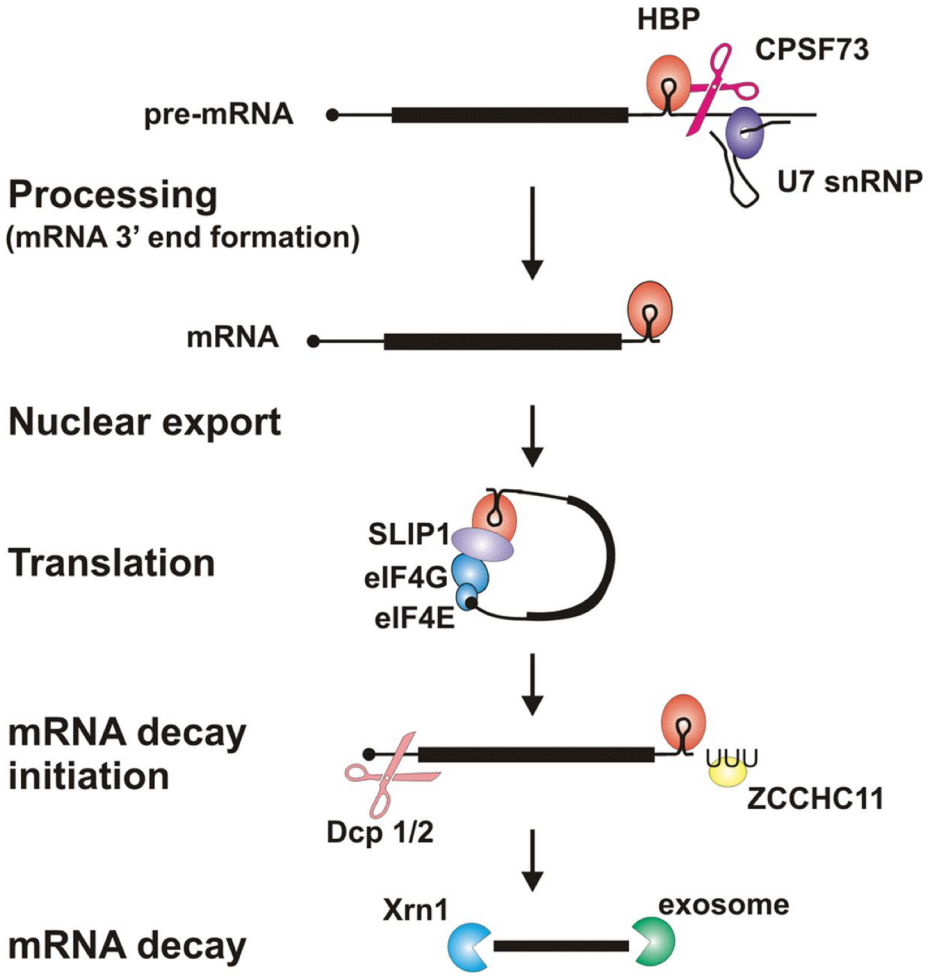
After CyclinE/Cdk2 activity has reached its peak in early S-phase, CyclinE/Cdk2 activity drops due to the degradation of the essential CyclinE component

Activation of histone H4. HiNF-P binds to SSRE within H4 promoters and recruits NAPT and RNA polymerase II to activate gene transcription. NPAT recruits the Tip60 histone acetyltransferase complex to acetylate histone H4 at the G1/S-phase transition. At the end of S phase, the tyrosine kinase WEE1 is recruited to histone promoters to phosphorylate H2B tyrosine 37, which evicts NPAT and RNA polymerase II and instead recruits HIRA to repress histone gene expression.



NPAT, Nuclear Protein Ataxia-Telangiectasia Locus; RNAPII, RNA polymerase II; TRRAP, transformation/transactivation domain-associated protein; SSRE, subtype-specific regulatory elements; OCA-S, Oct-1 co-activator in S-phase; HiNF-P, histone nuclear factor P; TBP, TATA-box binding protein.

HBP (SLBP) - REGULATION OF RNA METABOLISM



HBP (SLBP) protein itself is cell cycle regulated.
SLBP mRNA is synthesized constantly throughout the cell cycle, but **HBP becomes translated just prior to S-phase entry and the protein is degraded at the end of S- phase**

Histone mRNA 3'-end processing requires the RNA-binding protein **HBP (also called SLBP)** , which binds to the conserved hairpin structure in histone pre-mRNA, and the U7snRNP, which binds to a sequence element downstream of the cleavage site.

Together with other factors they position the nuclease CPSF73 for cleavage to produce histone mRNA ending immediately after the stem loop.

