Network algorithms
for molecular biology

Course organised by members of Inria European Team ERABLE

Physically located and also part of
Laboratory of Biometry and Evolutionary Biology
CNRS UMR 5558 / University Lyon 1

And including researchers at
University La Sapienza of Rome
Universities of Florence and Pisa
Center for Mathematics and Computer Science (CWI) Amsterdam
Free University of Amsterdam

Organisation of the course – Who will be teaching

Myself – Marie-France Sagot, Director of Research Inria

Arnaud Mary, Associate Professor UCL

Blerina Sinaimeri, Junior Researcher Inria

All three members of ERABLE

Organisation of the course – Program

Before Christmas: 12 courses of 2 hours each, on Thursday mornings, 8-10am
In January: Final evaluation (see next slide)

Schedule (in black: MFS; in red: AM; in blue: BS)

Sept 17	Presentation of the course and general introduction to biology
Sept 24	General overview of networks (graphs) in biology & associated algorithms
Oct 1	General introduction to enumeration algorithms
Oct 8	Enumeration + Motifs in networks
Oct 15	Motifs in networks
Oct 22	Cycles and *st*-paths in NGS-related graphs
Nov 5	Cycles and *st*-paths in NGS-related graphs
Nov 12	Phylogenetic networks
Nov 19	Co-phylogenetic networks
Nov 26	Metabolic networks and precursor sets (as a prelude to species interactions)
Dec 3	Metabolic networks and precursor sets (as a prelude to species interactions)
Dec 17	Metabolic stories

Organisation of the course – Evaluation

Two types of evaluation – May be adapted depending on how many attend!

Continuous

Will consist mainly in exercises to be done at home possibly accompanied by short presentations to be done in class

Final

Report + presentation of a paper with open problem(s) and attempts to address such

or

Report + presentation of an algorithmic project developed on a topic related to those given in the course

In both cases, choice should be discussed with us and made before December 1st

Master research training period

In case of an interest in doing the Master research training in computational biology

Apart from our own Inria research group that can greet Master students

There are other groups in Lyon, France or abroad who might interest you

In the first case, talk to us

In the second case, we can give you suggestions of appropriate groups, so talk also to us

Notice that in France, there is a rather large community of persons working in computational biology, including from a computer science perspective

There is even, since 2000, an annual conference called JOBIM

Today

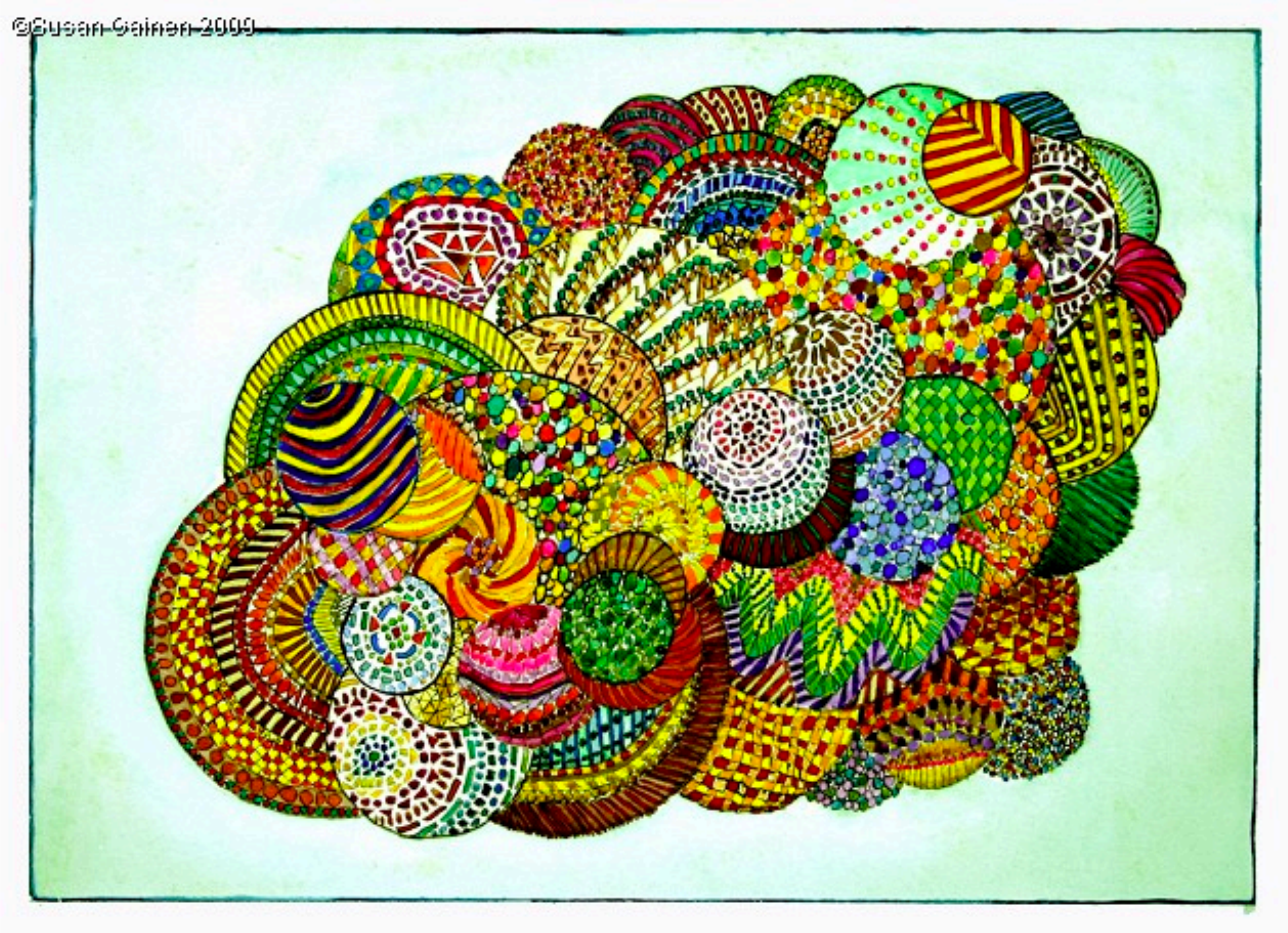
General introduction to biology

The idea is to give you just a very broad overview that will enable you to acquire the basic vocabulary and concepts

Next week, I'll get more in detail on the various uses of networks/graphs in biology before we focus on the algorithmics of some more specific cases

Some more biological concepts will be introduced later as needed

Basics of molecular biology



The cell

Cells are the fundamental working units of every living system
Smallest structural unit of an organism capable of independent functioning

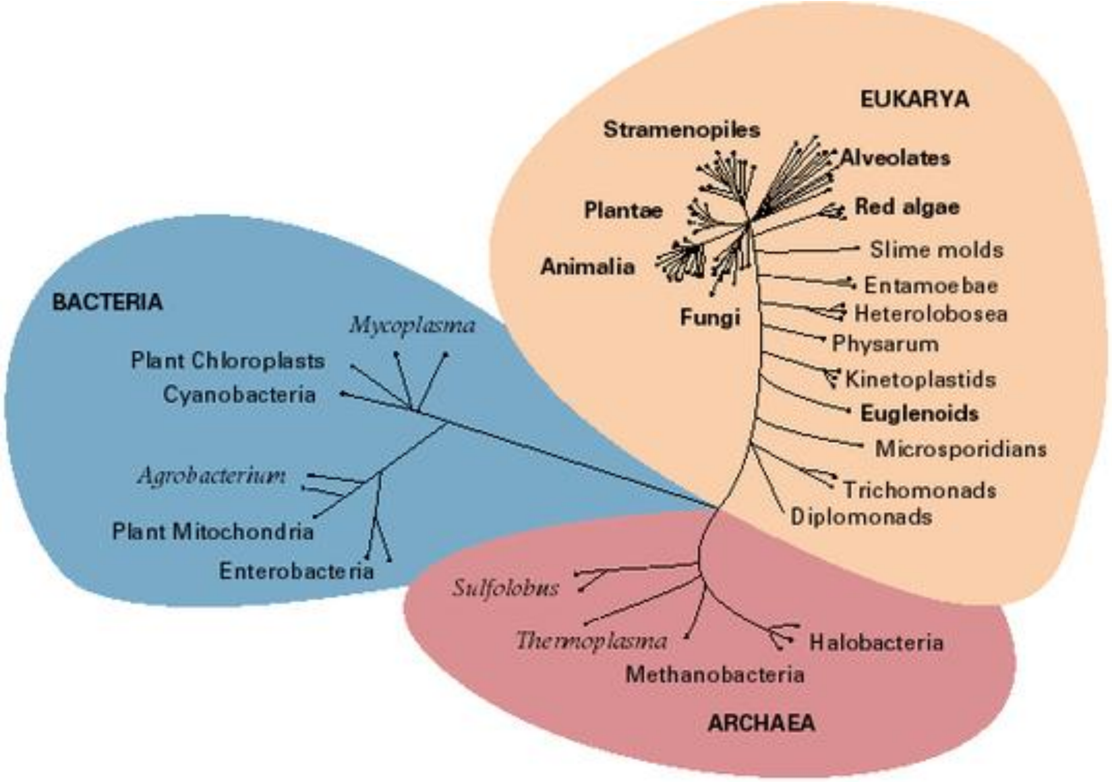
All organisms are made of 1 or more cells

