Structure of eukaryotic genome, its replication and gene expression

Molecular Biology

RNDr. Jan Hošek, Ph.D.
hosekj@vfu.cz

Structure of eukaryotic genome

Characteristic features of eukaryotic cell

- Nucleus consists of chromatin
  - dsDNA
  - histones
  - nonhistones = proteins of non-histonic character
- Chromosomes contain linear dsDNA
- Nucleus is divided by mitosis
- Cell wall consists of cellulose (plants) or chitin (fungi), animal cells have no cell wall
- They have organelles – mitochondria and plastids
- Endoplasmic reticulum, Golgi system, and lysosomes

Genom of eukaryotic organisms

Animal cells: nucleus and mitochondria

Plant cells: nucleus, mitochondria and chloroplasts

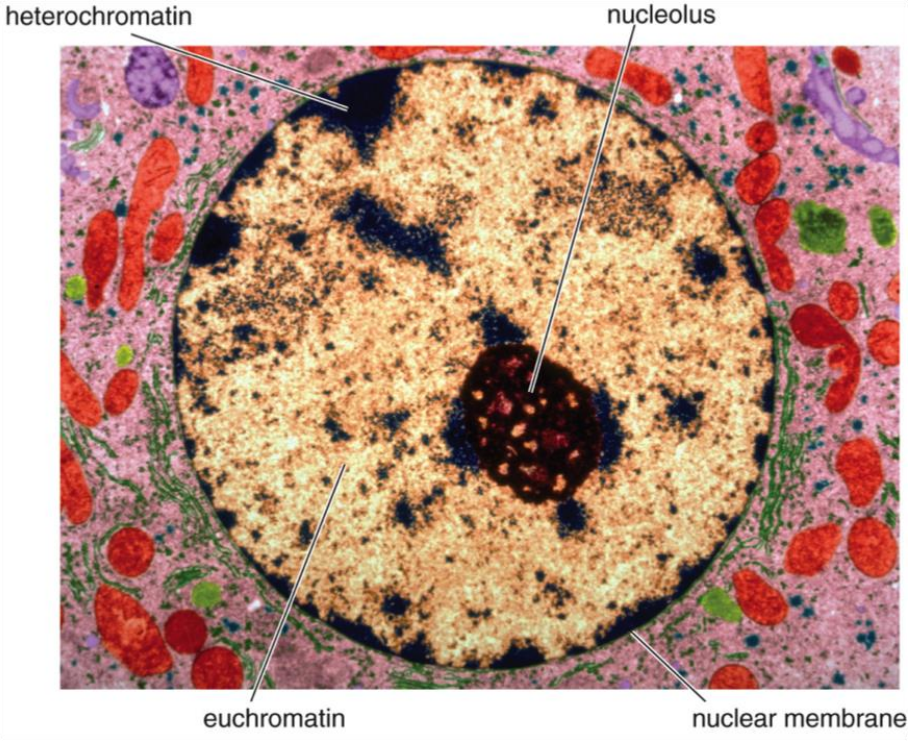
- **chromosomal (AKA nuclear) DNA (nDNA)**
- **mitochondrial DNA (mtDNA)**
- **chloroplast DNA (ctDNA)**
- **plasmids**

Chromatin

- Stainable material, which forms the nucleus of eukaryotic cells
- dsDNA, histones, nonhistones

According to the ability to be stained by basic dyes and degree of condensation we distinguish:

- **euchromatin** – weakly stainable, decondensed, "transcriptionally active"
- **heterochromatin** – strongly stainable, condensed, "transcriptionally inactive"



https://www.studyblue.com/notes/note/n/3-chromosomes/deck/4743713

Heterochromatin

Constitutive

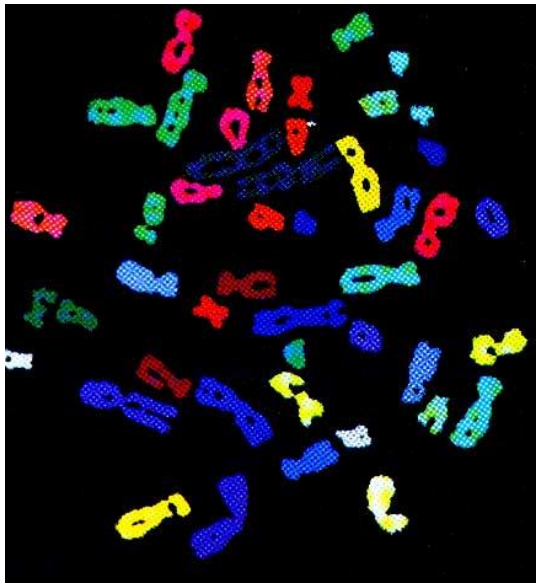
- Constantly in heterochromatin stage
- centromeres and telomeres
- One of X chromosome in women

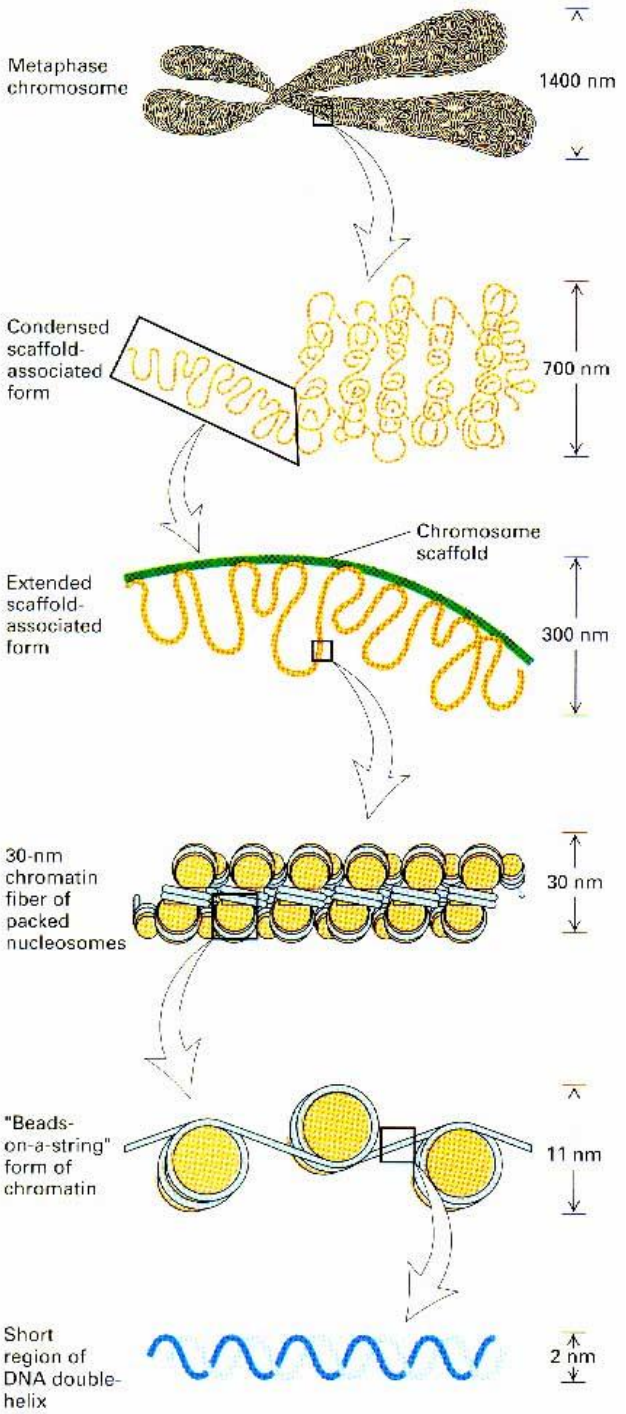
Facultative

- Switches between heterochromatin and euchromatin on the base of oncogenetic development of organism

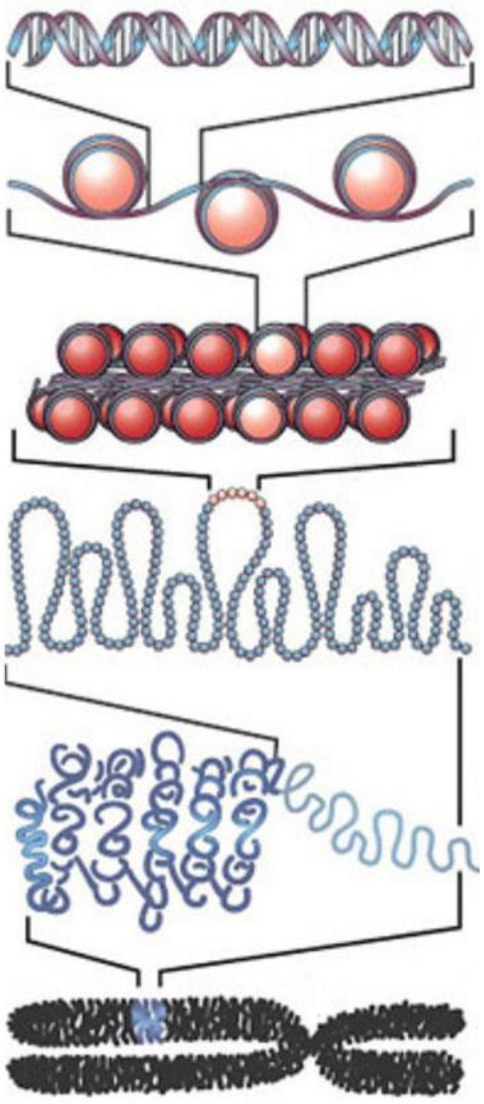
Chromatin condensation

1) Basic structure = interphase = decondensed 10-nm chromatin fiber (beads-on-a-string)
2) 30-nm chromatin fibre
3) Chromatin in mitotic phase = mitotic chromosomes





Chromatin condensation



„naked“ dsDNA

“beads-on-a-string“ form nucleosomes

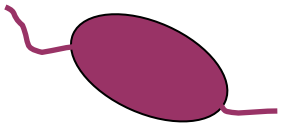
30 nm solenoid

relaxed form of chromosome

condensed region of chromosome

mitotic chromosome

Chromatin components



1) Histones

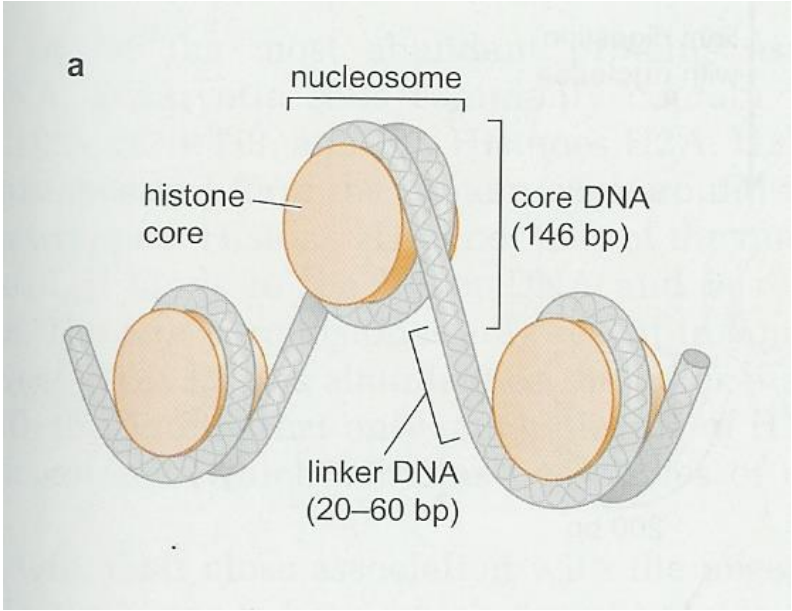
- Centre is globular, ends are flexible and filamentous
- High content of arginine and histidine
- 5 species = H1, H2A, H2B, H3 and H4