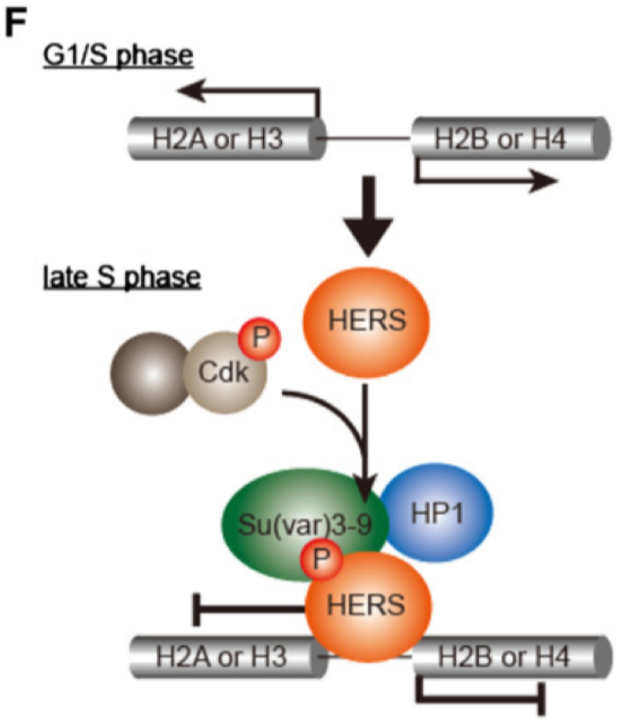
After nuclear export, HBP interacts with SLIP1 and other translation initiation factors to form a closed-loop structure for efficient translation.

This structure is disrupted, presumably when histone mRNA decay is initiated, for example at the end of S-phase.

Addition of an oligo(U) tail by the terminal uridylyl transferase ZCCHC11 is an early step in decay, which involves decapping followed by 5'→3' degradation by Xrn1 or 3'→5' decay by the exosome.

CONTROL OF HISTONE EXPRESSION IN S-PHASE

DROSOPHILA: HERS and Su(var)3-9



Drosophila: The histone gene-specific epigenetic repressor in late S-phase (HERS) protein becomes phos- phorylated by the late S-G2 Cyclin complex CyclinA/Cdk1, which localizes it to the histone genes where it acts to silence histone genes after S-phase .

Cdk-activated HERS silence histone gene expression in late S phase through recruitment of Su(var)3-9/HP1 repressor complex.

Ito et al. 2012 Mol Cell. Epigenetic Silencing of CoreHistone Genes by HERS in Drosophila

REPLICATION COUPLED HISTONES

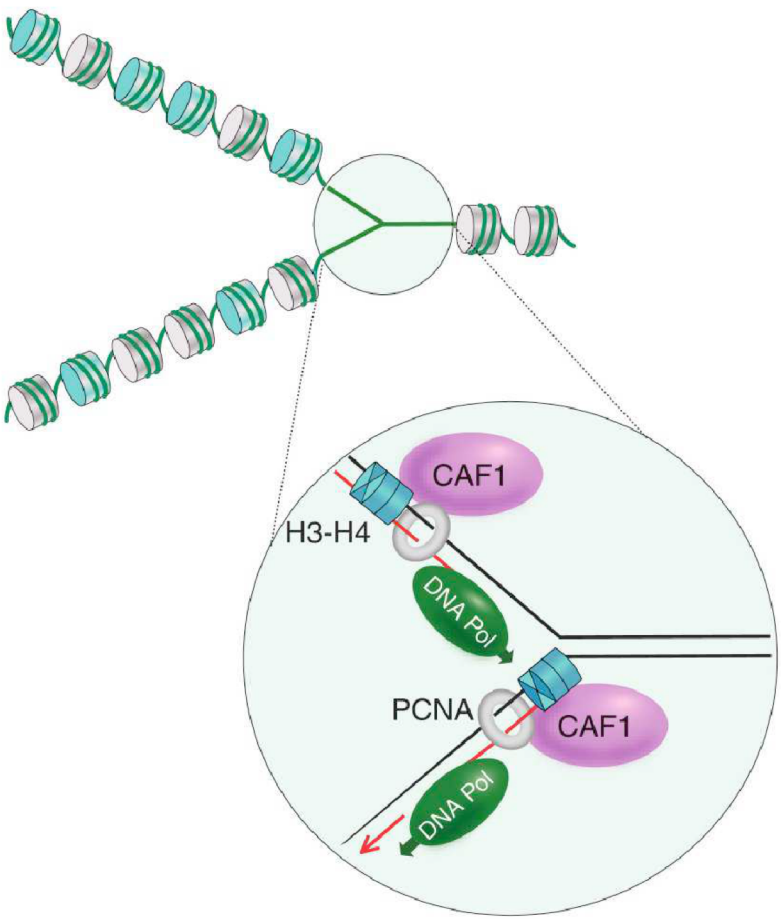


Figure 4. Distribution of old and new nucleosomes at a replication fork. Old nucleosomes (gray disks) are randomly distributed behind the replication fork and new nucleosomes (cyan disks) are deposited in the gaps. CAF-1-mediated nucleosome assembly is depicted on the leading and lagging strand in magnification. DNA polymerase (green); replication processivity clamp, PCNA (gray ring); histone H3-H4 tetramers (cyan); newly synthesized DNA (red lines).

