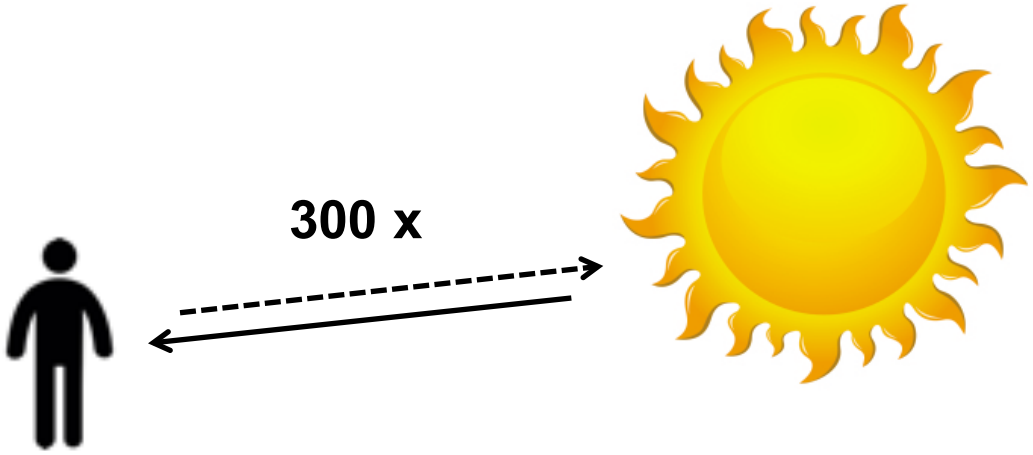
Corso di Biologia Molecolare per Fisica
5 Aprile 2016

Monica Ballarino

Chromatin structure and Transcription

Each diploid cell contains about 2 meters of DNA

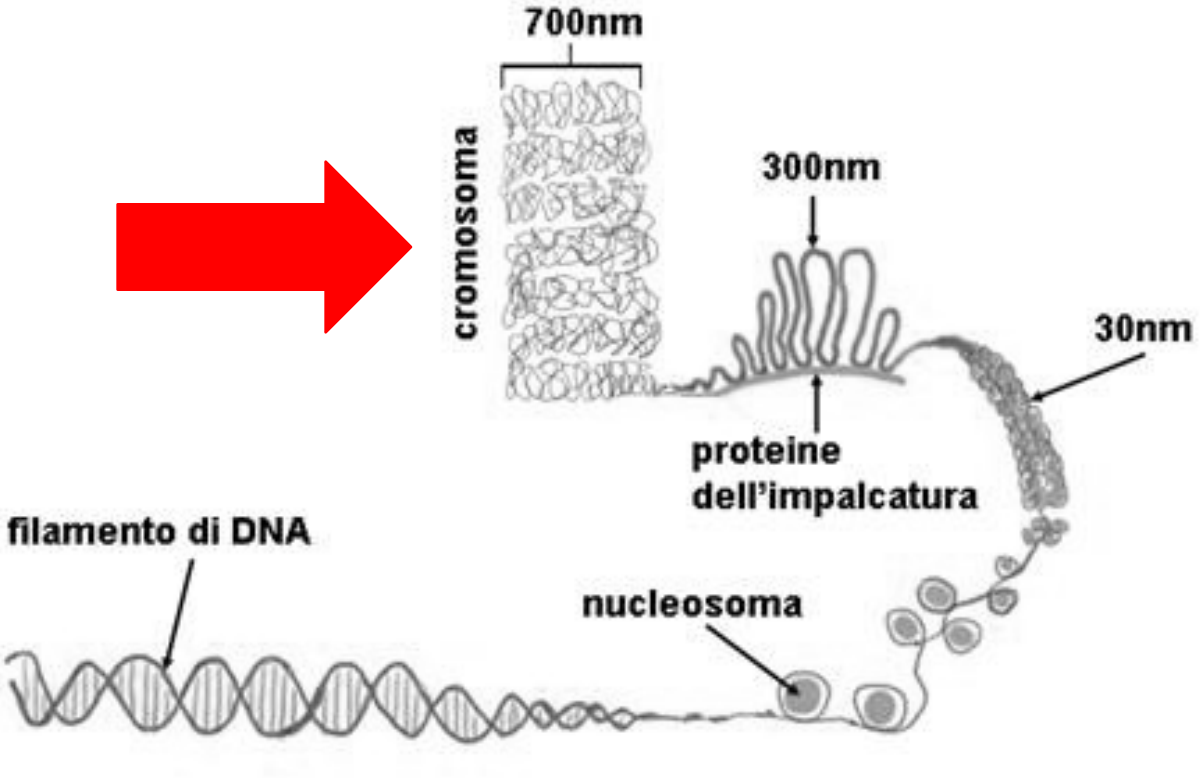
Each of us has enough DNA to go from here to the Sun and back more than 300 times, or around Earth's equator 2.5 million times!



How is this possible?

The answer to this question lies in the fact that certain proteins compact chromosomal DNA into the microscopic space of the eukaryotic nucleus. These proteins are called ***histones***, and the resulting DNA-protein complex is called ***chromatin***.

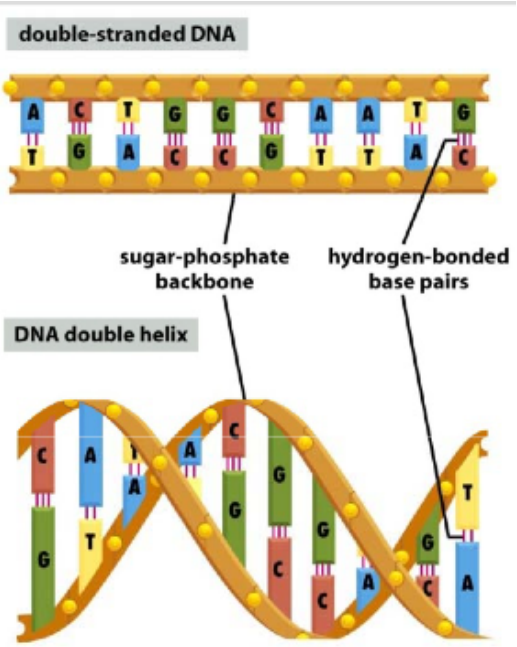
Packaging is essential for several reasons:



- without such packaging, DNA molecules would be too long to fit inside cells
- *damage* protection
- during cell division, it is essential that DNA remains intact and evenly distributed among cells. Chromosomes are a key part of the process that ensures DNA is accurately *copied* and distributed in the vast majority of cell divisions.

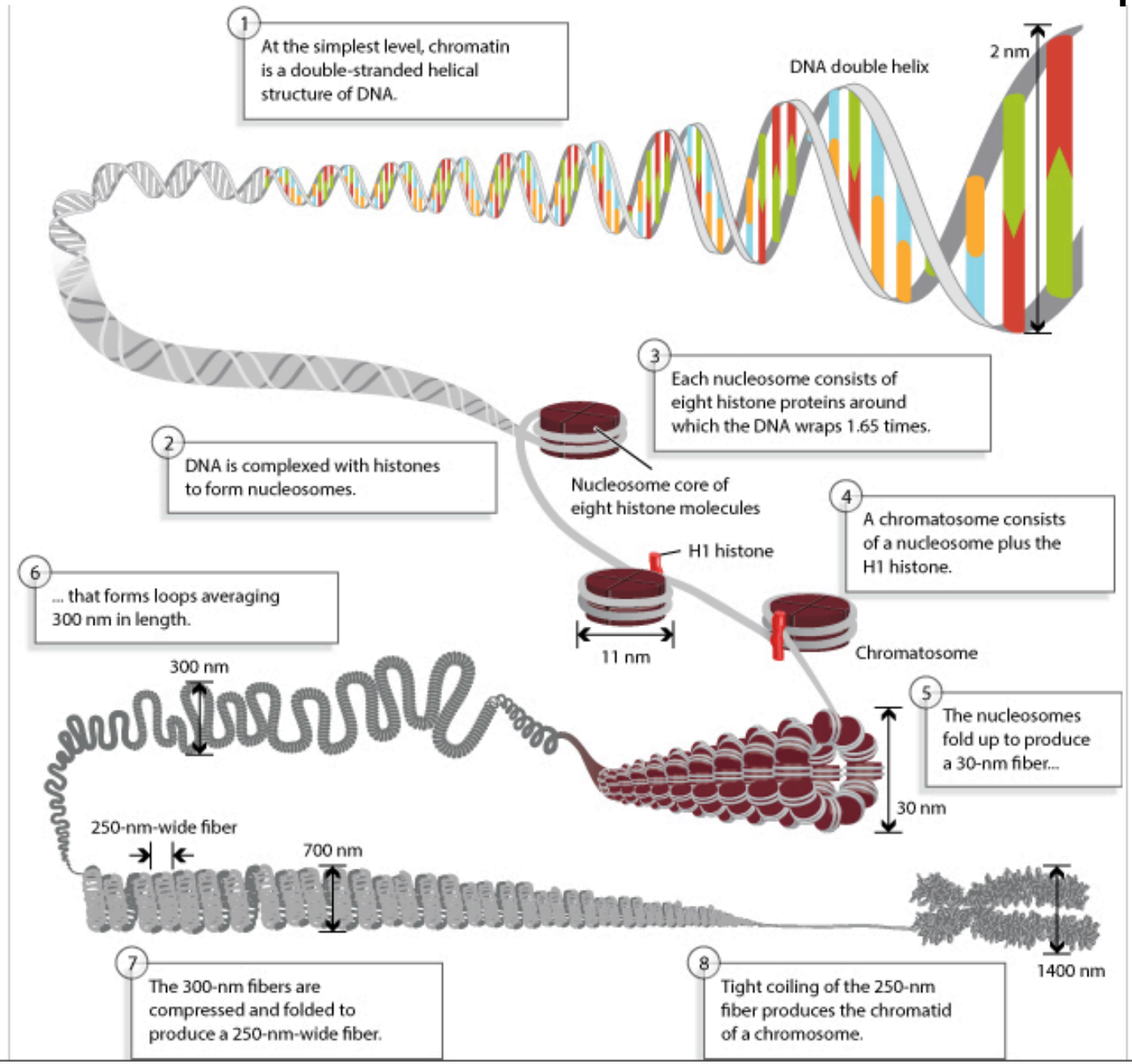
The problem: DNA lenght is always higher than the dimension of the compartment in which it is stored

- Spacing between base pairs ≈3.4Å
- For human genome, approximately 3.2 billion base pairs
- Total length ≈ $3.4\times10^{-10}\times3.2\times10^{9}\times2 \approx 2.2$m
- Diameter of a nucleus: $5\sim10\times10^{-6}$m



Compartment	Shape	Dimensions	Type of nucleic acid	Length
TMV	filament	0.008 × 0.3 μm	1 single-stranded RNA	2 μm = 6.4 kb
Phage ϕd	filament	0.006 × 0.85 μm	1 single-stranded DNA	2 μm = 6.0 kb
Adenovirus	icosahedron	0.07 μm diameter	1 double-stranded DNA	11 μm = 35.0 kb
Phage T4	icosahedron	0.065 × 0.10 μm	1 double-stranded DNA	55 μm = 170.0 kb
E. coli	cylinder	1.7 × 0.65 μm	1 double-stranded DNA	1.3 μm = 4.2×10^3 kb
Mitochondrion (human)	oblate spheroid	3.0 × 0.5 μm	~10 identical double-stranded DNAs	50 μm = 16.0 kb
Nucleus (human)	spheroid	6 μm diameter	46 chromosomes of double-stranded DNA	1.8 m = 6×10^6 kb

Packing ratio



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

$1{,}2 \times 10^4$

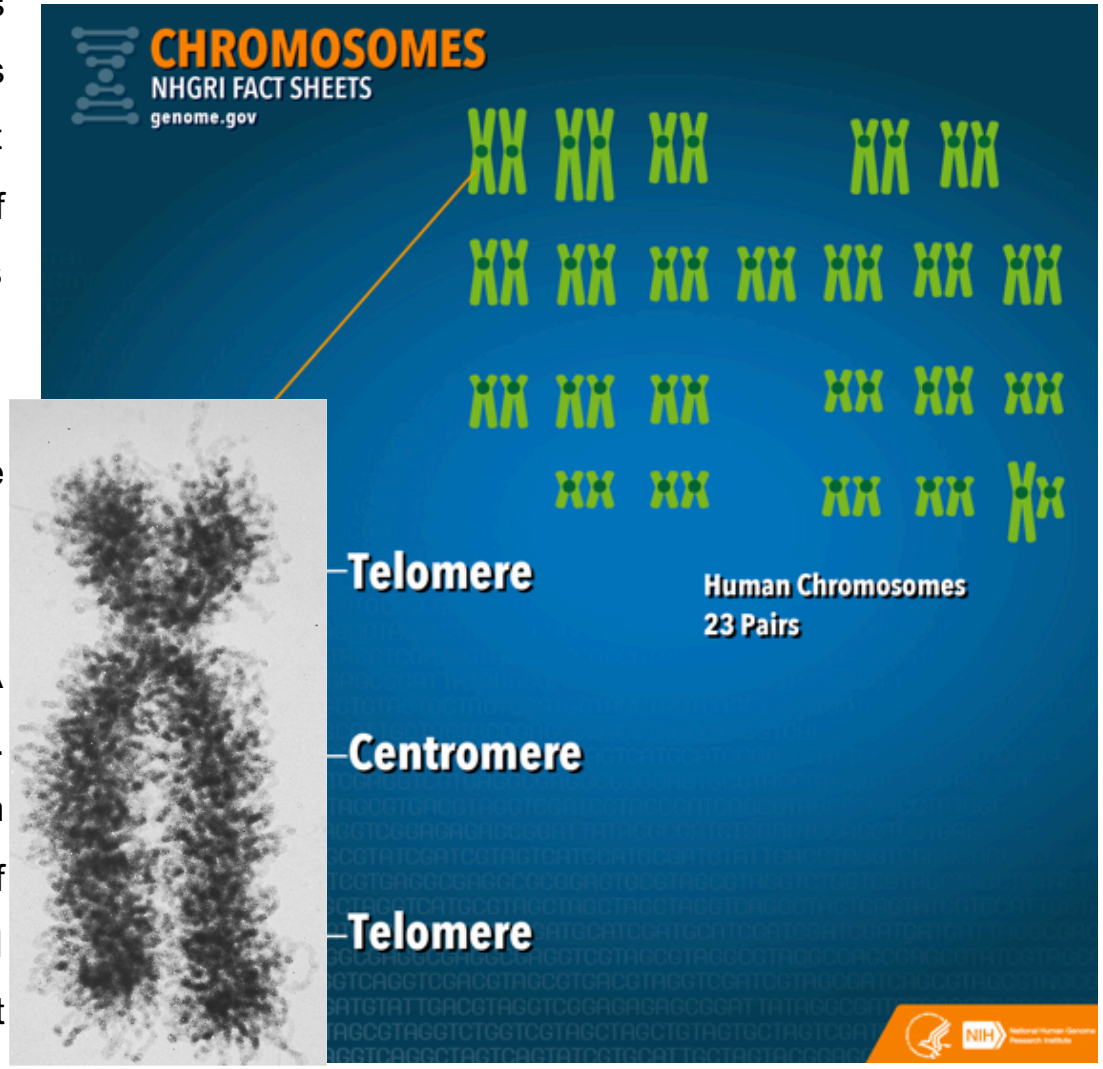
The packaging of the genetic material in Eukariotes

A ***chromosome*** is a packaged and organized structure. Each chromosome is made up of DNA tightly coiled many times around proteins called *histones* that support its structure.

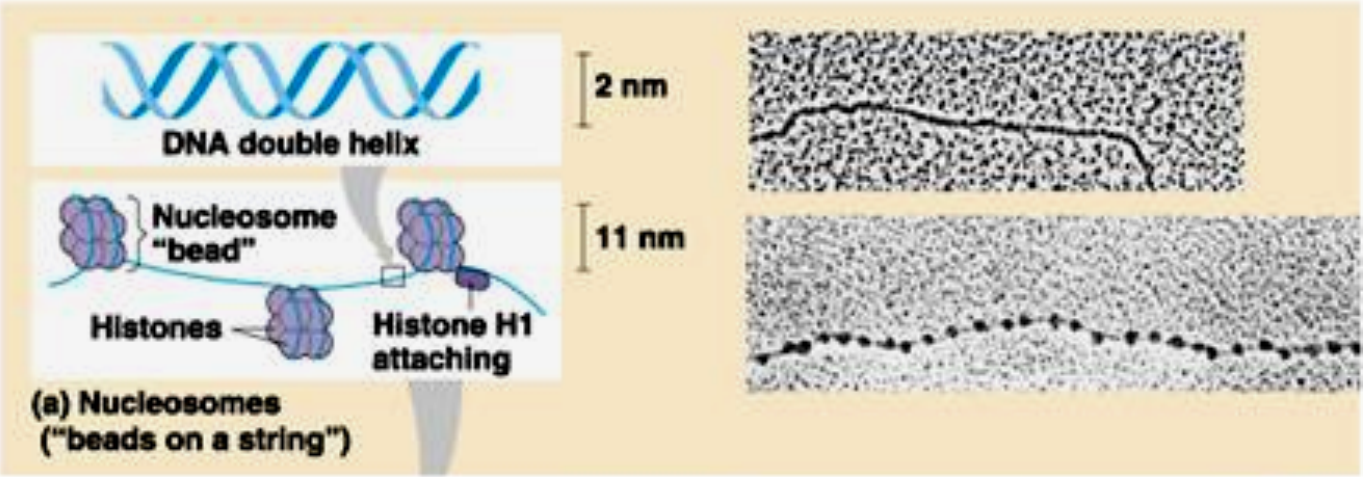
The constricted region of linear chromosomes is known as the ***centromere***. Although this constriction is called the centromere, it usually is not located exactly in the center of the chromosome and, in some cases, is located almost at the chromosome's end.

The regions on either side of the centromere are referred to as the chromosome's ***arms***.

Telomeres are repetitive stretches of DNA located at the ends of linear chromosomes. They protect the ends of chromosomes. In many types of cells, telomeres lose a bit of their DNA every time a cell divides: when all of the telomere DNA is gone, the cell cannot replicate and dies.

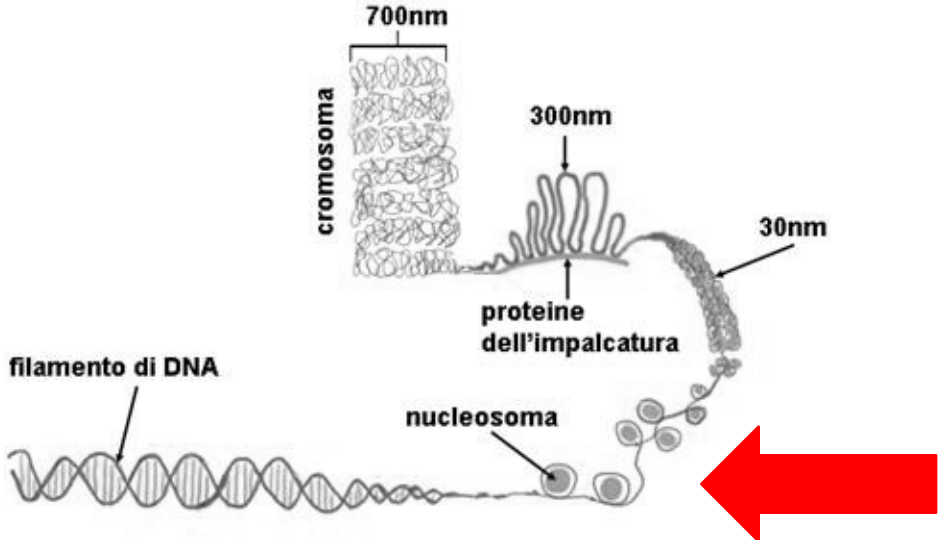


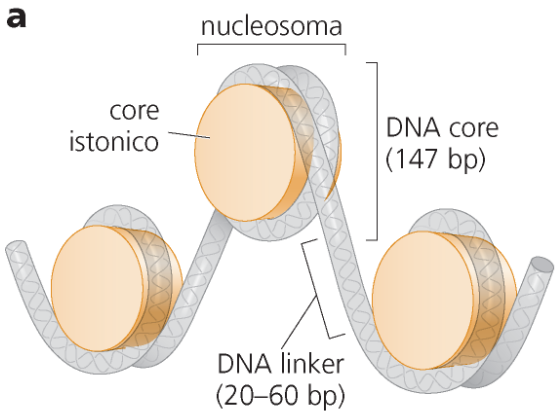
The 10 nm fiber



The observation by electron microscopists that chromatin appeared similar to **beads** on a string

The basic repeating structural (and functional) unit of chromatin is the ***nucleosome***, which contains eight histone proteins and about 146 base pairs of DNA (Van Holde, 1988; Wolffe, 1999).

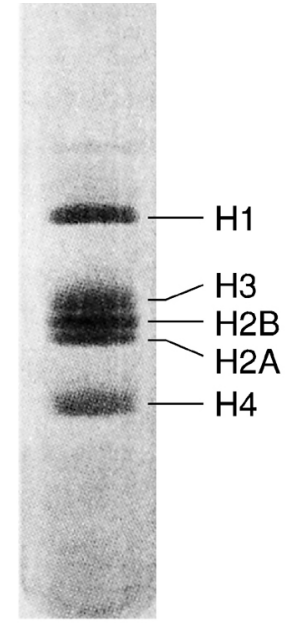




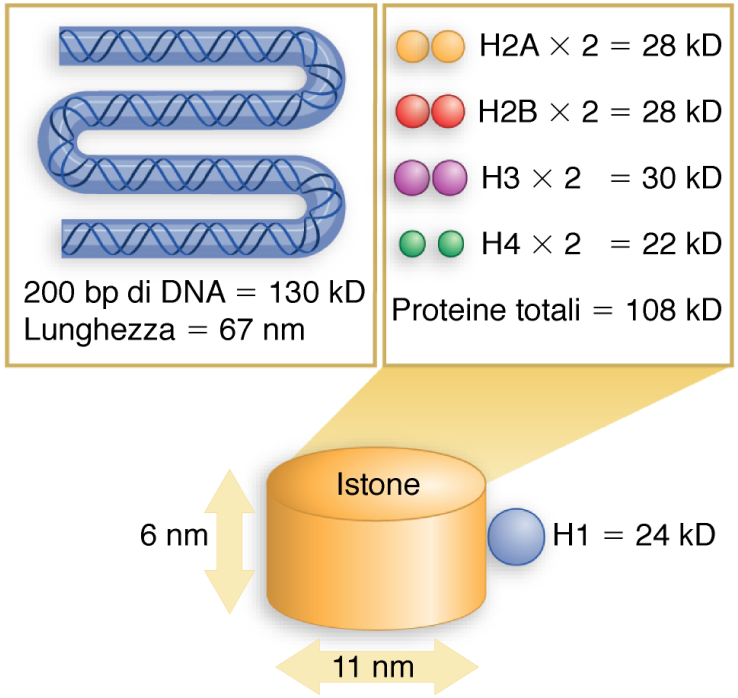
Nucleosomes are structured as follows:

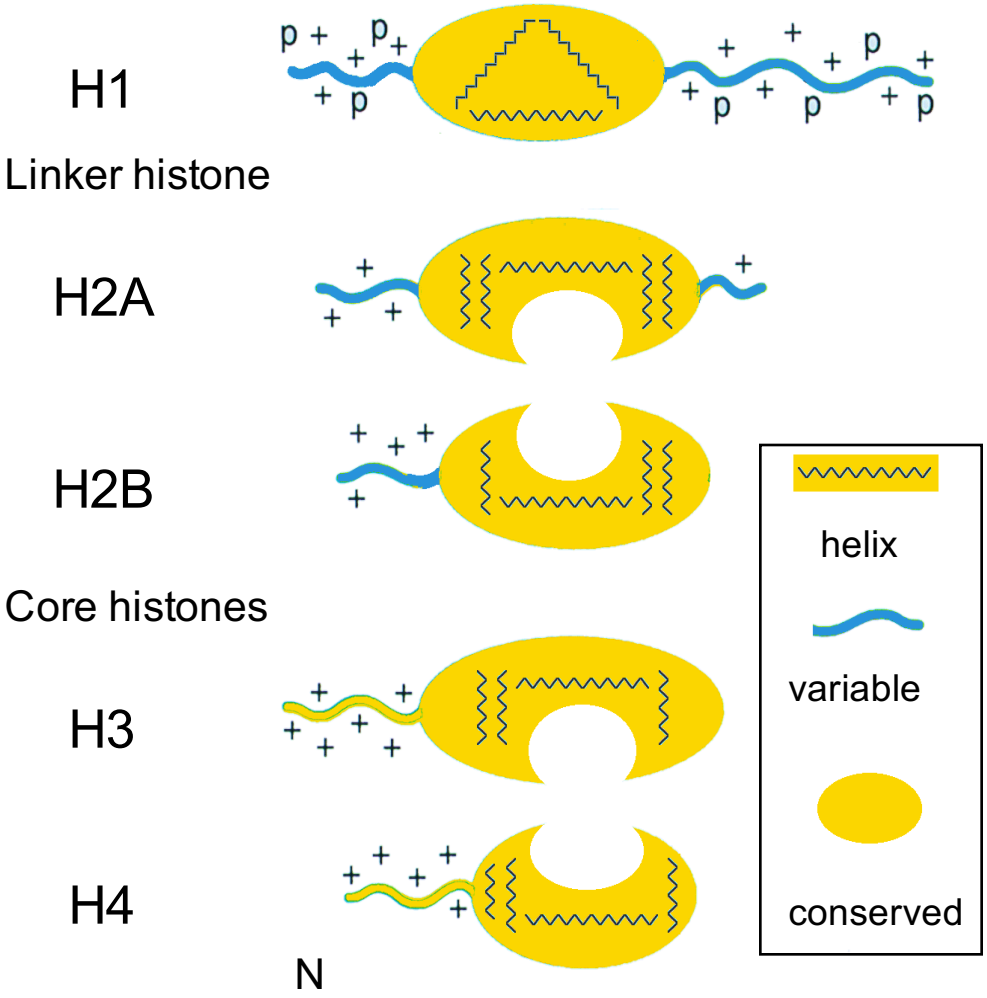
- two each of the histones H2A, H2B, H3, and H4 come together to form a histone octamer, which binds and wraps approximately 1.7 turns of DNA, or about 146 base pairs.

