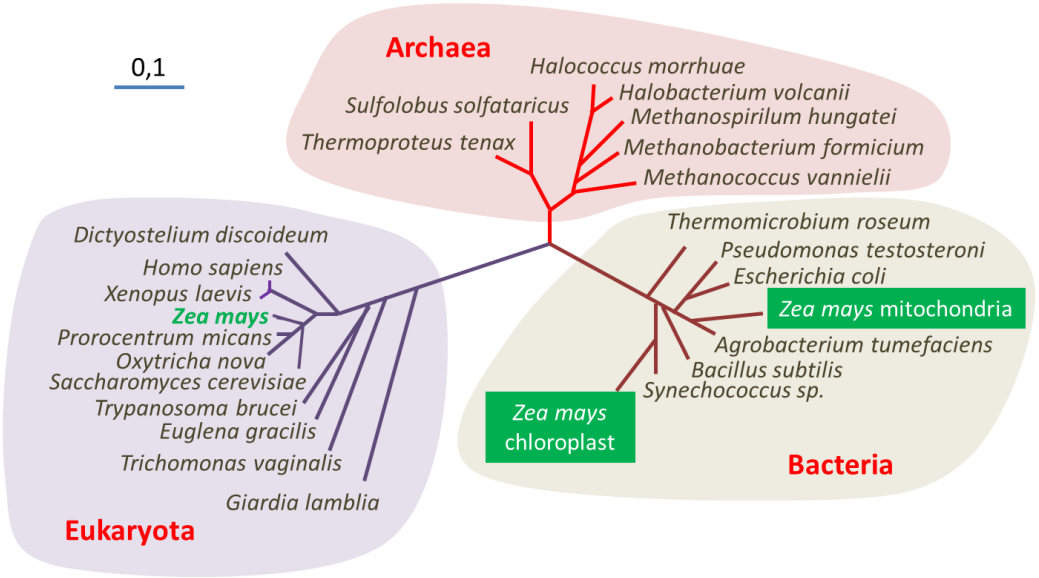
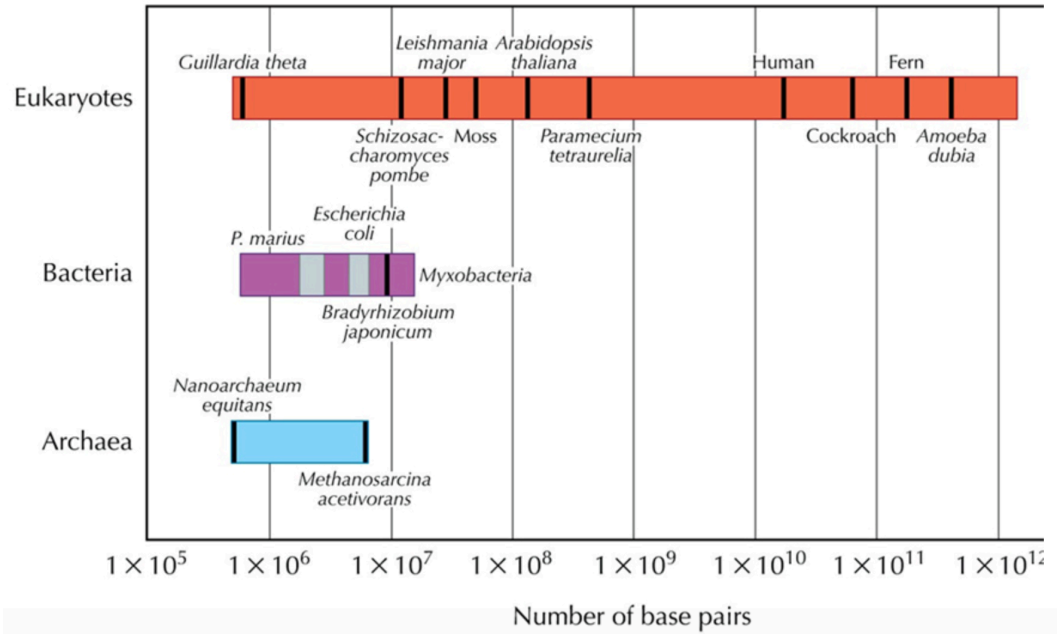
Organization of chromatin - Histones

Genomes in the 3 domains of living organisms





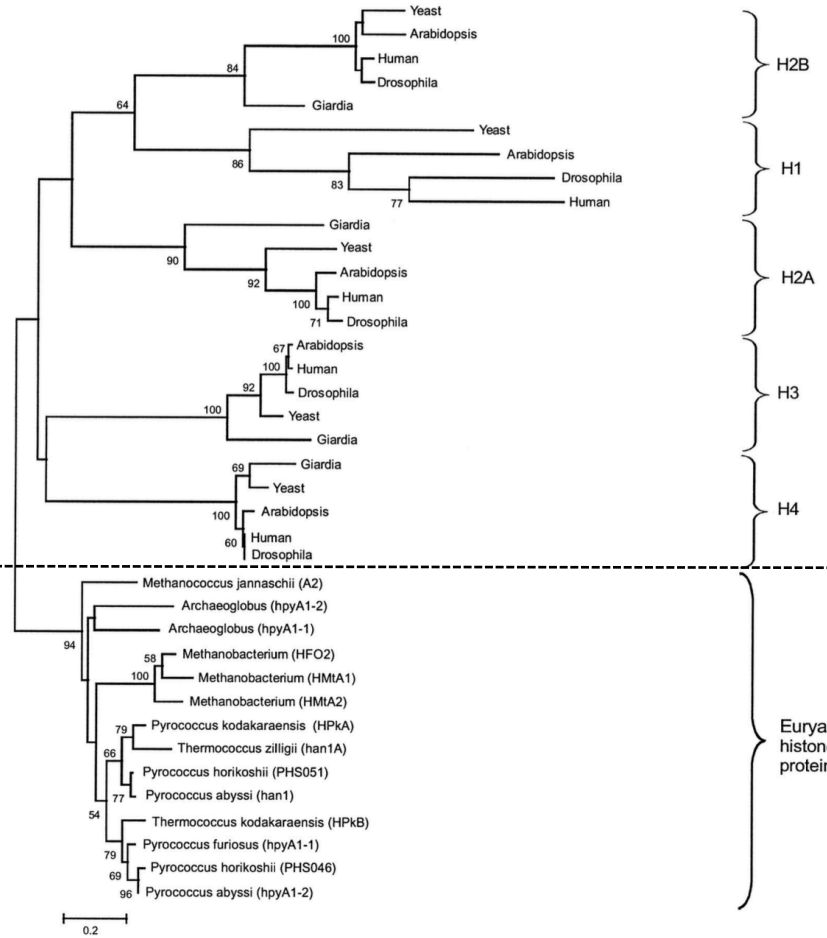
The unrooted phylogenetic tree of the three domains of living organisms, produced using a gene from the small ribosomal subunit

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

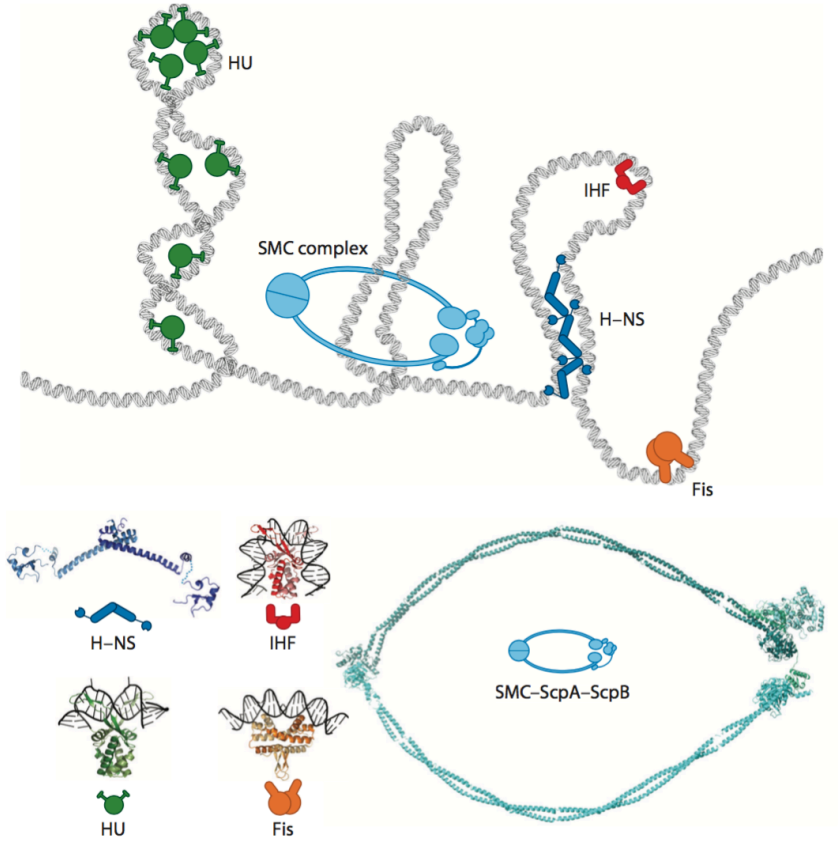


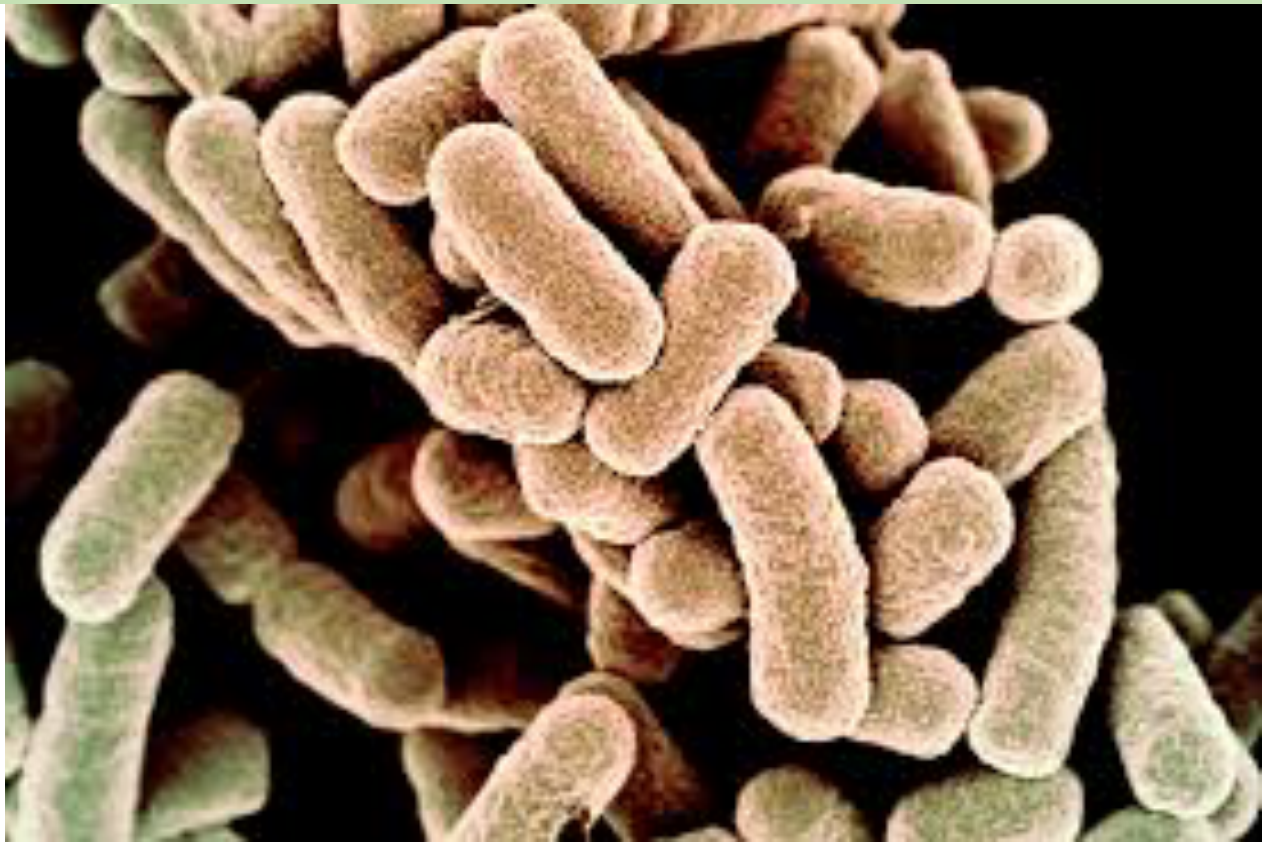
H4: very highly conserved between species
(98% identical between cows and peas)
~ 1% change in 600 million years (2 changes, Val to Ile, Lys to Arg)