Prokaryotes and eukaryotes

According to most recent evidence, there are three main branches to the tree of life

Prokaryotes which include Archaea ("ancient ones") and Bacteria

Eukaryotes (Eukarya) which include plants, animals, fungi, and certain algae



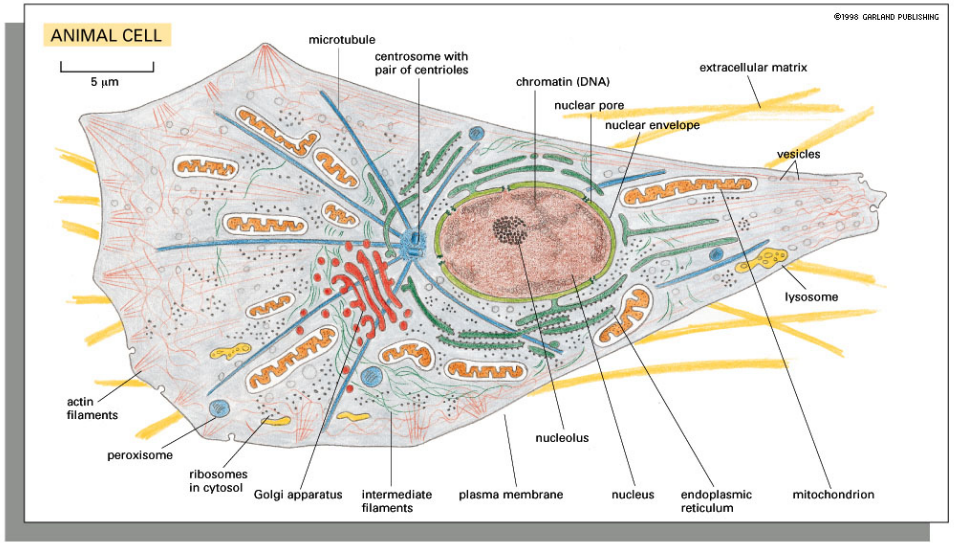
Main differences

Prokaryotes	Eukaryotes
Single cell	Single or multi cell
No nucleus	Nucleus
One single piece of circular DNA = one single chromosome	Chromosomes
No organelles	Organelles

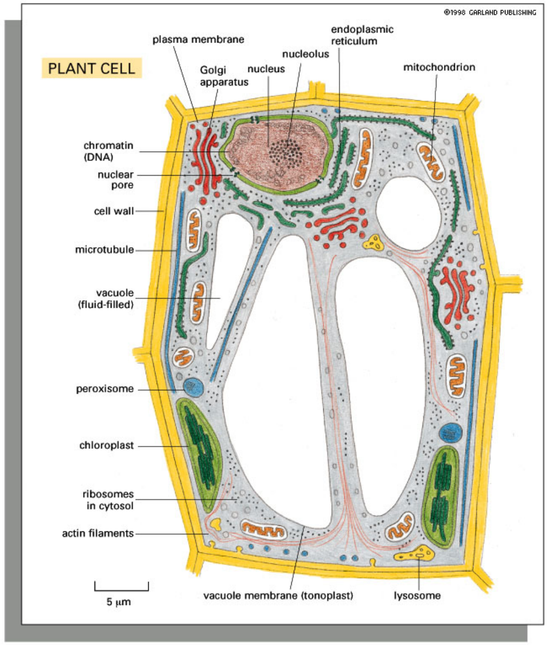
Organelle = Specialised compartment with a specific function

