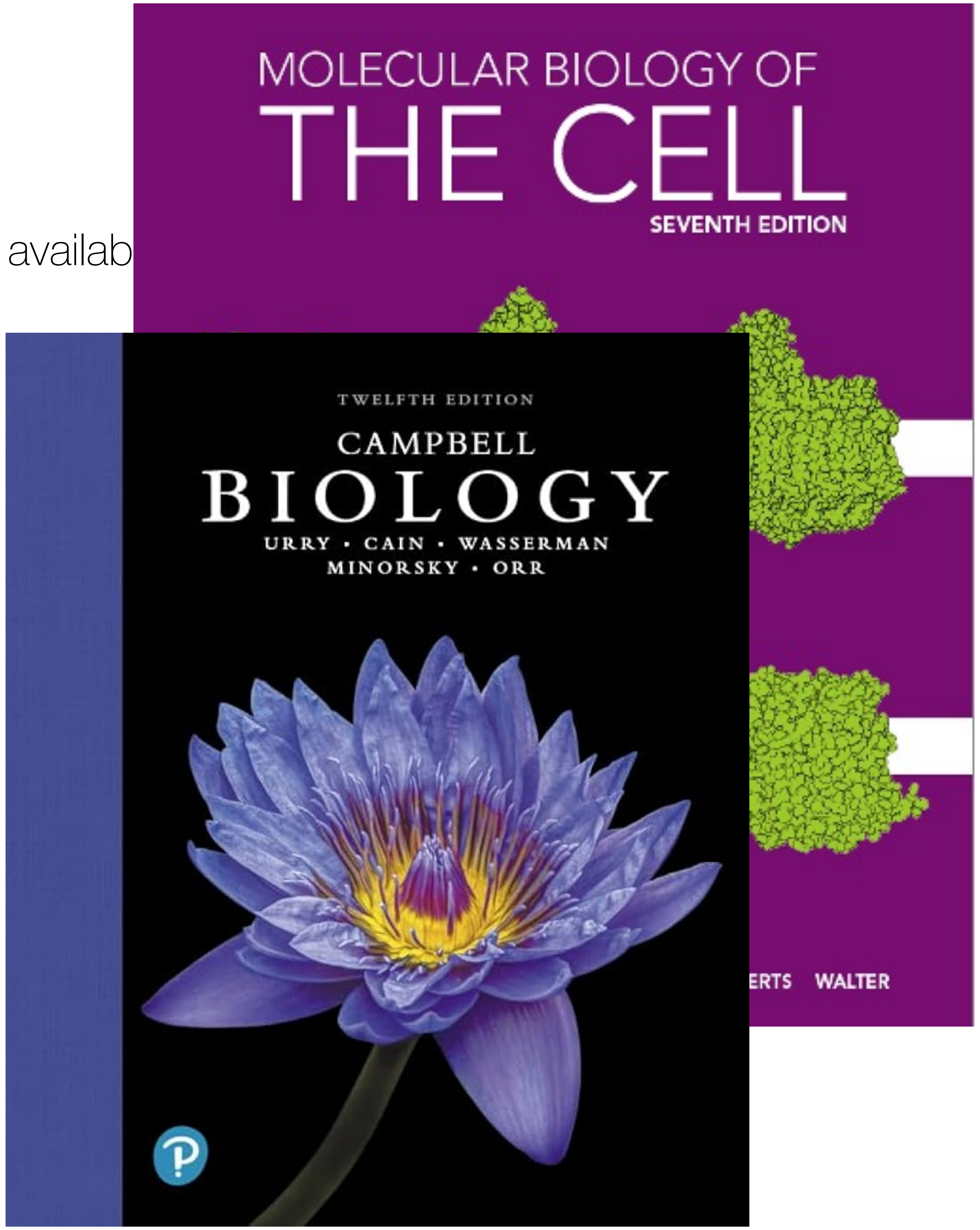
Welcome to Cellular and Molecular Biology I

BIO-205-1

Prof. Camille Goemans
EPFL-SV-GHI

General information about this course

- Slides and lectures are in **English**
- Reference Books are **Molecular Biology of the Cell (chapters 4-9) and Biology,** both availab at the library - not needed
- **Images** are primarily from both books
- All **references** available upon request
- Background is **Biologie Générale** with some overlap
- There will be **no recordings** of the lectures
- **Questions**:
    - during the break/after lecture
    - on forum (**ED** or Moodle) - last resort
- **Interesting resources**: https://www.biointeractive.org/classroom-resources



General information about this course

- **Exam** (English+French) on lectures+exercise material; **without** open book, similar to exercises
    - 30 **Multiple Choice** questions (only 1 good answer = 2 pts), no negative points
    - 30 **TRUE or FALSE** questions (+1 for right answers, -1 for wrong answers)
- Interactive **quizz** during the course: please participate
- Students' reps?

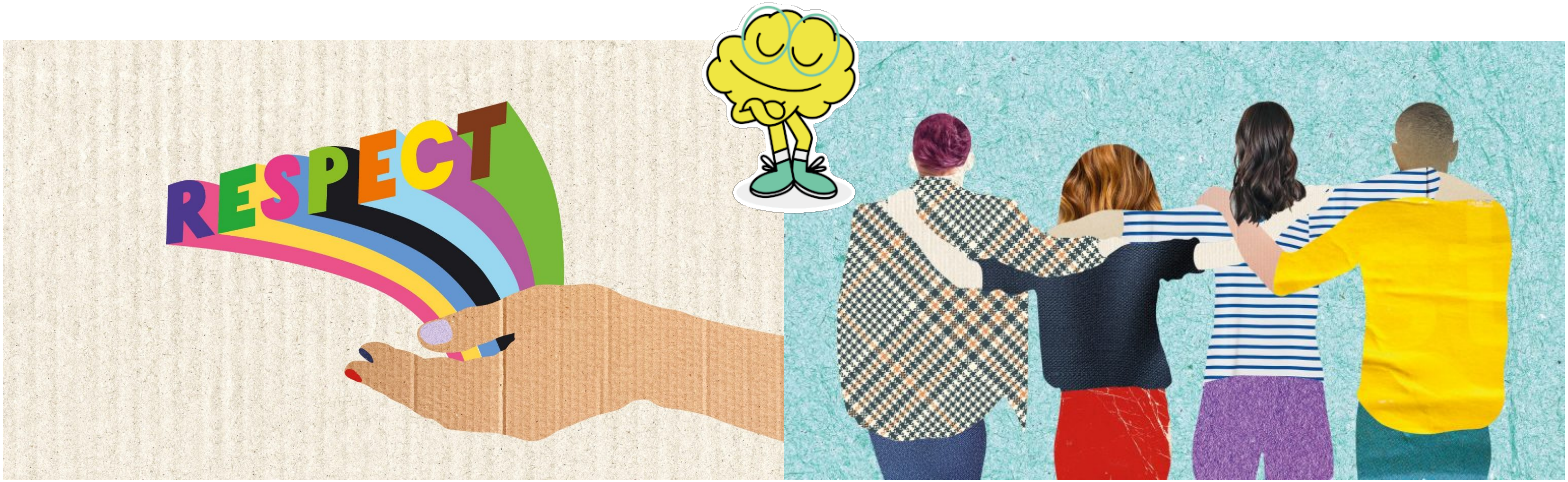
General information about this course

- Every **Friday** on **Moodle**:
  - **Slides** for the next class
  - **Exercises** for the next session
- Every **Tuesday** on **Moodle**:
  - **Answers** to yesterday's exercises

Exercise sessions with TAs

- Exercises (English): Mondays 17.15-19.00
- Room CE1100 : students A-F (last name)
- Room CE1101 : students G-O (last name)
- Room CE1105: students P-Z (last name)
- **Make sure to attend the exercise sessions as they prepare you for the exam !**

Let’s enjoy our time together



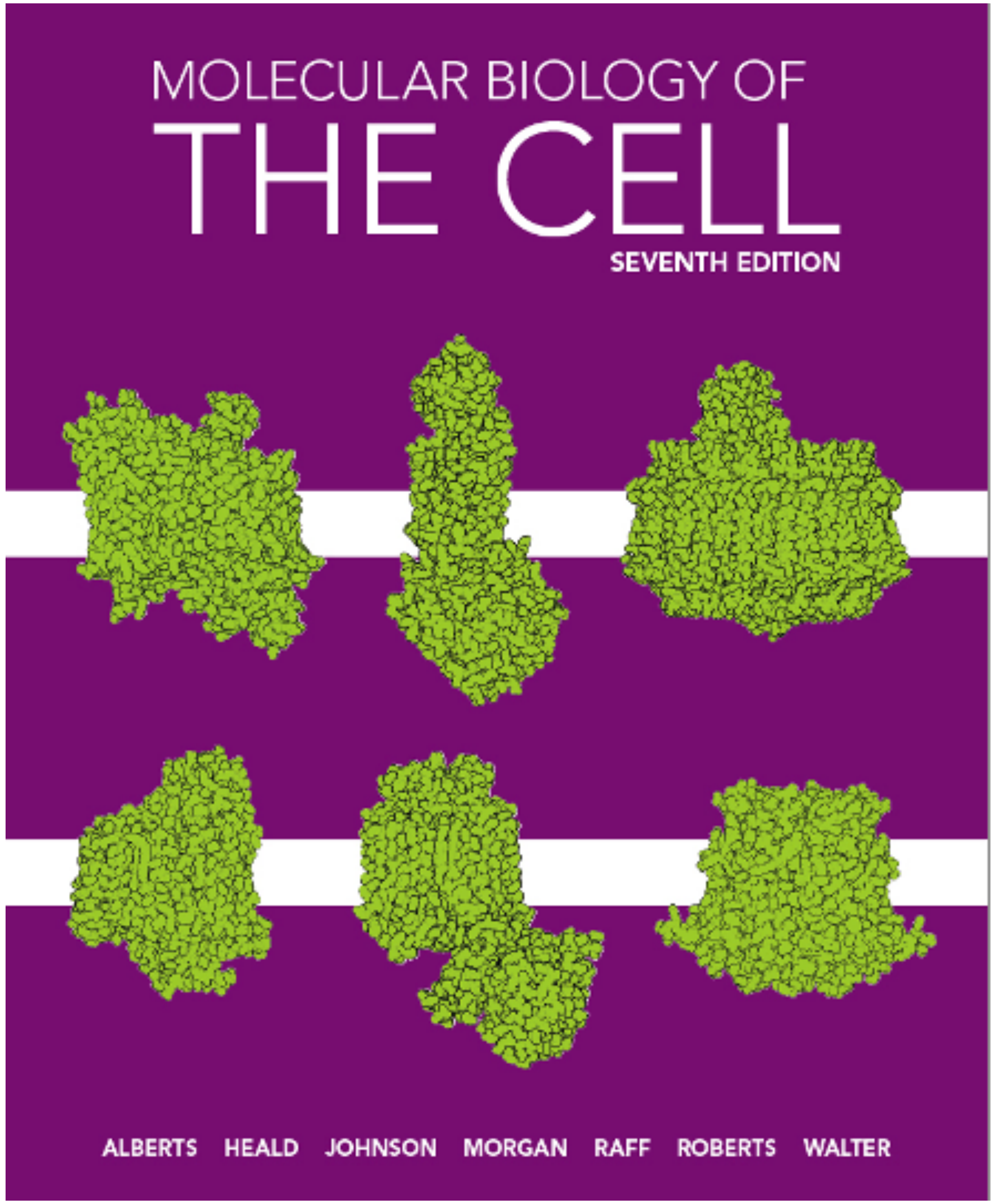
- I show up on time
- I prepare my classes so they are (hopefully) enjoyable
- I am open to feedback

- You show up on time/listen to the class
- You participate to the class when needed
- You provide me with feedback when necessary
- You don’t film the lectures

General objectives of this course

- Understand **DNA organization**, **gene expression** and **protein function**
- **General principles** and **experimental approaches**

Now, let's start!



Chapter 4

DNA, Chromosomes, and Genomes

Now, let's start!

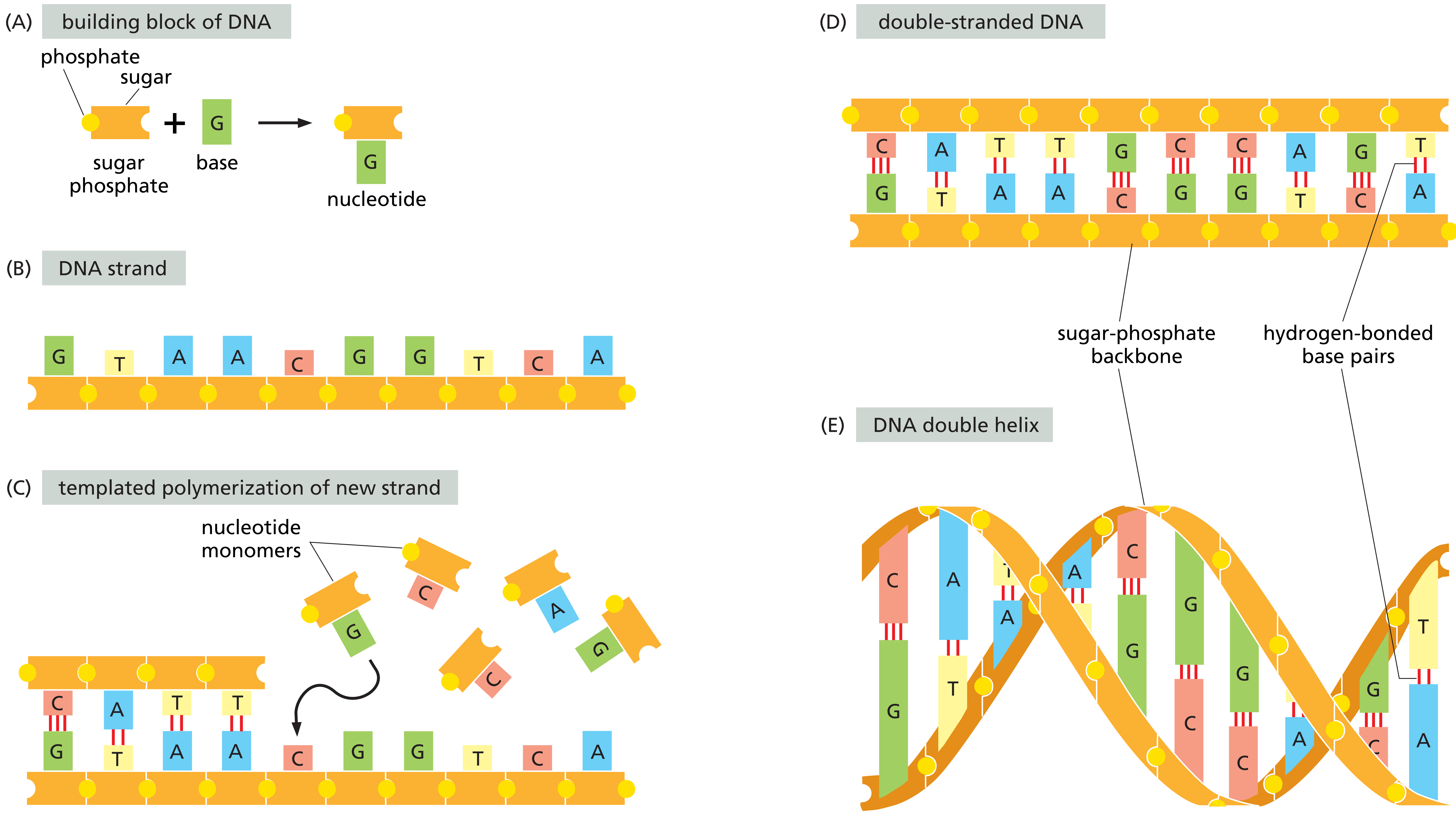
Plan

- Brief introduction on biology basics
- The structure and function of DNA
- Chromosomal DNA and its packaging into chromatin
- The effect of chromatin structure on DNA function

Brief introduction on biology basics

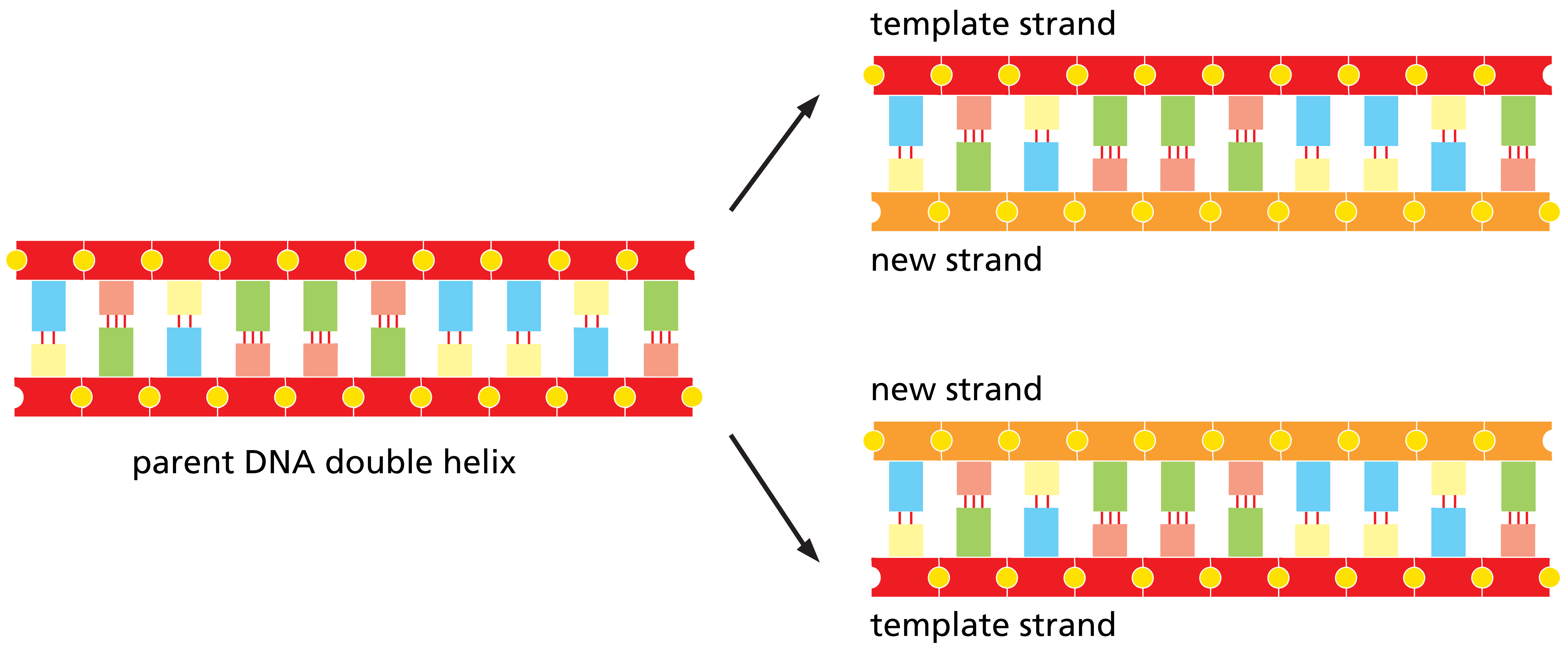
- Living organisms are **diverse**
- They are all made of **cells**
    - Membrane-enclosed units
    - Filled with a concentrated aqueous solution of chemicals
    - Ability to grow and divide in two
- All cells store their **hereditary information** in the same chemical linear code: **DNA**

What is DNA ?


(A)
building block of DNA
phosphate
sugar
G
sugar
phosphate
base
G
nucleotide
(B)
DNA strand
G T A A C G G T C A
(C)
templated polymerization of new strand
nucleotide
monomers
C A T T
G T A A C G G T C A
(D)
double-stranded DNA
C A T T G C C A G T
G T A A C G G T C A
sugar-phosphate
backbone
hydrogen-bonded
base pairs
(E)
DNA double helix

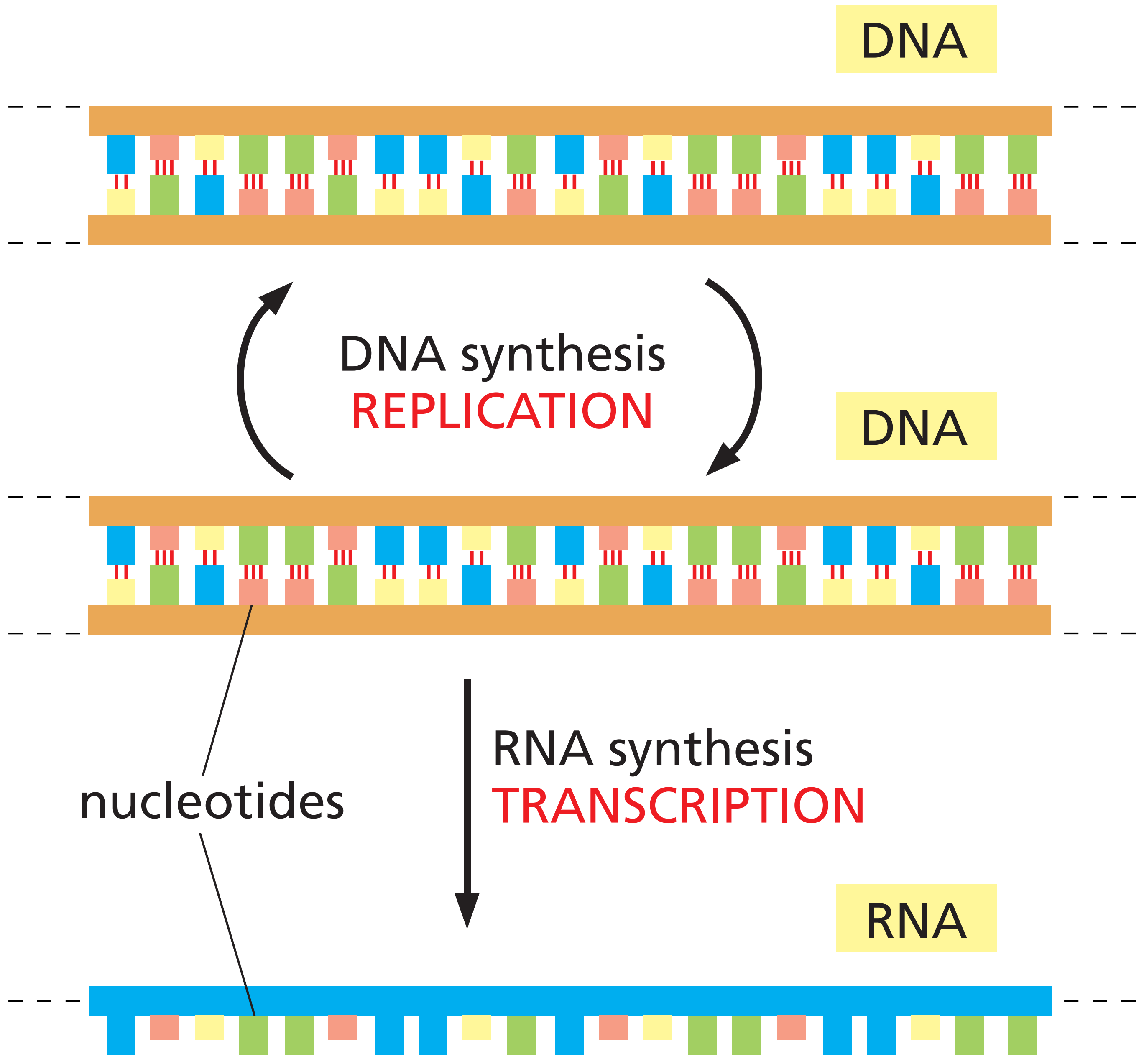
DNA replication

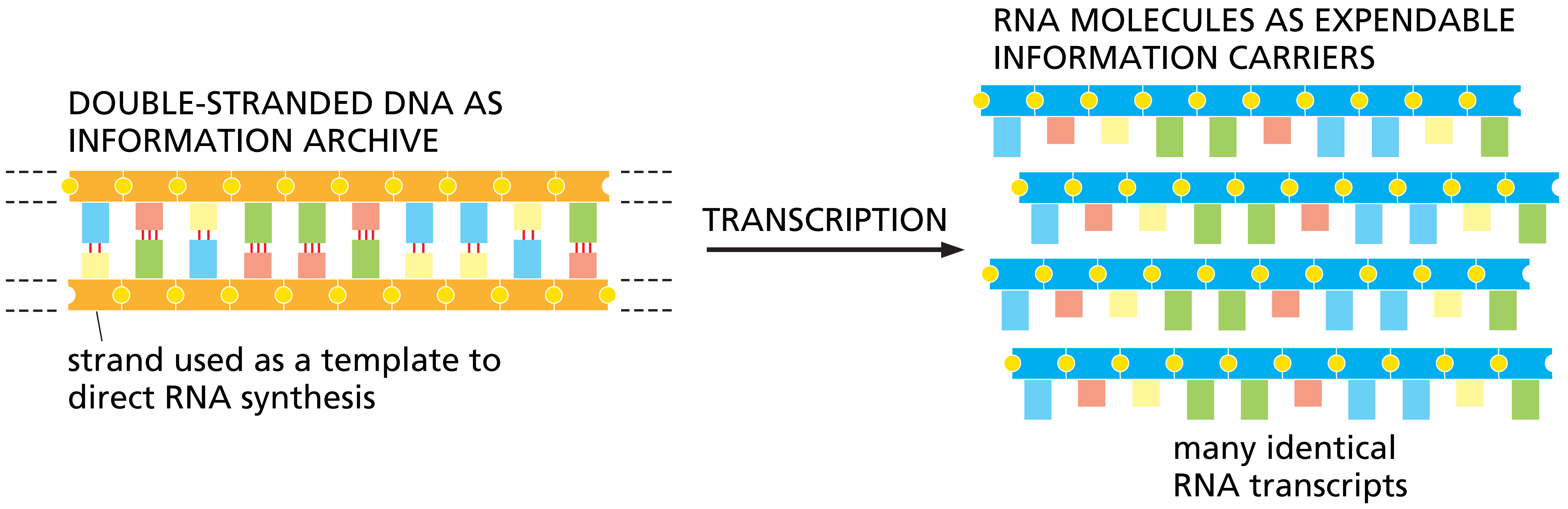
- All cells replicate their DNA by **templated polymerization**