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

Histone-like proteins
In archaebacteria

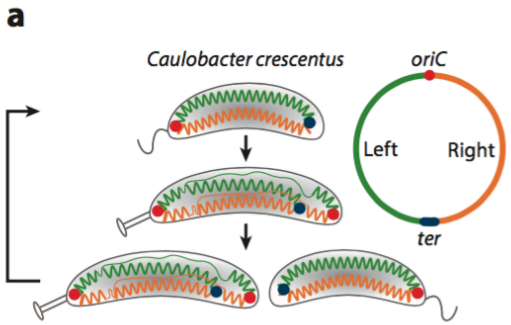
Genome organization in bacteria



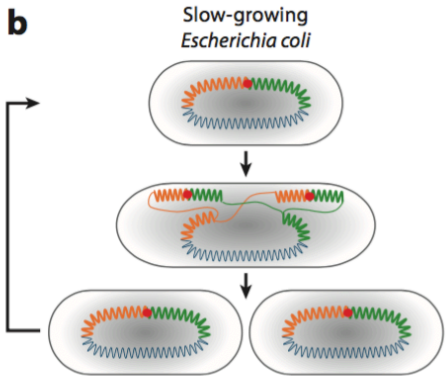


Helical fold of DNA in bacteria

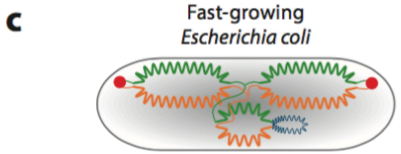
A helix-like conformation of DNA has been observed in replicating bacteria In addition, whole The biological significance of a helical fold is unknown but may represent an energy-minimal configuration for fitting chromosomes within rod-shaped cells.



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**. 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



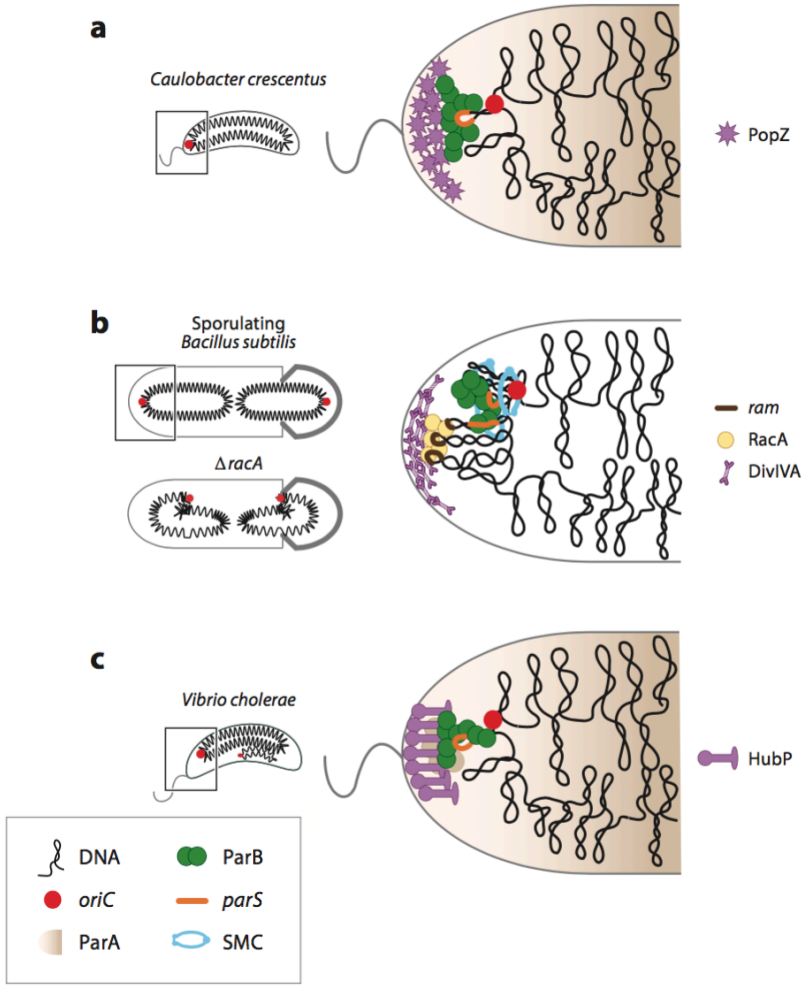
Escherichia coli
The origin resides near midcell, with the two chromosomal arms on opposite sides of the cell and the terminus typically near midcell, in a so-called left-*ori*-right configuration (Nielsen et al. 2006b, Wang et al. 2006). DNA replication and segregation of the origins to cell quarter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



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

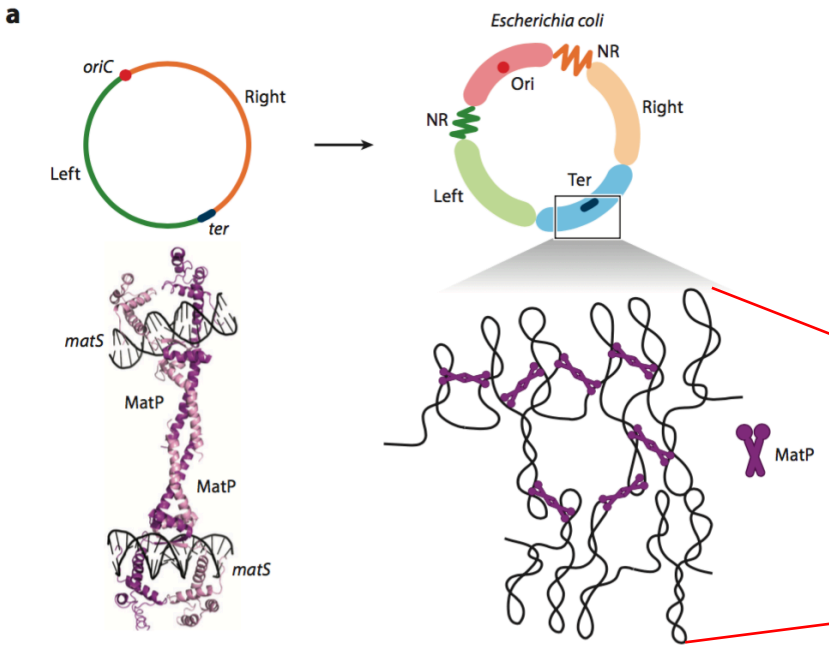
In *Vibrio cholerae*, a membrane-associated protein called HubP anchors the origin of the large chromosome, ChrI, to the pole. HubP interacts with ParAI, which likely interacts with ParBI, which in turn binds a *parS* site near the ChrI origin. HubP has a peptidoglycan-binding LysM domain, which is required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

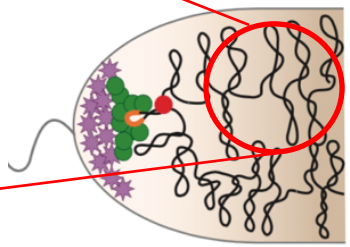
Macrodomains in bacteria

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

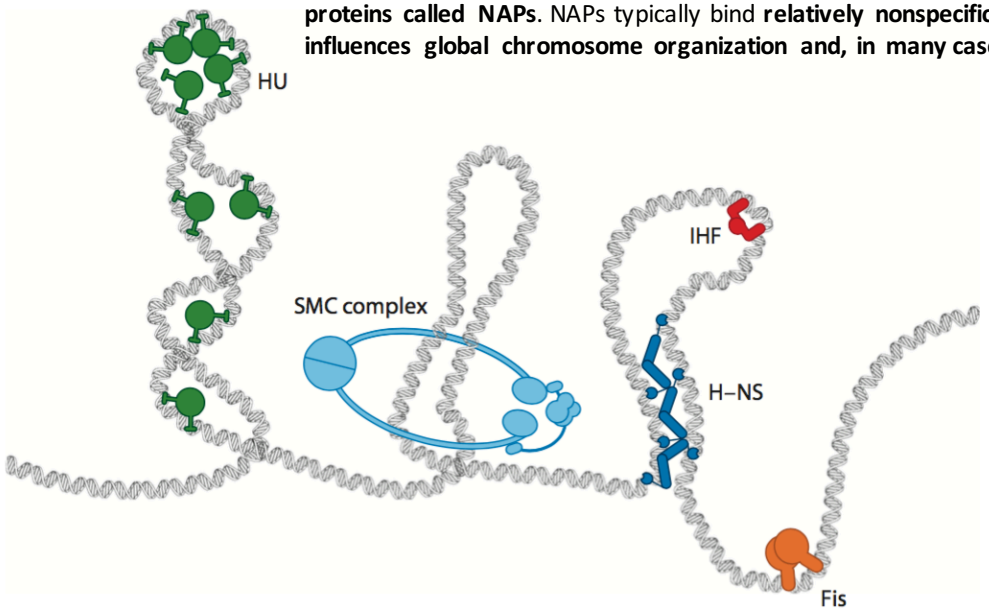
The *E. coli* protein MatP binds to 13-bp *matS* sites present exclusively in the ~800-kb Ter macrodomain (Mercier et al. 2008). A MatP dimer bound to one *matS* site can form a tetramer with a MatP dimer bound at another site bringing different sites in the genome together.