Examples of cells

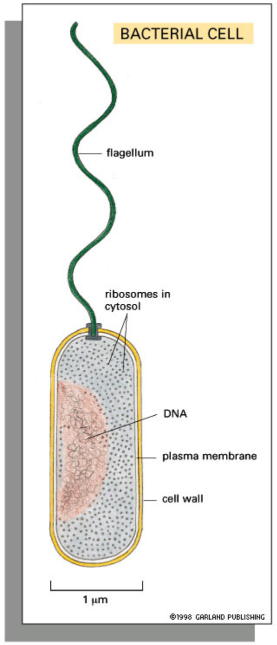
Animal



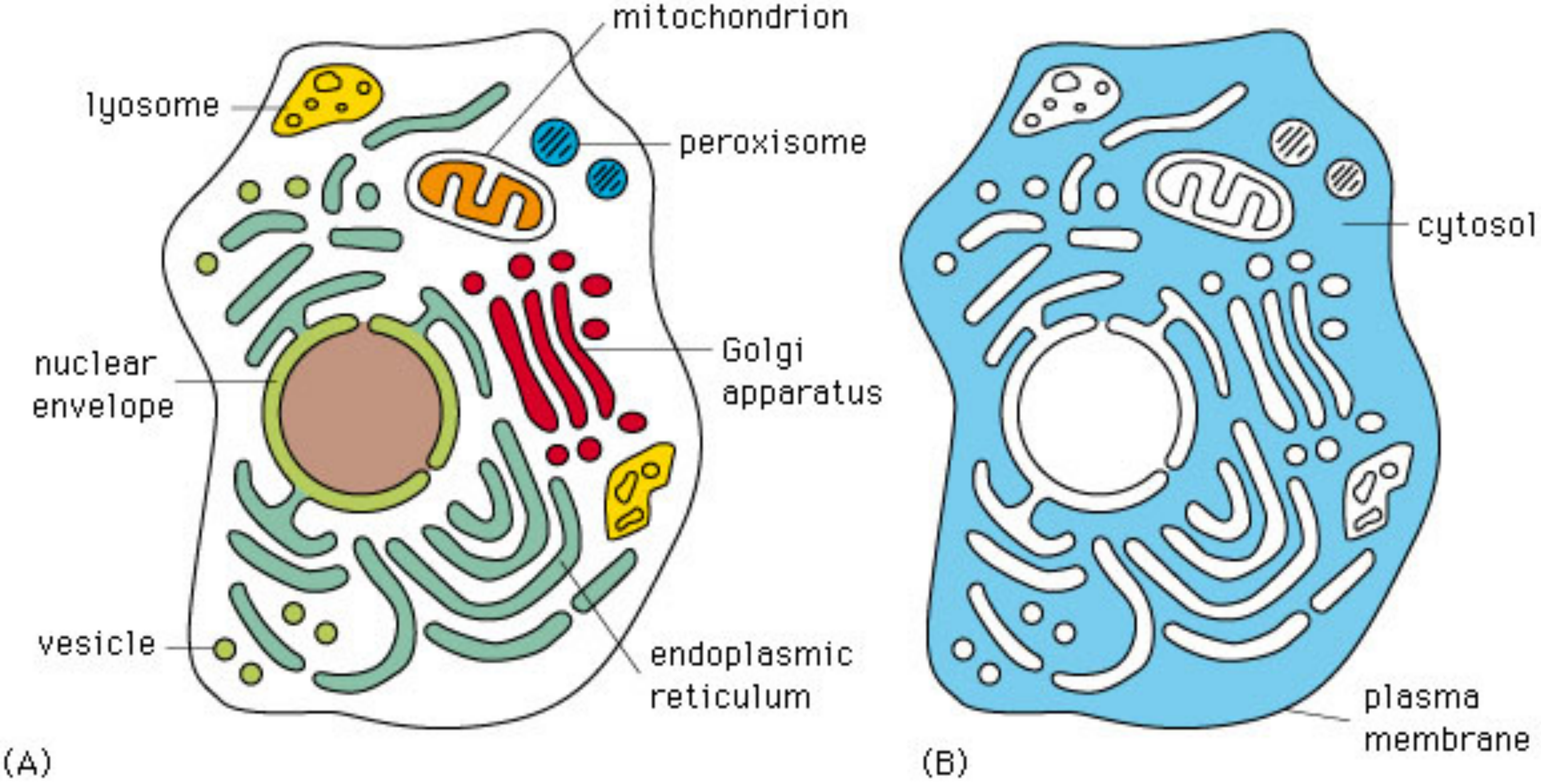
Plant



Bacterium



Compartmentation of the eukaryote cell: Various organelles

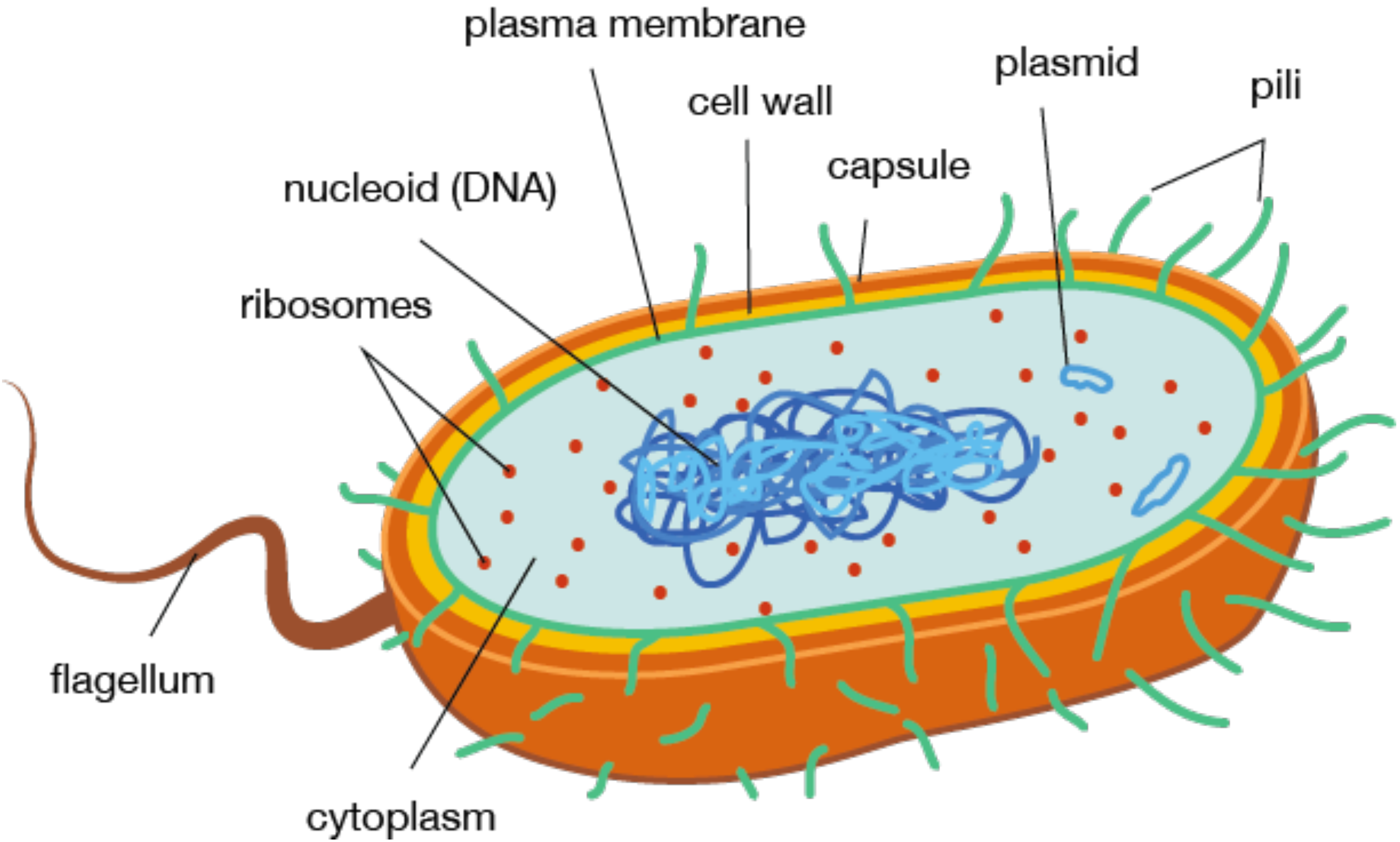


©1998 GARLAND PUBLISHING

Main functions of the different compartments just to show that they vary greatly

Nucleus	contains main genome DNA and RNA synthesis
Cytosol	contains many metabolic pathways protein synthesis
Endoplasmic reticulum (ER)	synthesis of most lipids synthesis of proteins for distribution to many organelles and plasma membrane
Golgi apparatus	modification, sorting, and packaging of proteins and lipids for either secretion or delivery to another organelle
Lysosomes	intracellular degradation
Endosomes	sorting of endocytosed material
Mitochondria	ATP synthesis by oxidative phosphorylation
Chloroplasts	ATP synthesis and carbon fixation by photosynthesis
Peroxisomes	oxidation of toxic molecules

Meanwhile in prokaryotic cells, the picture is much different



In both cases:
Main composition of a cell

70% water

23% macromolecules
- Proteins
- Polysaccharides
- Lipids

7% small molecules
- Salts
- Lipids
- Amino acids
- Nucleotides

Genetic information is stored in DNA – Deoxyribonucleic Acid

Consists of two biopolymer strands coiled around each other to form a double helix

The structure and the four genomic letters of the DNA code for all living organisms

The letters, called "nucleotides" or also "bases", are:

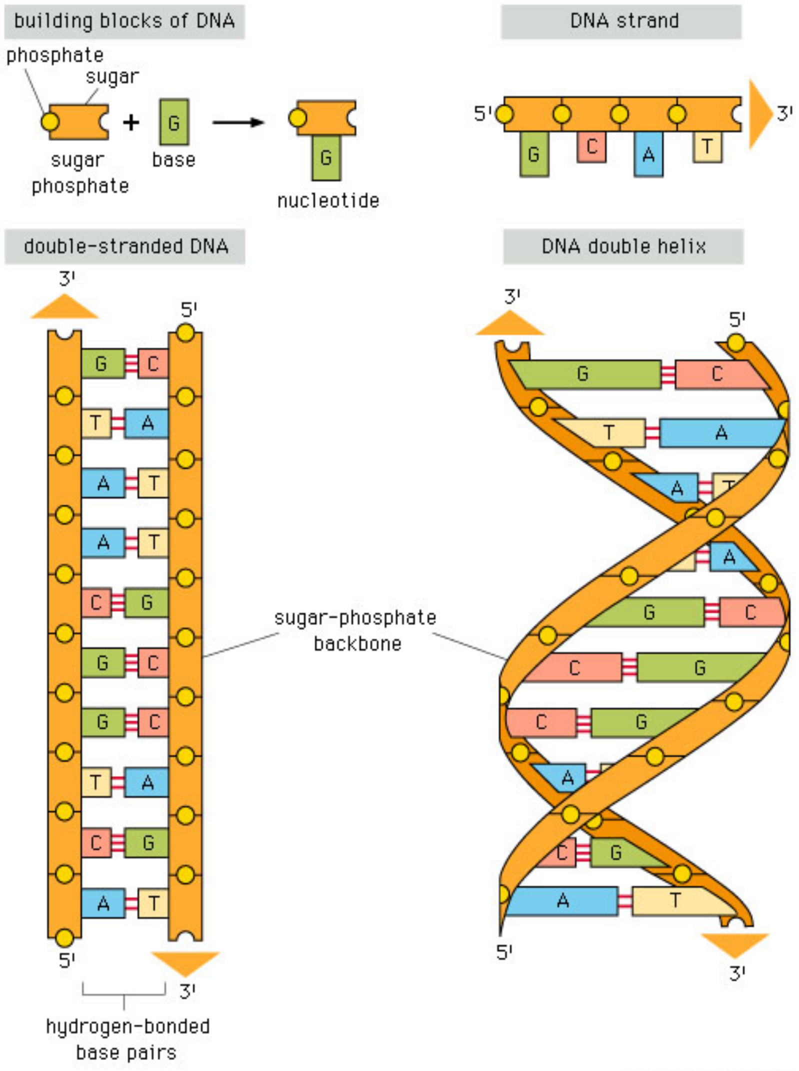
Adenine – A

Guanine – G

Thymine – T

Cytosine – C

which pair A with T and C with G on the complementary strands



©1998 GARLAND PUBLISHING

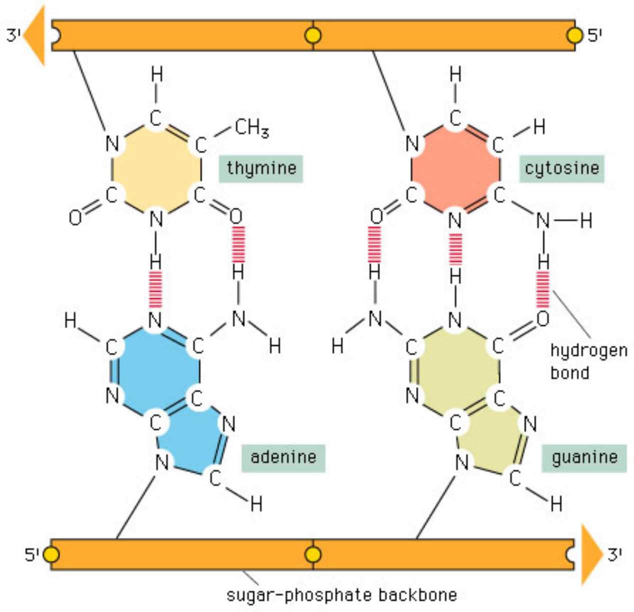
DNA has an orientation

Actually, the double helix structure of DNA is composed of

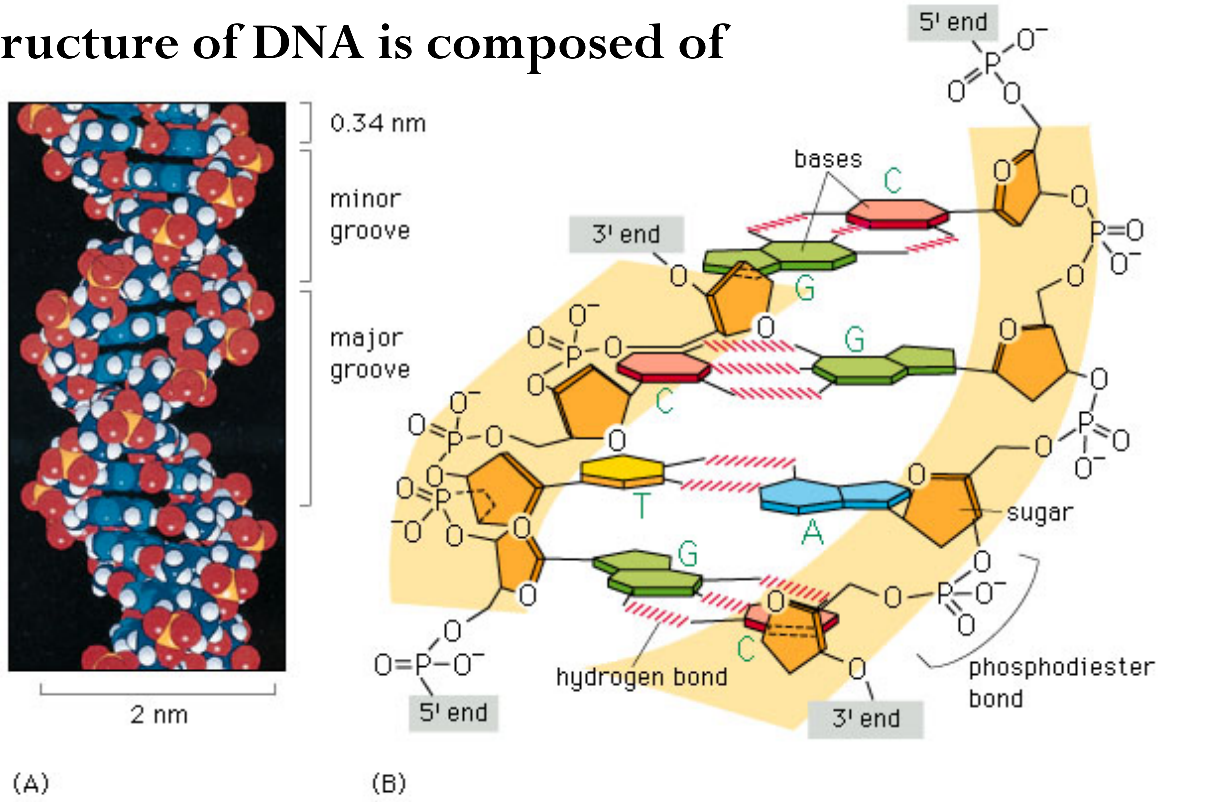
a base (A,C,G,T)

a sugar molecule

a phosphate group



©1998 GARLAND PUBLISHING



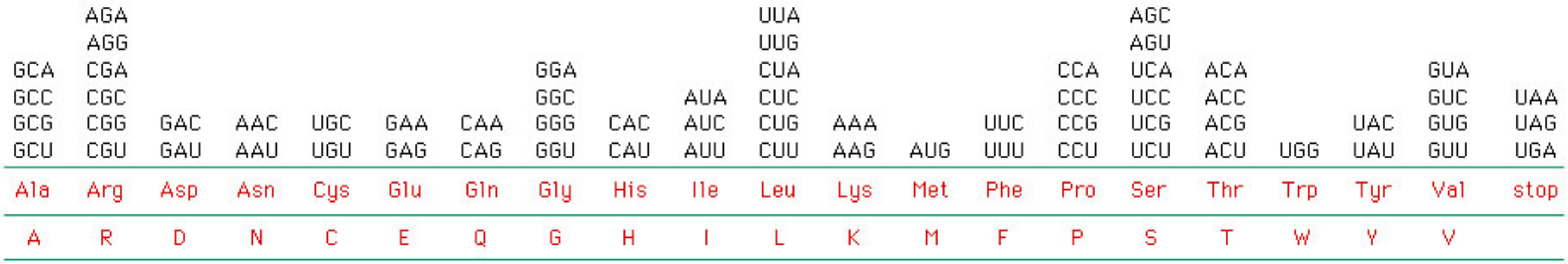
©1998 GARLAND PUBLISHING

DNA always reads from 5' end to 3' end for transcription replication (see later)

5' ATTTAGGCC 3'

