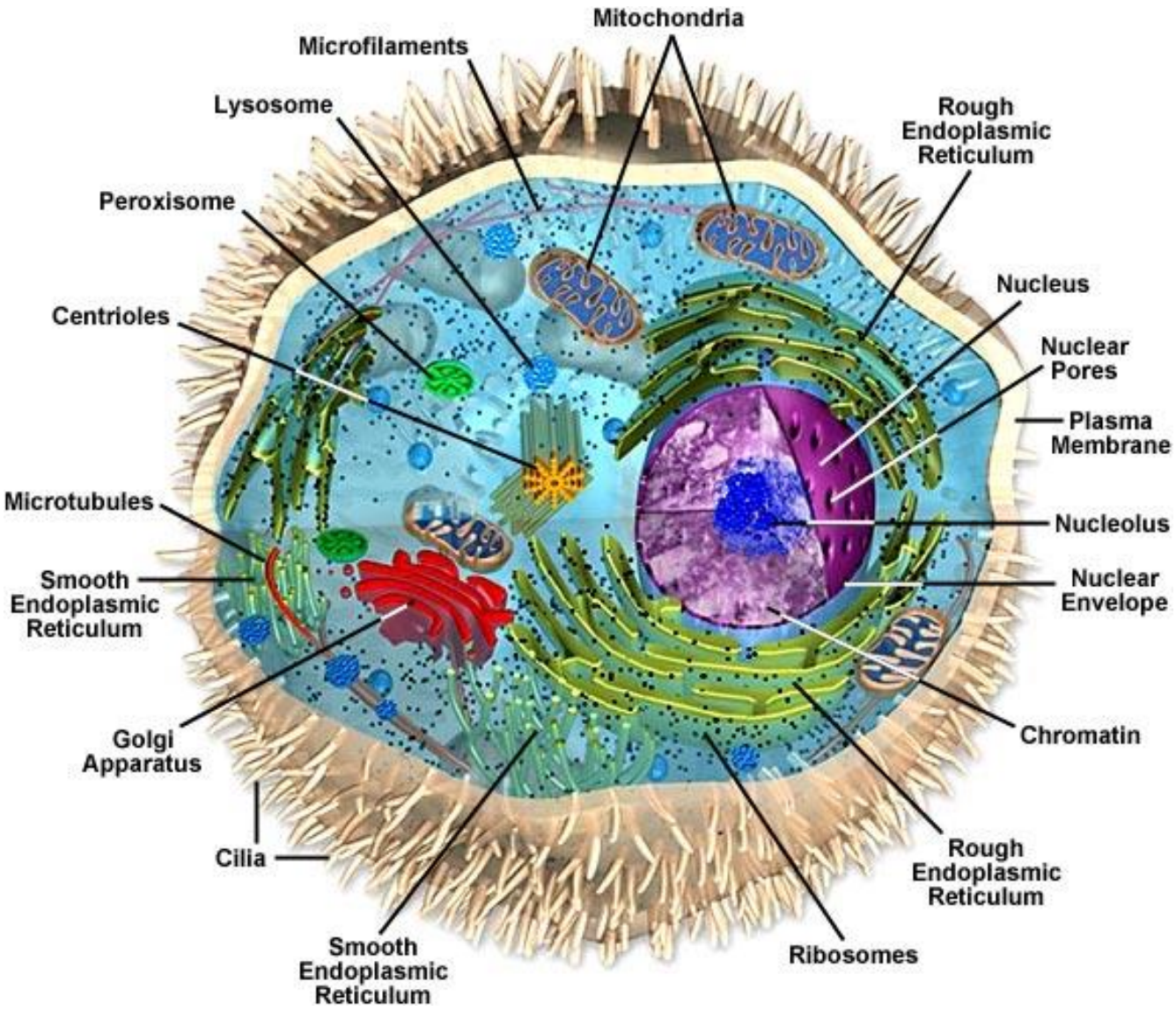
Structure of eukaryotic genome, replication and gene expression in eukaryotes

Molecular Biology

1. Structure of eukaryotic genome

- Eukaryotic cells:
  - cell wall consists of cellulose (plants) or chitin (fungi), animal cells have no cell wall
  - they contain organelles (mitochondria, plastids)
  - nucleus is divided by mitosis
  - nucleus consists of chromatin:
    - dsDNA
    - histones
    - non-histone proteins
  - chromosomes contain linear dsDNA

- Eukaryotic cells:
  - reproduction: asexual (unicellular organisms), sexual
  - replication, transcription and translation are more complicated processes than in prokaryotes
  - genes usually contain introns



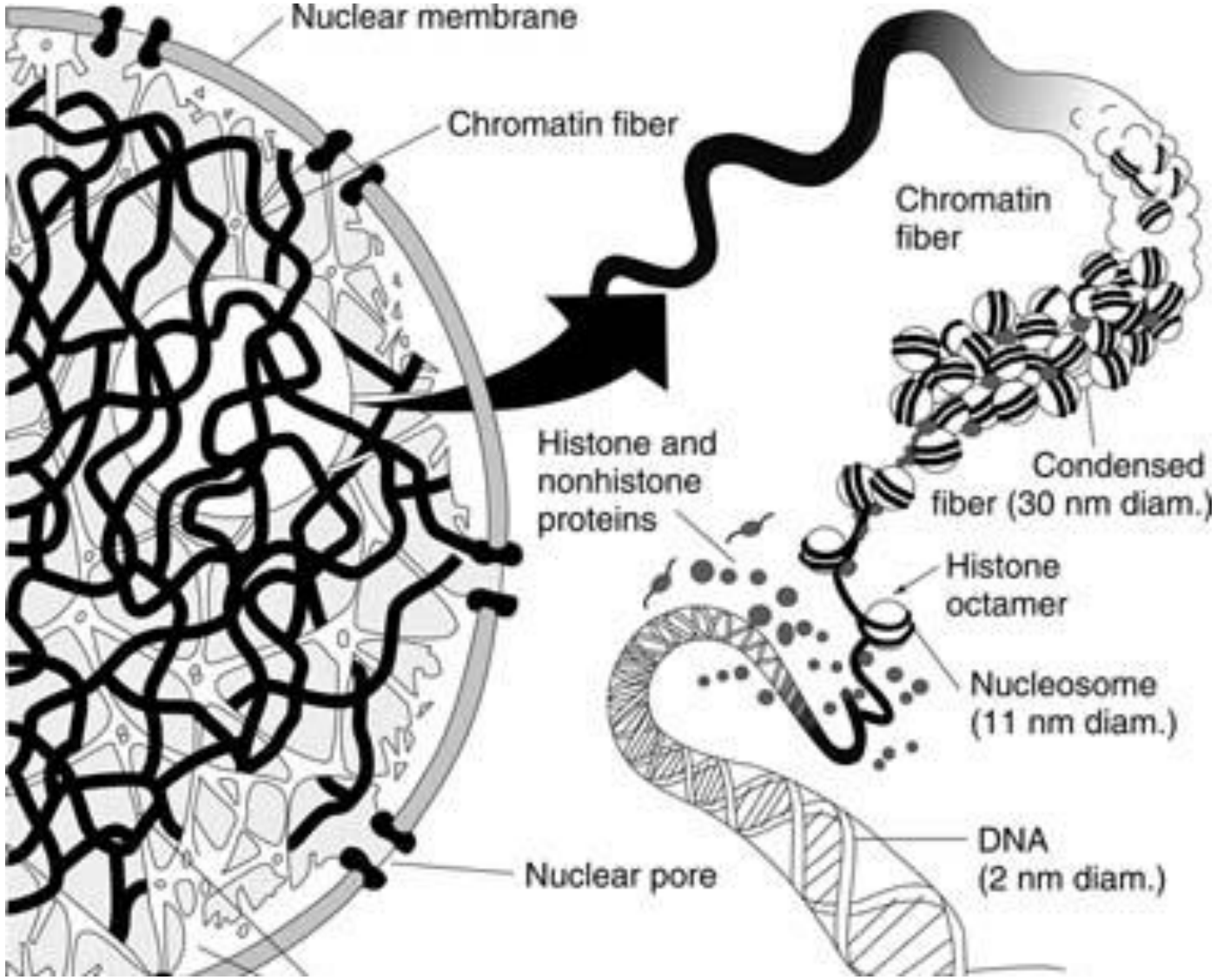
http://pulpbits.net/7-eukaryotic-cell-structure/structures-of-eukaryotic-cells/

- Genome of eukaryotic cells:
  - animal cells: nucleus and mitochondria
  - plant cells: nucleus, mitochondria a plastids

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

➢ Chromatin

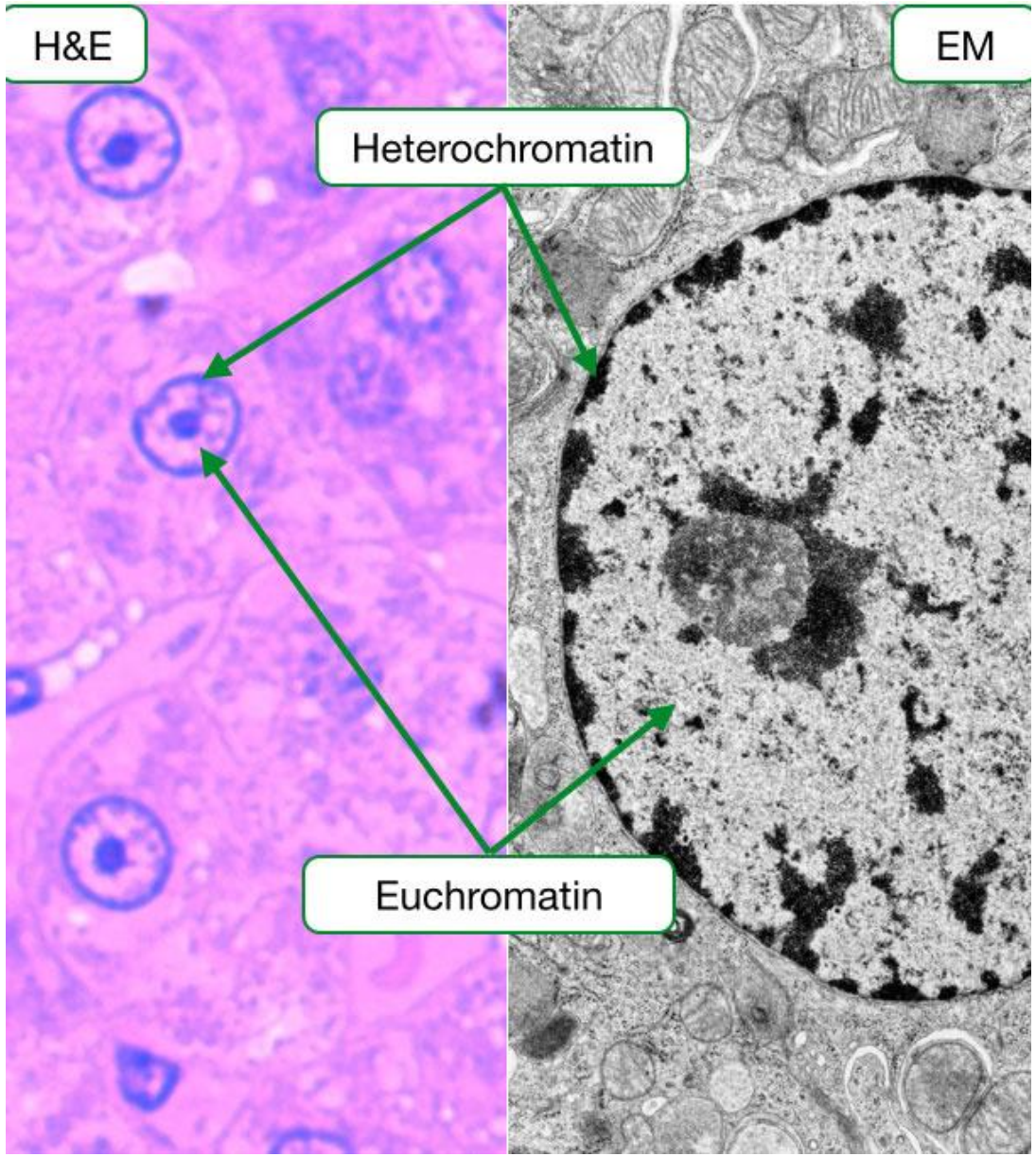
- material which forms the nucleus
- consists of DNA and two types of proteins: histones and non-histone proteins
- level of chromatin condensation is dependent on cell cycle phase
- depending on the level of condensation and the ability to be stained by basic dyes, we distinguish two types of chromatin: **euchromatin** (weakly stainable, decondensed, transcriptionally active) and **heterochromatin** (strongly stainable, condensed, transcriptionally inactive)



https://www.cliffsnotes.com/study-guides/biology/biochemistry-ii/eukaryotic-genes/structure-of-chromatin

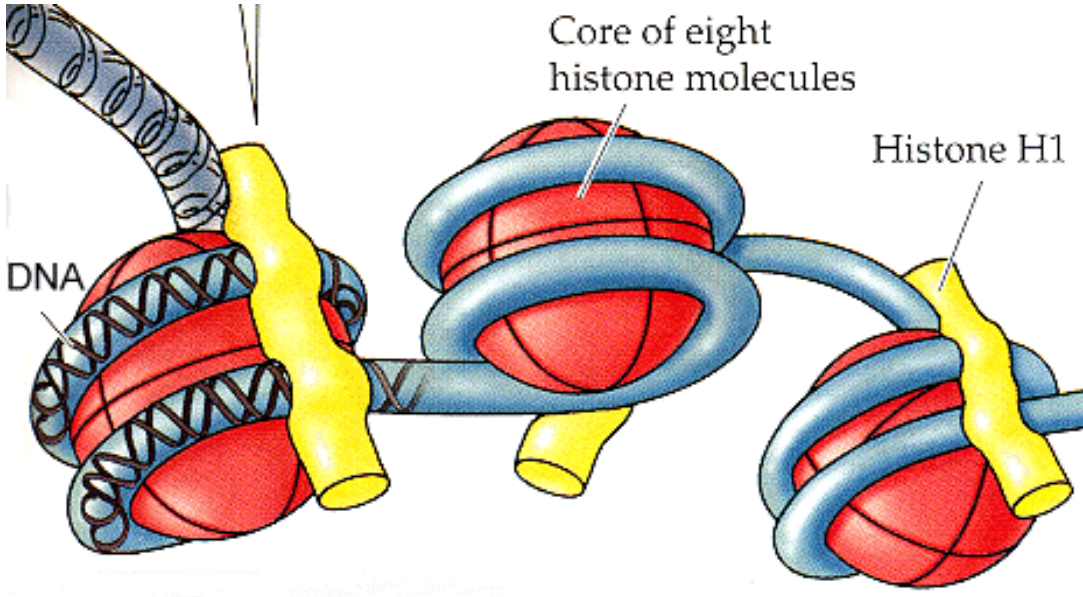
Heterochromatin:

- Constitutive
  - constantly in heterochromatin state
  - centromeres, telomeres
  - one X chromosome in women
- Facultative
  - switches between heterochromatin and euchromatin states during ontogenetic development



http://histology.med.yale.edu/histological_features_of_cells/histological_features_of_cells_reading.php

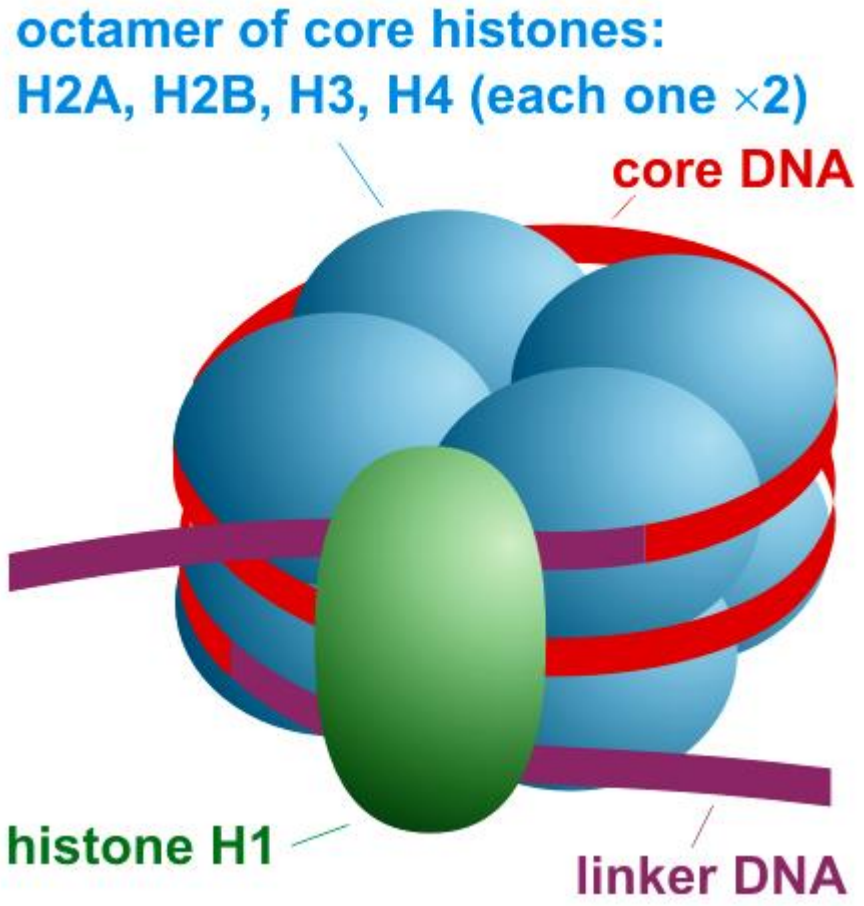
- Chromatin components:
  - **Histones:**
    - small basic proteins (positively charged) which bind to negatively charged DNA
    - 5 types: H1, H2A, H2B, H3, H4
    - high content of arginine and histidine
  - **Non-histone proteins:**
    - RNA polymerases and other enzymes of the transcriptional apparatus
    - HMG1 and HMG2 (*high mobility group proteins*)
    - HMG3 and HMG4 bind to the histone core especially in transcriptionally active regions



http://biology.kenyon.edu/courses/biol114/Chap01/chrom_struct.html

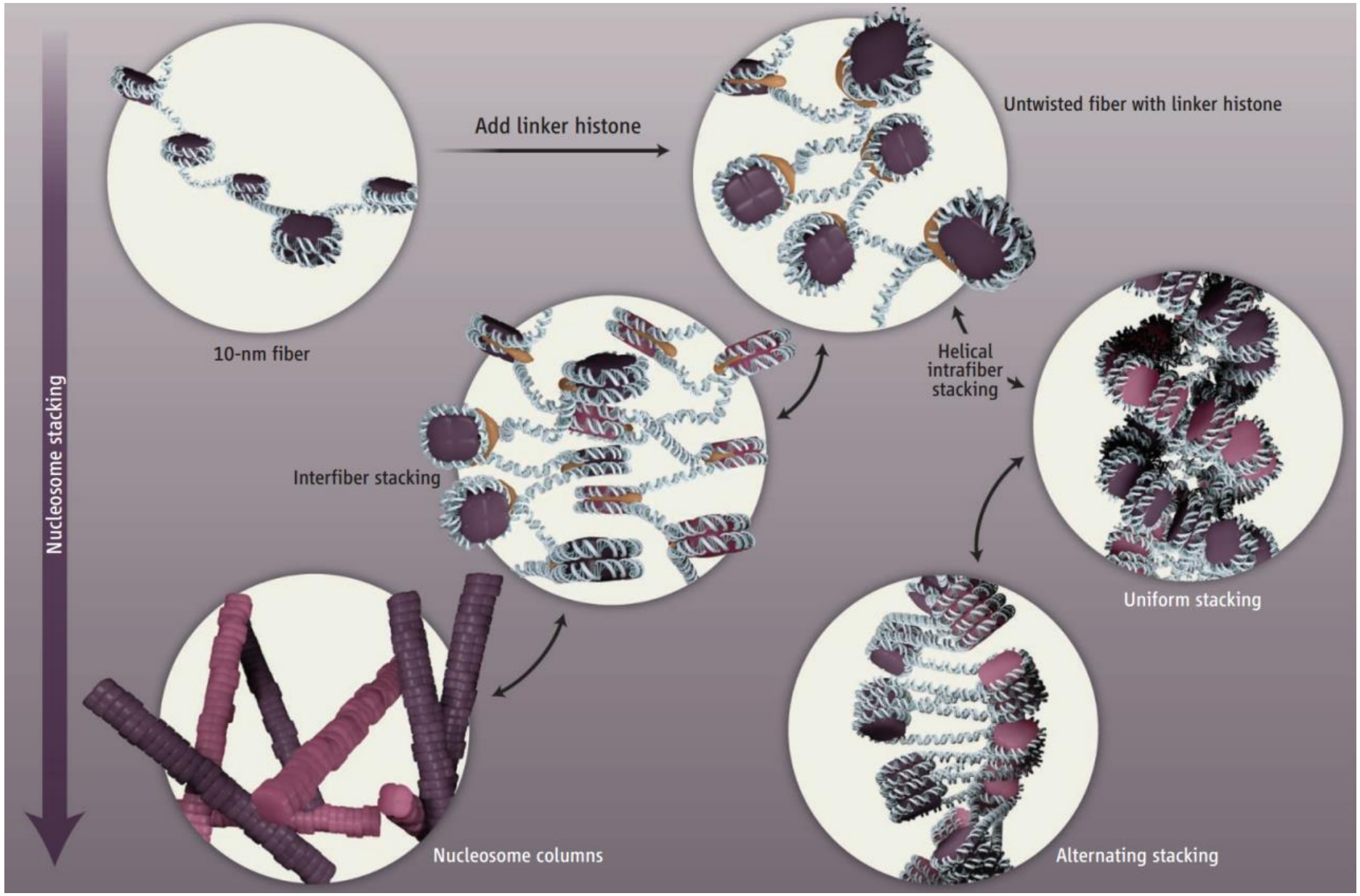
Nucleosome:

- basic unit of chromatin
- histone octamer (H2A, H2B, H3,H4)$_2$
- 1 molecule of histone H1
- DNA segment of 200 bp, which is wound around the histone octamer twice
- nucleosome fibre (10 nm) is visible by EM



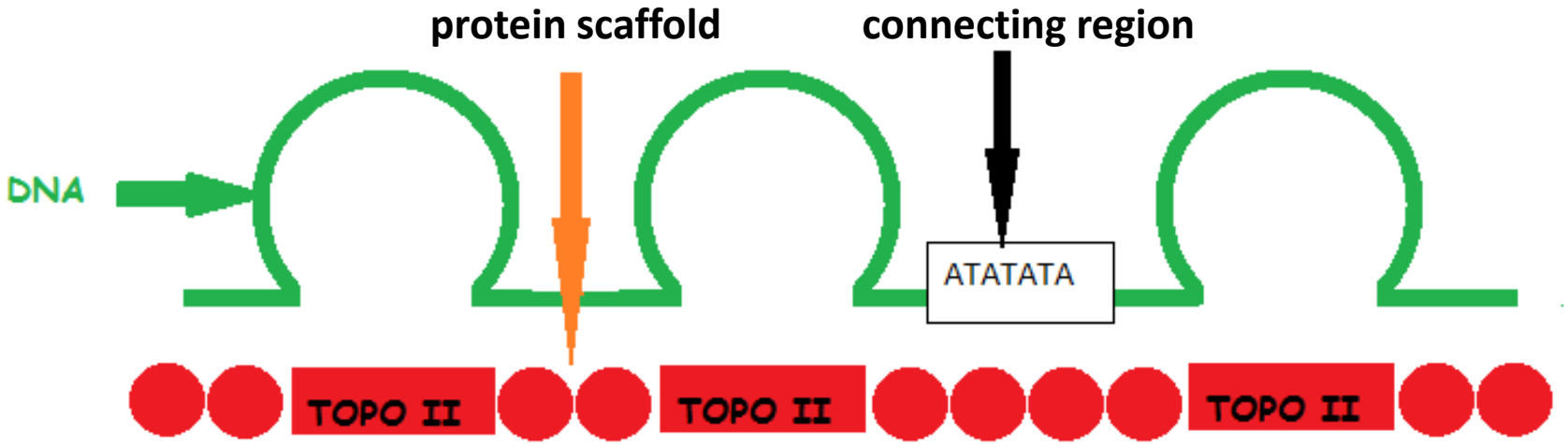
Stryer, Lubert (1995). *Biochemistry* (fourth ed.). New York - Basingstoke: W. H. Freeman and Company. ISBN 978-0716720096

Discovery of nucleosome stacking in 2014



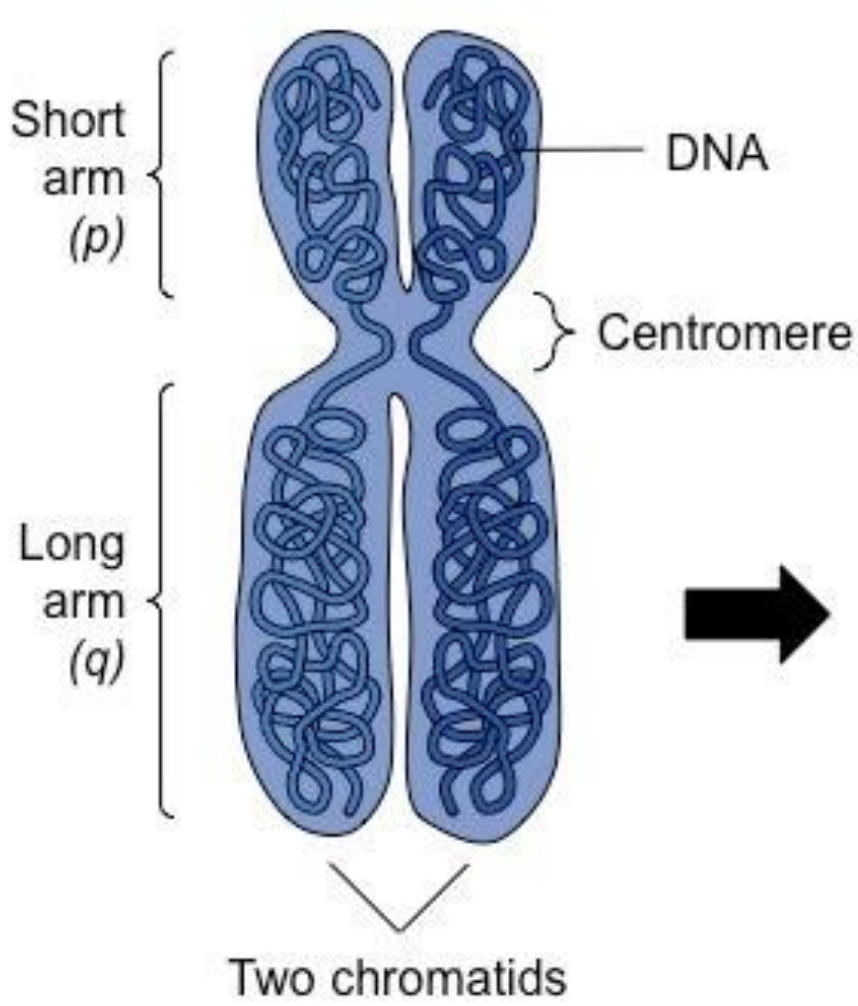
Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science. 2014 Apr 25;344(6182):376-80. doi: 10.1126/science.1251413.