Transcription

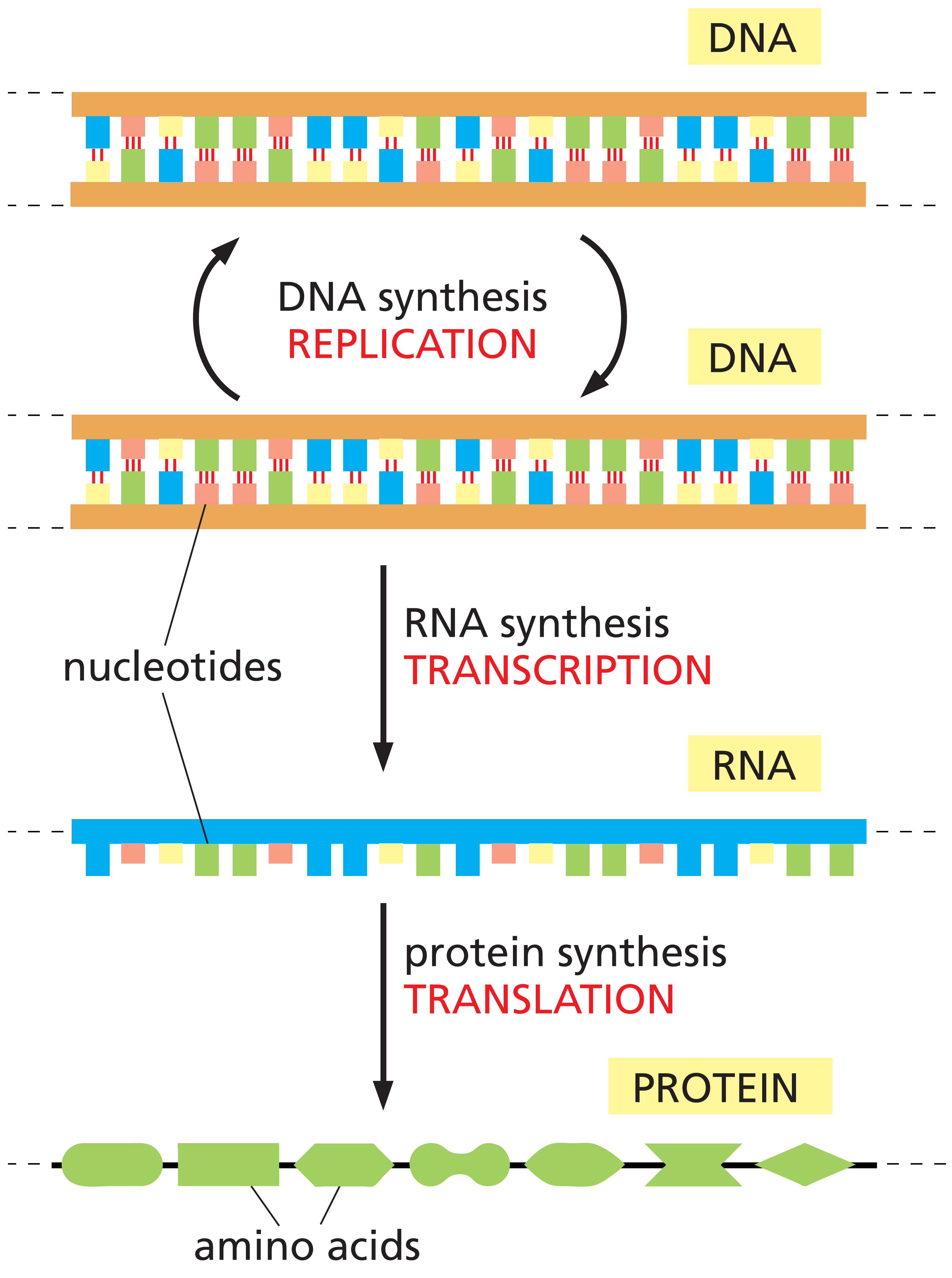
- All cells transcribe portions of their **DNA** into **RNA**





Translation

- All cells **translate RNA into proteins** the same way

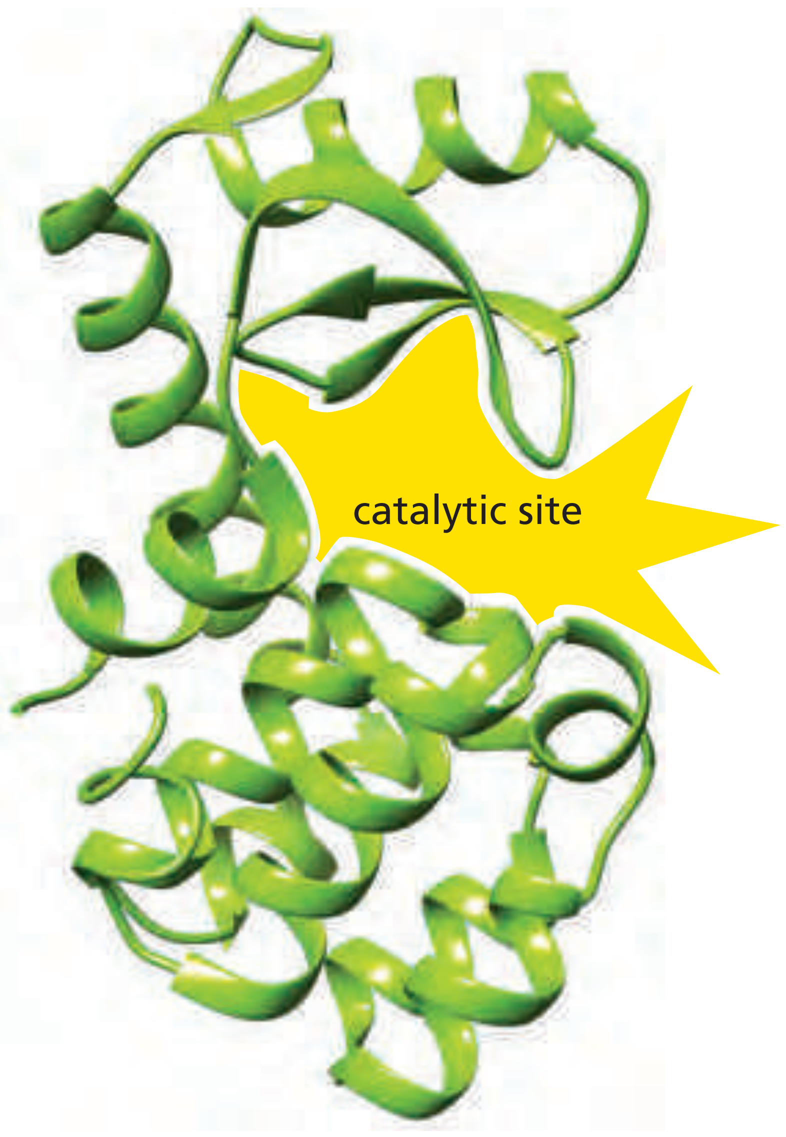


What are proteins?

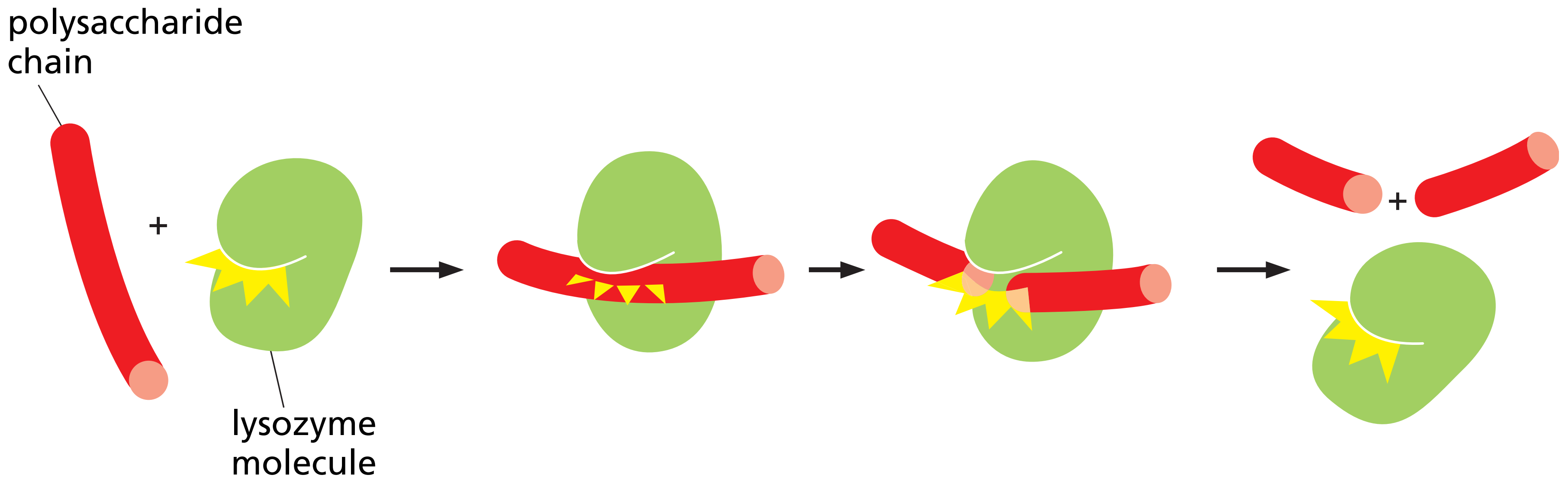
Proteins are

- Long polymers of **amino acids** (20 types) = polypeptides
- Important **3D structure** linked to their function
- Can act as **enzymes, structural function, movement generation, signals sensors,**…

As example



(A) lysozyme



(B)

Figure 1–7 **How a protein molecule acts as a catalyst for a chemical reaction.** (A) In a protein molecule, the polymer chain folds up into a specific shape defined by its amino acid sequence. A groove in the surface of this particular folded molecule, the enzyme lysozyme, forms a catalytic site. (B) A polysaccharide molecule *(red)*—a polymer chain of sugar monomers—binds to the catalytic site of lysozyme and is broken apart, as a result of a covalent bond-breaking reaction catalyzed by the amino acids lining the groove (see Movie 3.9). (PDB code: 1LYD.)

Plan

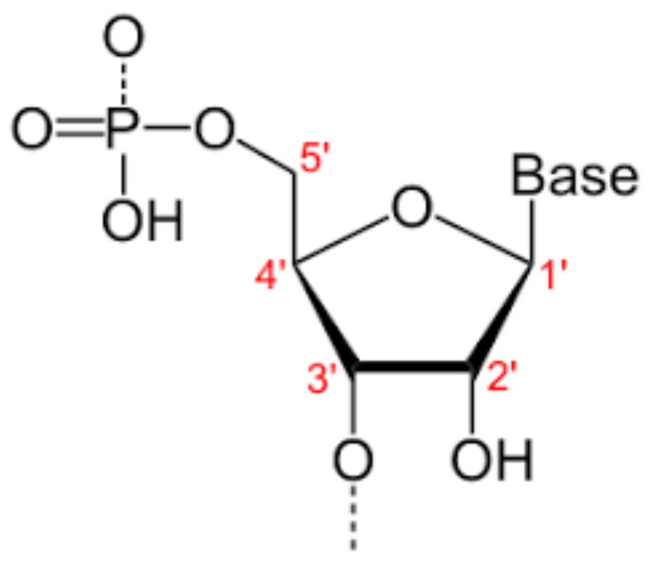
- Brief introduction on biology basics
- <u>The structure and function of DNA</u>
- Chromosomal DNA and its packaging into chromatin
- The effect of chromatin structure on DNA function

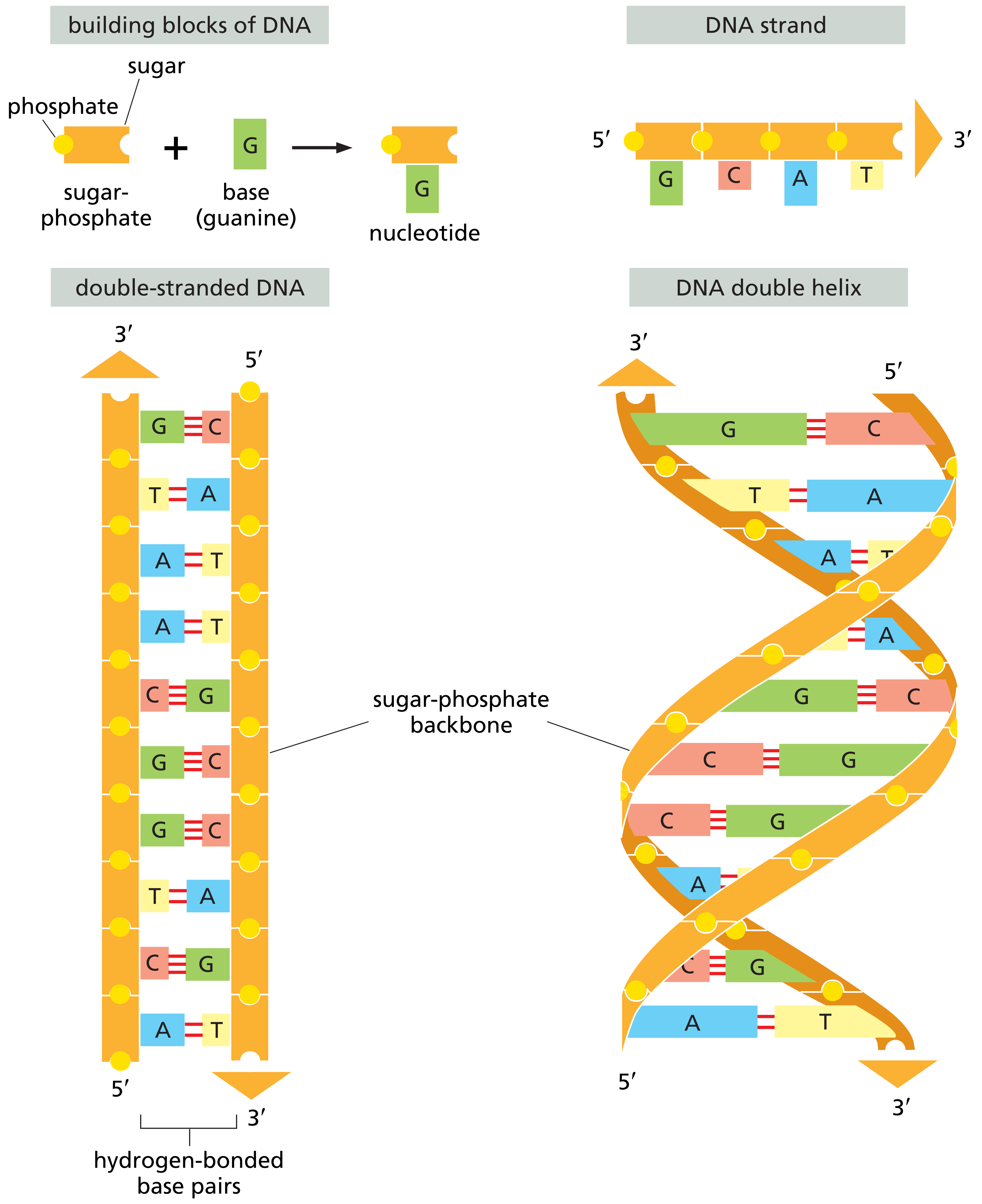
The structure of DNA

- DNA is passed on from one cell to its daughter cells at **cell division** (remember **cell cycle** in general biology)
- Instructions are stored as **genes**, which define the characteristics of a **species**
- In the 1940s, it was difficult to imagine that a molecule as **simple** as DNA could be the genetic material
- In the 1950s, x-ray diffraction analysis showed it had **two strands and formed a helix**
- 1953: **model of DNA** by Watson and Crick

The chemical structure of DNA

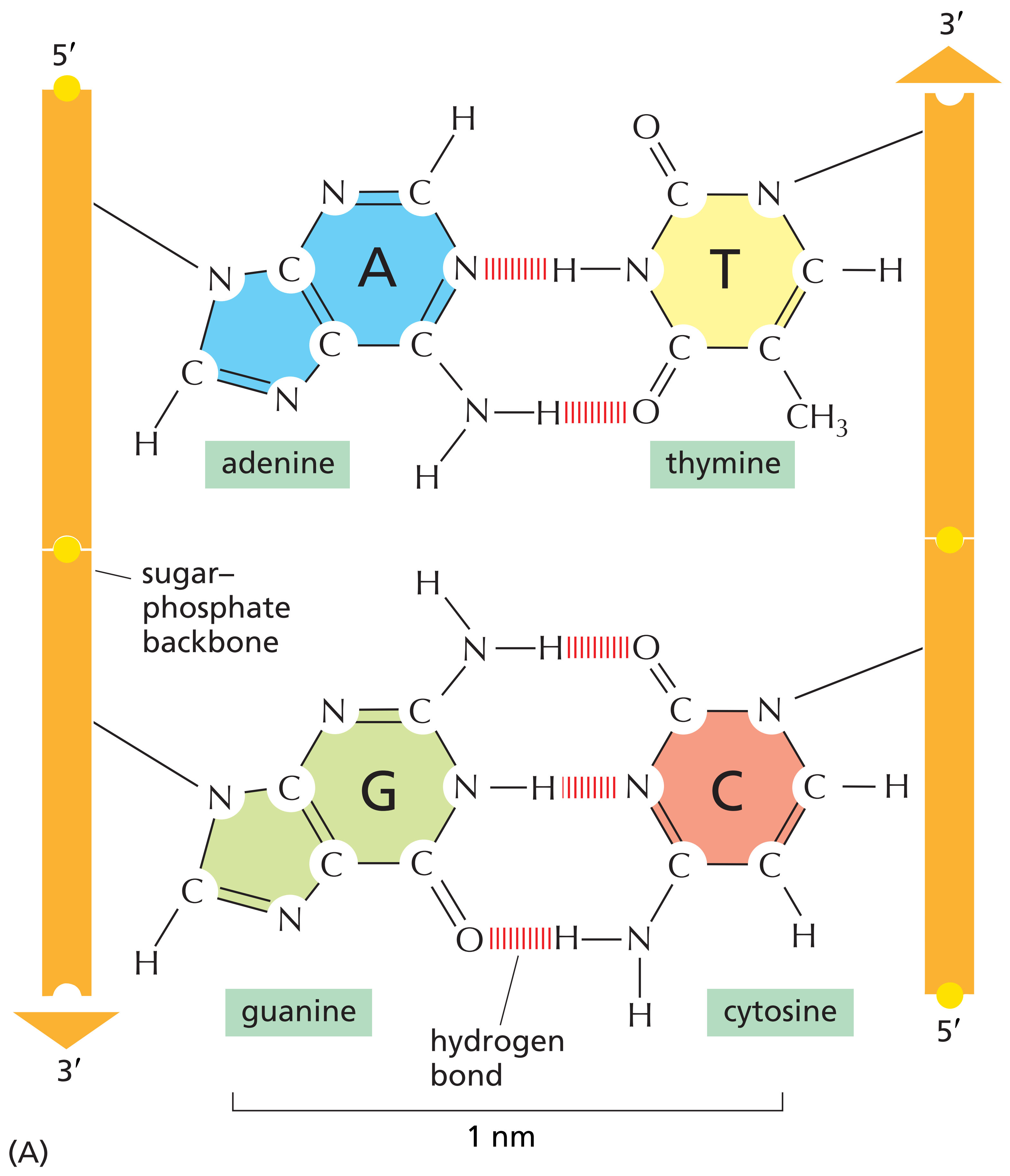
- **DNA** = Deoxyribonucleic Acid
- **Nucleotides** are the building blocks of DNA
- **Nucleotides** are formed by a **sugar** (deoxyribose) linked to a **phosphate** group and a nitrogen-containing **base** (adenine, cytosine, guanine, thymine)
- **Covalent link** to form the **sugar-phosphate backbone** = DNA strand
- Two long **polynucleotide chains**
- **Antiparallel** to each other (5’ end and 3’ end can be distinguished)
- Linked by **hydrogen bonds**