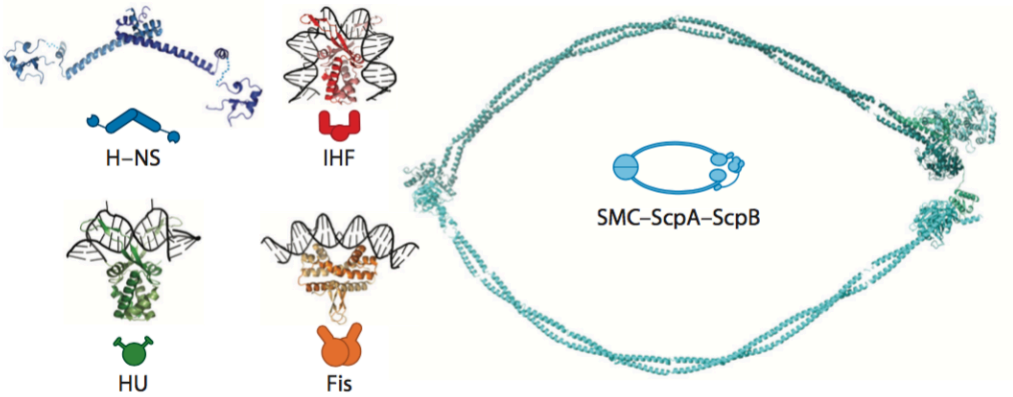


Macrodomains and chromosomal interaction domains. (*a*) Macrodomain organization of the *Escherichia coli* chromosome, (*left*) or with the four macrodomains, Ori, Ter, Left, and Right, The crystal structure of two MatP dimers, each bound to a *matS* recognition site, is shown.

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 (**repressoion**)

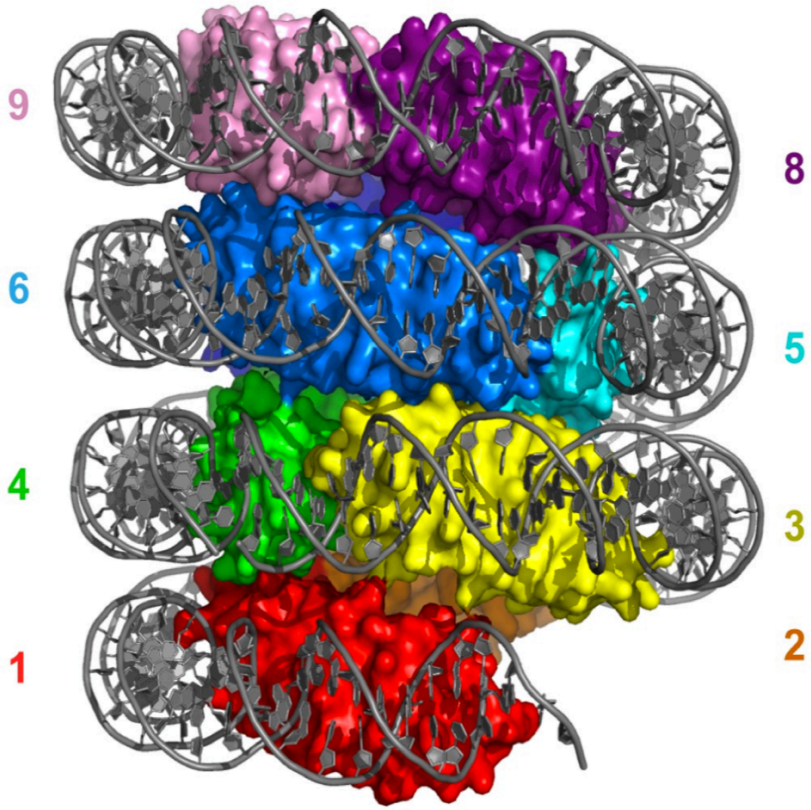
HU is another small (18 kDa), abundant (~30,000 copies/cell) NAP found in many bacteria that coats **and wraps chromosomal DNA around itself in a fashion grossly similar to that of histones**. Can coat 10% of DNA, aggregates, ha proposed role in DNA **compaction**, **promotes supercoiing**,

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 in- cluding an increase in anucleate** cells. SMC likely contributes to both **chromosome segregation and chromosome compaction.**
HOMOLOG IN EUKARYOTES: CONDENSIN

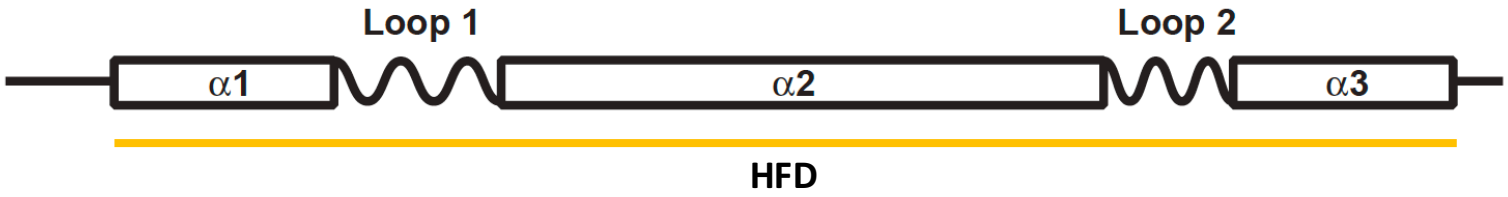
Note: the function of these proteins was found by introducting mutations and study DNA topology/gene expression/recombination…

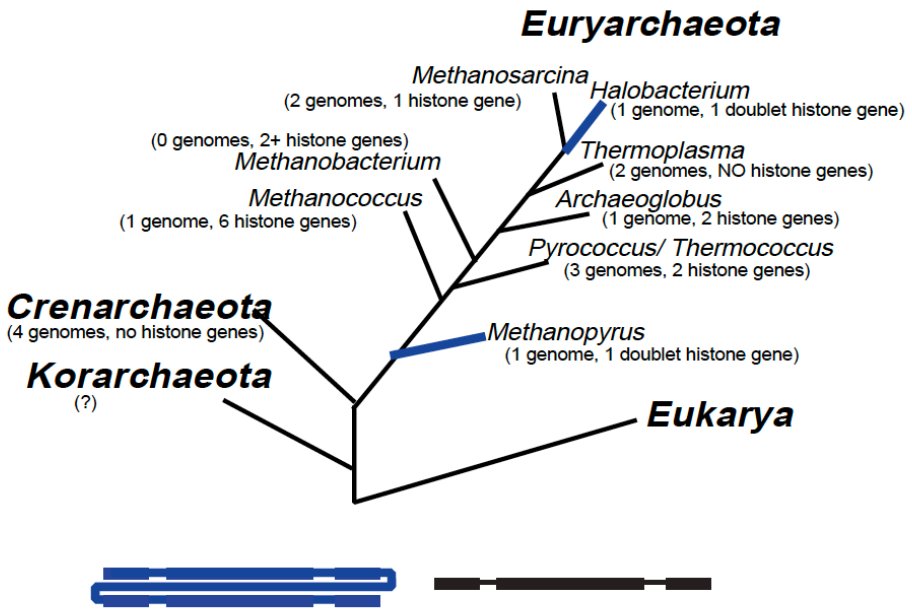
Genome organization in archaea bacteria



Halobacterium

The origin of histone proteins: Archaebacteria





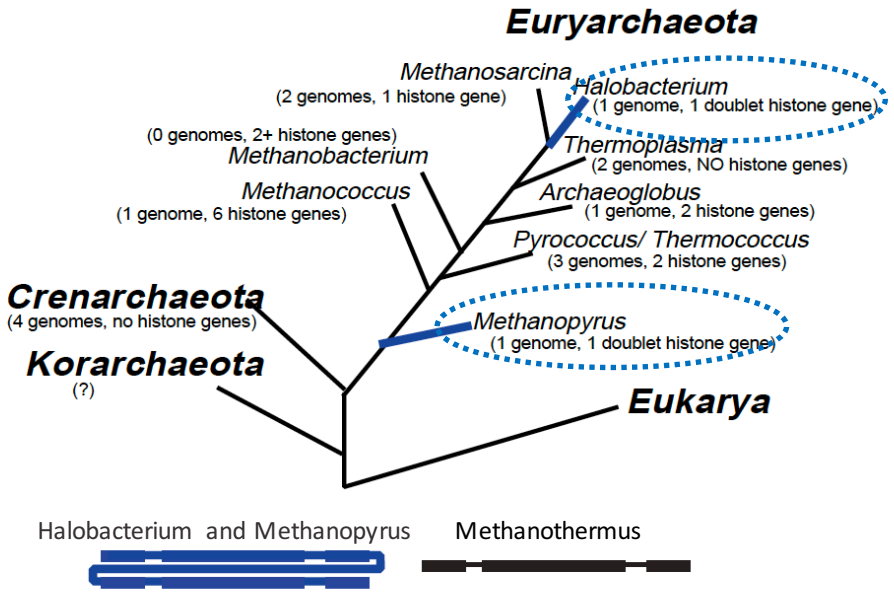
Histone proteins first evolved in Archaebacteria
→ Form «Archeal nucleosomes»

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.

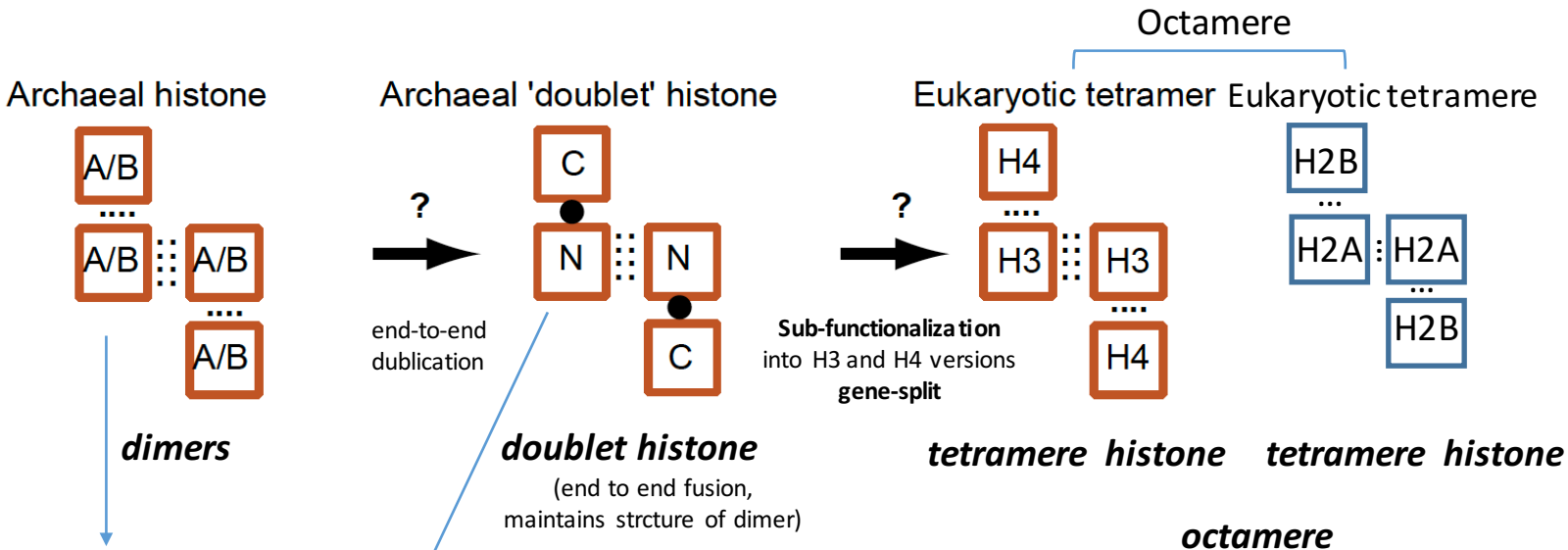
A tetramer of these histone proteins is able to protect 60 base pairs of packaged DNA from nuclease digestion, suggesting a **single wrap** of DNA around the tetramer