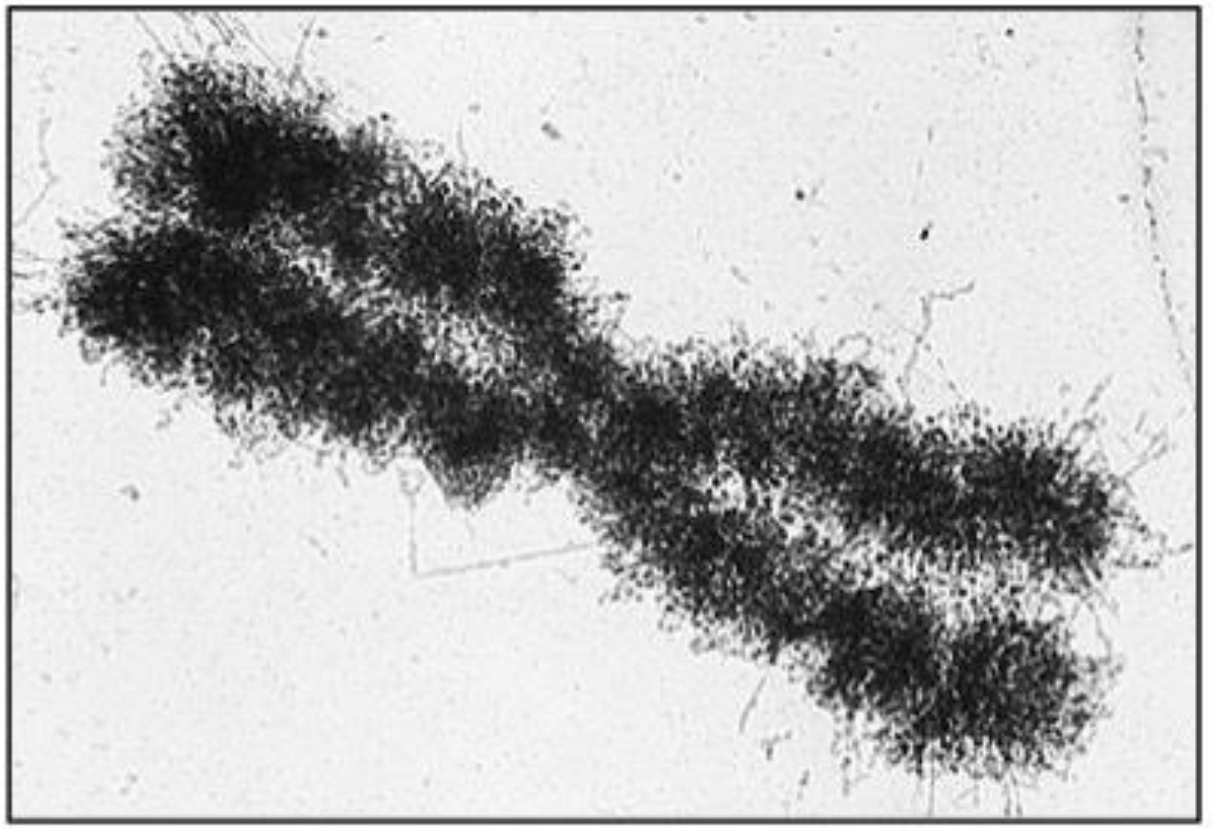
- Chromatin domains:
  - loops of 30nm chromatin fibre attached to protein scaffold (60-150 kbp)
  - one molecule of topoisomerase II in the base of each loop – change of topology during replication and transcription
  - each domain has one *ori* locus
  - one human chromosome contains approximately 2000 domains



- Mitotic chromosomes:
  - originate by condensation of 30 nm chromatin fibres
  - formed during mitosis or meiosis
  - condensation of 30 nm chromatin fibres into 600 - 700 nm fibres, which form the chromosome structure
  - in chromosomes, chromatin is in highly condensed state and is transcriptionally inactive

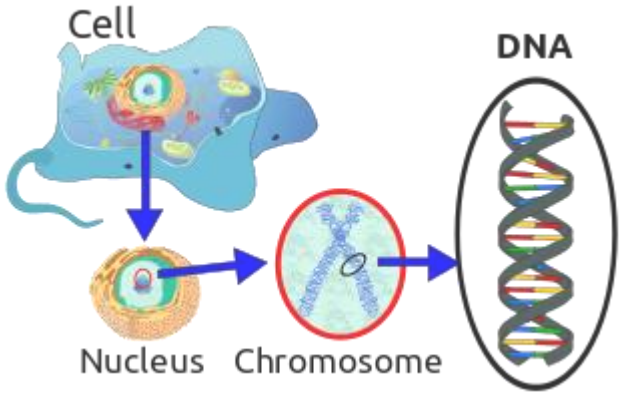
➢Mitotic chromosomes:





Chromosome as seen via an electron micrograph

http://ib.bioninja.com.au/standard-level/topic-1-cell-biology/16-cell-division/dna-supercoiling.html



- Chromosomal (nuclear) DNA:
  - 1 linear molecule of dsDNA
  - the number of bp per one chromosome in haploid cell is **$1{,}34 \times 10^7$ to $1{,}5 \times 10^{10}$**
  - only 1.5 % of mammal genome contains protein-coding genetic information
  - most structural genes are about 1×10^4 to 2×10^6 bp long – considerable part consists of regulatory elements

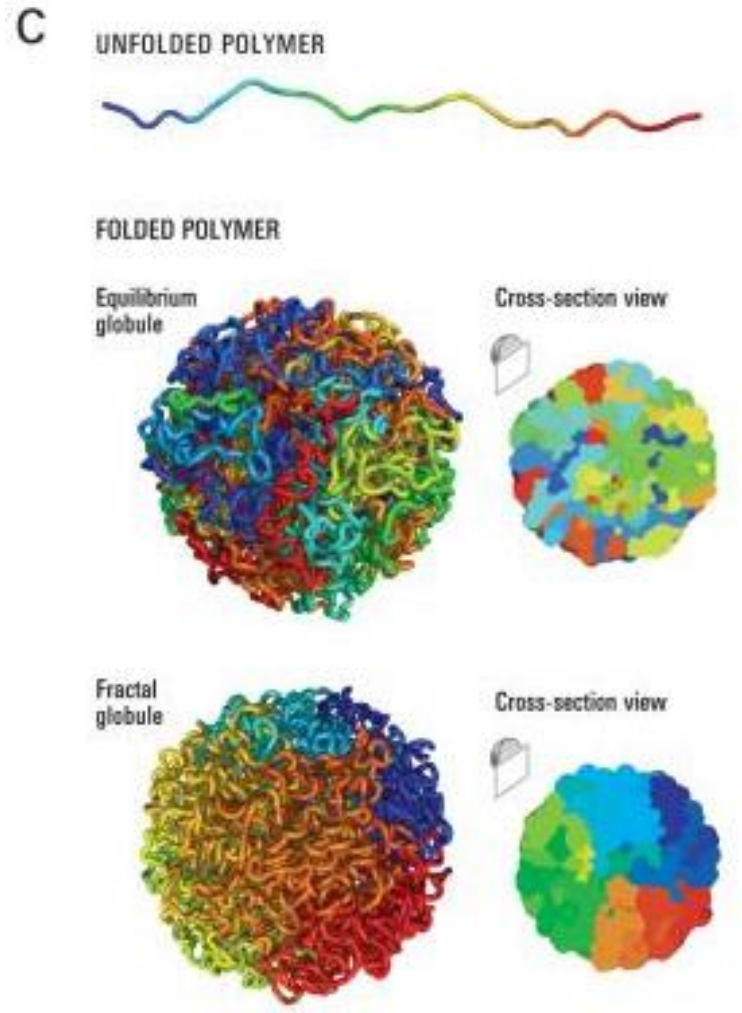
- Repeats in nuclear DNA:

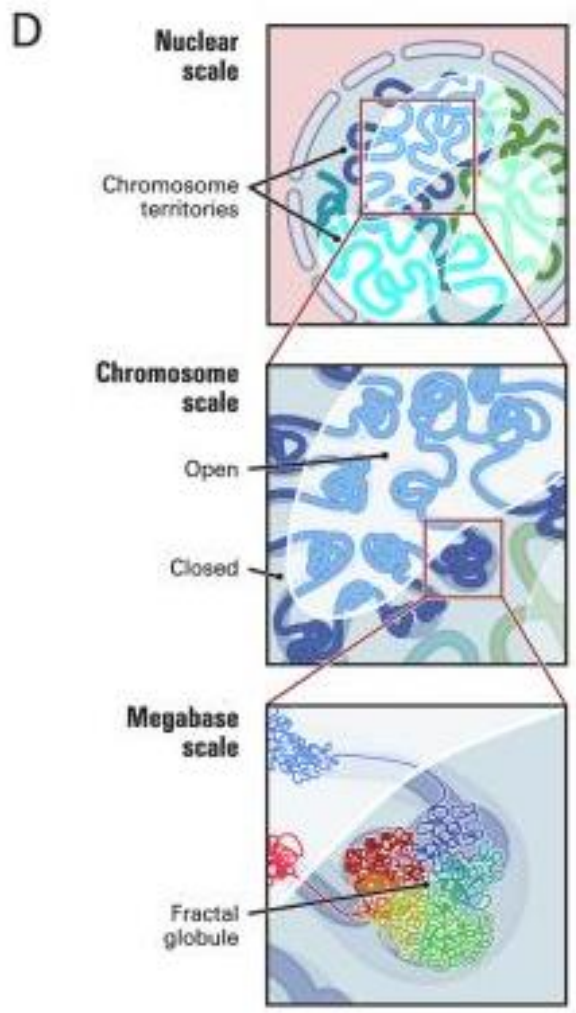
  - short tandem repeats- not in prokaryotes
  - dispersed repeats- not in prokaryotes

  - 25-50% of structural genes are **unique sequences**
  - the rest occurs as **gene repeats**

- Gene repeats:
  - **Gene family:** group of genes with similar sequence, from the same original gene, generally with similar functions (genes for hemoglobin subunits) or pseudogenes (dysfunctional genes)
  - **Tandem gene repeats:** directly adjacent, separated by spacers (intergenic sequence), genes for 5S-rRNA, genes for tRNA, genes for histones
  - **Dispersed gene repeats:** copies are dispersed in different locations in the genome (genes for tRNA, snRNA, etc.)

➢Chromatin organisation in the nucleus - fractal globules

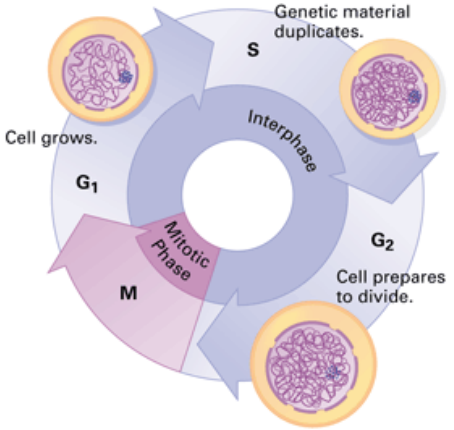




E. Lieberman-Aiden et al., Science 326, 289-293 (2009)

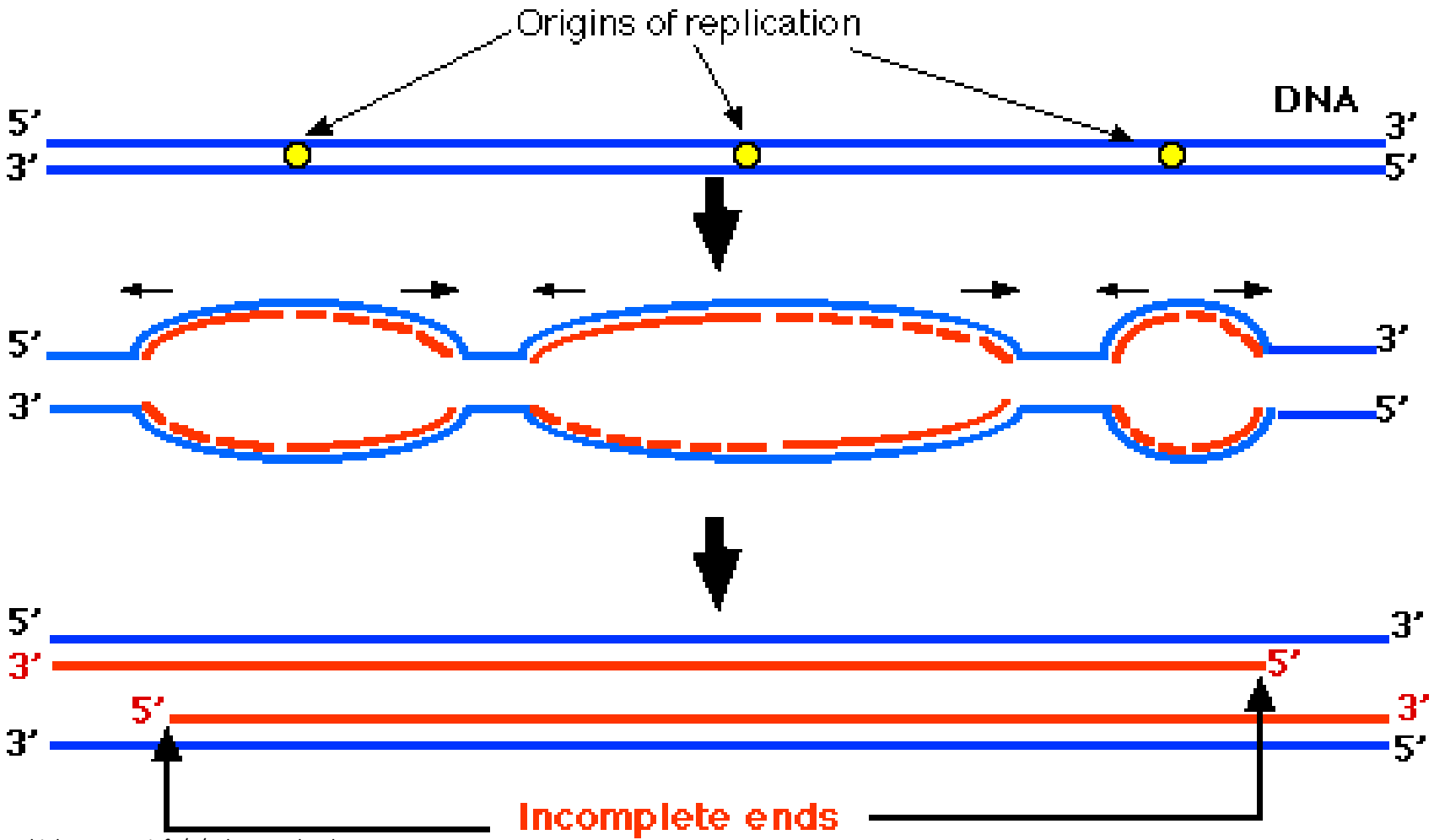
2. DNA replication in eukaryotes

- replication of mitochondrial and chloroplast DNA
- replication of nuclear chromosomes:
  - semiconservative (each new dsDNA contains one parental and one newly synthesised strand) and semidiscontinuous (leading and lagging strands)
  - initiation, elongation, termination
  - only in S-phase of cell cycle



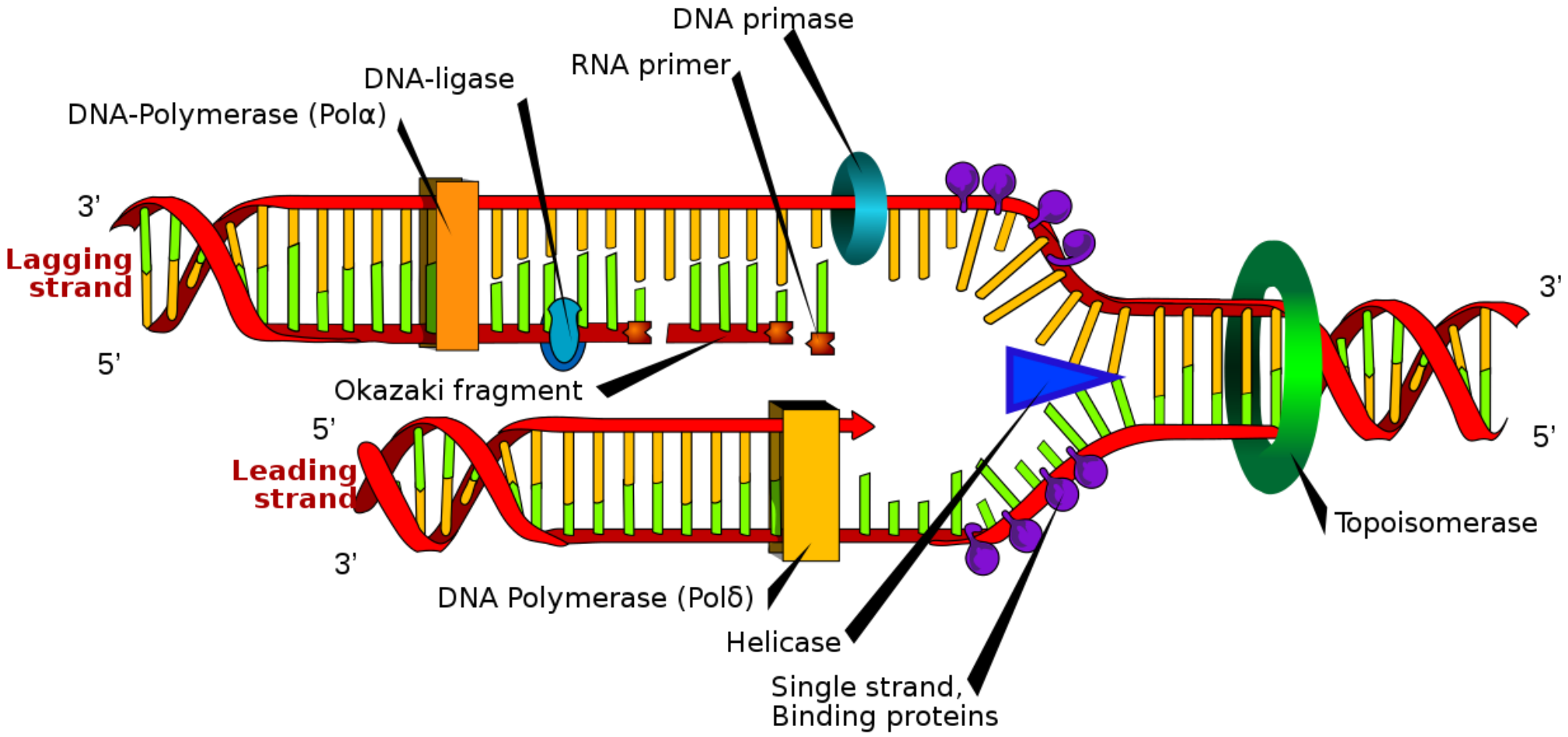
➢Replication of nuclear chromosomes:

- ➢proceeds in several places at once (in contrast to prokaryotes)
- ➢chomosome is a set of **replicons**, many *ori* sites (30-50k in mammals)
- ➢euchromatin is replicated earlier than heterochromatin



http://www.biology-pages.info/T/Telomeres.html

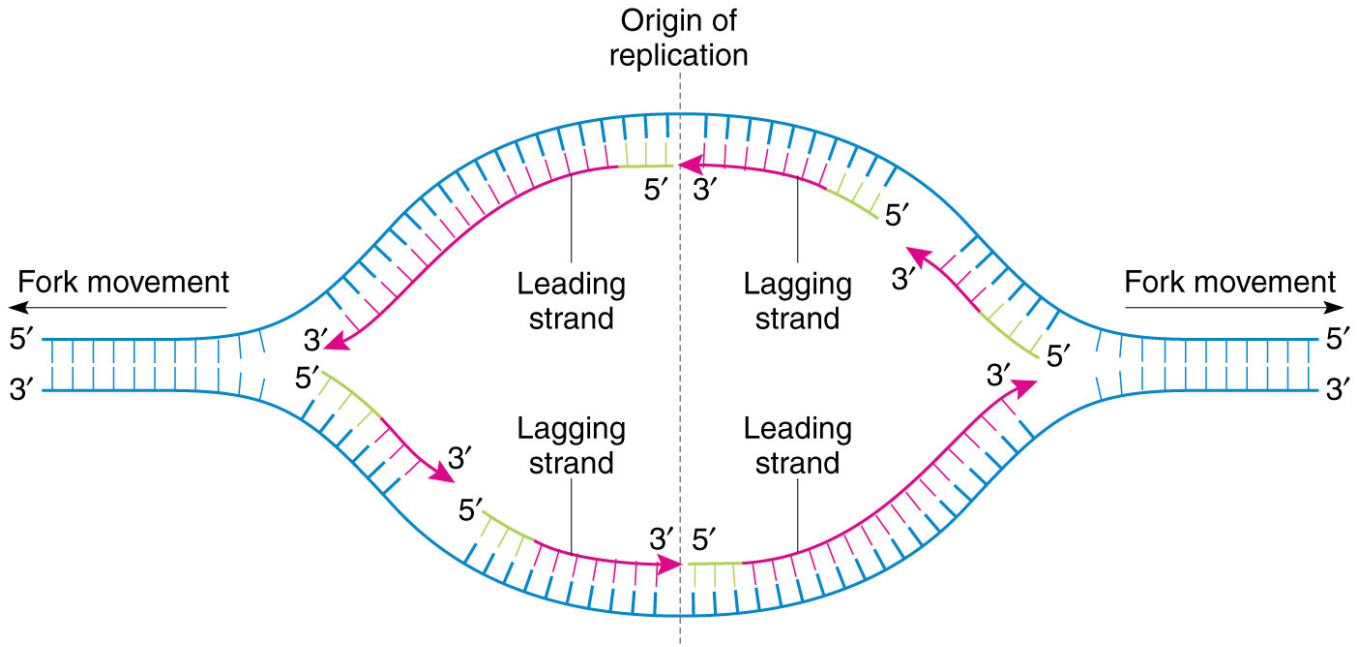
➢DNA replication process:



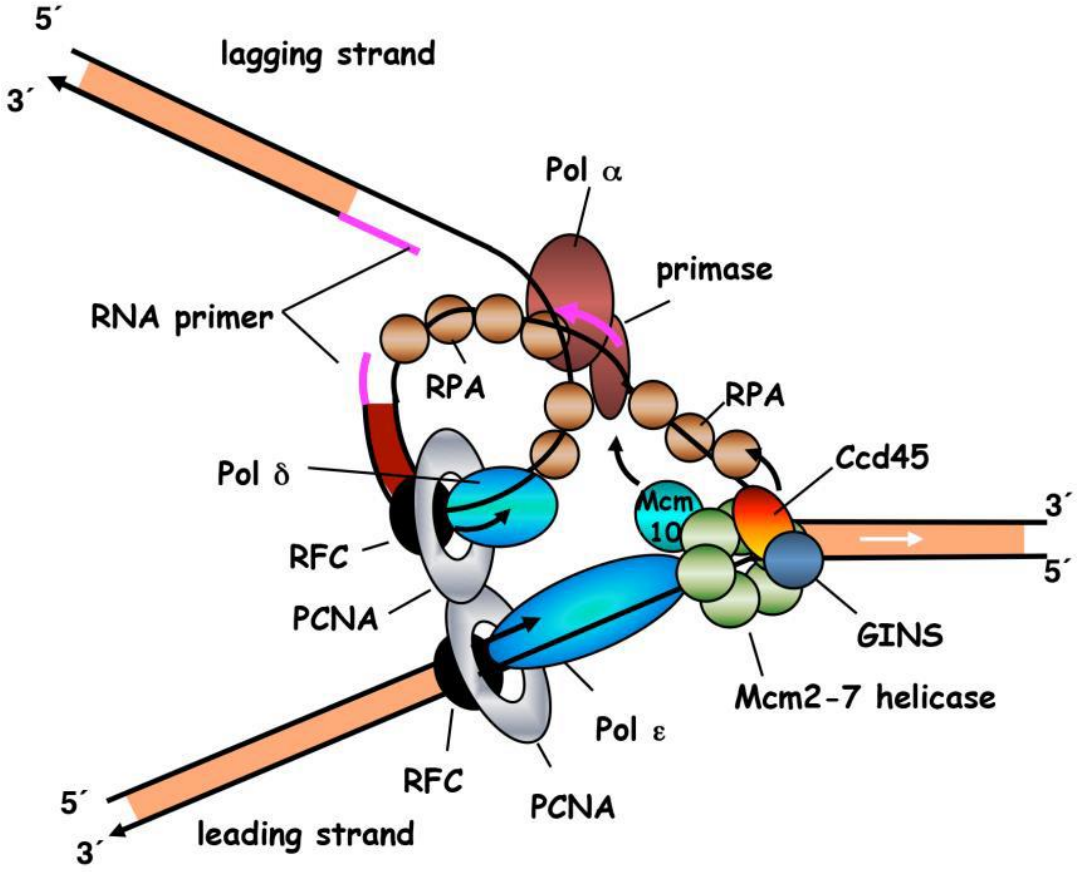
➢Eukaryotic DNA polymerases:

- ➢DNA polymerase α- in complex with primase synthesizes Okazaki fragments, does not possess 3‘-5‘ exonuclease activity (no *proofreading* activity)
- ➢DNA polymerase β- synthesizes short fragments during DNA repair
- ➢DNA polymerase γ- synthesis of mitochondrial DNA
- ➢DNA polymerase δ- synthesis of leading strand and completing the lagging strand
- ➢DNA polymerase ε- leading strand synthesis and DNA repair

➢Eukaryotic replication fork:

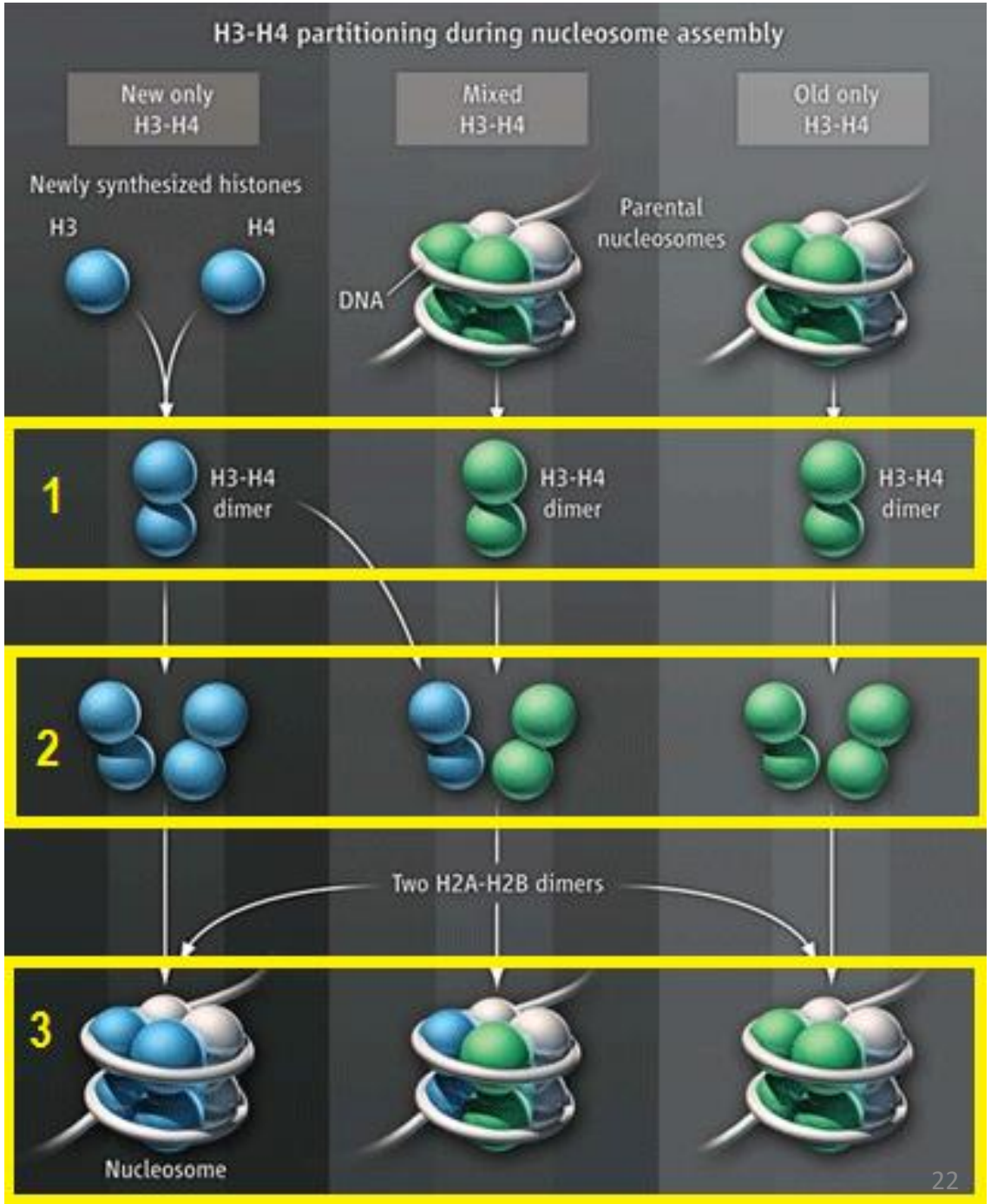


https://biology.stackexchange.com/questions/31585/does-dna-polymerase-always-go-the-same-direction