The chemical structure of DNA

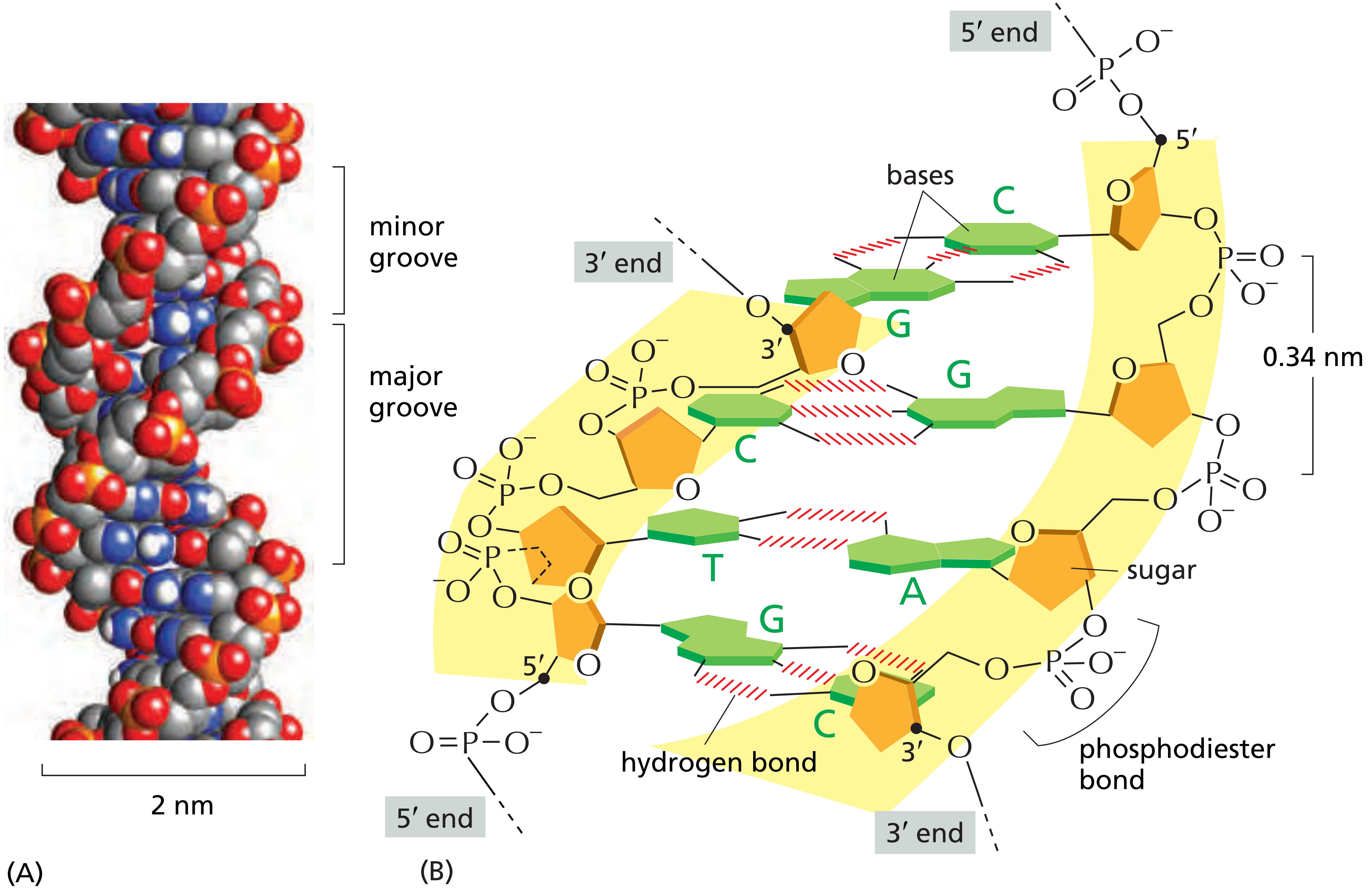
- A pairs with T; and G with C
- 2-ring base = purine
- 1-ring base = pyrimidine



! You don’t need to be able to draw them, but you should recognise them

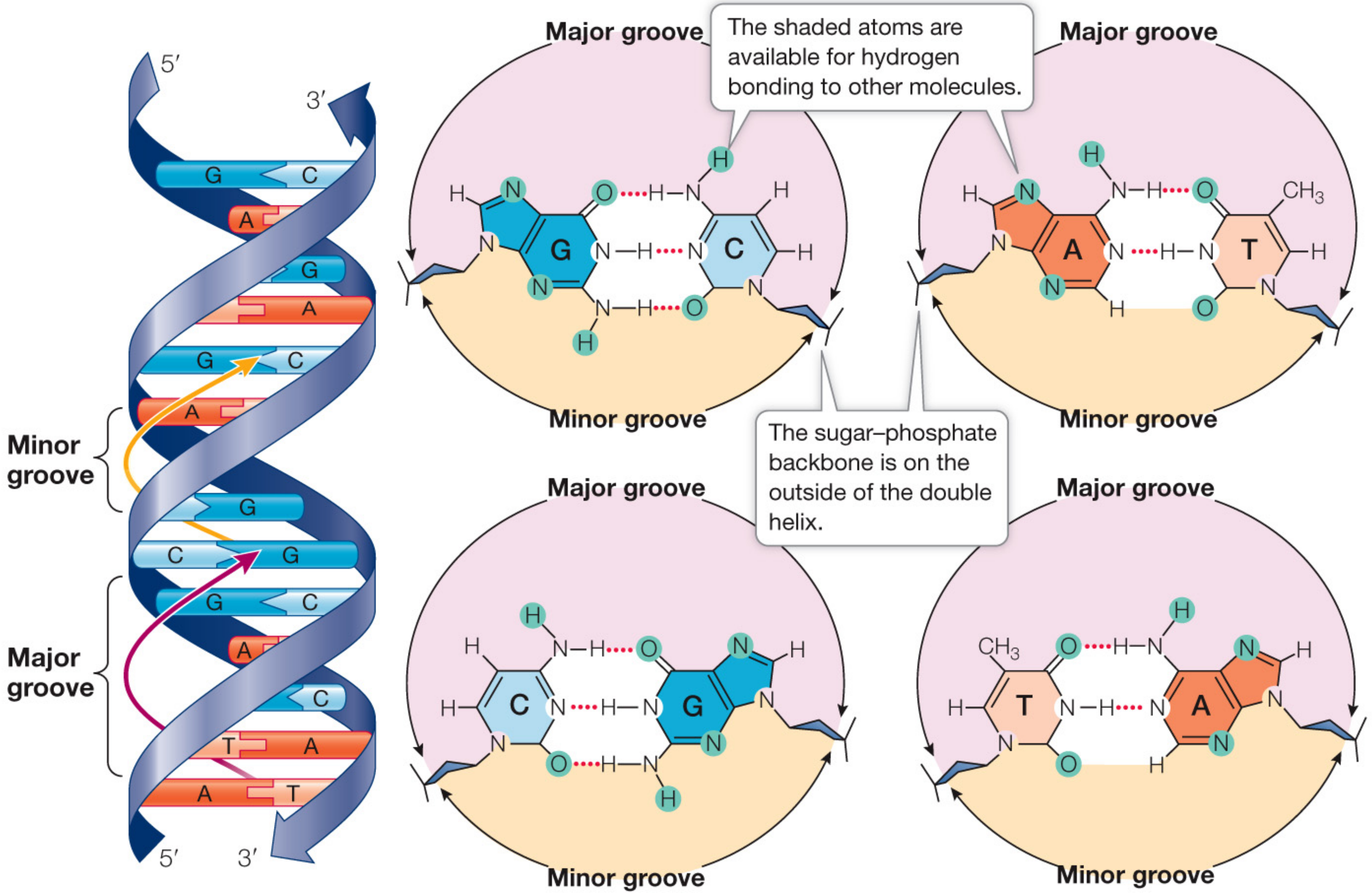
The chemical structure of DNA

- The two backbones wind around each other to form a **helix** with a turn per 10 base pairs



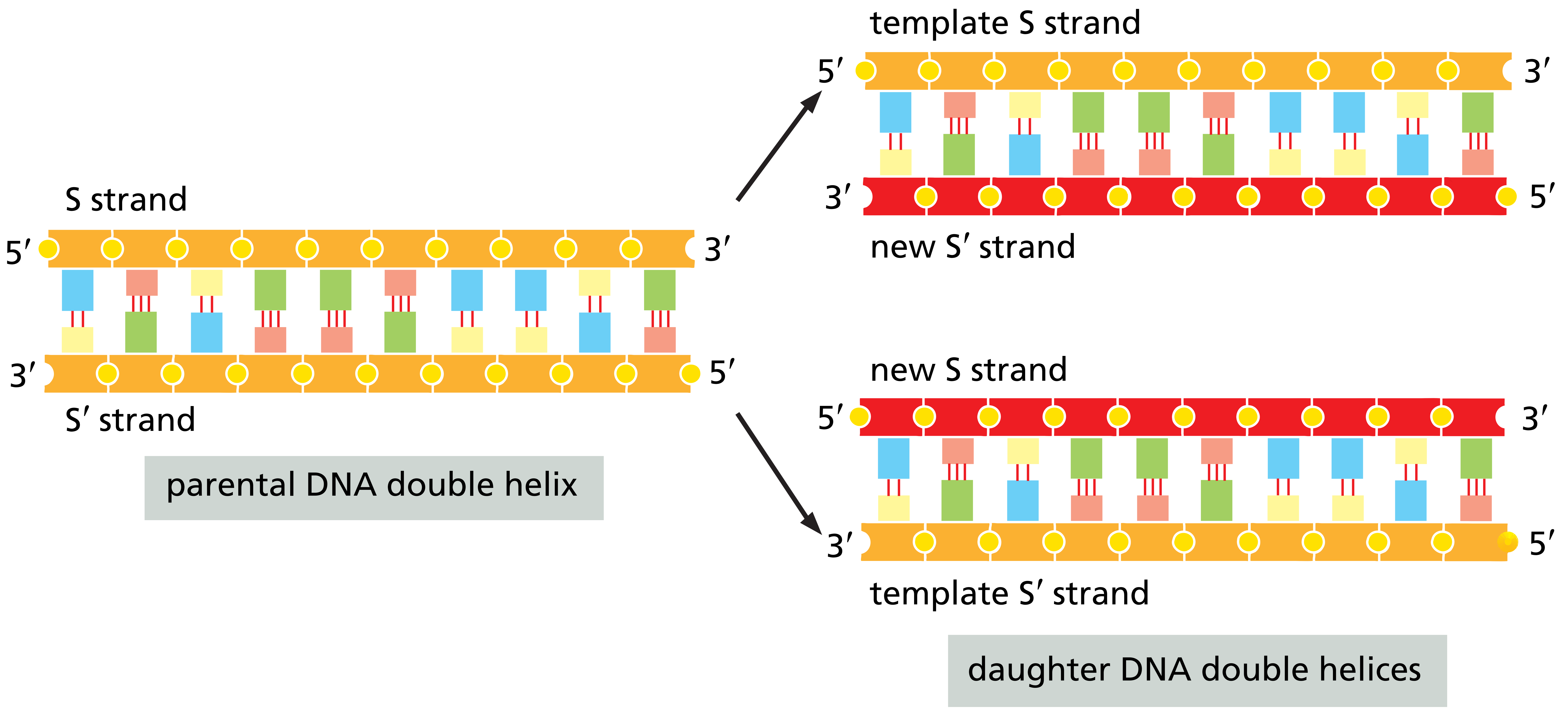
The chemical structure of DNA

- Major and minor grooves of the DNA



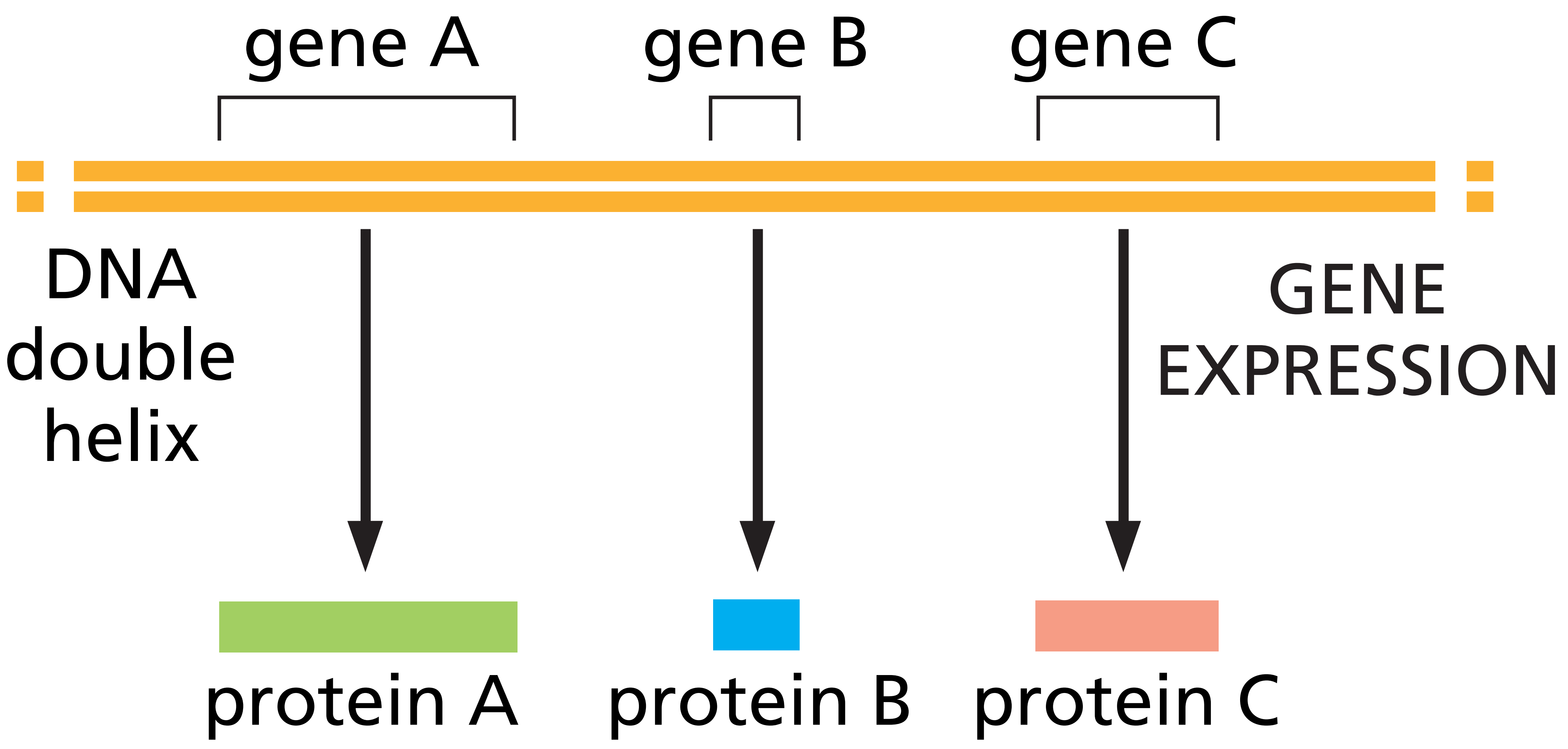
LIFE: THE SCIENCE OF BIOLOGY 11e, Figure 13.7

How is information copied from generation to generation?



How is information stored?

- **Genes** encode the information to make **proteins**
- If genes are made of DNA, DNA can be somehow turned into proteins
- **Genetic code**: correspondance between DNA and amino-acid
- The complete store of information of an organism is called **genome**

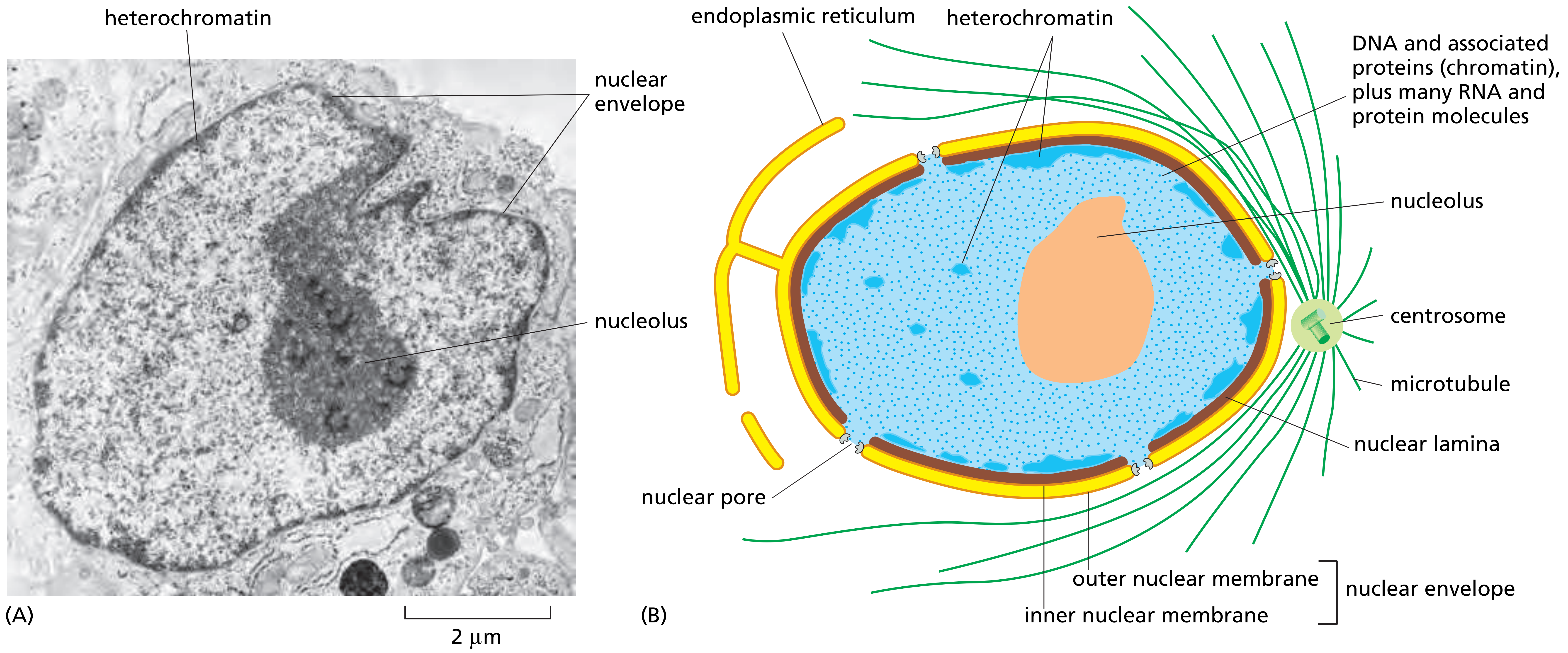


Plan

- Brief introduction on Biology basics
- The structure and function of DNA
- <u>Chromosomal DNA and its packaging into chromatin</u>
- The effect of chromatin structure on DNA function

Chromosomal DNA and chromatin

- In Eukaryotes, DNA is packed in a **nucleus** (10% of the volume of the cells), separated by **nuclear envelope**
- Nuclear envelope is a **two concentric lipid bilayer membranes** punctured by large **nuclear pores**
- Mammalian **DNA is 2-3 m** long but the nucleus has a **diameter of ~ 10 um**

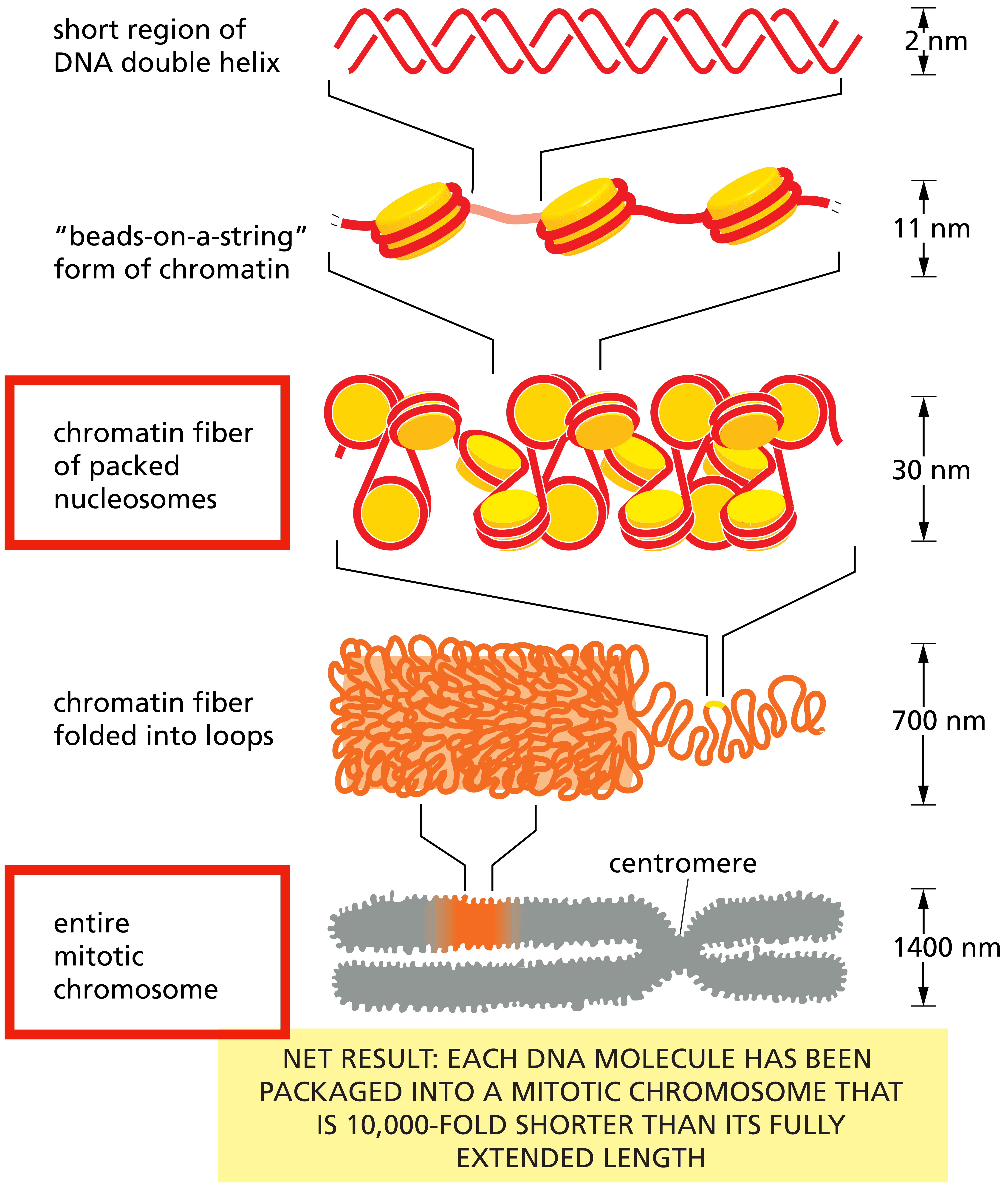


How is DNA packed?