REPLICATION COUPLED (RCs) HISTONES:
H2A, H2B, H3, H4
Are incorporated into new and old DNA strand during DNA replication

Chromatin assembly factor 1 (CAF-1) is a HISTONE CHAPERON that is associated with PCNA. → Facilitates the formation of new nuclosomes
The assembly of a nucleosome consists of the loading of an (H3-H4)2 tetramer (tetrasome) that is followed by the addition of 2 H2A-H2B dimers.

REPLICATION INDEPENDENT (RIs) HISTONES:
Are incorporated independently of DNA replication

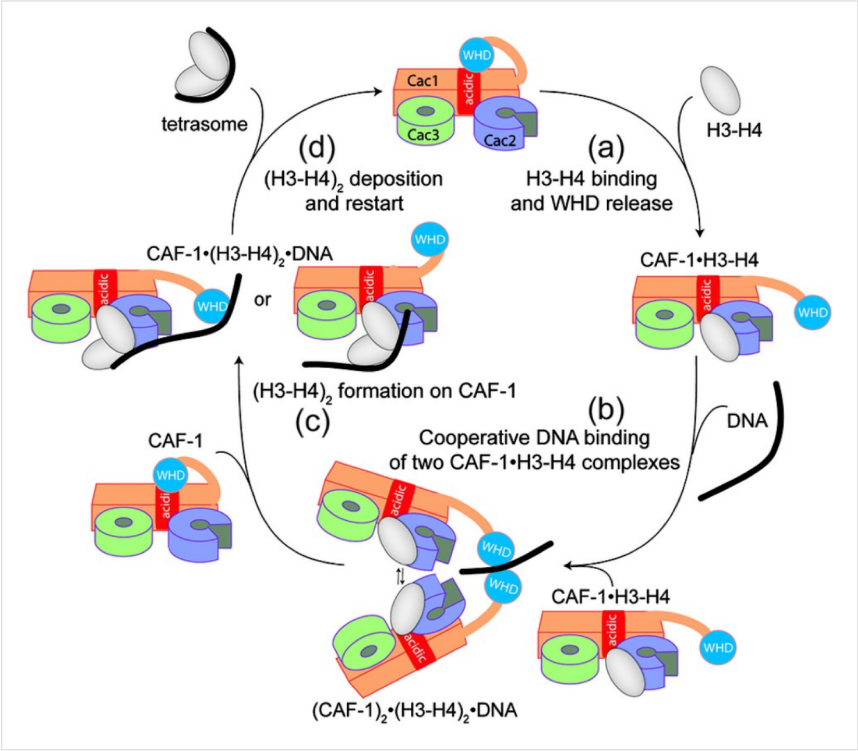
RI histones require the displacement of a preexisting nucleosome unit (active displacement or “loss”)

RI histones can **reset the epigenetic state of a pre-exising nucelosome**

Histone variants with high similarity to ”normal” histone are also incorporated into cannonical octamers during replication. However concentration of variants is very low → no big effect on chromatin structure

HOWEVER: HISTONE CHAPERONS EXIST THAT ENSURE CONCENTRATED INCORPORATION OF HISTONE VARIANTS AT DEFINED SITES → CONTROLLED LOCAL CONCENTRATION → ALTERATION OF CHROMATIN STRUCTURE

Incorporation of replication coupled histones by CAF-1 histone chaperone



Humans CAF-1 subunits: p150, p60, and p48,
Budding yeast CAF-1 subunits: Cac1, Cac2, and Cac3

CAF-1 p150 interacts with PCNA

(a) The nucleosome assembly mechanism of CAF-1 is activated by H3-H4 binding, which releases the WHD domain from an intramolecular interaction with the acidic region on Cac1.
(b) DNA binding promotes the association of two CAF-1•H3-H4 complexes to join the histones into a (H3–H4)2 tetramer
(c) In the presence of DNA of sufficient length, the (H3–H4)2 histones are directly sequestered from CAF-1.
(d) (H3–H4)2 are transferred to the DNA to form the **tetrasome**, and the WHD rebinds to the now free acidic region, resulting in its dissociation from DNA.

H2A-H2B can spontaneously associate with tetrasomes in vitro and because CAF-1 itself has significantly lower affinity for H2A-H2B compared to H3-H4, it appears that the primary role of CAF-1 is to promote the formation of an ordered $(H3\text{-}H4)_2$•DNA complex, the tetrasome

Mattiroli F; Elife. 2017; PMID: 28315523