- the addition of one H1 protein wraps another 20 base pairs, resulting in two full turns around the octamer, and forming a structure called a chromatosome. This joining DNA is referred to as *linker* DNA.



chem. & Biophys. 130, 1969, f. 6A, p. 343. Reprinted with

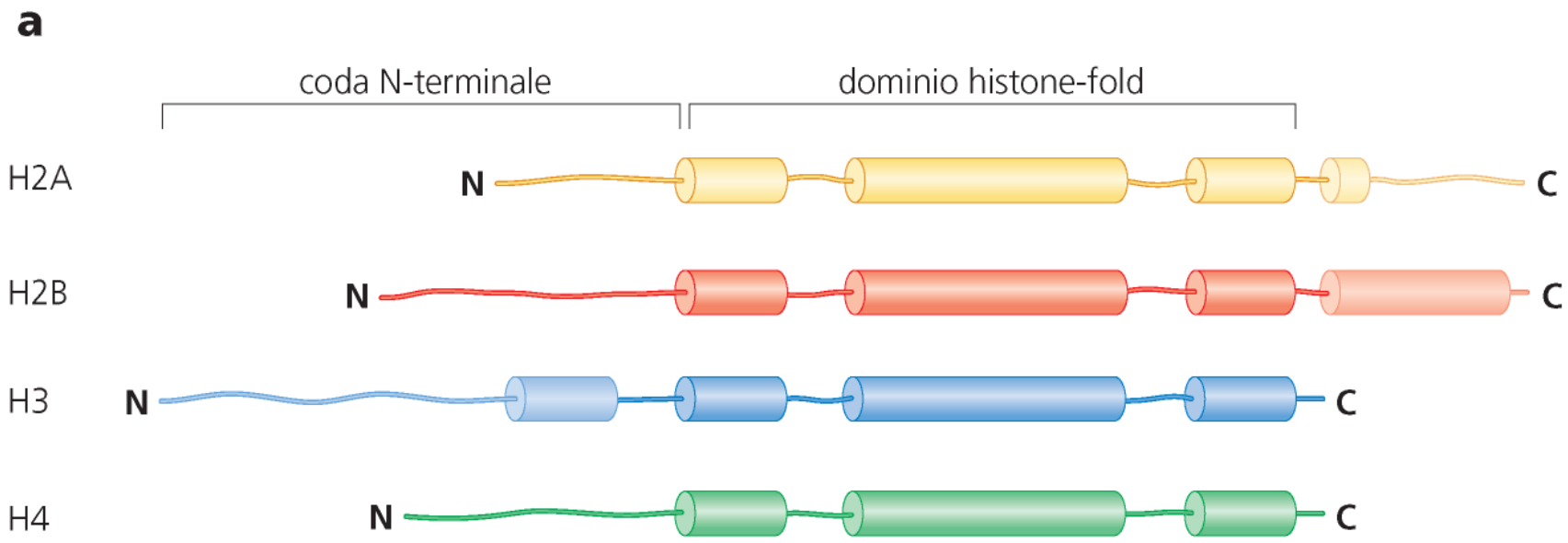




Histones are a family of *small*, *positively* charged proteins termed **H1**, **H2A**, **H2B**, **H3**, and **H4** (Van Holde, 1988).

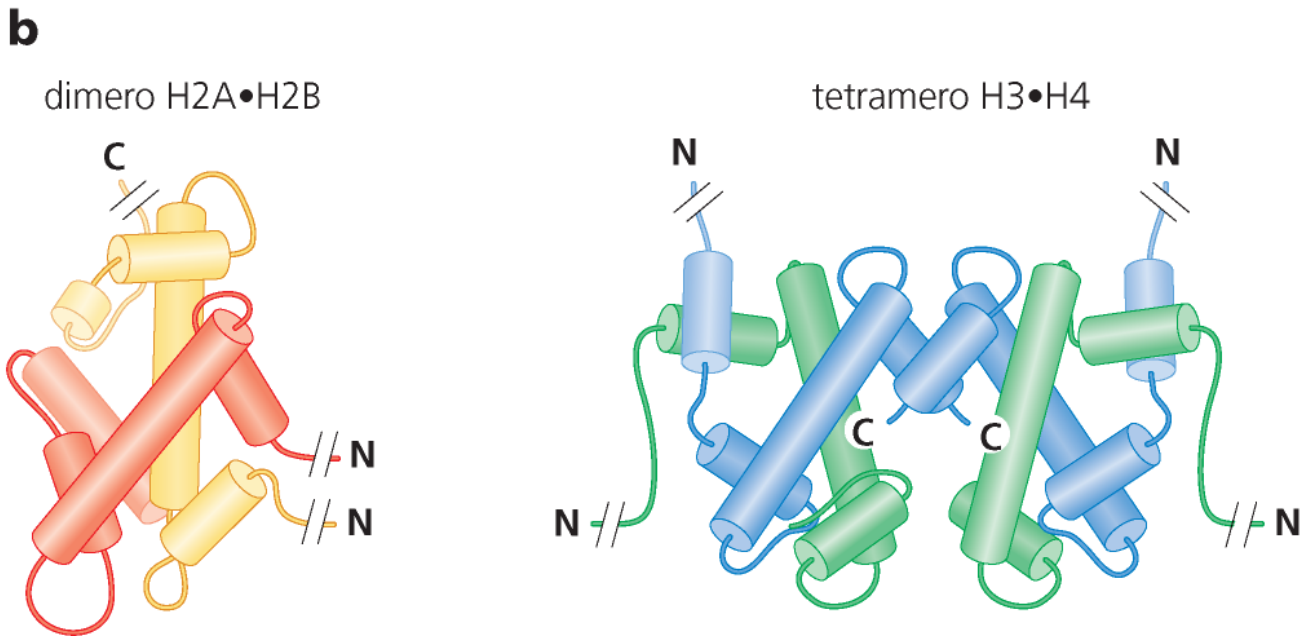
DNA is negatively charged, due to the phosphate groups in its phosphate-sugar backbone, so histones bind with DNA very tightly.

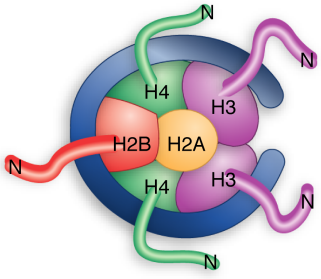
Histone Type	Molecular Weight	Number of Amino Acids	Approx. Content of Basic Amino Acids
H1	17,000–28,000	200–265	27% lysine, 2% arginine
H2A	13,900	129–155	11% lysine, 9% arginine
H2B	13,800	121–148	16% lysine, 6% arginine
H3	15,300	135	10% lysine, 15% arginine
H4	11,300	102	11% lysine, 4% arginine



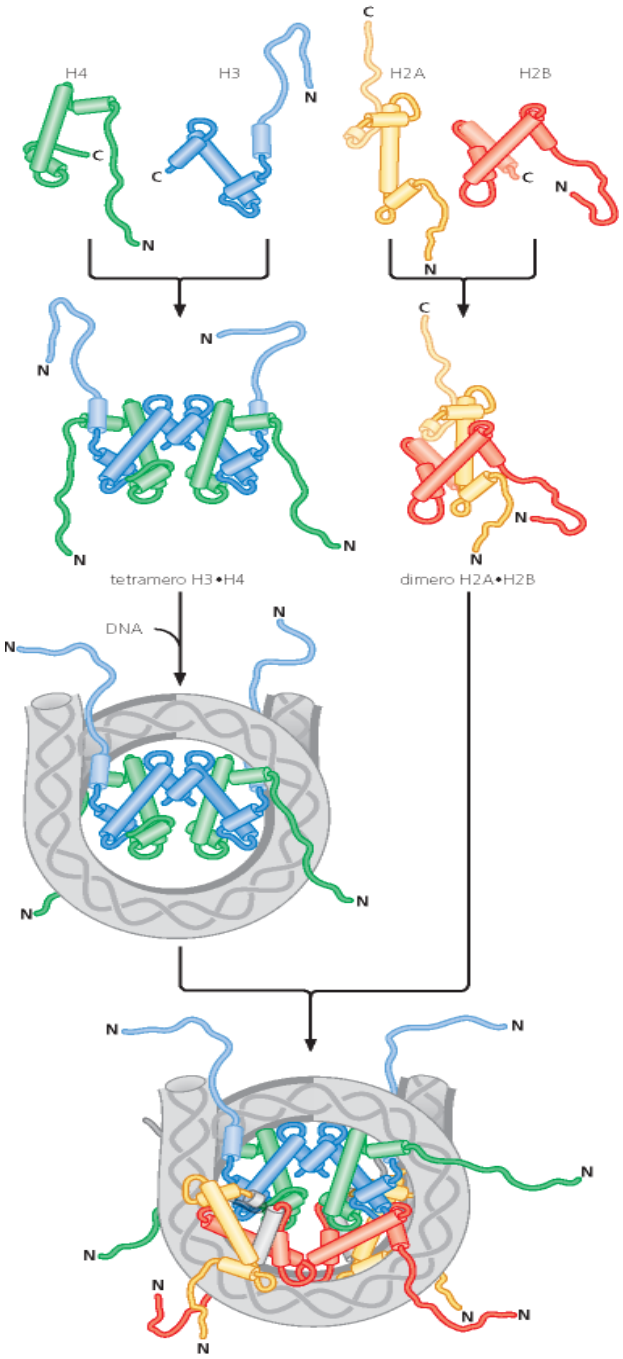
The *histone fold* domain helps the formation of the globular core.

The *N-term tails* protrude outside and interact with DNA





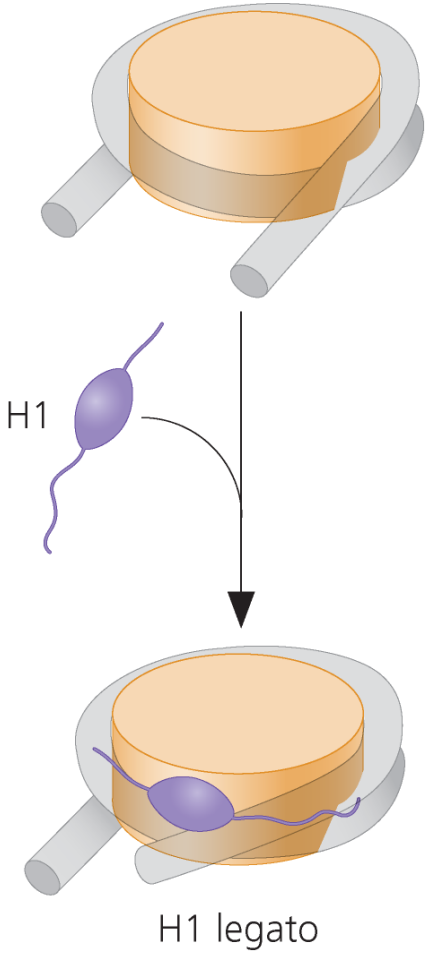
Nucleosome assembly

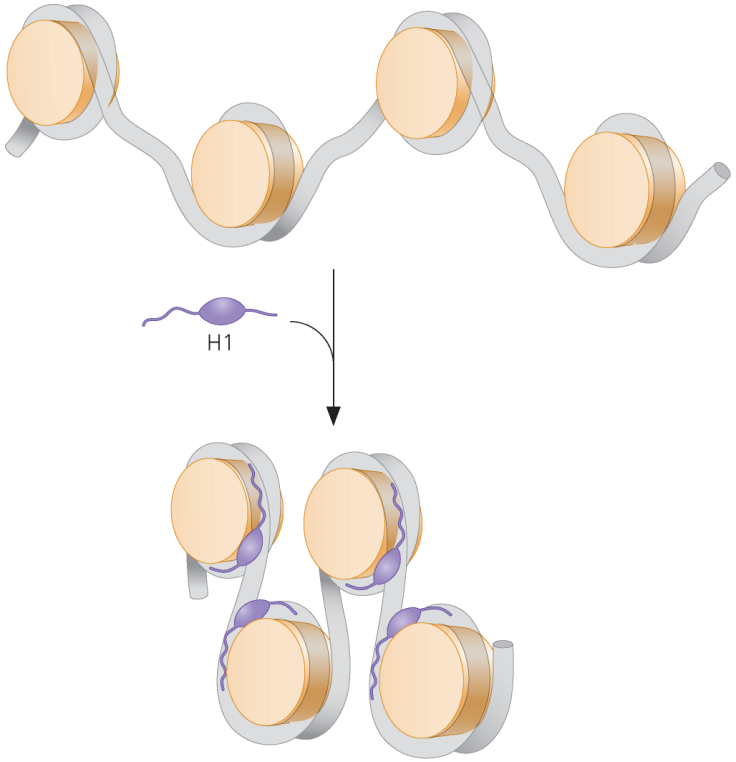


1) Tetramers H3/H4 and dimers H2A/H2B form indipendentently

2) H3/H4 tetramer binds DNA

3) H2A/H2B dimers are recruited at the end





Accorciamento della lunghezza del DNA associato al nucleosoma

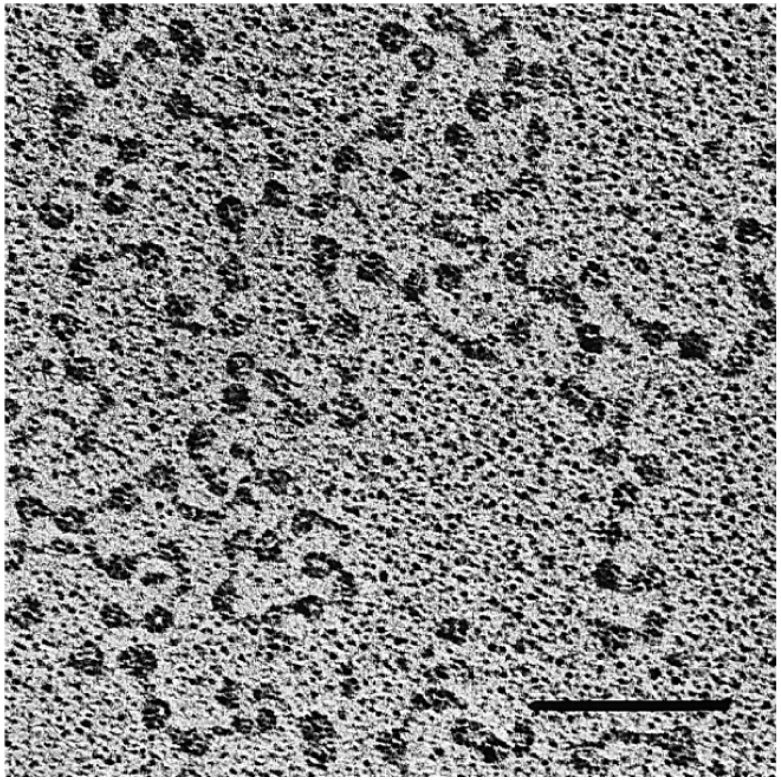
H1 is outside the nucleosome and binds DNA at the level of entry/exit points

H1 contributes to the formation of higher ordered structures: the 30 nm fiber.

Effect of H1 absence on nucleosomes packaging (Electron Microscopy)