"Normal state"

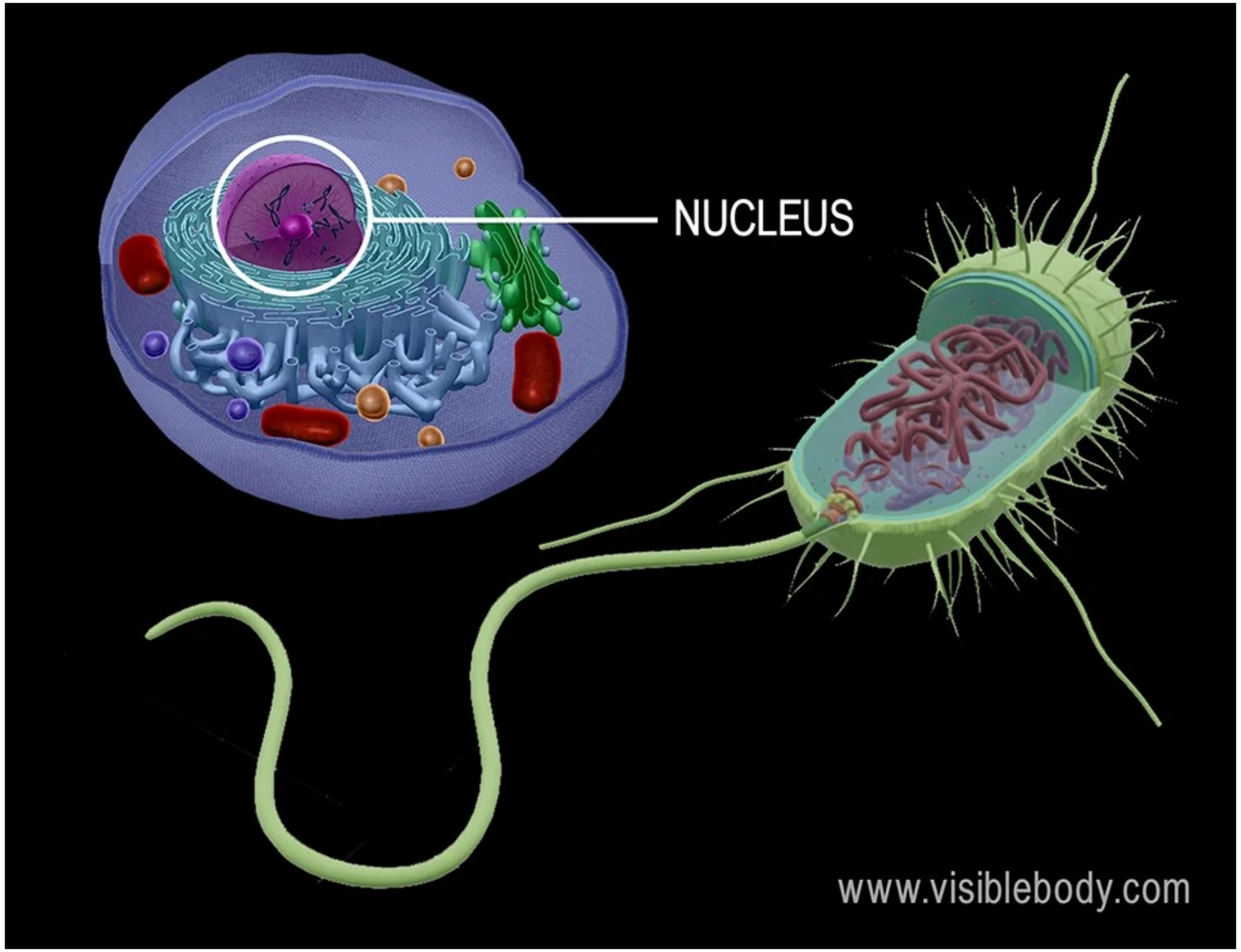
- Level 1: DNA - **double helix**
- Level 2: DNA and proteins form **chromatin**
- Level 3: Chromatin is condensed into **chromosomes**
- Each **chromosome** is a single long DNA molecule with proteins that pack it into the right structure
- Chromosomes reach their highest level of compaction during **mitosis**
- **Bacteria** do not have a nucleus, no chromatin and typically have one chromosome (different from the ones in eukaryotes)

At division only



Prokaryotes and eukaryotes

- **Prokaryotes:**
  - **Circular** DNA molecule
  - Associated with a **few** proteins
  - Inside the **cell** (no nucleus)
- **Eukaryotes:**
  - **Linear** DNA molecules
  - **Large amount** of proteins
  - Inside the **nucleus**

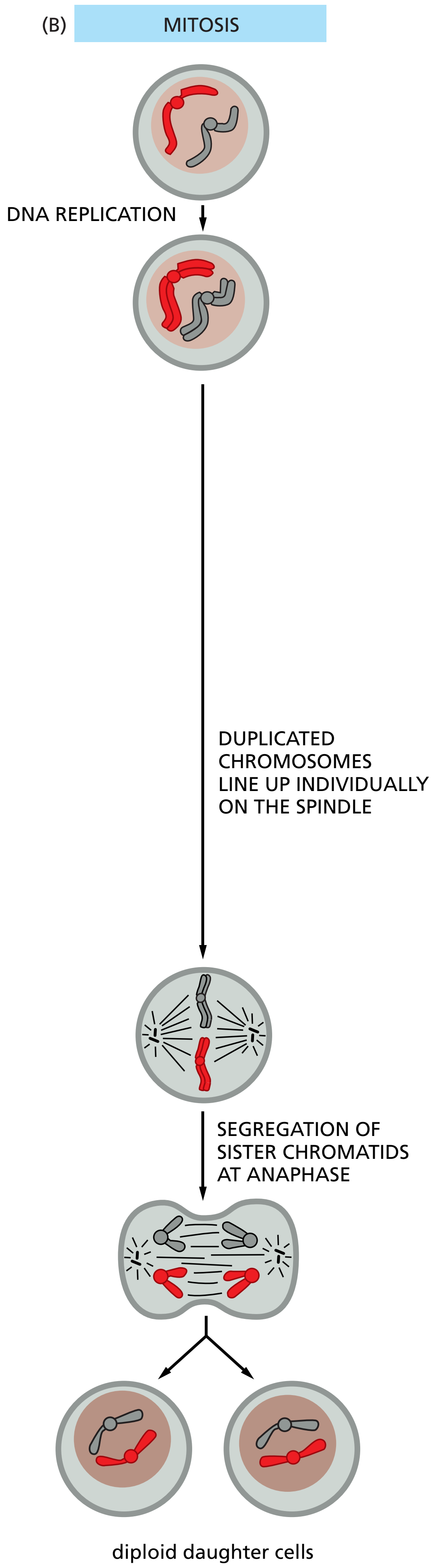


Human chromosomes

- Human cells (except gametes and highly-specialised cells like red-blood cells) have **two copies of each chromosome** (one from the father, one from the mother)
- The maternal and paternal chromosomes are called **homologous chromosomes**
- The **sex** chromosomes in males are **non homologous**
- Each human cell comprises **46 chromosomes: 22 pairs of autosomes and 2 sex chromosomes**

Human chromosomes

- Reminder - number of chromosomes



- Each human cell comprises **46 chromosomes:** 22 pairs of homologous autosomes and 2 sex chromosomes

- When those are **replicated** (right before division), there are still **46 chromosomes** (each with **2 chromatids**)

Human chromosomes

- Two DNA molecules produced during **DNA replication,** each called a **sister chromatid**
- Linked at the **centromere** (highly-condensed regions)

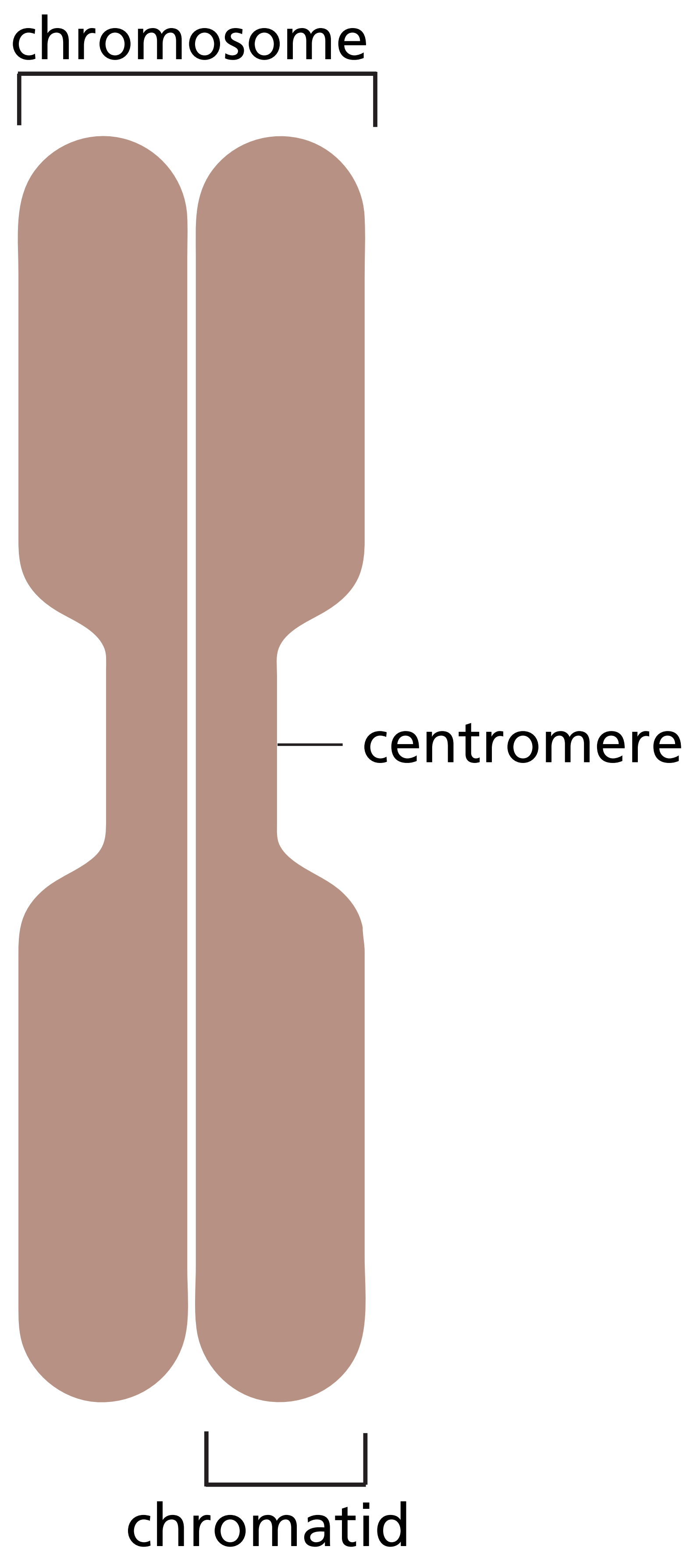
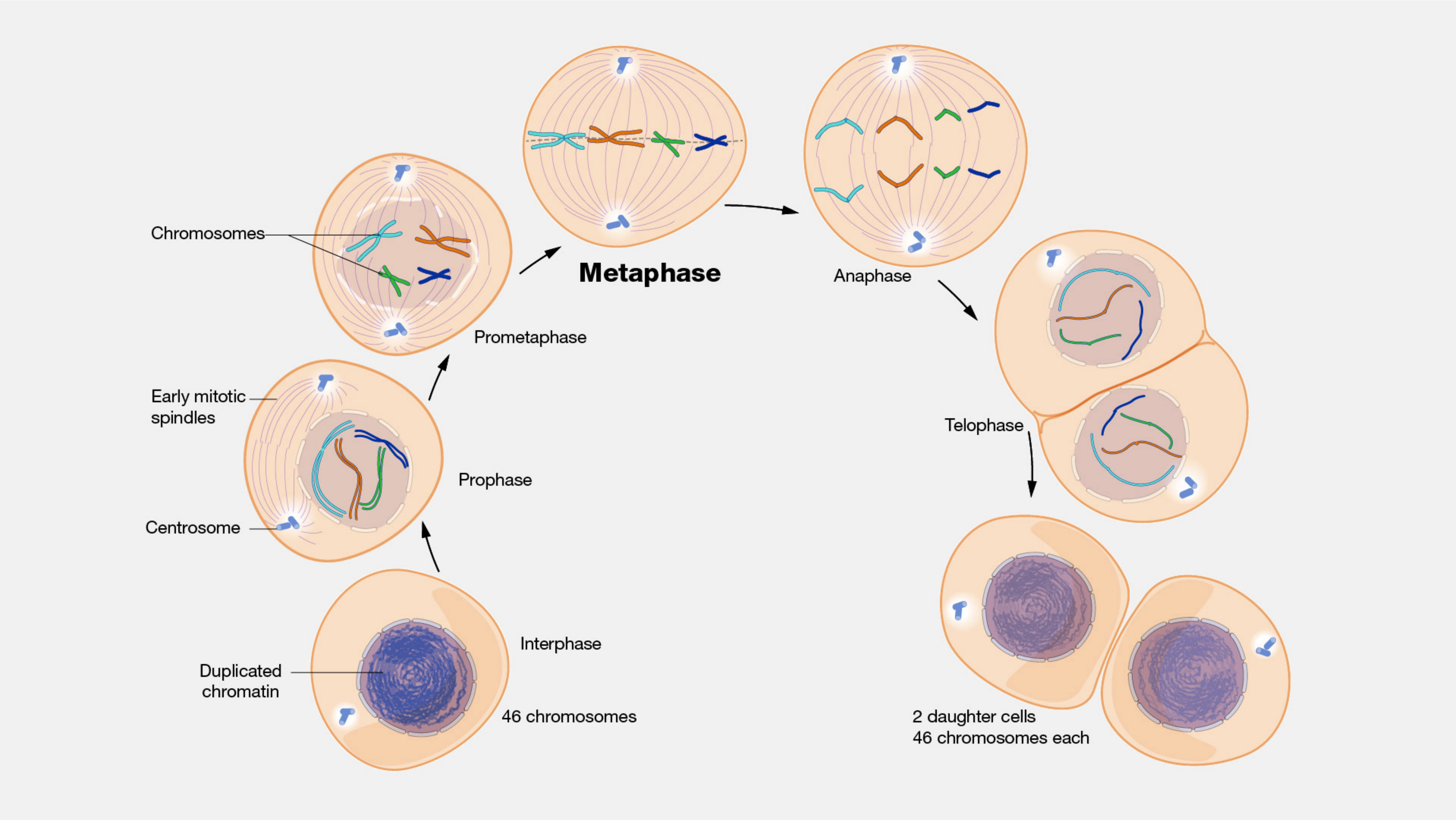


Figure 4–59 A typical mitotic chromosome at metaphase. Each sister chromatid contains one of two identical sister DNA molecules generated earlier in the cell cycle by DNA replication (see also Figure 17–21).

Human chromosomes

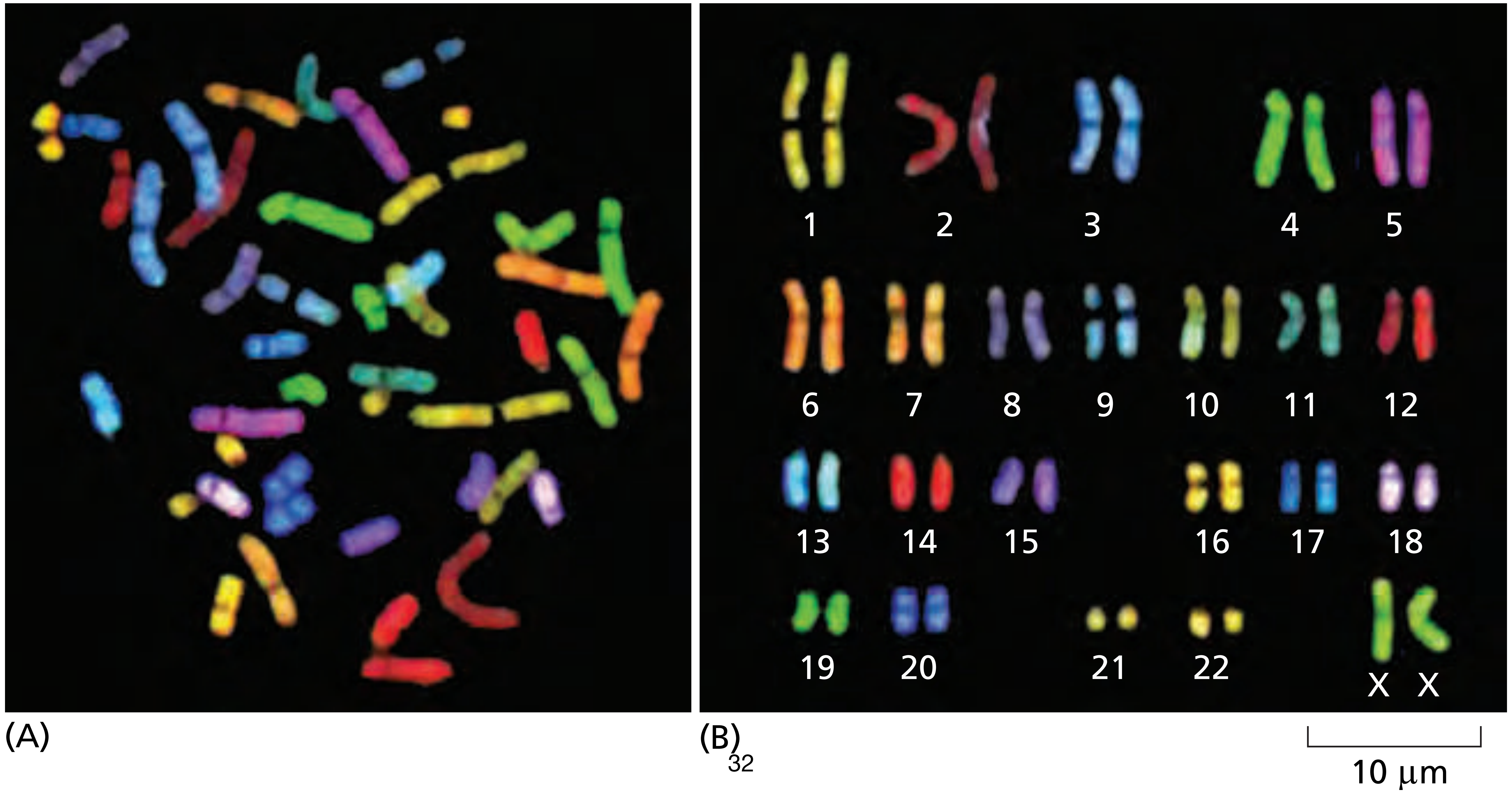


- During **mitosis**, **genes are silenced**
- Chromosomes are **highly compacted** with help of proteins (condensins and cohesins)
- The **centromere** is the region where the **mitotic spindle** will attach to the chromosomes to separate them in two cells

https://www.genome.gov/sites/default/files/media/images/tg/Metaphase.jpg

Human chromosomes

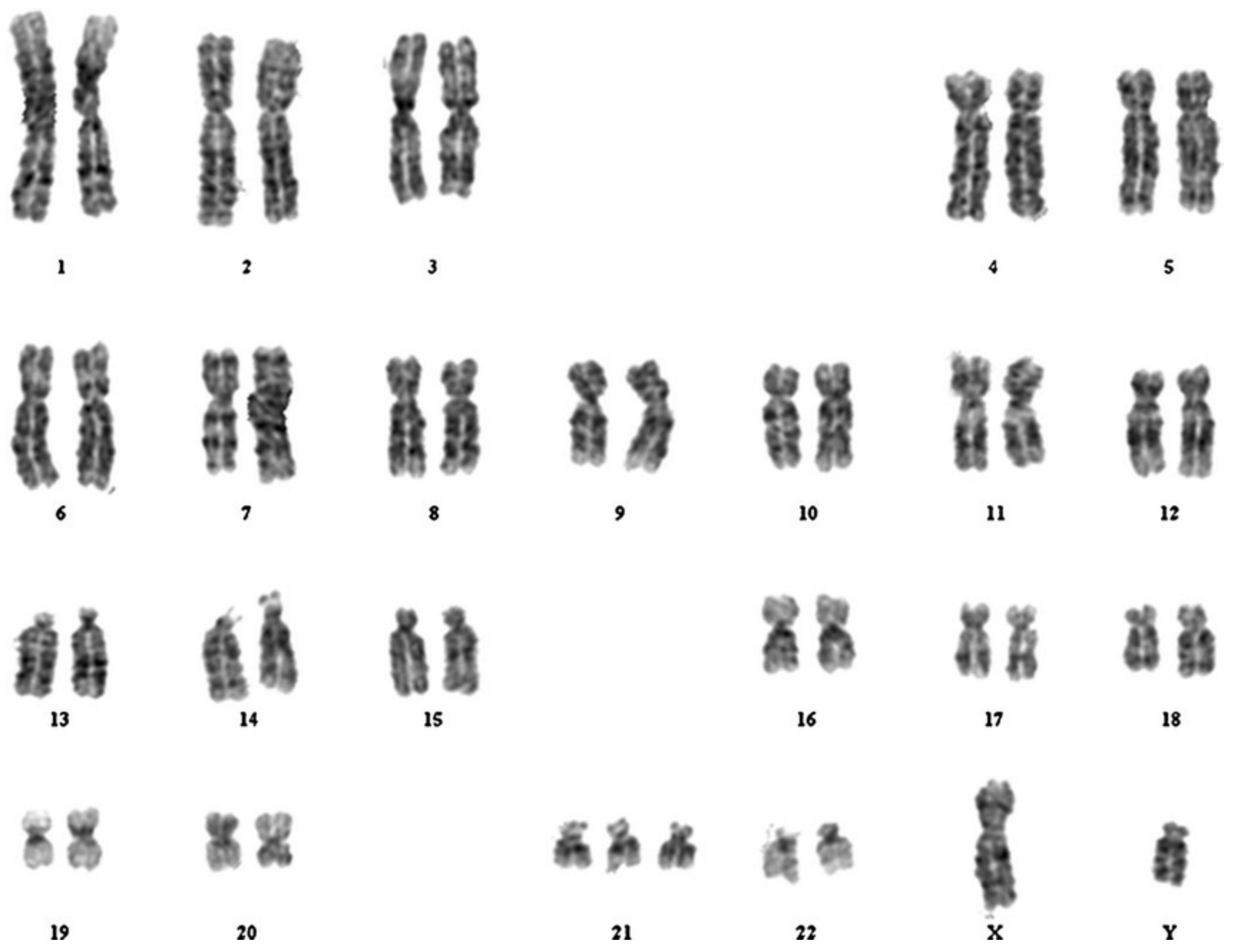
- Chromosomes can be distinguished by **“painting” or DNA hybridisation** (see later)
- A short strand of nucleic acid with a fluorescent dye serves as a probe that binds a complementary DNA sequence
- The display of the 46 chromosomes at mitosis is called **karyotype** and is used to detect inherited chromosome abnormalities