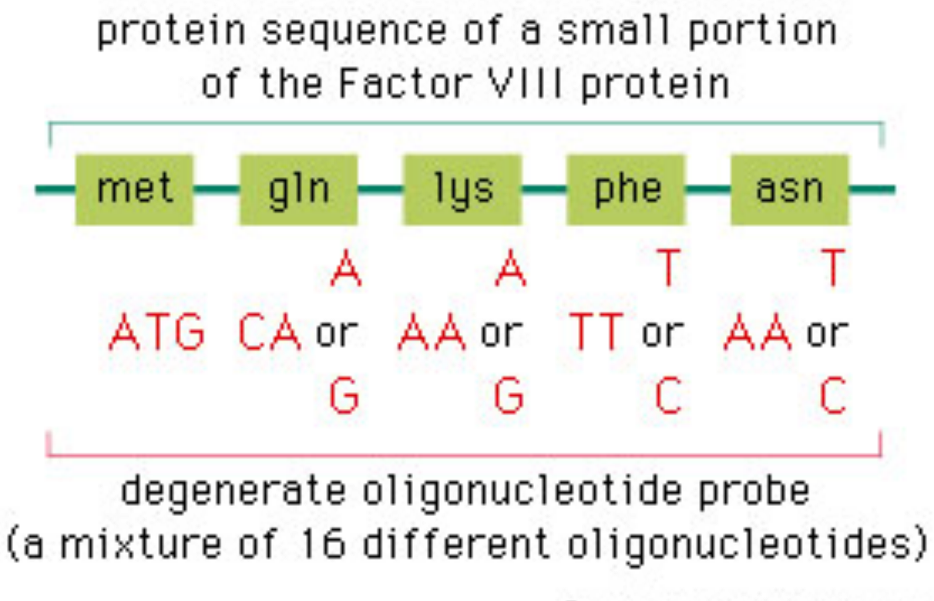
3'TAAATCCGG 5'

DNA encodes proteins



GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

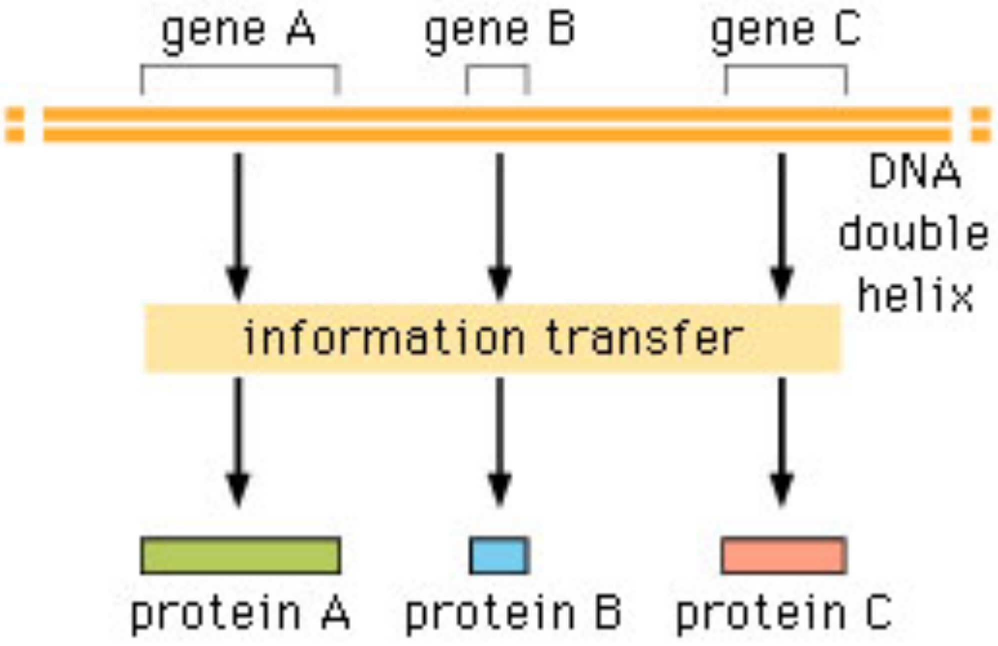
©1998 GARLAND PUBLISHING


protein sequence of a small portion of the Factor VIII protein
met
gln
lys
phe
asn
A A T T
ATG CA or AA or TT or AA or
G G C C
degenerate oligonucleotide probe
(a mixture of 16 different oligonucleotides)

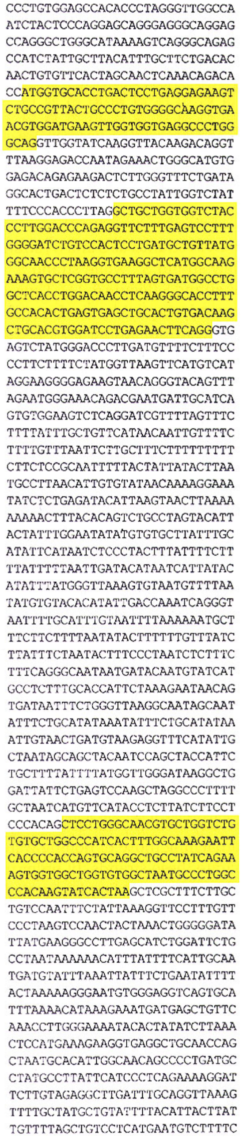
©1998 GARLAND PUBLISHING

Some more details will be given later on the process of going from DNA to protein ("genetic dogma")

Actually portions of chromosomes called genes encode proteins



©1998 GARLAND PUBLISHING

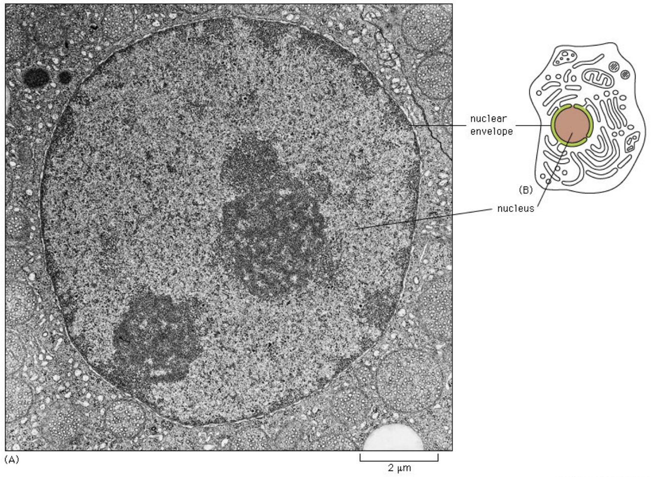


```
CCCTGTGGAGCCACACCCTAGGGTTGGCCA
ATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAG
CCATCTATTGCTTACATTTGCTTCTGACAC
AACTGTGTTCACTAGCAACTCAAACAGACA
CCATGGTGCACCTGACTCCTGAGGAGAAGT
CTGCCGTTACTGCCCTGTGGGGCAAGGTGA
ACGTGGATGAAGTTGGTGGTGAGGCCCTGG
GCAGGTTGGTATCAAGGTTACAAGACAGGT
TTAAGGAGACCAATAGAAACTGGGCATGTG
GAGACAGAGAAGACTCTTGGGTTTCTGATA
GGCACTGACTCTCTCTGCCTATTGGTCTAT
TTTCCCACCCTTAGGCTGCTGGTGGTCTAC
CCTTGGACCCAGAGGTTCTTTGAGTCCTTT
GGGGATCTGTCCACTCCTGATGCTGTTATG
GGCAACCCTAAGGTGAAGGCTCATGGCAAG
AAAGTGCTCGGTGCCTTTAGTGATGGCCTG
GCTCACCTGGACAACCTCAAGGGCACCTTT
GCCACACTGAGTGAGCTGCACTGTGACAAG
CTGCACGTGGATCCTGAGAACTTCAGGGTG
AGTCTATGGGACCCTTGATGTTTTCTTTCC
CCTTCTTTTCTATGGTTAAGTTCATGTCAT
AGGAAGGGGAGAAGTAACAGGGTACAGTTT
AGAATGGGAAACAGACGAATGATTGCATCA
GTGTGGAAGTCTCAGGATCGTTTTAGTTTC
TTTTATTTGCTGTTCATAACAATTGTTTTC
TTTTGTTTAATTCTTGCTTTCTTTTTTTTT
CTTCTCCGCAATTTTTACTATTATACTTAA
TGCCTTAACATTGTGTATAACAAAAGGAAA
TATCTCTGAGATACATTAAGTAACTTAAAA
AAAAACTTTACACAGTCTGCCTAGTACATT
ACTATTTGGAATATATGTGTGCTTATTTGC
ATATTCATAATCTCCCTACTTTATTTTCTT
TTATTTTTAATTGATACATAATCATTATAC
ATATTTATGGGTTAAAGTGTAATGTTTTAA
TATGTGTACACATATTGACCAAATCAGGGT
AATTTTGCATTTGTAATTTTAAAAAATGCT
TTCTTCTTTTAATATACTTTTTTGTTTATC
TTATTTCTAATACTTTCCCTAATCTCTTTC
TTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAG
TGATAATTTCTGGGTTAAGGCAATAGCAAT
ATTTCTGCATATAAATATTTCTGCATATAA
ATTGTAACTGATGTAAGAGGTTTCATATTG
CTAATAGCAGCTACAATCCAGCTACCATTC
TGCTTTTATTTTATGGTTGGGATAAGGCTG
GATTATTCTGAGTCCAAGCTAGGCCCTTTT
GCTAATCATGTTCATACCTCTTATCTTCCT
CCCACAGCTCCTGGGCAACGTGCTGGTCTG
TGTGCTGGCCCATCACTTTGGCAAAGAATT
CACCCCACCAGTGCAGGCTGCCTATCAGAA
AGTGGTGGCTGGTGTGGCTAATGCCCTGGC
CCACAAGTATCACTAAGCTCGCTTTCTTGC
TGTCCAATTTCTATTAAAGGTTCCTTTGTT
CCCTAAGTCCAACTACTAAACTGGGGGATA
TTATGAAGGGCCTTGAGCATCTGGATTCTG
CCTAATAAAAAACATTTATTTTCATTGCAA
TGATGTATTTAAATTATTTCTGAATATTTT
ACTAAAAAGGGAATGTGGAGGTCAGTGCA
TTTAAAACATAAAGAAATGATGAGCTGTTC
AAACCTTGGGAAAATACACTATATCTTAAA
CTCCATGAAAGAAGGTGAGGCTGCAACCAG
CTAATGCACATTGGCAACAGCCCCTGATGC
CTATGCCTTATTCATCCCTCAGAAAAGGAT
TCTTGTAGAGGCTTGATTTGCAGGTTAAAG
TTTTGCTATGCTGTATTTACATTACTTAT
TGTTTTAGCTGTCCTCATGAATGTCTTTTC
```