2) Nonhistones

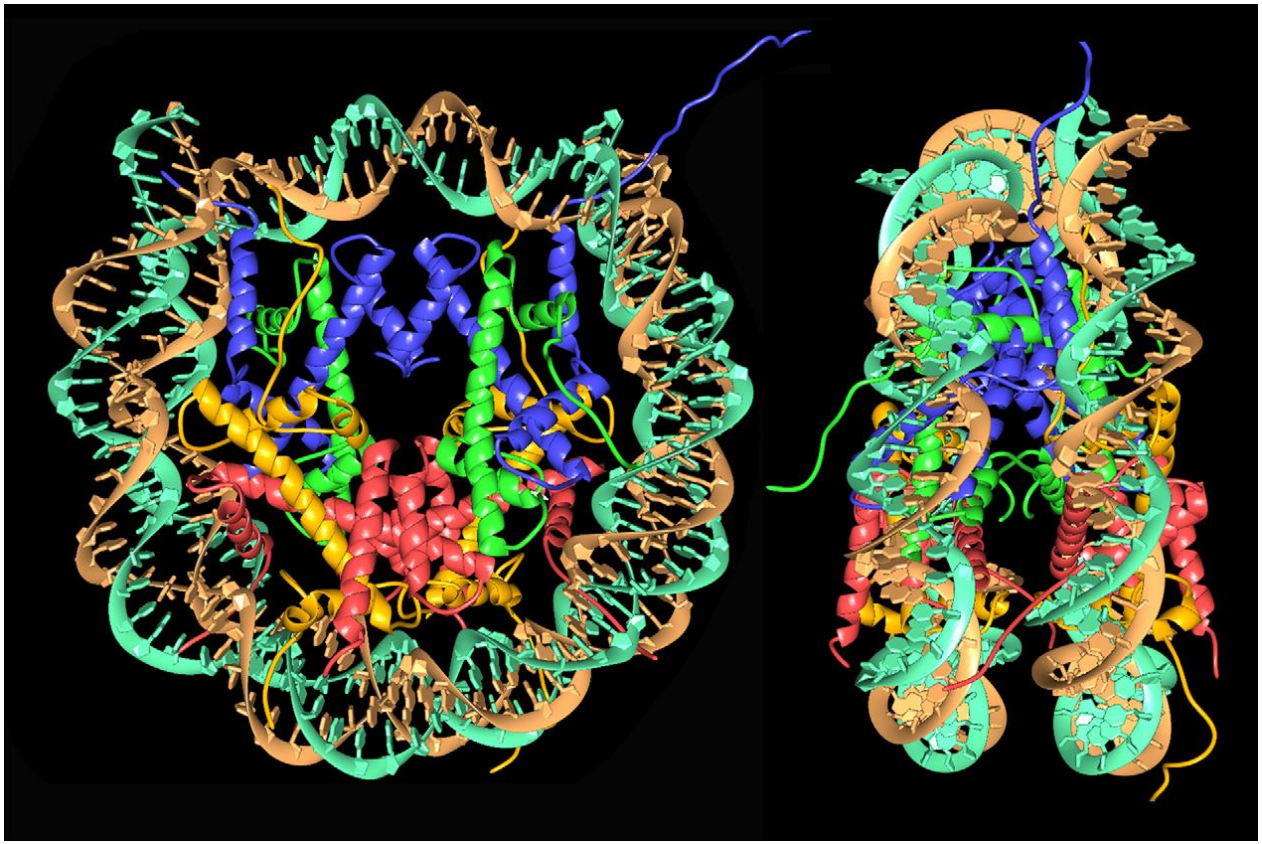
- RNA polymerase and other enzymes usable in transcription
- HMG1 and HMG2 – bind to unusual DNA structures
- HMG3 a HMG4 – bind to histone core especially in transcriptionally active regions

Nukleosome

- **The basic unit of chromatin**
- **octamer of histones $(H2A, H2B, H3, H4)_2$**
- **One molecule of histone H1**
- **DNA segment 200 bp long, which is wound about 2 times around the octamer of histones**



Nucleosome structure

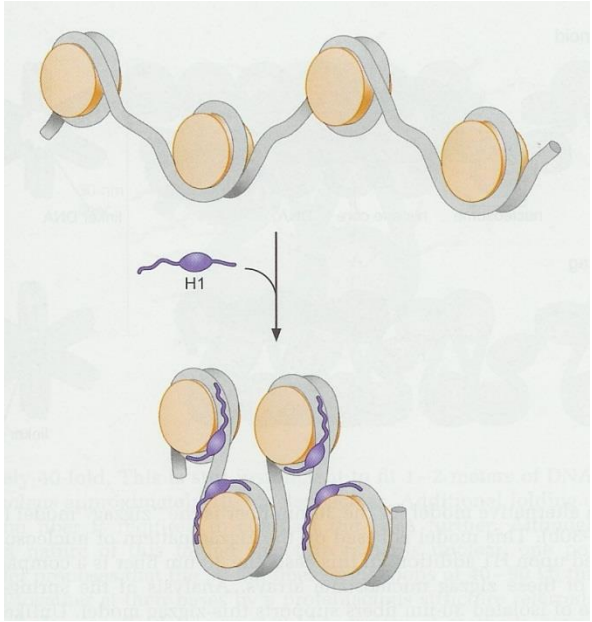


H2A – yellow, H2B red, H3 blue, H4 green

Nature 389: 251–260 (1997)

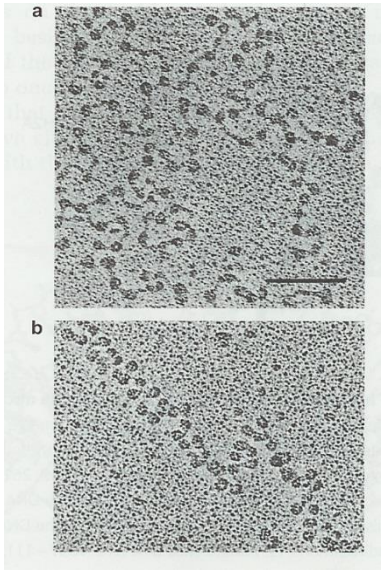
Nucleosome fibre

- 10nm chromatin fibre
- its individual items form nucleosome cores connected by long linear dsDNA
- visible by microscope



- H1

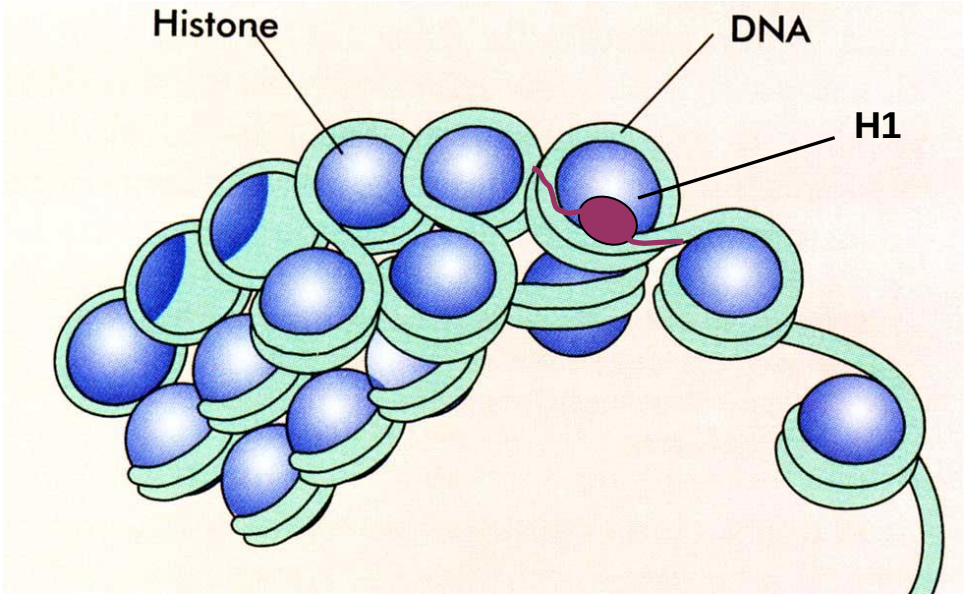
+ H1



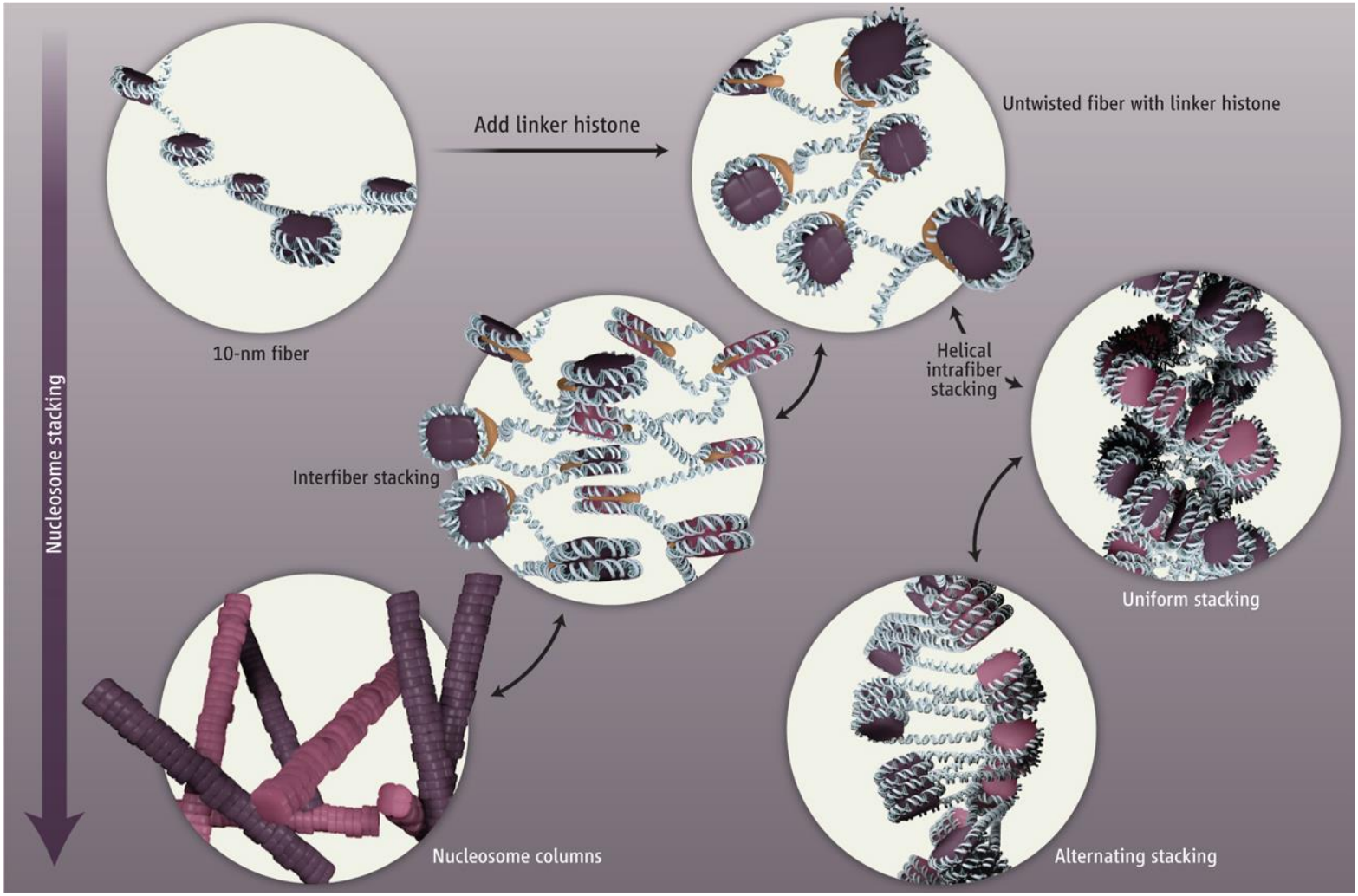
30nm chromatin fibre

30nm chromatin fibre

- **It is created by the condensation of nucleosome fibre caused by histone H1**
- **It binds to protein scaffold (nonhistones, e.g. topoisomerase II)**



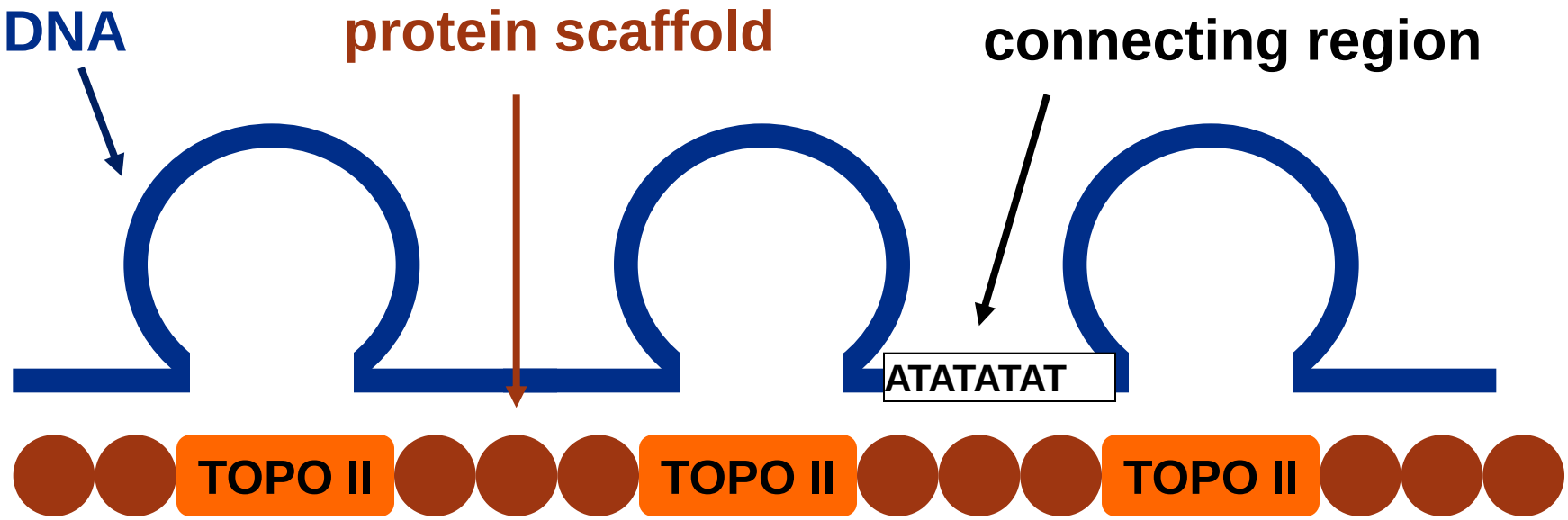
Stacking for packing.Possible modes of how a chromatin fiber folds into helical or columnar forms are shown.