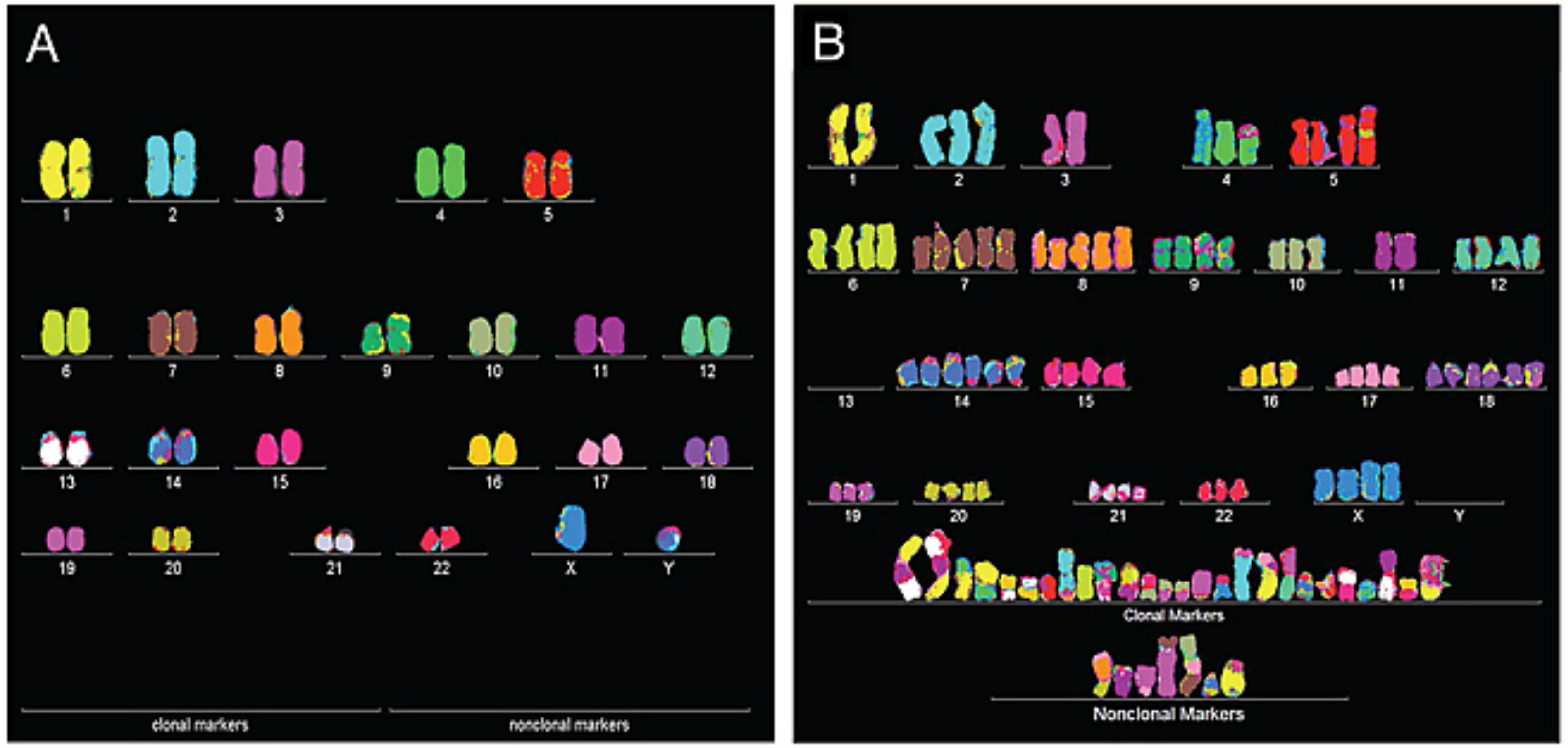


Chromosome aberrations

- **Genetic** (trisomy) vs. **Somatic** diseases (cancer)

- **Genetic** = mutation/aberration in germ cells, will be passed down to offspring
- **Somatic** = mutation/aberration in somatic cells, is only affecting the host

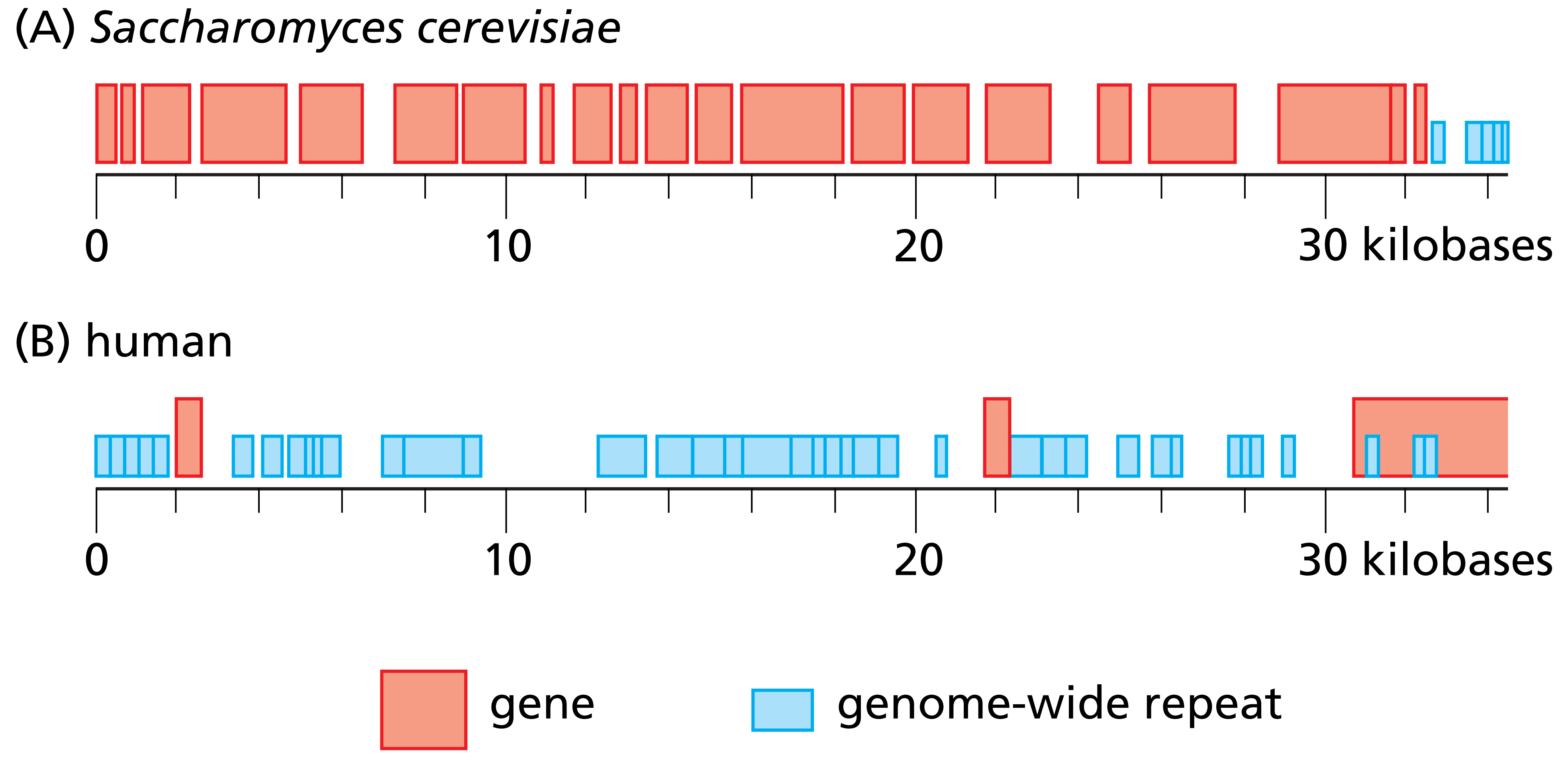




https://vcresearch.berkeley.edu/sites/default/files/news_images/karyograms670.jpg

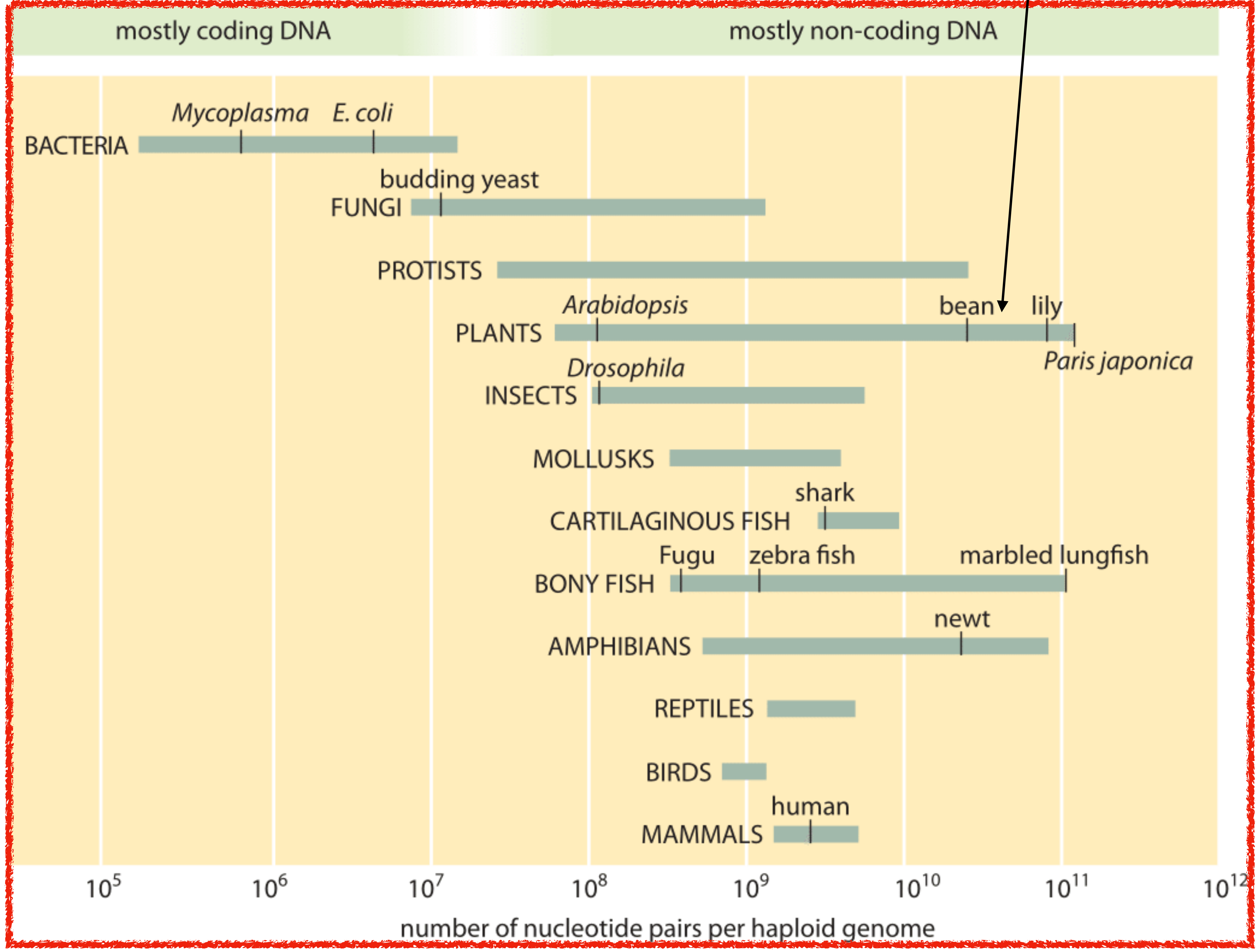
Zooming into the chromosomes

- Chromosomes carry **genes** = segments of DNA that contain the instructions to make **particular proteins**
- Some genes are **RNA genes, with RNA as a final product**
- Correlation between the **complexity of an organism and its number of genes** (~3K in bacteria vs. 30K in humans, incl. 21K coding for proteins)
- ! This is not true for **genome size** (non-coding DNA)
- In multicellular organisms, large quantity of **interspersed DNA between genes** whose function is poorly understood - crucial for the control of **gene expression** = **non-coding DNA**



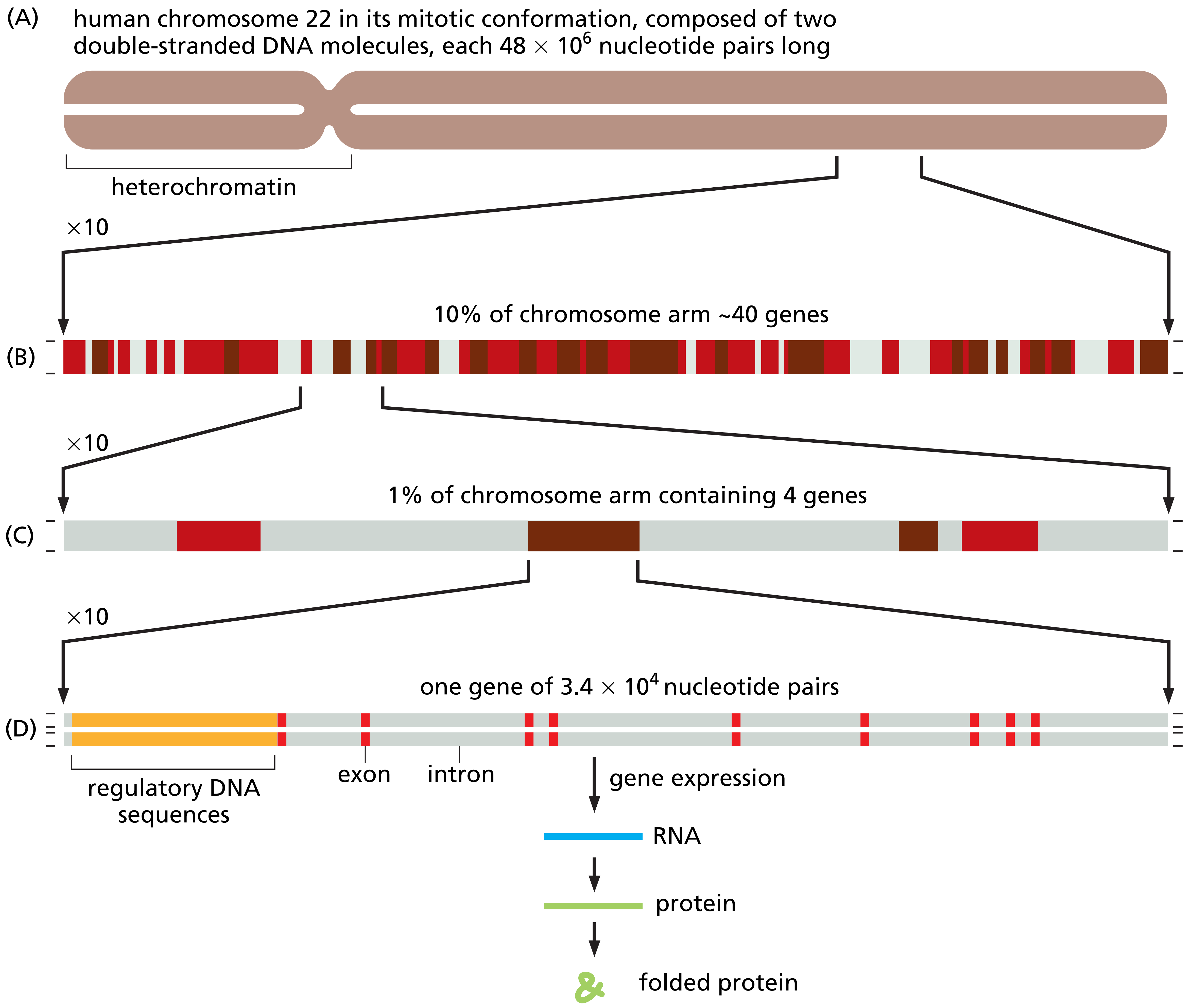
Genome size

Polyploidy and non-coding DNA



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

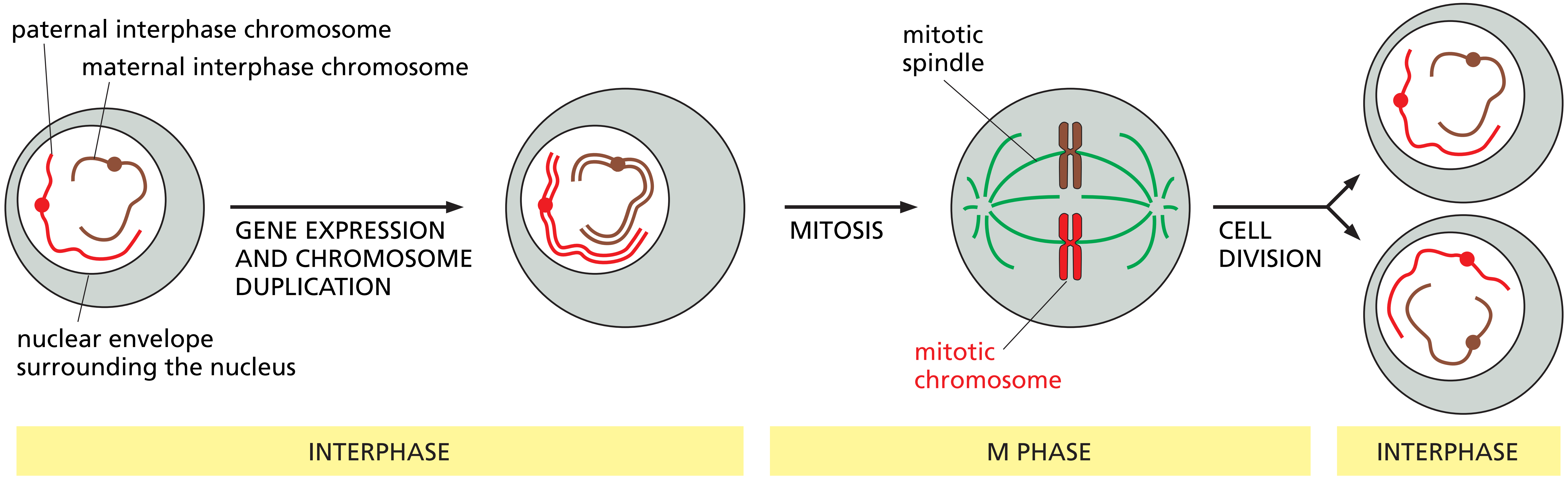
Zooming into the chromosomes



- In 2004, full sequence of the **human genome**
- We see **how genes are arranged** within chromosomes
- Only **1.5% of DNA** is coding for proteins
- ~50% is made of **mobile DNA**, inserted gradually over evolutionary time
- Average gene size is 27Kbp, which is long: **introns (non-coding DNA) between exons (coding sequences)**

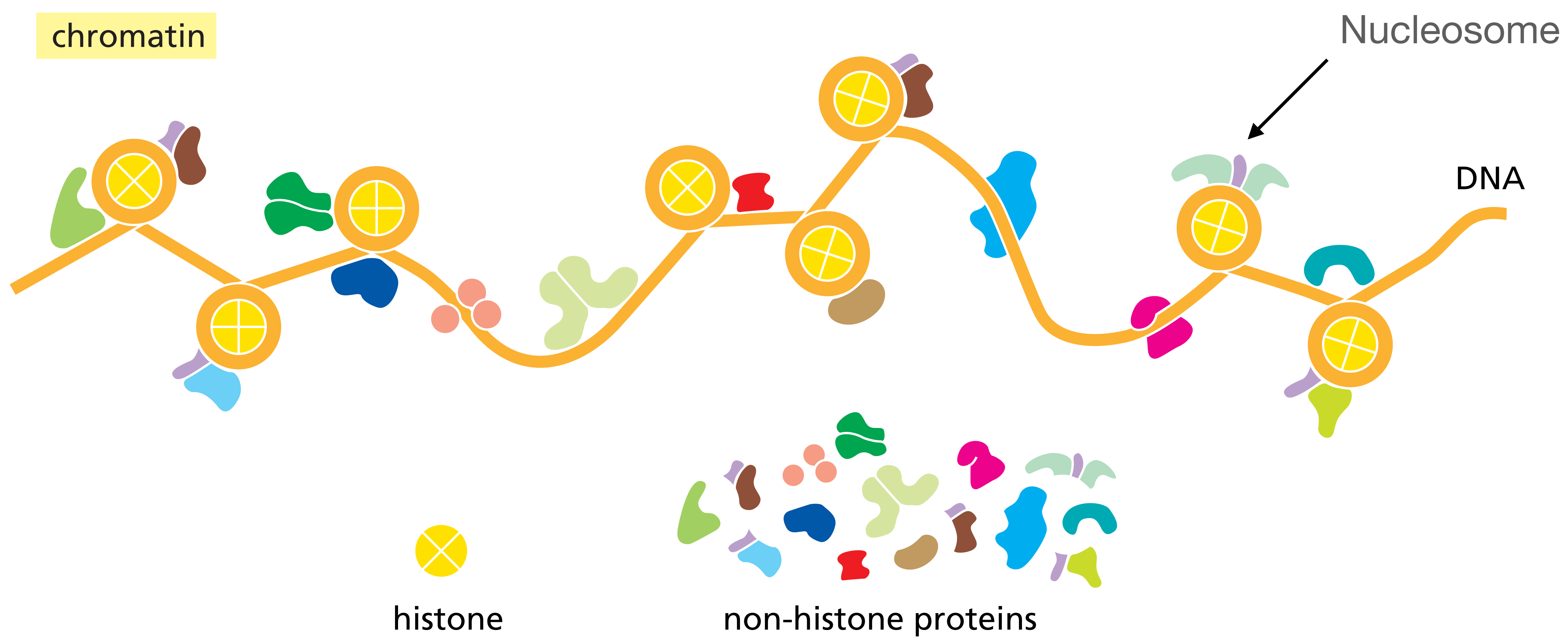
Chromosomes and cell cycle

- A **functional chromosome** must be able to **replicate**, be **separated** and reliably **partitioned** into daughter cells at each division = cell cycle
- During the **Interphase**, **genes are expressed and chromosomes replicated**; the two sister chromatids stay together
- At **mitosis**, chromosomes are **condensed** (= mitotic chromosomes)



Chromatin organization

- DNA molecules are **highly condensed** in chromosomes
- **Proteins** that coil and fold the DNA into higher levels of organization
- Chromosome structure is dynamic - they **decondense for replication, gene expression or DNA repair**
- Proteins that bind to DNA to form eukaryotic chromosomes are the **histones and the non-histone chromosomal proteins**
- Histones+ non-histone chromosomal proteins+DNA = **chromatin**



Nucleosomes

- Histones are responsible for the first level of chromosome packing, the **nucleosome**
- **“Beads on a string”** structure