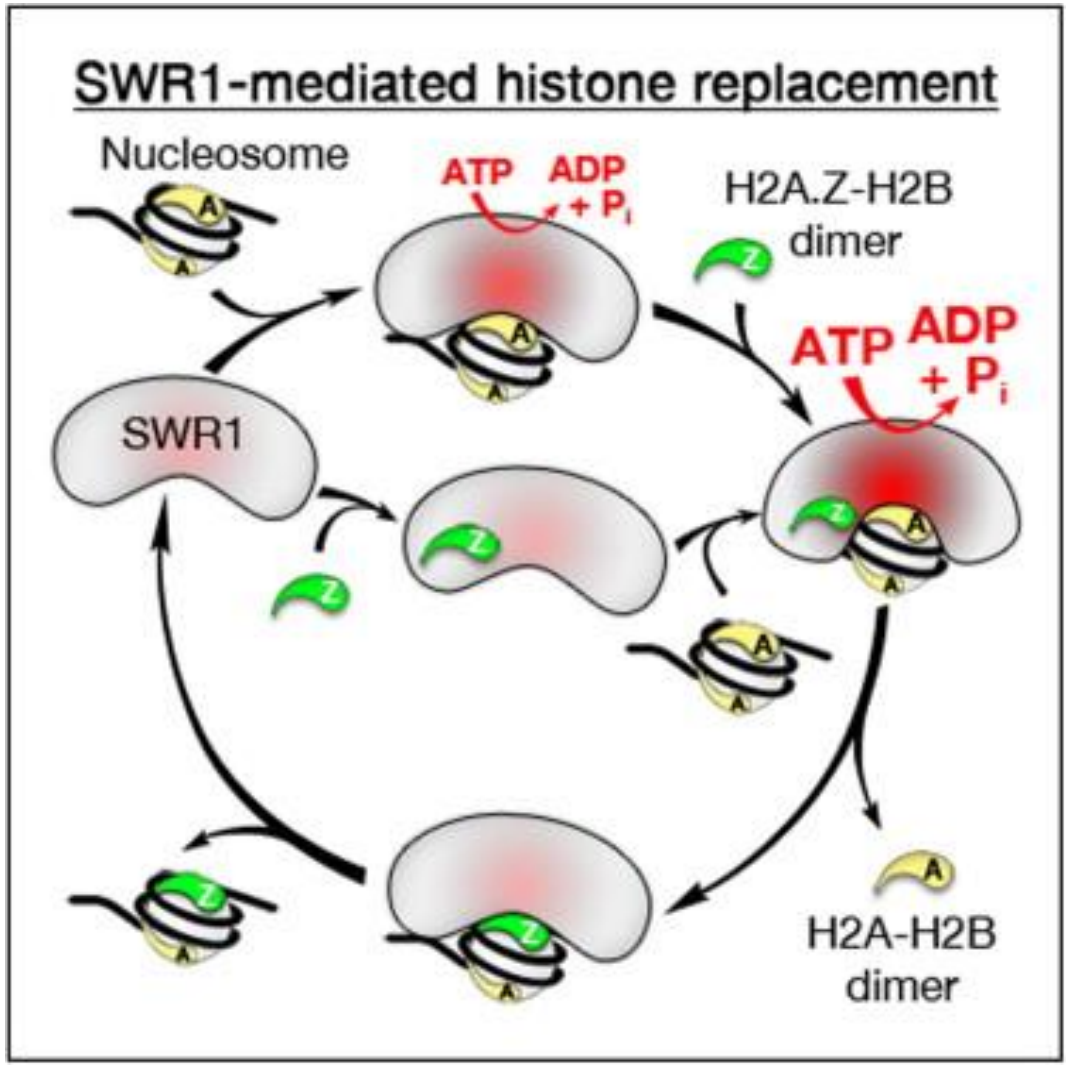
Histone partitioning:

- H3-H4 dimers are formed by parental or newly synthesized histone monomers
- newly synthesized dimers are associated together or mixed with parental dimers
- H2A-H2B dimers are added and nucleosome is formed



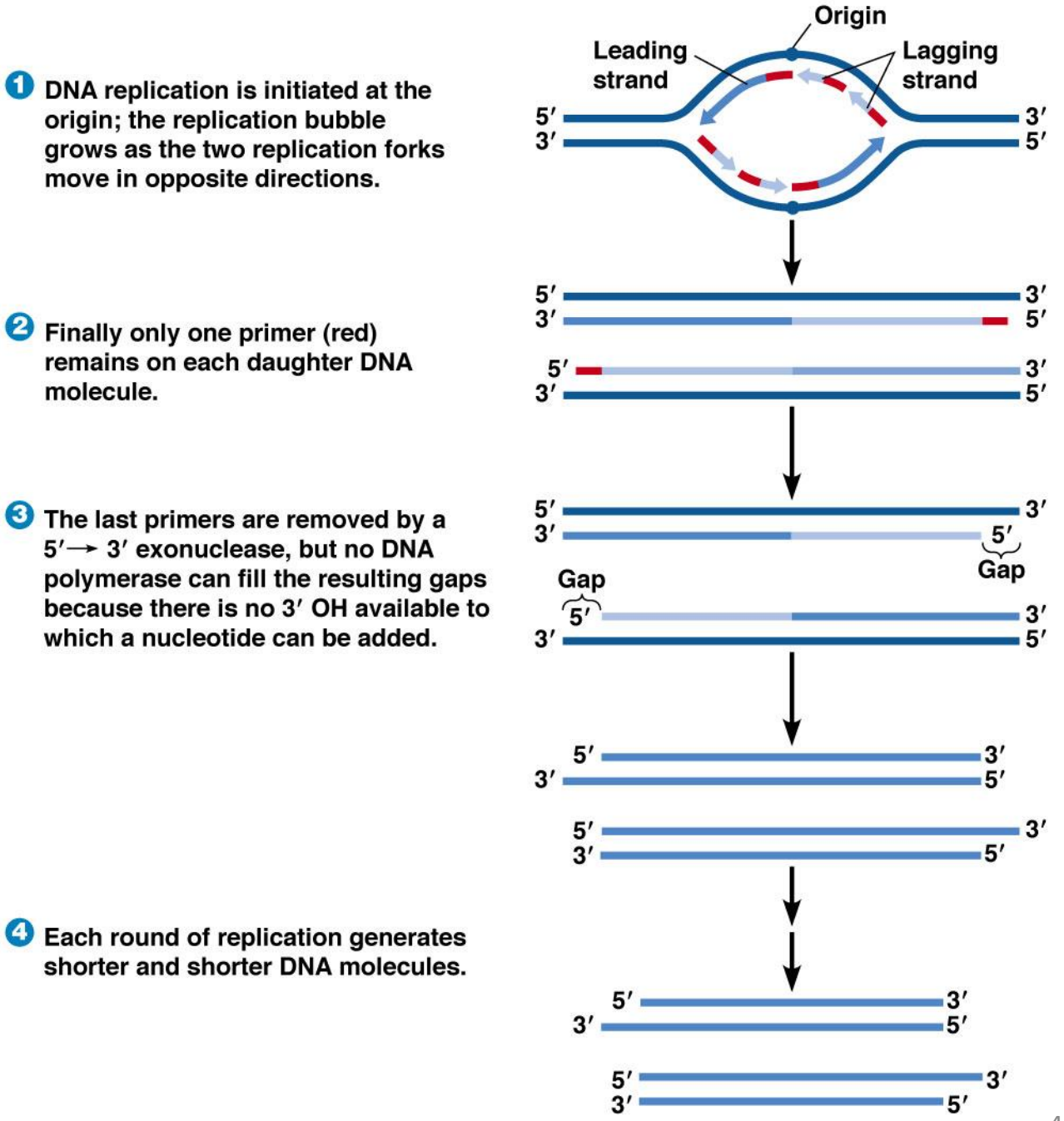
D. Ray-Gallet et al., Science 328,56-57 (2010)

- Nucleosome assembling is associated with hydrolysis of ATP



E. Luk et al., Cell 143,725-736, Nov 24,2010

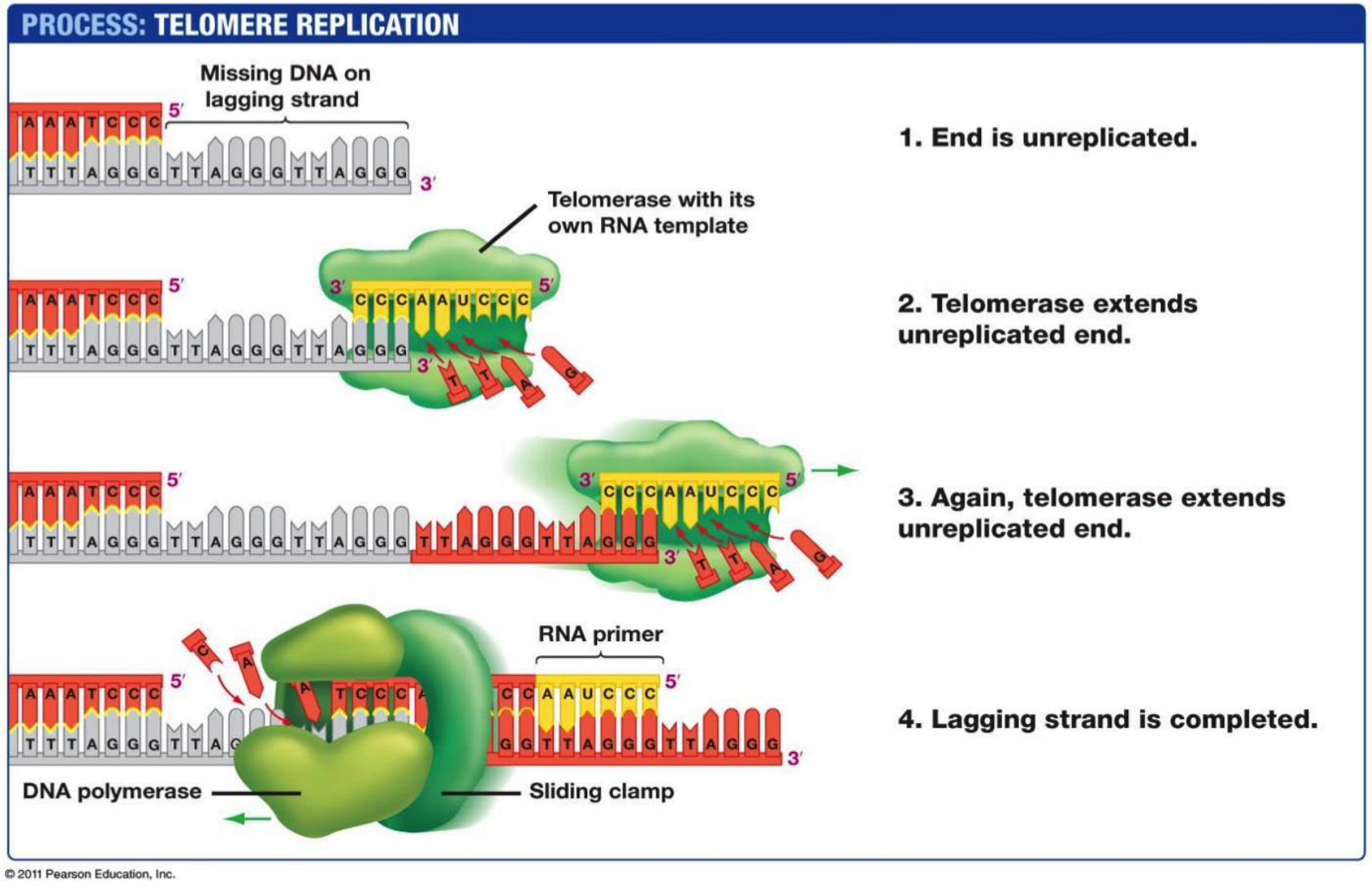
- Replication of linear molecules:
  - the end replication problem
  - telomerase = ribonucleoprotein - RNA is a template, protein has catalytical function



https://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-19/CB19.html

- Telomeres:
  - ends of eukaryotic chromosomes
  - repeats of TTAGGG/CCCTAA sequence (in vertebrates), several thousands of repeats
  - single-stranded overhang of 50 – 200 nucleotides
  - protection of chromosomes against degradation (by exonucleases) or fusions
  - telomeres associate with nuclear membrane

- Filling of missing 3‘-ends:
  - telomerase elongates 3‘-ends
  - formation of hairpin and RNA primer
  - replication of complementary strand and removing of hairpin



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

➢Sequences of telomeres:

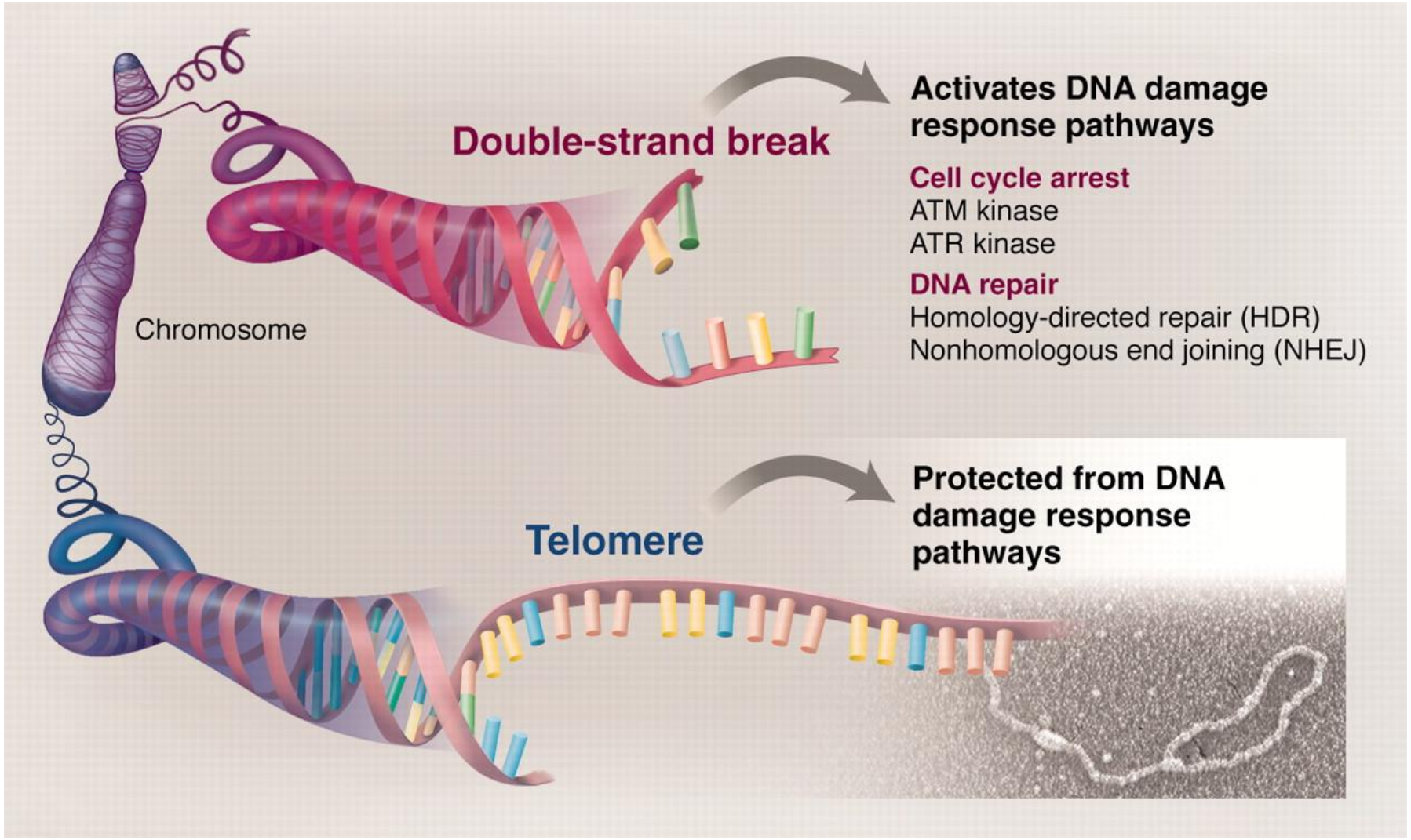
TABLE 11.5

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

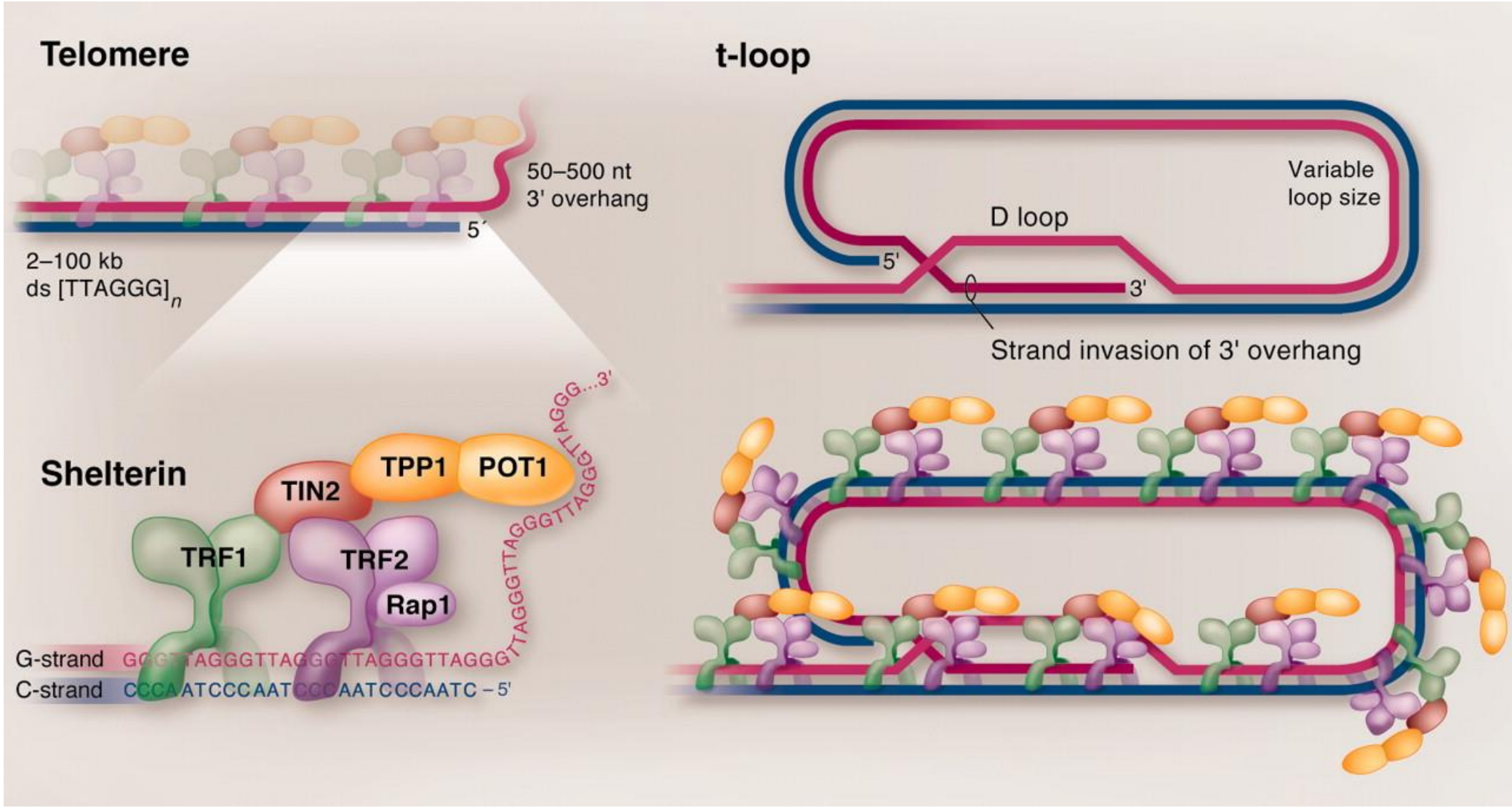
http://reasonandscience.heavenforum.org/t2263-the-telomerase-enzyme

➢Protection of chromosome ends:



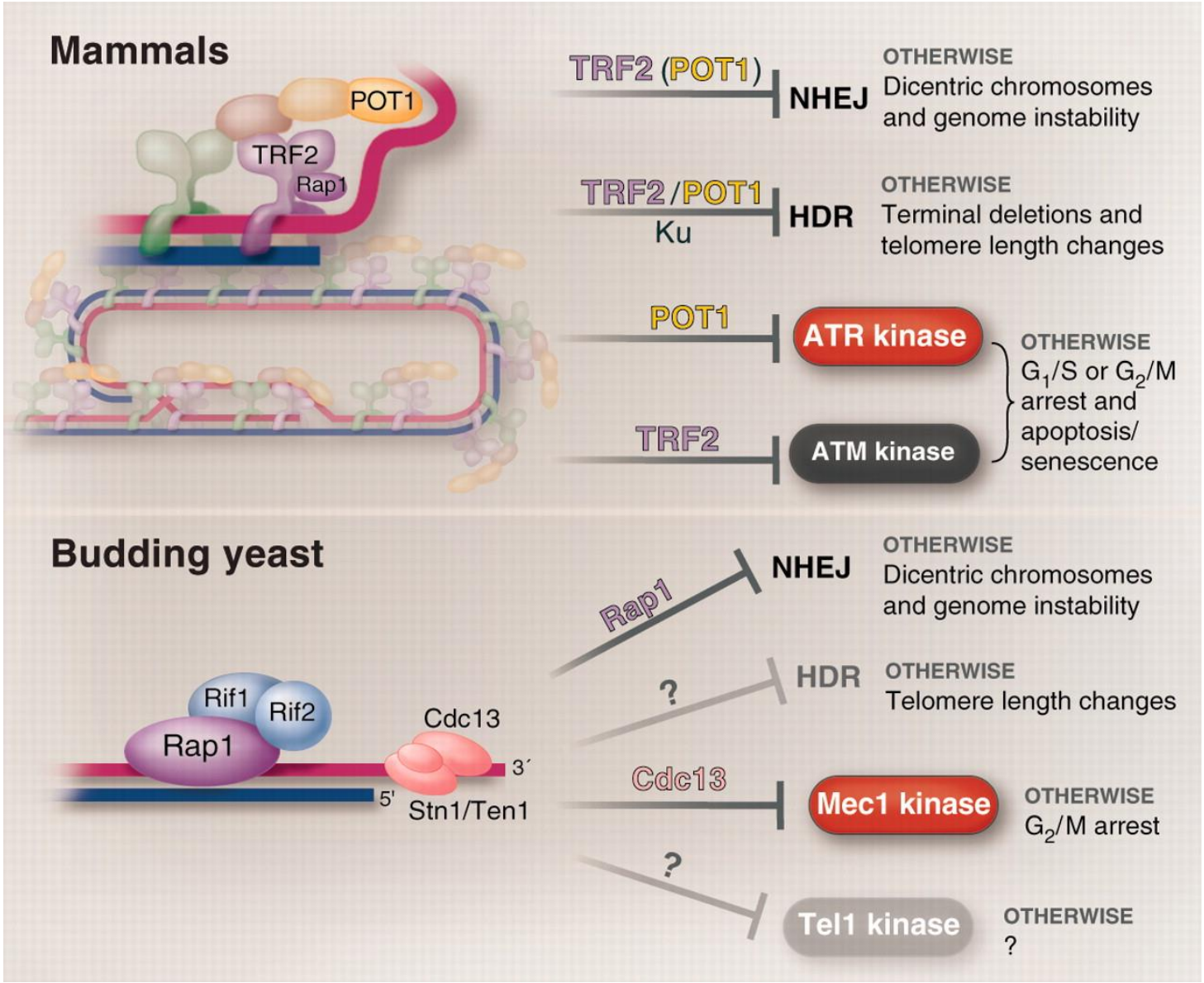
T. De Lange, Science 326, 948-952 (2009)

➢Mammalian telomeres:



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

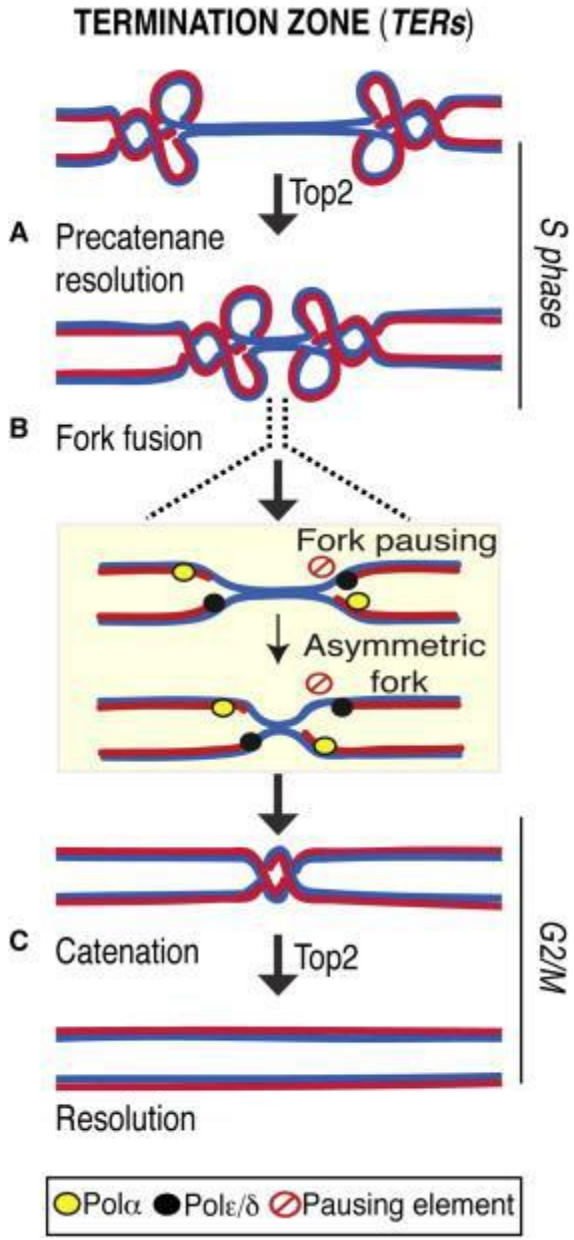
➢Different outcomes of end-protection problem:



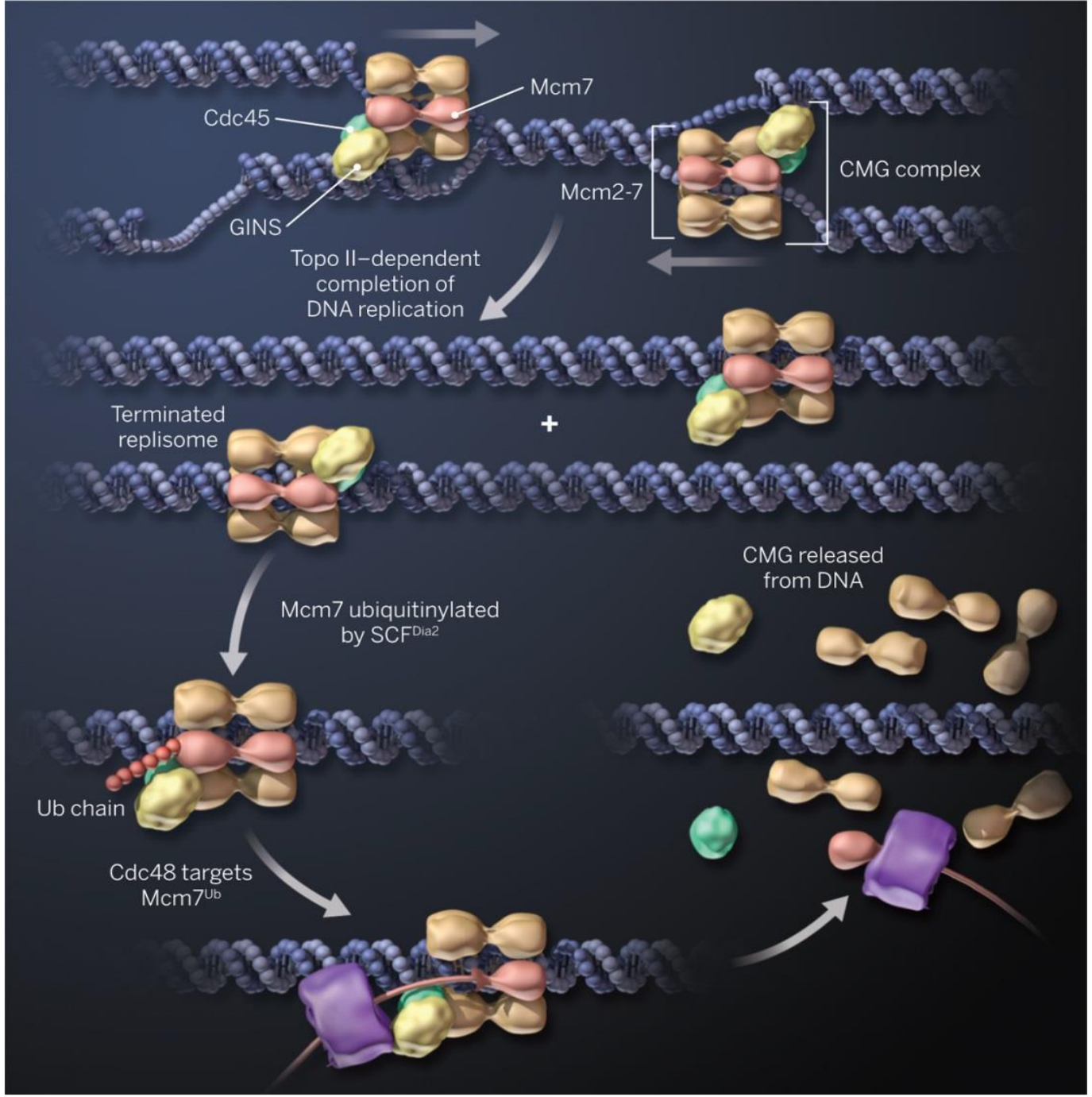
T. De Lange, Science 326, 948-952 (2009)

➢Termination of DNA replication in eukaryotes:

- ➢ replication of prokaryotic chromosome ends in specific sequences – termination zones
- ➢other factors involved in termination of DNA replication in eukaryotes – topoisomerase II participates and the process is regulated by ubiquitination



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



S.P Bell, Science 2014, 346: 418-419

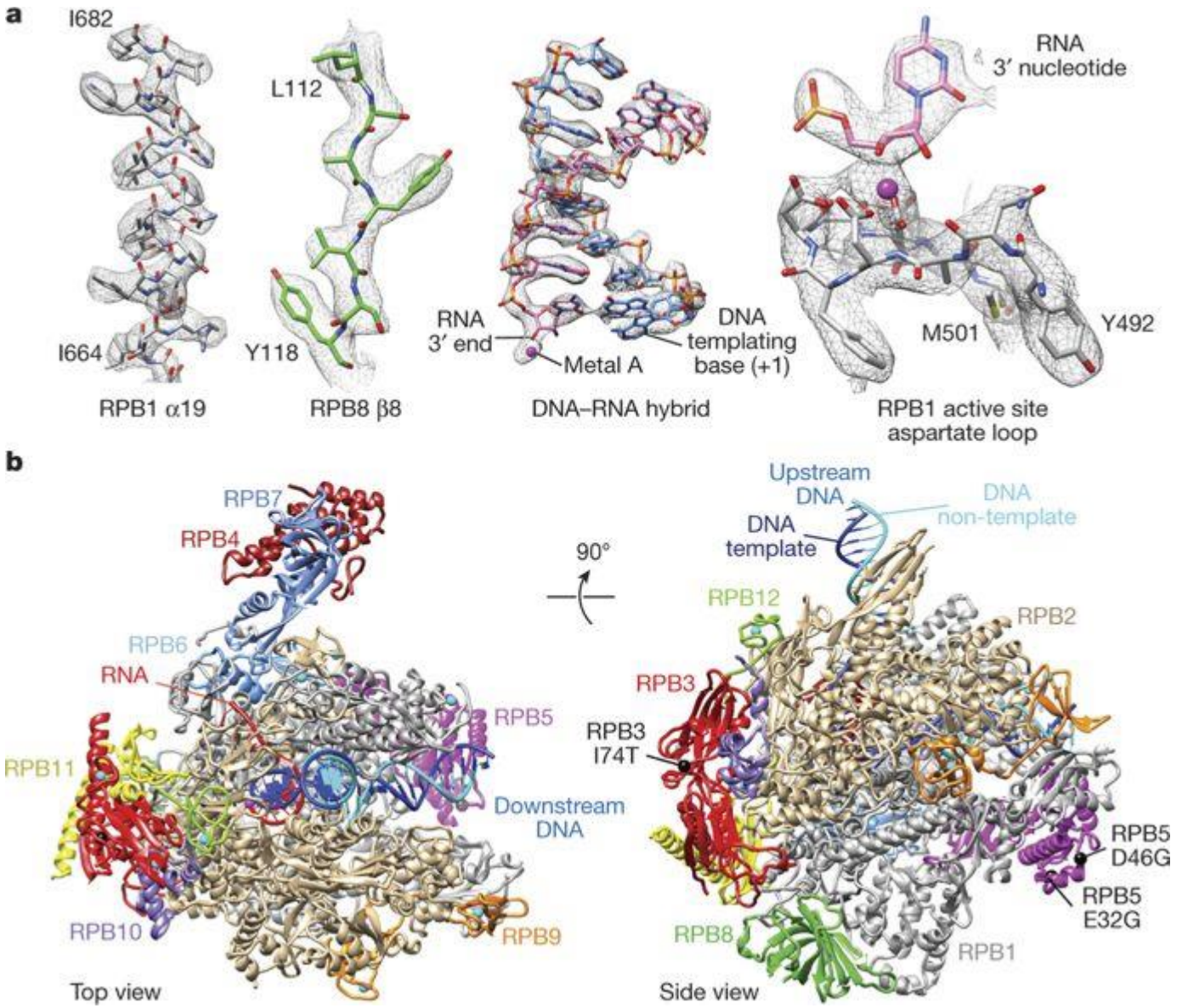
3. Transcription in eukaryotes

- Primary transcripts:
  - precursor mRNA (pre-mRNA)
  - heterogenous nuclear RNA (hnRNA)= pre-mRNA forming in nucleus
  - precursor ribosomal RNA (pre-rRNA)
  - precursor transfer RNA (pre-tRNA)
  - 5S-rRNA
  - small RNAs (snRNA, snoRNA, scRNA)
- Eukaryotic DNA-dependent RNA-polymerase:
  - RNA-polymerase I, II, III
- Transcription factors

- Eukaryotic RNA-polymerases:
  - RNA polymerase I:
    - synthesis of pre-rRNA
    - only in nucleolus
    - insensitive to α-amanitin
  - RNA polymerase II:
    - synthesis of hnRNA and some snRNA
    - sensitive to α-amanitin
  - RNA polymerase III:
    - synthesis of pre-tRNA, 5S-rRNA and some snRNA

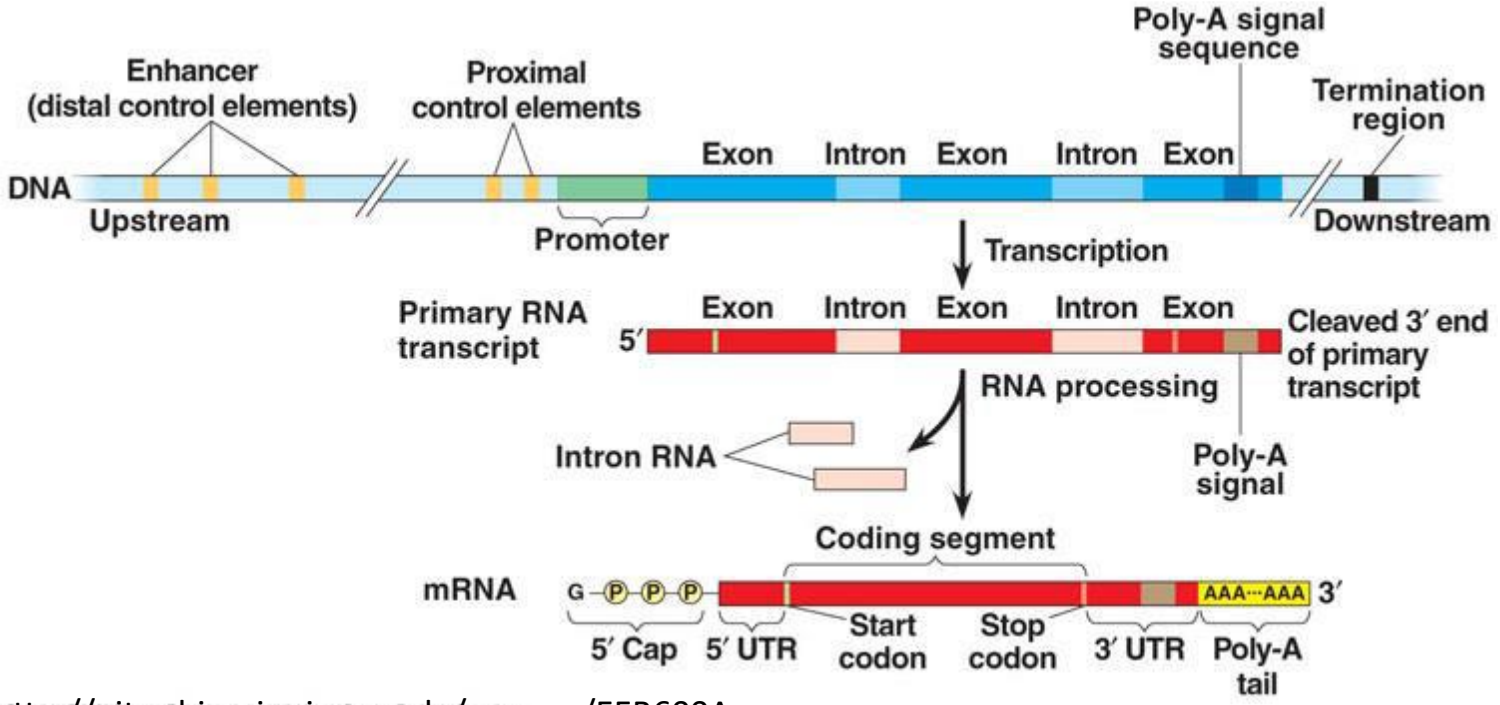
➢RNA polymerase II

a, Representative regions of the cryo-EM density for EC1 with the refined model superimposed. Depicted are (from left to right) RPB1 helix α19, RPB8 strand β8, the DNA–RNA hybrid, and the active site aspartate loop with the bound catalytic metal ion A and the 3′-nucleotide of the RNA transcript. **b**, Ribbon model. The views correspond to the previously used ‘top’ and ‘side’ views of yeast Pol II and are related by a 90° rotation around a horizontal axis. Black spheres indicate the location of residues that are not identical between bovine and human Pol II, bovine indicated second (three out of seven residues, the remaining four are disordered). The final model lacked several short surface loops and flexible N-terminal residues.



Structure of transcribing mammalian RNA polymerase II
Nature **volume 529**, pages 551–554 (28 January 2016)

- Transcription unit:
  - **monocistronic character**
  - contains:
    - promoter
    - leading sequence
    - polyadenylation signal
    - terminator

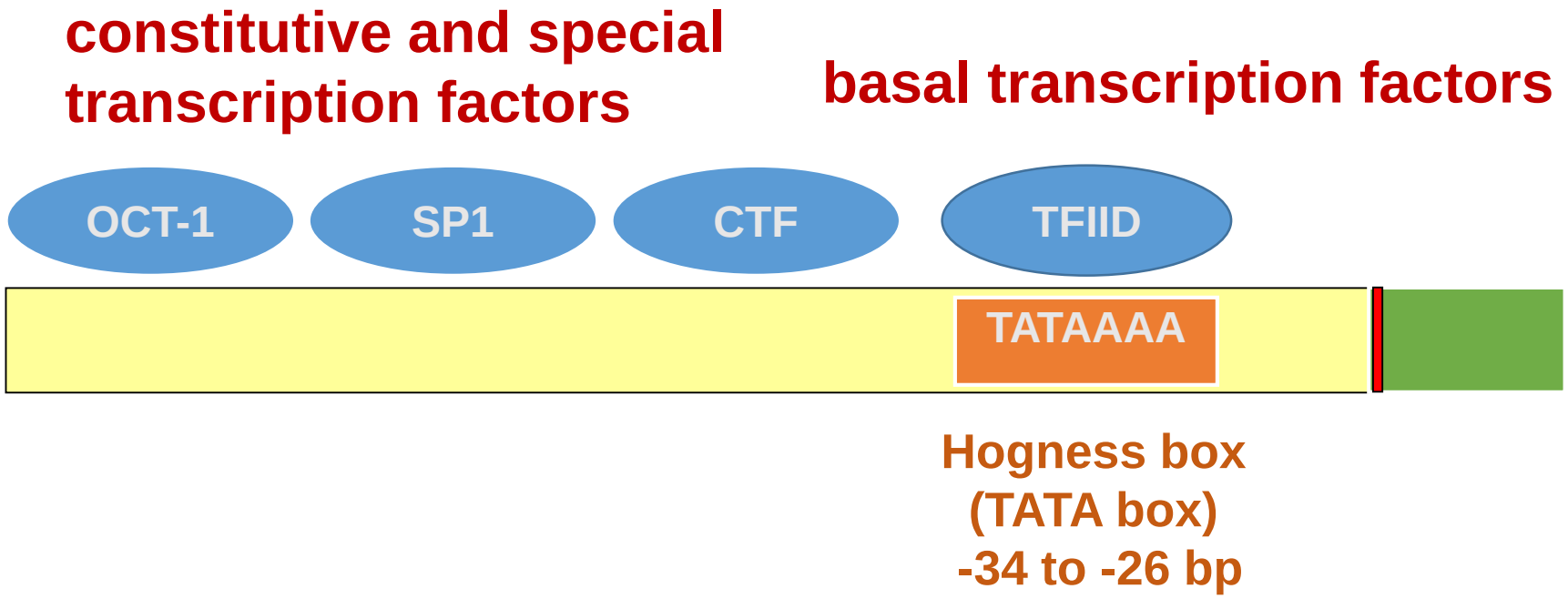


http://nitro.biosci.arizona.edu/courses/EEB600A-2003/lectures/lecture24/lecture24.html

- Transcription factors:
  - necessary for transcription inititation
  - bind to promoter sequence
  - General TFs:
    - present in all cells, necessary for initiation
    - Basal - low activity, minimal cell requirements
    - Constitutive - inrease the basal activity according to the cell type
  - Special TFs:
    - only in specific cells in specific time
    - function in inducible transcription

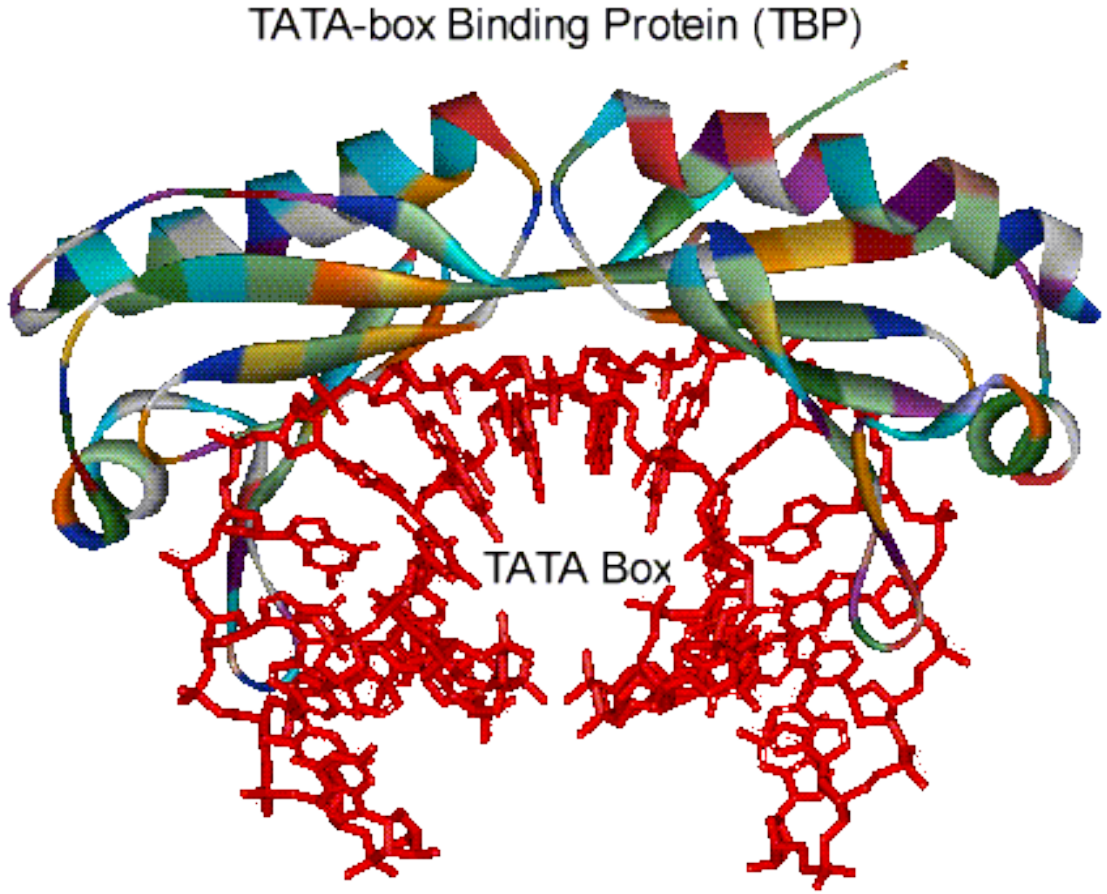
➢Eukaryotic promoter:

➢ transcription by RNA polymerase II



➢Binding to TATA box:

- ➢recognised by basal transcription actor **TFIID**
- ➢**TBP Protein** (*TATA binding protein*) is a part of TFIID, present in all eukaryotes



http://www.web-books.com/MoBio/Free/Ch4E.htm

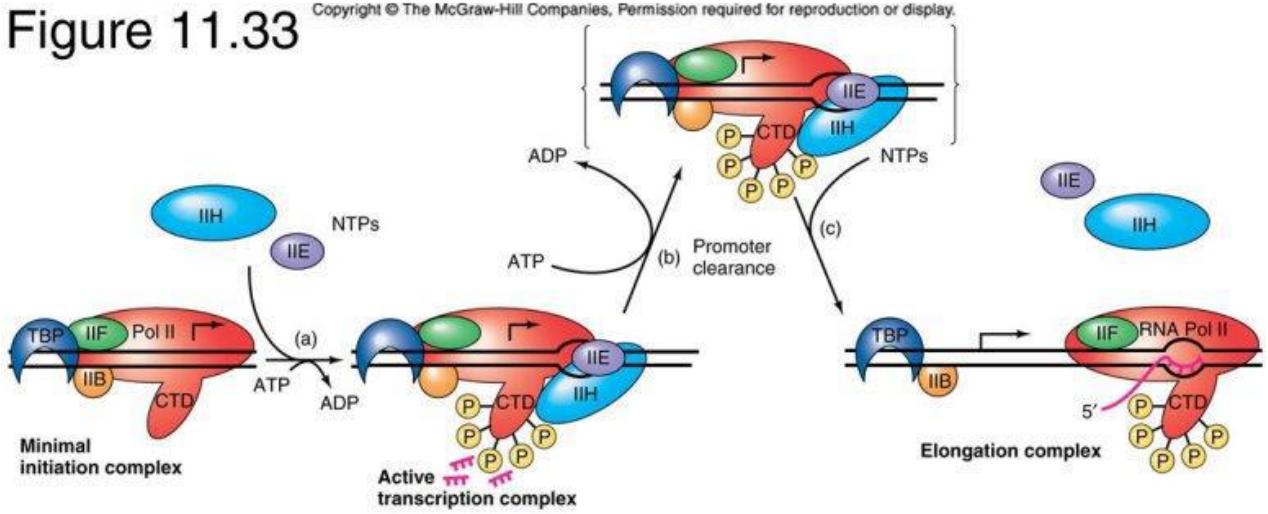
➢Transcription of hnRNA:

1. decoupling of transcription and translation
2. hnRNA is capped and methylated (binding to ribosome)
3. in the 3‘-end (after *stop codon*) AAUAAA sequence is present, hnRNA is cleaved in this region
4. hnRNA is polyadenylated on 3‘-end (stabilization in cytoplasm)
5. introns are removed and exons are connected to form mRNA

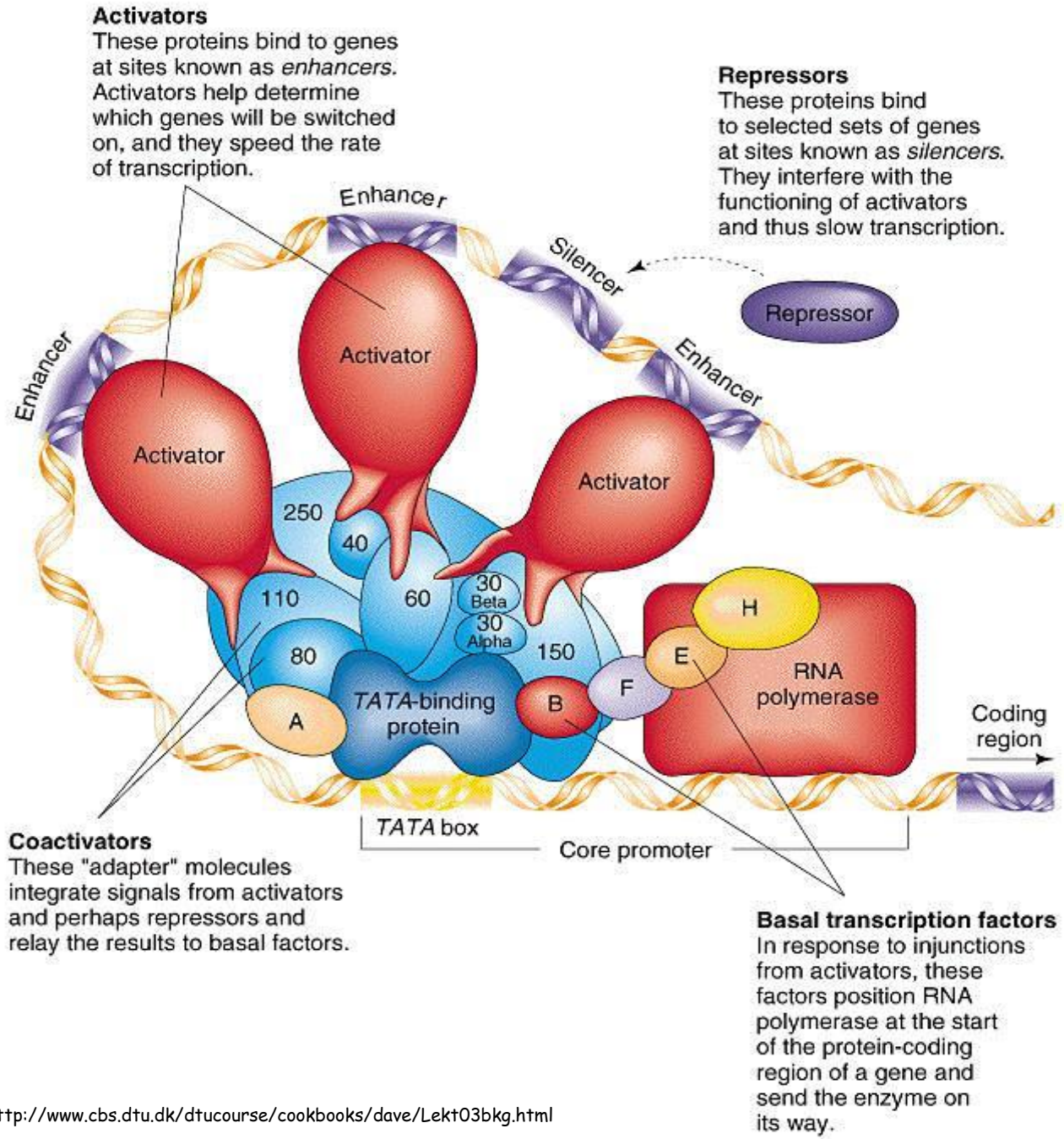
➢Initiation of transcription:

1. binding of transcription factors to TATA box and other regulatory sequences - preinitiation complex
2. binding of RNAP II - closed initiation complex
3. phosphorylation of CTD domain of RNAP II by transcription factor TFIIH (helicase and kinase activity) - RNAP II activation and unwinding of dsDNA - open initiation complex
4. dissociation of RNAP II from other TFs (apart from TFIIF) and start of RNA synthesis

Figure 11.33



- Parts of eukaryotic promoter:



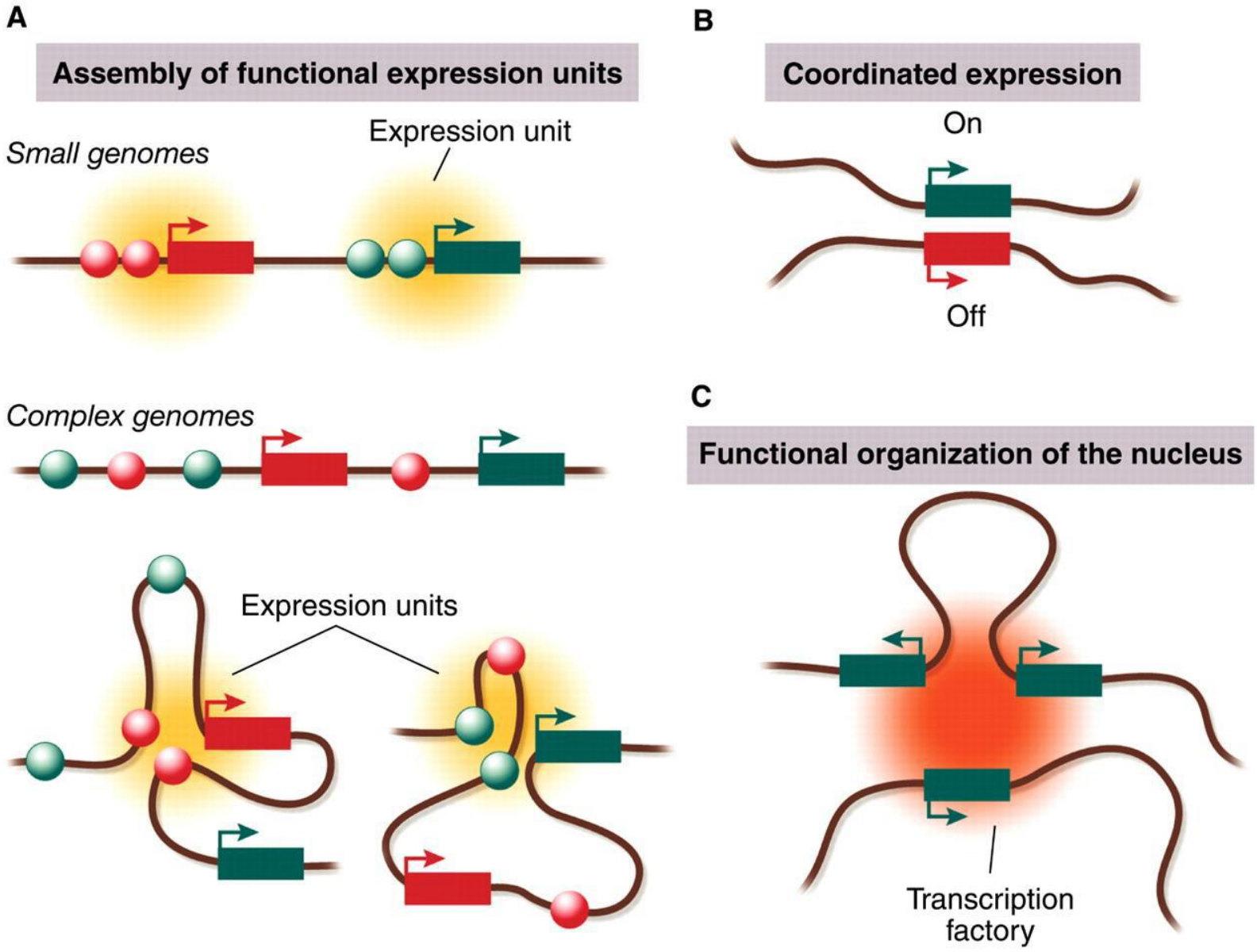
http://www.cbs.dtu.dk/dtucourse/cookbooks/dave/Lekt03bkg.html

➢Spatial assemblies during transcription:

(A)Linearly defined expression units in compact genomes and spatially assembled expression units in complex genomes.