A Travers Science 2014;344:370-372

Chromatin domains

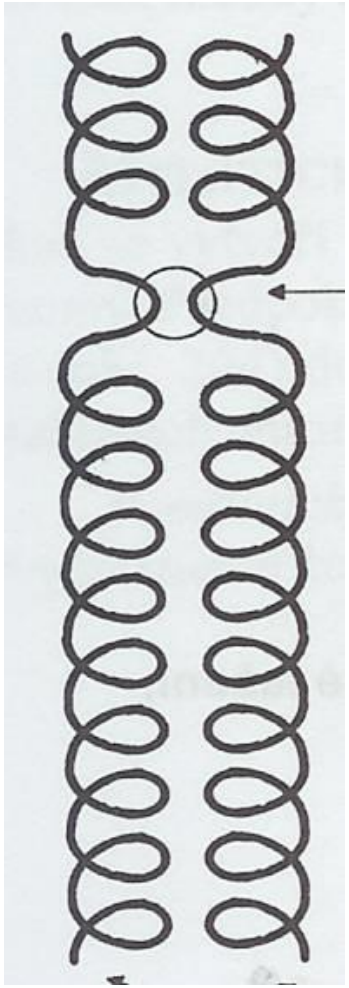
- loops of 30nm chromatin fibre attached to protein scaffold
- there is one molecule of topoisomerase II in base of each loop = change of topology during replication and transcription
- each domain has one *ori* locus

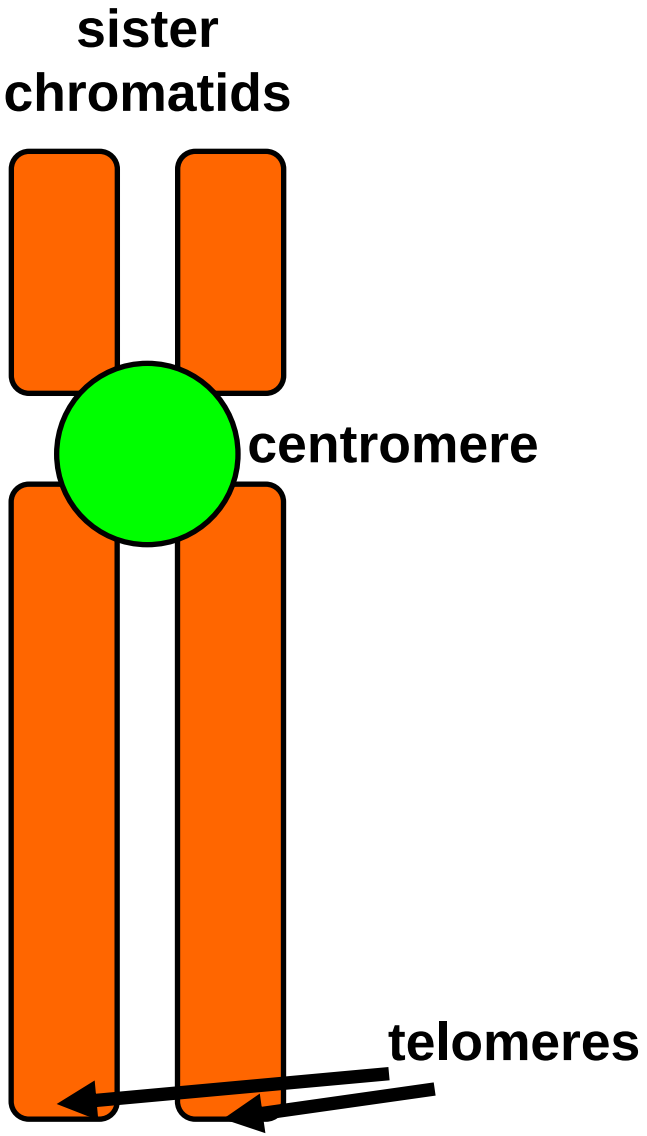


Mitotic chromosomes

- They originate by condensation of 30nm chromatin fibres
- They are formed during mitosis or meiosis
- Condensation of 30nm to 600-700nm chromatin fibres, which constitute the structure of chromosomes
- In chromosomes, the chromatin is in the stage of the highest condensation and is transcriptionally inactive

Mitotic chromosomes





Chromosomal (nuclear) DNA

- a single linear molecule of dsDNA
- the number of bp of haploid cell is about $1.34 \times 10^7 – 1.5 \times 10^{10}$
- only 10% of mammal genome bears the genetic information
- most structural genes are from 1×10^4 to 2×10^6 bp long – considerable part is created by regularoty sequences

Repeats in nuclear DNA

- **short tandem repeats – not in prokaryotes**
- **dispersed repeats – not in prokaryotes**
- **25% - 50% structural genes as unique sequence**
- **the rest as gene repeats = repetitions has function of a gene**

Gene repeats

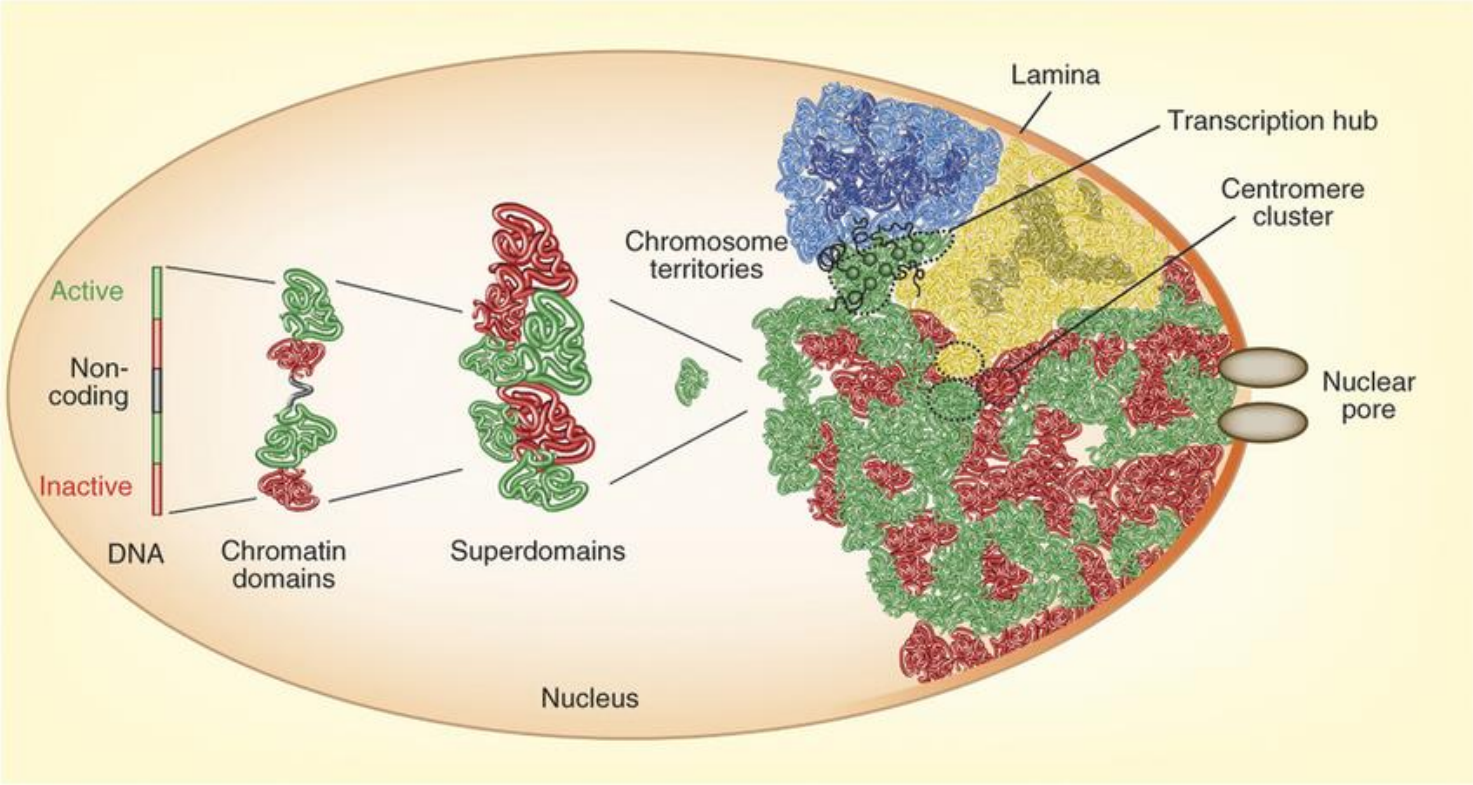
- **Gene family = is a set of several similar genes, formed by duplication of a single original gene, and generally with similar biochemical functions. E.g. the genes for haemoglobin subunits. The term pseudogene.**
- **Tandem gene repeats = directly adjacent, the term spacer (intergenic sequence, separates genes or a group of genes), genes transcribed to 5S-rRNA, genes for tRNA and genes for histones. The repeats provide sufficient amounts for the cell requirements.**
- **Dispersed gene repeats = their copies are dispersed at different positions of haploid genome, genes transcribed to tRNA, snRNA and others.**

Pseudogene

- is a dysfunctional relative of known genes that lost their protein-coding ability or are otherwise no longer expressed in the cell
- usually has no introns

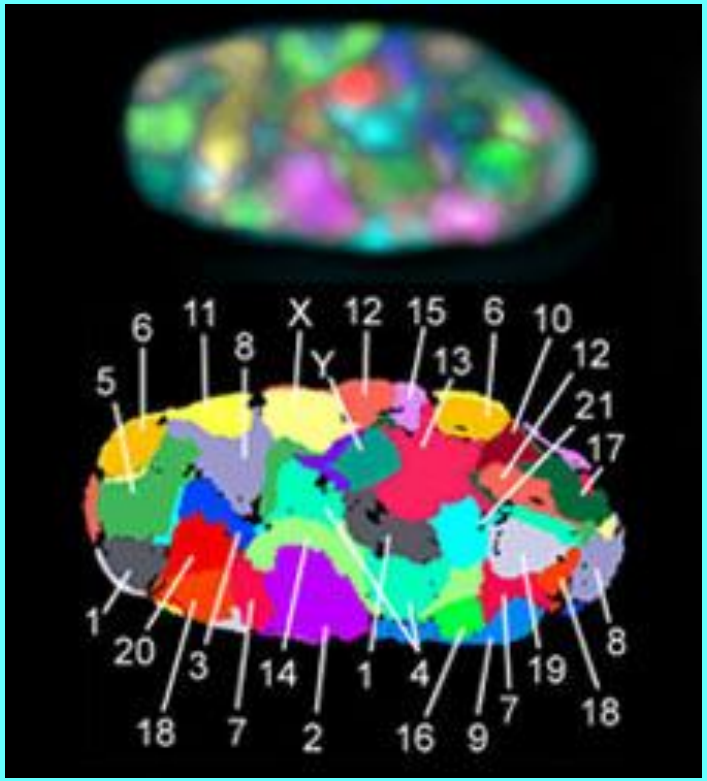
Organisation of chromatine in nucleus

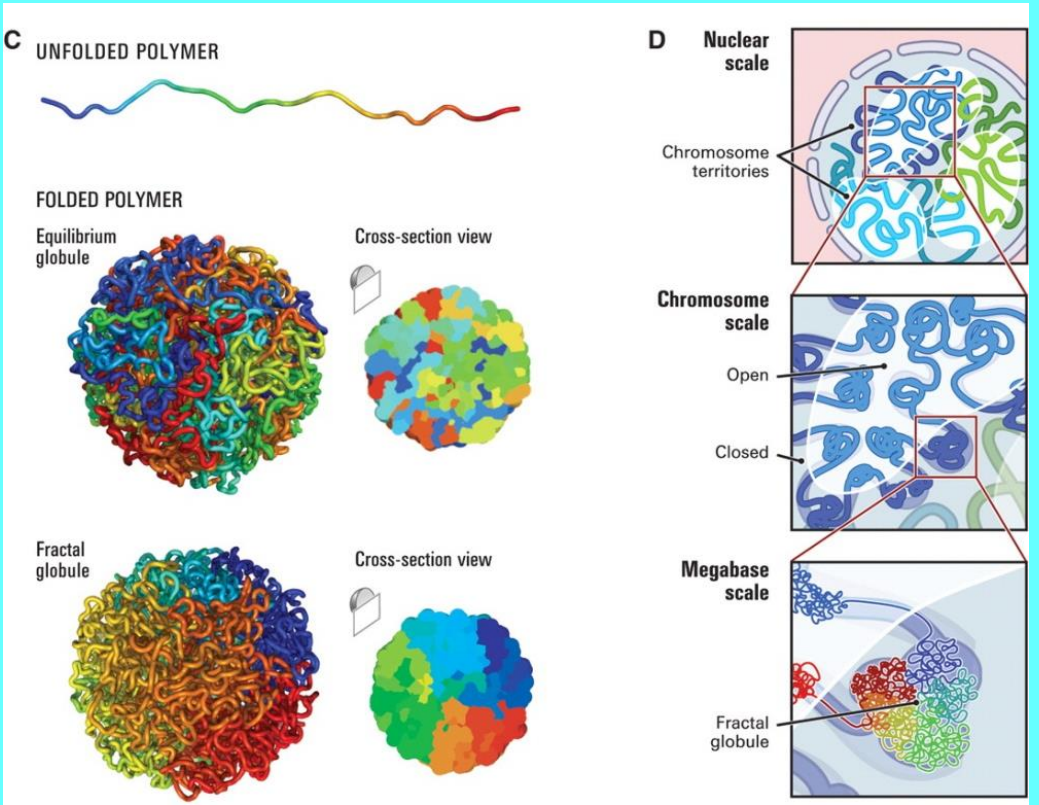
- Localisation of chromosomes is not accidental
- Crowding of region with similar function or activity