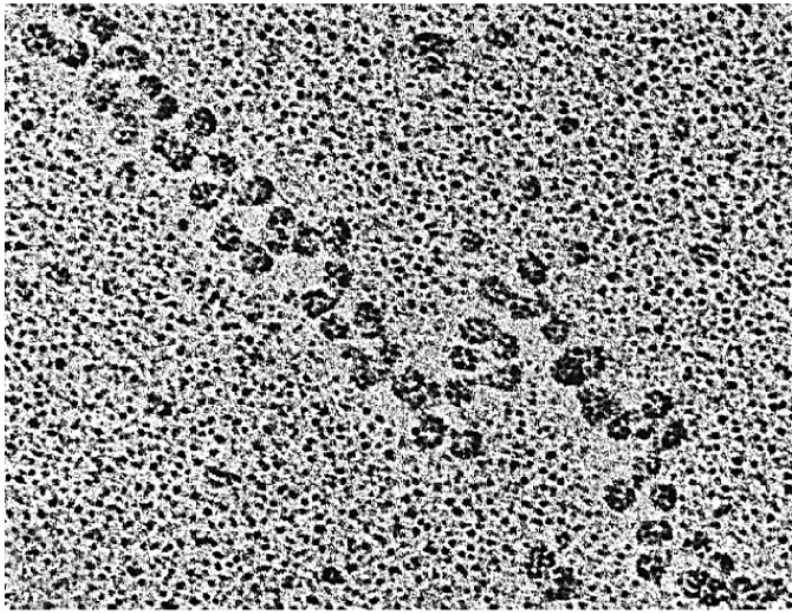
Without di H1

a

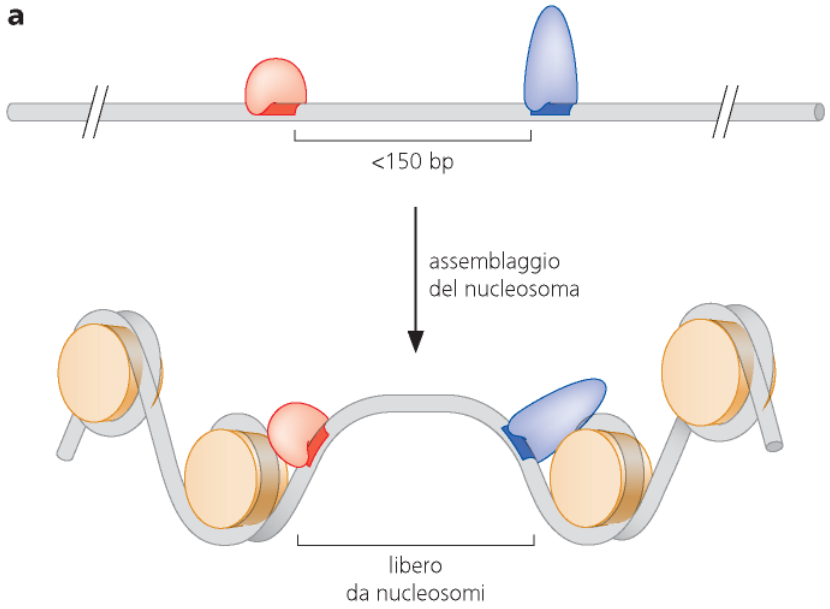


With H1

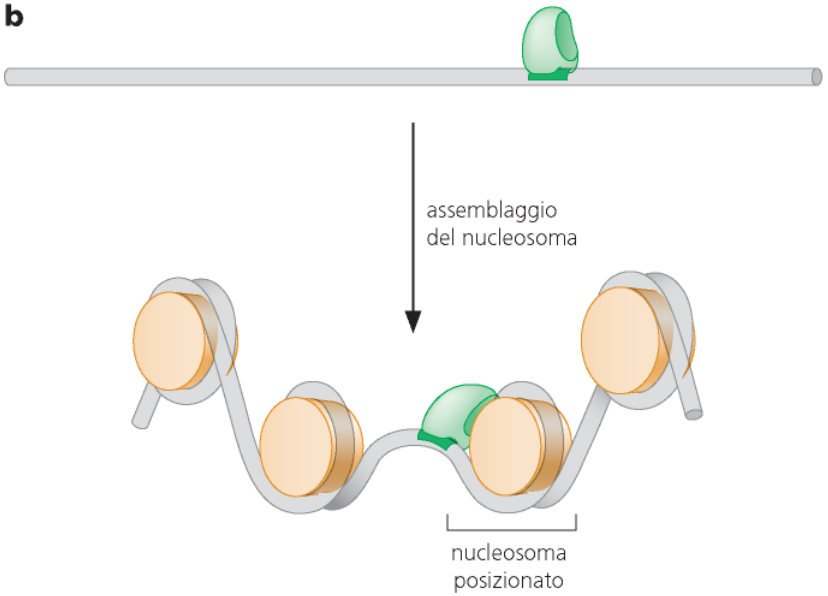
b



Proteins can assist nucleosomes positioning

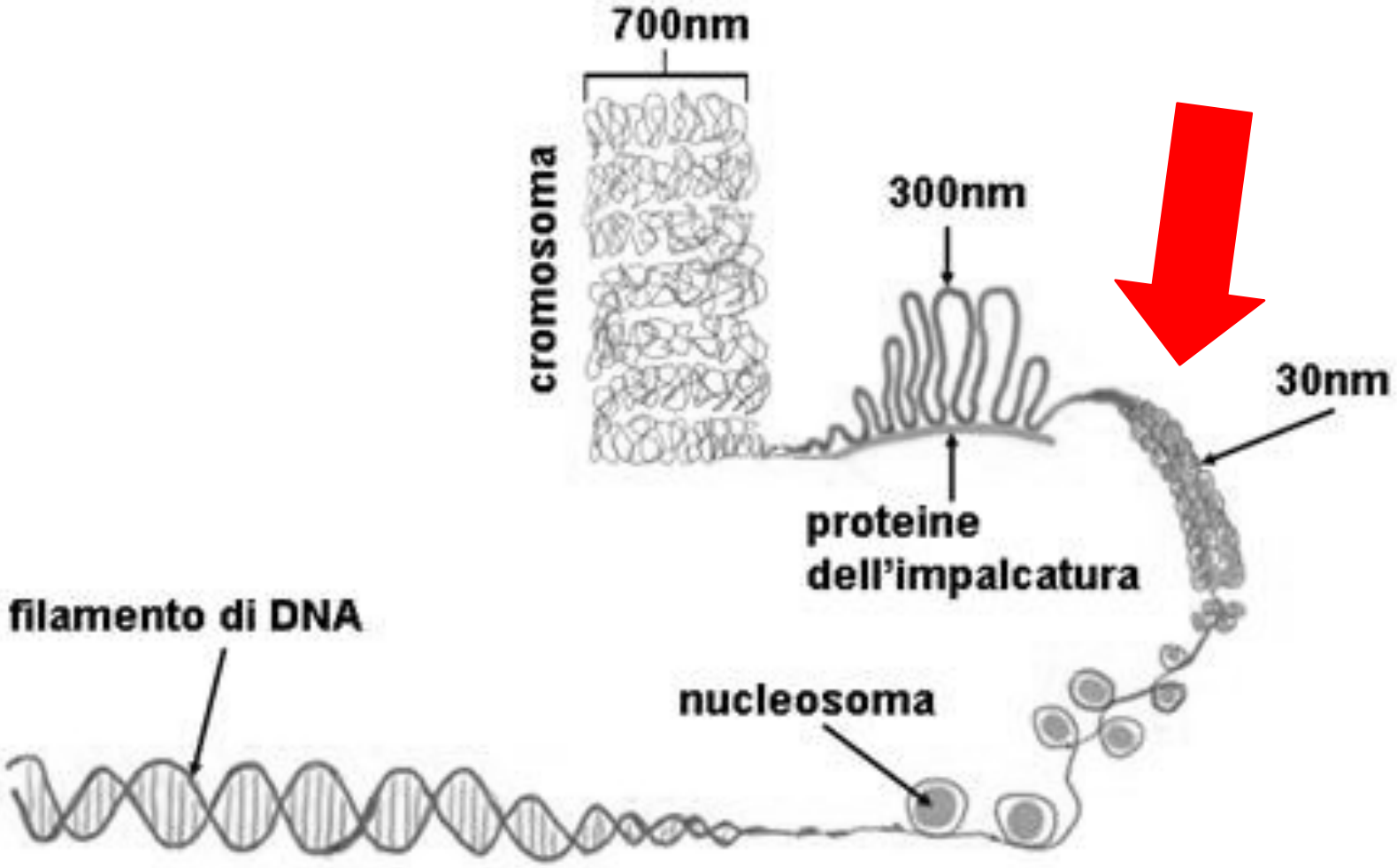


..to form *nucleosome free* regions

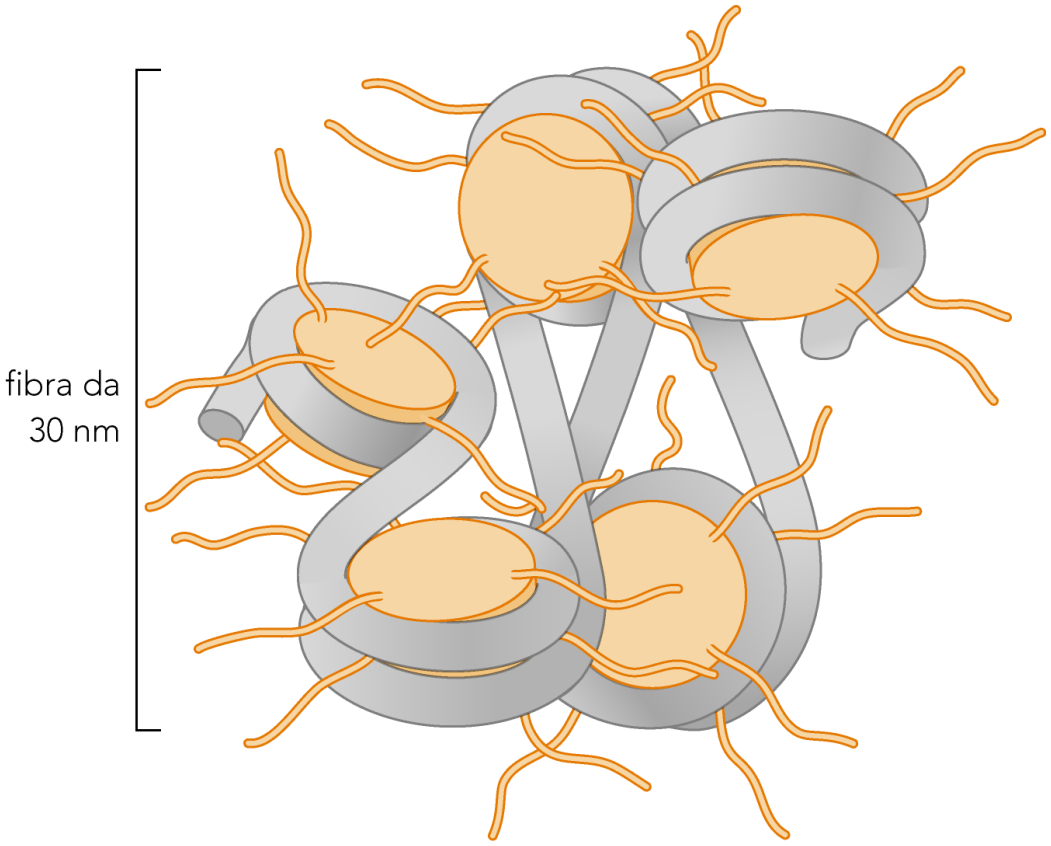


…to guide nucleosomes in specific positions

The 30 nm fiber



Histone tails play a role in the formation of the 30 nm fiber

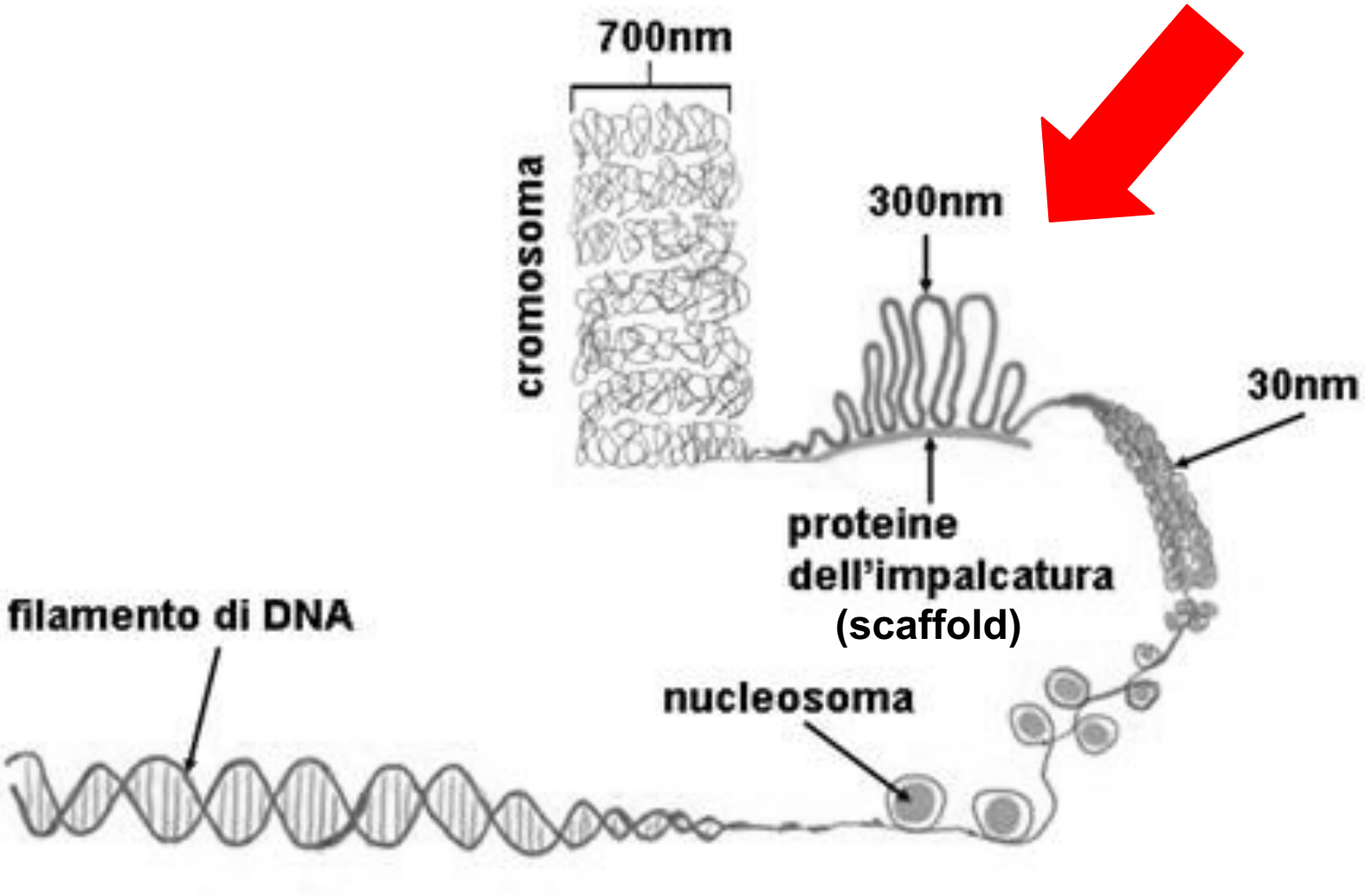


Istoni, privati di coda N-terminale, sono *incapaci* di formare la fibra 30 nm.

Nel cristallo si osservano legami idrogeno fra la coda di H4 e parti di H2A-H2B.

E' probabile che le code istoniche, in quanto frequentemente modificate, influenzino la formazione della fibra da 30 nm e altre strutture d'ordine superiore.

Higher-order structures: the chromatin loops



Chromatin is an highly dynamic structure

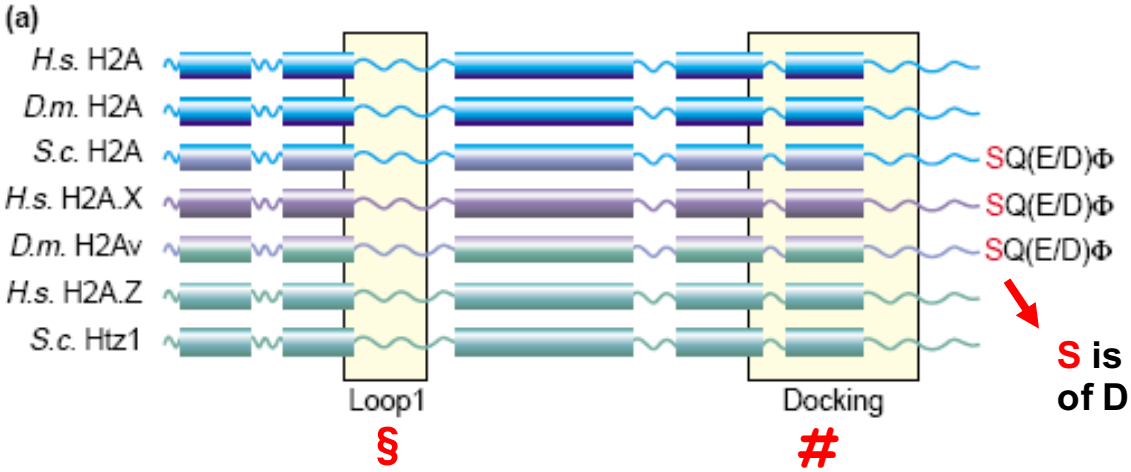
Different ways to modify the chromatin exists which regulates gene expression, DNA repair, replication and recombination.

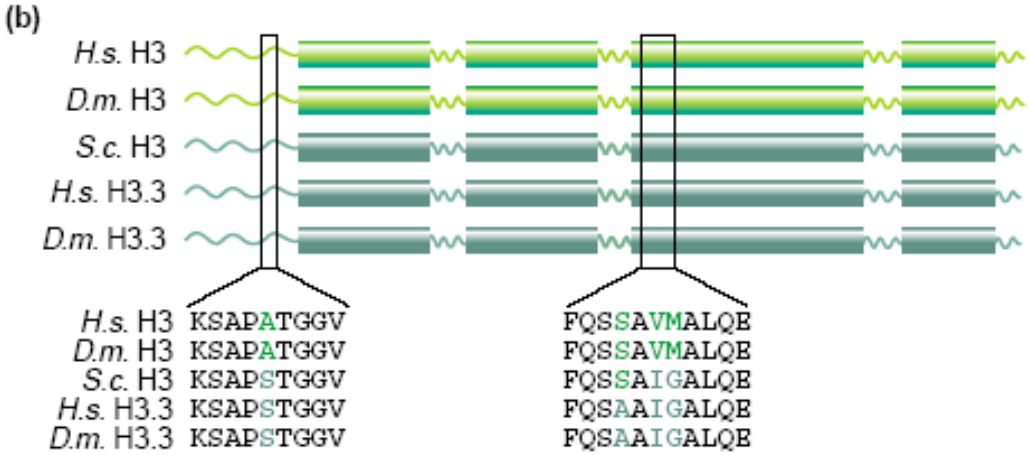
1) *Histone variants* (i.e. H2A e H3)

2) *Epigenetic post-translational histone modifications*: acetilation, methylation etc.…

3) Chromatin remodelling changes the stability and the position of nucleosomes

1) Histone variants



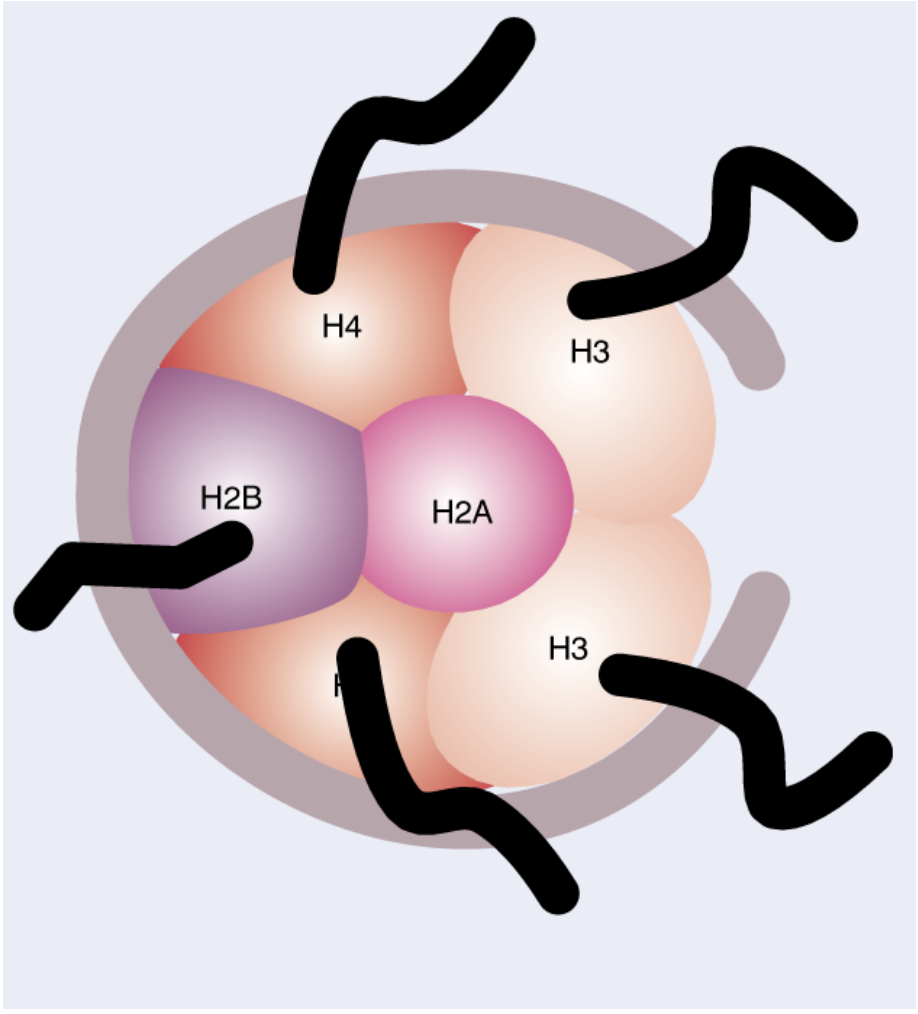


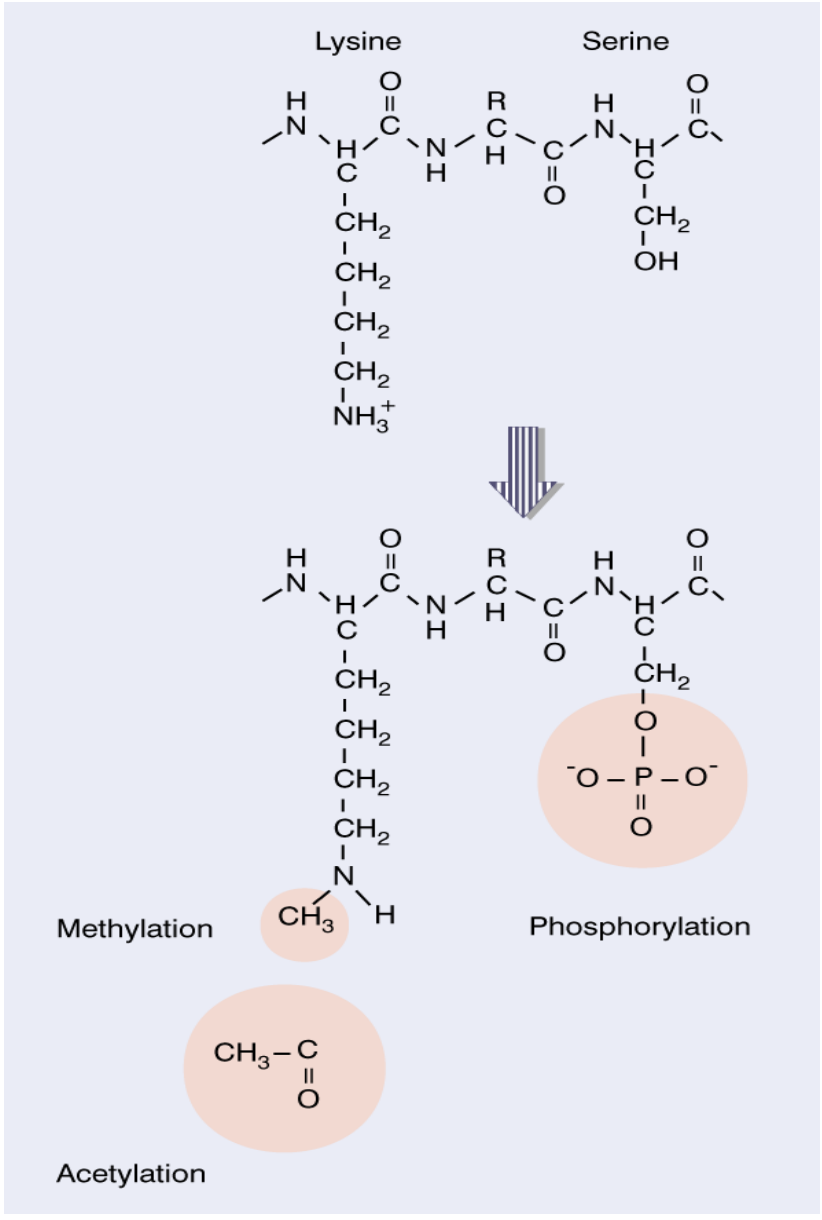
S is phosphorylated in consequence of DNA damage

§ : differences in this loop prevent the assembly of nucleosomes containing H2A e H2A.Z

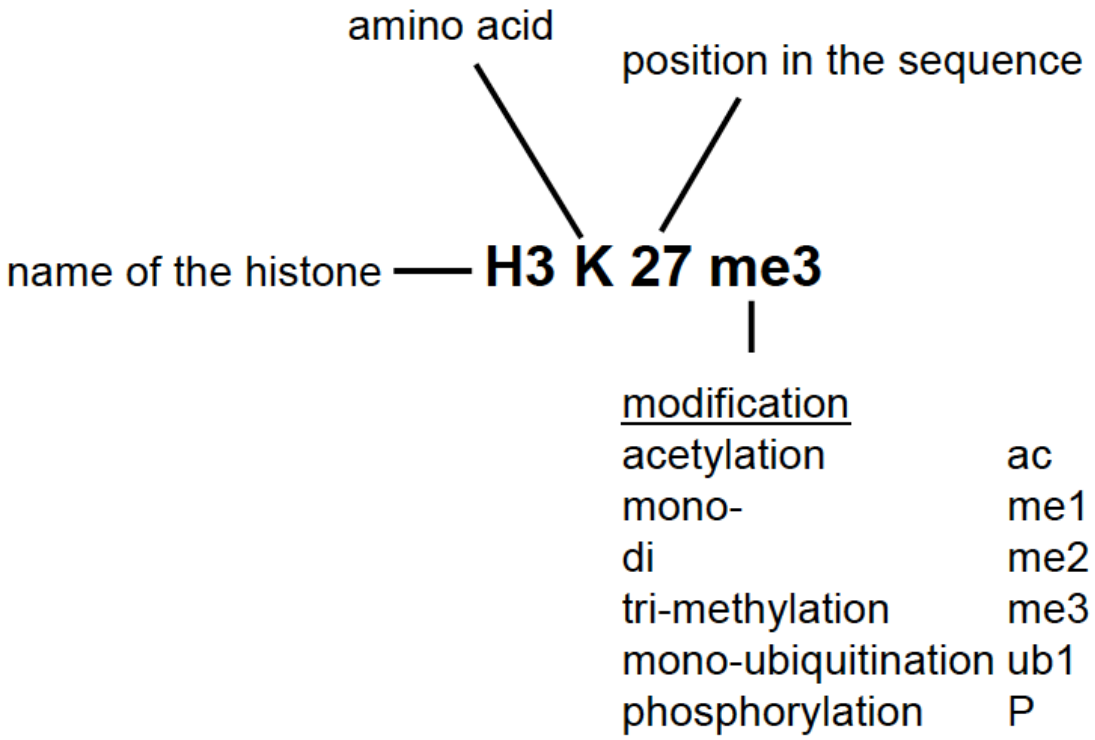
#: this domain binds $(H3\text{-}H4)_2$. Veriations influence nucleosome stability.

2) Covalent modifications of N-term tails





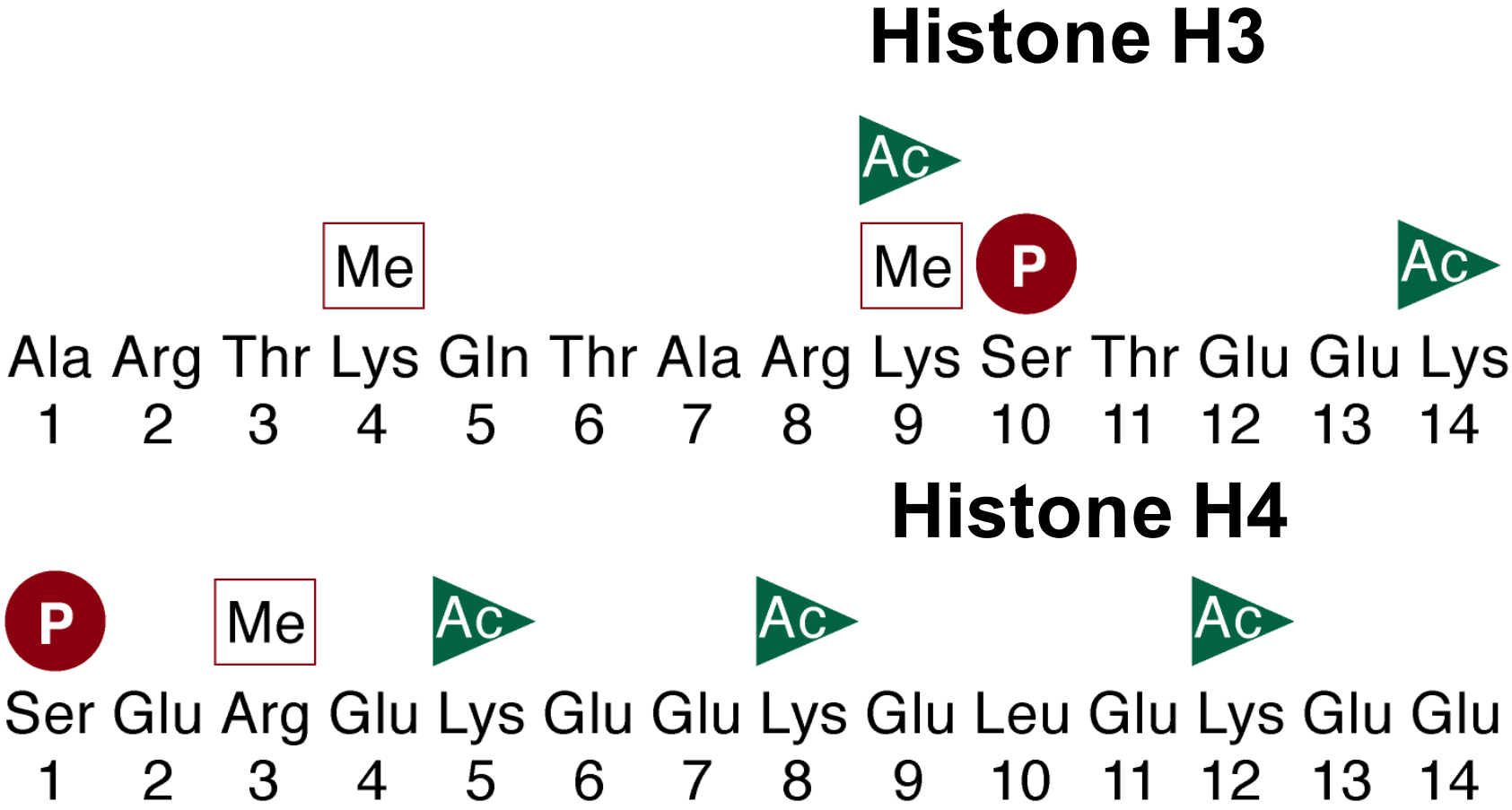
2) Covalent modifications of N-term tails: NOMENCLATURE



2) Covalent modifications of N-term tails: NOMENCLATURE

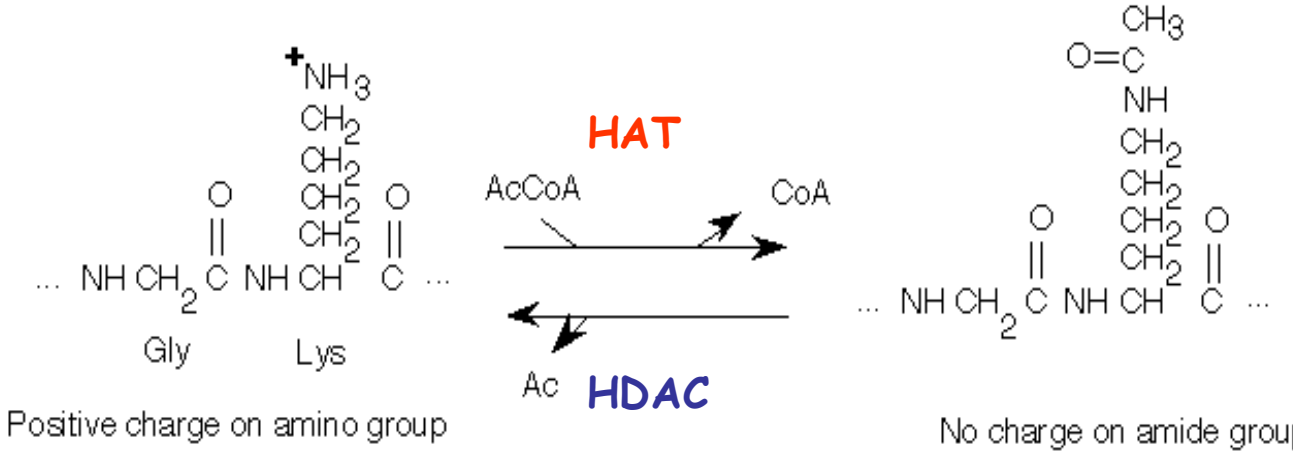
1) Histone acetylases (HATs)
2) Histone de-acetylases (HDACs)
3) Histone methylases (HMTs)
4) Histone de-methylases (HDMs)
5) Histone ubiquitin ligases