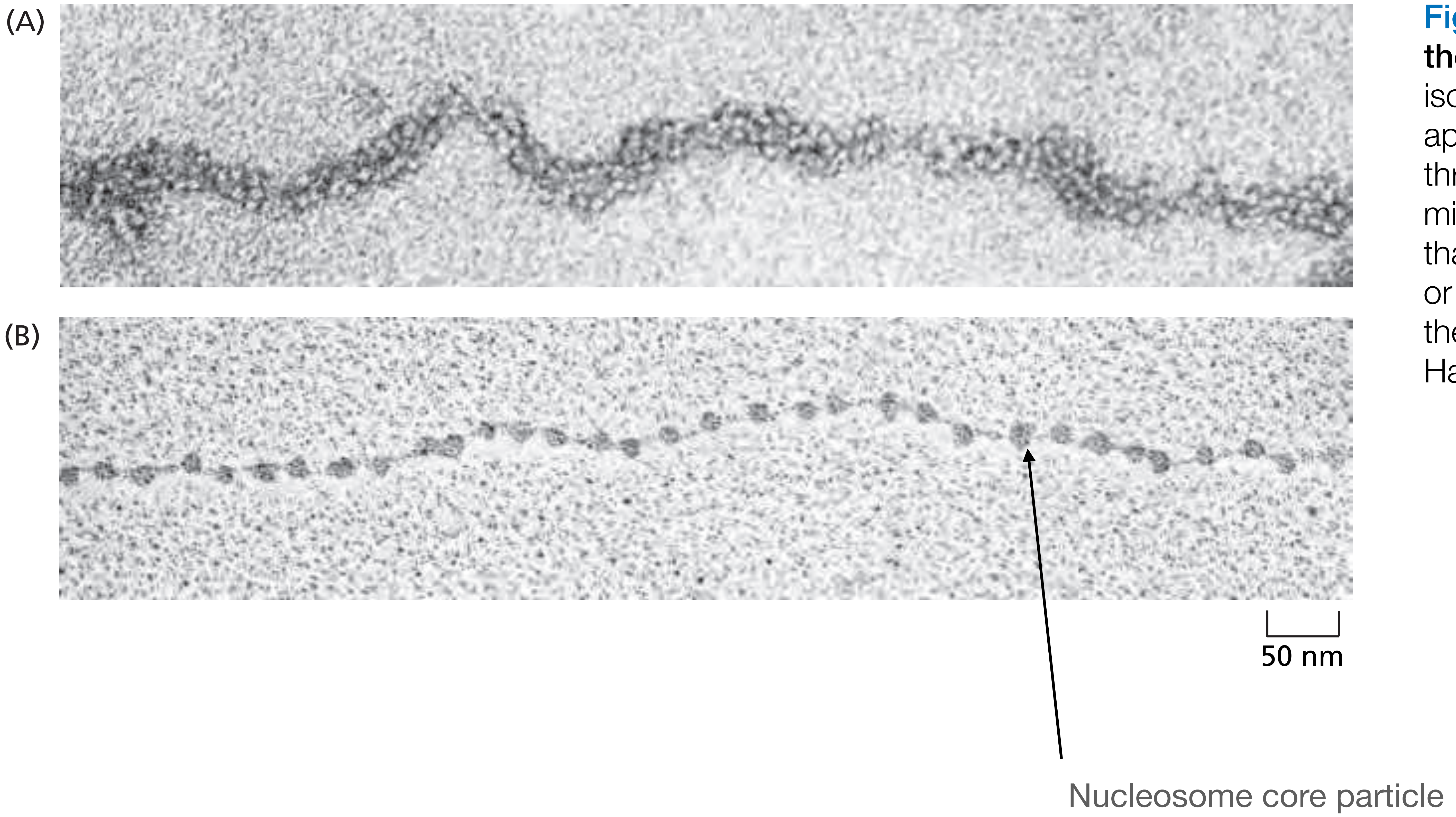
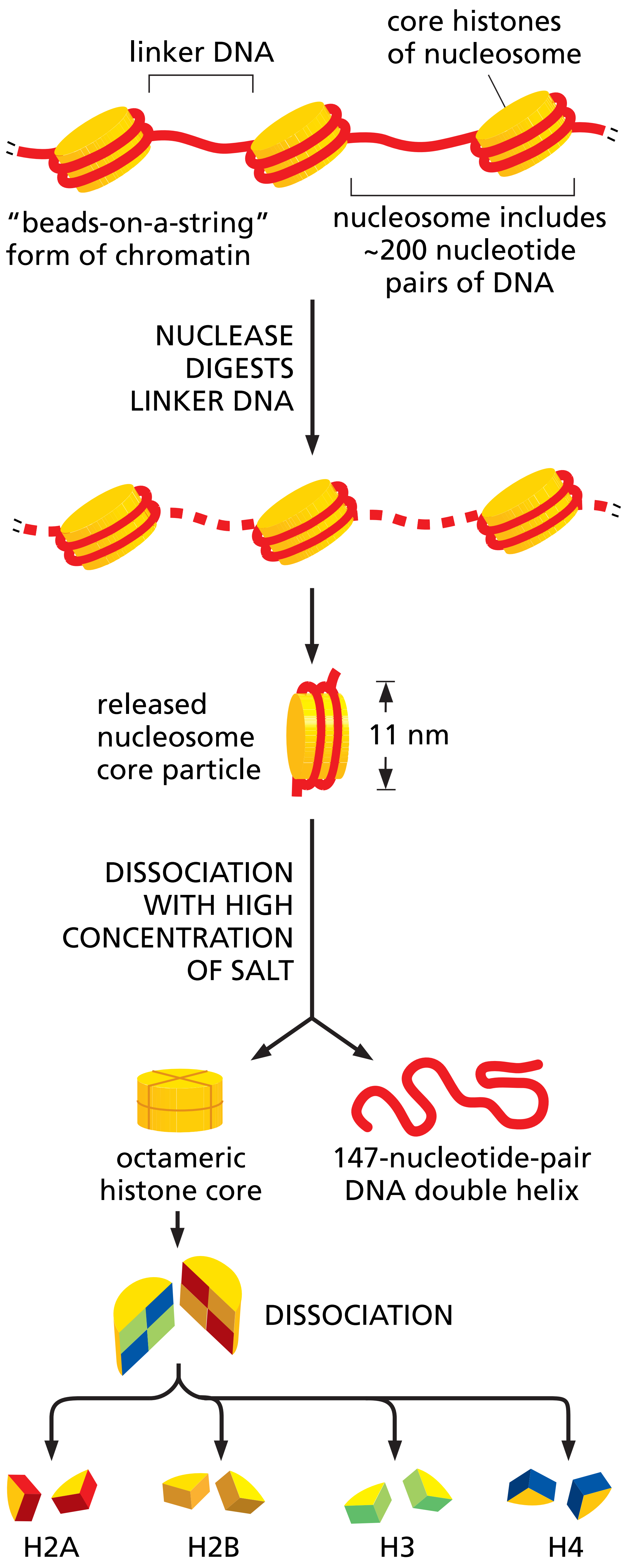


Figure 4–21 Nucleosomes as seen in the electron microscope. (A) Chromatin isolated directly from an interphase nucleus appears in the electron microscope as a thread about 30 nm thick. (B) This electron micrograph shows a length of chromatin that has been experimentally unpacked, or decondensed, after isolation to show the nucleosomes. (A, courtesy of Barbara Hamkalo; B, courtesy of Victoria Foe.)

Nucleosome structure

- 1 **nucleosome =** a complex of **8 histone proteins** and **double-stranded DNA**
- 4 different histones in duplicate = **octamer**
- Each histone octamer forms a **protein core** around which the DNA is wound
- Nucleosomes repeat at intervals of **~200 nucleotide** pairs
- **Linker DNA** is the DNA located between two nucleosomes
- **Compacts** DNA ~**3x**



Linker DNA and histone-1

- Nucleosomes are packed together into a **compact chromatin fiber** (diameter 30 nm) "Normal state"
- How this happens is unclear but involves **histone tails** and an additional **histone H1**

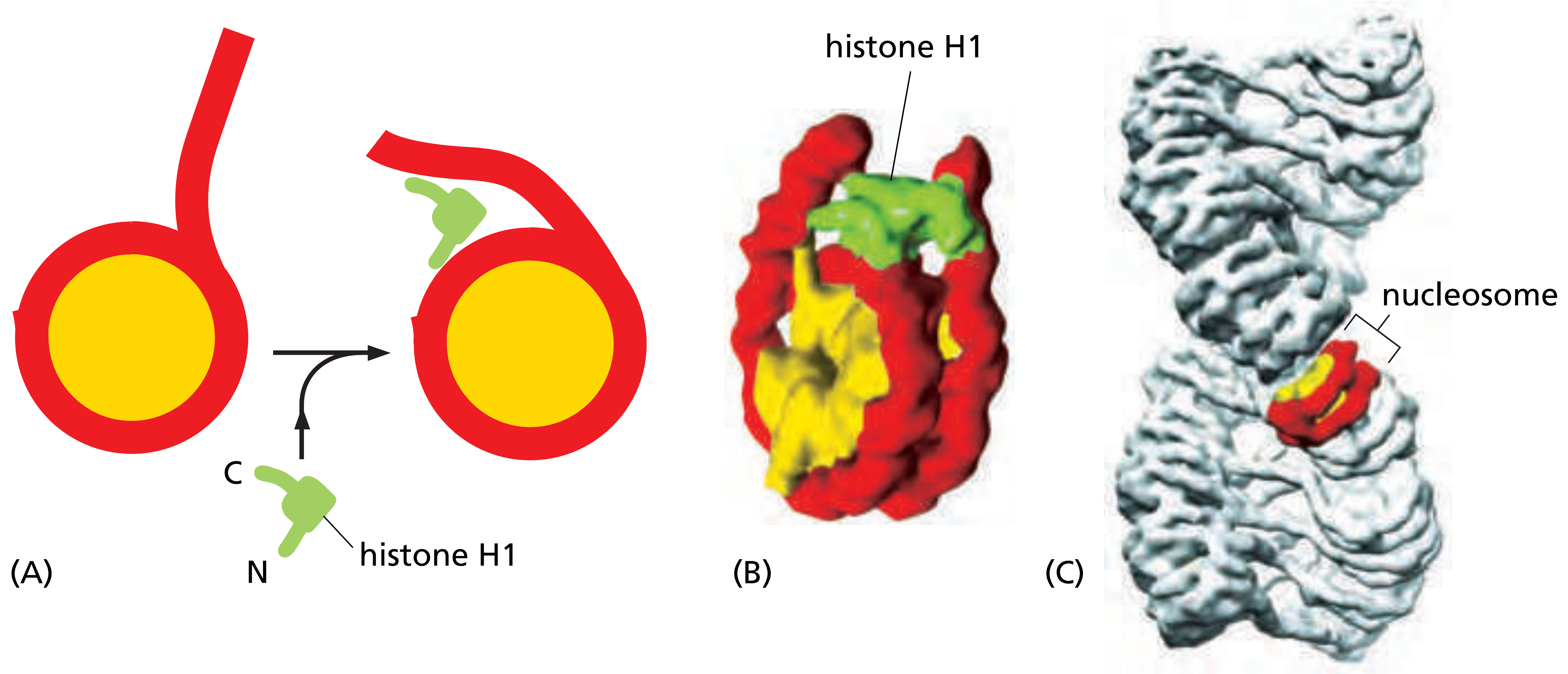


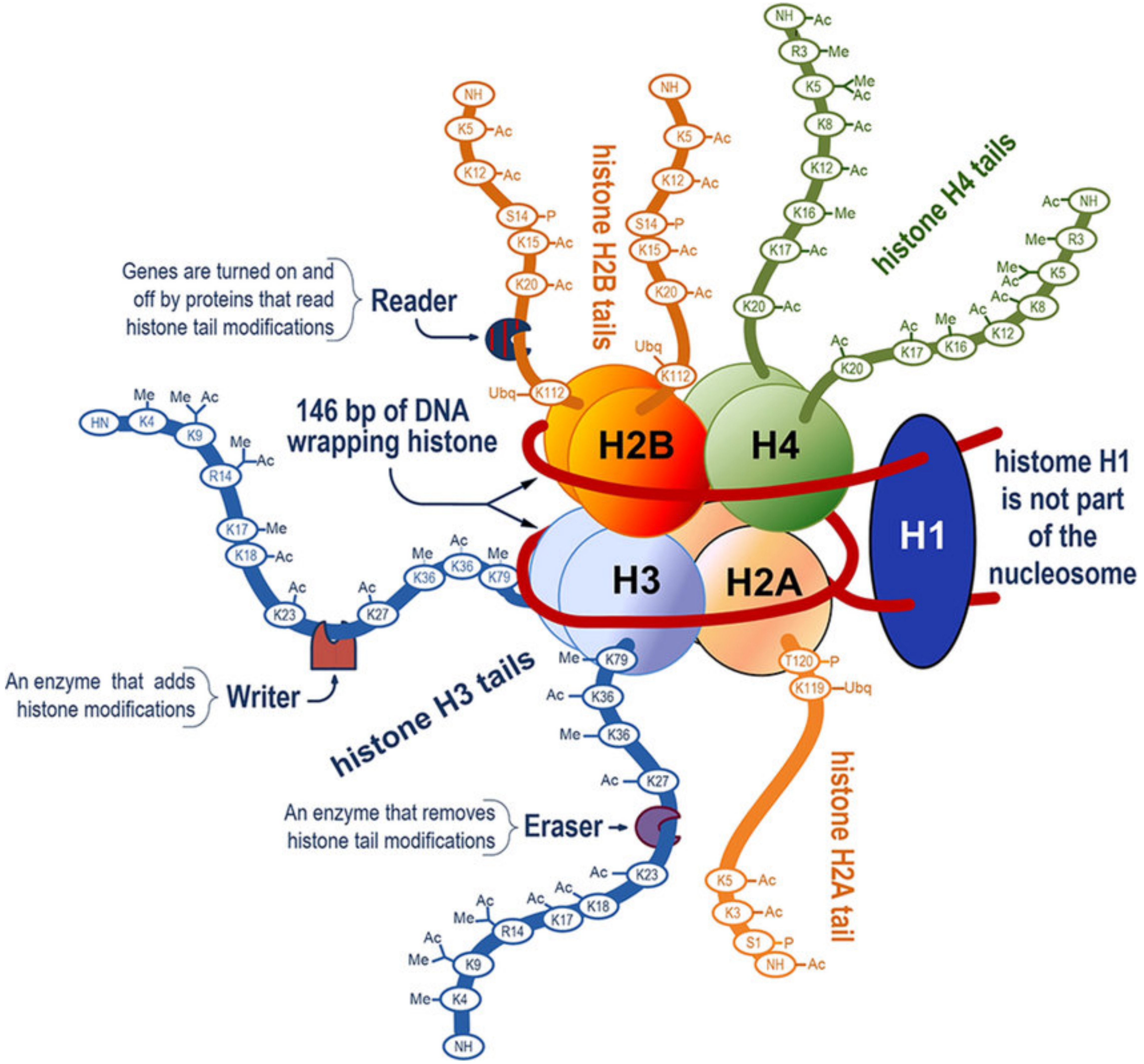
Figure 4–30 How the linker histone binds to the nucleosome. The position and structure of histone H1 is shown. The H1 core region constrains an additional 20 nucleotide pairs of DNA where it exits from the nucleosome core and is important for compacting chromatin. (A) Schematic, and (B) structure inferred for a single nucleosome from a structure determined by high-resolution electron microscopy of a reconstituted chromatin fiber (C). (B and C, adapted from F. Song et al., *Science* 344:376–380, 2014.)

Plan

- Brief introduction on Biology basics
- The structure and function of DNA
- Chromosomal DNA and its packaging into chromatin
- <u>The effect of chromatin structure on DNA function</u>

Histones characteristics

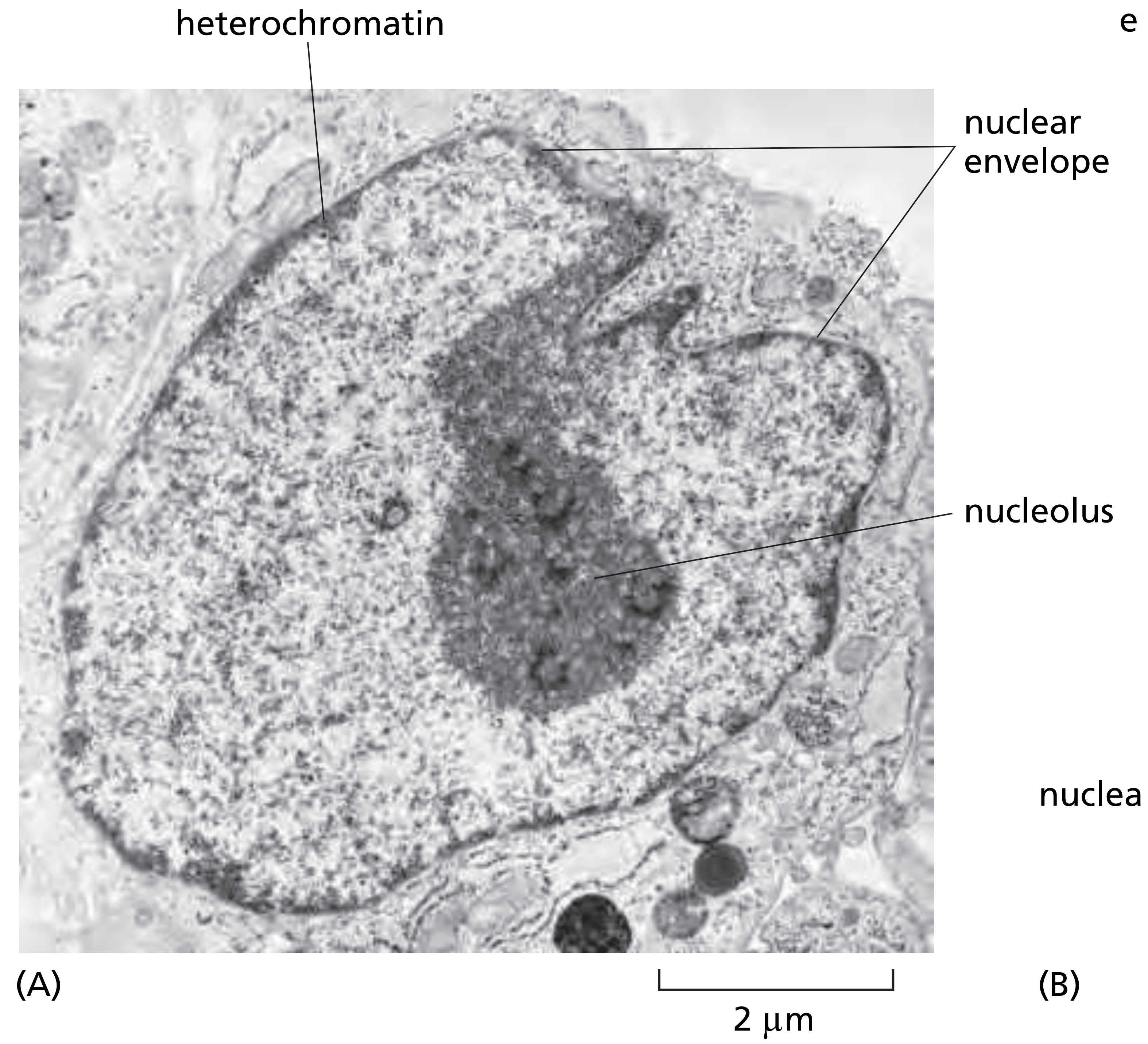
- **Positively charged** - neutralise negative charges from DNA
- Core histones are highly **conserved** across evolution
- Nucleosomes are **dynamic** to allow replication and transcription
- **N-terminal tails** of histones stick out from the nucleosome and are subject to **histone modifications**

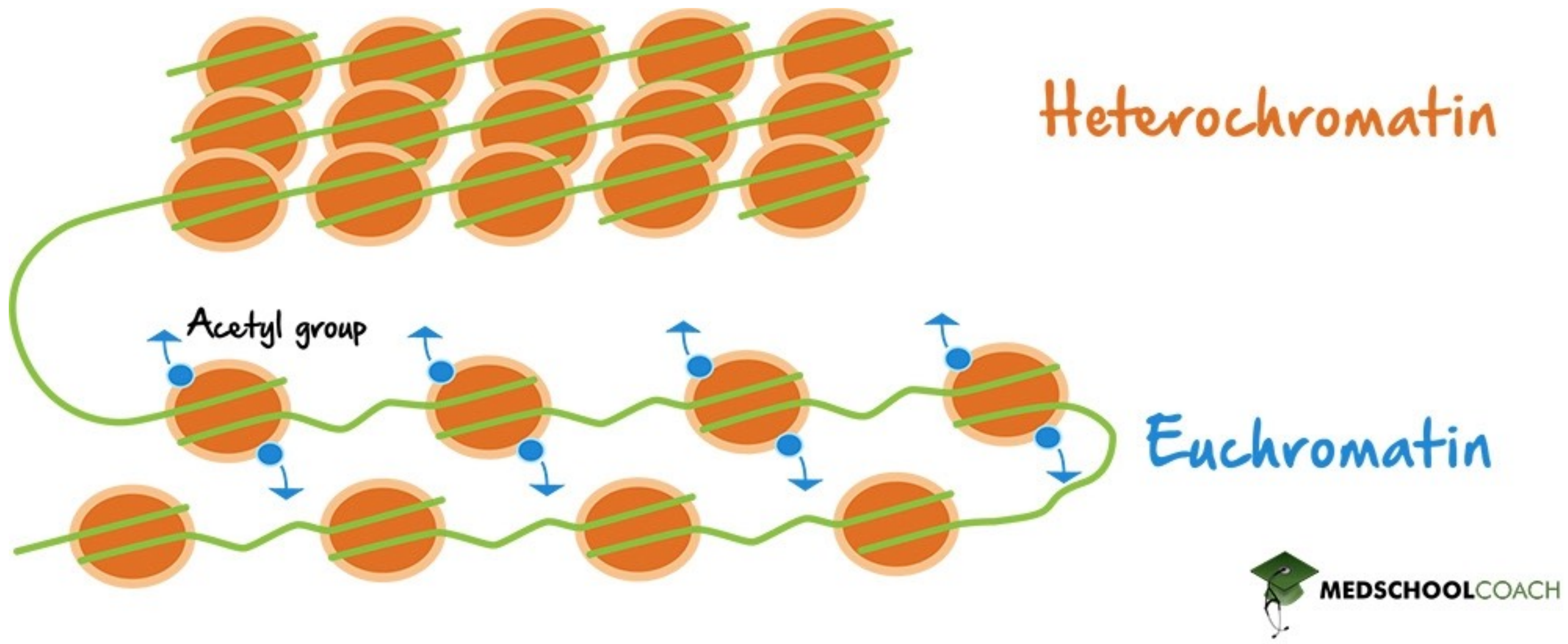


DOI:10.1021/acs.analchem.7b05007

Chromatin dynamically changes

- In the 1930s, light microscopy distinguished two types of chromatin: **heterochromatin (highly condensed- inactive) and euchromatin (less condensed- active)**
- Heterochromatin is highly concentrated in **certain regions** (centromeres and telomeres), but also present in other regions **depending on the physiological state of the cell**
- Heterochromatin can be **constitutive** (= regions that are always in heterochromatin) or **facultative** (=regions that are sometimes on, sometimes off)
- Heterochromatin contains **few genes**
- Through chromosome breakage, if a part of **euchromatin ends up in heterochromatin**, this causes the **silencing of normally active genes** = **position effect**