DOI: 10.1038/nsmb.2474

The local packing of chromatin is consistent with the behavior of a fractal globule





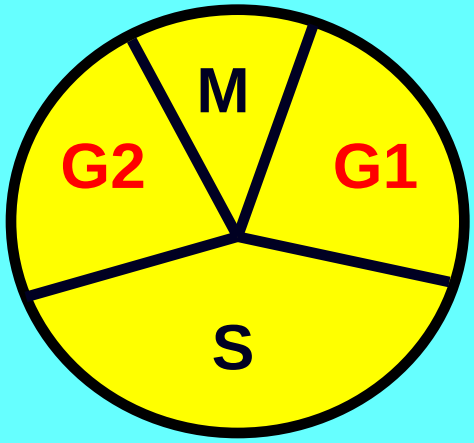
E. Lieberman-Aiden et al., Science 326, 289-293 (2009)

https://unlockinglifescode.org/the-genome-ball

DNA replication in eukaryotes

Replication of eukaryotic genome

- replication of mitochondrial and chloroplat DNA
- replication of nuclear chromosomes
  - semiconservative and semidiscontinuous
  - initiation, elongation, and termination
  - only in S phase of the cellular cycle

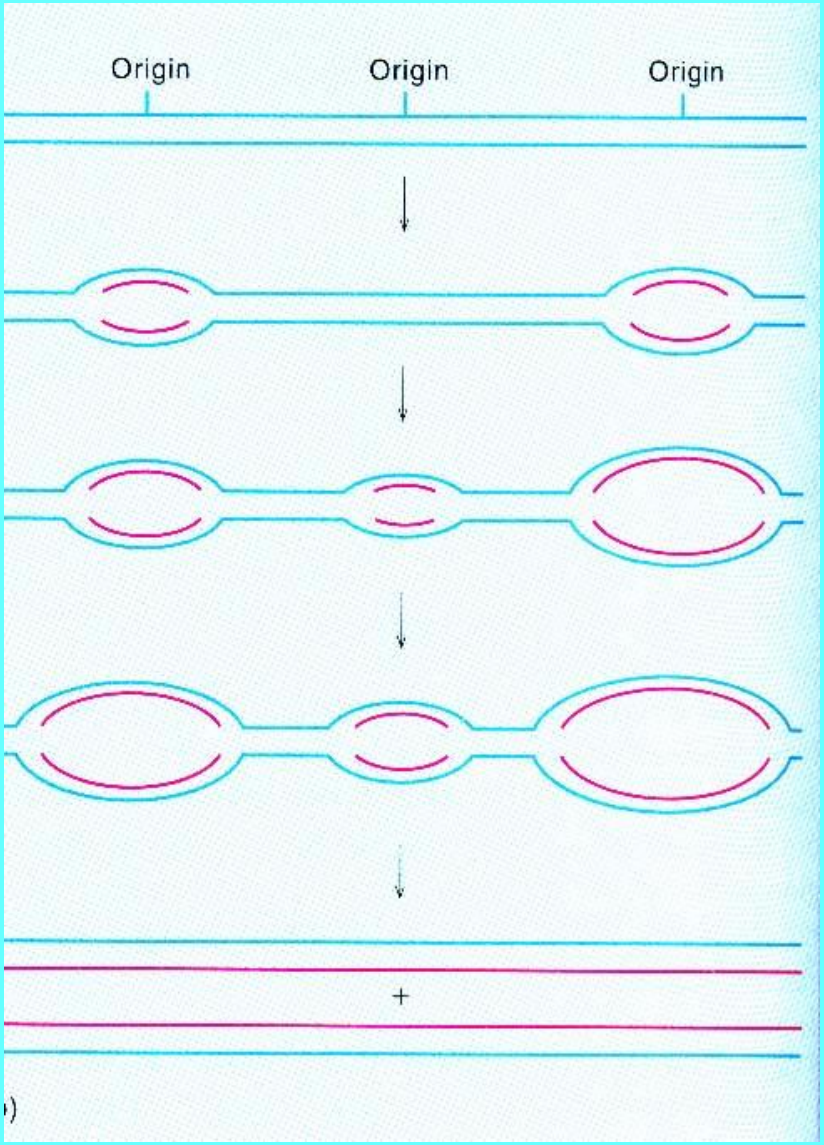


transcription,

translation,

metabolism

Replication of nuclear chromosomes

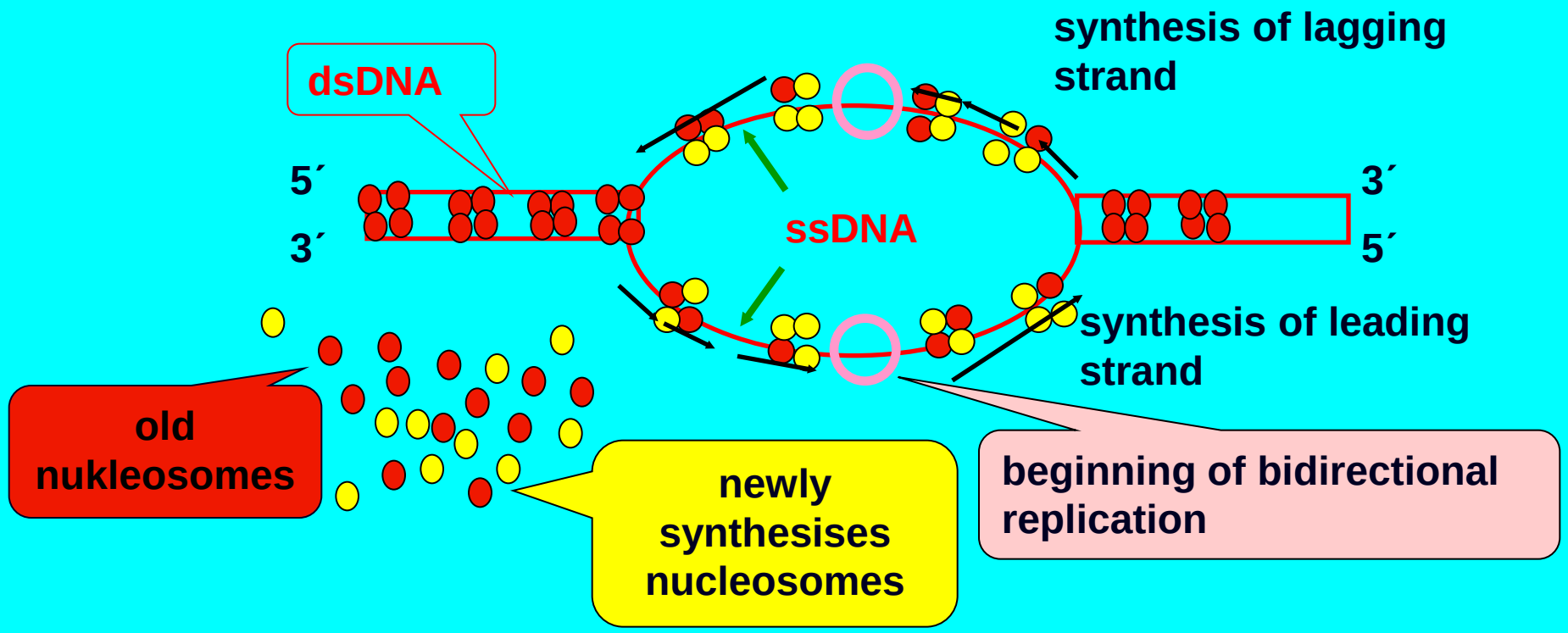


In contrast to prokaryotic cells the eukaryotic replication proceeds on several places in time

Chromosome is a couple of replicons, it has more *ori* sequences (mammals 30.000-50.000)

Euchromatin replicates earlier to heterochromatin

Process of replication

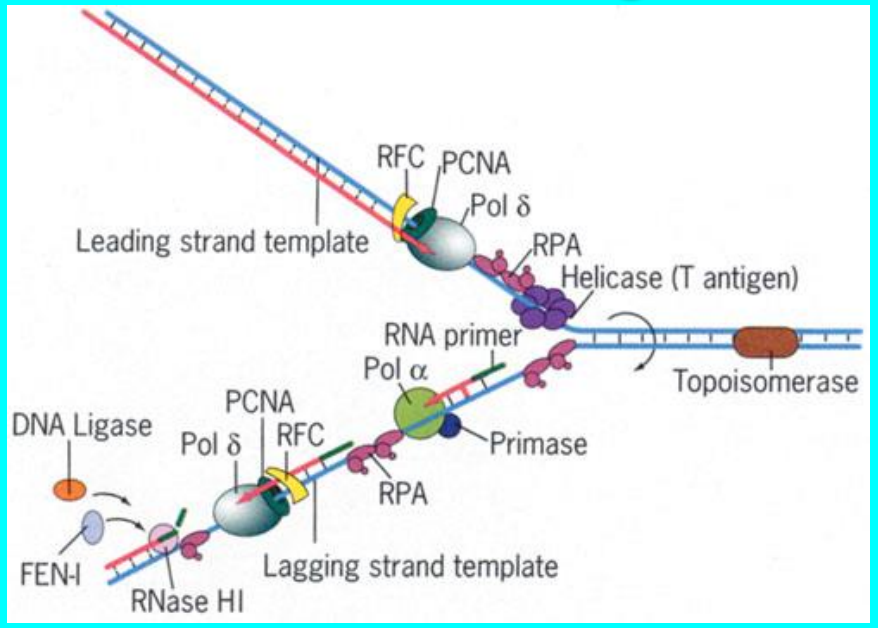


Eukaryotic DNA-polymerase: α, β, γ, δ and ε

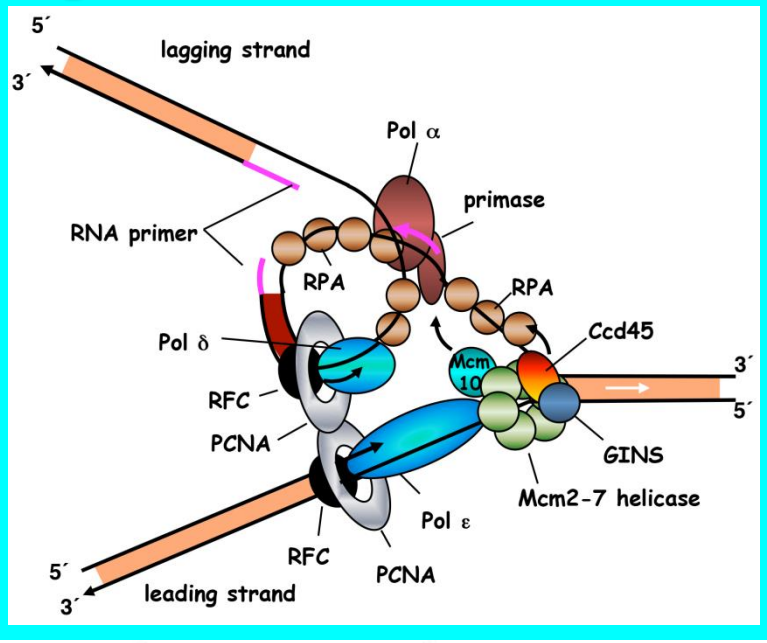
Eukaryotic DNA-polymerase

- DNA polymerase α – in the complex of primase synthetises Okazaki fragments, it does not possess 3´-5´exonuclease activity (proof reading)
- DNA polymerase β – synthesis of short fragments during DNA reparation
- DNA polymerase γ – synthesis of mitochondrial DNA
- DNA polymerase δ – synthesis of leading strand and completing of lagging strand
- DNA polymerase ε – probably synthesis of leading strand