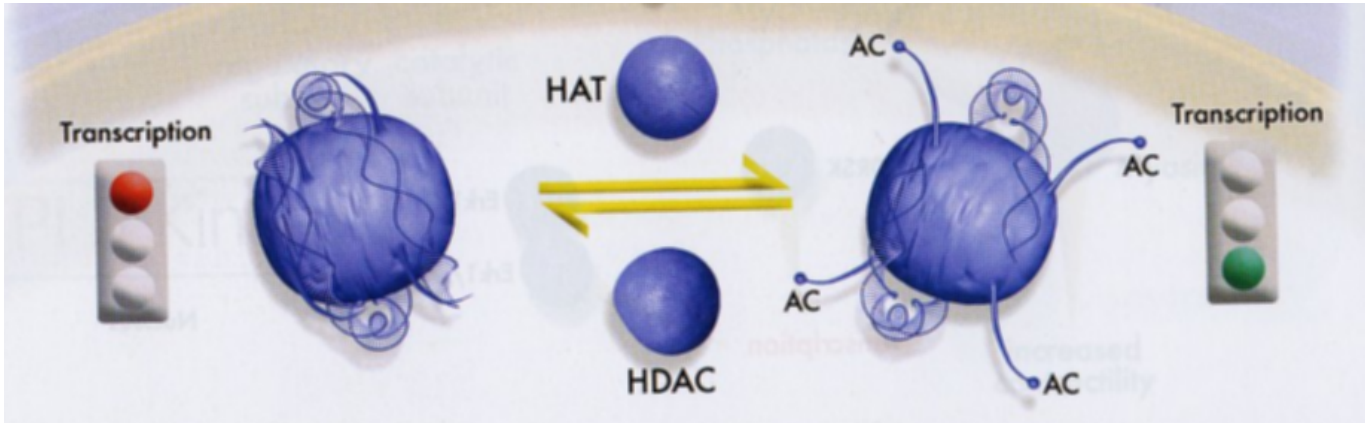
Histone N-terminal ends contain several sites of modification



Histone acetylation

It is a reversible modification of the lysines of the N-terminal tails of the histone core

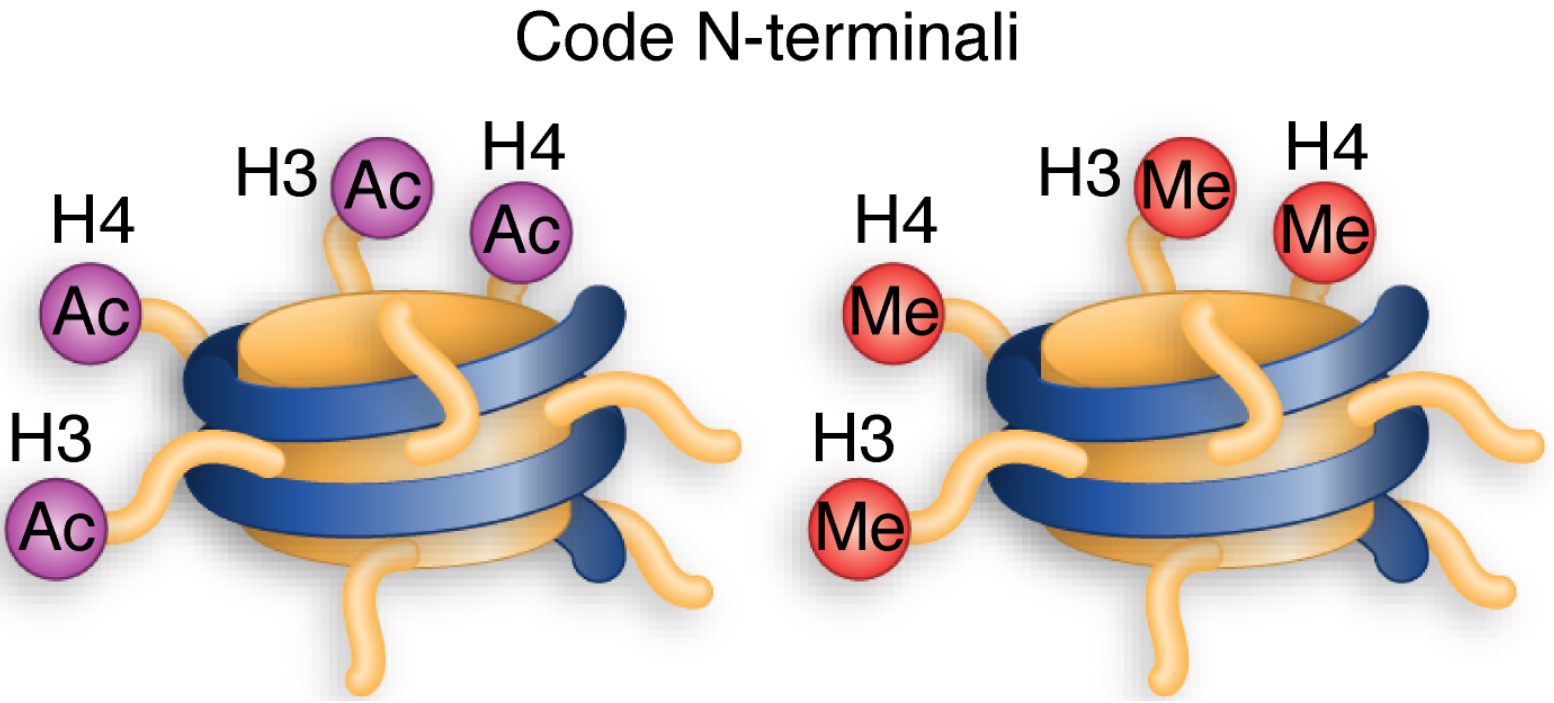




This modification helps transcription as it:

- produces a *decondensation* of the 30nm fiber
- favours the access of transcription factors to chromatin
- acts as a signal for non-histone proteins

Histone modifications control the activity of the genes

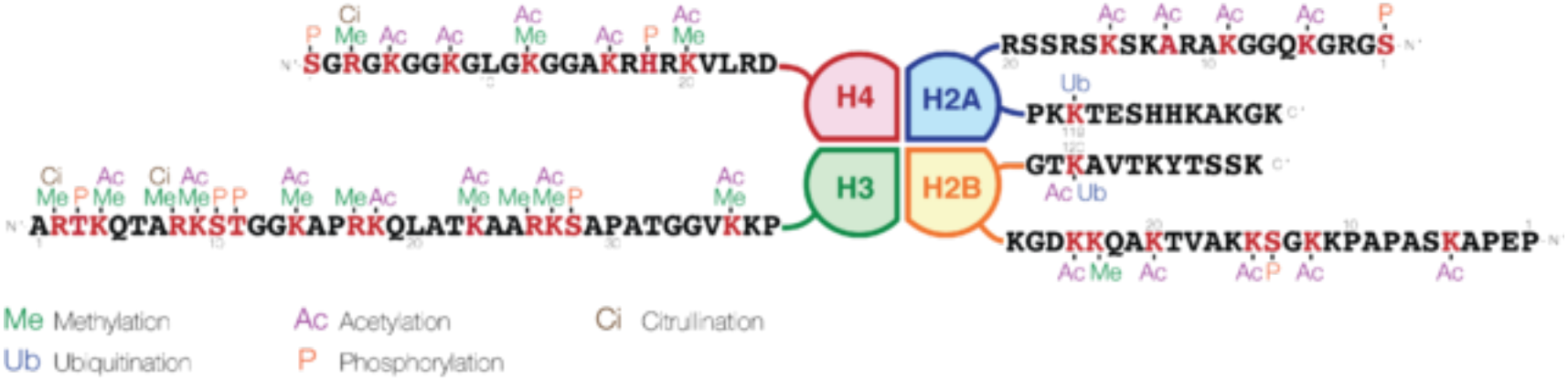


ACTIVE chromatin

INACTIVE chromatin

The histone CODE

A huge catalogue of histone modifications have been described, but a functional understanding of most is still lacking. Collectively, it is thought that histone modifications may underlie a histone code, whereby combinations of histone modifications have specific meanings.



A very basic summary of the histone code for gene expression status is given below (histone nomenclature is described here):

Type of modification	Histone						
	H3K4	H3K9	H3K14	H3K27	H3K79	H4K20	H2BK5
mono-methylation	activation[6]	activation[7]		activation[7]	activation[7][8]	activation[7]	activation[7]
di-methylation	activation	repression[3]		repression[3]	activation[8]		
tri-methylation	activation[9]	repression[7]		repression[7]	activation,[8] repression[7]		repression[3]
acetylation		activation[9]	activation[9]				

- H3K4me3 is found in actively transcribed promoters, particularly just after the transcription start site.<citation needed>
- H3K9me3 is found in constitutively repressed genes.
- H3K27me is found in facultatively repressed genes.[7]
- H3K36me3 is found in actively transcribed gene bodies.
- H3K9ac is found in actively transcribed promoters.
- H3K14ac is found in actively transcribed promoters.

The histone code

- Several combinations of histone modifications establish a "**histone code**" able to *demarcate* distinct regions within ***enhancers***, core ***promoters*** and ***ORFs*** in a way that is critical for the regulation of chromatin-related processes.

- The different types of modifications are deposited by a variety of well-characterized enzymes, which include two main systems of chromatin-modifying activities:

Polycomb (PcG) and **Trithorax** (TrxG) groups of proteins

How do these enzymes

which *lack* DNA binding capacity

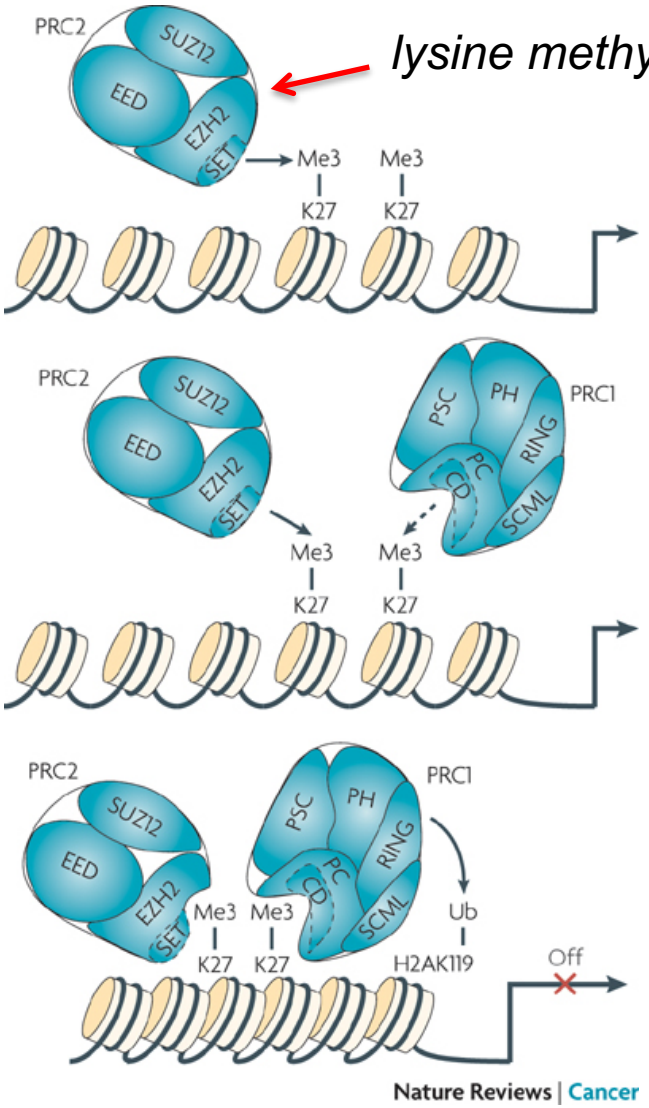
recognise their target genes in the various cell types

Polycomb group proteins (PcG)

Polycomb repressive complexes (PRCs), repress transcription by a mechanism that involves chromatin *modification*. *Two* major Polycomb repressive complexes (PRCs) have been described:

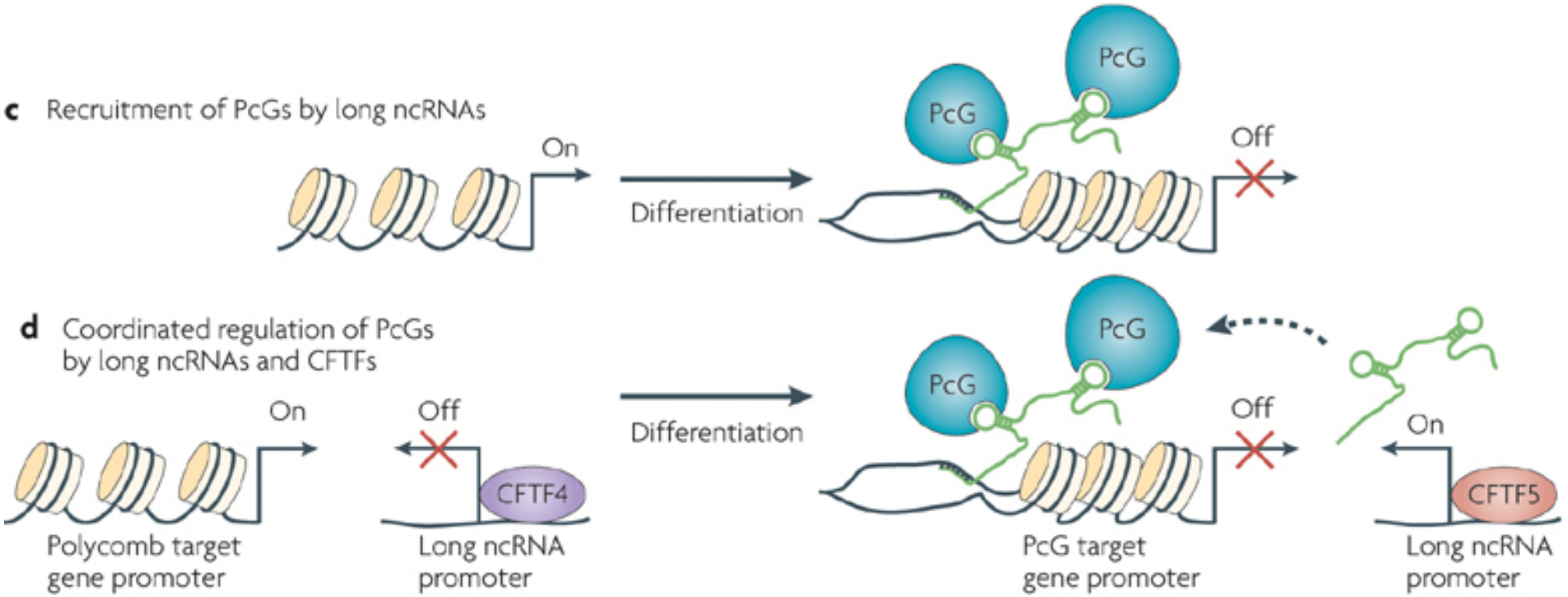
- The **PRC2** contains the histone methyltransferase **EZH2**, which together with **EED** and **SUZ12** catalyses the **H3K27me3 (trimethylation of H3 at lysine K27)**.

- The **PRC1** complexes are recruited by the affinity of chromodomains in chromobox (Cbx) proteins to the H3K27me3 mark. PRC1 recruitment results in the **ubiquitylation of histone H2A on lysine 119**, which is thought to be important for transcriptional repression.

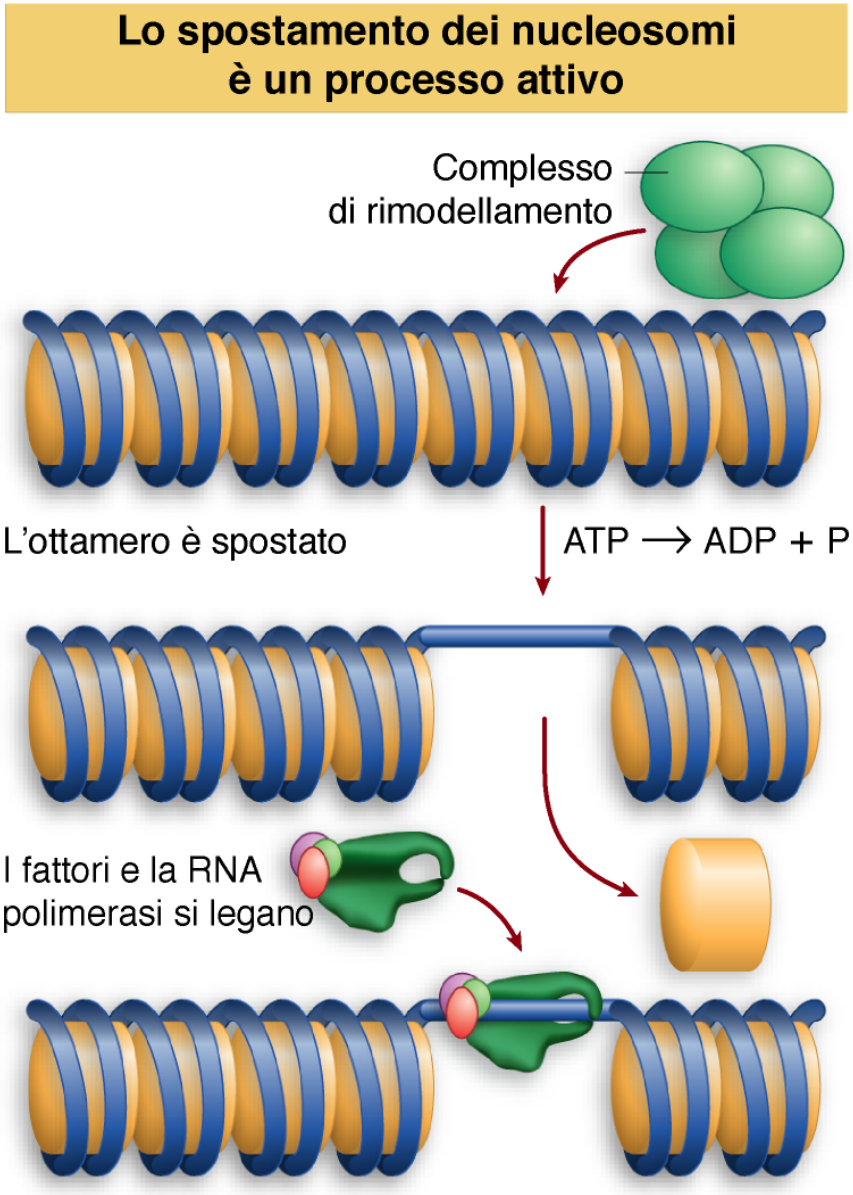


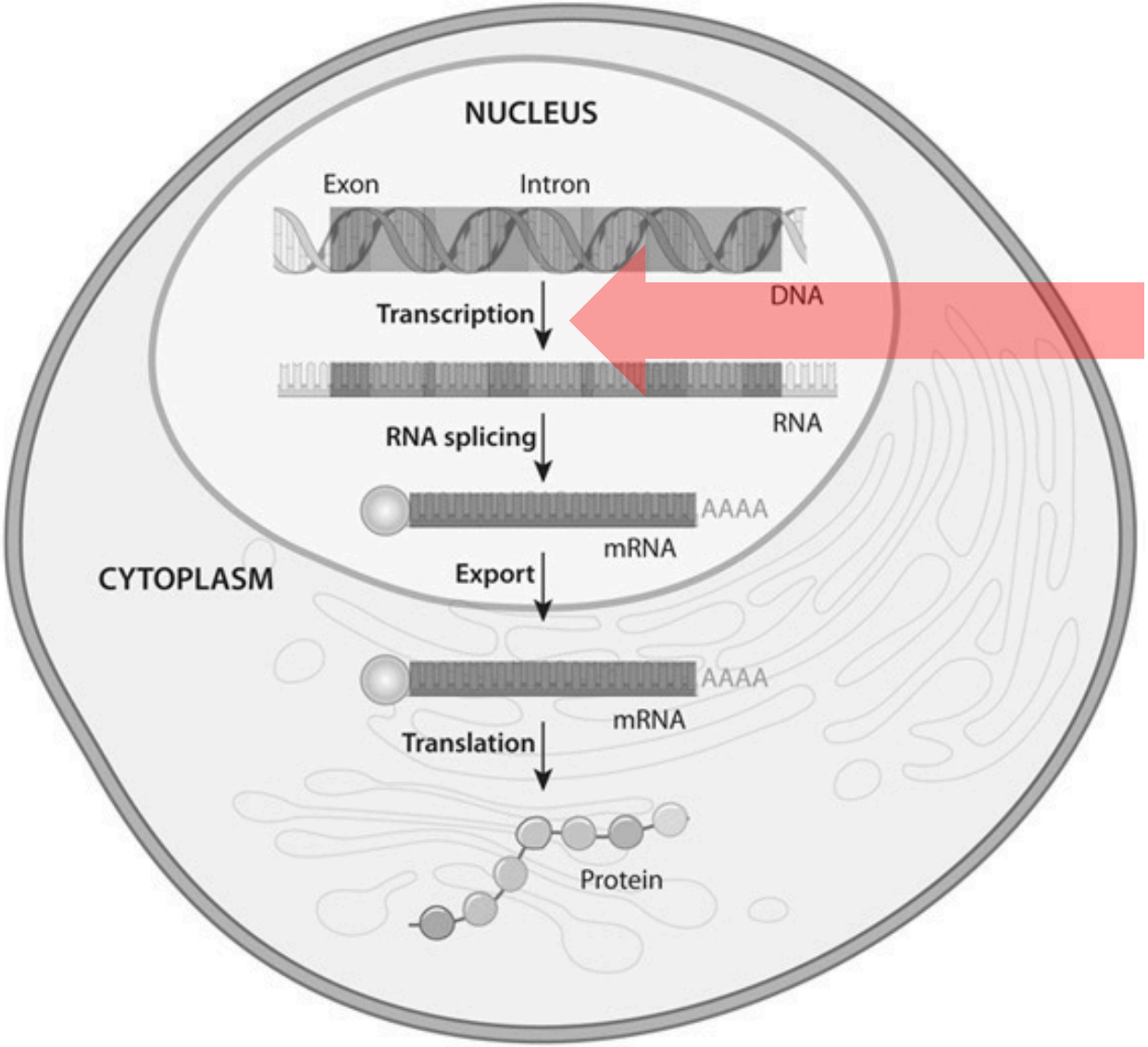
PcG recruitment to target genes

Long non coding RNAs may also regulate PRC binding to specific promoters



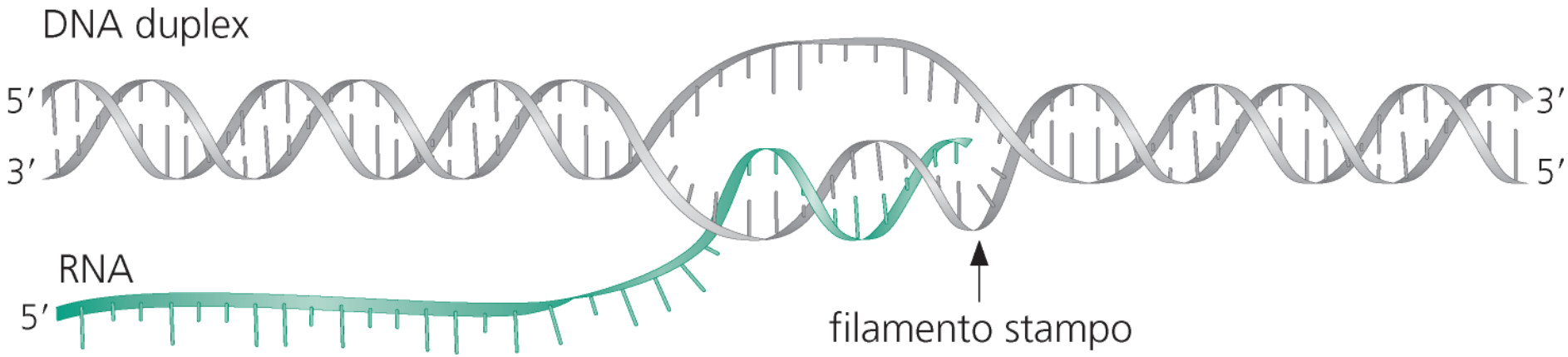
3) Histone modifications favour the binding of chromatin remodelling factors





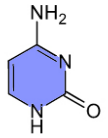
Transcription is the *first* step of gene expression and it is the process by which the information contained into DNA is *converted* in RNA.

Transcription consists in the *synthesis* of an RNA chain from a DNA template.