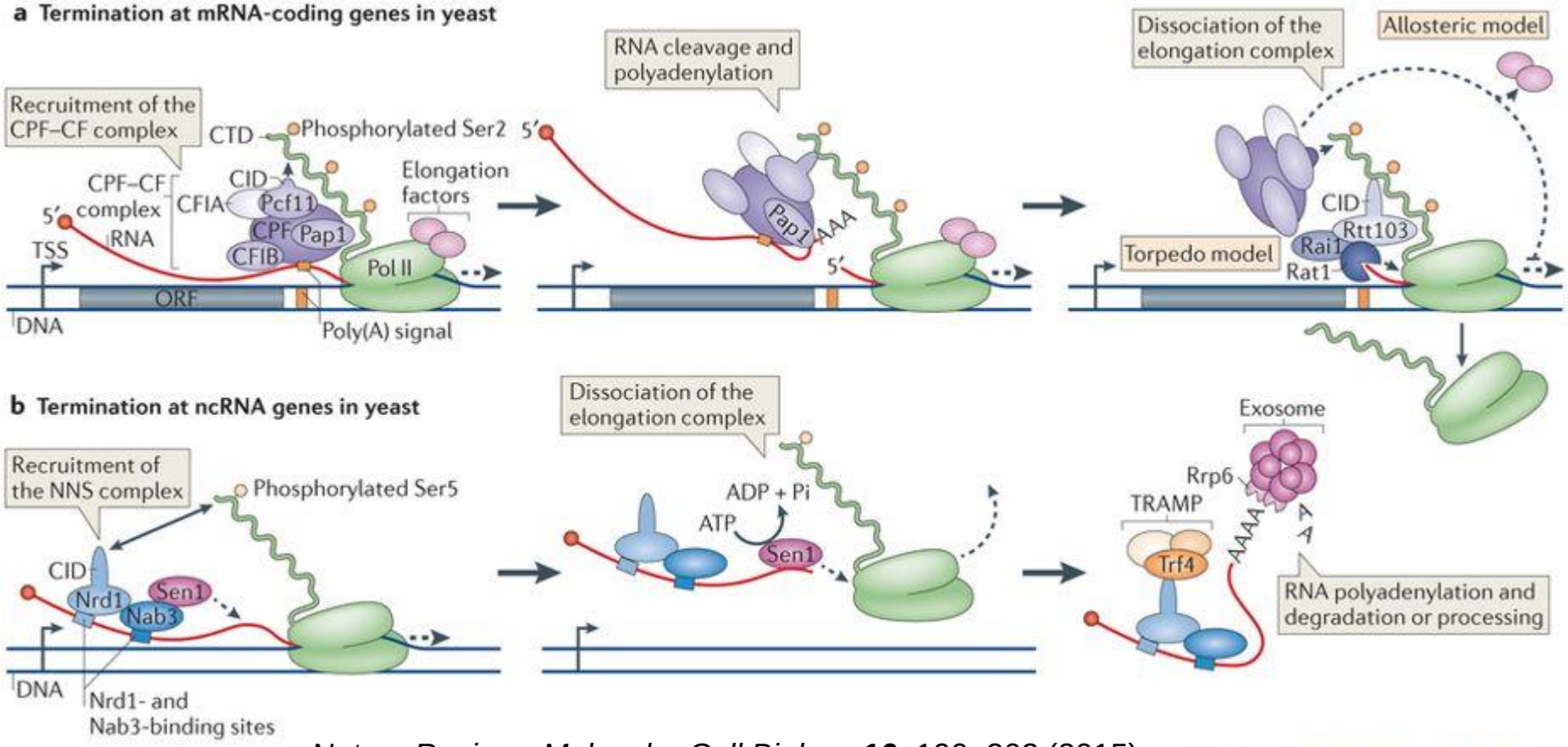
(B)Association between coordinately expressed genes.

(C)Colocalization of genes at subnuclear structures, such as transcription factories.



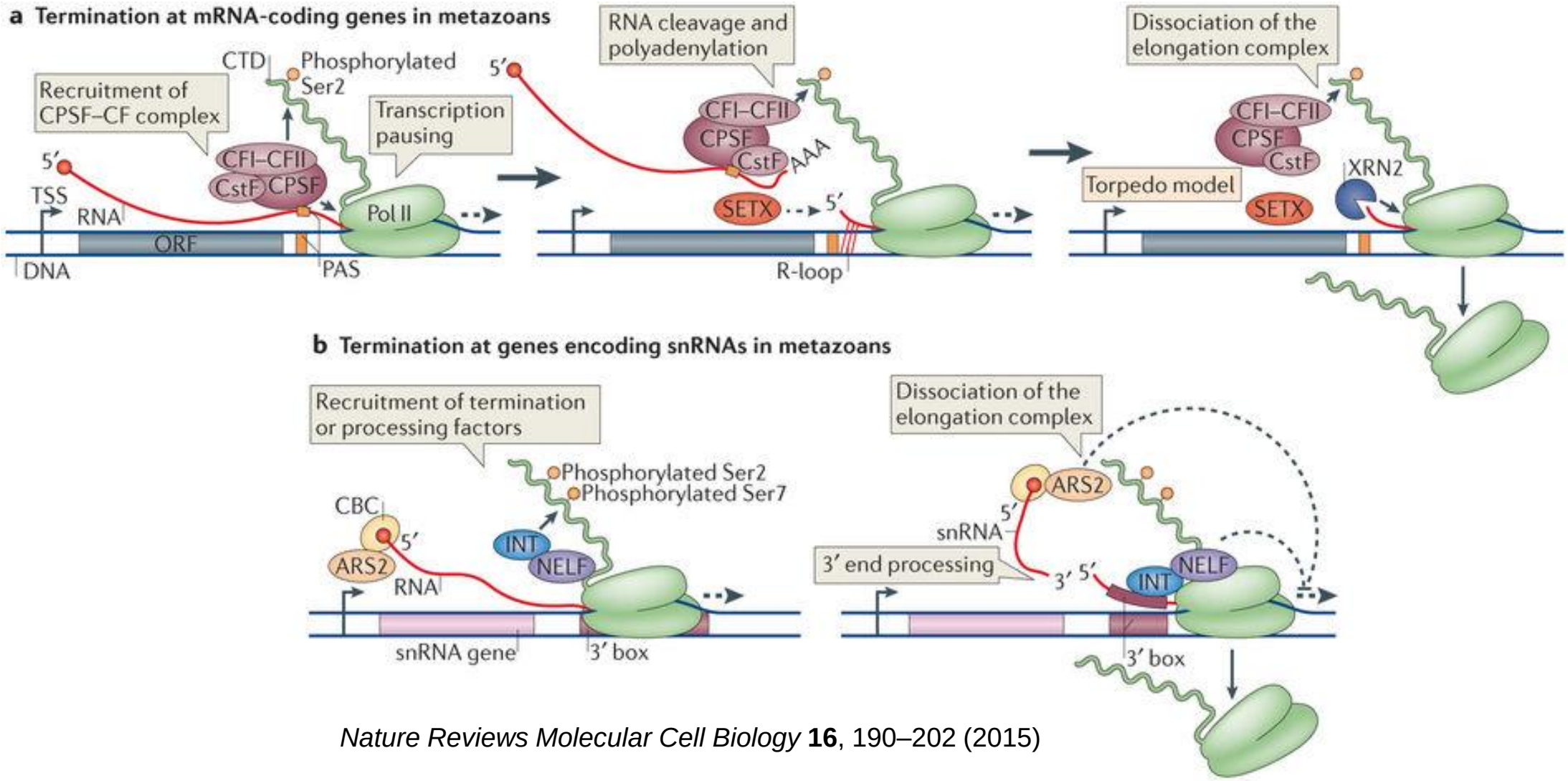
Dekker J: Science 319, 1793-1794 (2008)

- Termination of transcription:
  - terminator contains AATAAA sequence - polyadenylation signal
  - polyadenylation signal in hnRNA is recognized by protein complex, which cleaves hnRNA 10-30 nt towards 3‘- end
  - RNAP II dissociates from DNA and the rest of hnRNA is degraded



Nature Reviews Molecular Cell Biology **16**, 190–202 (2015) Nature Reviews | Molecular Cell Biology

➢Termination of transcription:



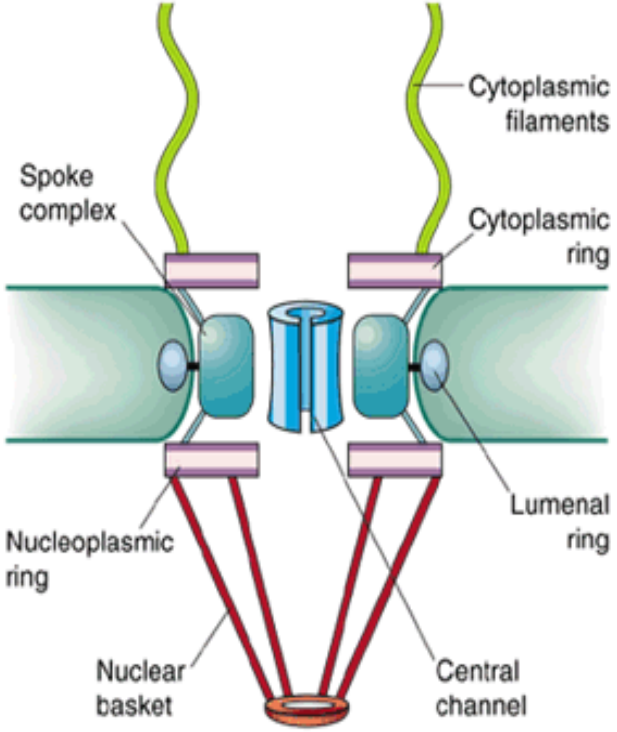
Nature Reviews Molecular Cell Biology **16**, 190–202 (2015)

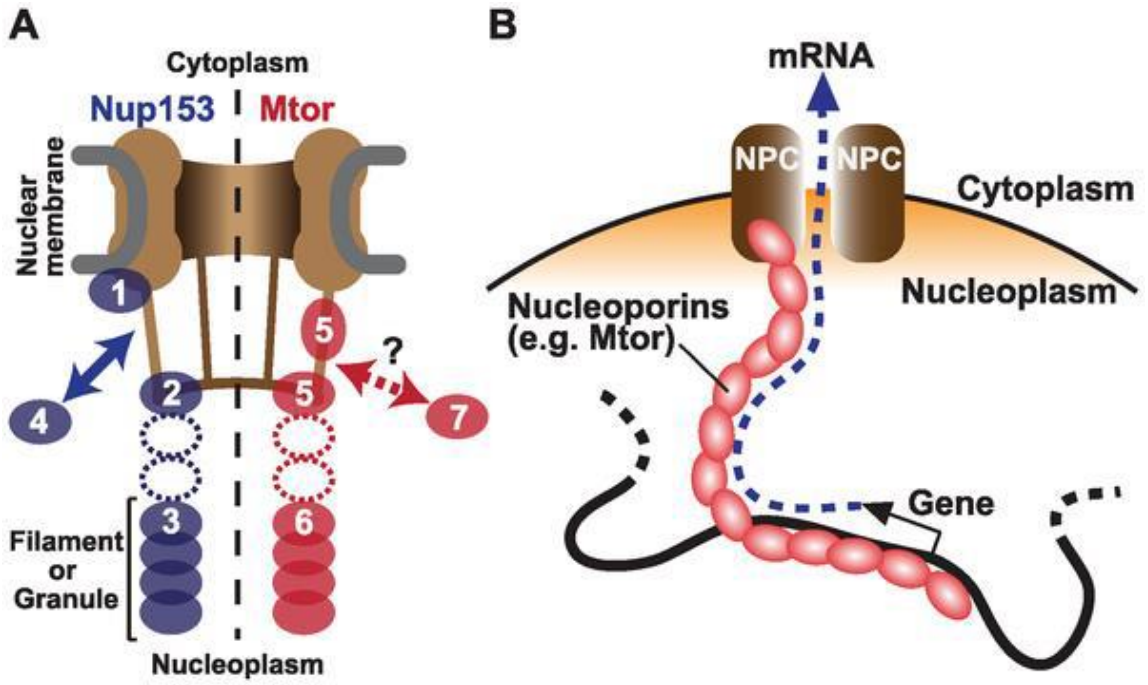
- Transcription and nucleoporins:
  - in yeast, frequently transcribed genes are located near nuclear pores
  - after activation of transcription, activated regions are transported to nuclear surface
  - multicellular organisms contain lamins, which are localized in inner surface of nucleolema (not yeast)

Ikegami, K. a Lieb, J. D. Plos Genetics 6 (2), 1-2 (February 2010)

➢Transcription and nucleoporins:

- ➢nuclear pore complexes (NPC) selectively transmit macromolecules
- ➢complexes of more than 400 proteins (nucleoporins) in 30 subunits
- ➢nucleoporins Nup153 and Mtor form filamentous structures which transport DNA from inner part of nucleus towards nuclear pores





http://cellbio.emory.edu/lab/powers/Research.html

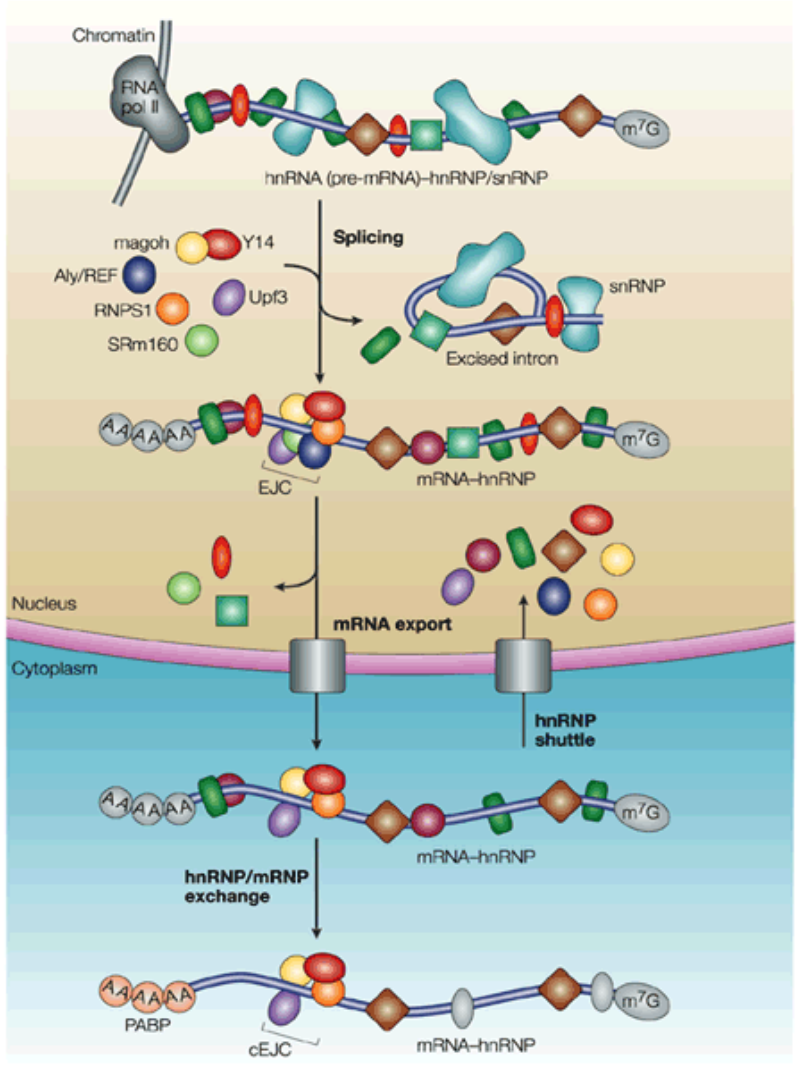
Ikegami, K. a Lieb, J. D. Plos Genetics 6 (2), 1-2 (February 2010)

- Posttranscriptional RNA processing:
  - hnRNA modifications:
    - Formation of hnRNA-protein complexes
    - Adding cap to 5‘- end
    - Polyadenylation of 3‘- end
    - Splicing of hnRNA

➢Formation of hnRNP complexes:

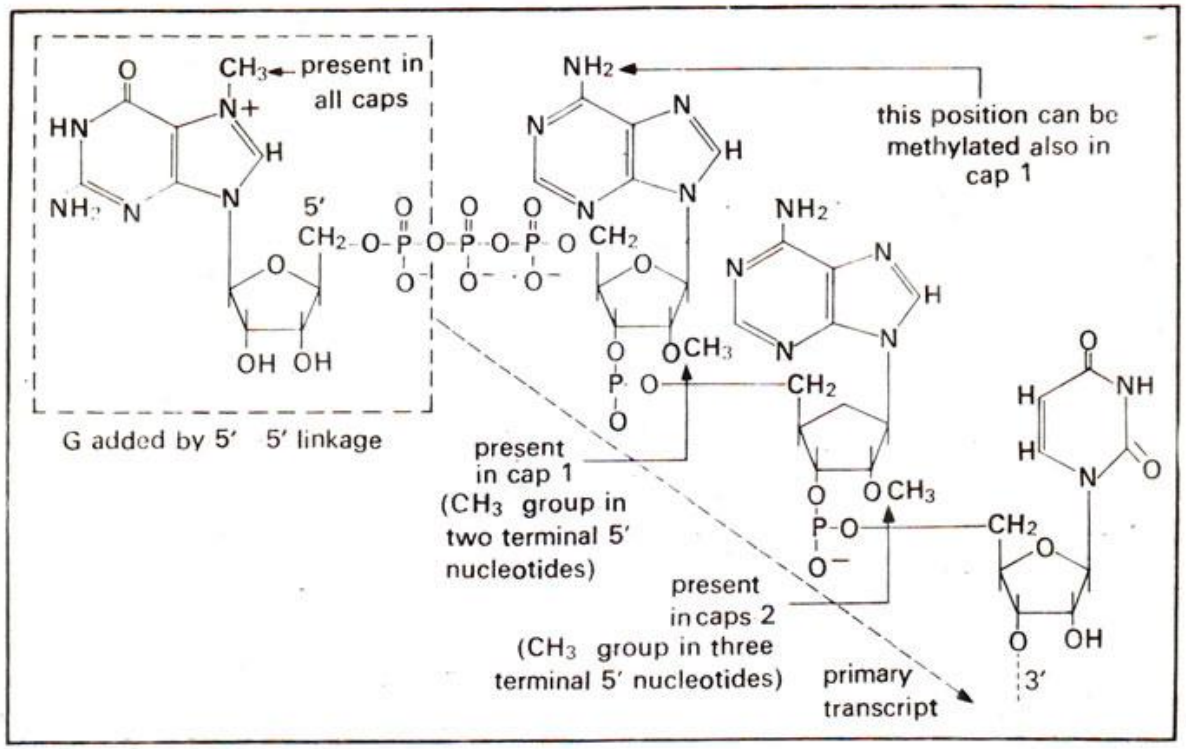
- ➢Proteins which specifically bind hnRNA- hnRNP proteins
- ➢Proteins which specifically bind to small nuclear RNA (snRNA)- snRNP proteins
- ➢snRNP proteins + snRNA= snRNP particles
- ➢hnRNA + hnRNP proteins + snRNP particles= hnRNP complex
- ➢snRNP particles bind to introns and form spliceosome
- ➢hnRNP proteins participate on transport of mRNA to the cytoplasm

➢Formation of hnRNP complexes



Nature Reviews | Molecular Cell Biology *Nature Reviews Molecular Cell Biology* **3**, 195-205 (March 2002)

- Adding cap to 5‘-end of hnRNA:
  - Binding of 7-methylguanosine (m^7G) via three phosphate groups to 5‘-end of hnRNA via 5‘-5‘ linkage
  - Also two or three 5‘-end nucleotides can be methylated
  - m^7G plays and important role in initiation of translation



http://www.biocyclopedia.com/index/genetics/expression_of_gene_protein_synthesis_rna_processing_rna_splicing_rna_editing_and_ribozymes/addition_of_caps_m7g_and_tails_polya_for_mrna.php

➢Polyadenylation of 3‘-end:

- ➢Addition of 150-250 adenosines to 3‘-end= poly(A)sequence
- ➢Catalysed by poly(A)-polymerase
- ➢poly(A)-polymerase is a part of complex which binds to polyadenylation signal on hnRNA
- ➢poly(A)-end is crucial during transport of mRNA to cytoplasm and its stabilization

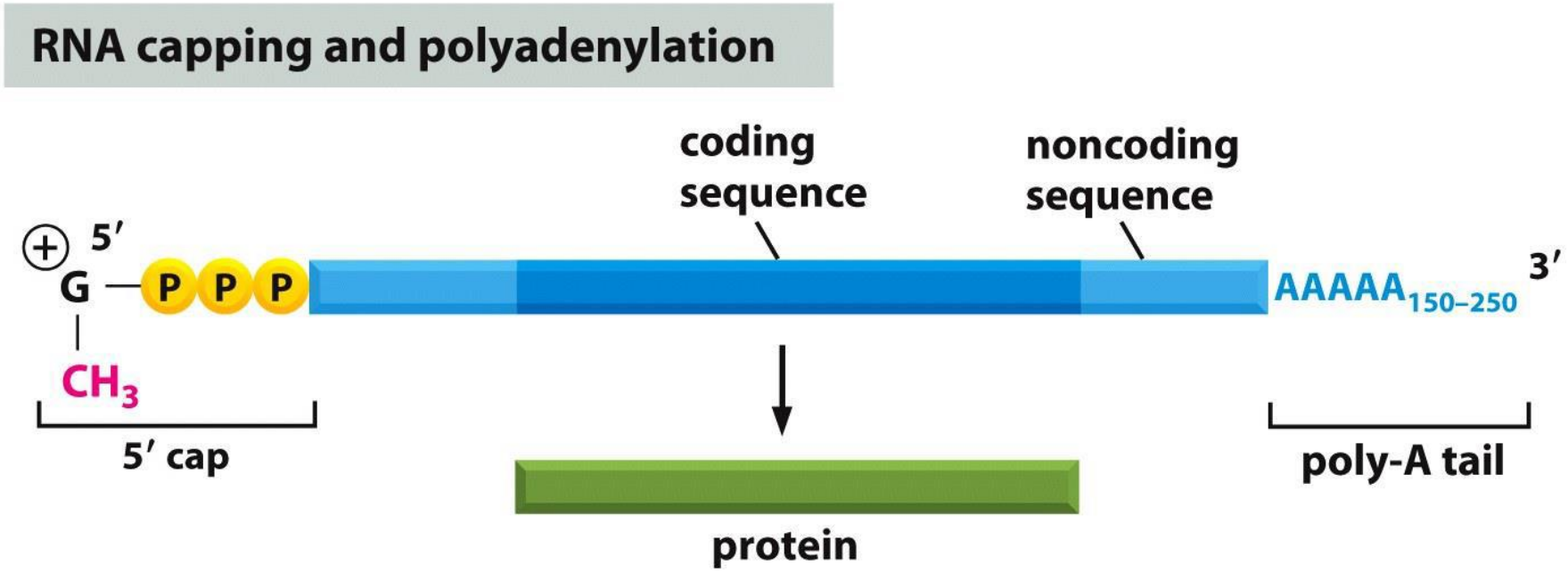
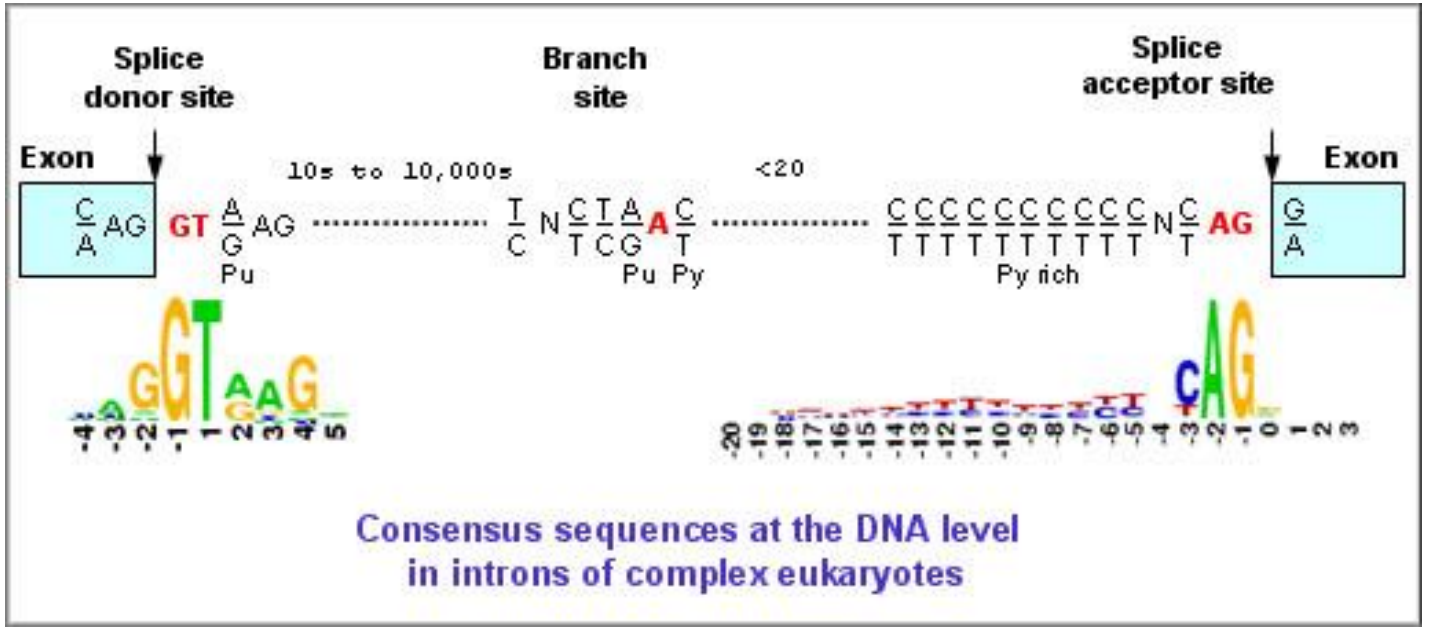