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 if eukaryotic histones octamers containing 4 core histone components



b) homodimer of HMfA from M. fervidus

c) doublet histone from M. kandleri

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

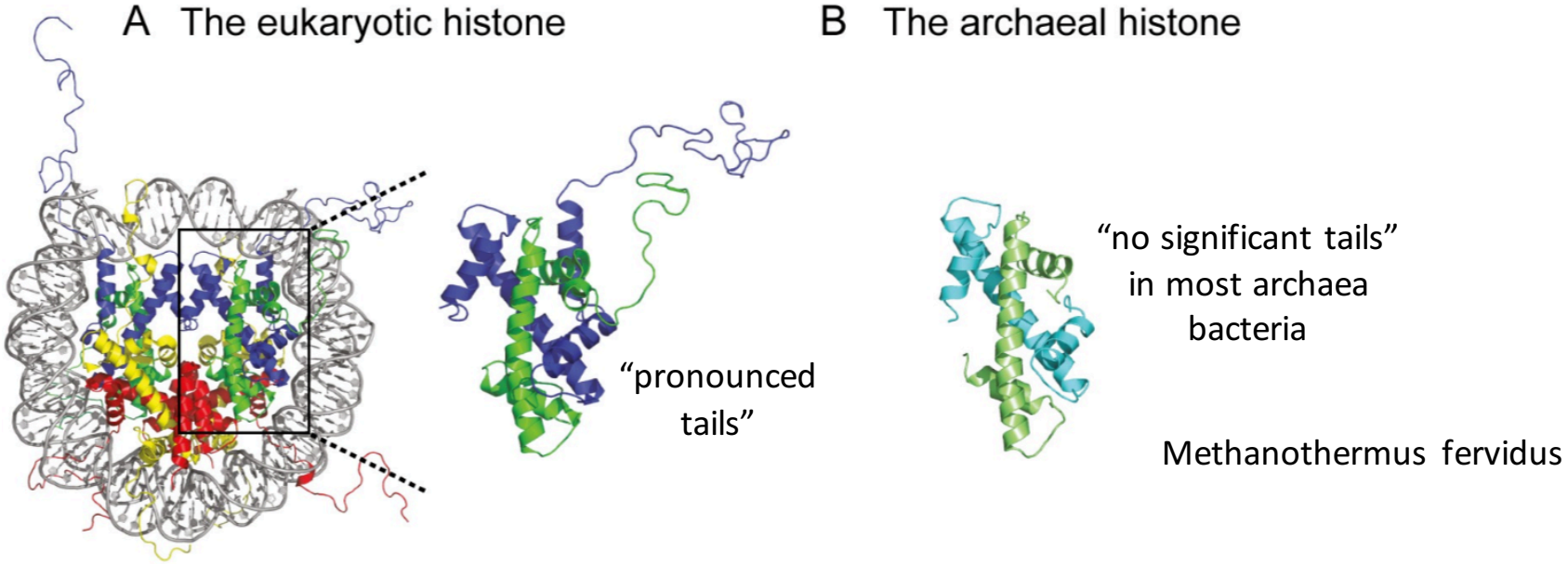


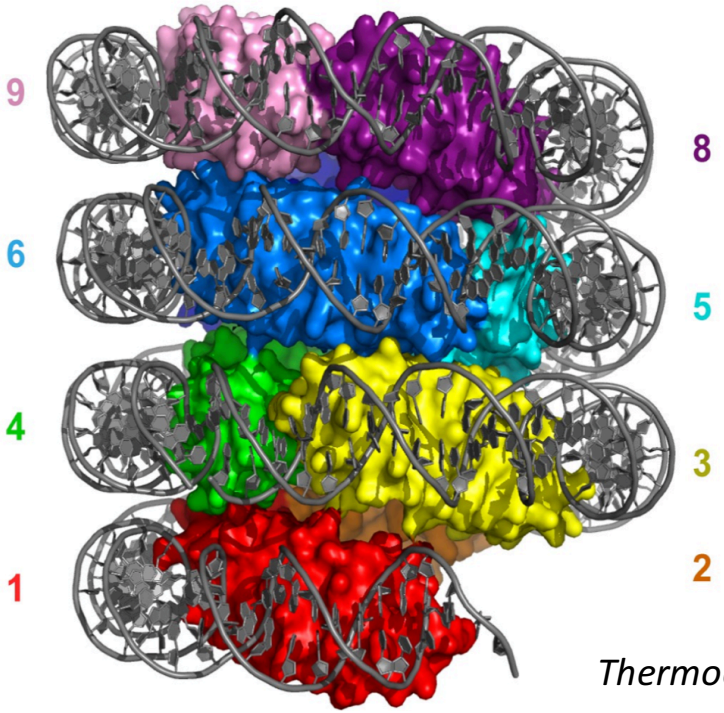
Fig 1. Eukaryotic and archaeal histones. (A) Eukaryotic nucleosome consisting of DNA wrapped around a core of a $(H3\text{-}H4)_2$ tetramer and two H2A-H2B dimers. Yellow, H2A; red, H2B; blue, H3; green, H4. (B) Archaeal histone homodimer of HMfB. HMfB, Histone B from *Methanothermus fervidus*.

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



Thermococcus kodakarensis

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

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

Histone tails in Archaea bactreia hypernucleosomes

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

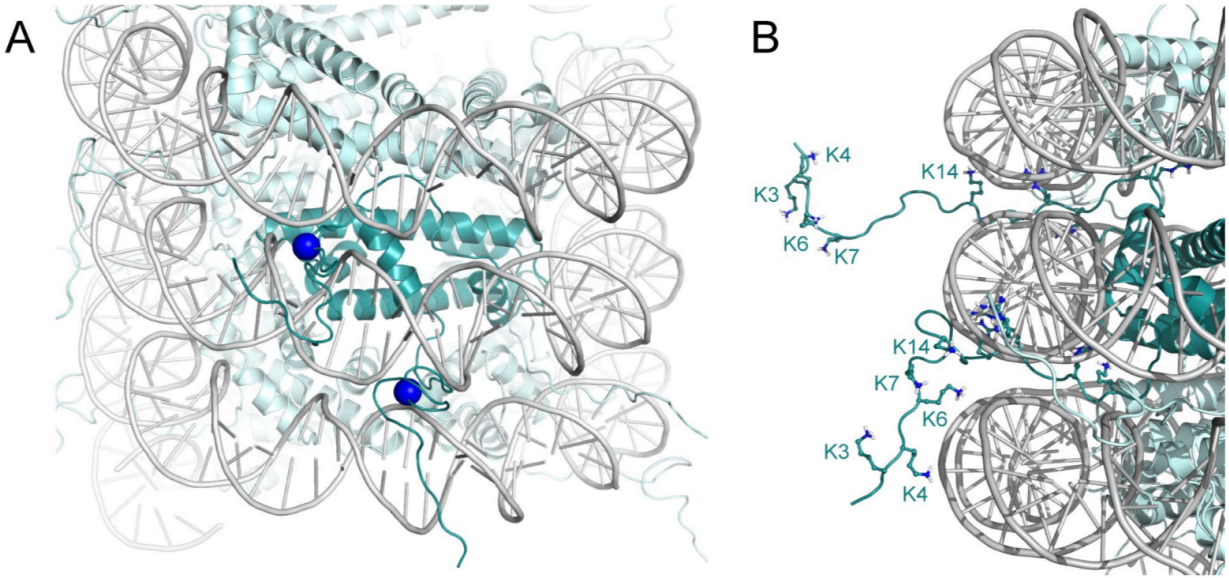


Fig 4. Model of a Heimdall LC_3 hypernucleosome with N-terminal tails. (A) View showing histone tails protruding through the DNA minor grooves. The R17 Cα-atom is shown as a blue sphere to mark the exit point of the tail. (B) Close up of the histone tails with lysine and arginine residues shown as sticks, 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

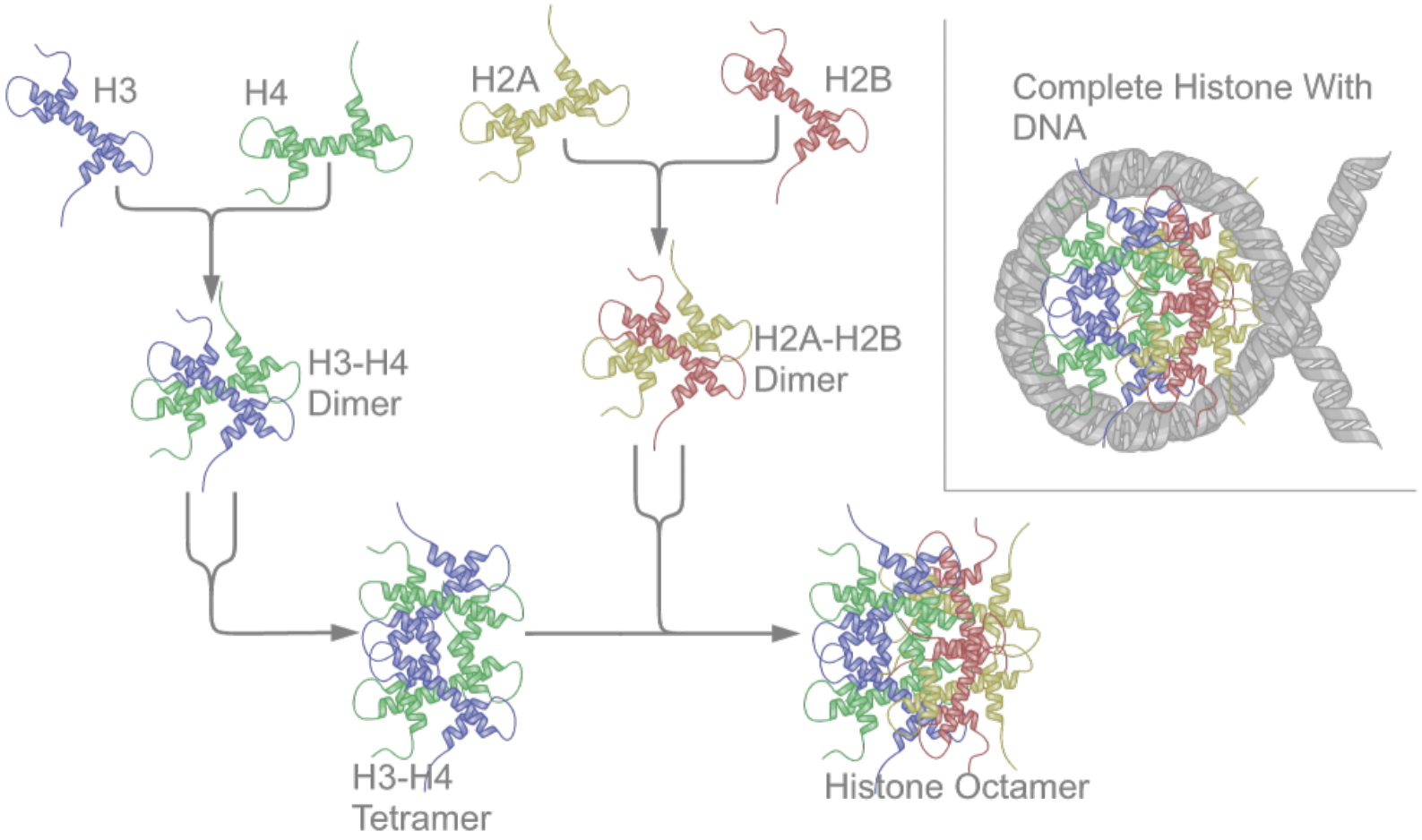
10-fold compaction

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*.