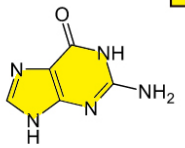


Substrates

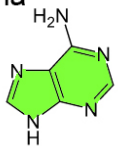
Citosina **C**



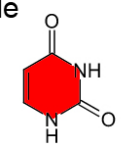
Guanina **G**



Adenina **A**



Uracile **U**



Nitrogenous bases

Enzyme

RNA Polymerase

Products

tRNA

rRNA

mRNA

Gene

Template

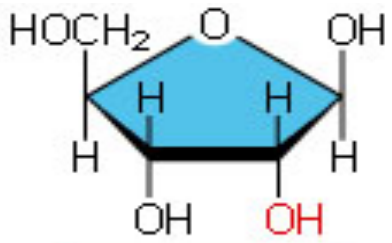
The RNA chain is chemically different from DNA!!!

Why?

- *It is single-stranded.* Like DNA, RNA is composed of its **phosphate group**, five-carbon **sugar** (the less stable ribose), and four nitrogen-containing **nucleobases**.

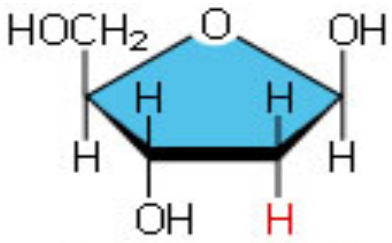
- *Nucleobases* contain *Uracile* (U) instead of Timine (T). Uracil links to Adenine (A-U) and cytosine links to guanine (C-G).

- *nucleotides* contain ribose instead of deoxyribose. Ribose sugar is *more reactive* because of C-OH (hydroxyl) bonds. Not stable in alkaline conditions. *See next…*



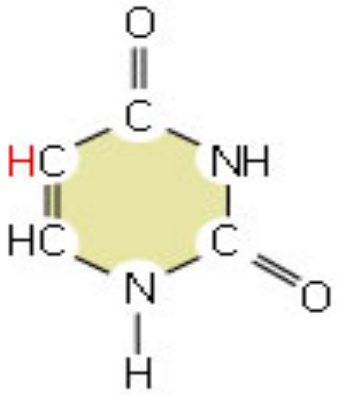
ribose

used in ribonucleic acid (RNA)



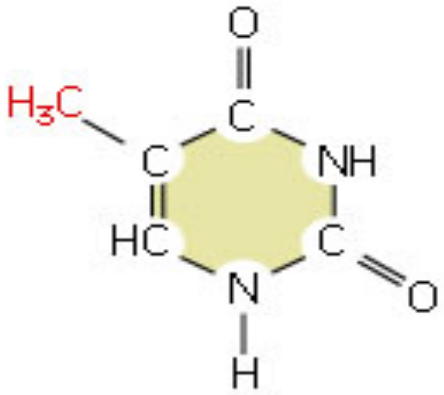
deoxyribose

used in deoxyribonucleic acid (DNA)



uracil

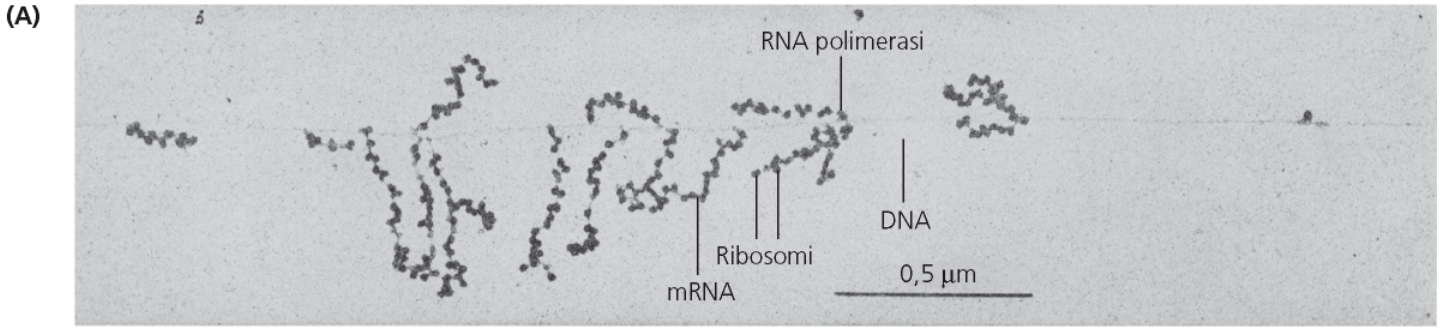
used in RNA

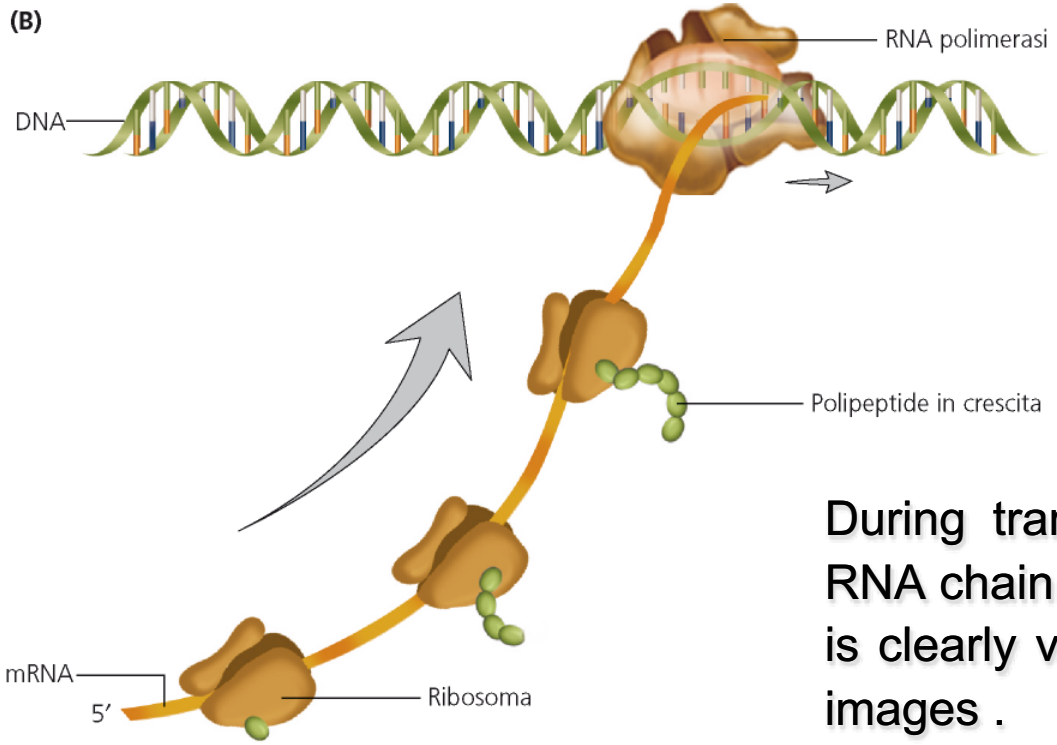


thymine

used in DNA

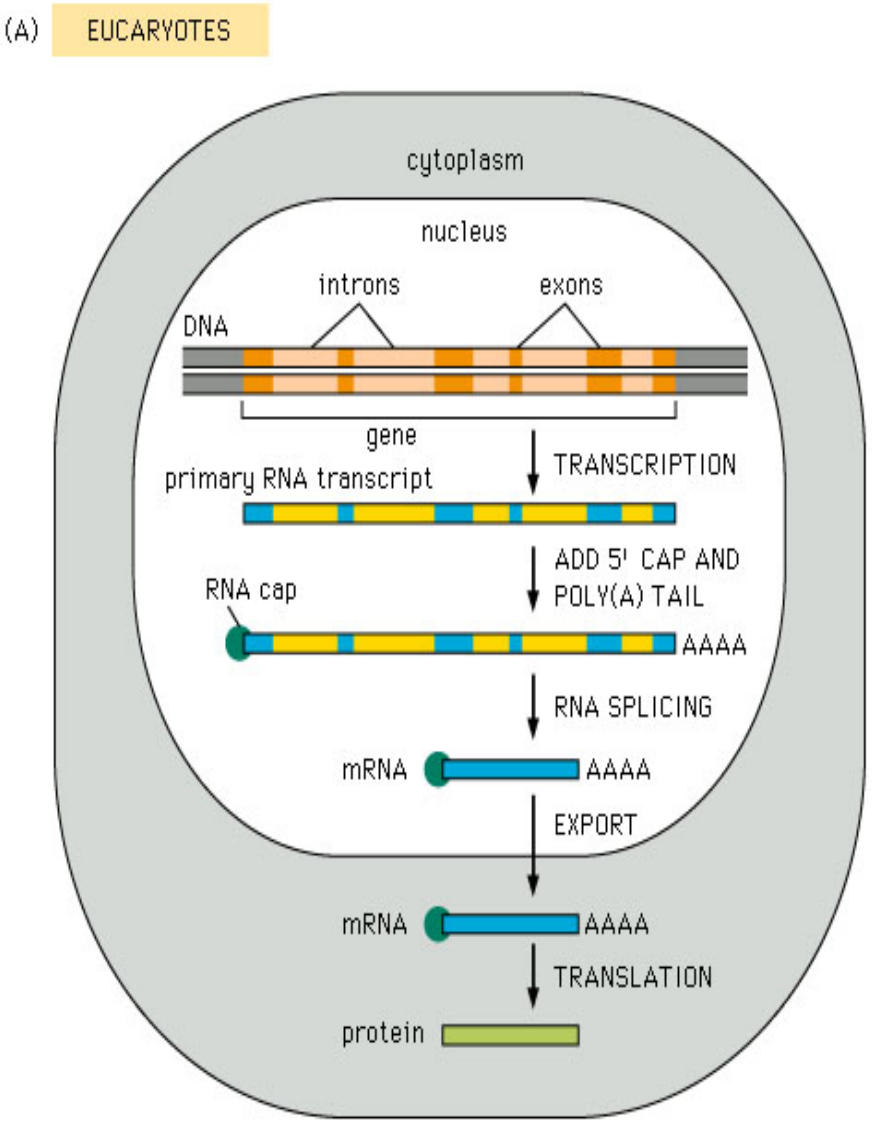
In Prokaryotes …

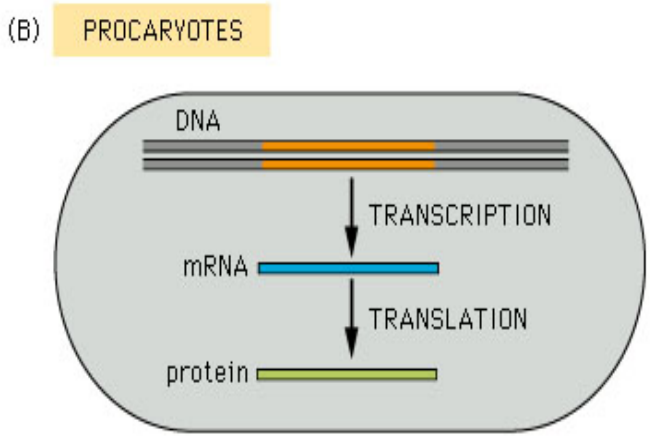




During transcription ribosomes reach the RNA chain to start protein translation, as it is clearly visible from electron microscopy images .

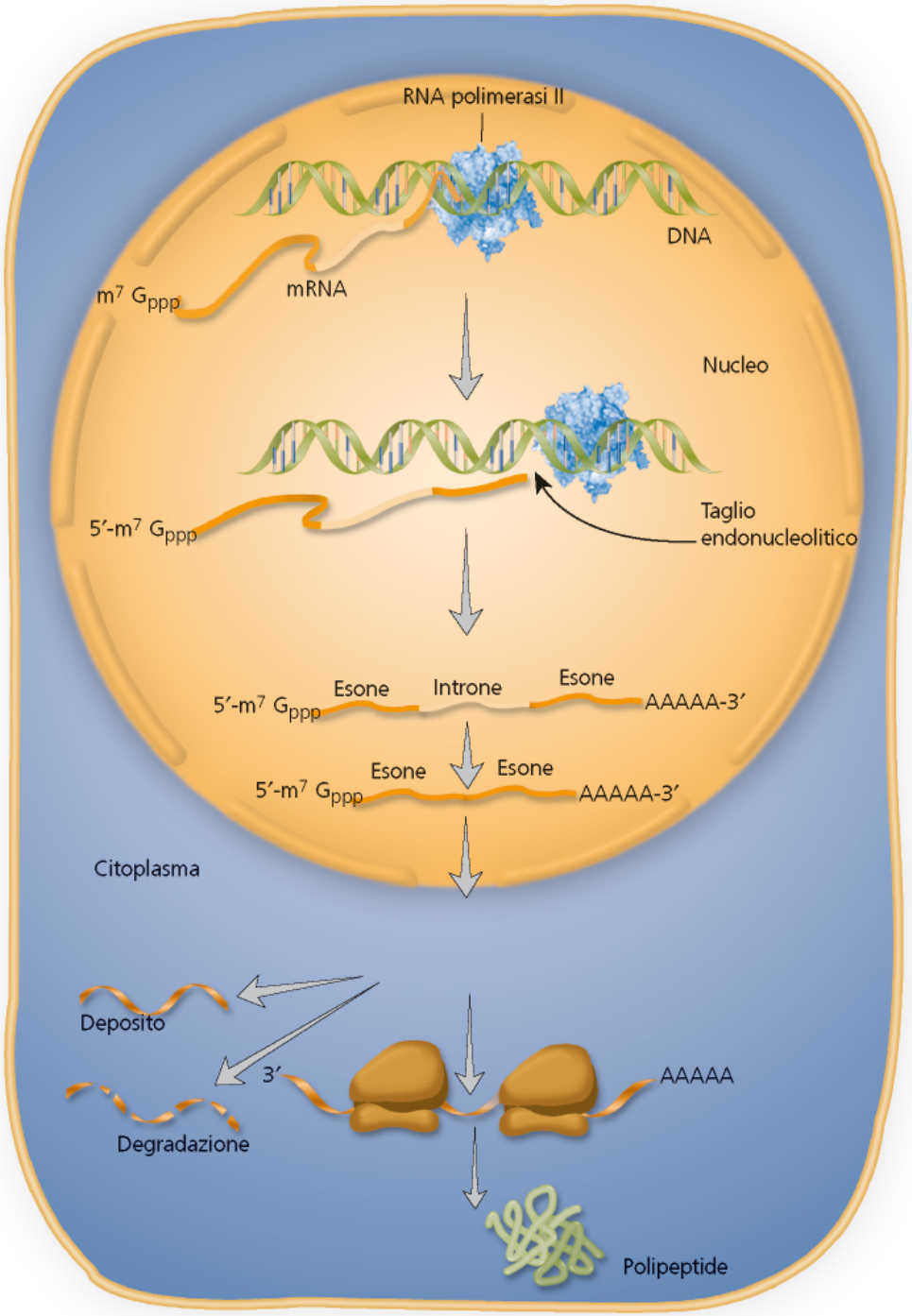
Prokaryotic and Eucaryotic Transcription





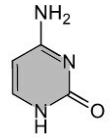
- In Eukaryotes **transcription** and **translation** are temporally and spatially defined events.
- In Prokaryotes **transcription** and **translation** occurr inside the *same* cellular compartment and are coupled.

In Eukaryotes…

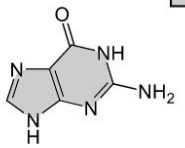


Substrates

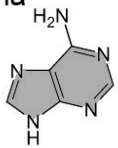
Citosina **C**



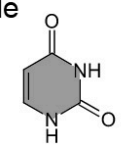
Guanina **G**



Adenina **A**



Uracile **U**



Nitrogenous bases

Enzyme

RNA Polymerase

Products

tRNA
rRNA
mRNA

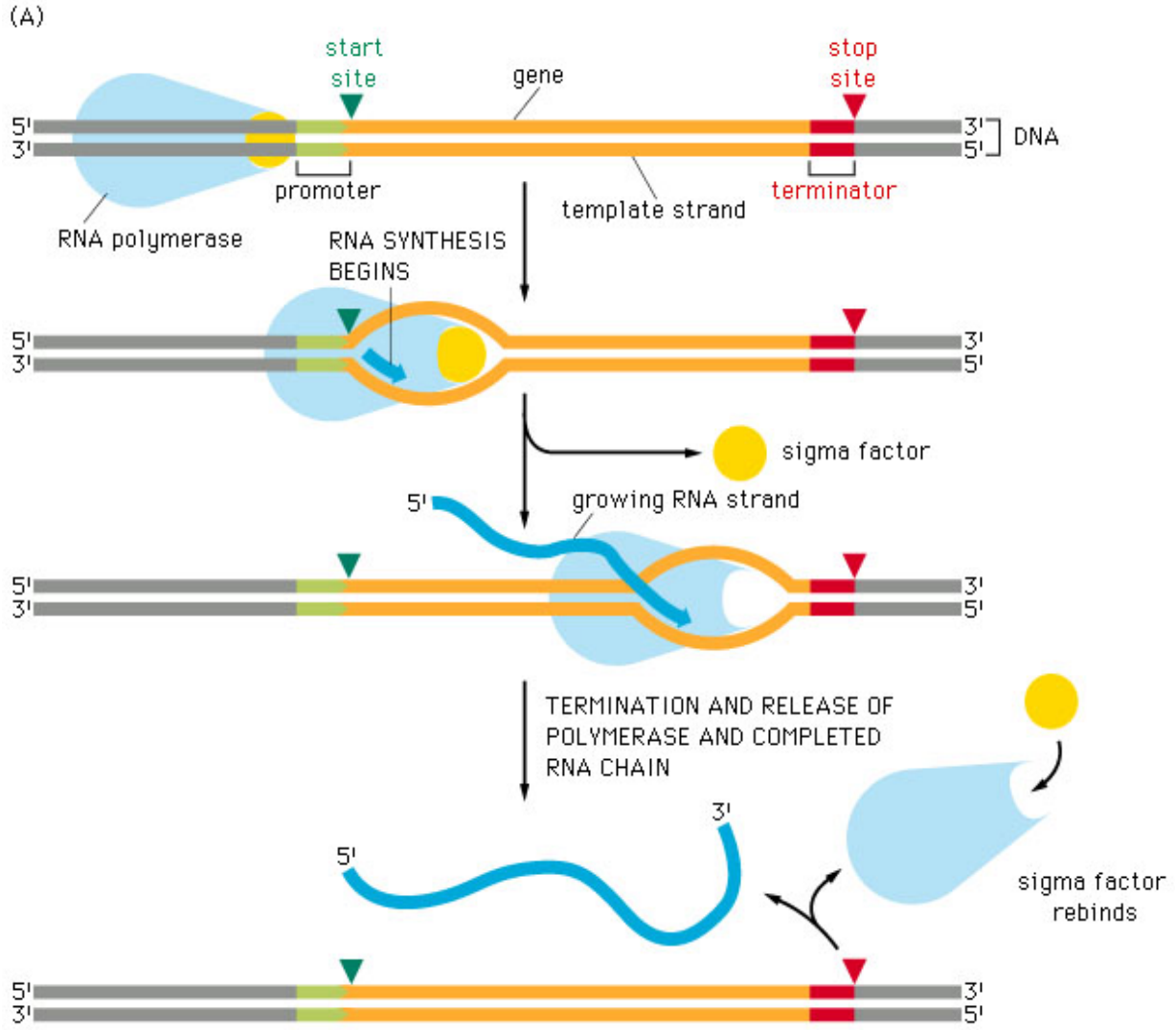
(The Transcription UNIT)

Gene

Template

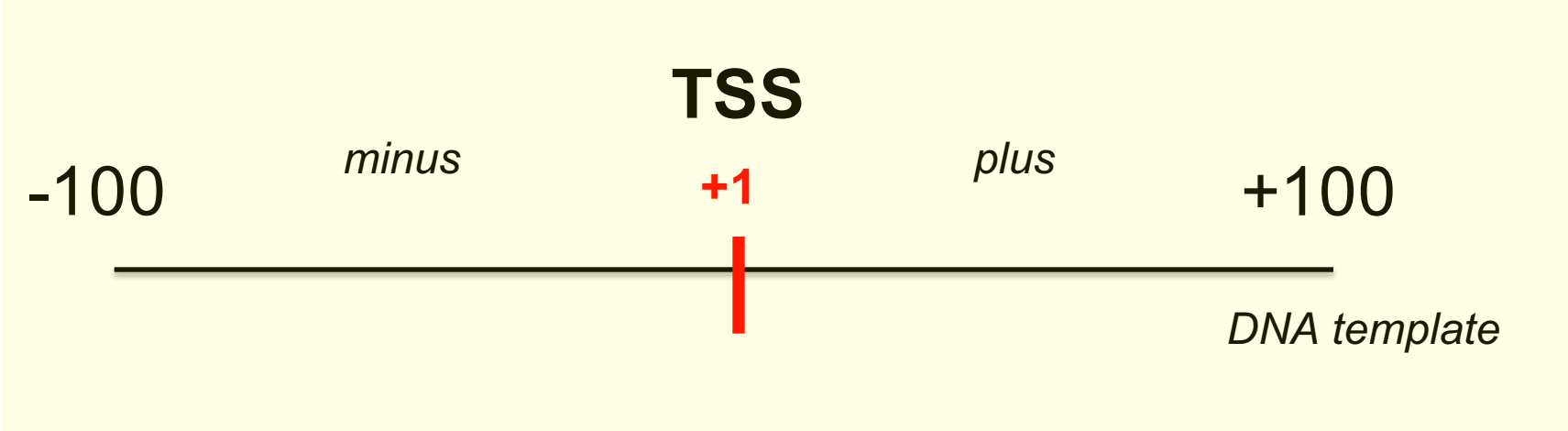
The Transcription UNIT

Sequence of nucleotides in DNA that codes for a single RNA molecule, along with the sequences necessary for its transcription; normally contains a promoter, an RNA-coding sequence, and a terminator.



Transcription starts by the recognition by the enzymatic machinery of specific regions of DNA located at the 5'-end of a gene (UPSTREAM). These sequences generally identify the ***PROMOTER*** of a gene.

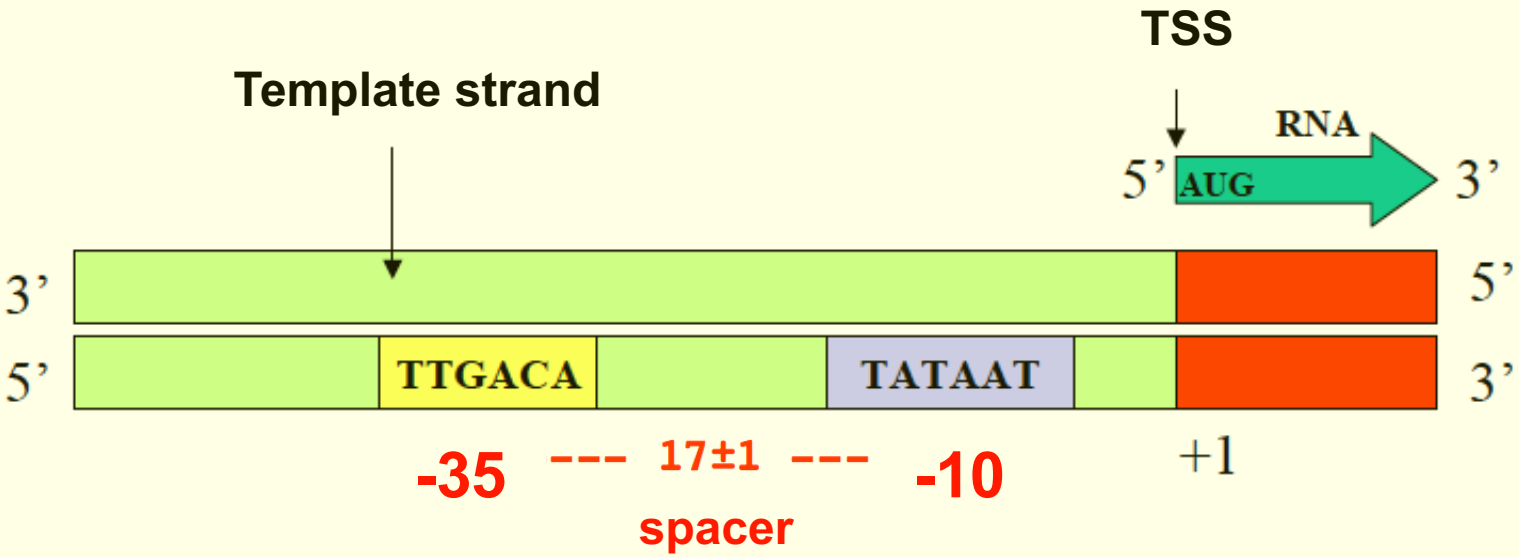
Molecular Biologist use a numbering system which has no zero! The first nucleotide of the RNA transcript is numbered **+1** and correspond to the Transcription Start Site or **TSS**; the nucleotide immediately upstream from that is numbered -1.



PROMOTERS contain *cis elements* which are important to guide the RNA polymerases to recognize the TSS and to start transcription from the right place.

For istance: the comparison of many E. coli promoters has revealed three main *conserved* boxes (or *consensus* sequences): **-35**, **-10**, and the **spacer**.