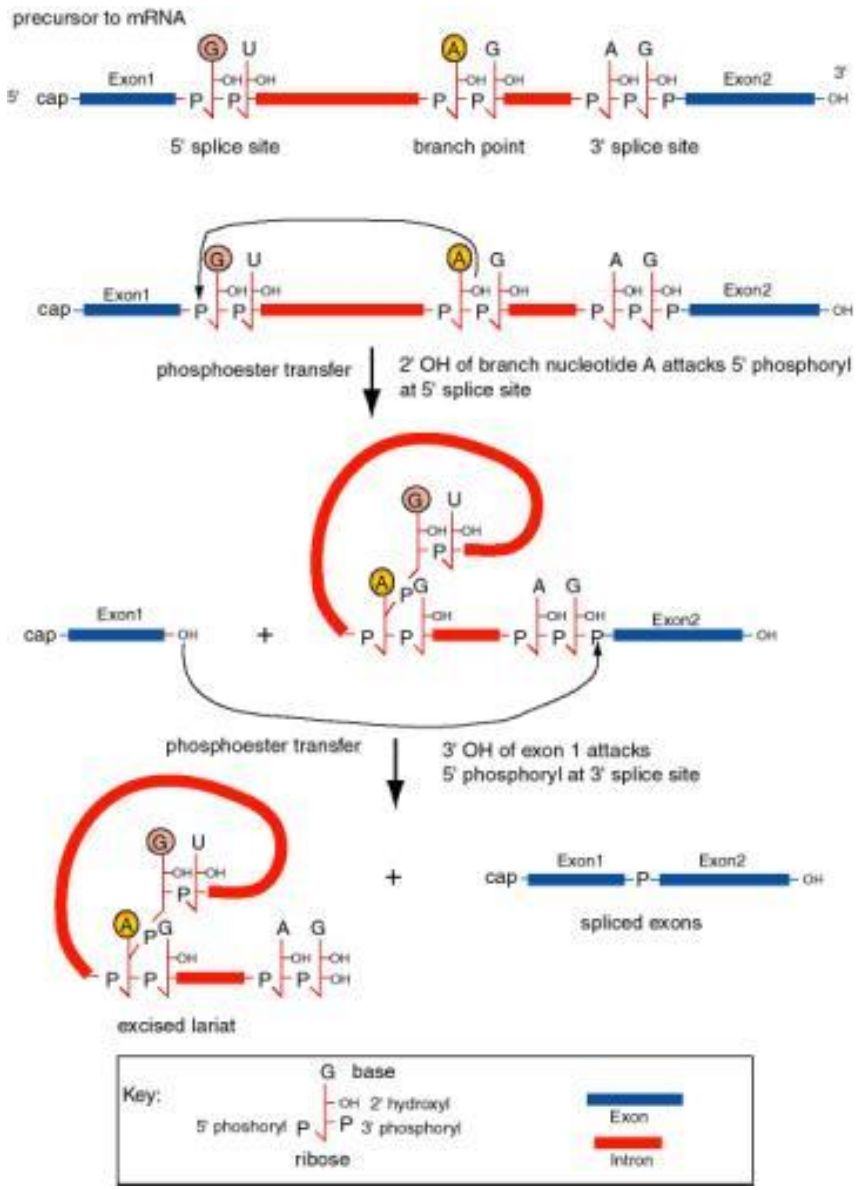
Figure 7-16a Essential Cell Biology 3/e (© Garland Science 2010)

- Splicing of hnRNA:
  - Introns are cleaved out of hnRNA to form mRNA
  - Structure of intron:
    - GU-AG rule (donor and acceptor sites)
    - Branch site



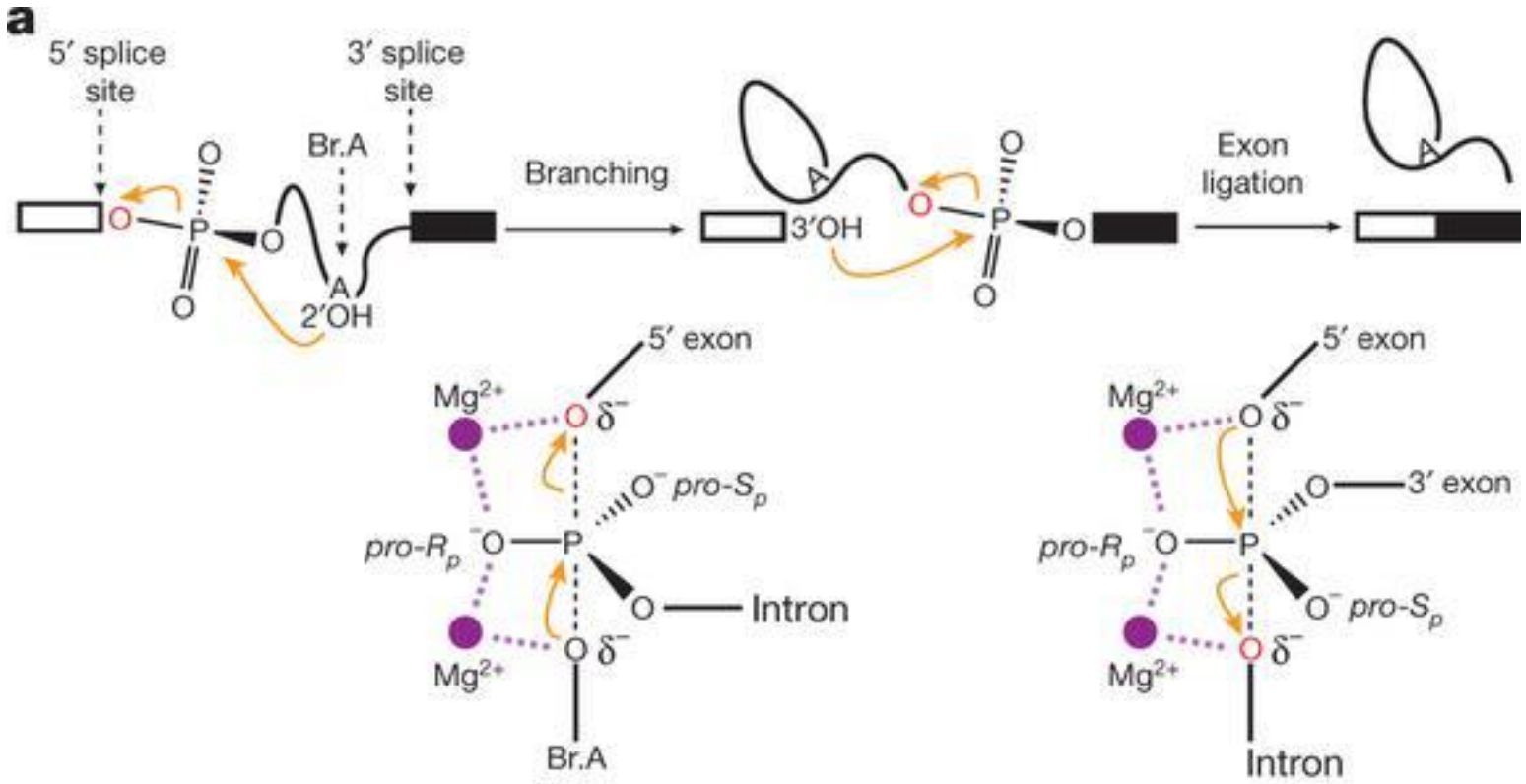
http://www.geneinfinity.org/sp/sp_coding.html

- Splicing of hnRNA:
  - transesterification – no energy from ATP or GTP is needed
  - snRNA and snRNP particles play crucial role
  - intron is cut out in the form of lariat RNA



http://www.bx.psu.edu/~ross/workmg/RNAProcessingCh12.htm

➢Role of Mg^{2+} ions during splicing:

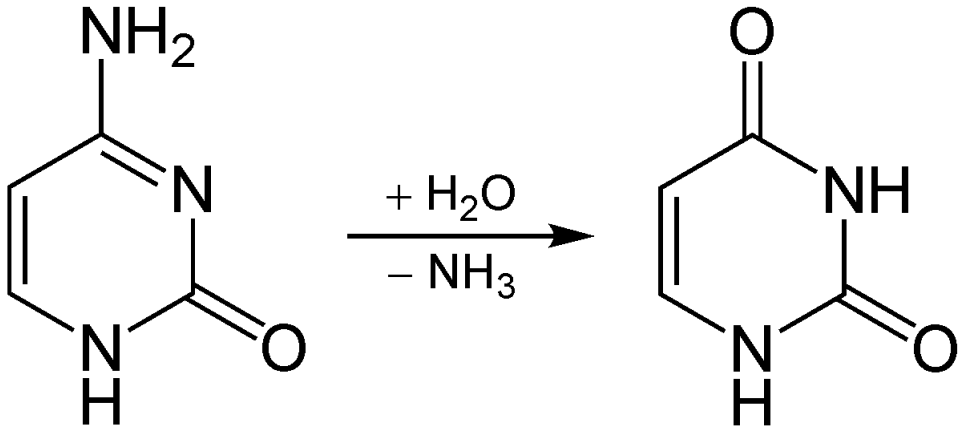


Nature **503**, 229–234 (14 November 2013) doi:10.1038/nature12734

- Self-splicing:
  - Rare autocatalytic process of hnRNA splicing
  - No proteins (enzymes) are needed
  - Digestion and ligation of RNA during self-splicing is catalyzed by ribozymes (ribonucleic acid enzymes)

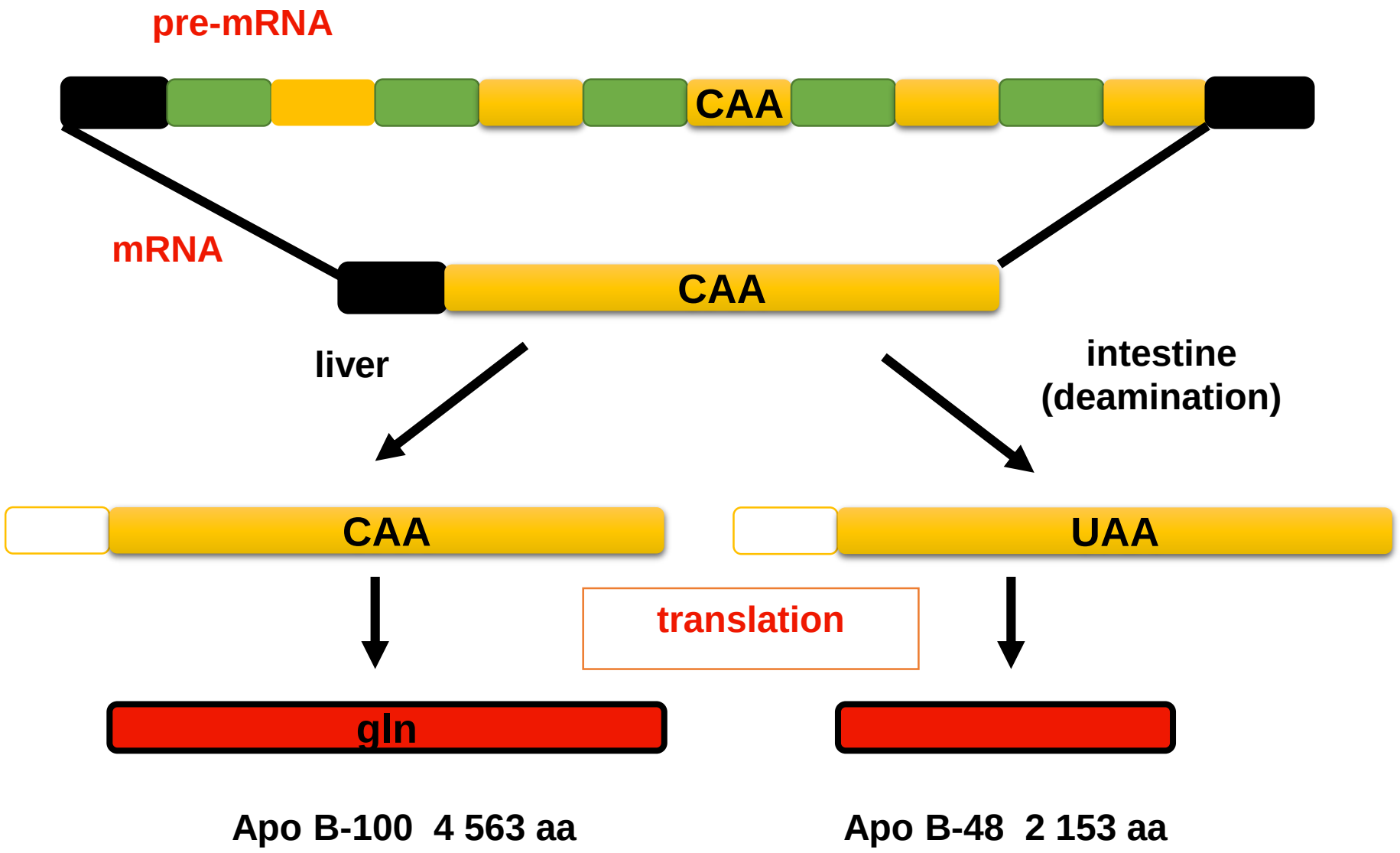
- RNA editing:
  - posttranscriptional insertion or deletion of nucleotides in RNA or conversion of one base to another
  - results in RNA transcript which does not correspond to original coding sequence in DNA
  - Two types:
    1. site-specific deamination
    2. gRNA (guide RNA)-directed editing

- Deaminaton C → U:
  - In specific mRNAs only in certain tissues or cell types
  - Two forms of apolipoprotein B
    - Liver: long Apo B-100
    - Intestine: short Apo B-48
  - Formation of stop codon UAA

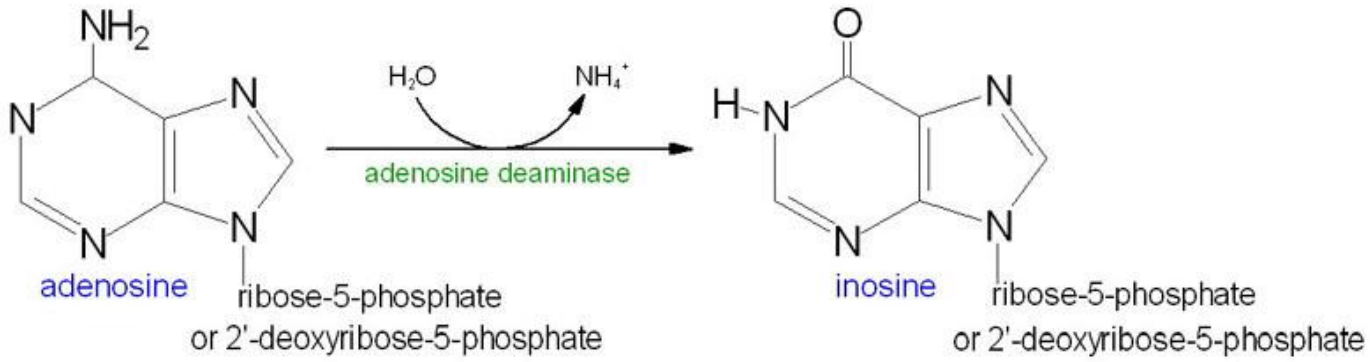


Cytidine deaminase

➢mRNA editing in apolipoprotein B gene:

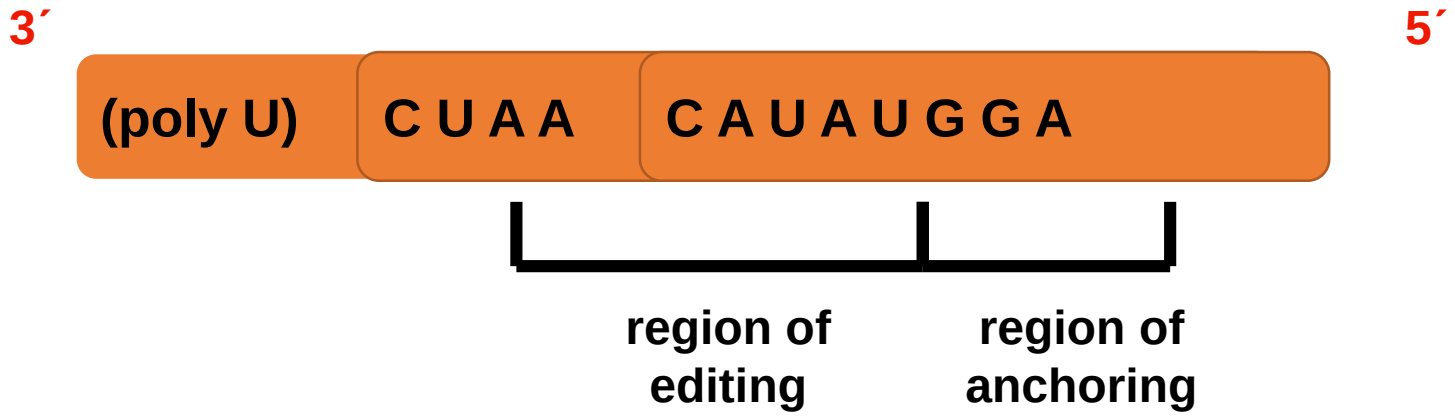


- Deamination A → I:
  - In ion channels in mammalian brain
  - Single nucleotide conversion changes the coded aminoacid
  - This changes the permeability of ion channel to Mg^{2+} ions

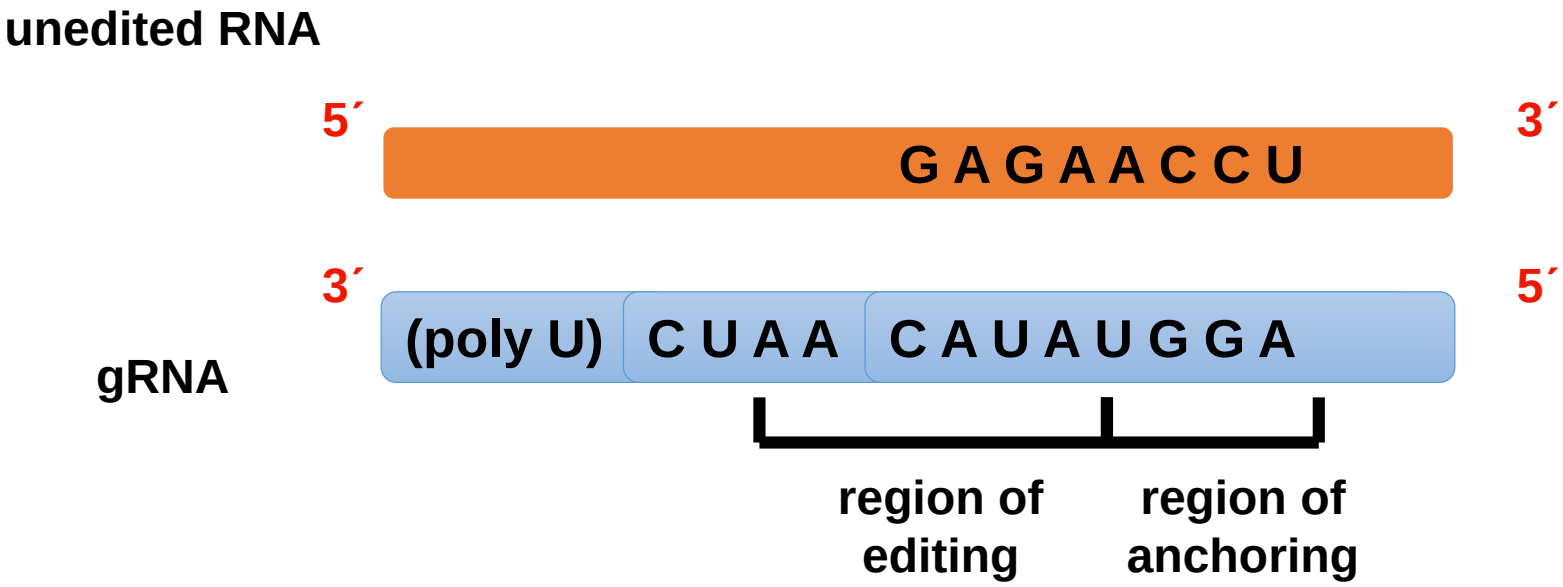


- gRNA-directed editing:
  - gRNAs (guide RNA) are 40-80 nucleotides long
  - First described in *coxII* gene in *Trypanosome*
  - Enable adding of U in specific regions of transcripts
  - gRNAs bind to mRNA, enable their splicing, adding the missing nucleotides and linking of the spliced fragments
  - Resulting mRNAs contain large segments of added U and lose several U from original sequence
  - Insertions of U can be present in up to 50 % edited mRNAs

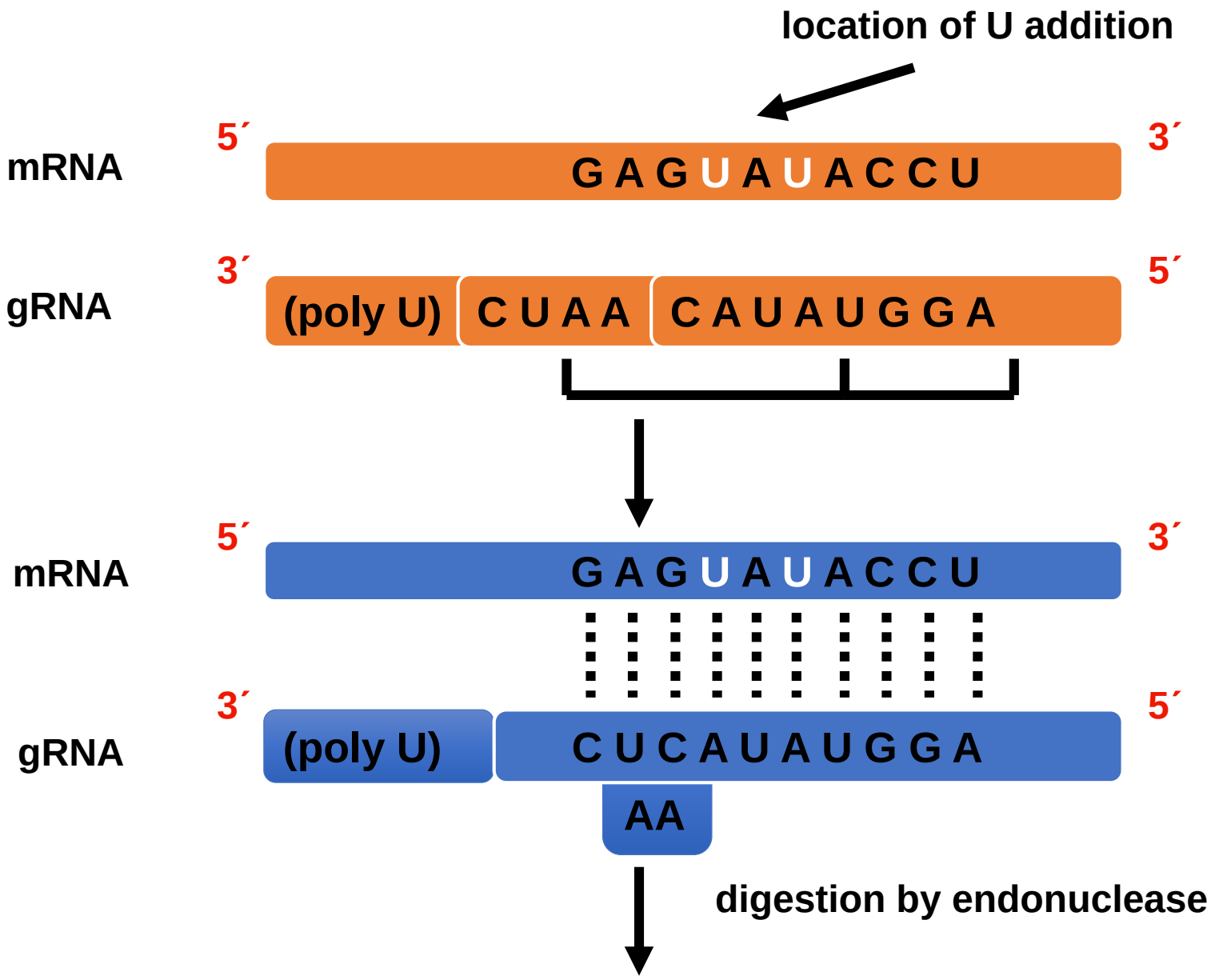
- gRNA structure:
  - each gRNA has 3 regions:
    1. first, in 5‘-end (*anchor*), enables binding of gRNA to region of mRNA editing
    2. the second directs which nucleotides will be inserted
    3. polyU sequence at 3‘-end