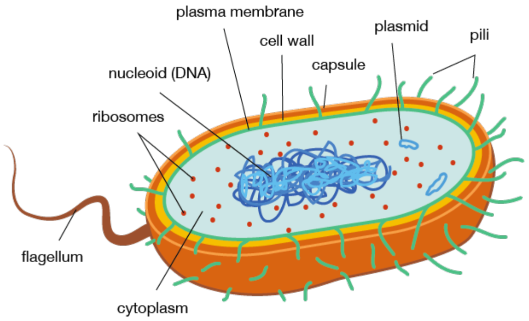
©1998 GARLAND PUBLISHING

Where genetic information is

In eukaryotes, the DNA is in the nucleus



In prokaryotes, the DNA is in the cytoplasm



A very simplified and abstract view up to now

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

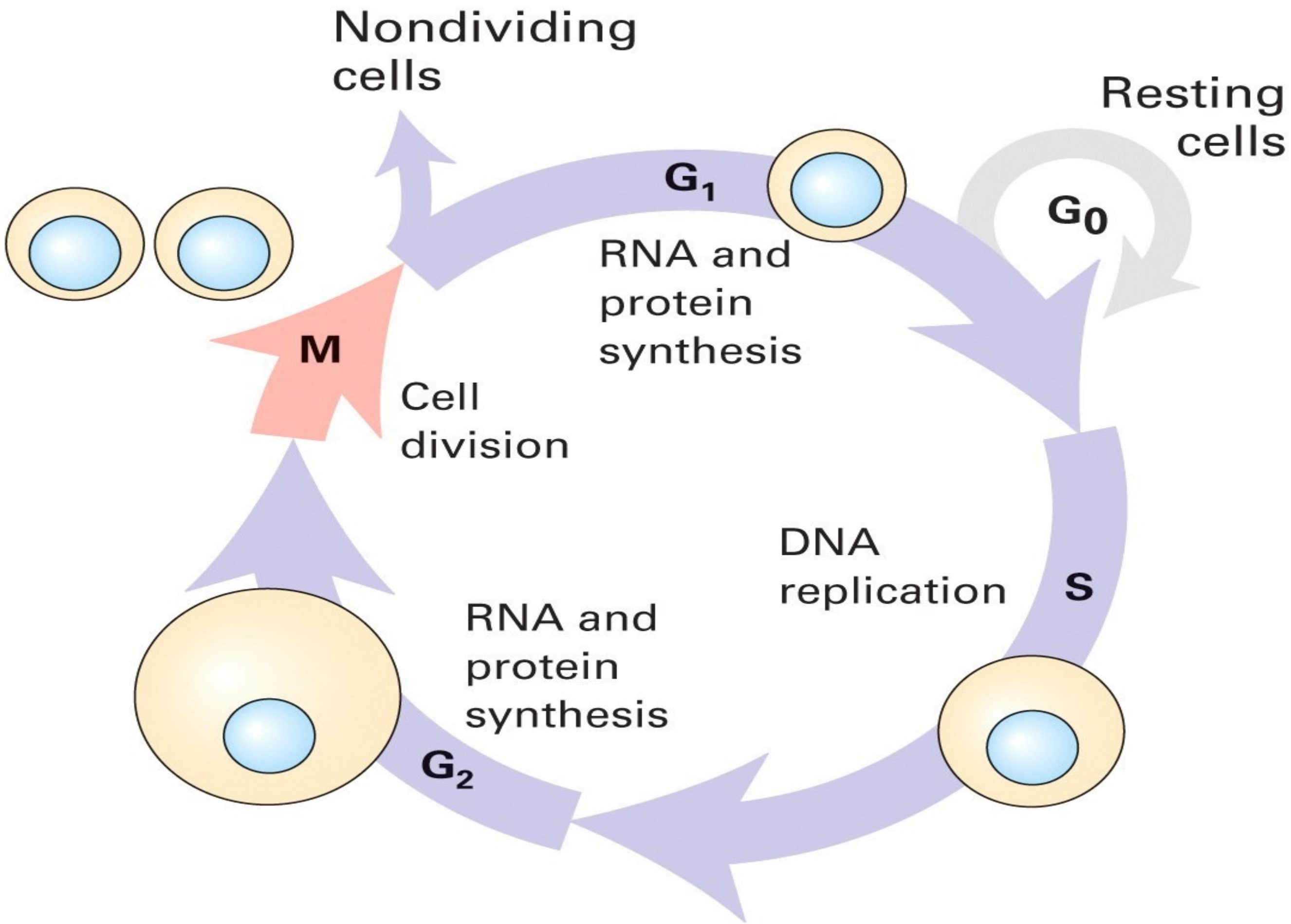
BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