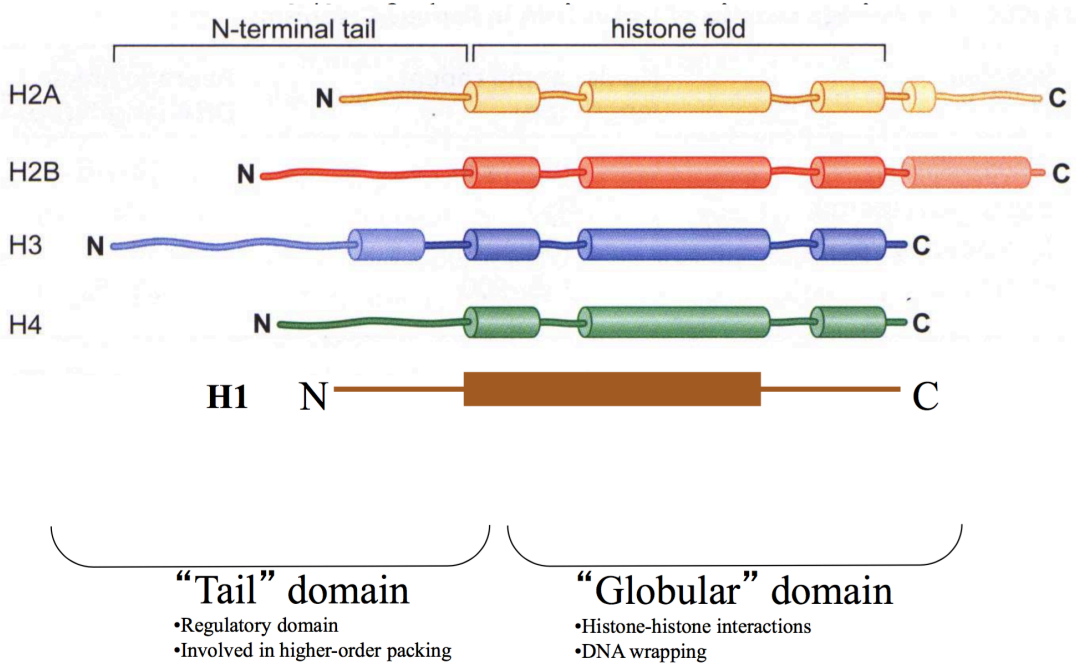
Like **the H4 tail that is subject to acetylation of lysines K5, K8, K12, and K16, lysines in the Heimdallarchaeal histone tail may well be subject to acetylation**. Archaeal genomes are known to **have several candidate lysine acetyltransferase and deacetylase enzymes**.

Heterodimerization + tetramerization = eukaryotic octamere formation



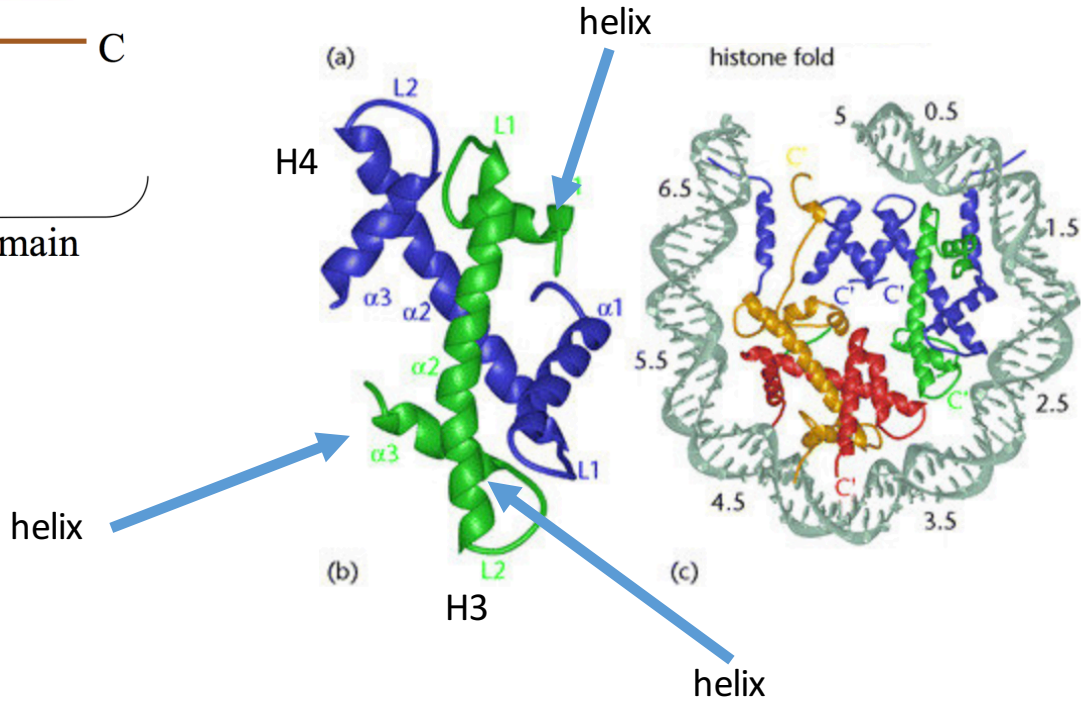
Nucleosome
2 wraps of DNA around Octamere:

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

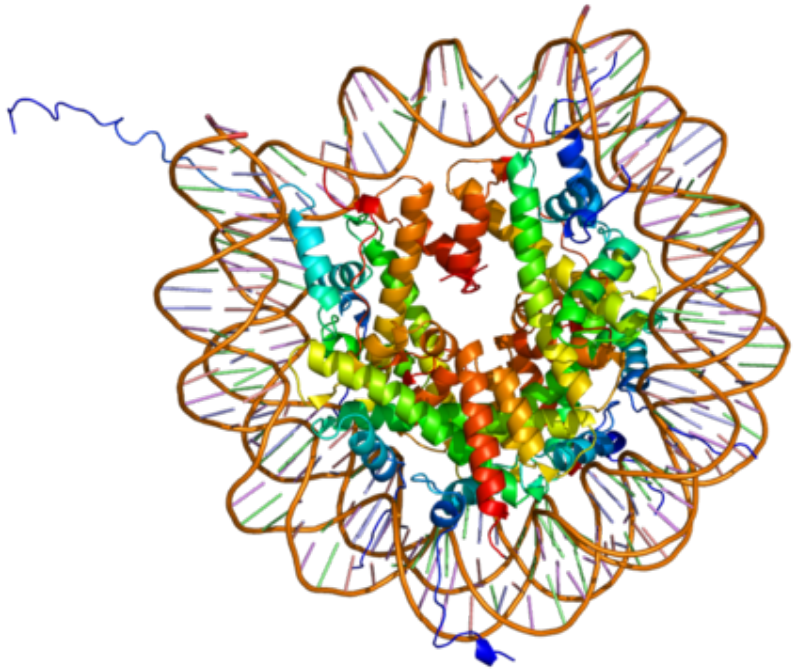
Types and Properties of Histones

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

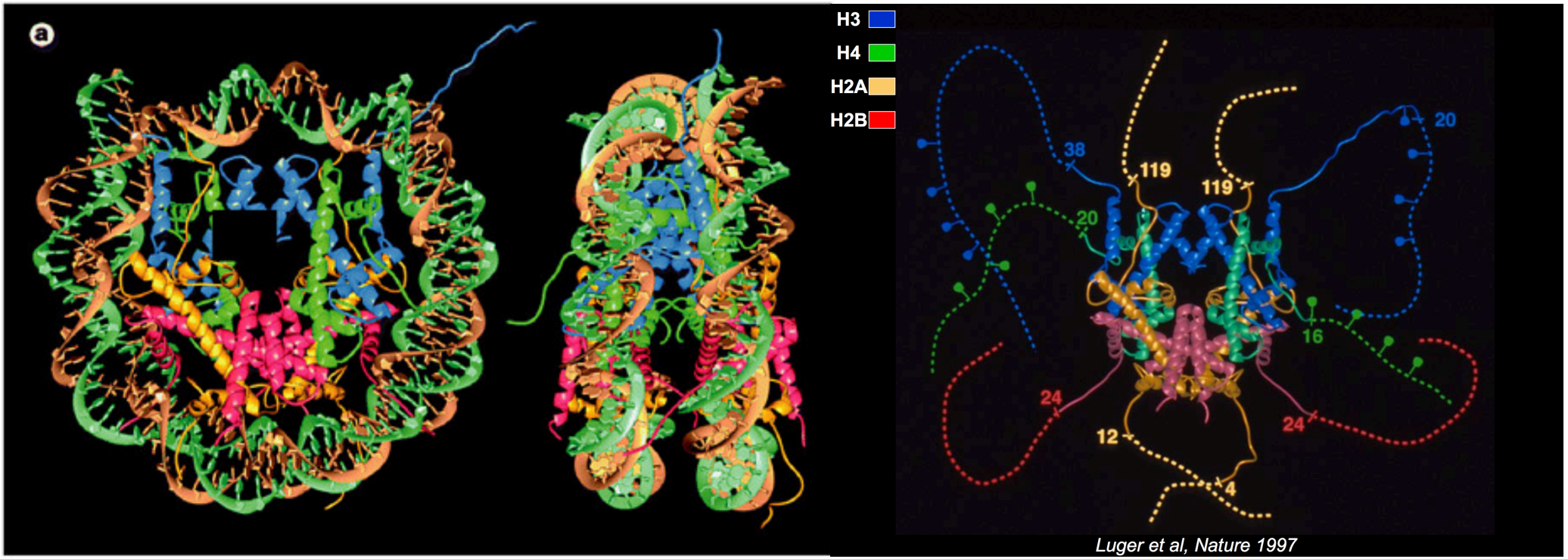
- **Helix-dipoles form alpha-helixes in H2B, H3, and H4 cause a net positive charge to accumulate at the point of interaction with negatively charged phosphate groups on DNA**
- Hydrogen bonds between the DNA backbone and the amide group on the main chain of histone proteins
- Non-polar interactions 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