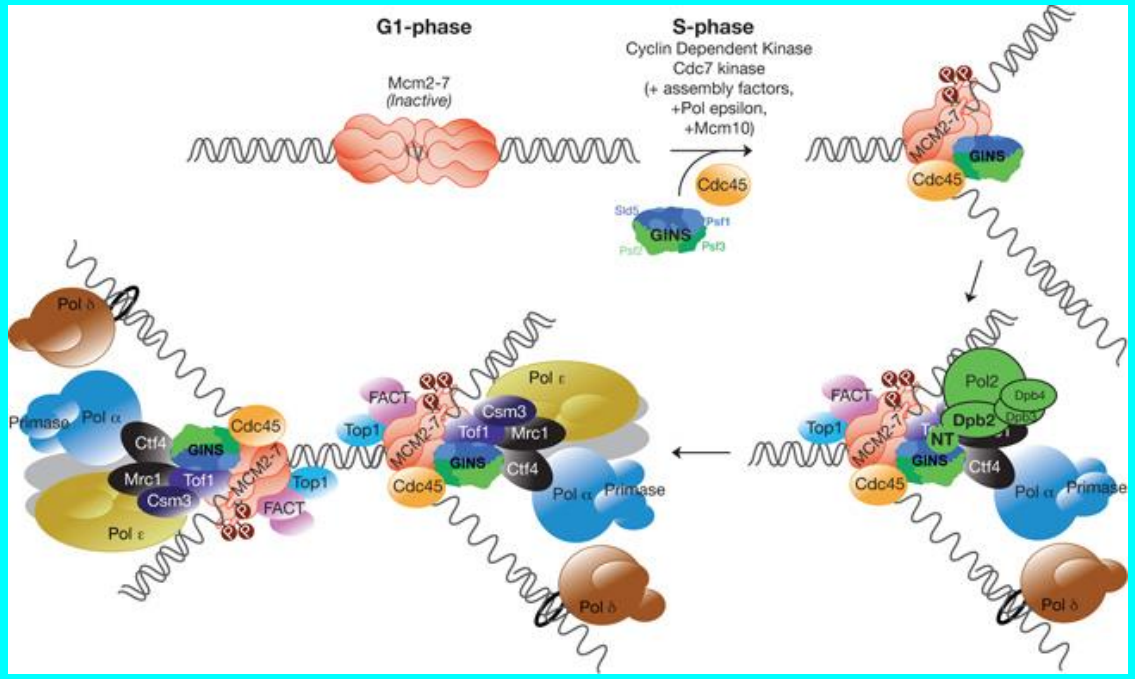
Eukaryotic replication fork



http://yxsj.baiduyy.com/

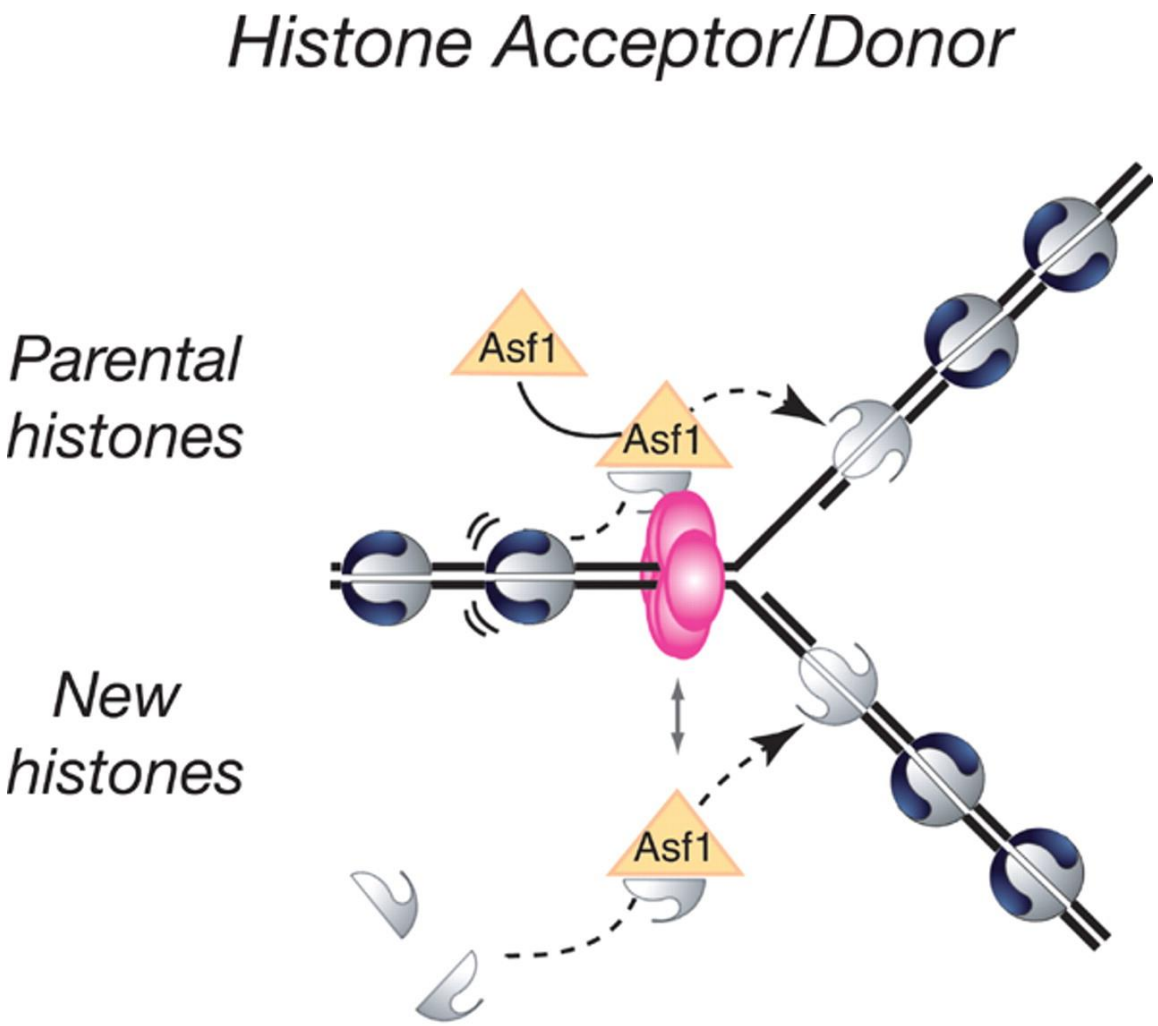


http://www.leibniz-fli.de/research/research-groups/grosse/



http://www.ppu.mrc.ac.uk/research/?pid=1012&sub1=research

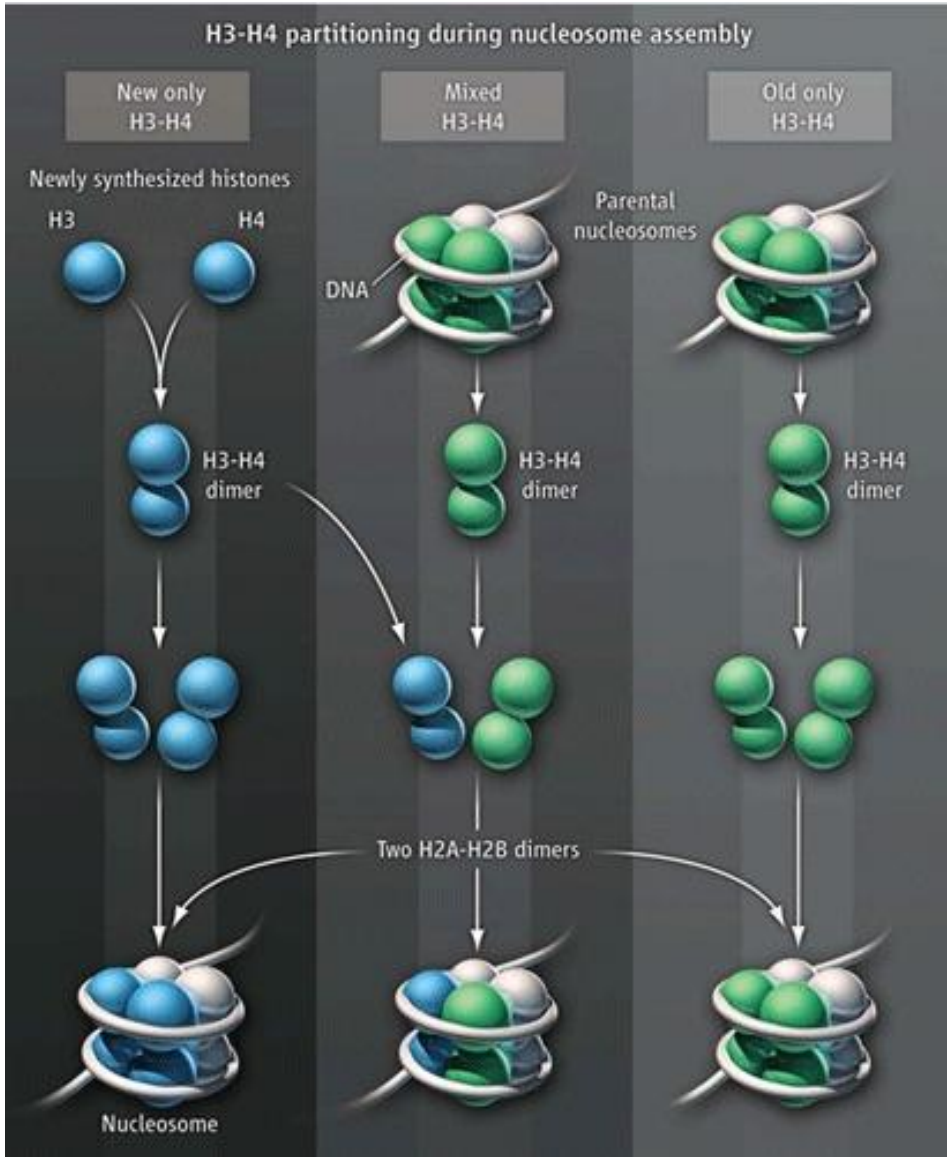
Model for Asf1 function in replication as a histone acceptor and donor



A. Groth et al., Science 318, 1928 -1931 (2007)

Science AAAS

Histone partitioning

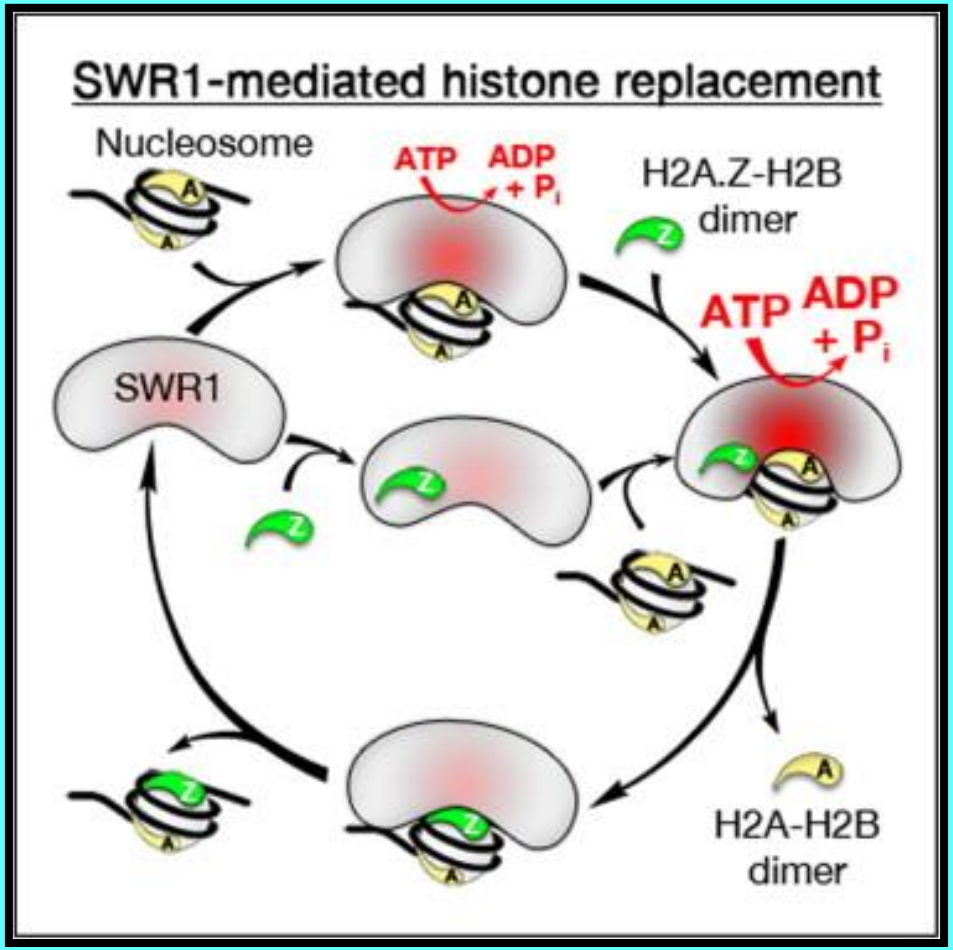


D. Ray-Gallet et al., Science 328, 56-57 (2010)

New nucleosomes result from de novo assembly using newly synthesized H3 and H4 histones in the form of two H3-H4 dimers; after association with two H2A-H2B dimers, the result is a nucleosome containing only new H3-H4 dimers. Mixed particles will form using a newly synthesized H3-H4 dimer and an H3-H4 dimer recycled from a disrupted parental nucleosome. Old nucleosomes will form either by self-reassociation of two H3-H4 dimers recycled from a transiently disrupted parental nucleosome, or according to the generally accepted view, by inheritance of a stable H3-H4 tetramer from a parental nucleosome.