=

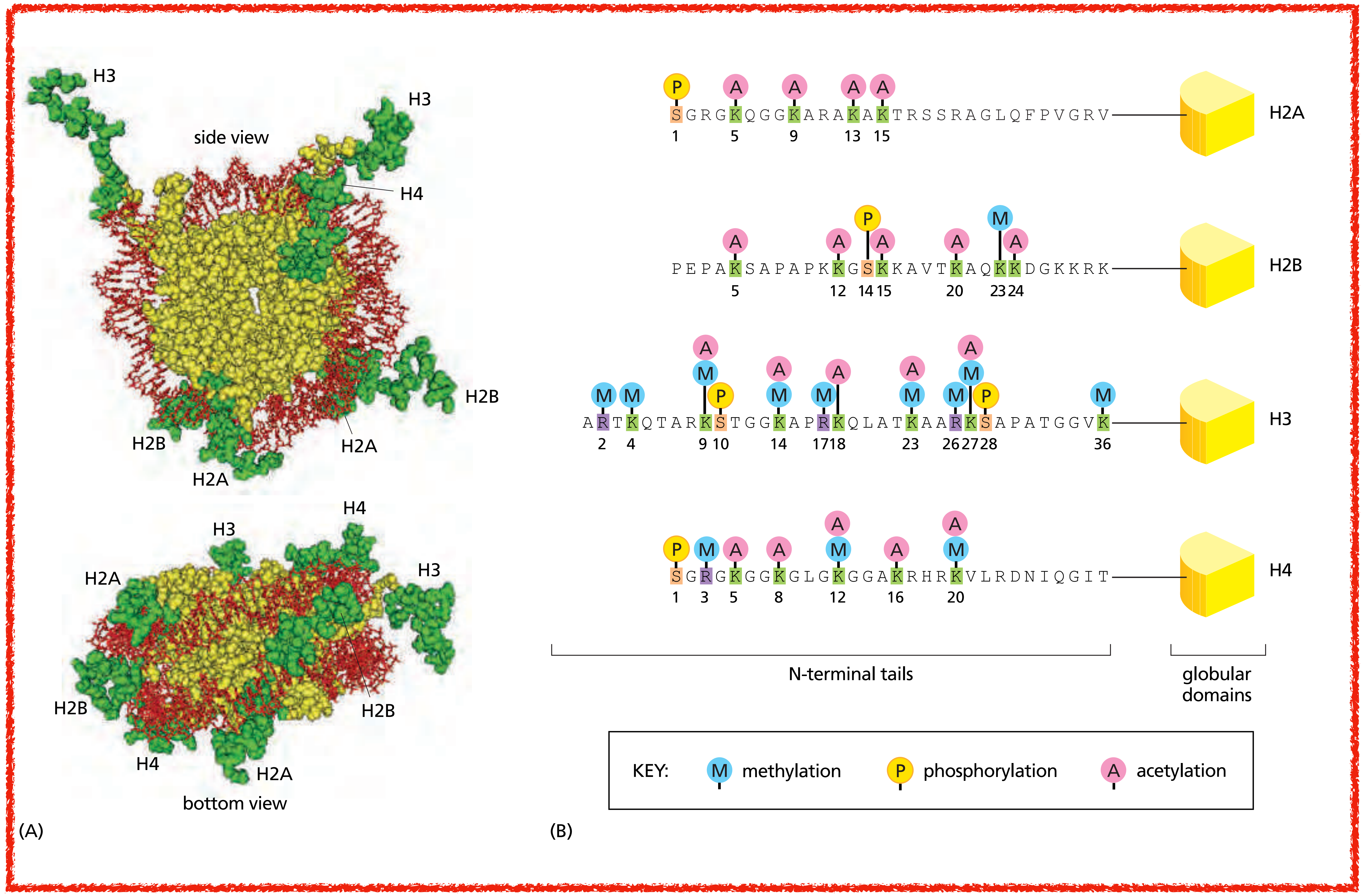
Chromatin dynamically changes

To make it simple…

- Regions of chromosomes that **do not contain genes** are typically organised as **heterochromatin**
- Regions of chromosomes that **do contain genes** can be switching between **euchromatin** and **heterochromatin**
- Regions of chromosomes that **contain genes** that need to be always expressed are part of **euchromatin**

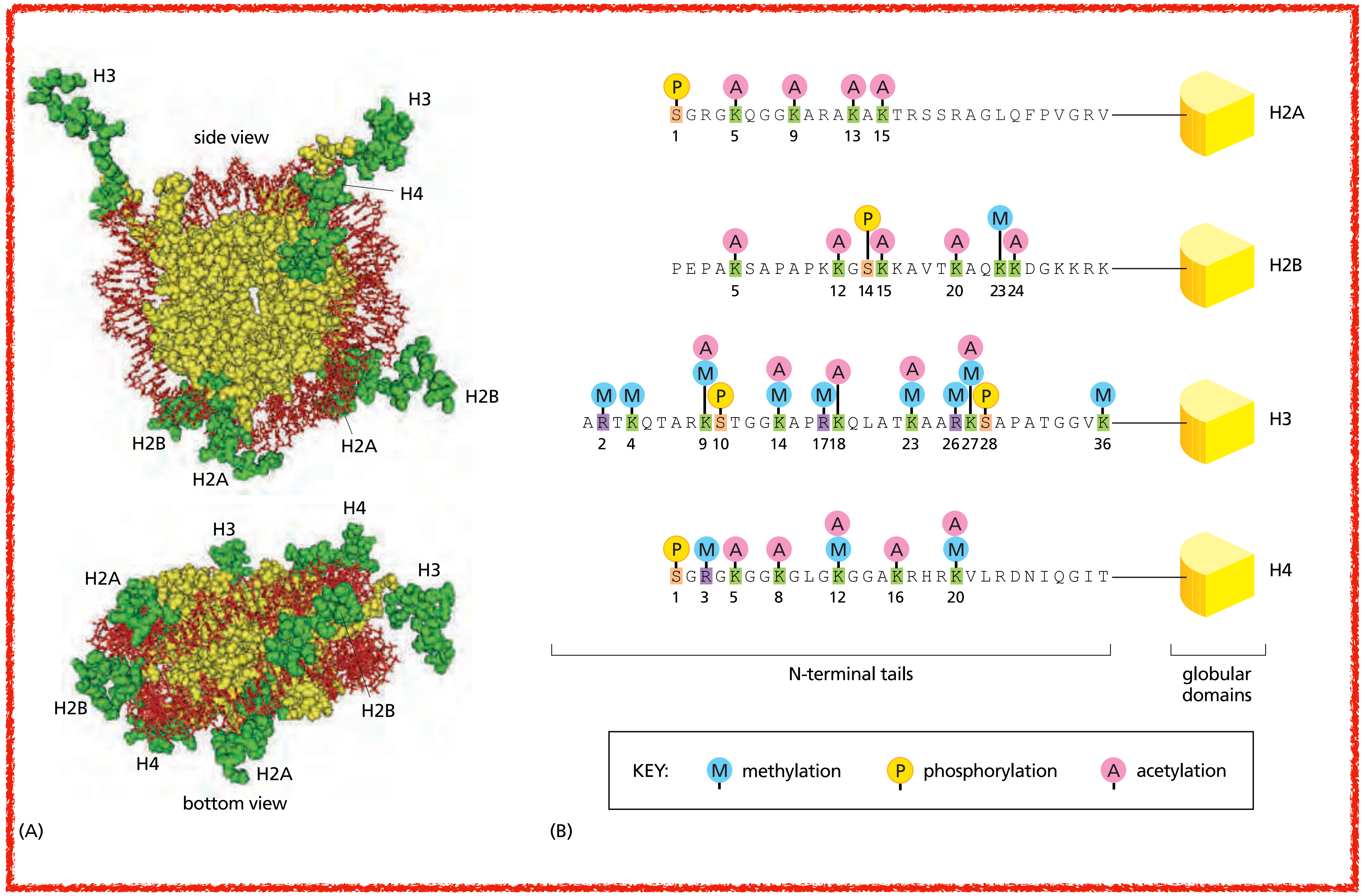
Histone modifications

- The **amino-acid side-chain of the core histones** are subjected to many covalent mutations (acetylation of lysines, methylation of lysines or phosphorylation of serines)
- Primarily on **histone tails** but also on **histone** core
- Not on **histone H1** (not part of the nucleosome)
- Modifications are reversible with **dedicated enzymes** to add or remove them (e.g. HATs and HDACs)
- Results in change in **chromatin organization**
- Are referred to as the **"histone code"**

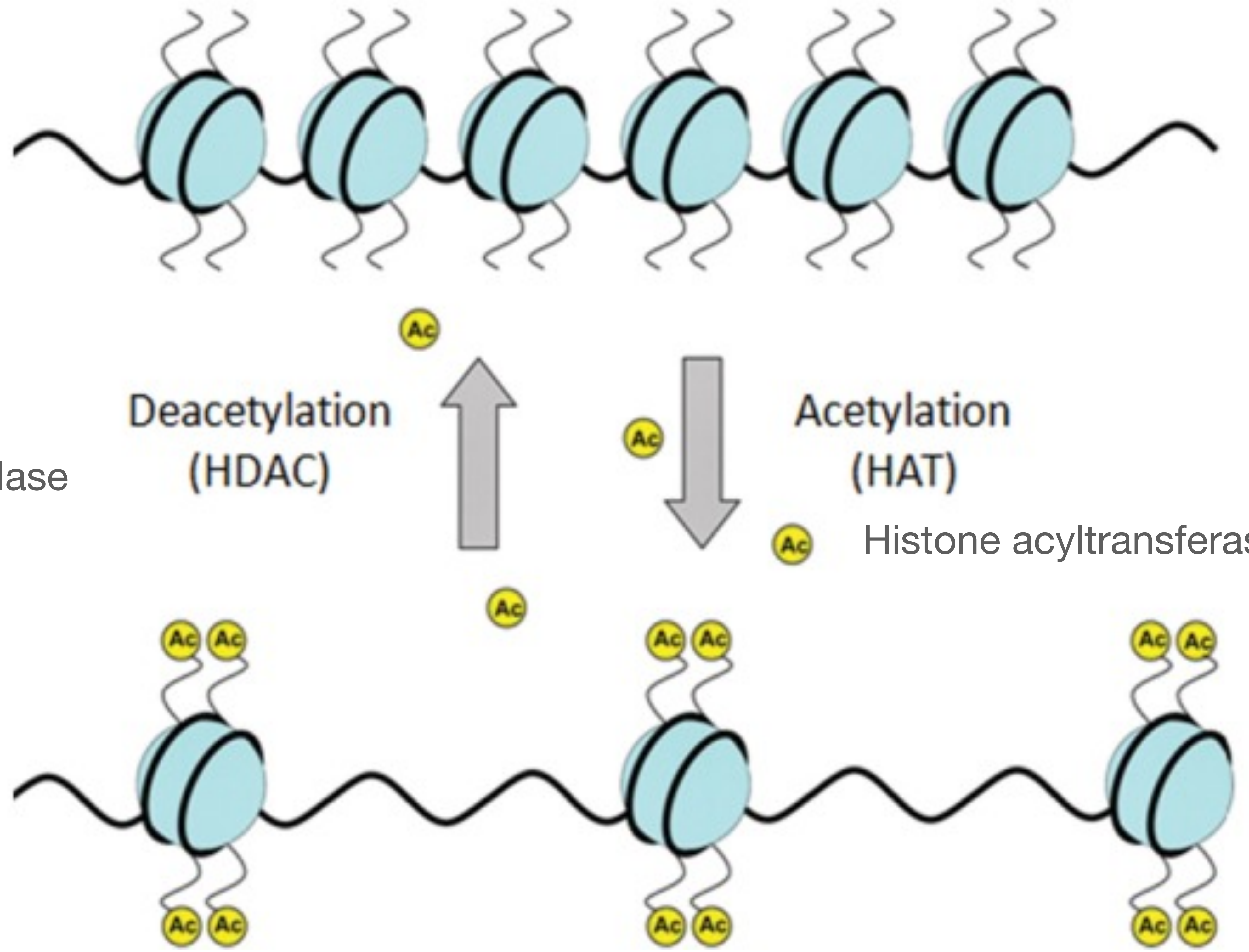


Histone modifications

- Modification enzymes bind to **DNA sequences**, so **DNA determines** how histones are modified
- What would happen if the **histone proteins** were recognised by the modification enzymes?
- Some modifications **persist in time**