Structure of a canonical prokaryotic Promoter

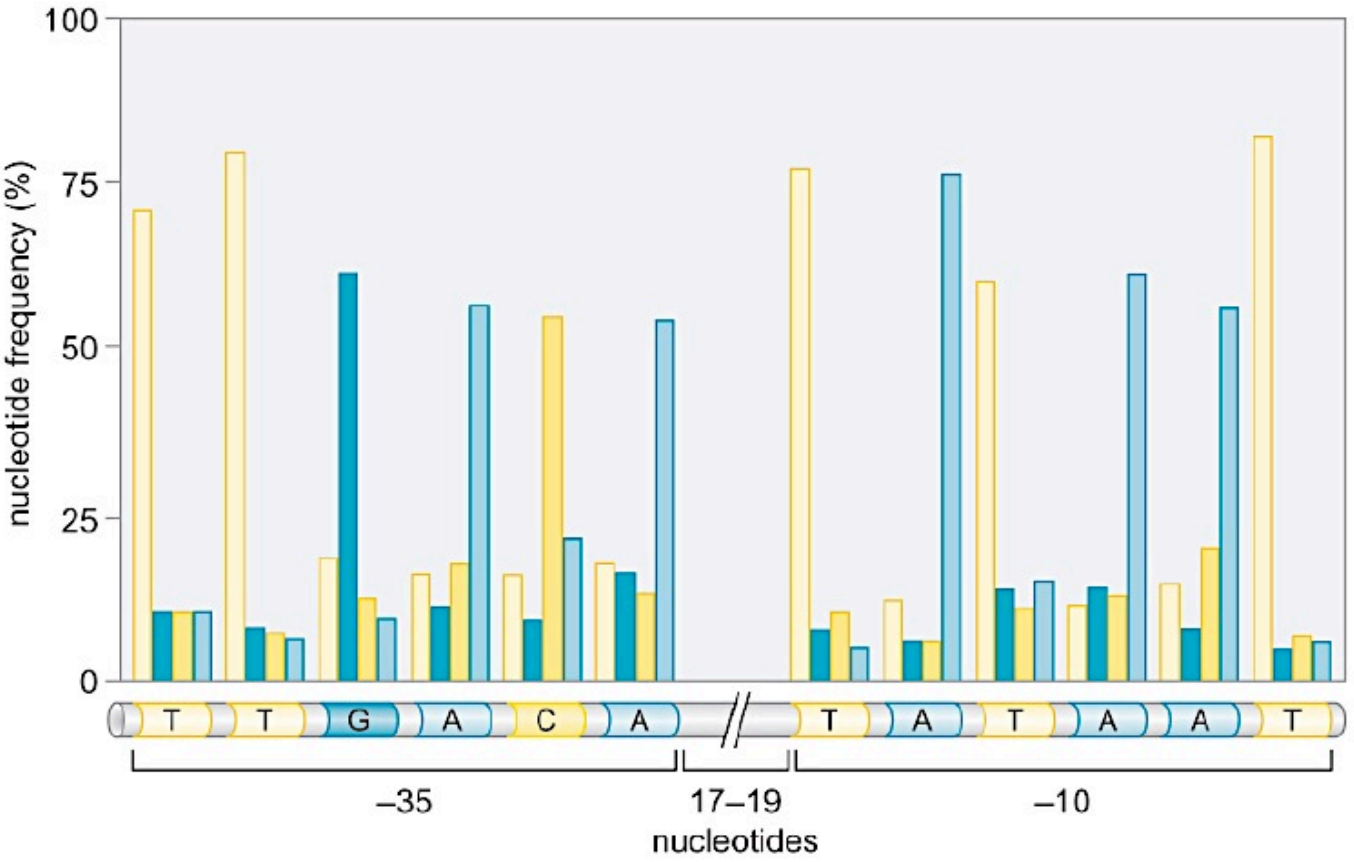


Promoter = *Consensus*

A ***consensus* sequence** is a genetic sequence found in widely divergent organisms or genetic locations with *minor* variations and (probably) *similar* functions.

It represents the residues which are more represented when a lot of sequences are aligned. Take care!!! The consensus sequence is NOT a real sequence but represents the most common nucleotides: it is a **statistical creature**!!!

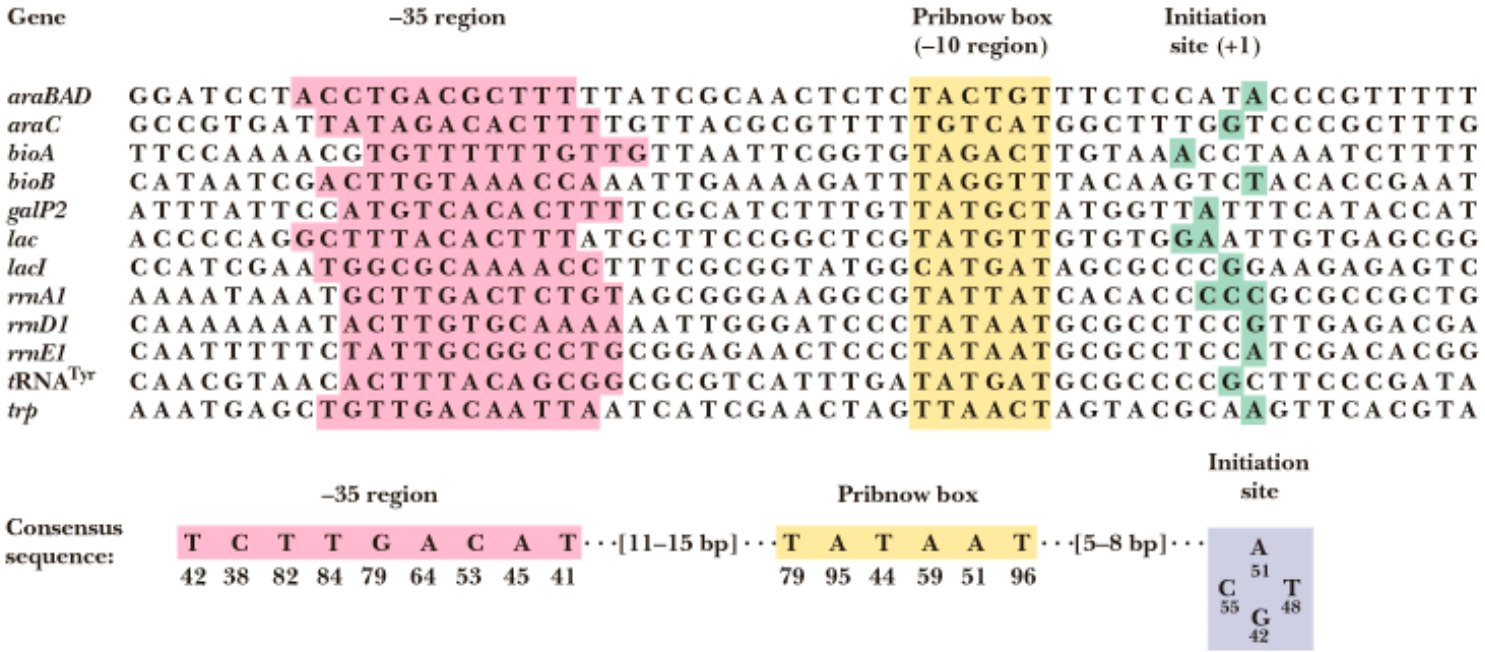
The *consensus sequence* of *E. coli* promoter was found by alignment of 300 sequences interacting with σ^{70}



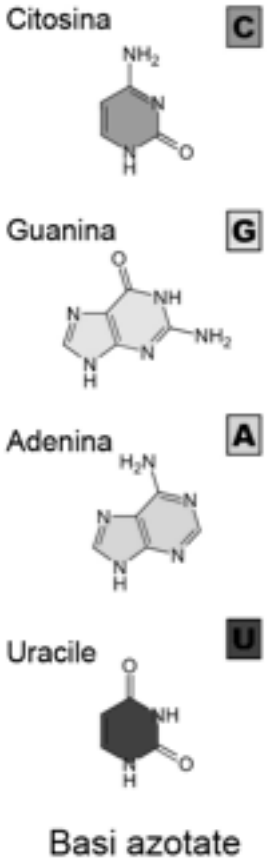
Promoter "*strength*" is defined as the number of transcripts made/unit of time.
It is generally a matter of:

- How *tightly* RNA polymerase binds Promoter (which depends on the consensus sequences)
- *Isomerization* efficiency
- How rapidly the RNA polymerase leaves the Promoter

Based on their *strenght* Promoters can be **STRONG** or **WEAK**



Substrates



Basi azotate

Enzyme

RNA Polymerases

Products

tRNA
rRNA
mRNA

Gene

Template

The RNA polymerase enzymes

RNA polymerases are able to *recognize* and *selectively* transcribe genes by interacting, with the help of other proteins (*trans* elements), with specific sites of DNA (*cis* elements) inside the Promoters.

Once recognized the *cis* elements, the RNA polymerases begins the synthesis of RNA.

- they *synthetize* RNA in 5′→3′ direction by using the 4 *riboNTP* as precursors
- They copy the template DNA
- They do not need start primers
- They have no proofreading activity

Procaryotes		*Eucaryotes*		
Bacteria	**Archei**	**RNAP I**	**RNAP II**	**RNAP III**
Core	Core	Pol I	Pol II	Pol III
β	A’/A’’	RPA1	RPB1	RPC1
β’	B	RPA2	RPB2	RPC2
α^{I}	D	RPC5	RPB3	RPC5
α^{II}	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	[+ other 6]	[+ other 9]	[+ other 7]	[+ other 11]
		transcribes **rRNA**	transcribes **mRNA**	transcribes **tRNA**, small RNA and the **RNA 5S**

Trascription in Eukaryotes

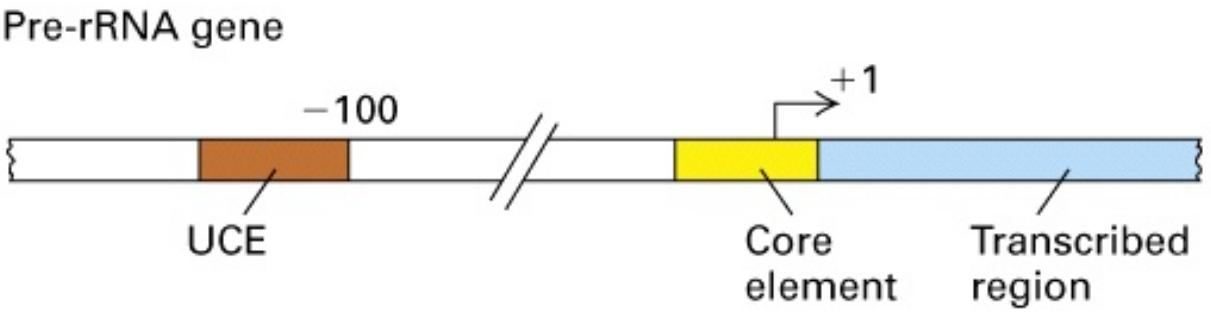
Nucleus of Eucaryotic cells contains three different kind of RNA polymerases DNA-dependent, all of them are omologs to bacterial RNA polymerases.

Each kind of eukaryotic RNA polymerases recognise *different* promoters (sequence and position) to synthetize specifically *different* kind of RNA.

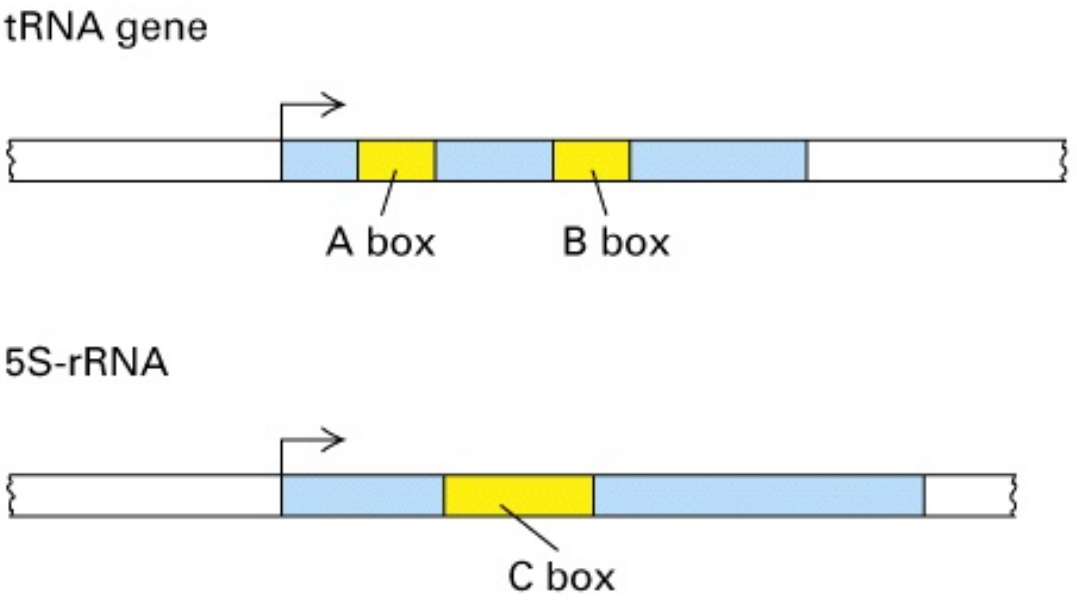
		α-amanitin
RNA polymerase I	**rRNA**	-
RNA polymerase II	**mRNA**	++
RNA polymerase III	**tRNA**	-/+

General structure of Pol I e Pol III Promoters

RNA Pol I

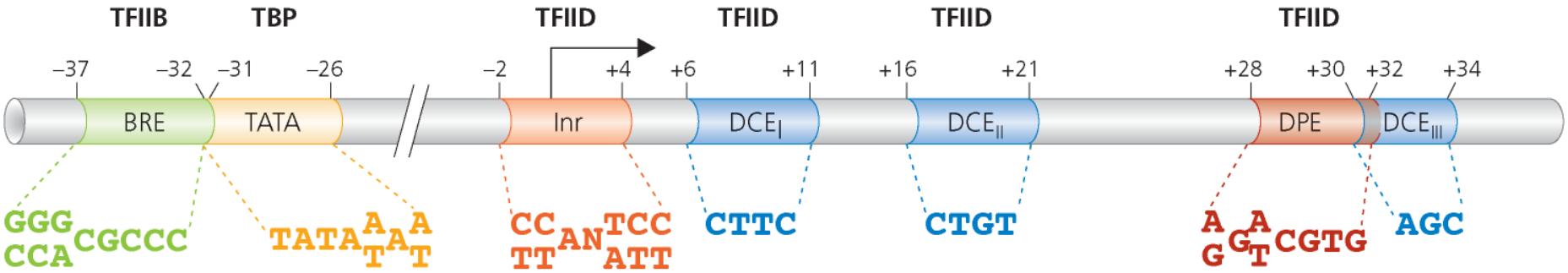


RNA Pol III



RNA Pol II promoters

Each gene can be categorized on the basis of the presence of specific DNA elements at its promoter, the core promoter elements (CPEs)



"Core" promoter: 40-60 bp

-25: **TATA** box, similar to bacterial -10. The presence or absence of a TATA box is used broadly to classify genes as TATA-containing or TATA-less promoters

The sequences immediately flanking the TATA box can contain the elements recognized by the general transcription factor TFIIB. These elements contact general transcription factors (**GTFs**).

-35: **BRE** (TFIIB Recognition Element)

-2+4: **Inr** element (Initiator)

+28-+30: **DPE** (Downstream Promoter Element)

Different positions: **DCE** (Downstream Core Element)

- They are not always present *together* in the Pol II promoters
- There are also other regulatory elements differently located upstream them (enhancers, LCR, insulators)

Whether prokaryotic or eukaryotic…**transcription** consists of three main events:

1) ***Initiation*** - binding of RNA polymerase to double-stranded DNA; this step involves a transition to *single-strandedness* in the region of binding; RNA polymerase binds at a sequence of DNA called the promoter.

2) ***Elongation*** - the covalent *addition* of nucleotides to the 3' end of the growing polynucleotide chain; this involves the development of a short stretch of DNA that is transiently single-stranded

3) ***Termination*** - the recognition of the transcription termination sequence and the release of RNA polymerase

1) RNA Polymerases search for specific sequences (promoters)

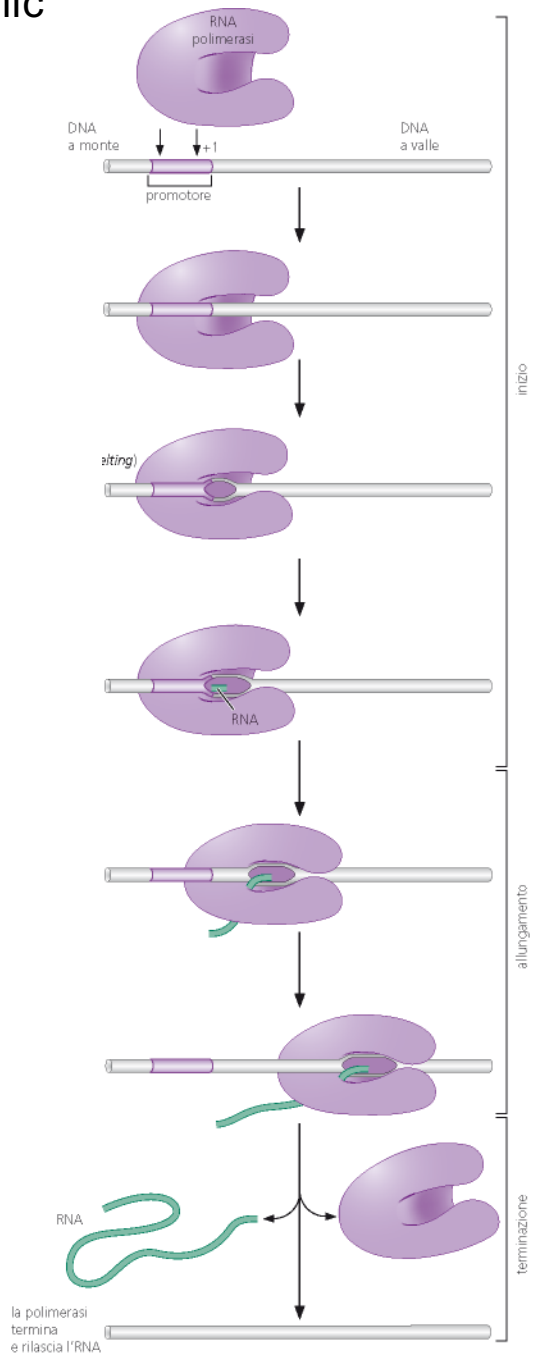
2) RNA Polymerase binds promoter: *closed* complex

3) Promoter melting: *open complex*

4) INITIATION

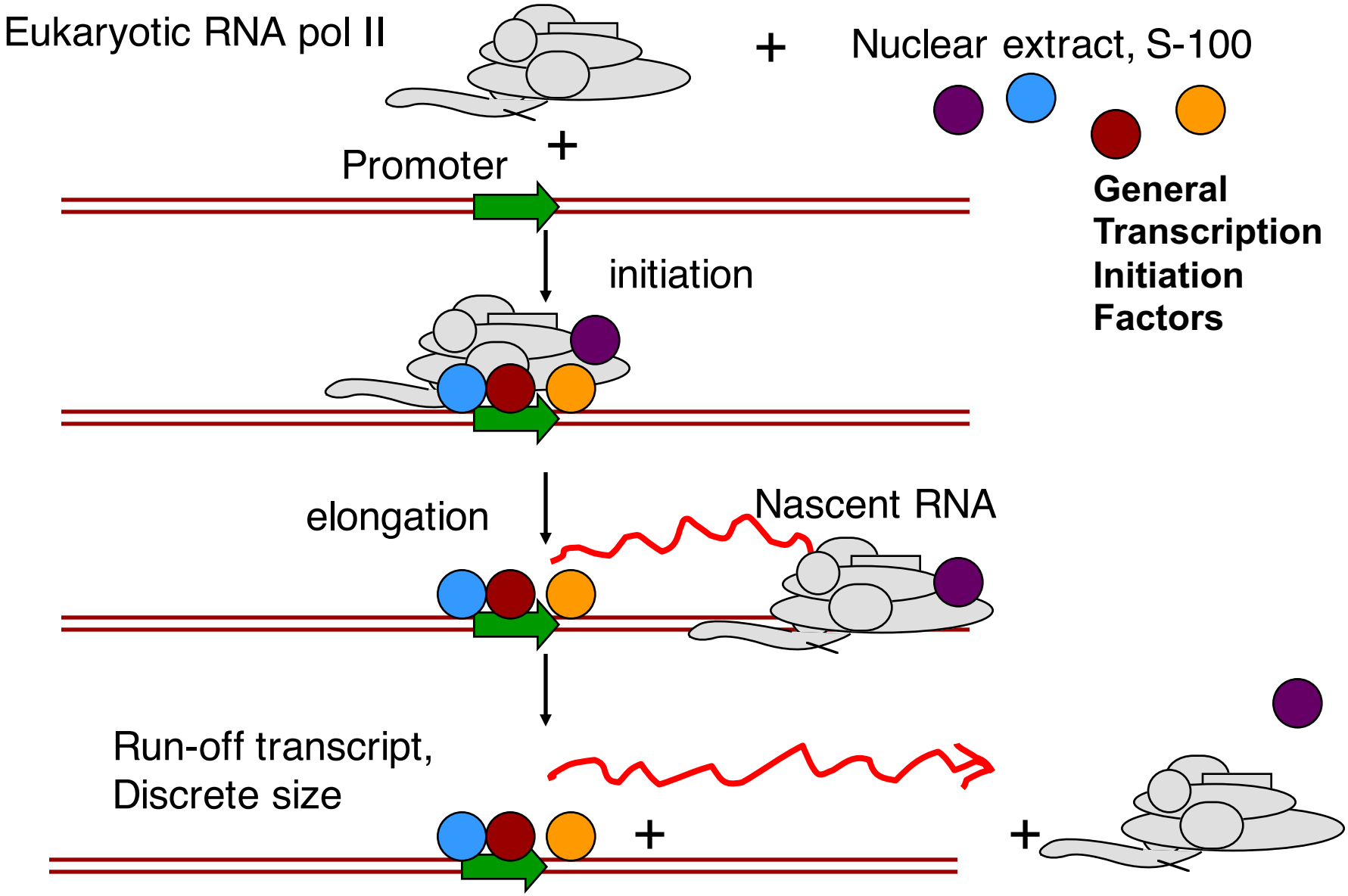
5) ELONGATION

6) TERMINATION and RNA release

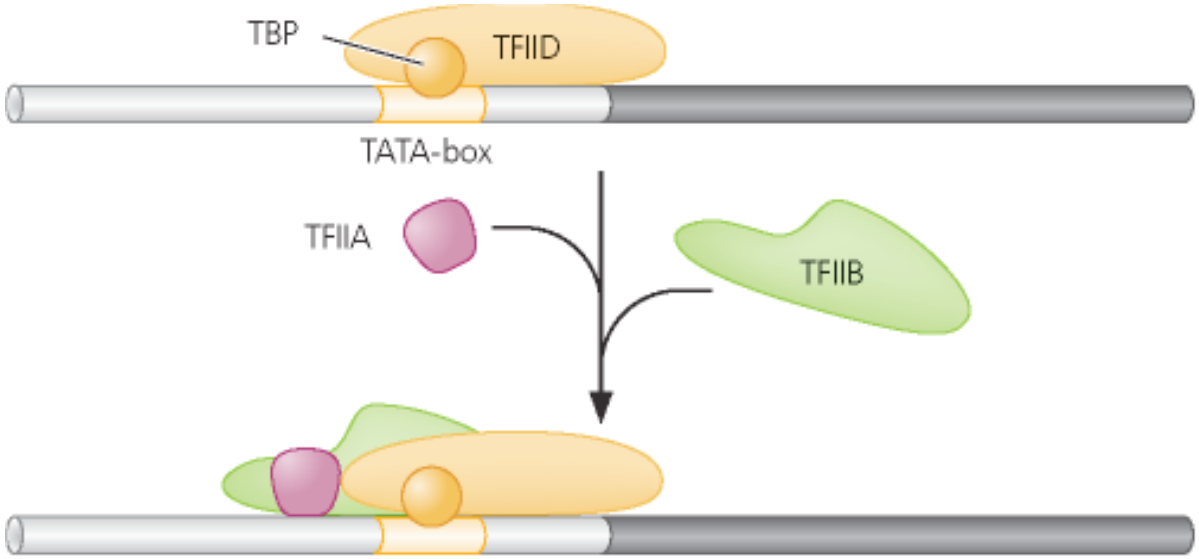


- Only one of the two strands acts as template.
- Transcription only proceeds in 5' - 3' direction.
- Transcripts of less than 5 nt are unstable, resulting in a high frequency of *abortive* initiation.