The structure of the eukaryotic nucleosome



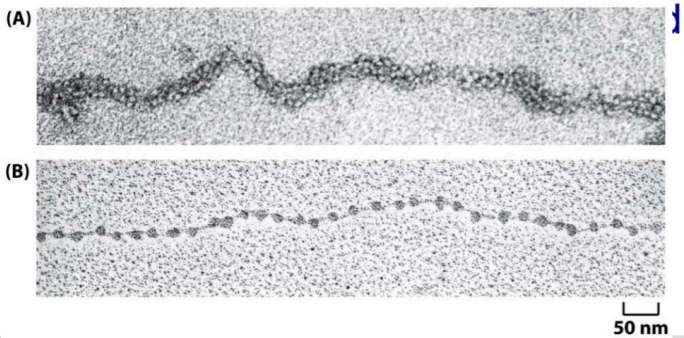
Luger et al., 1997

DNA Packaging by histones

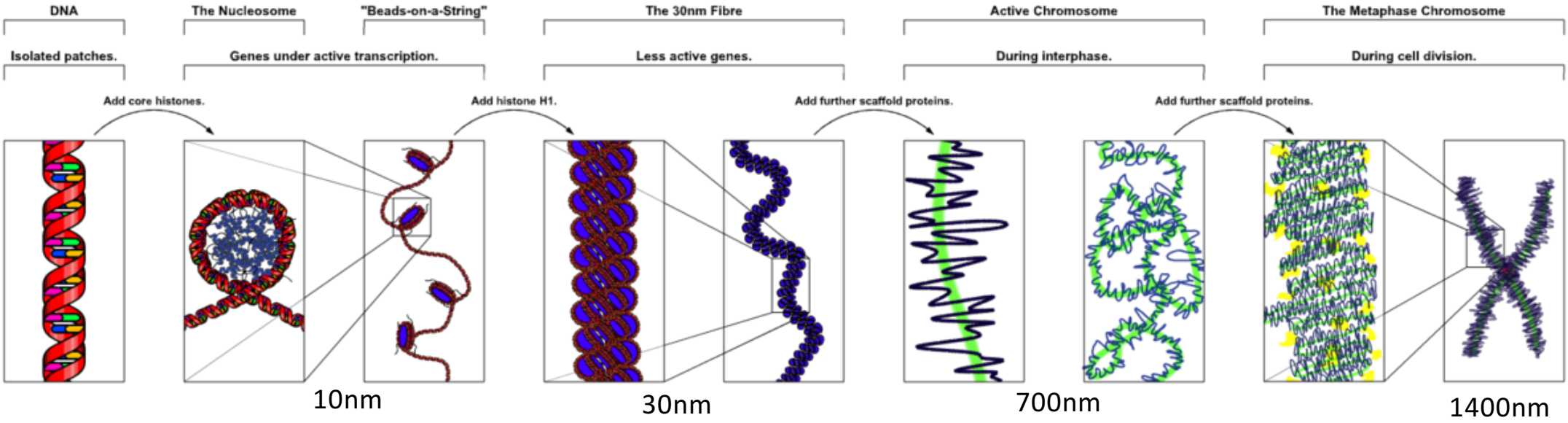
- Long sequence of DNA must be stored within the geometry of a nucleus

  - 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.

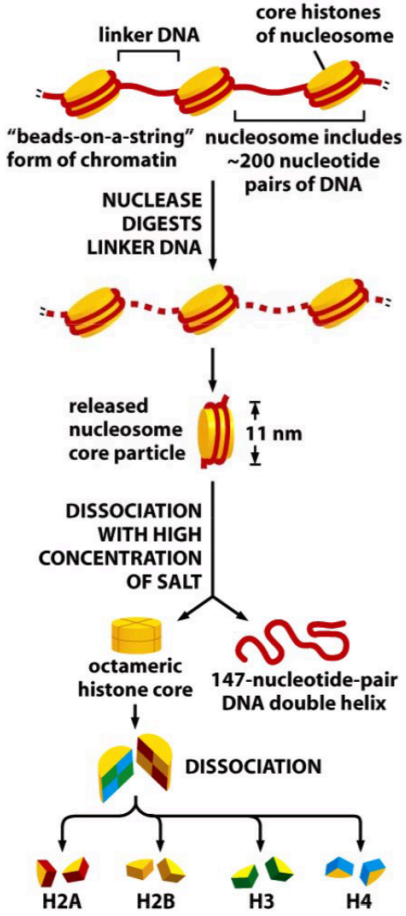


The major structures in DNA compaction



The 10nm fibre – Nucleomes

- 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.

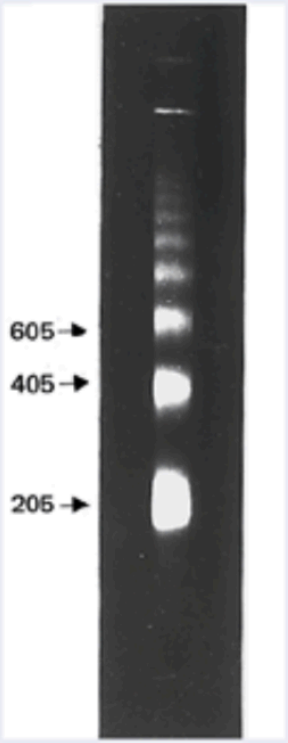
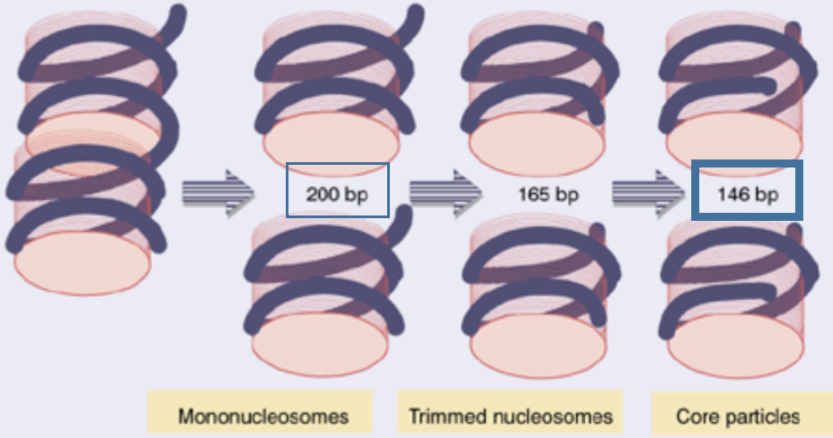
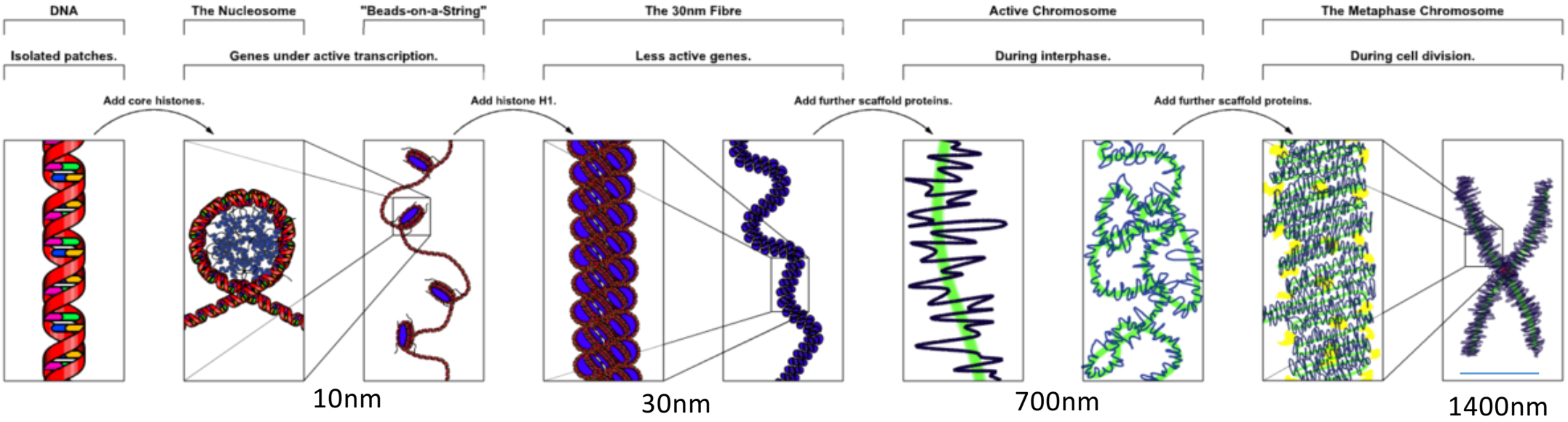


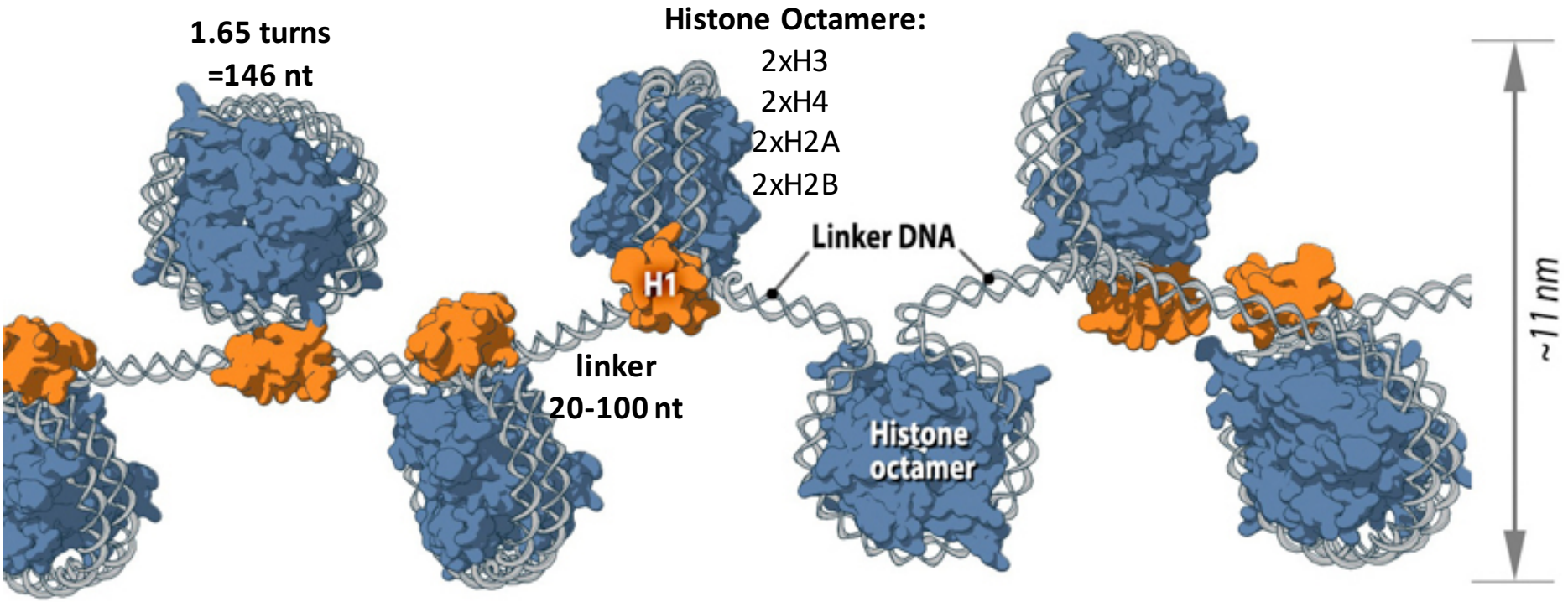
Figure 19.10 Micrococcal nuclease initially cleaves between nucleosomes. Mononucleosomes typically have ~200 bp DNA. End-trimming reduces the length of DNA first to ~165 bp, and then generates core particles with 146 bp.