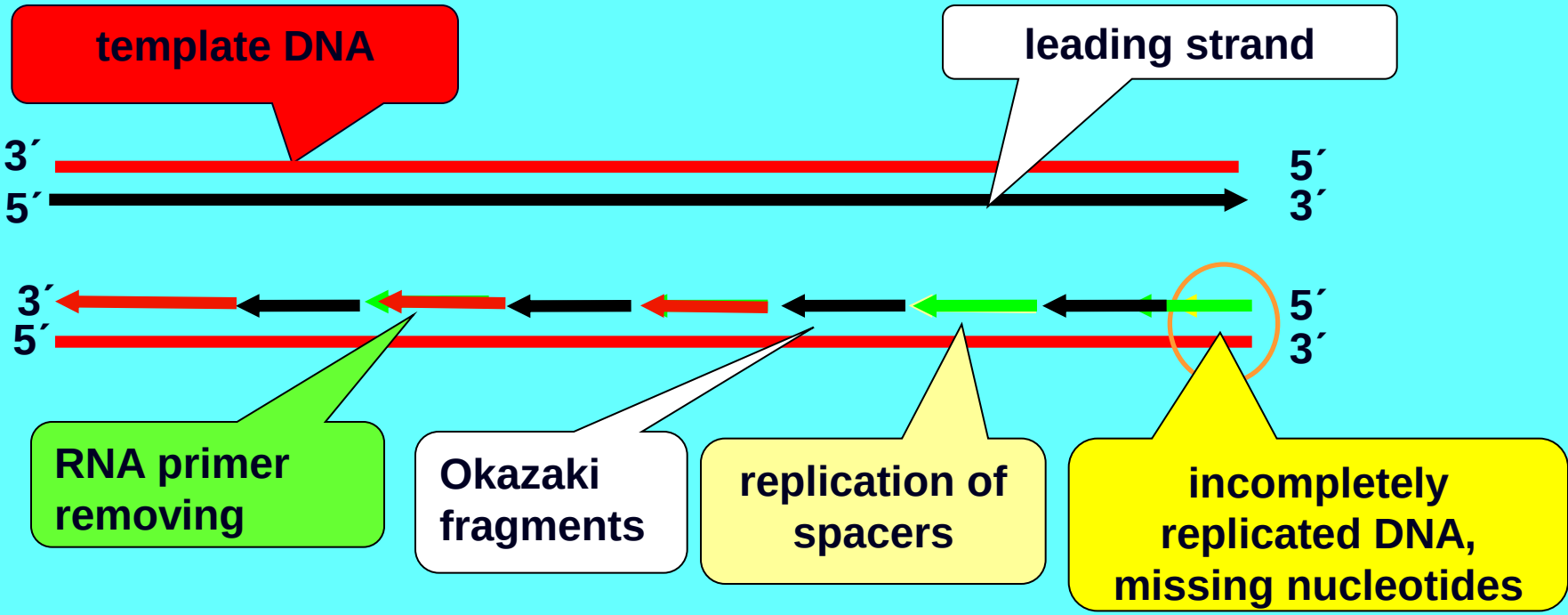
Science AAAS

Nucleosome assembling is associated with hydrolysis of ATP



E. Luk et al., Cell 143, 725–736, November 24, 2010

Scheme of replication of linear molecules

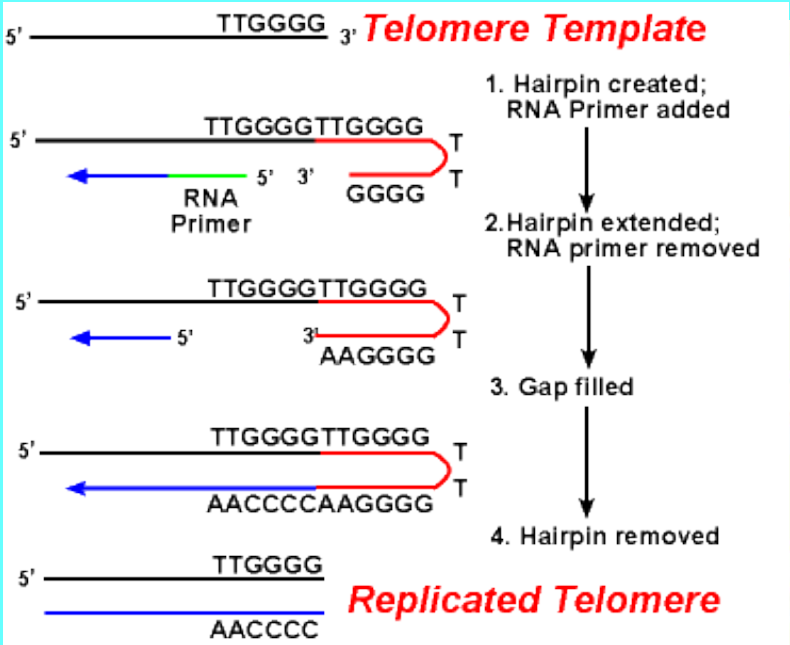


the end replication problem

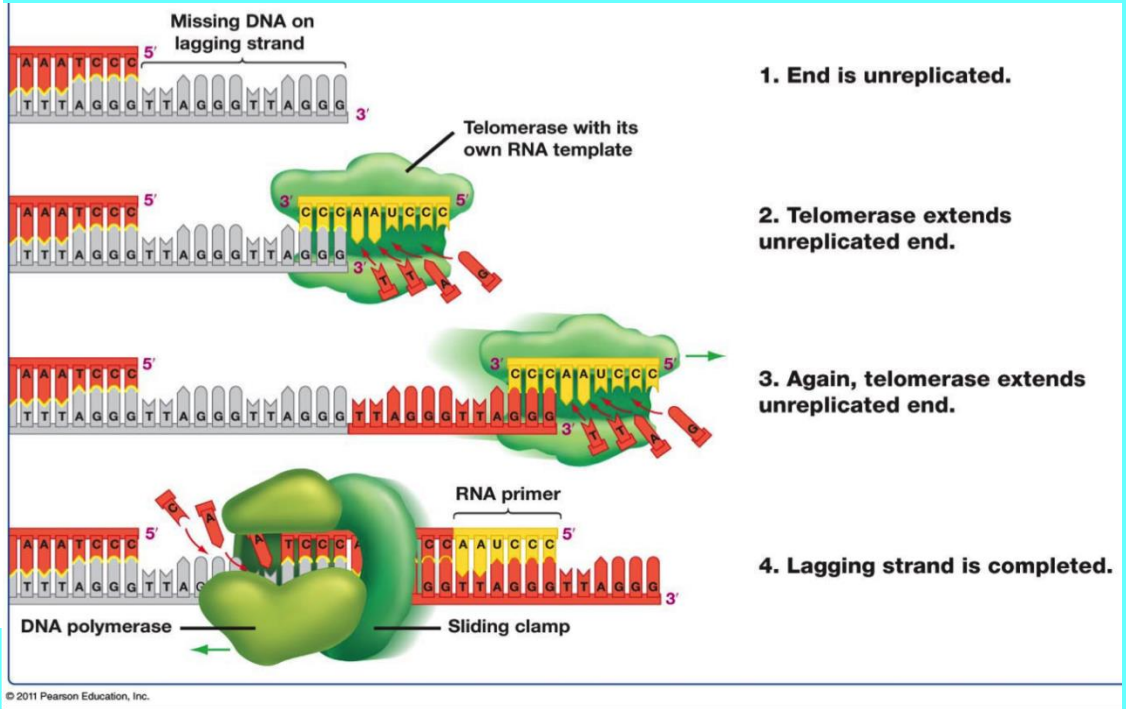
Telomerase = ribonukleoprotein - RNA acts as a template, protein has catalytic function

Filling of missing 3´-ends

- Telomerase elongates the 3´-end
- Formation of hairpin and RNA primer
- Replication of complementary strand and removing of hairpin



https://www.ndsu.edu/pubweb/~mcclean/plsc431/eukarychrom/eukaryo3.htm



http://masteringyourwaytomedschool.blogspot.cz/p/bio-1000-dna-shortening.html

Sequences of telomers

TABLE 11.5

Telomeric Repeat Sequences Within Selected Organisms

Group	Examples	Telomeric Repeat Sequence
Mammals	Humans	TTAGGG
Slime molds	*Physarum, Didymium*	TTAGGG
	Dictyostelium	$AG_{(1–8)}$
Filamentous fungi	*Neurospora*	TTAGGG
Budding yeast	*Saccharomyces cerevisiae*	$TG_{(1–3)}$
Ciliates	*Tetrahymena*	TTGGGG
	Paramecium	TTGGG(T/G)
	Euplotes	TTTTGGGG
Higher plants	*Arabidopsis*	TTTAGGG

http://reasonandscience.heavenforum.org/t2263-the-telomerase-enzyme

The end-protection problem

VIEWPOINT

How Telomeres Solve the End-Protection Problem

Titia de Lange

The ends of eukaryotic chromosomes have the potential to be mistaken for damaged or broken DNA and must therefore be protected from cellular DNA damage response pathways. Otherwise, cells might permanently arrest in the cell cycle, and attempts to "repair" the chromosome ends would have devastating consequences for genome integrity. This end-protection problem is solved by protein-DNA complexes called telomeres. Studies of mammalian cells have recently uncovered the mechanism by which telomeres disguise the chromosome ends. Comparison to unicellular eukaryotes reveals key differences in the DNA damage response systems that inadvertently threaten chromosome ends. Telomeres appear to be tailored to these variations, explaining their variable structure and composition.

Of the three major questions in telomere biology, two were solved in the 1980s. First, the nature of the DNA sequences that confer telomere function onto chromosome ends was revealed when Blackburn and Szostak showed that the short G-rich repeats from the ends of yeast chromosomes were sufficient to stabilize a linear plasmid (*1*, *2*). Since then it has become clear that G-rich repeats cap the ends of most eukaryotic chromosomes, including mammalian chromosomes that end in TTAGGG repeats.

Second, the mechanism by which telomeric DNA is maintained was resolved when Blackburn and Greider showed that telomeric DNA is synthesized by telomerase. Telomerase is a reverse transcriptase that adds telomeric repeats to the 3′ ends of each chromosome (*3*). In doing so, telomerase makes up for the shortcomings of semiconservative DNA replication, which cannot complete the synthesis of chromosome ends. Other solutions to this end-replication problem exist, notably in *Drosophila* and other dipterans, but it is now clear that telomerase is the main method by which eukaryotes avoid sequence loss at the ends of their chromosomes.

It has been suggested that early eukaryotes used a primitive form of telomeres without telomerase to solve the end-replication problem (*4*). The later acquisition of telomerase not only solved the end-replication problem but ensured the presence of the same sequence at all chromosome ends. Once all telomeres in the cell had the same sequence, telomeric DNA binding factors could evolve, thereby enabling cells to distinguish natural chromosome ends from sites of DNA damage.

The End-Protection Problem

Research on the third major issue in telomere biology, how telomeres solve the end-protection problem, stagnated until the 1990s. The end-protection problem first surfaced early last century, when Muller and McClintock observed a critical distinction between the behavior of broken chromosome ends and telomeres. Muller found that chromosomes lacking their natural ends were unstable; McClintock documented the propensity of broken ends, but not telomeres, to fuse. However, the full extent of the end-protection problem remained obscure until the principles of the DNA damage response were revealed in the 1980s.