All cells have common cycles

They are born, they eat, they replicate, and they die



Some cell-cycle times
just to give an idea of how much they differ

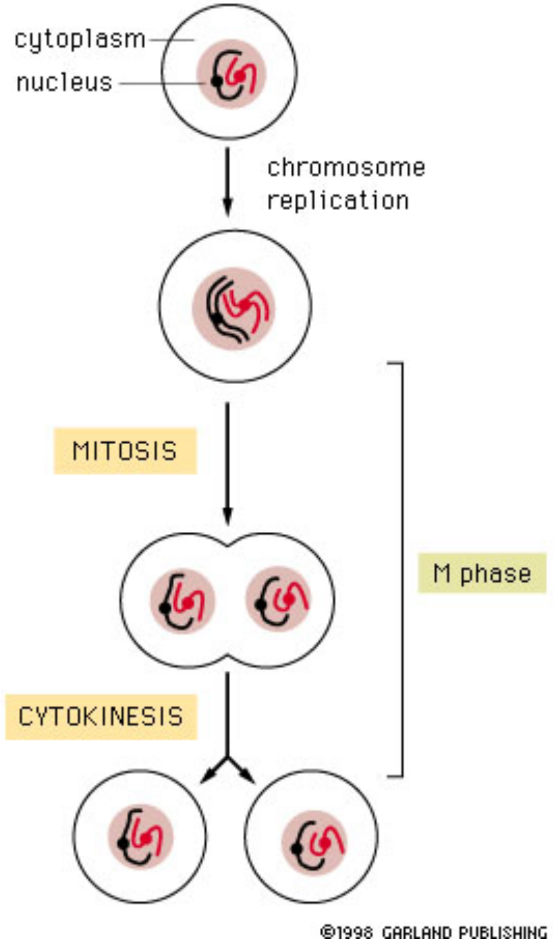
Eukaryotic cell-cycle times

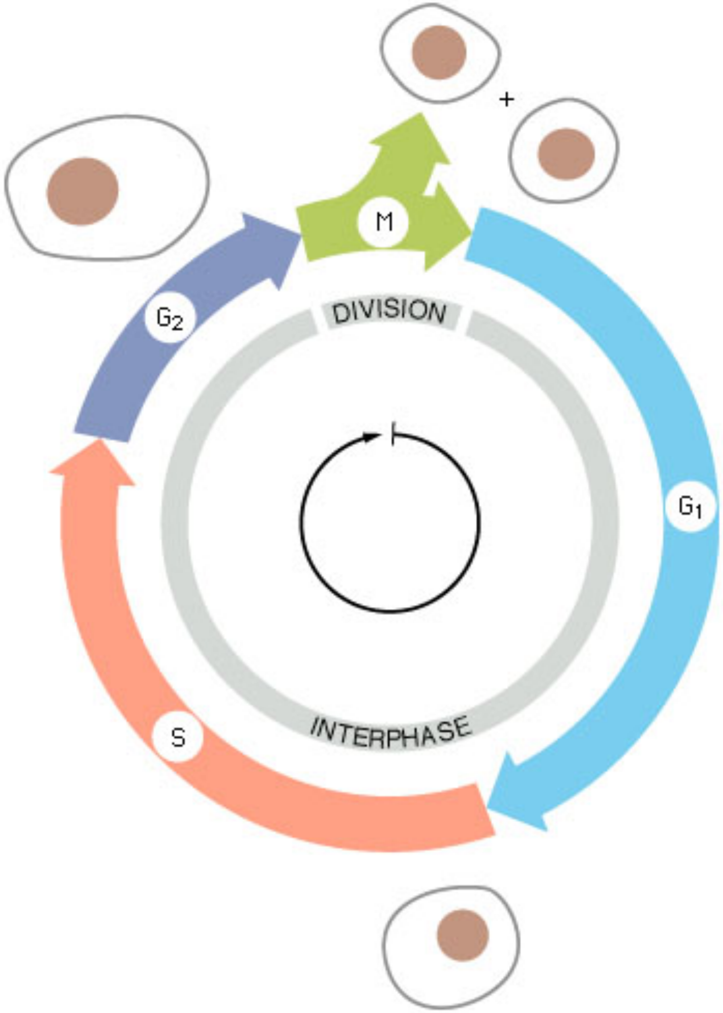
Cell Type	Cell-Cycle Times
Early frog embryo cells	30 minutes
Yeast cells	1.5-3 hours
Intestinal epithelial cells	about 12 hours
Mammalian fibroblasts in culture	about 20 hours
Human liver cells	about 1 year

Prokaryotic cell-cycle times

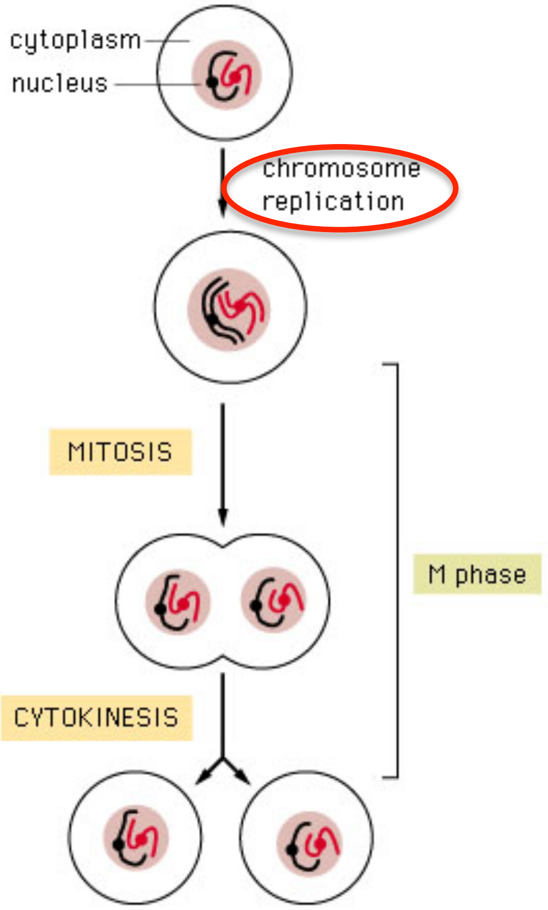
Rate at which bacteria grow and divide depends on the nature of the microbe, the ingredients of the medium in which it is grown, and the environmental conditions. E. coli, when grown in a rich medium, with plenty of aeration at 37°C is capable of dividing every 20 min