The major structures in DNA compaction



Histone H1 compacts chromatin to the 30nm fibre



7-fold shortening of DNA

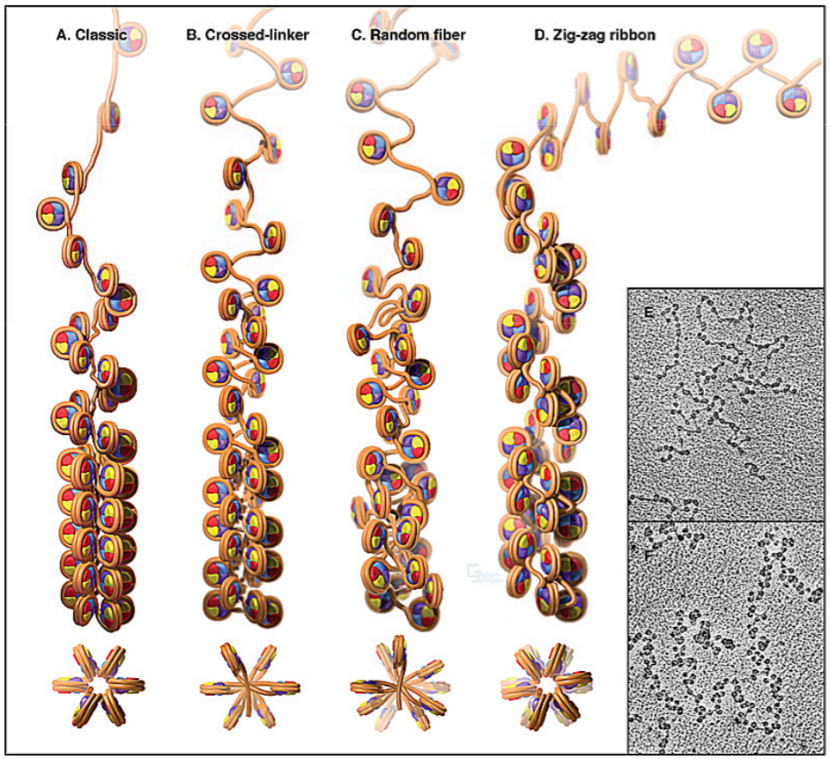
Nucleosome core particle: 146nt DNA + 2xH3, 2xH4, 2xH2A, 2xH2B

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.

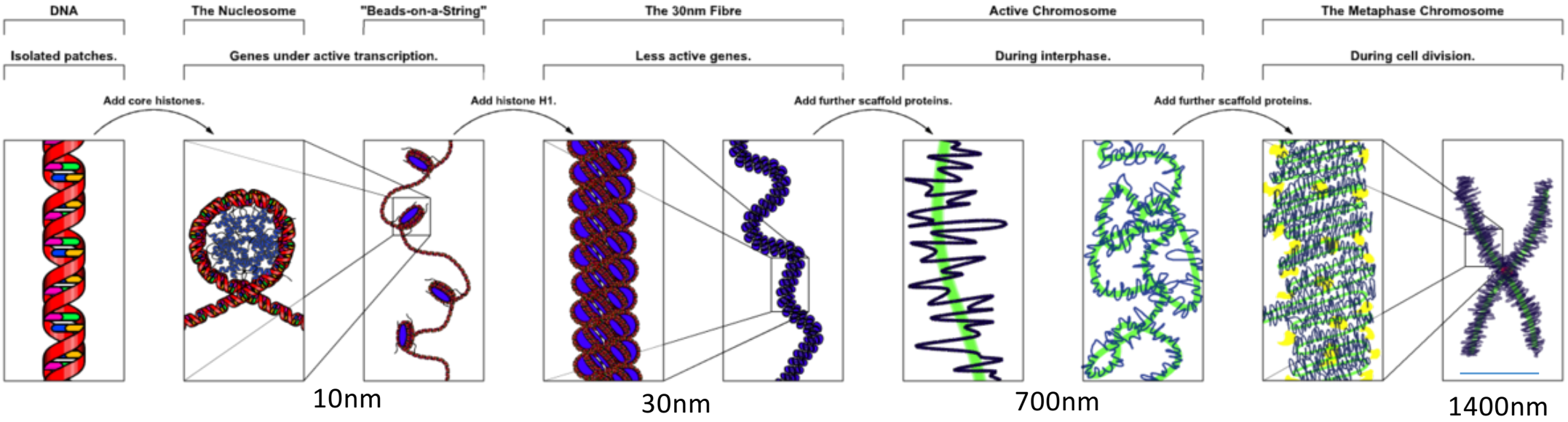
The structure of the 30-nm chromatin fiber remains unclear



Different models of the 30-nm fiber

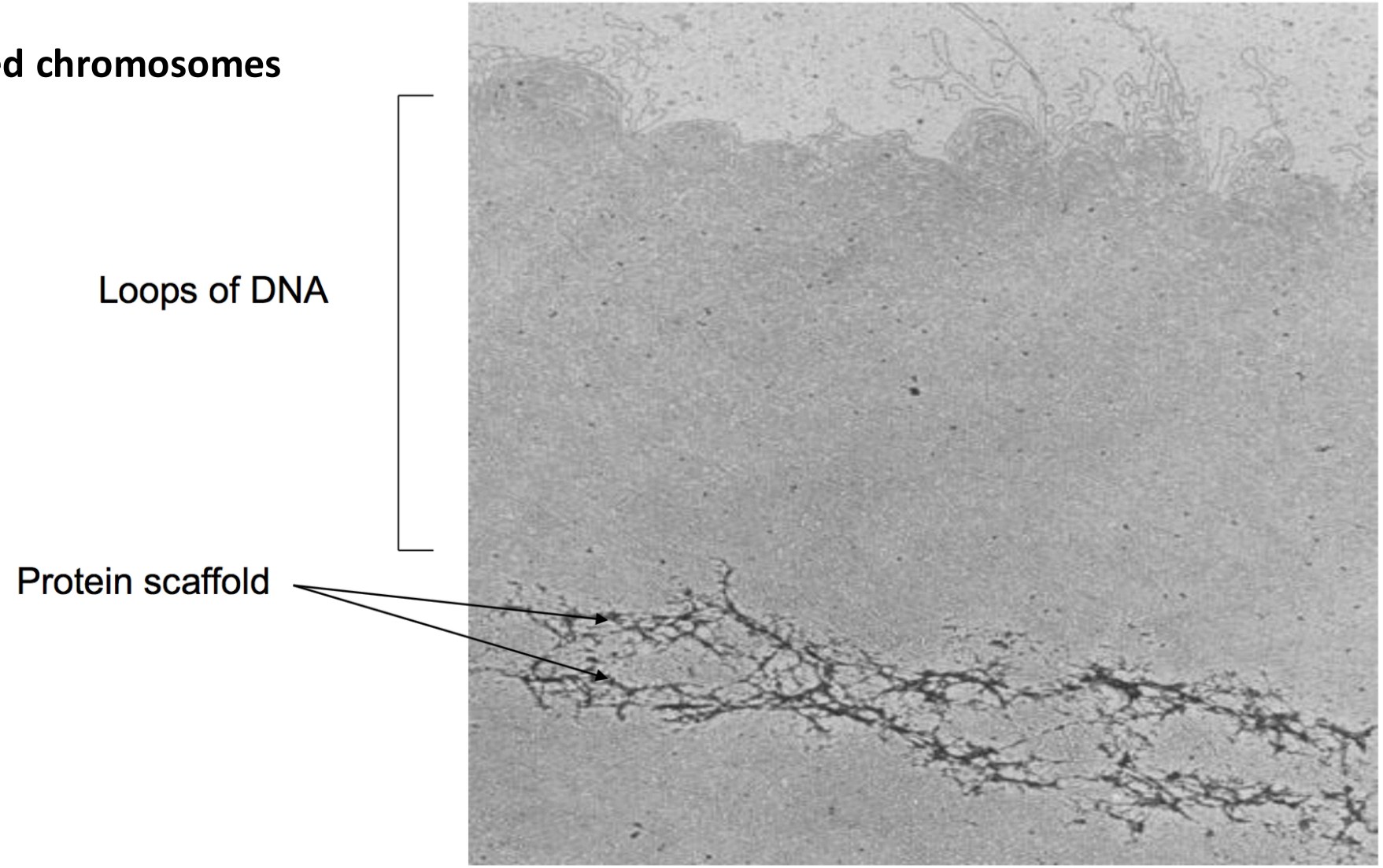
30-nm fiber most likely represents a delicate balance of different configurations that can be modified by a number of factors, including but not limited to, deposition of linker histones and high mobility group (HMG) proteins, incorporation of variant histones, modifications of histone tails, activity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