Accurate INITIATION step for RNA polymerase II+GTF



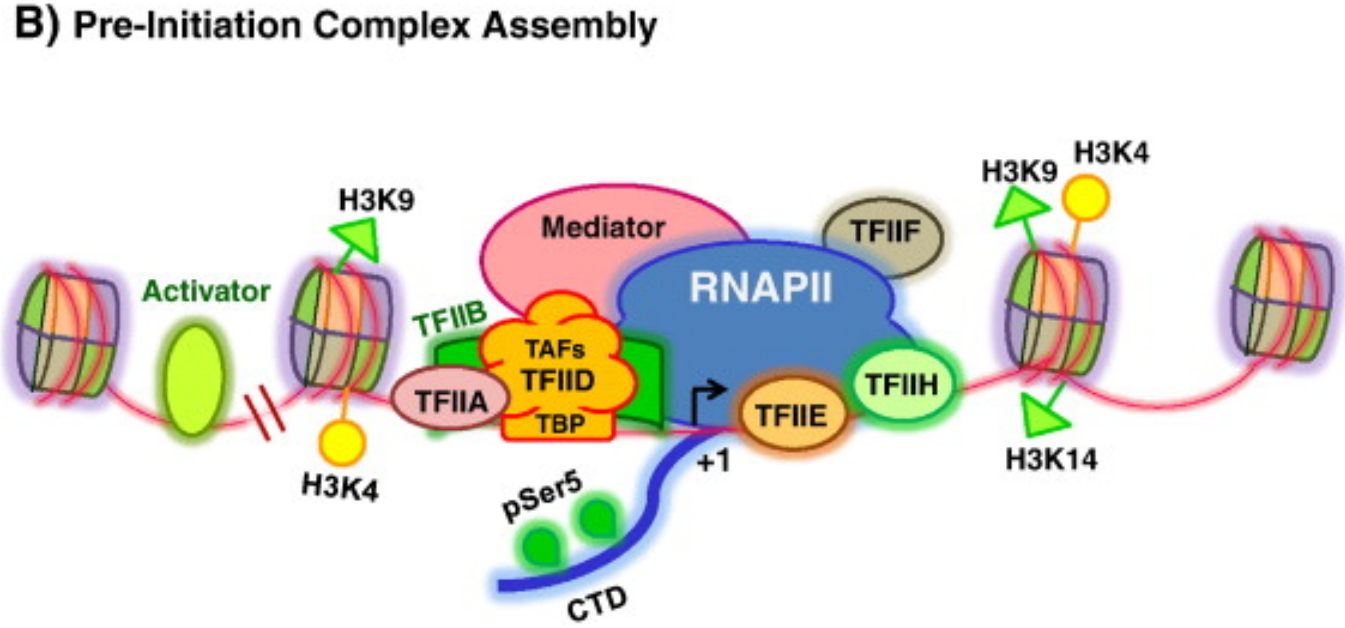
Thus…the general transcription factors (**GTFs**) *help* RNA polimerase to bind promoter



TFIID is the first complex which binds promoter and contains:

- **TBP** (TATA-Binding Protein)
- **TAFs** (TBP-associated factors). Le TAFs si associano all'iniziatore e agli elementi a valle DPE

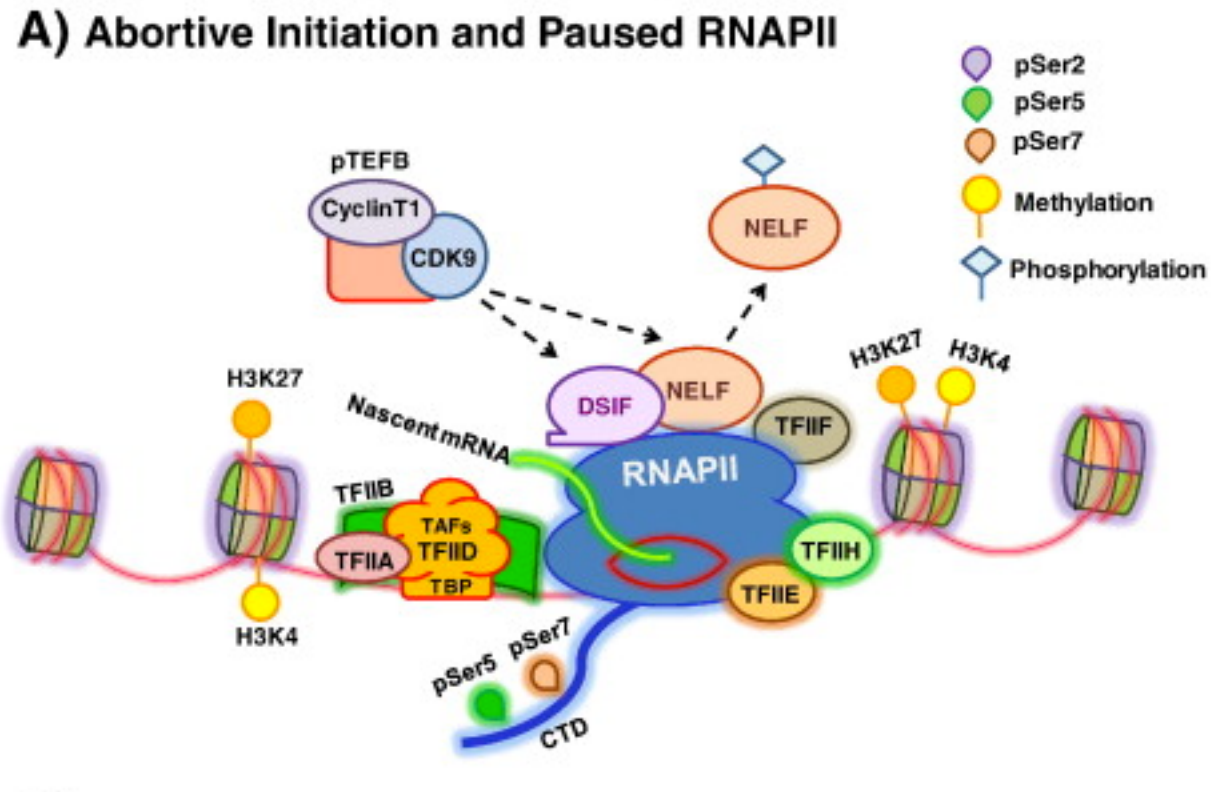
TFIIA and **B** are recruited later. **TFIIB** acts as a bridge between TBP and the Polimerase that is arriving.



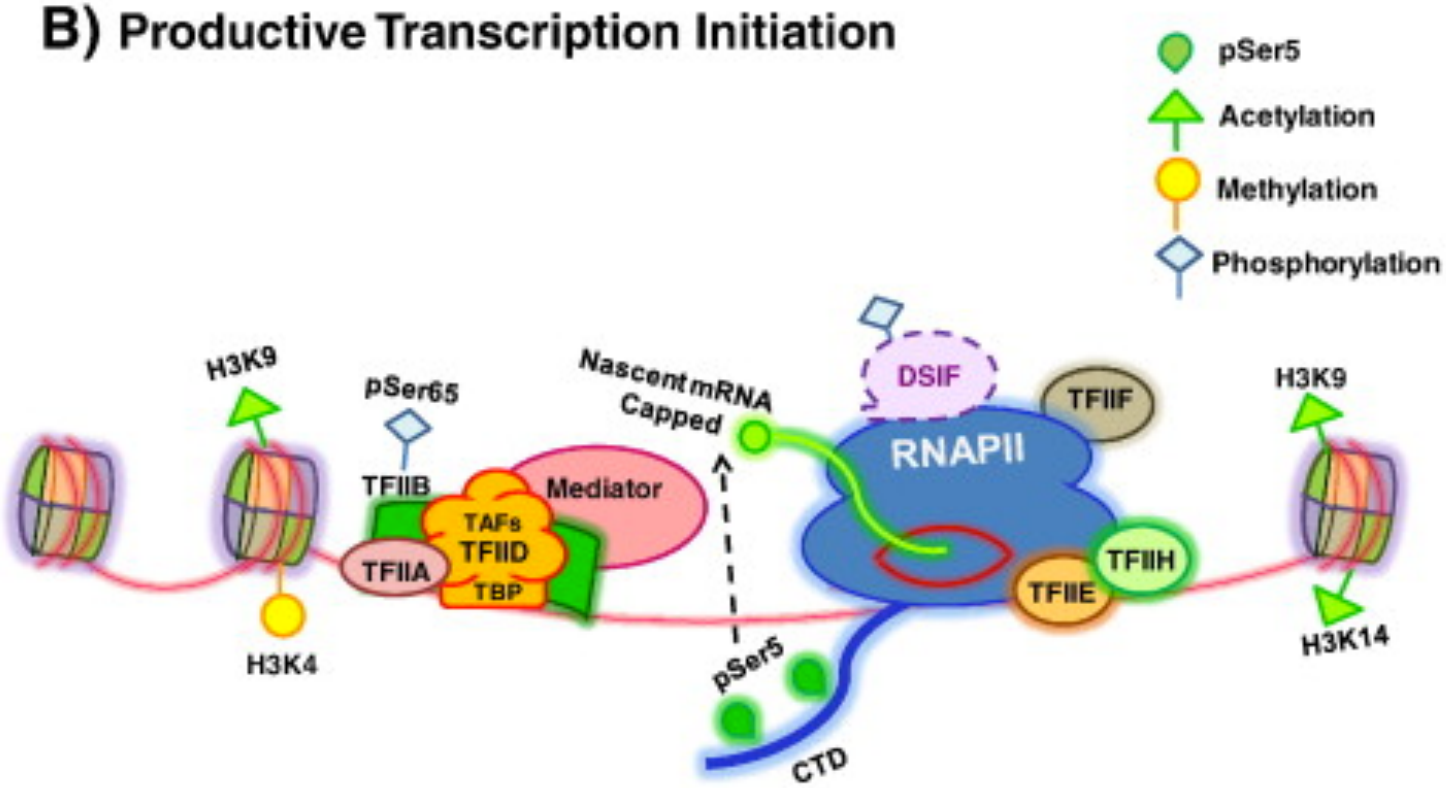
The stability of an assembled PIC is characterized by the presence of essential GTFs, RNAPII and the mediator complex at the promoter.

The nucleosomes flanking the promoter have distinct *histone code* which marks the promoter for active transcription. These epigenetic modifications include acetylation of histone H3 at lysines 9 and 14 (H3K9/14) (green triangles) and methylation of histone H3 at lysine 4 (H3K4) (yellow circles). The CTD repeat of RNAPII is also enriched with phospho-serine 5 mark (green balloon).

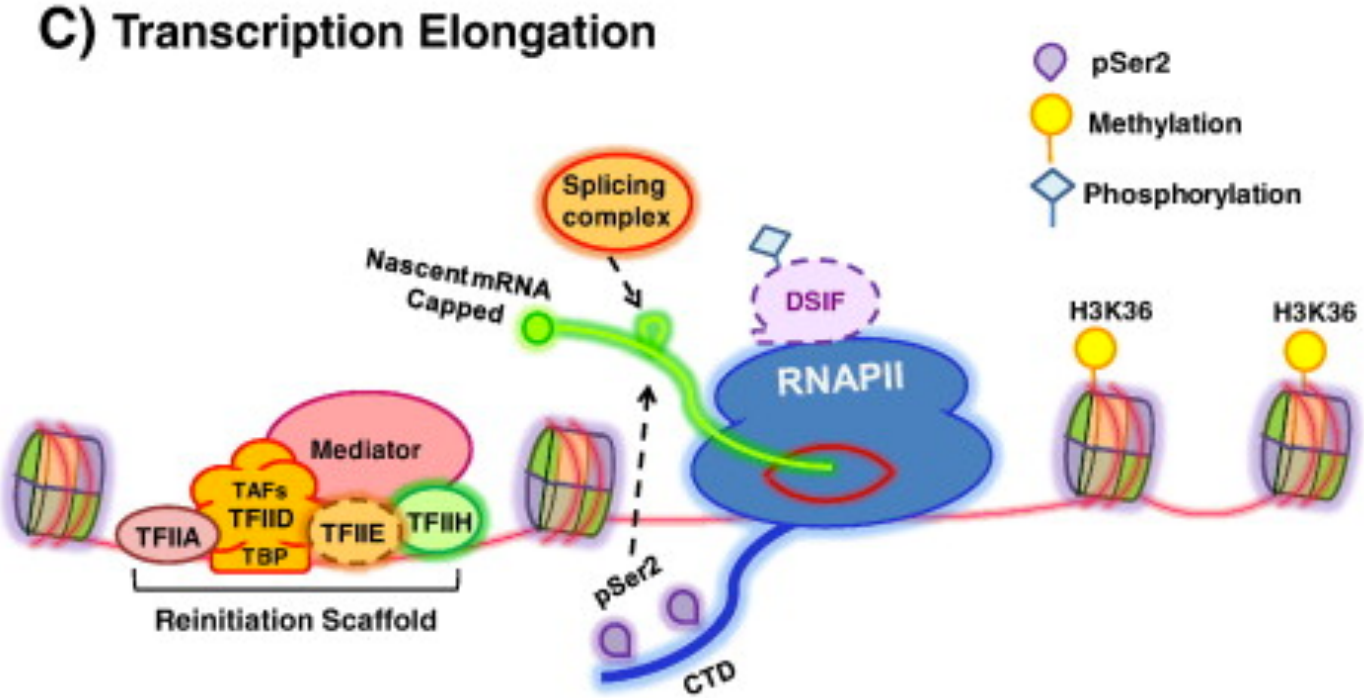
The transcription cycle



(A). Depending on the transcriptional competence of RNAPII, it can potentially enter a *paused* state. Presence of negative factors such as NELF and DSIF inhibits productive transcription initiation resulting in abortive transcription. The kinase activity of the CDK9 subunit of the pTEFb complex alleviates this repression via phosphorylation of NELF and DSIF which results in dissociation of phosphorylated NELF from RNAPII, while DSIF may move along with the elongating RNAPII.

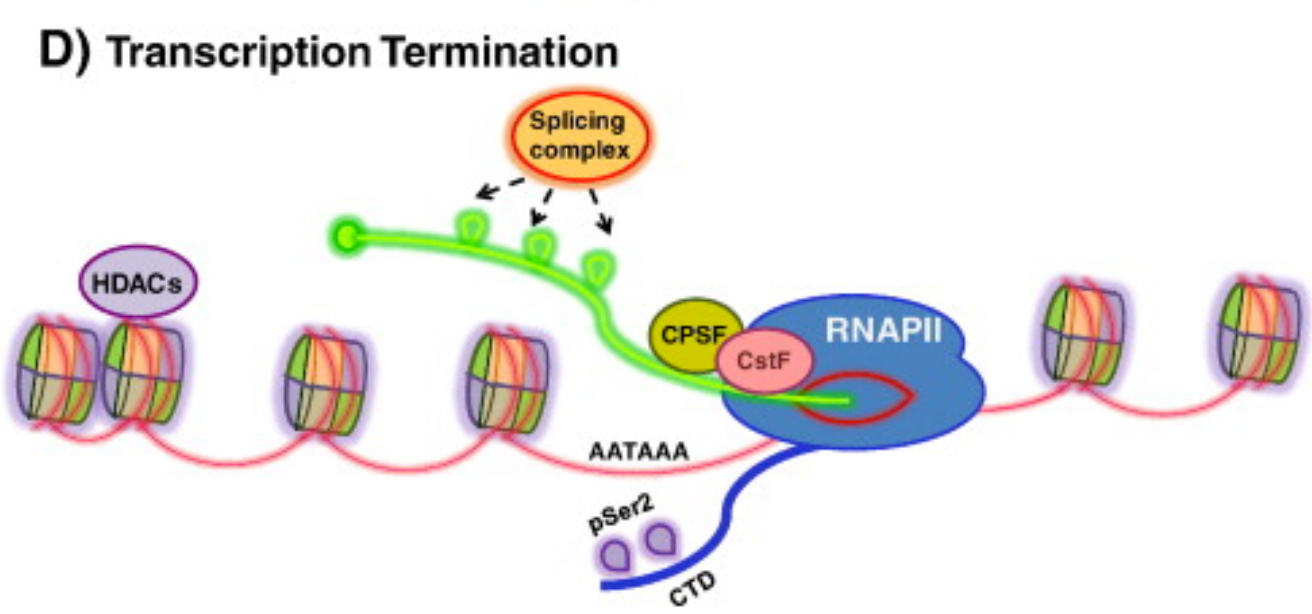


RNAPII can switch to productive initiation mode when all the inhibitory signals/factors are overcome. Active initiation is dependent on TFIIH-mediated promoter melting (red bubble) and phosphorylation of the CTD repeats at serine 5 (green balloon). Along with the phosphorylation of Ser5-CTD, productive transcription initiation also requires the phosphorylation of TFIIB at serine 65. The phospho-Ser5-CTD recruits capping enzyme to the 5′ region of nascent mRNA (green string) which triggers RNAPII-escape from the promoter to the open reading frame (ORF) of the gene



Following promoter clearance, RNAPII proceeds for elongating the transcript while a part of the PIC components remains associated at the promoter forming a reinitiation scaffold. Meanwhile, GTFs such as TFIIB, TFIIF and likely TFIIE fall off.

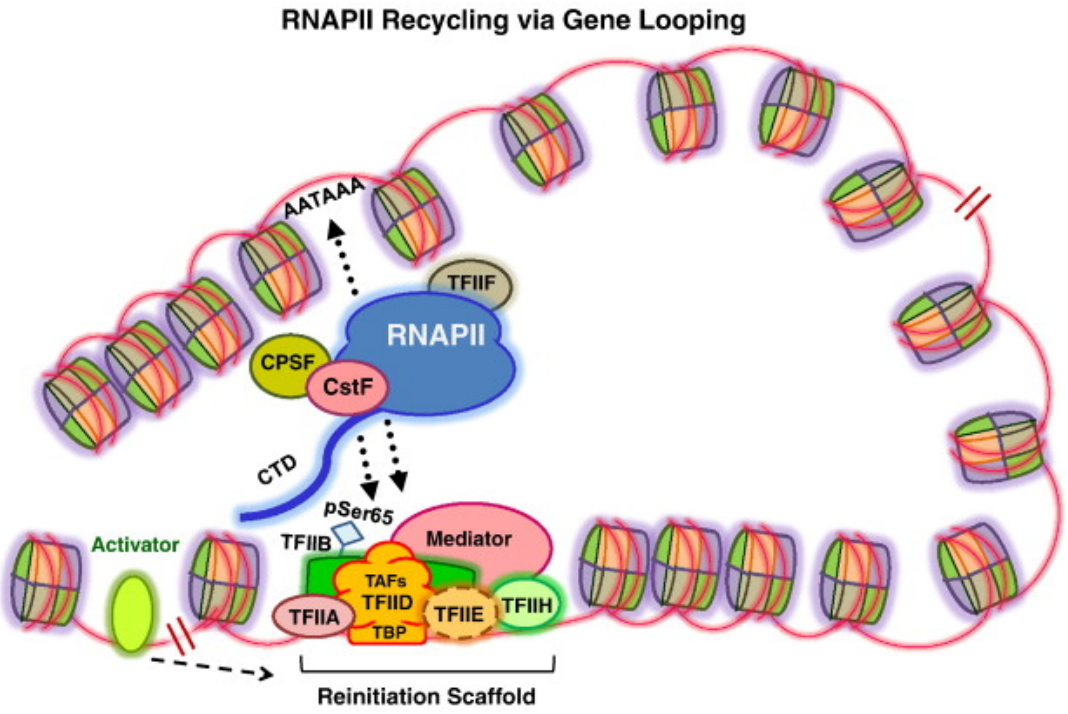
The elongating RNAPII CTD repeat is progressively phosphorylated at serine 2 by cyclin-dependent kinase 9 (CDK9), while the SSU72 phosphatase removes the phospho-Ser5 mark. Phospho-Ser2-CTD (purple balloon) recruits mRNA splicing complex for co-transcriptional splicing of nascent mRNA. The nucleosomes occupying the ORF are enriched with methylation of histone H3 at lysine 36 which is a mark of active elongation (D).



In eukaryotes, most of the protein coding mRNA precursors have a highly conserved poly (A) signal, 5′-AAUAAA-3′ which is followed by a G/U-rich sequence towards the 3′ end. Hence the 3′ end processing of mRNA is a poly (A) coupled process. During this process, once RNAPII transcribes the poly (A) signal there is a marked reduction in its processivity, which leads to pausing further downstream. This results in the endoribonucleolytic cleavage of the nascent transcript followed by polyadenylation of the upstream cleavage product and degradation of the downstream cleavage product.

Once the RNAPII reaches a pause signal (poly A) at the gene terminal, 3′ end processing and termination specific complexes such as CPSF and CstF are recruited. The CTD repeat is hyperphosphorylated at serine 2 at the gene terminus. Phospho-Ser7 (orange balloon) regulates the termination of snRNA genes by recruiting the Integrator complex.

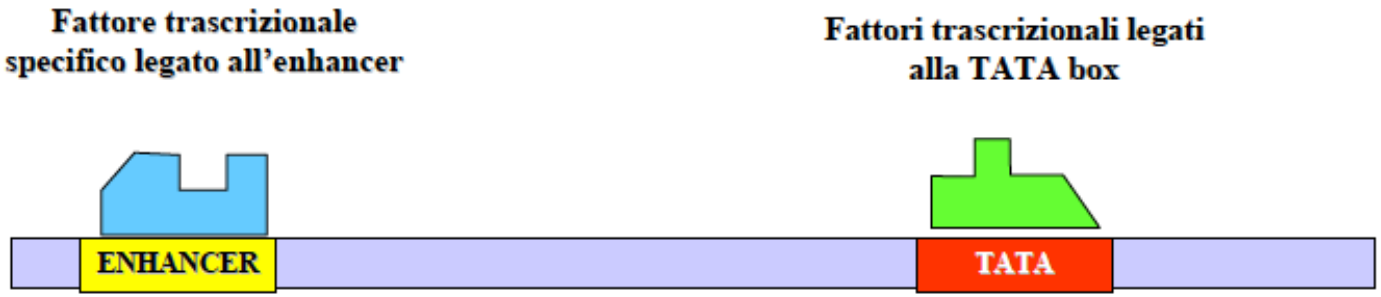
The end is a new beginning: gene looping

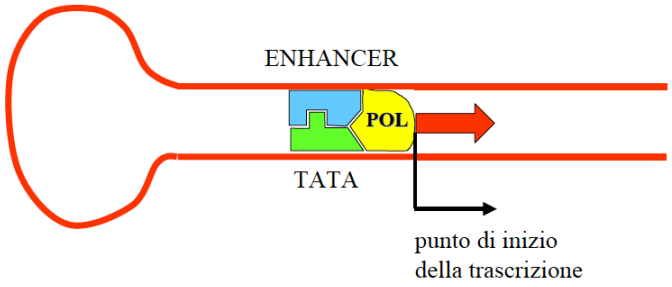


The terminator and promoter regions of a gene juxtapose during active transcription in order to facilitate RNAPII recycling and multiple rounds of transcription. The prebound GTFs and mediator in the reinitiation scaffold stabilized by the activator along with phospho-TFIIB interact with RNAPII and the termination complexes such as CPSF, CstF, and mediate such promoter–terminator contacts known as gene looping and thereby increase the efficiency of reinitiation by RNAPII.