The first insight came when Szostak, Rothstein, and Orr-Weaver found that linear DNA introduced into eukaryotic cells is unstable because the DNA ends recombine with the genome (*5*). It is now clear that introduced linear DNA falls victim to two important DNA repair pathways that mend broken chromosomes: homology-directed repair (HDR) and nonhomologous end joining (NHEJ). The observation that DNA ends (also known as double-strand breaks) are processed by these DNA repair reactions raised the question of whether the natural ends of chromosomes are also attacked by HDR and NHEJ, and if not, why not.

A second question arose from the work of Hartwell and Weinert, who found that budding yeast lacking the *RAD9* gene failed to arrest the cell division cycle in response to double-strand breaks (*6*). This experiment, and earlier observations on fission yeast and mammalian cells (*7*), revealed that the cell cycle arrest normally associated with DNA damage is not due to the DNA damage itself. Rather, cells arrest because of the activation of a pathway that detects DNA damage and blocks cell cycle progression in response. Why, then, are these pathways not activated by the natural ends of linear chromosomes?

These findings on how eukaryotes respond to DNA damage shaped the current molecular definition of the end-protection problem: How do telomeres prevent the activation of the DNA damage signaling pathways, and why are they resistant to the repair pathways that act on DNA ends?

In the context of mammalian cells, the end-protection problem can be rephrased in more precise terms, based on current knowledge of the molecular pathways that recognize and repair double-strand breaks (Fig. 1). Mammalian cells have two independent signaling pathways that are activated by double-strand breaks: (i) the ATM (ataxia telangiectasia mutated) kinase path-

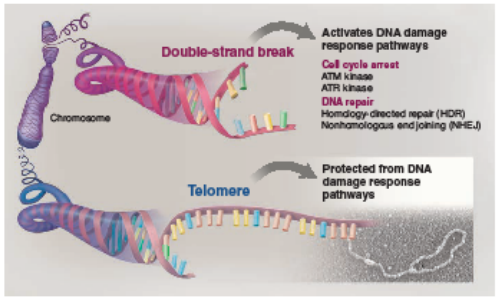
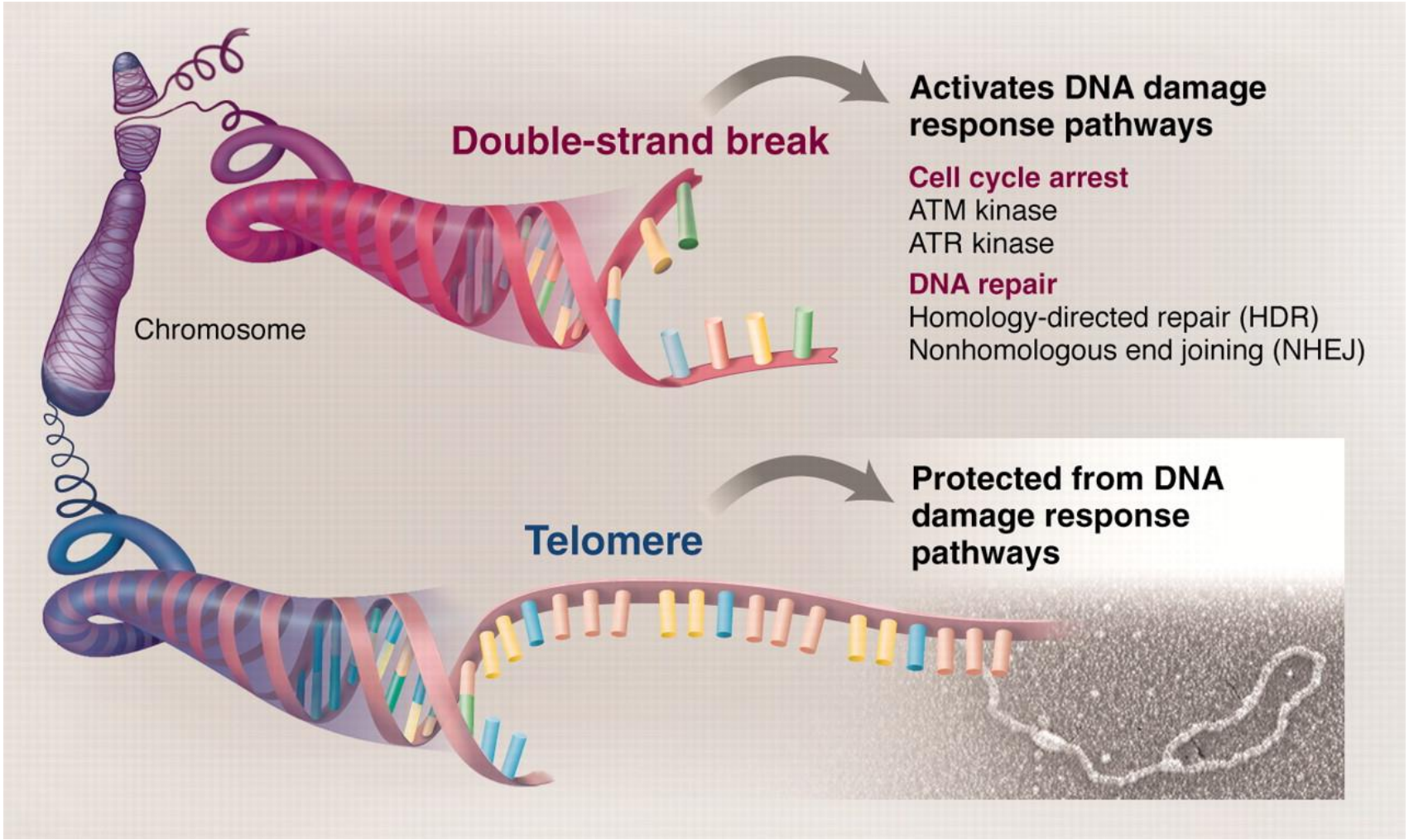


Fig. 1. The end-protection problem. When a mammalian chromosome breaks (top), the exposed DNA ends can activate two signaling pathways (the ATM and ATR kinase pathways) that arrest the cell division cycle and can induce cell death. The broken chromosome is usually repaired by one of two different DNA repair pathways (NHEJ and HDR), allowing cells to continue their divisions with an intact genome. The presence of these DNA damage response pathways poses a problem for the ends of linear chromosomes (telomeres, bottom) because activation of DNA damage signaling or DNA repair at telomeres would be disastrous. Mammalian telomeres solve this end-protection problem through the use of a telomere-specific protein complex (shelterin) and an altered structure (the t-loop) that together ensure that all four pathways remain blocked.

Laboratory of Cell Biology and Genetics, Rockefeller University, New York, NY 10021, USA. E-mail: delange@mail.rockefeller.edu