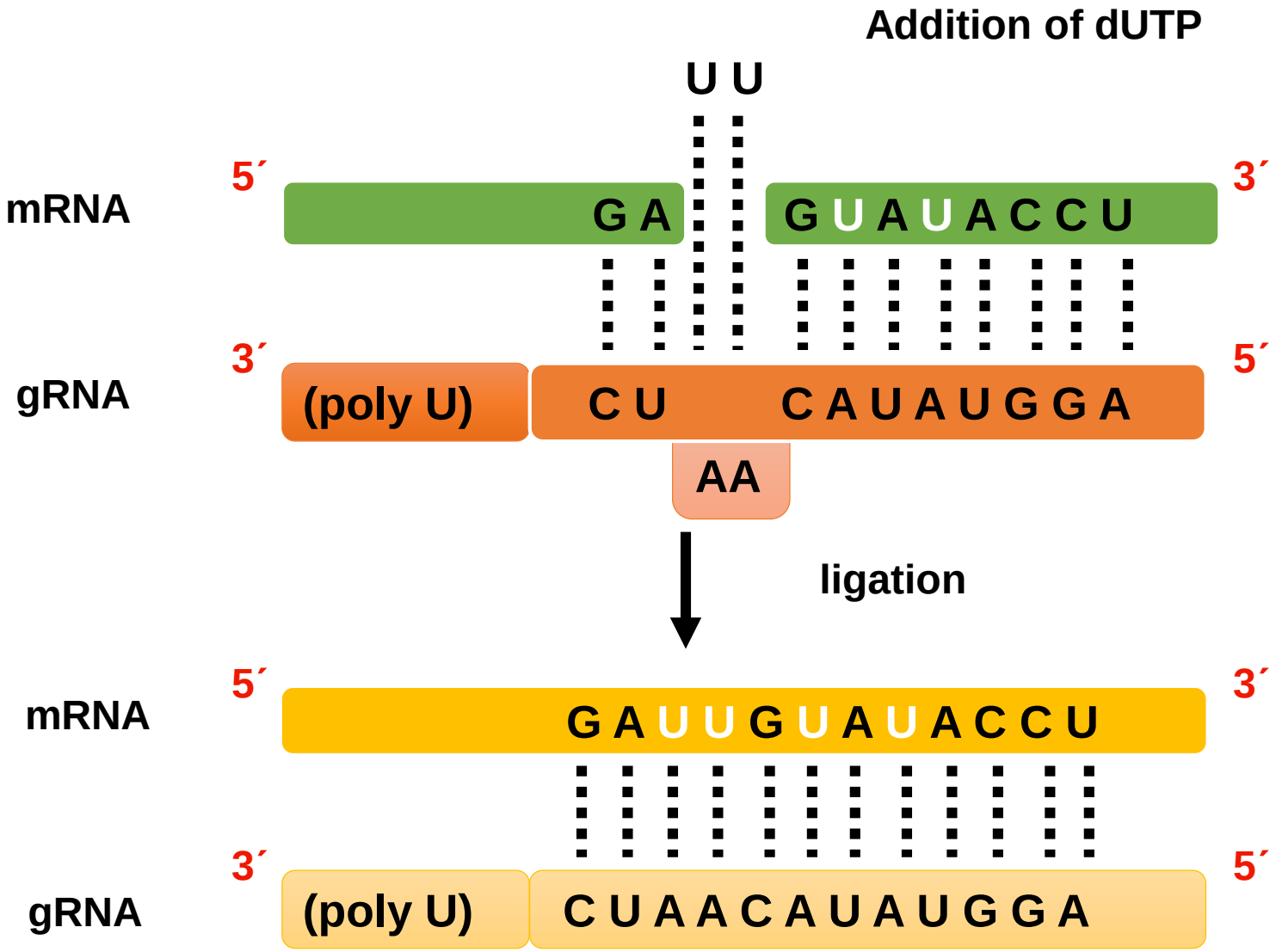
➢ Sequence of gRNA and unedited mRNA



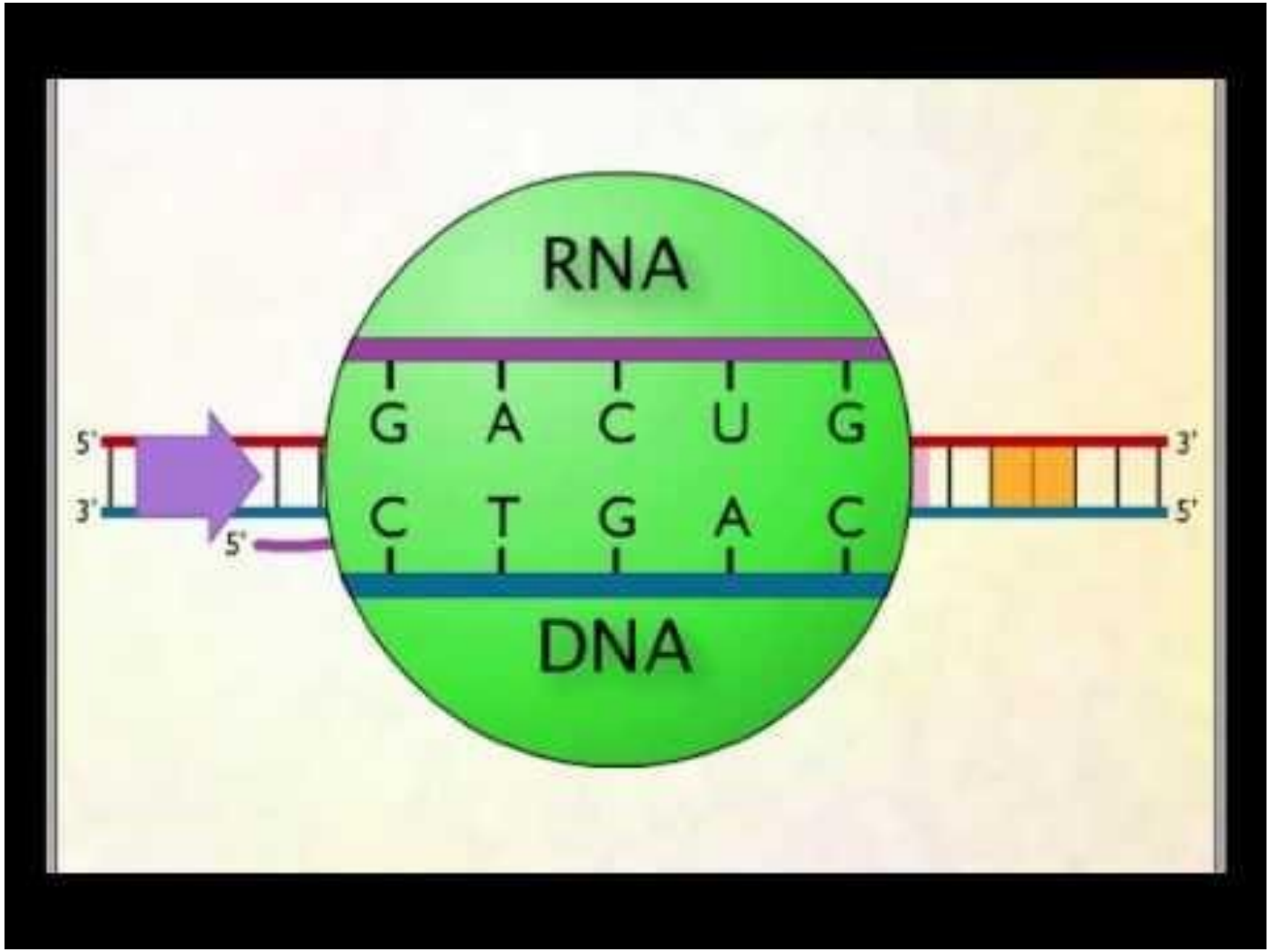
➢Process of editing I:



➢Process of editing II:

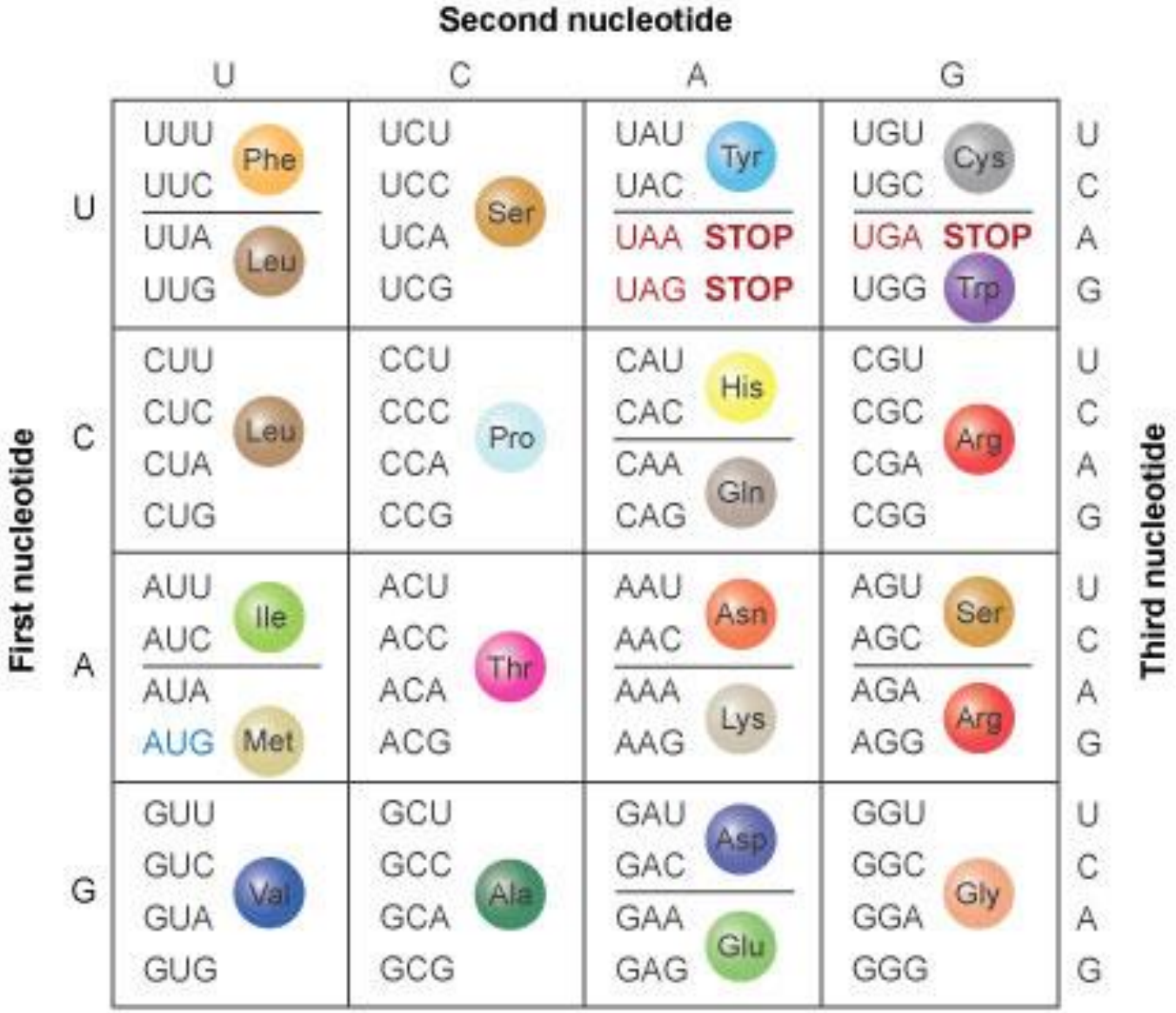


➢Transcription:



https://www.youtube.com/watch?v=WQbE7I-tV44

➢Codons:



https://www.nature.com/scitable/content/the-amino-acids-specified-by-each-mrna-6903567

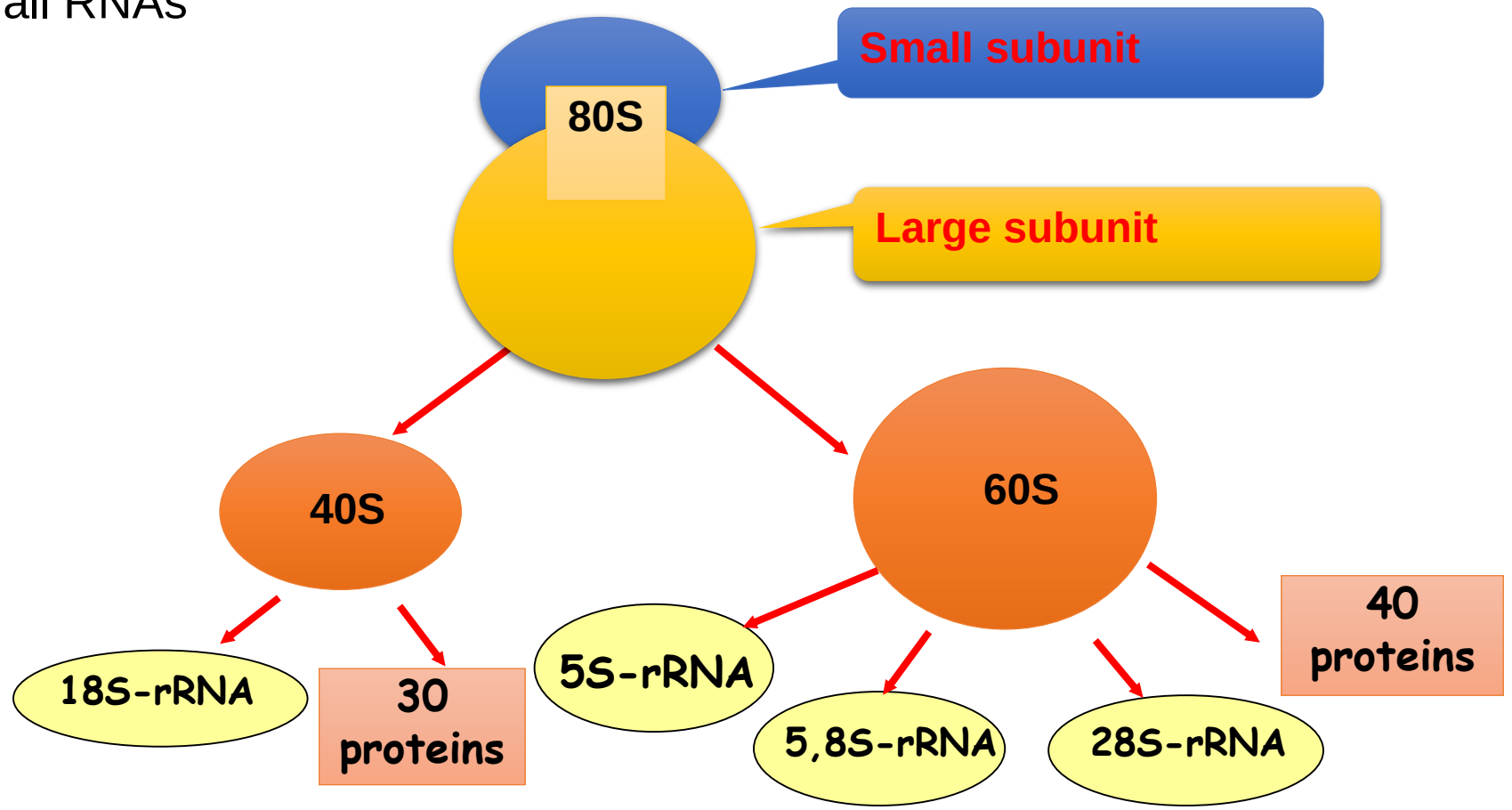
4. Translation in eukaryotes

- Differences from prokaryotic translation:
  - it takes place in 2-3 compartments – cytoplasm, mitochondria, chloroplasts
  - the initial AA is not fMet but Met, which binds to a specific initiation tRNAiMet that recognizes the initiation codon AUG
  - the number of initiation factors required for translation initiation is higher than in prokaryotes

- Translation in eukaryotes:
  - Similar to translation in prokaryotes
  - initiation, elongation a termination
  - Individual complexes are more complicated
  - Higher number of initiation factors
  - Genetic code in mammalian mitochondria has a different meaning for some codons, 22 mitochondrial tRNA genes
  - Eukaryotic cells possess 45 tRNAs differing in anticodons
  - Translation takes place at 1-20 AA/s, depending on organism and conditions

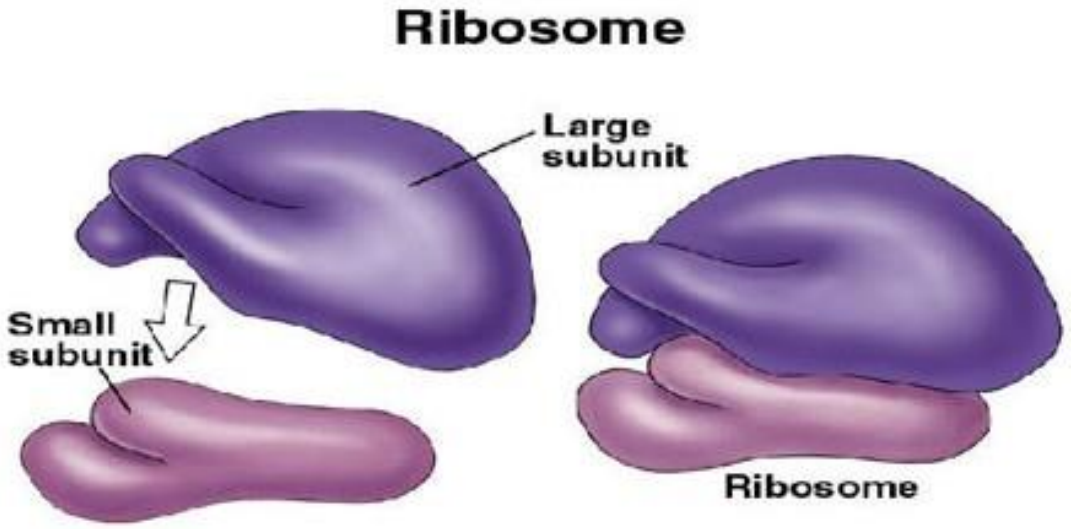
➢Cytoplasmic ribosomes:

- ➢ Also involved in the formation of ribosomes:
- ➢ 150 non-ribosome proteins
- ➢ 100 small RNAs



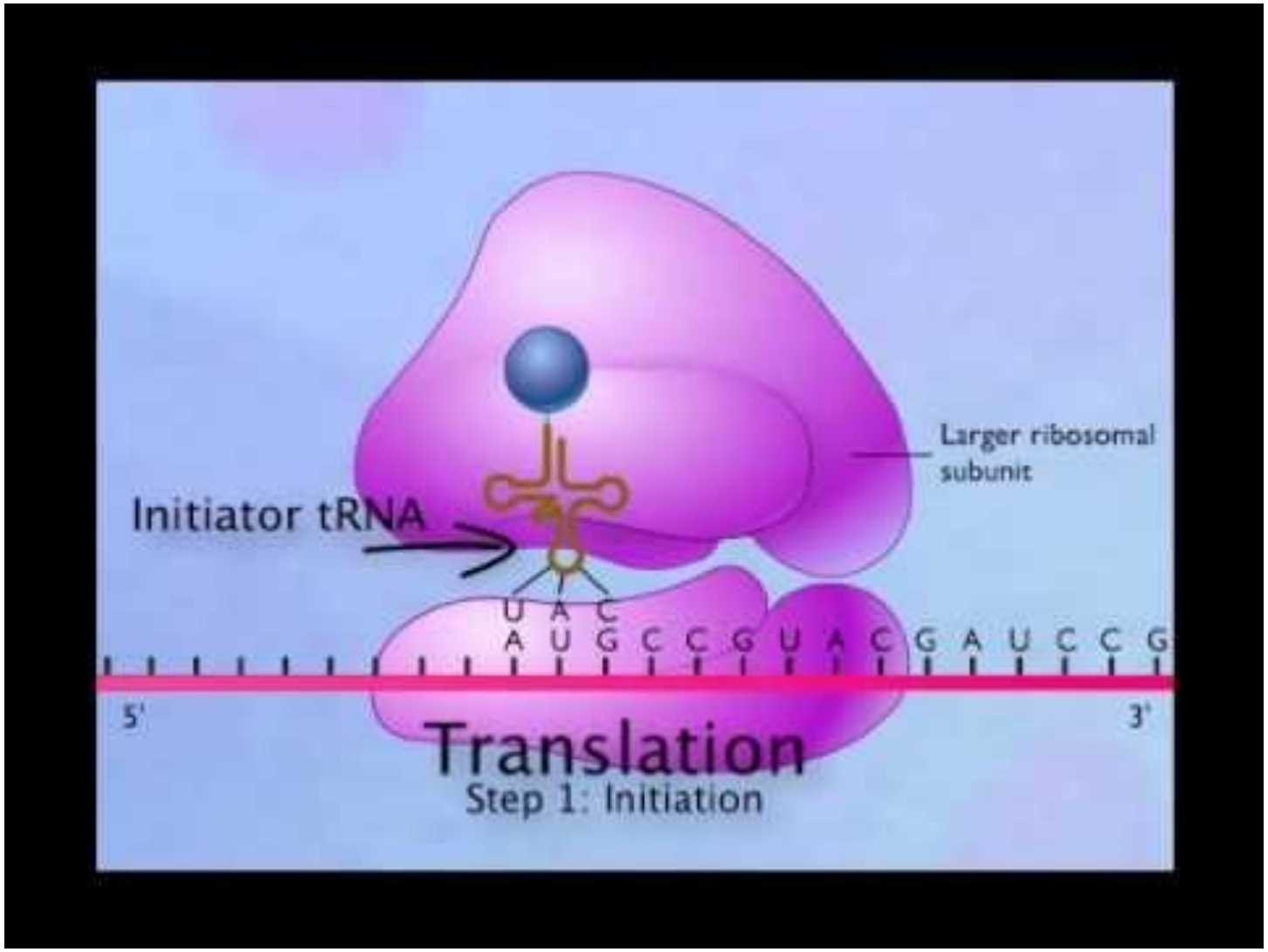
Ferreira-Cerca, S. et al. (2007): Analysis of the In Vivo Assembly Pathway of Eukaryotic 40S Ribosomal Proteins, Molecular Cell 28, 446-457, November 2007

- Bound and free ribosomes:
  - Free ribosomes occur in the cytoplasm
    - Intracellular protein synthesis
  - Others are bound to endoplasmic reticulum (ER)
    - Rough ER is covered with ribosomes
    - Smooth ER is ribosome-free

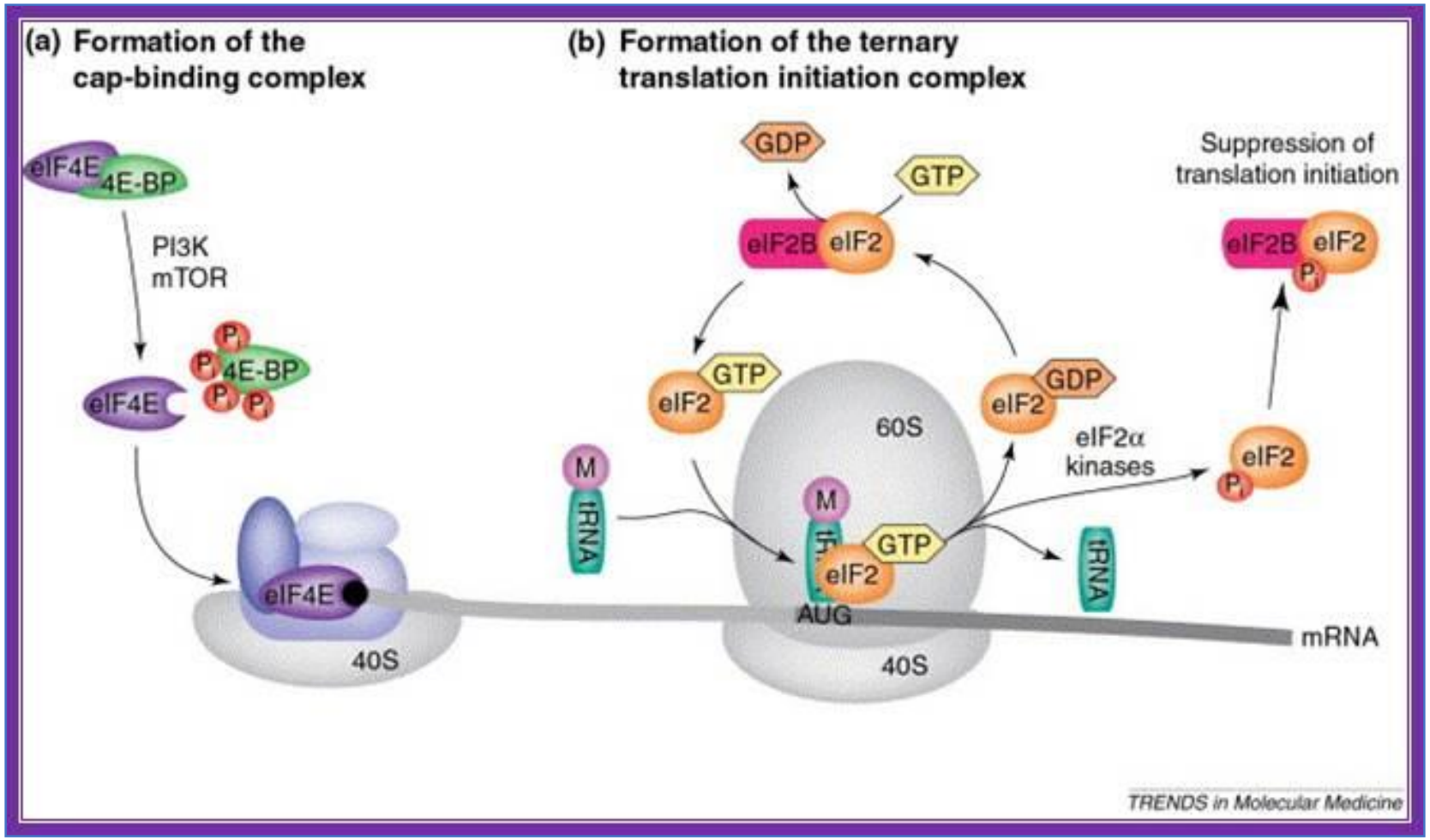


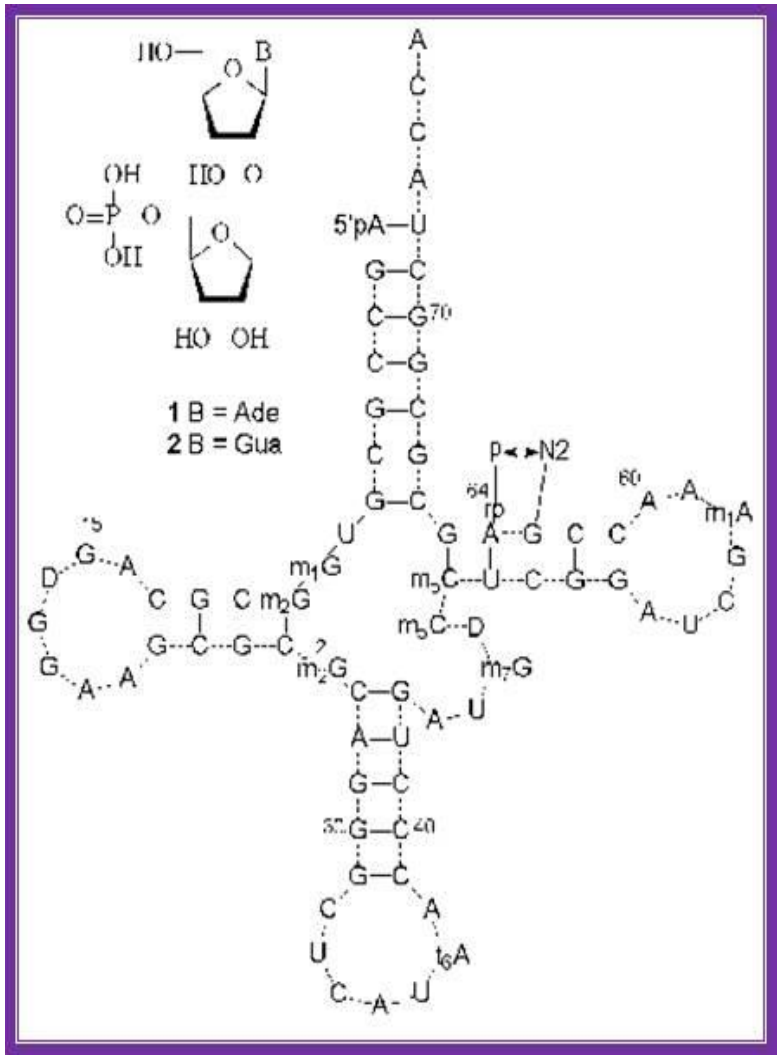
http://www.assignmentpoint.com/science/biology/about-ribosome.html