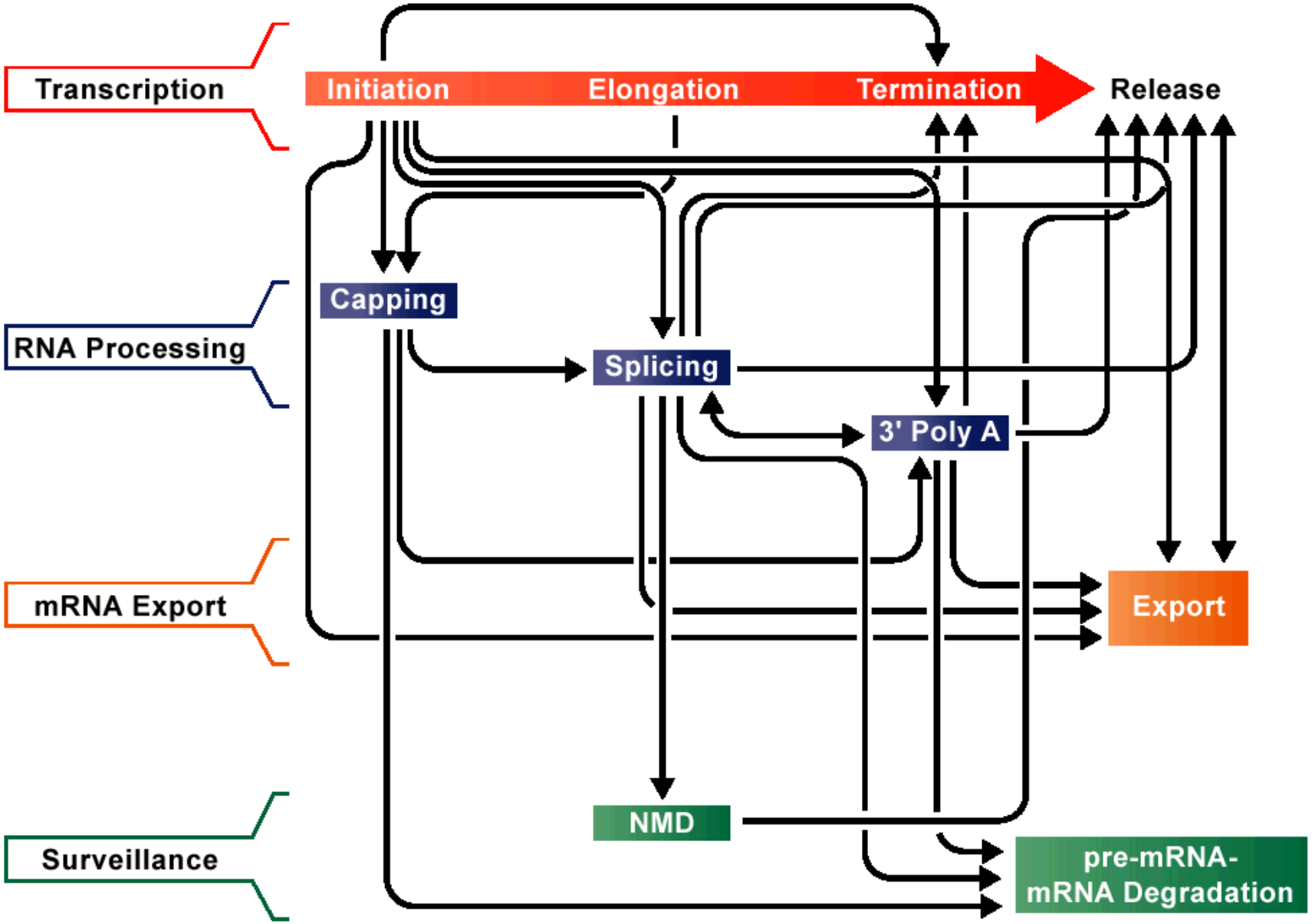
Il modello sinora illustrato garantisce un *livello basale* di attività trascrizionale. Il processo puo' venir favorito dall'interazione di GTFs *specifici* con sequenze a monte del Promotore.

es. Gli **Enhancer**





Network of *coupled* interactions in gene expression



The “mRNA Factory” model

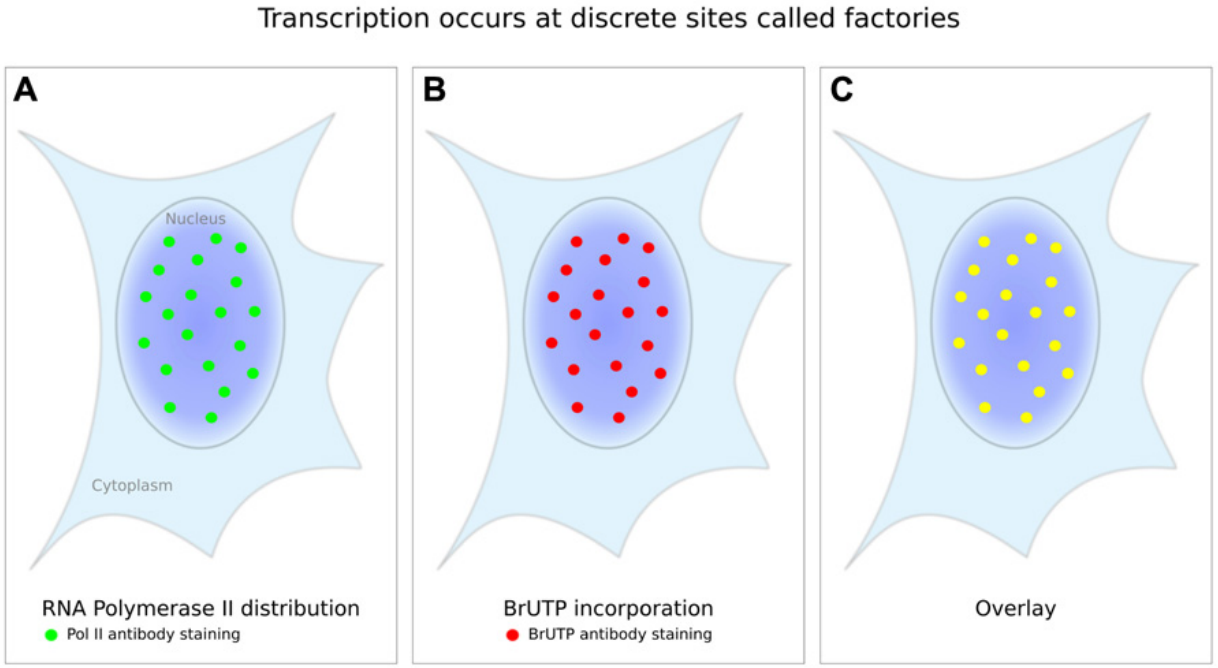
High throughput microarrays and next-generation sequencing technologies have revealed:

1) *temporal* coordination of gene transcription in response to developmental or environmental changes.

2) *spatially* coordination of gene transcription within each cell nucleus

When **RNA polymerase II** is detected by immunofluorescence a non-uniform staining pattern can be observed (*green* dots). **(B)** Labeling of **nascent RNA** by Br-UTP incorporation and subsequent immuno-staining (*red* dots) reveals a staining pattern that matches the polymerase staining as an overlay **(C)** shows (*yellow* dots).

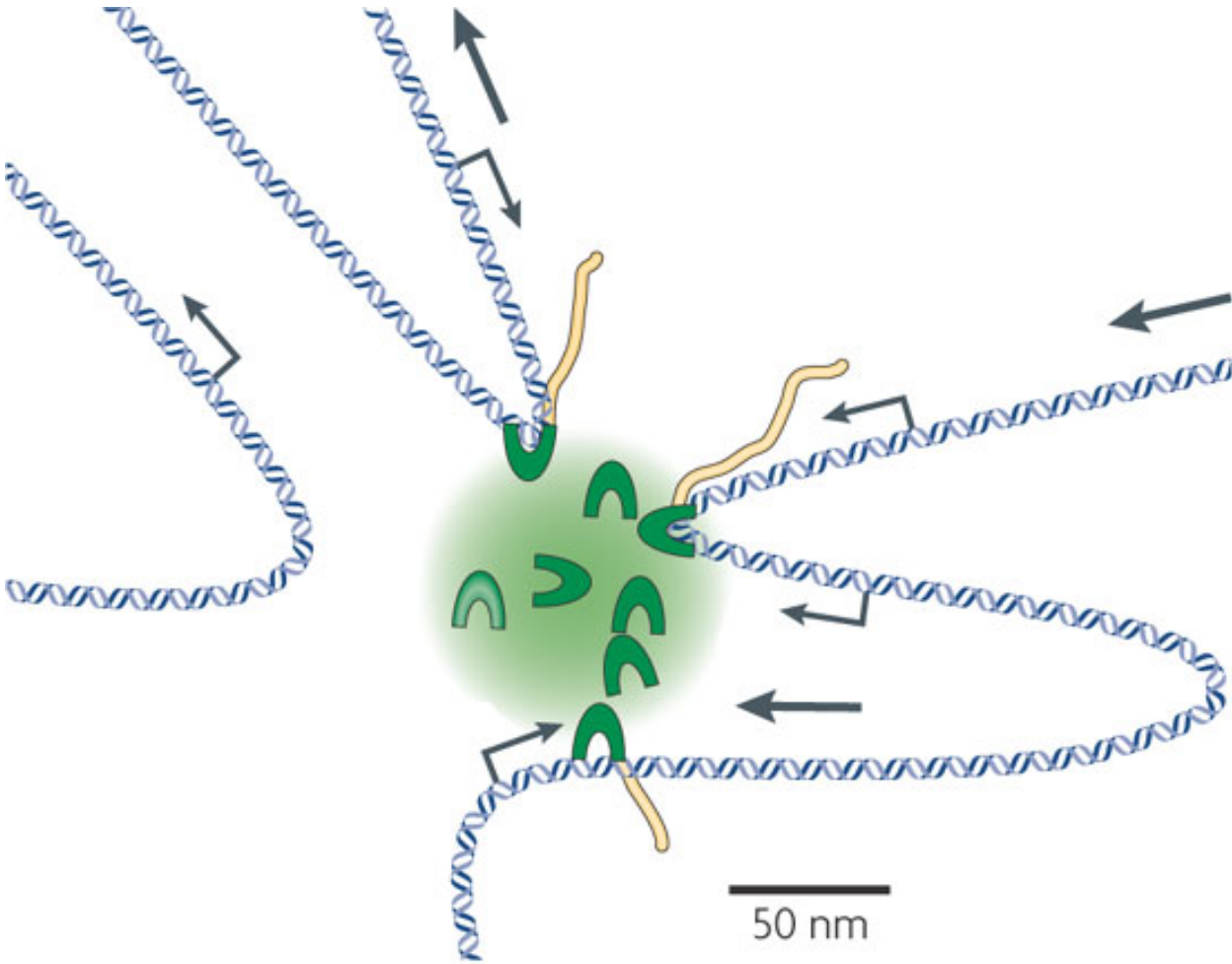
These discrete sites of active transcription are referred to as “***transcription factories***”.



Transcription occurs at discrete sites in the nucleus termed "***transcription factories***" where multiple active RNA polymerases are concentrated and anchored to a nuclear substructure.

It shows a ***transcription factory*** with a diameter of 70 nm that *contains eight RNA polymerase II* enzymes (green crescents). Genes are reeled through these polymerases (in the direction of the large arrows) as they are transcribed, and the nascent RNA (yellow) is extruded.

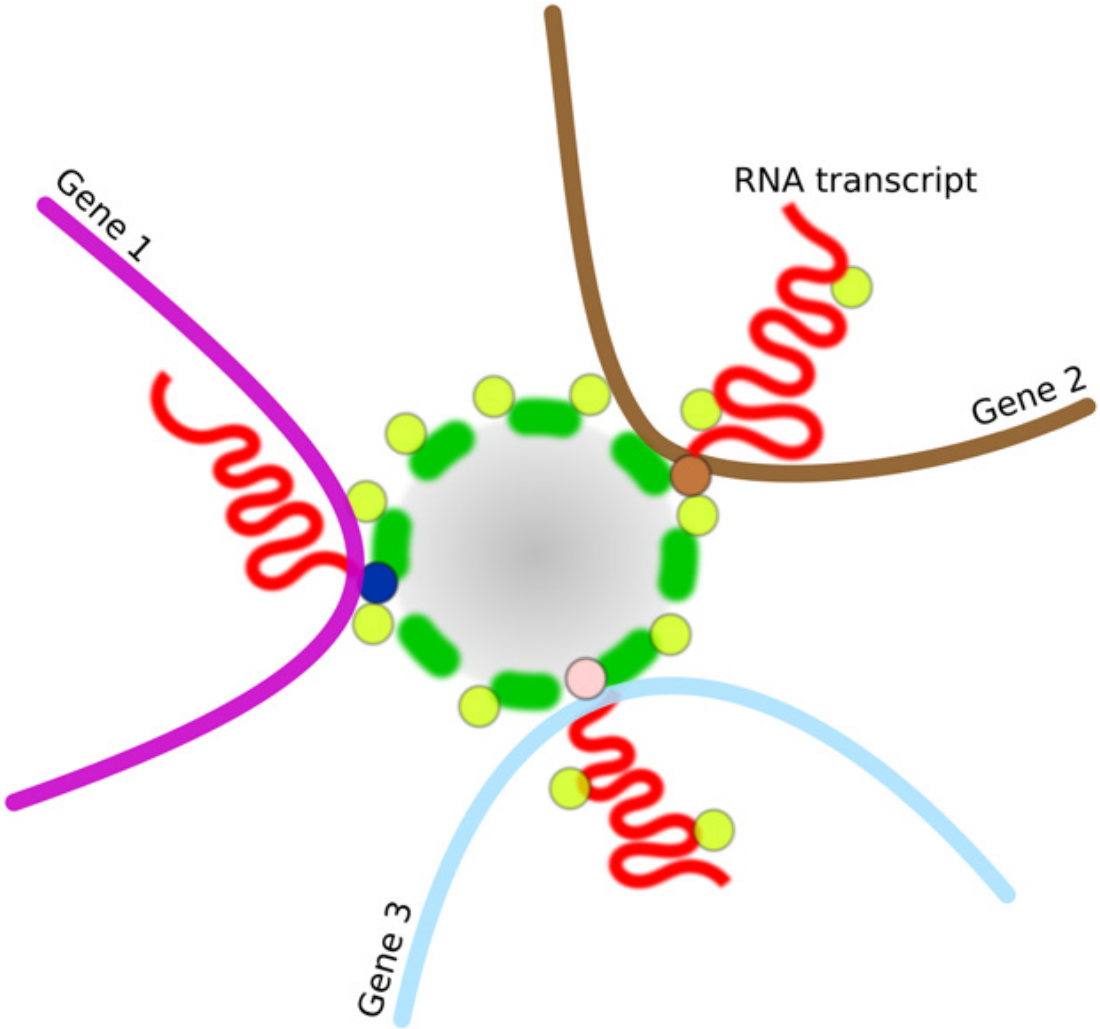
Genes from the *same* or from *different* chromosomes may associate with polymerases in the same factory. Small arrows indicate the direction of transcription at the transcription start site.

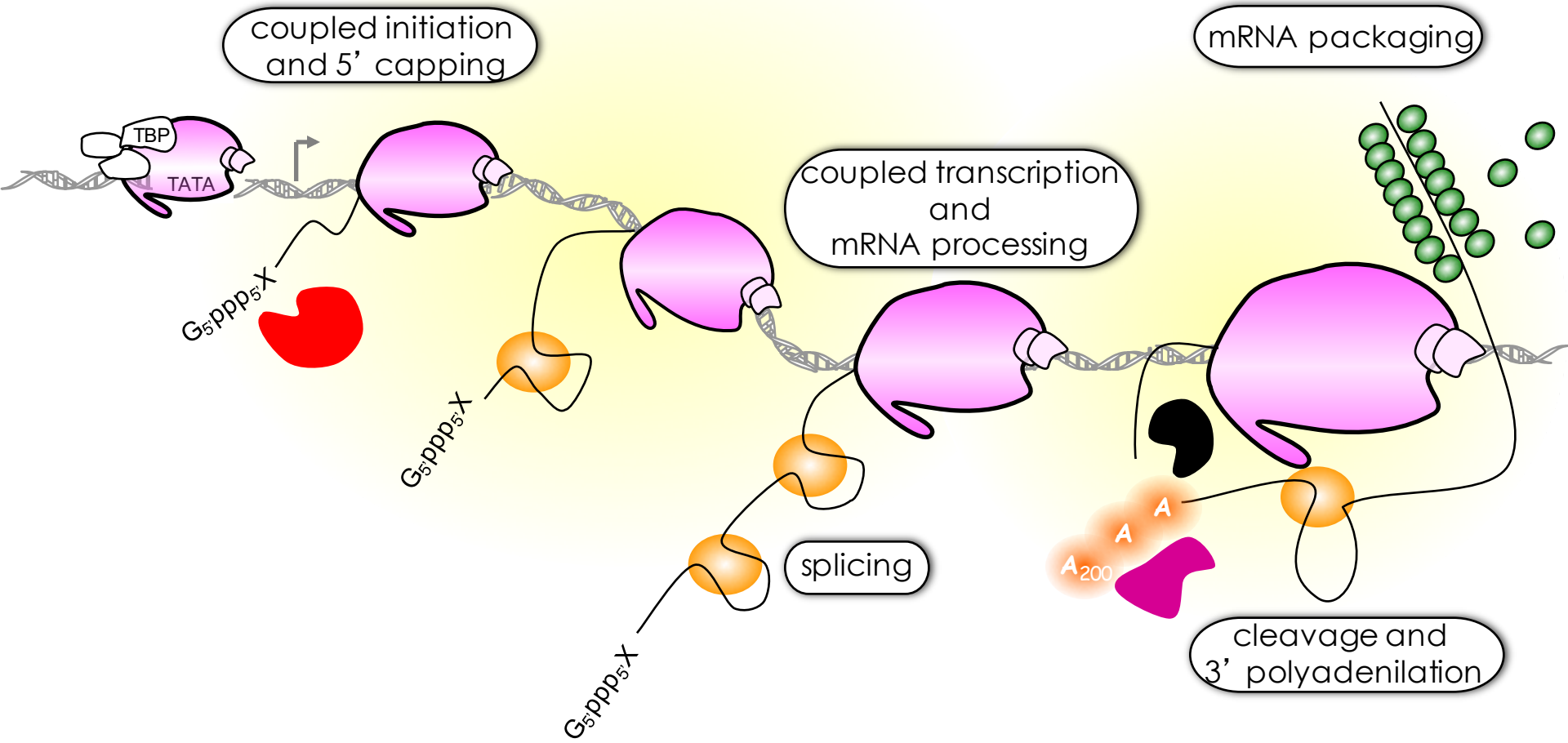


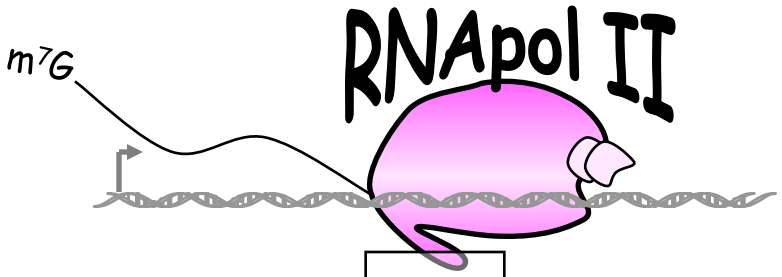
Nature Reviews | Genetics

Structure of a transcription factory

Each factory contains RNA polymerase II molecules which are located on the surface of a *protein-rich core* (87 nm in diameter, as determined by EFTEM in HeLa cells). These proteins include many factors involved in *transcription* such as co-activators, *chromatin remodelers*, transcription factors, histone modification enzymes, RNPs, RNA helicases, and *splicing* and *processing factors*. Multiple genes can be processed by the same factory (three are shown).

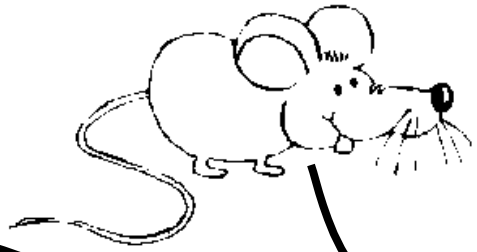



coupled initiation and 5’ capping
TBP
TATA
$G_{5'}ppp_{5'}X$
coupled transcription and mRNA processing
$G_{5'}ppp_{5'}X$
splicing
$G_{5'}ppp_{5'}X$
mRNA packaging
A
A
A_{200}
cleavage and 3’ polyadenilation

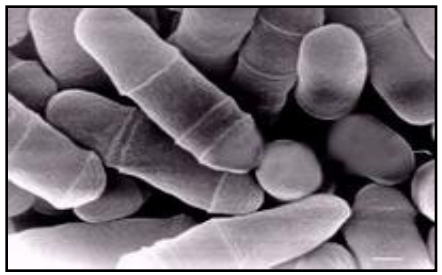


Carbossi Terminal Domain

Mouse

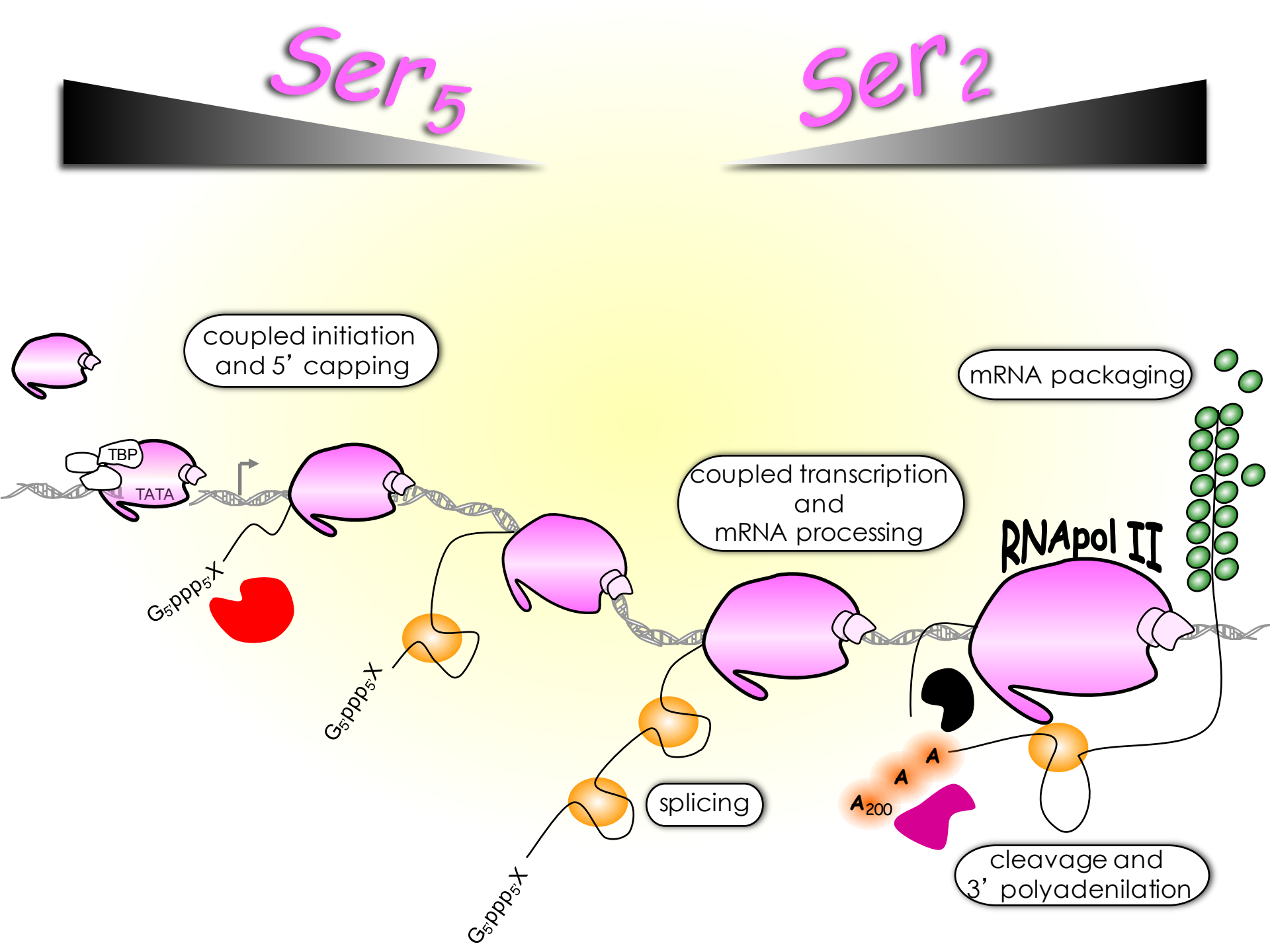


Yeast



	1	2	3	4	5	6	7
	Glu	Ala	Pro	Thr	Ser	Pro	Gly
1.	Phe	Gly	Val	Ser	Ser	Pro	Gly
2.	Phe	Ser	Pro	Thr	Ser	Pro	Thr
3.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
4.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
5.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
6.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
7.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
8.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
9.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
10.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
11.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
12.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
13.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
14.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
15.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
16.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
17.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
18.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
19.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
20.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
21.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
22.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
23.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
24.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
25.	Tyr	Ser	Pro	Thr	Ser	Pro	Gly
26.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
	Tyr	Ser	Pro	Lys	Gln	Asp	Glu Gln Lys His Asn Glu Asn Glu Asn Ser Arg

	1	2	3	4	5	6	7
	Glu	Gly	Ala	Met	Ser	Pro	Ser
1.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
2.	Tyr	Glu	Pro	Arg	Ser	Pro	Gly Gly
3.	Tyr	Thr	Pro	Gln	Ser	Pro	Ser
4.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
5.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
6.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
7.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
8.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
9.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
10.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
11.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
12.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
13.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
14.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
15.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
16.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
17.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
18.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
19.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
20.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
21.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
22.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
23.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
24.	Tyr	Thr	Pro	Thr	Ser	Pro	Ser
25.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
26.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
27.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
28.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
29.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
30.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
31.	Tyr	Ser	Pro	Ser	Ser	Pro	Arg
32.	Tyr	Thr	Pro	Gln	Ser	Pro	Thr
33.	Tyr	Thr	Pro	Ser	Ser	Pro	Ser
34.	Tyr	Ser	Pro	Ser	Ser	Pro	Ser
35.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
36.	Tyr	Thr	Pro	Thr	Ser	Pro	Ser
37.	Tyr	Ser	Pro	Ser	Ser	Pro	Glu
38.	Tyr	Thr	Pro	Ala	Ser	Pro	Lys
39.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
40.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
41.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
42.	Tyr	Ser	Pro	Thr	Thr	Pro	Lys
43.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
44.	Tyr	Ser	Pro	Thr	Ser	Pro	Val
45.	Tyr	Thr	Pro	Thr	Ser	Pro	Lys
46.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
47.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
48.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
49.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys Gly Ser Thr
50.	Tyr	Ser	Pro	Thr	Ser	Pro	Gly
51.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
52.	Tyr	Ser	Leu	Thr	Ser	Pro	Ala
53.	Ile	Ser	Pro	Asp	Asp	Ser	Asp Glu Glu Asn


Ser5
Ser2
coupled initiation and 5' capping
TBP
TATA
G5'ppp5'X
G5'ppp5'X
G5'ppp5'X
coupled transcription and mRNA processing
RNApol II
mRNA packaging
splicing
A
A
A200
cleavage and 3' polyadenilation