T. de Lange Science 326, 948-952 (2009)

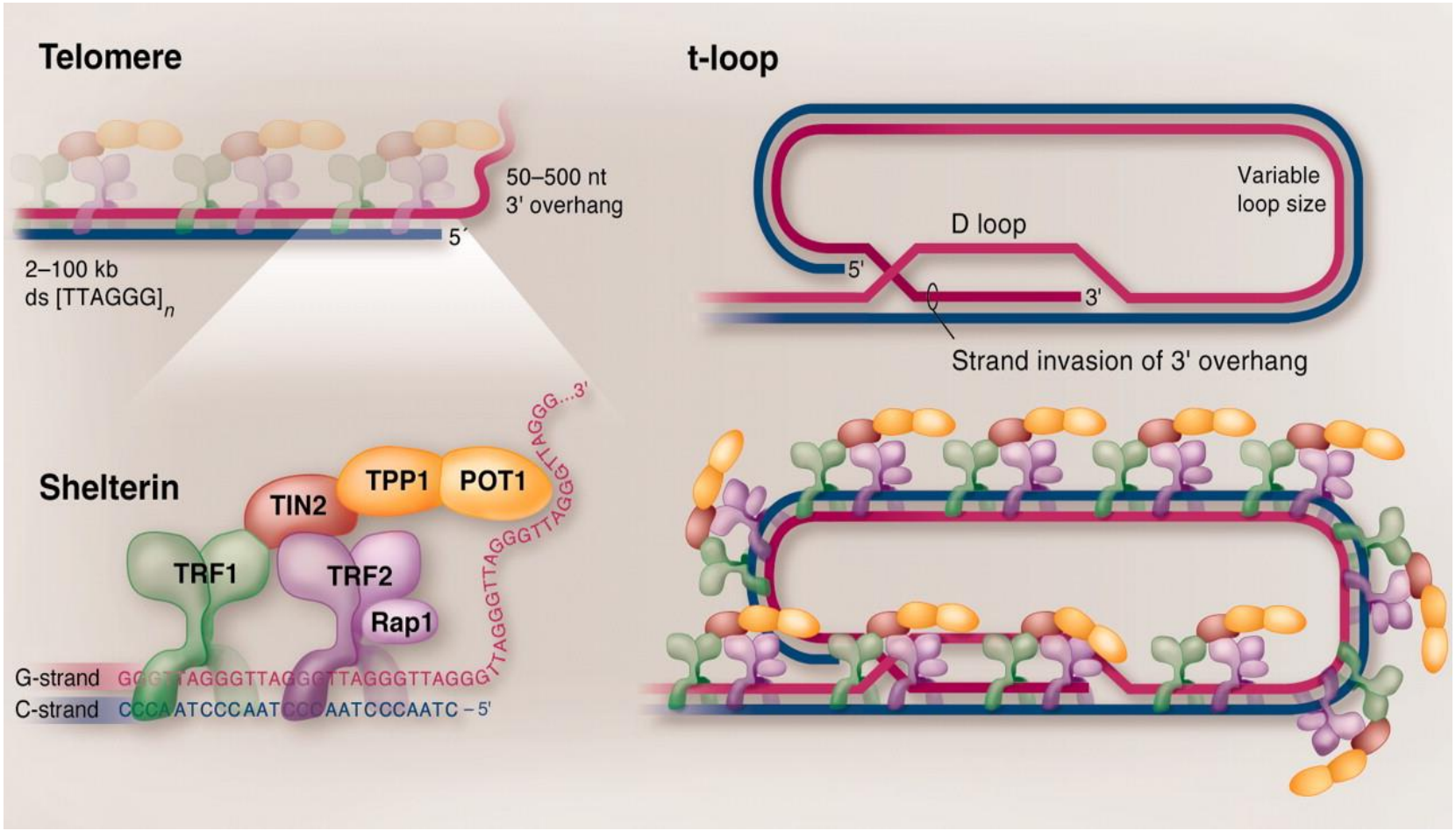
Fig. 1: The end-protection problem



T. de Lange Science 326, 948-952 (2009)

Science AAAS

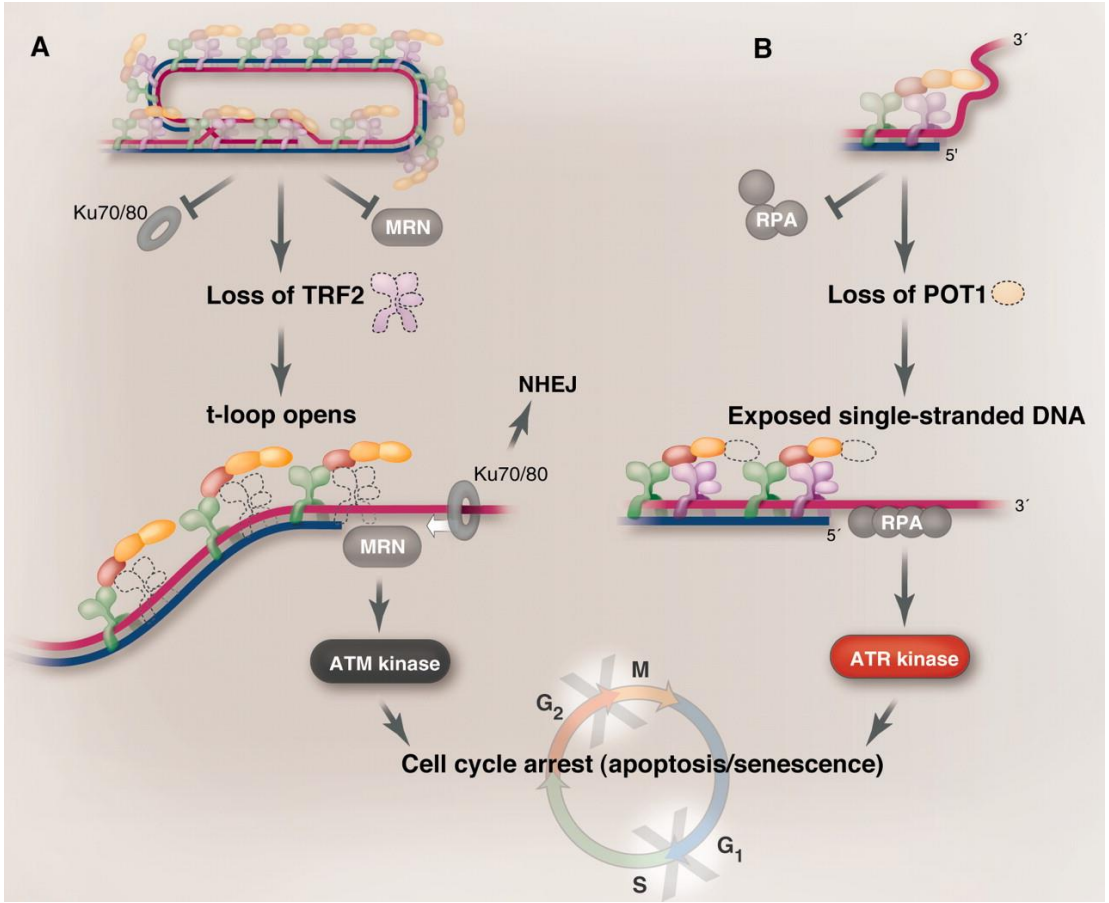
Fig. 2 Mammalian telomeres



T. de Lange Science 326, 948-952 (2009)

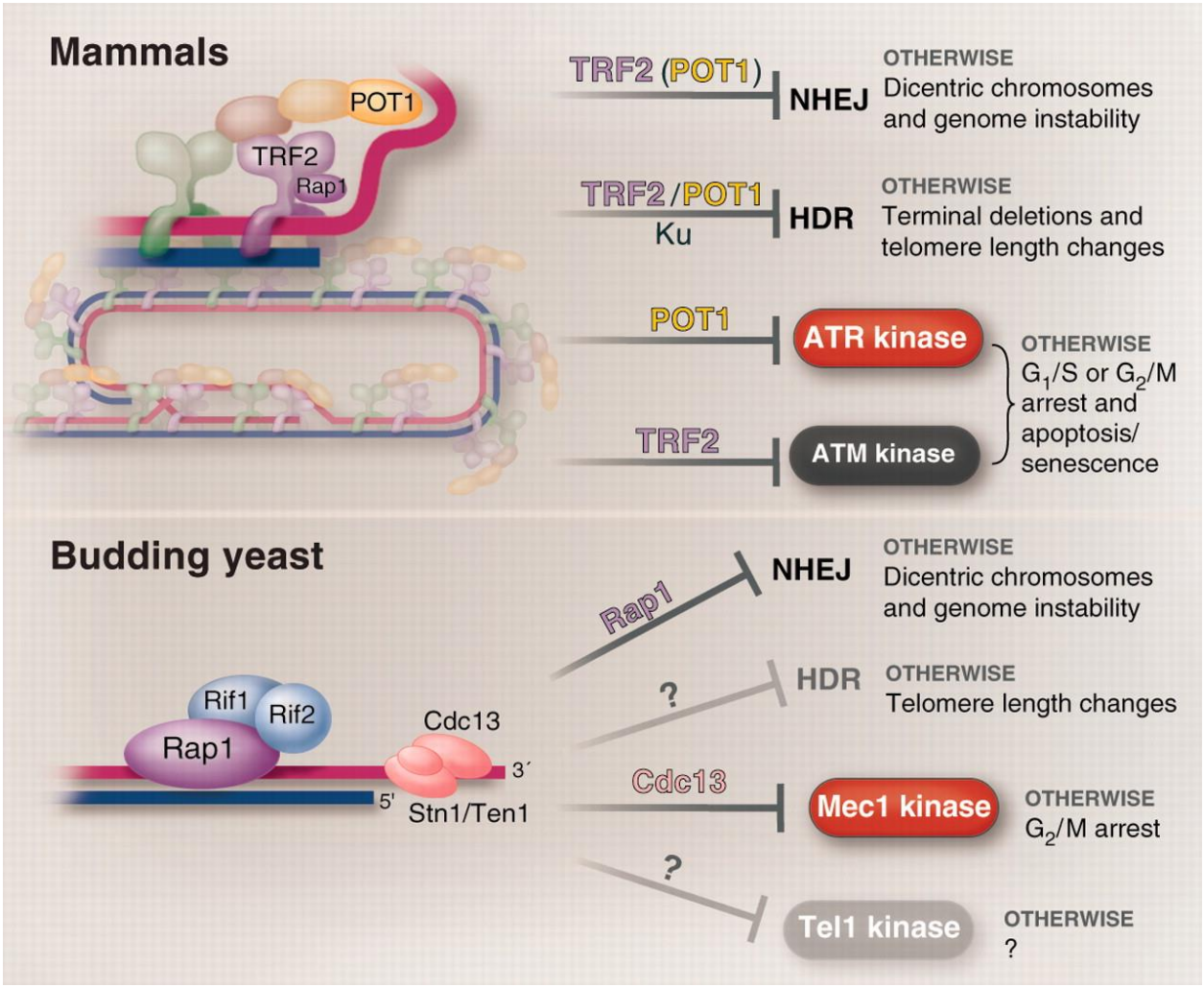
Science
AAAS

Fig. 3 Different components of shelterin are dedicated to different aspects of the end-protection problem



T. de Lange Science 326, 948-952 (2009)

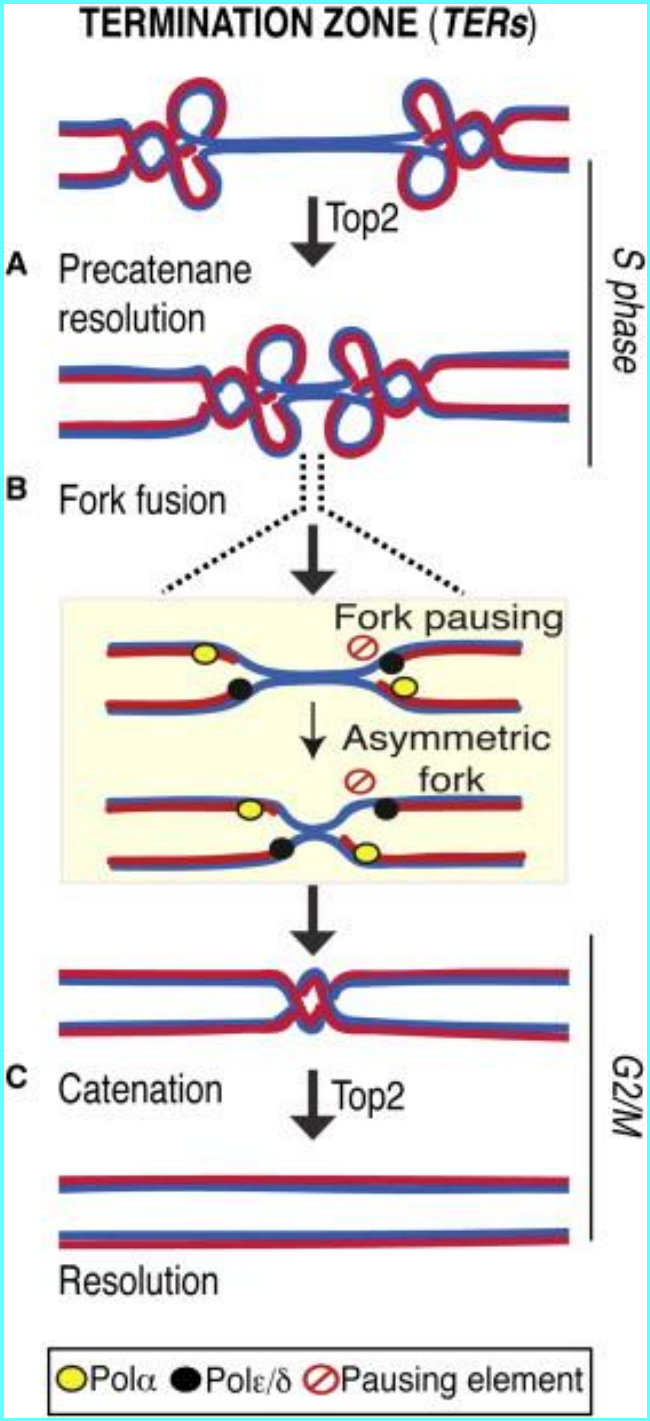
Fig. 4 Different solutions to the end-protection problem



T. de Lange Science 326, 948-952 (2009)

Science AAAS

Termination of eukaryotic DNA replication

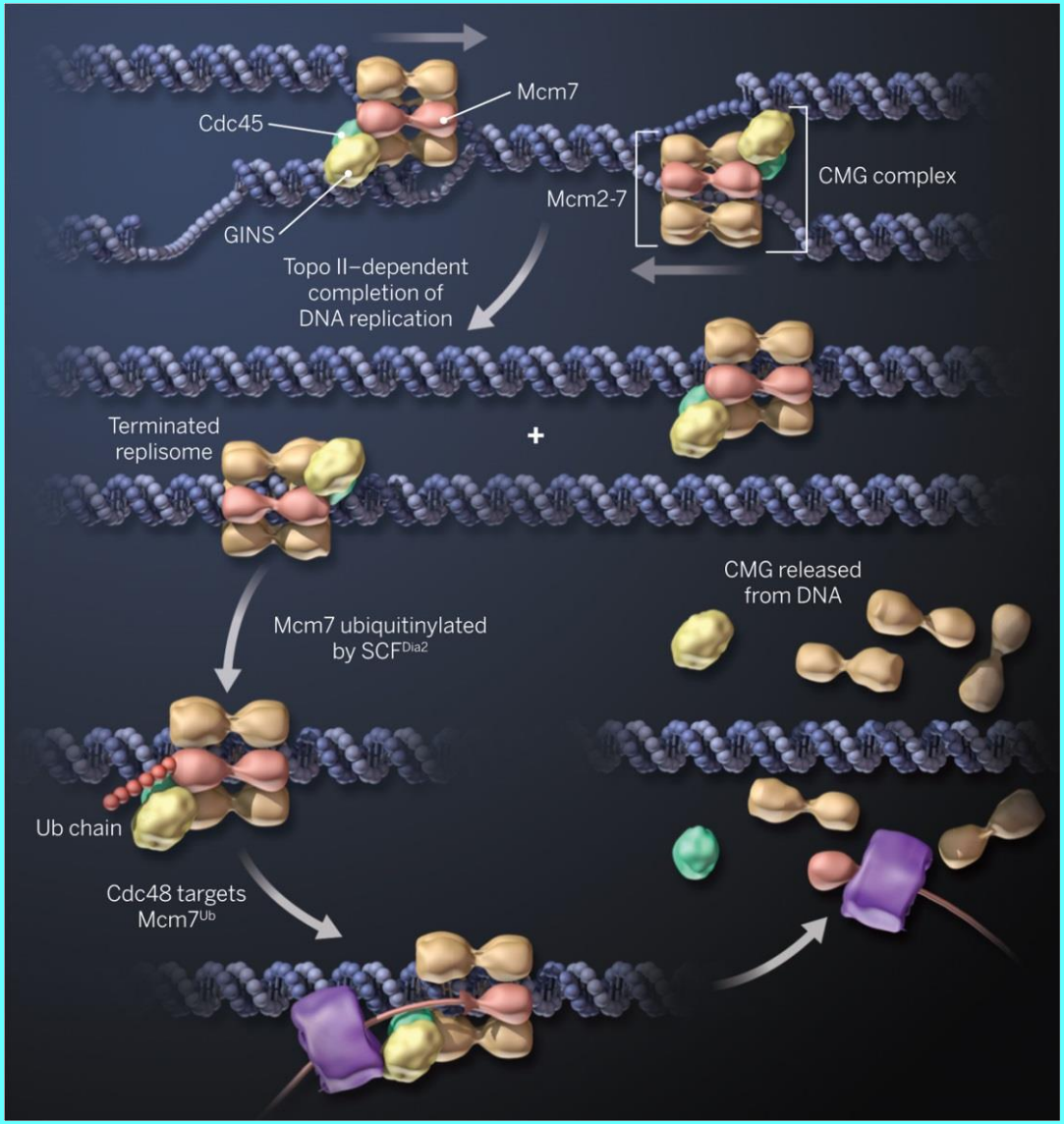


(A) Precatenane resolution: Top2 mediates fork progression at the *TER* zone by resolving precatenanes behind the forks.
(B) Fork fusion: the right fork stalls at a pausing site (pausing element, red symbol) and emerges with an asymmetric conformation. The leading polymerase (black oval) and the lagging apparatus (yellow oval) are shown.
(C) Catenation: Top2 then resolves the last catenation at *TERs* before DNA segregation, allowing chromosome resolution.

Other factors involve in termination of replication were described in 2014 – topoisomerase II participates on it and the proces is regulated by ubiquitination.

DOI: http://dx.doi.org/10.1016/j.molcel.2010.07.024

Disassembly required. Once two converging replisomes complete DNA replication, they are targeted sequentially by SCFDia2 and Cdc48, resulting in CMG disassembly.



S P Bell Science 2014;346:418-419