The major structures in DNA compaction

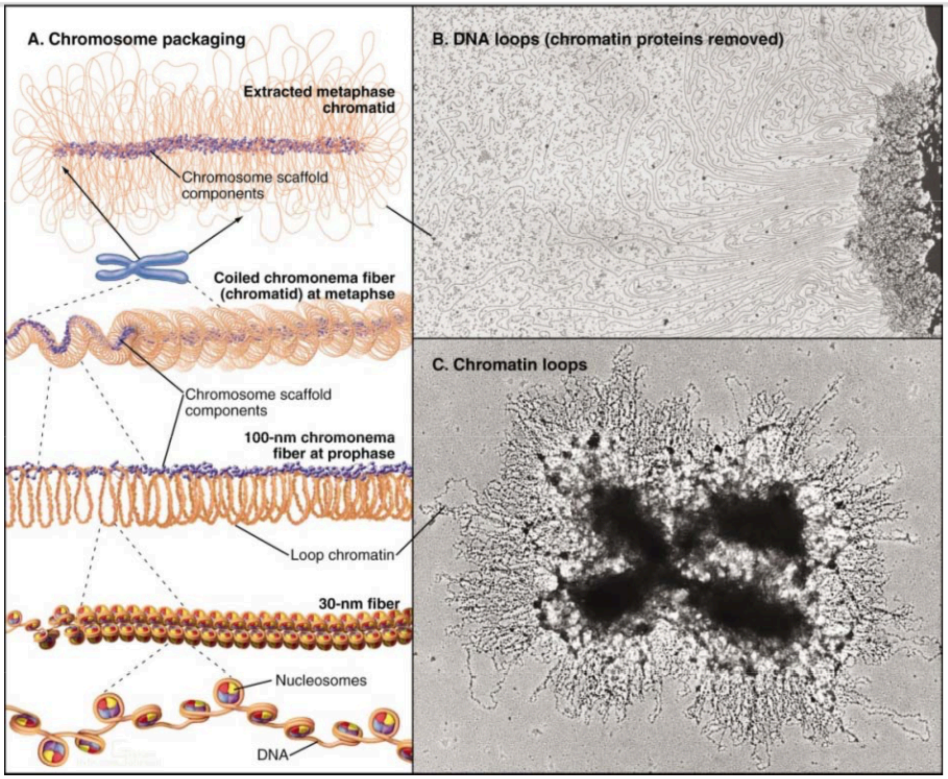


A protein scaffold is required to further compact the genome

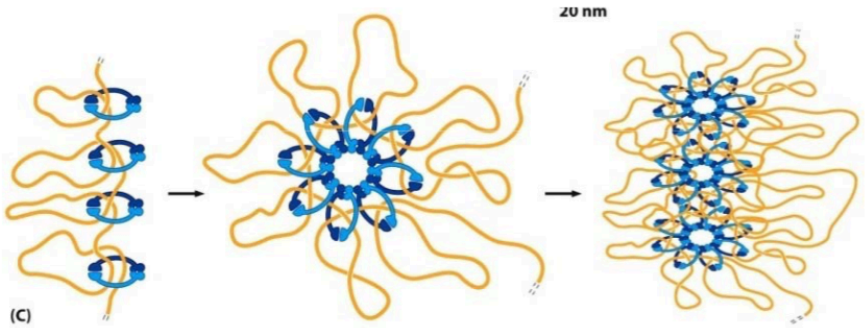
Histone depleted chromosomes



A protein scaffold is required to further compact the genome



Model of mitotic chromosomes



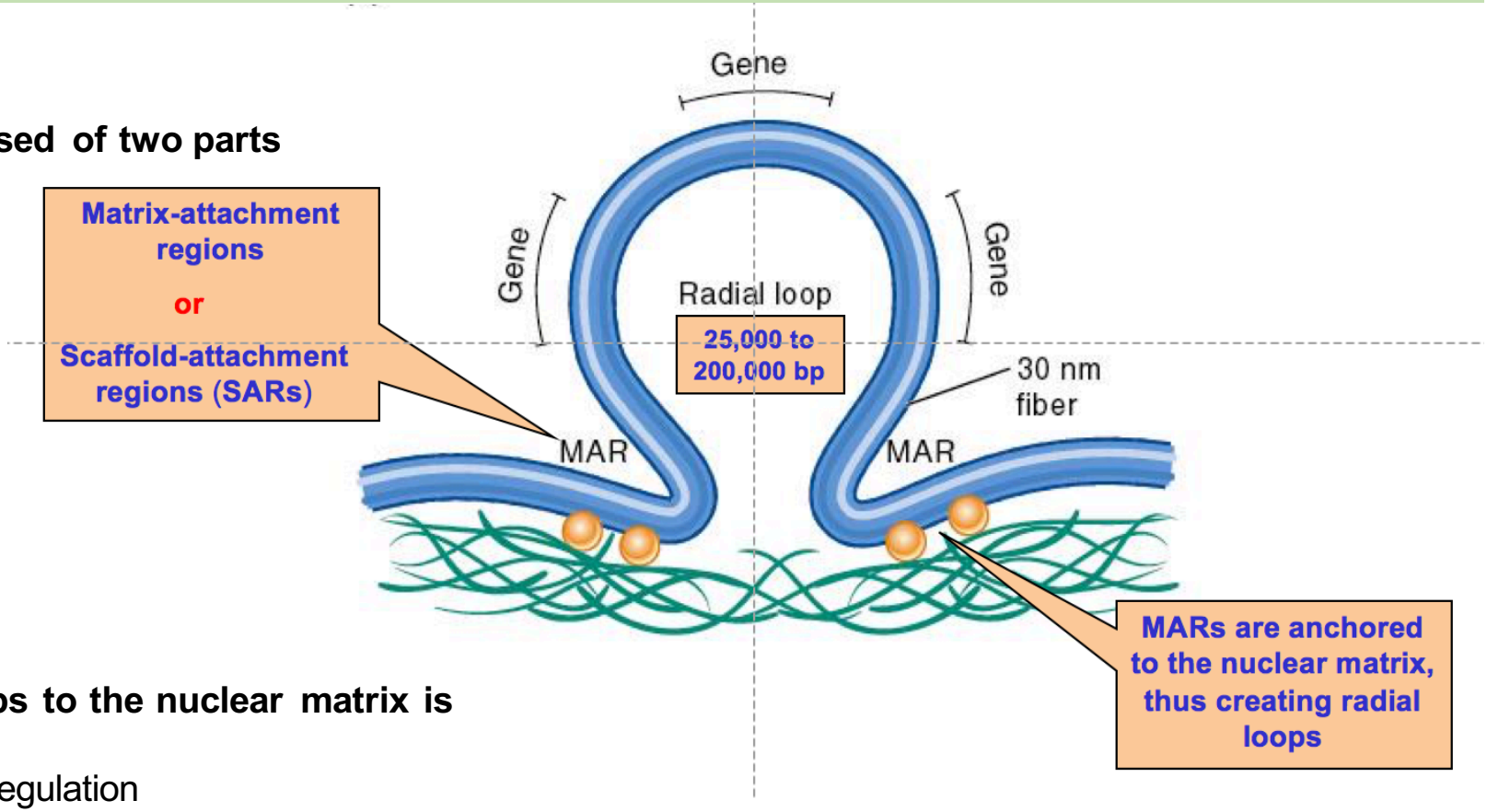
The axially-positioned chromosome scaffold of both chromatids mainly comprises non-histone proteins: so-called scaffold proteins, including condensin, topoisomerase IIα (Topo IIα) and kinesin family member 4

The nuclear mtatrix

The nuclear matrix is composed of two parts

Nuclear lamina

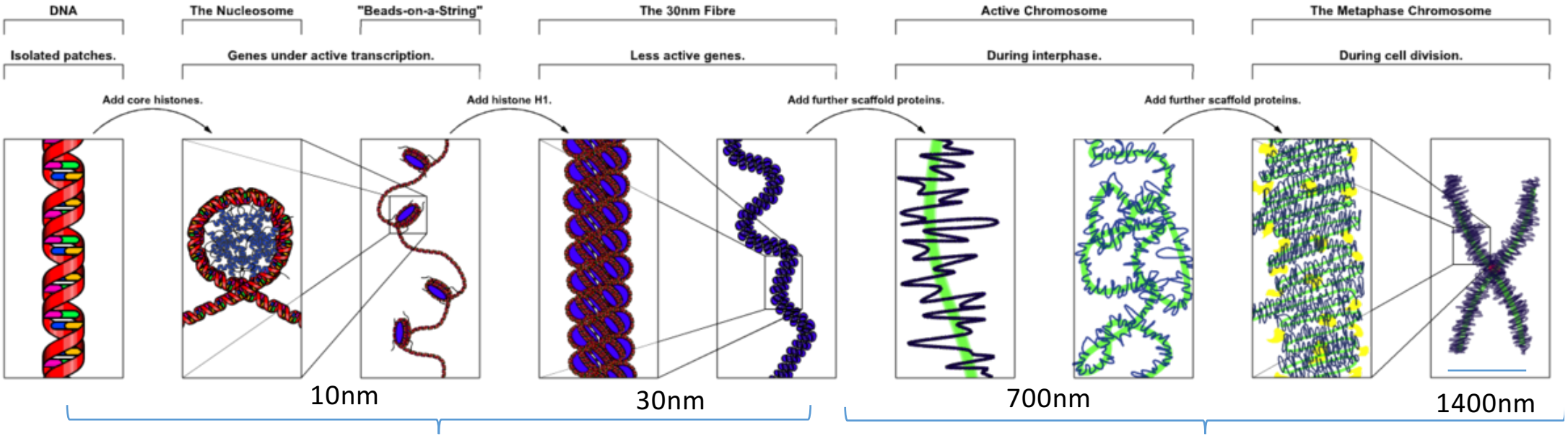
Internal matrix proteins



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**

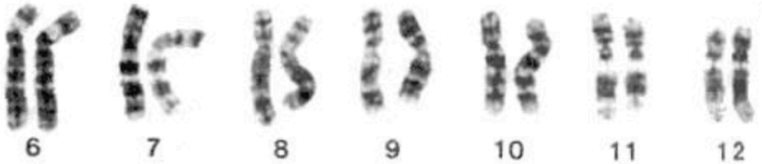
Transitions between the 30-nm and 10-nm fibers are thought to be essential for the control of chromatin transcriptional status.

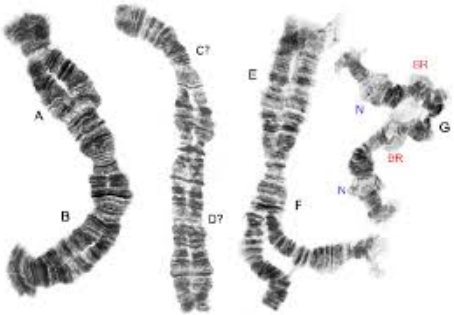
Euchromatin **Facultative heterochromatin** **Consitutive heterochromatin**

Chromatin comes in different flavors

Different types of chromatin

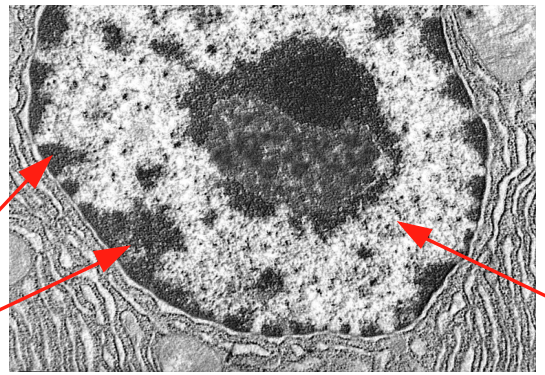
metaphase





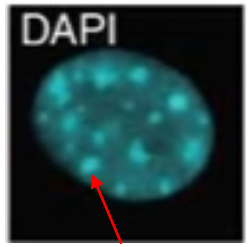
Giemsa

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"