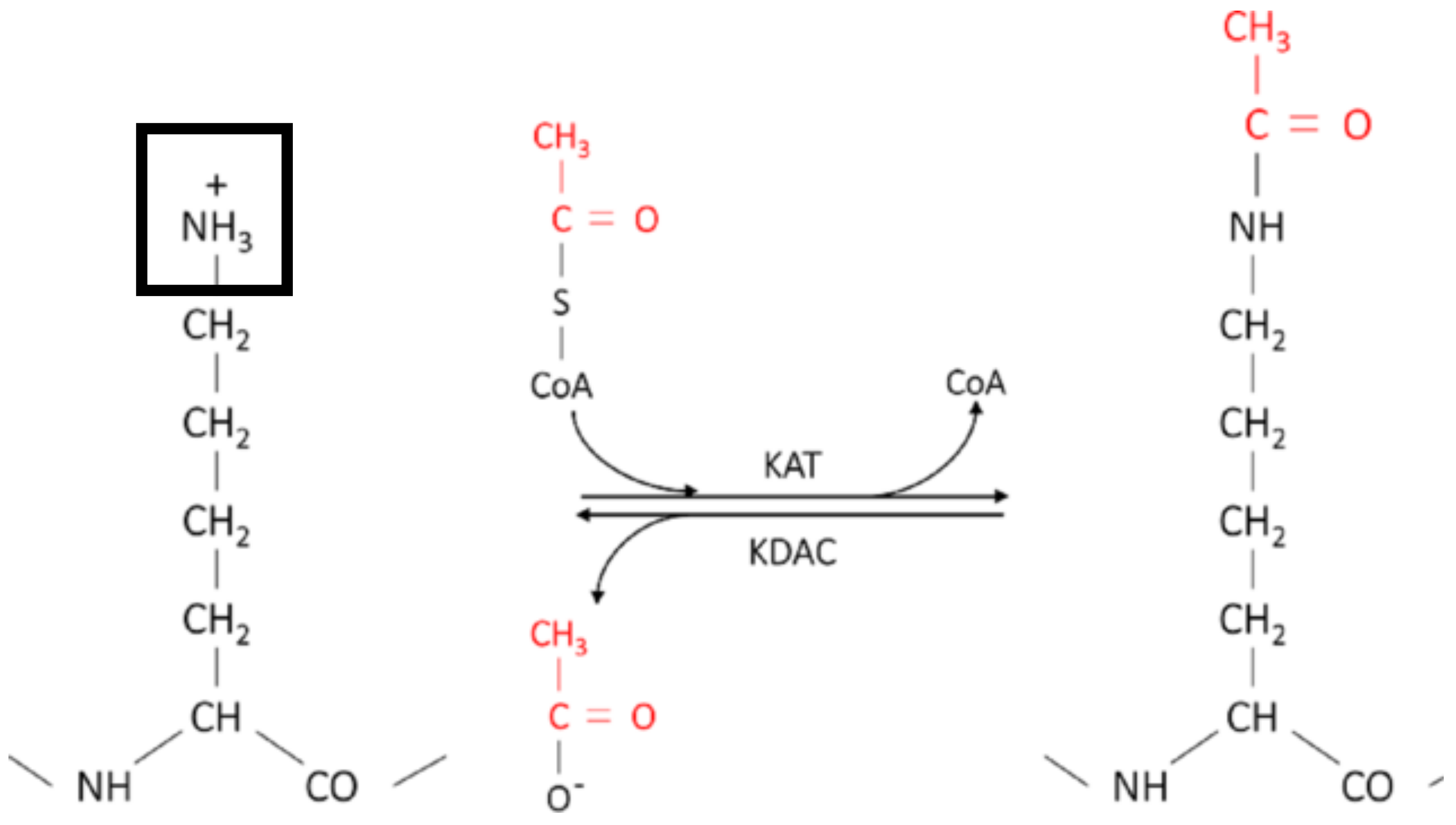
Example: Histone acetylation



Histone deacetylase

Histone acyltransferase

Masks the positive charge

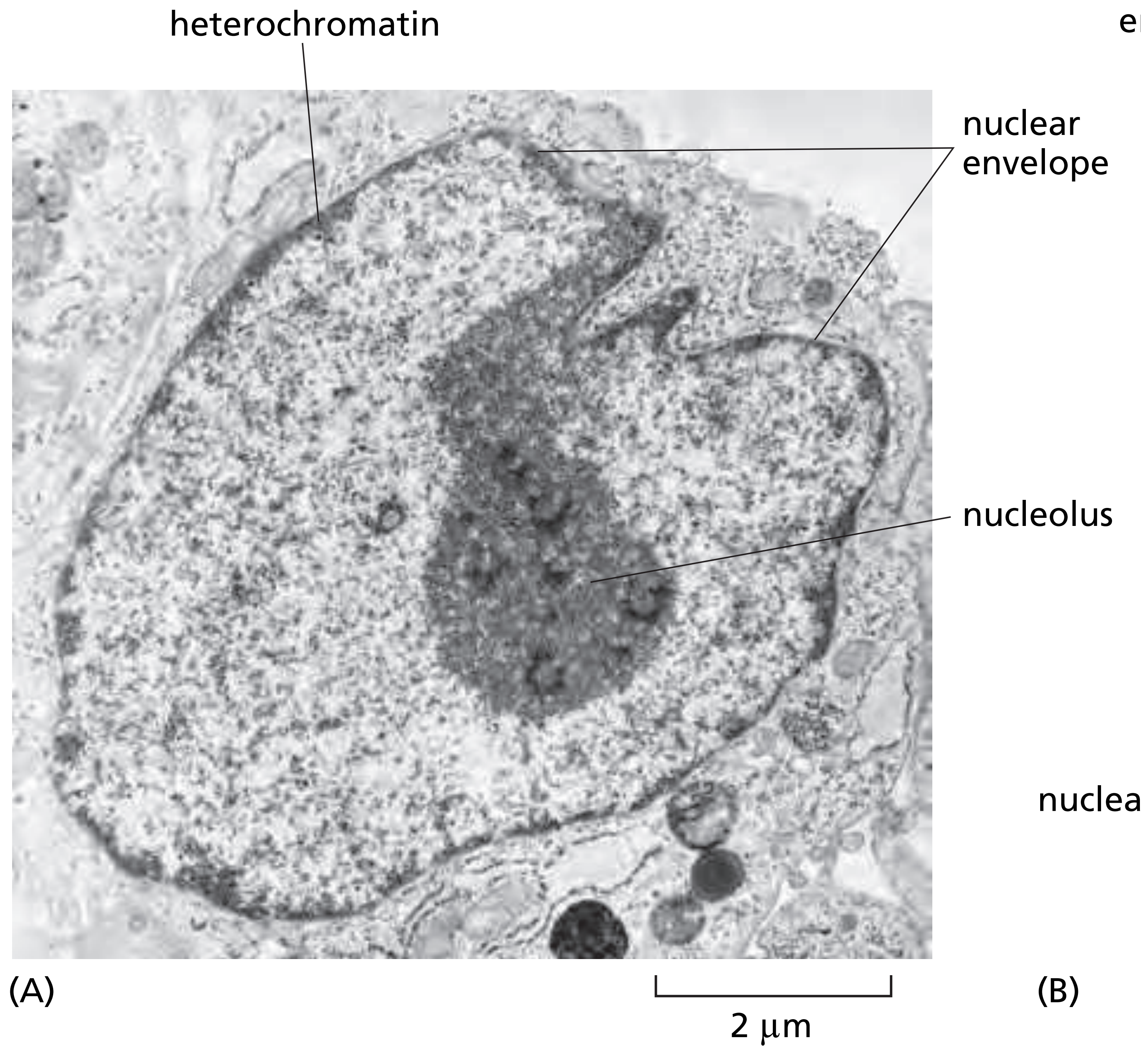


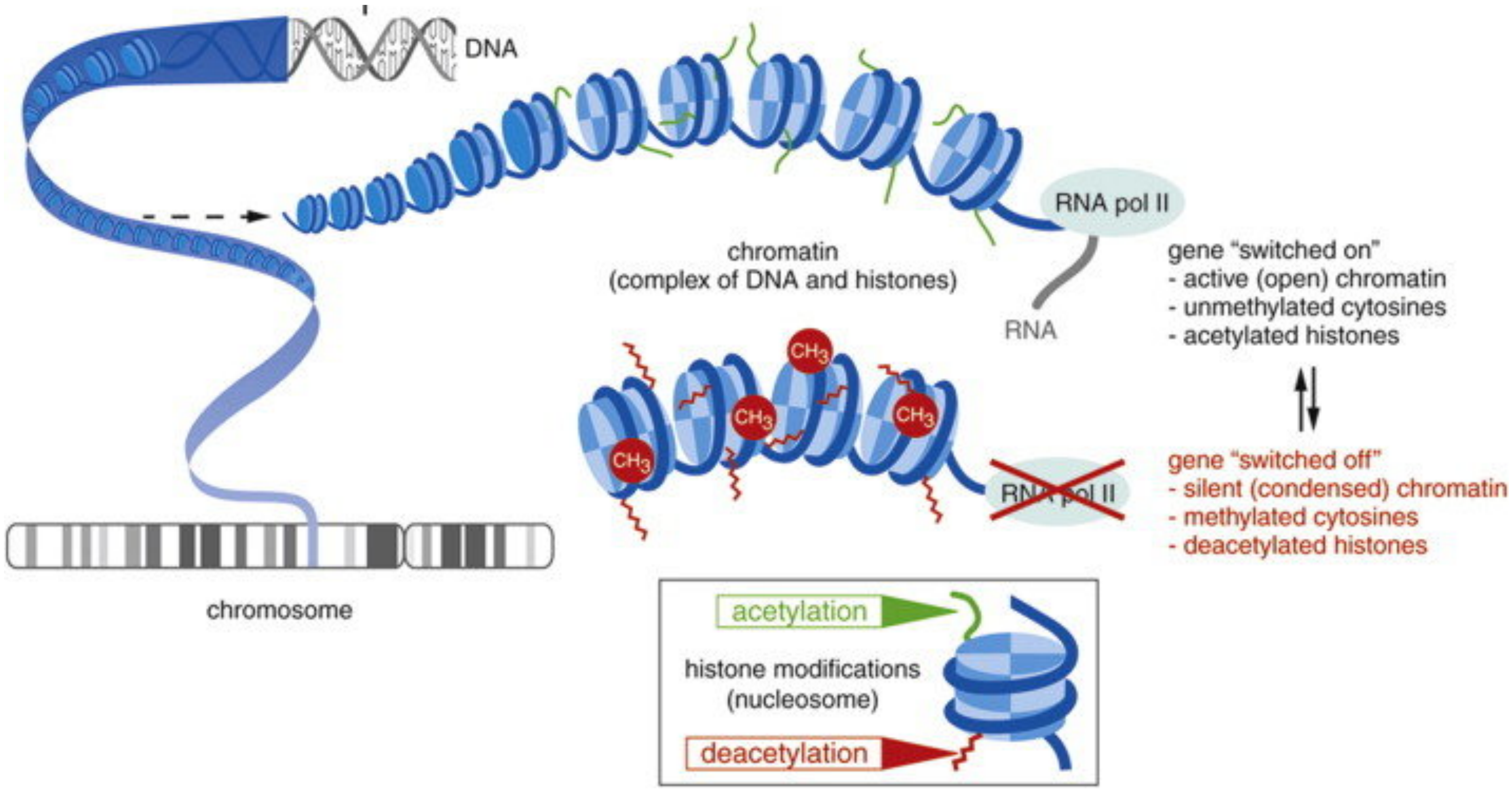
Lysine

PMCID: PMC3660019

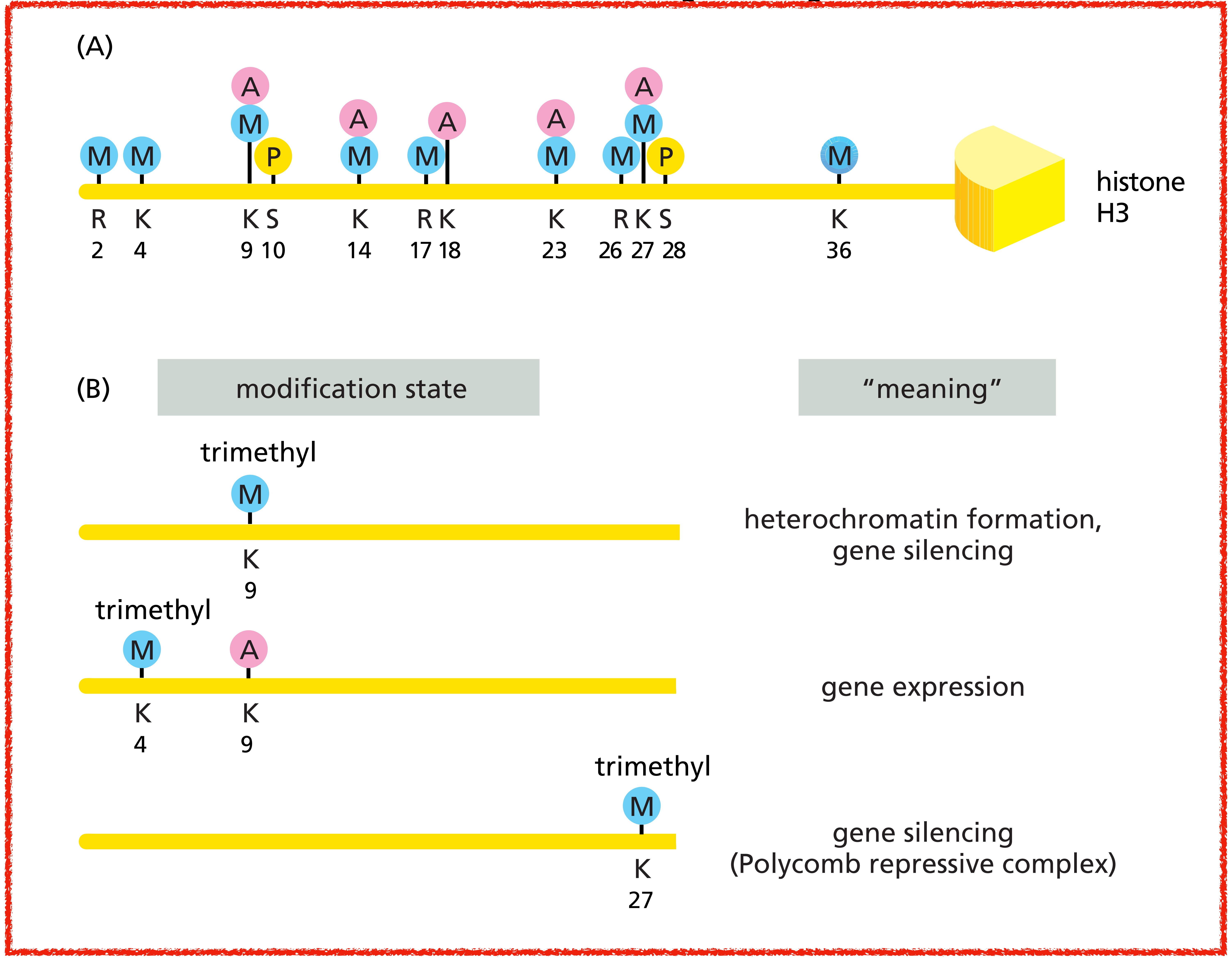
Histone modifications influence gene expression

- Acetylated histones - open chromatin (euchromatin)- transcriptional activity



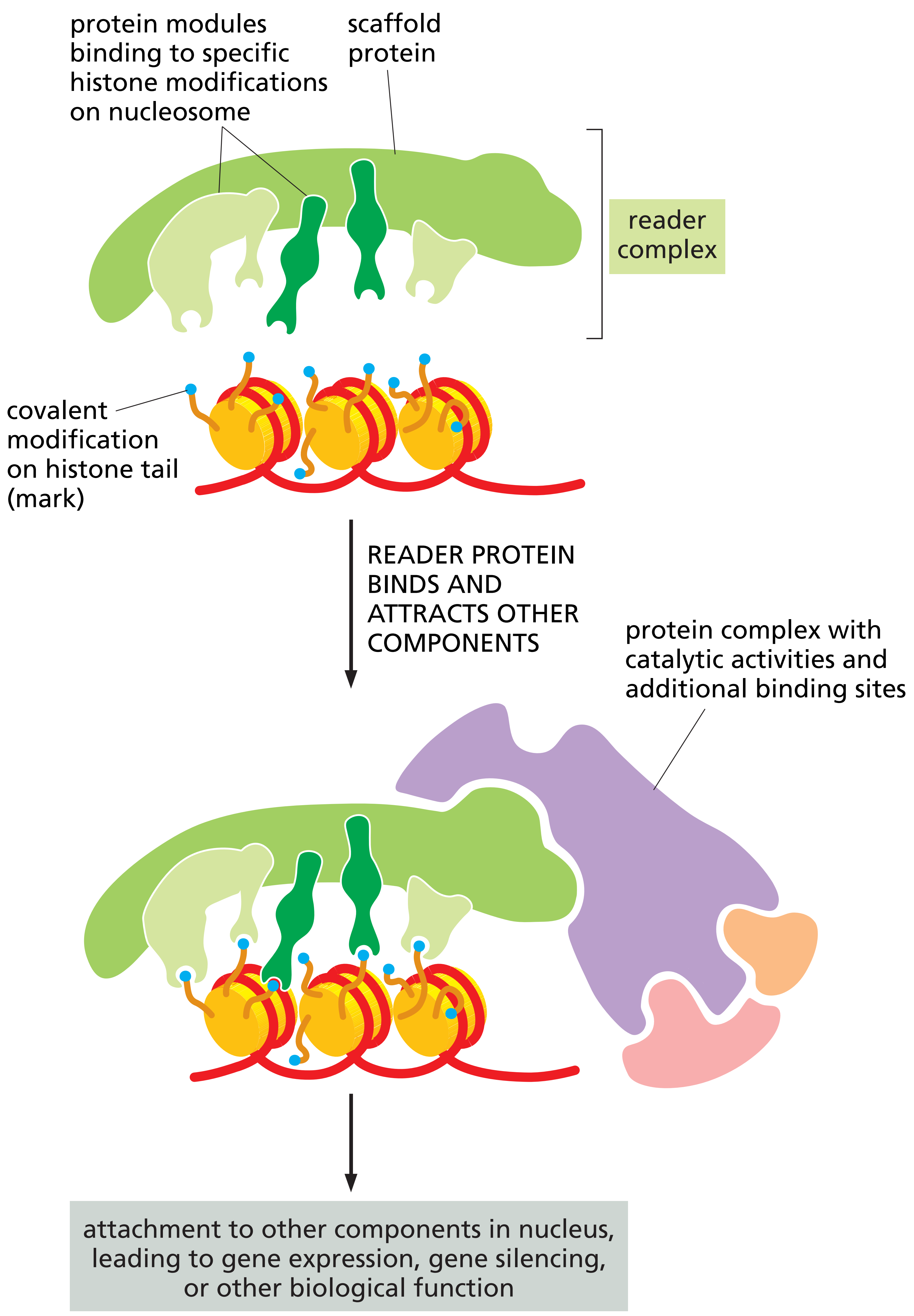


Example: histone code (H3)



Reader complexes read the histone code

- **Writer protein/complex**
  - An enzyme that deposits histone modifications, such as a methyltransferase or acetyltransferase.
- **Reader protein/complex**
  - A protein or domain that specifically recognizes and binds a given histone modification.
  - It can recruit additional proteins with specific activities

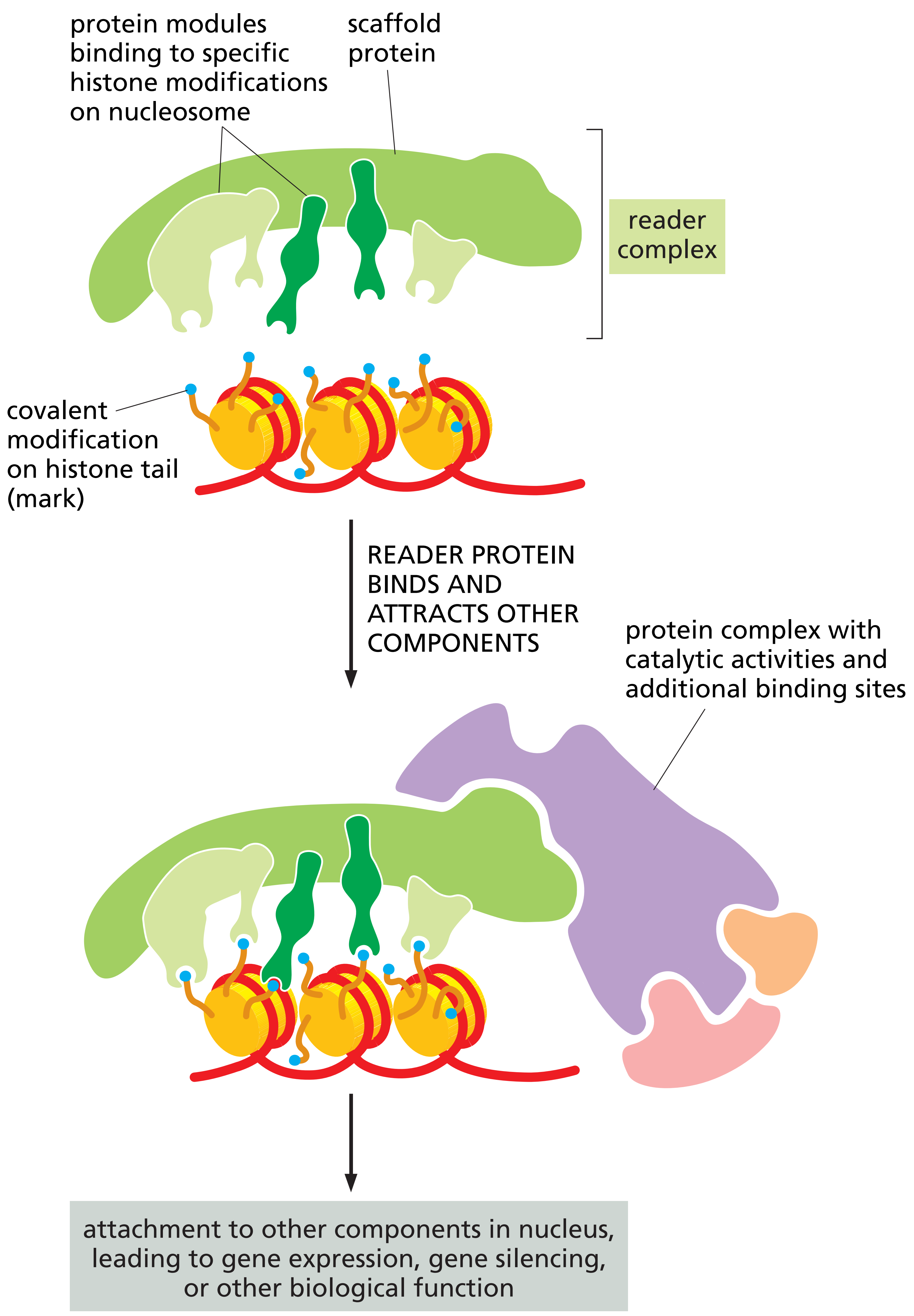


Reader complexes read the histone code

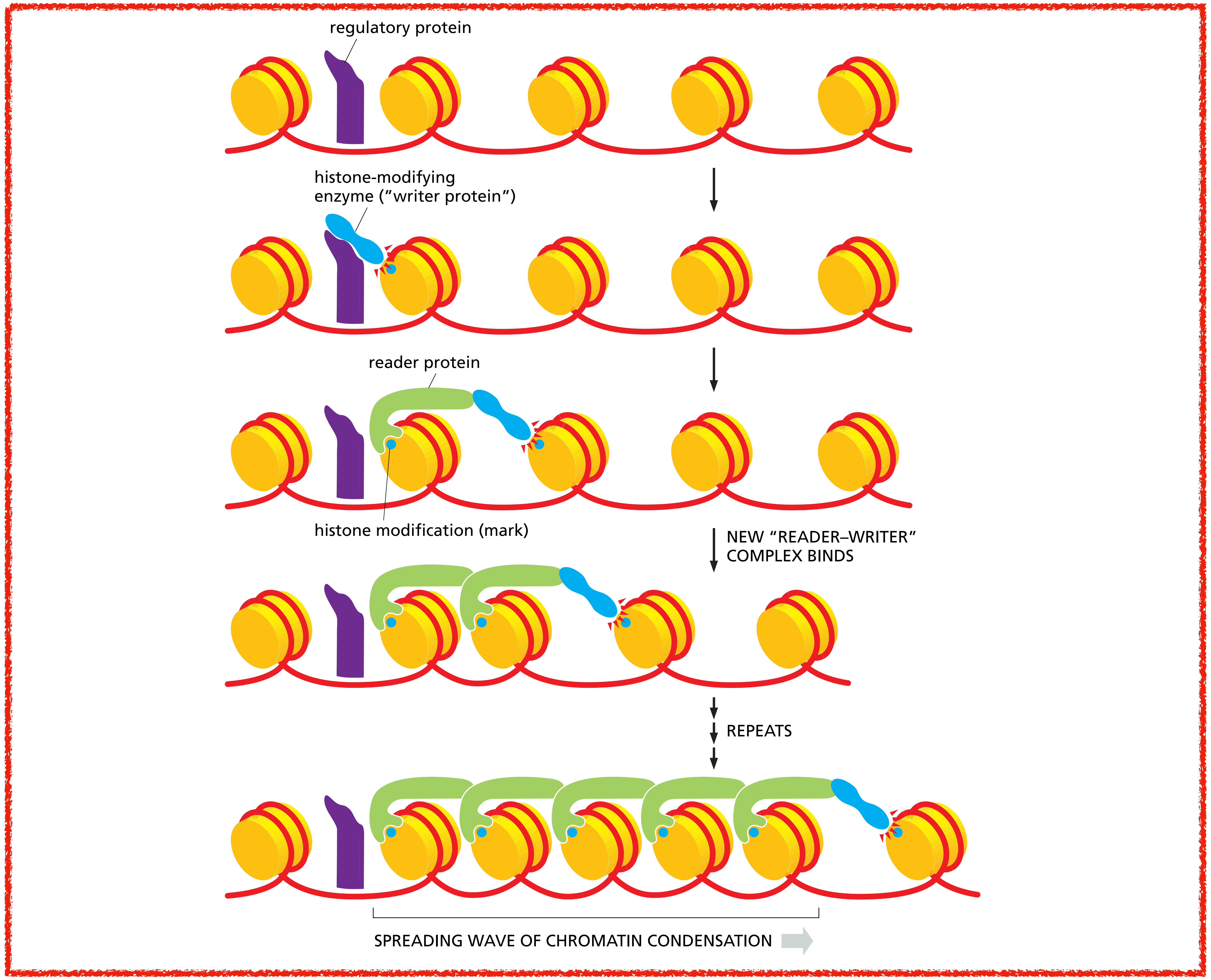
Reader–writer complexes refer to a functional unit in which one protein (or domain) reads an existing histone modification, while another protein (or domain) within the same complex writes (adds) the same modification onto nearby histones.

This mechanism allows for self-propagation and spreading of chromatin states across nucleosomes.

- **Reader–writer coupling**
    - When the reader recruits or stabilizes the writer at chromatin sites, the modification can be copied onto adjacent nucleosomes.
    - This creates a positive feedback loop that spreads and maintains chromatin domains (e.g., heterochromatin).



The spread of heterochromatin (position effect)



What blocks the spread of heterochromatin?

- This can be problematic as each chromosome is one long stretch of DNA
- **DNA sequences** have been identified and mark the boundaries of chromatin domains = barrier sequences. If those are deleted, euchromatin can be invaded by heterochromatin
- Cluster of binding sites for **acetylation enzymes**
- Acetylation of a lysine side-chain is **not compatible** with methylation of the same side-chain
- Specific **lysine methylations are required to spread heterochromatin** so histone **acetylases** are logical candidates for barrier regions

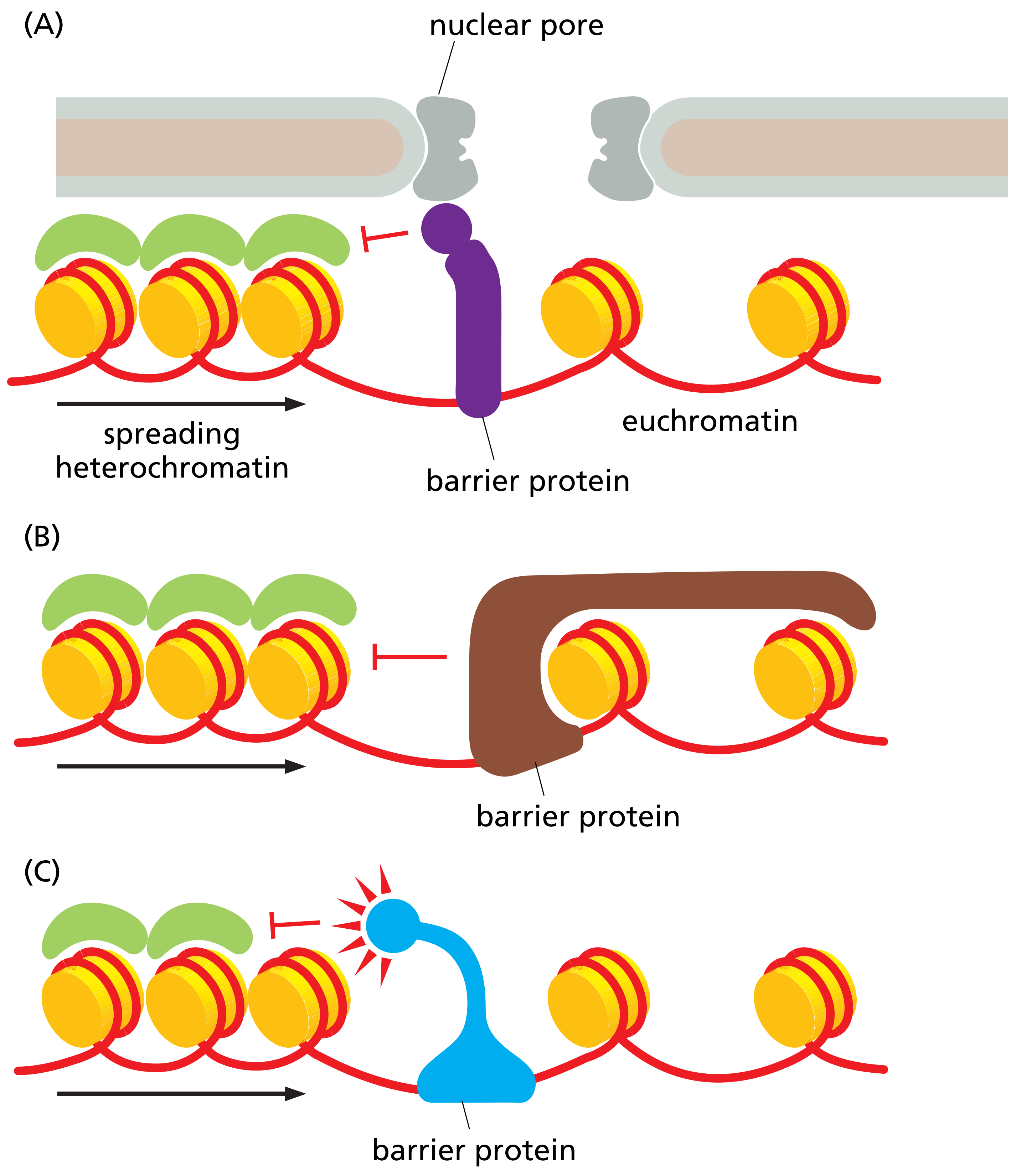
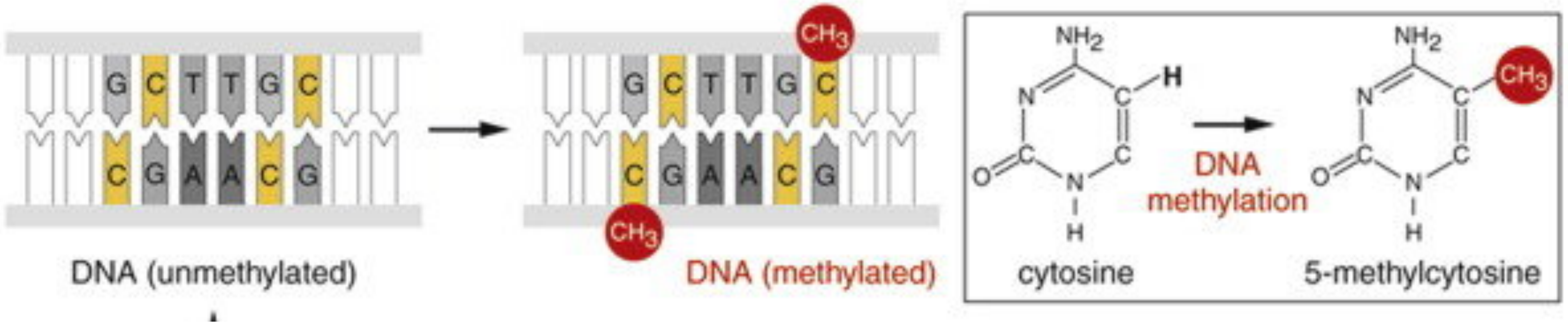


Figure 4–41 Some mechanisms of barrier action. These models are derived from experimental analyses of barrier action, and a combination of several of them may function at any one site. (A) The tethering of a region of chromatin to a large fixed site, such as the nuclear pore complex illustrated here, can form a barrier that stops the spread of heterochromatin. (B) The tight binding of barrier proteins to a group of nucleosomes can make this chromatin resistant to heterochromatin spreading. (C) By recruiting a group of highly active histone-modifying enzymes, barriers can erase the histone marks that are required for heterochromatin to spread. For example, a potent acetylation of lysine 9 on histone H3 will compete with lysine 9 methylation, thereby preventing the binding of the HP1 protein needed to form a major form of heterochromatin. (Based on A.G. West and P. Fraser, *Hum. Mol. Genet.* 14:R101–R111, 2005. With permission from Oxford University Press.)

In addition to histones, is DNA also modified?

- **DNA methylation** on **cytosine**
- DNA methylation influences **gene expression** but not directly the **chromatin organisation** (hetero- and euchromatin)



DOI:10.1016/j.pbiomolbio.2015.02.009

Genetics vs. Epigenetics

Genetic code	Epigenetic code
Conserved during division	Semi-conserved during division
Strong evolutionary conservation	Rapid evolutionary divergence
Identical in all cell types	Differs between cell types, responsive to environment

Have a nice day!