- Initiation of translation:
  - 40S subunit with tRNAiMet bound at the P site, initiation factors bind to the m^7G mRNA cap
  - The whole complex moves in the 5-3 direction before it hits the AUG initiation codon
  - Hydrolysis of GTP, the 60S subunit binds

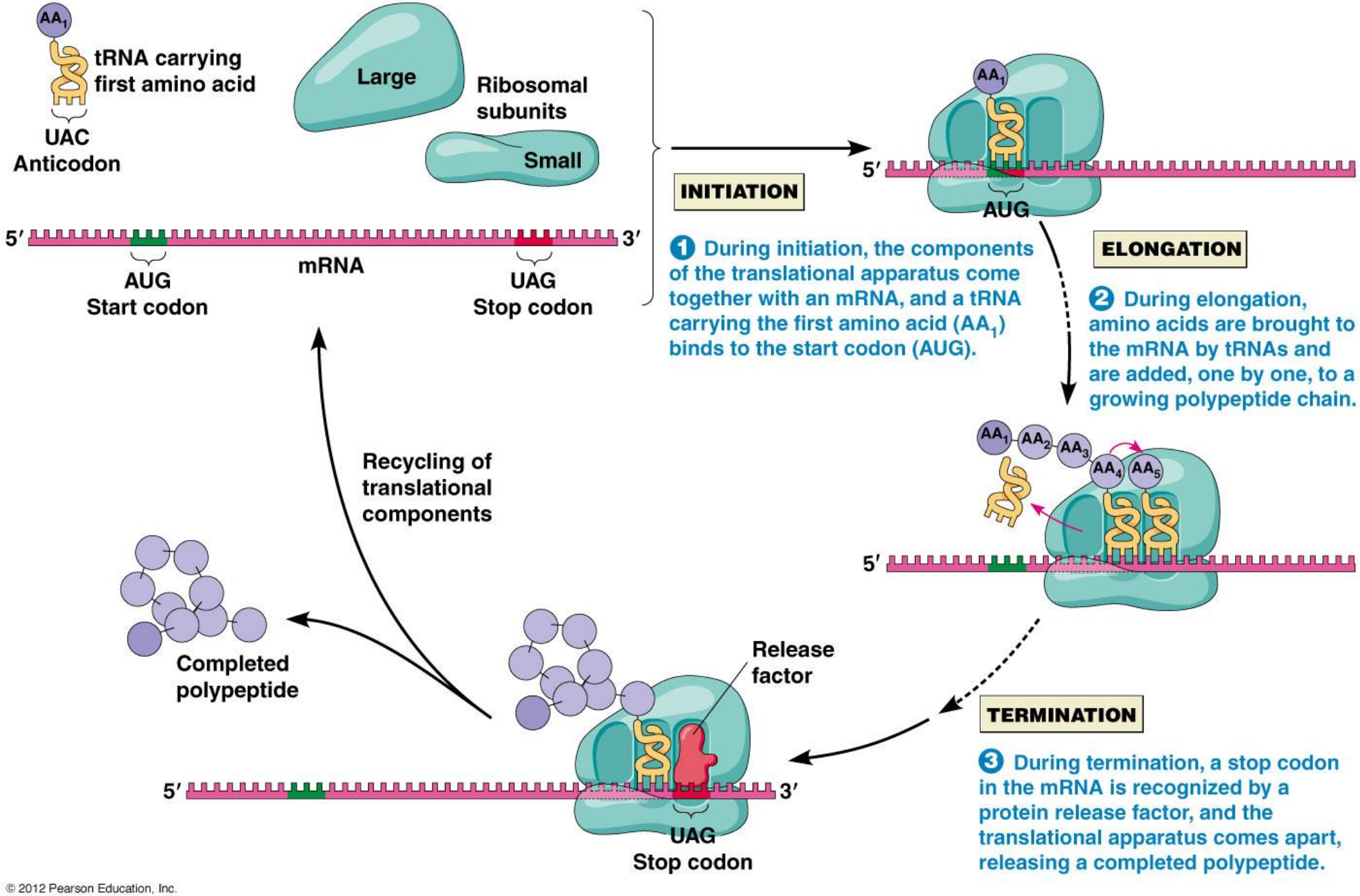


https://www.youtube.com/watch?v=kAeLta-Bst0





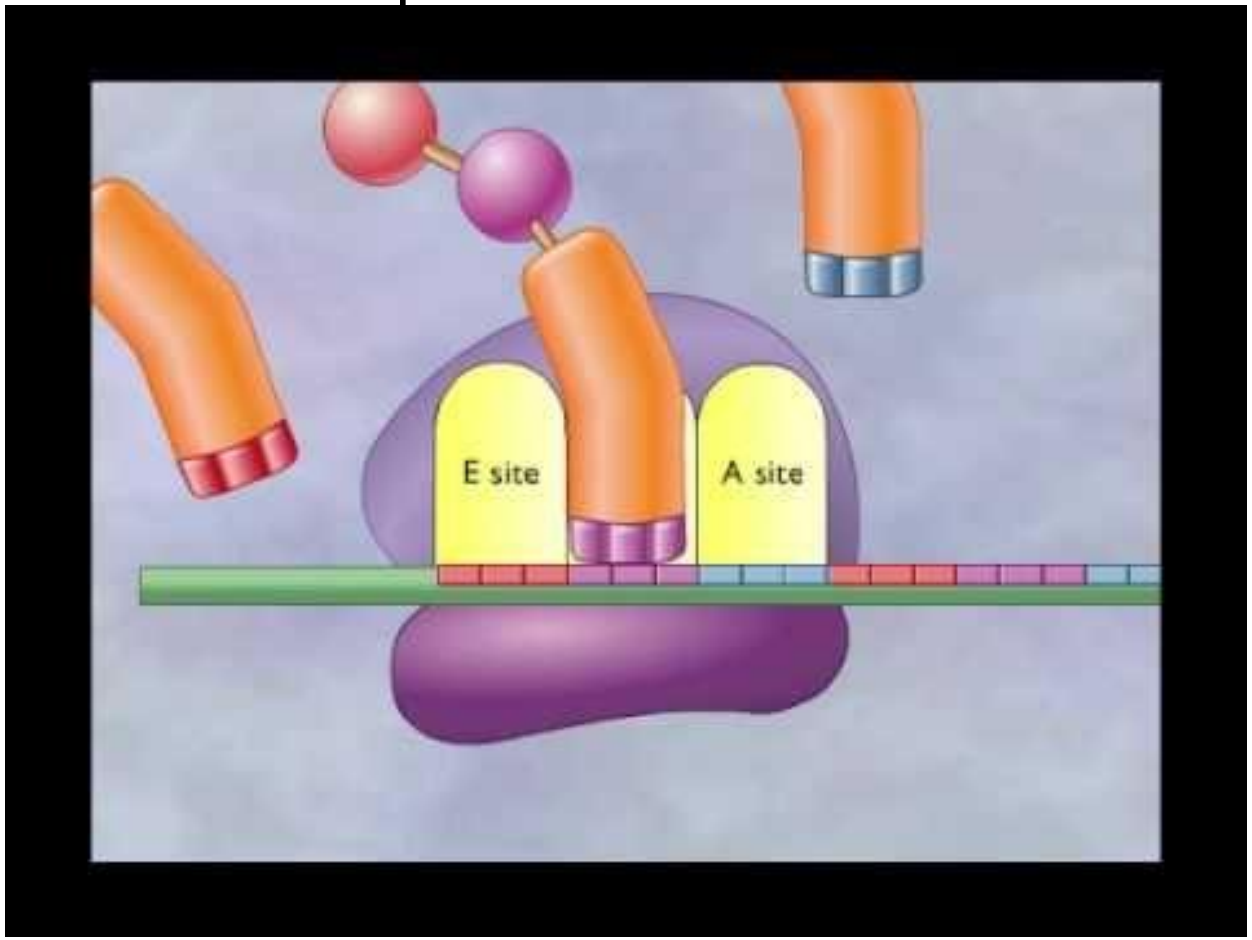
http://mol-biol4masters.masters.grkraj.org/html/Protein_Synthesis4-Eukaryotic_Mechanism_of_Synthesis.htm



https://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-22/22_07.jpg

➢Termination of translation:

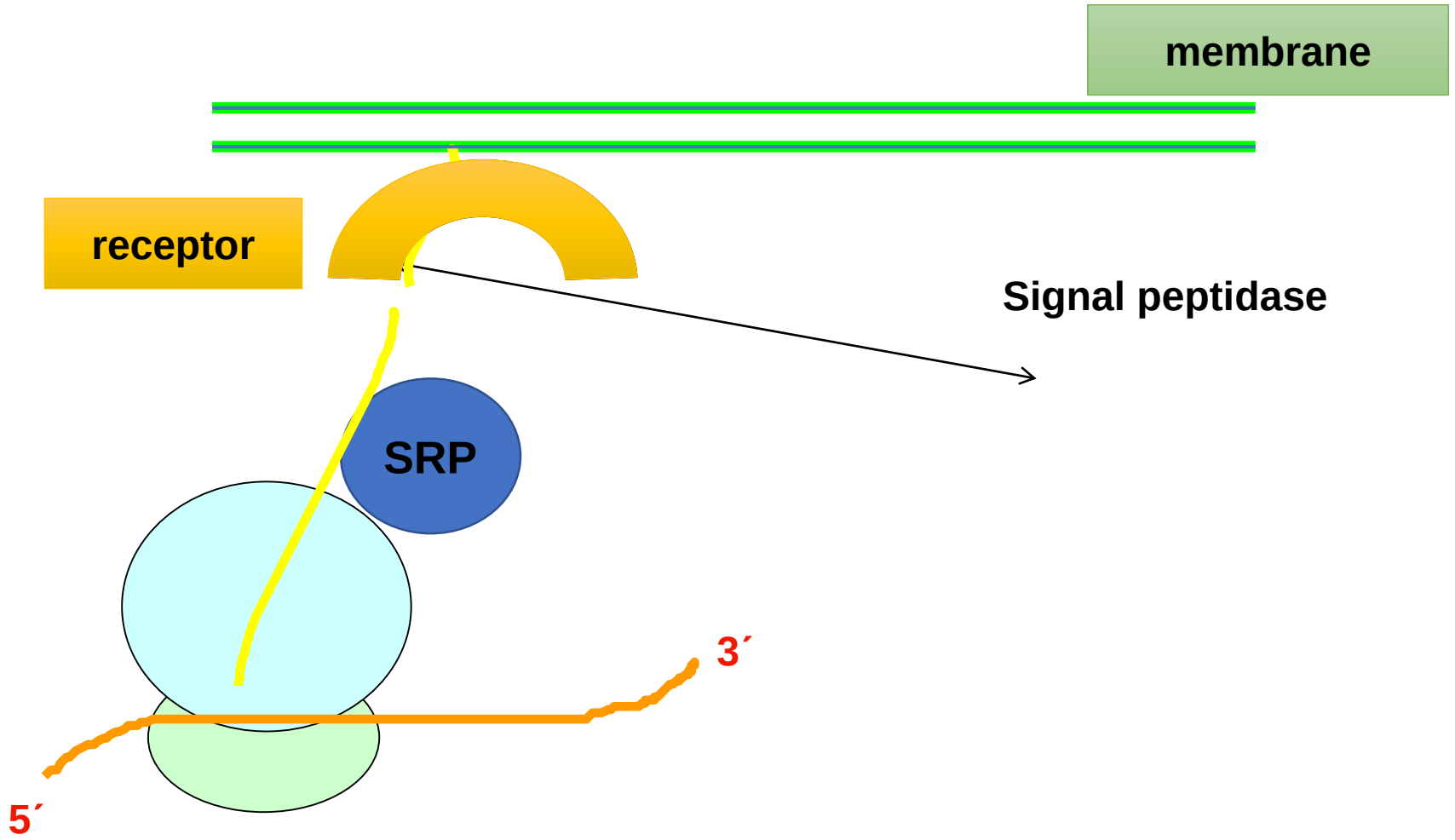
- ➢Eukaryotic translation termination factor eRF1 (release factor) recognizes all 3 stop codons
- ➢Energy from GTP is required to release the ribosome from mRNA



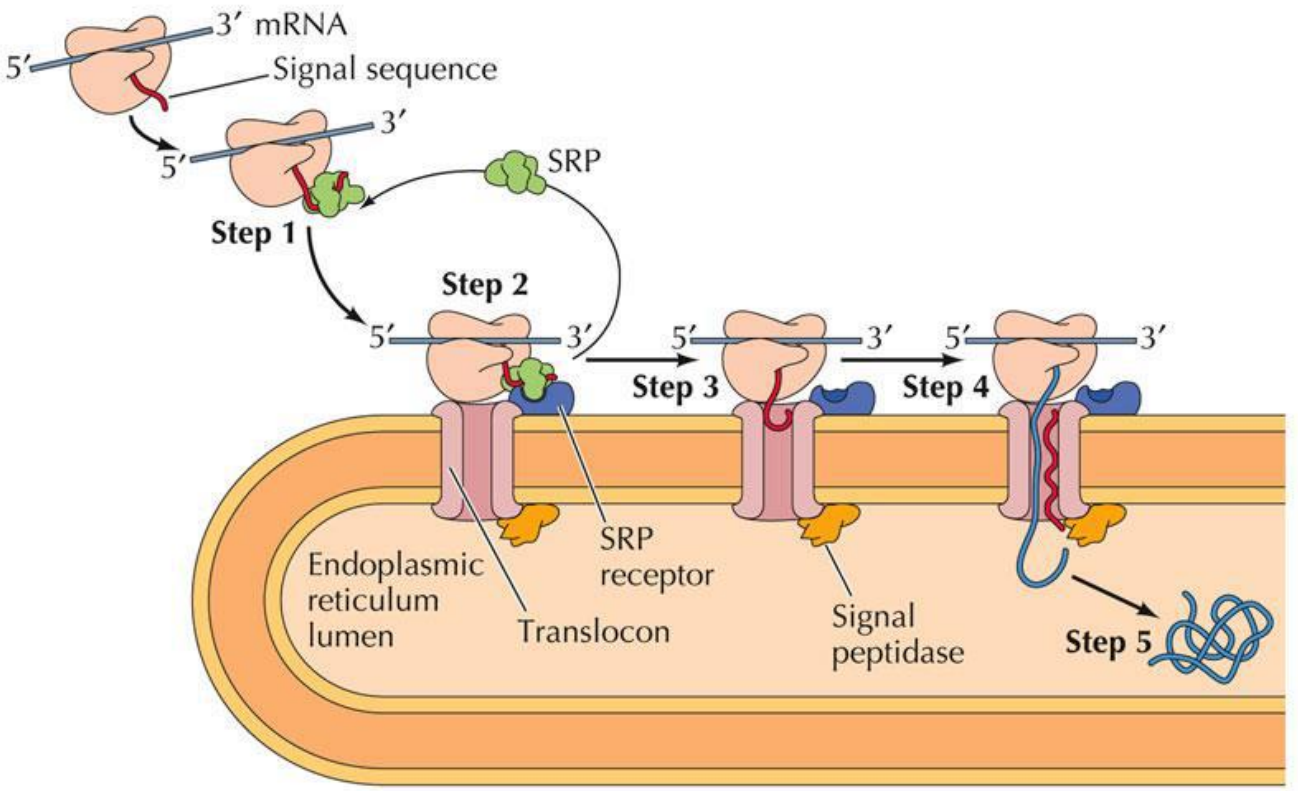
https://www.youtube.com/watch?v=C4QiMqBSDe4

- Extracellular and membrane proteins:
  - all extracellular and membrane proteins have a 15-25 AA signal peptide at the N-terminus
  - signal peptide binds to the signal recognition particle (SRP)
  - SRP stops translation on the ribosome
  - binding of SRP to the receptor on the membrane leads to cleavage of the signal peptide by signal peptidase and the translation continues

➢Extracellular and membrane proteins:



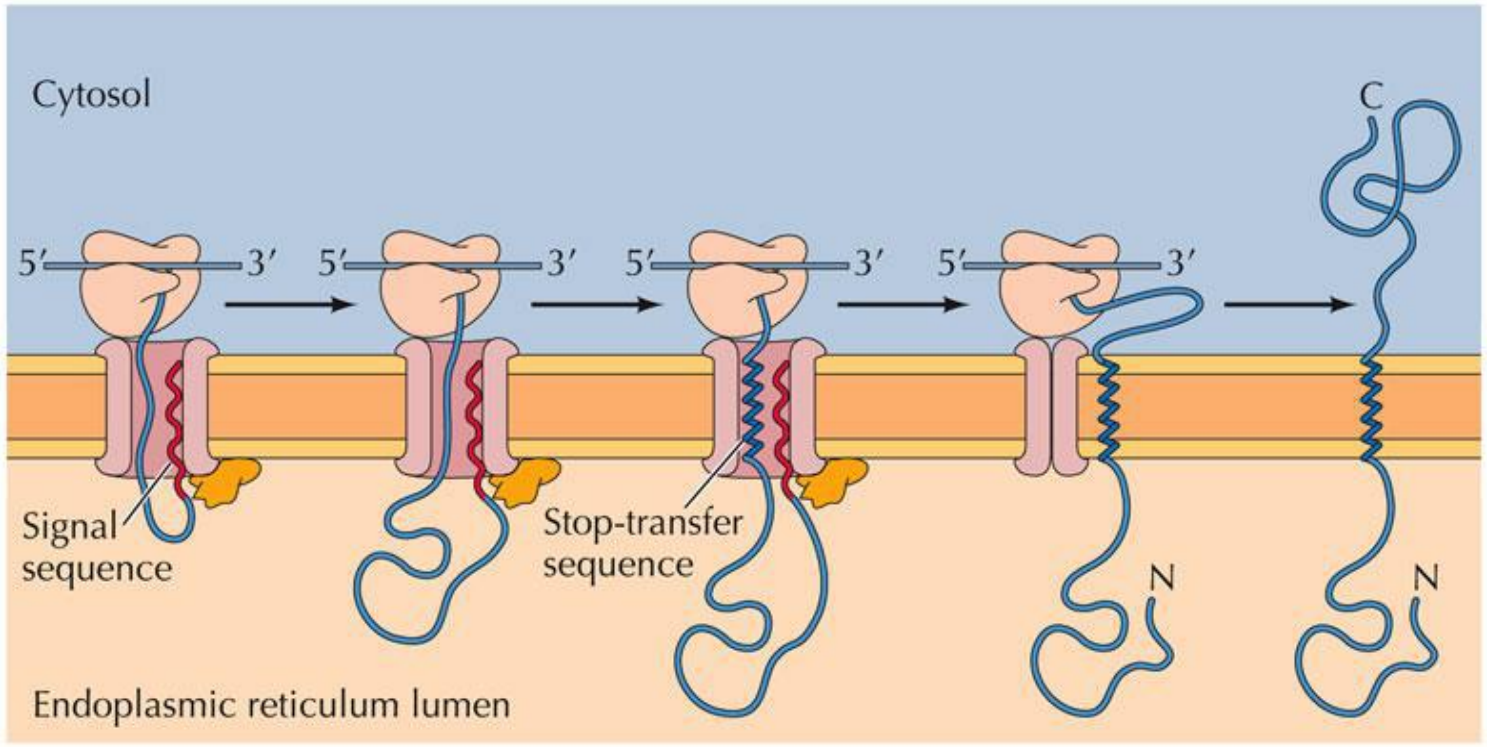
➢Translocation of extracellular proteins:



THE CELL, Fourth Edition, Figure 10.8 © 2006 ASM Press and Sinauer Associates, Inc.

The Cell, Fourt edition, Figure 10.8. 2006

➢Formation of membrane proteins:

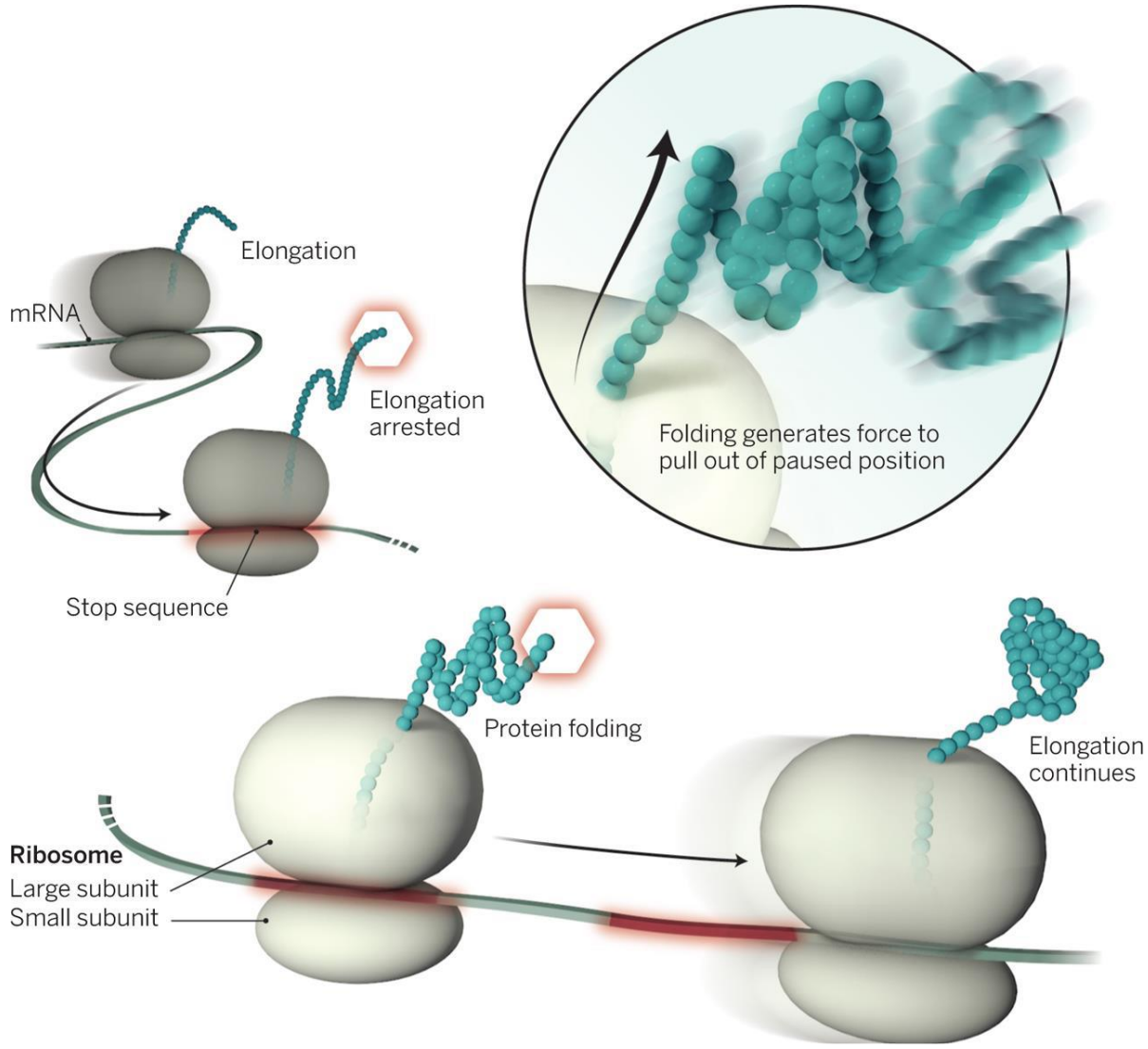


THE CELL, Fourth Edition, Figure 10.12 © 2006 ASM Press and Sinauer Associates, Inc.

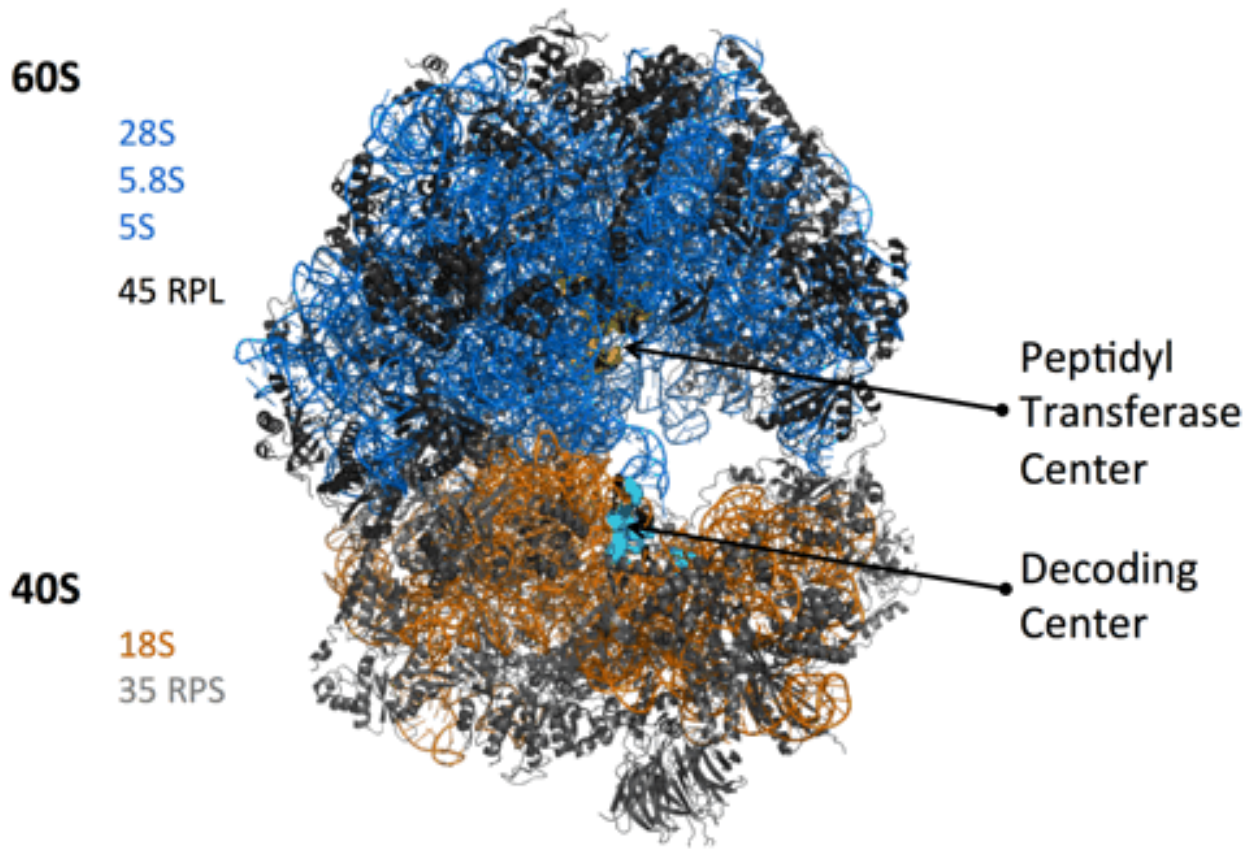
The Cell, Fourth edition, Figure 10.12, 2006

➢Translation and folding:

- ➢In some proteins, the ribosome shift briefly stops following synthesis of short oligopeptide
- ➢This will allow precise targeting of the protein to the exit tunnel in the large subunit
- ➢Precise targeting of the protein to the tunnel is associated with proper folding



Joseph D. Puglisi, Science 2015, 348,399-400



http://www.crcl.fr/311-Projets-en-GB.crcl.aspx?language=en-GB