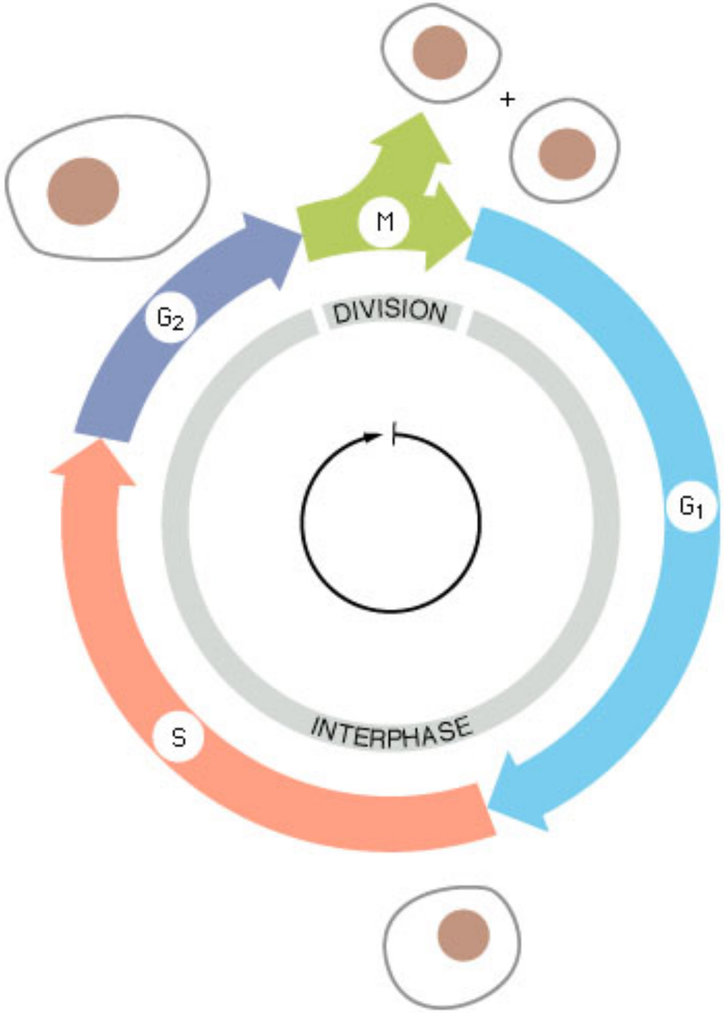
Eukaryotic cell cycle





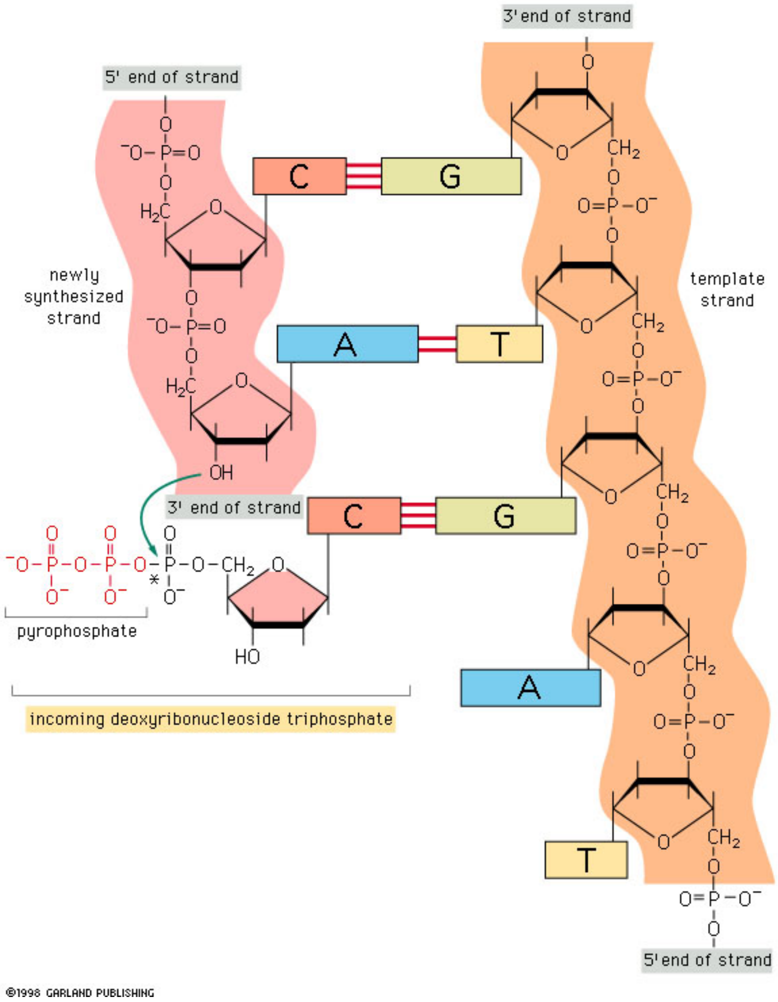
Eukaryotic cell cycle



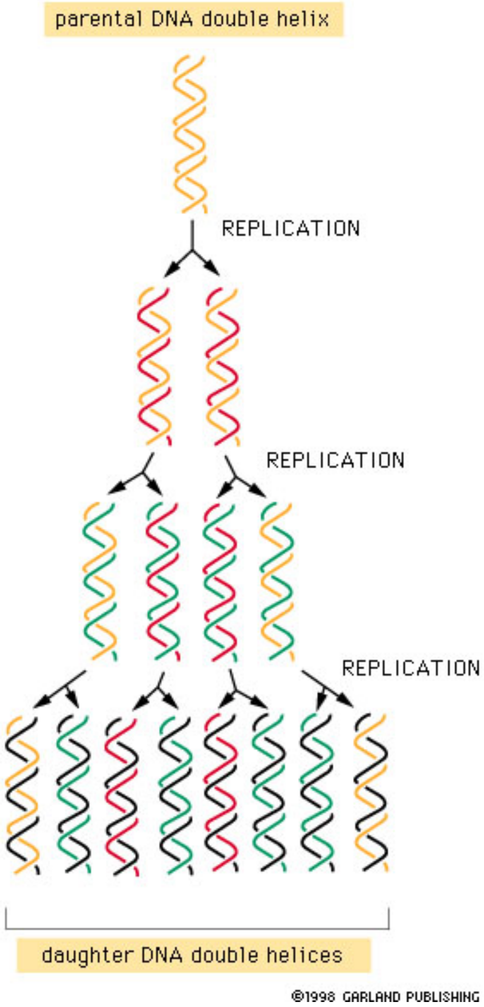


DNA replication

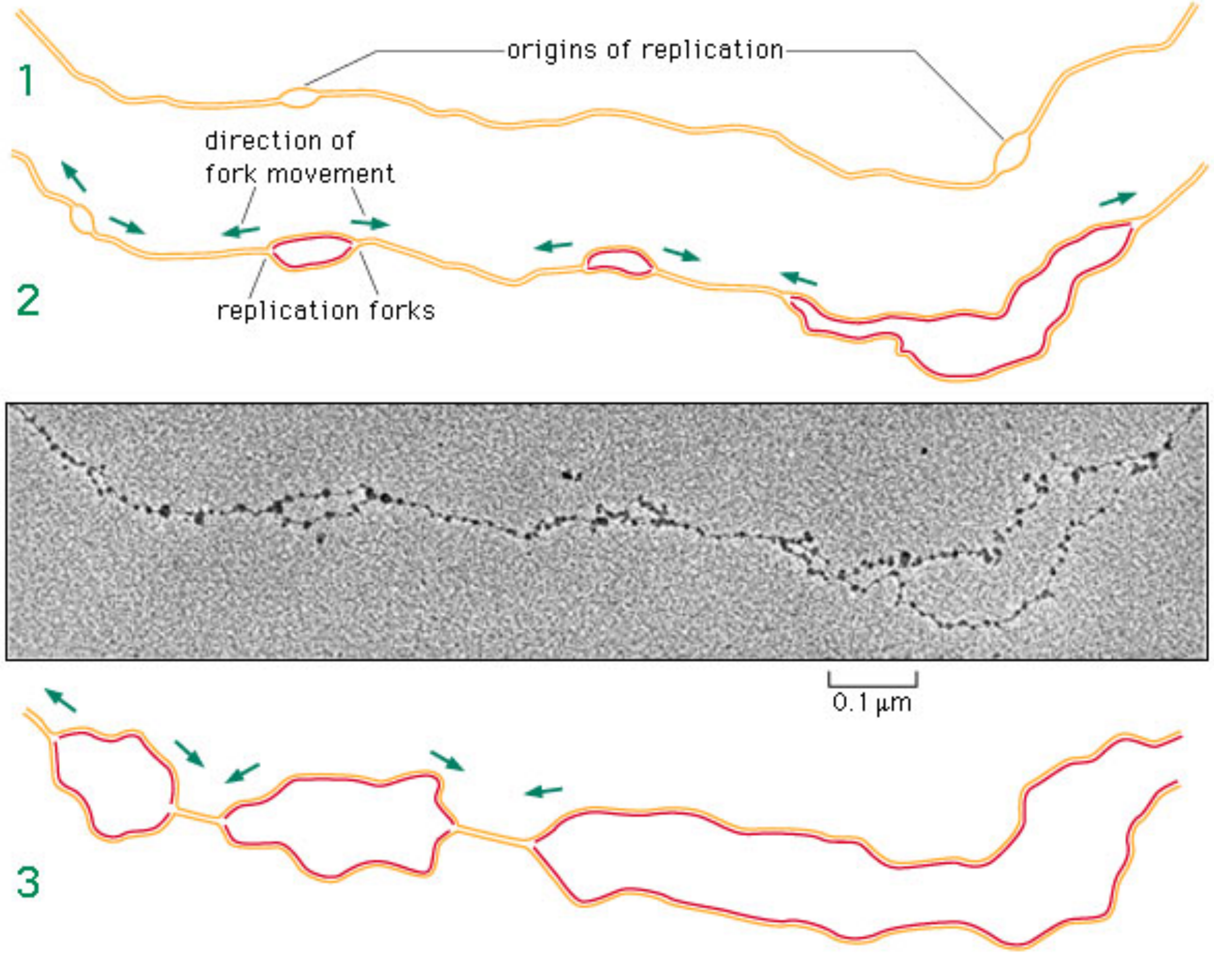
Each strand serves as template



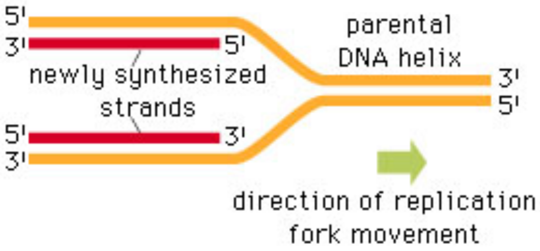
Process is semi-conservative



Replication of eukaryotic chromosomes

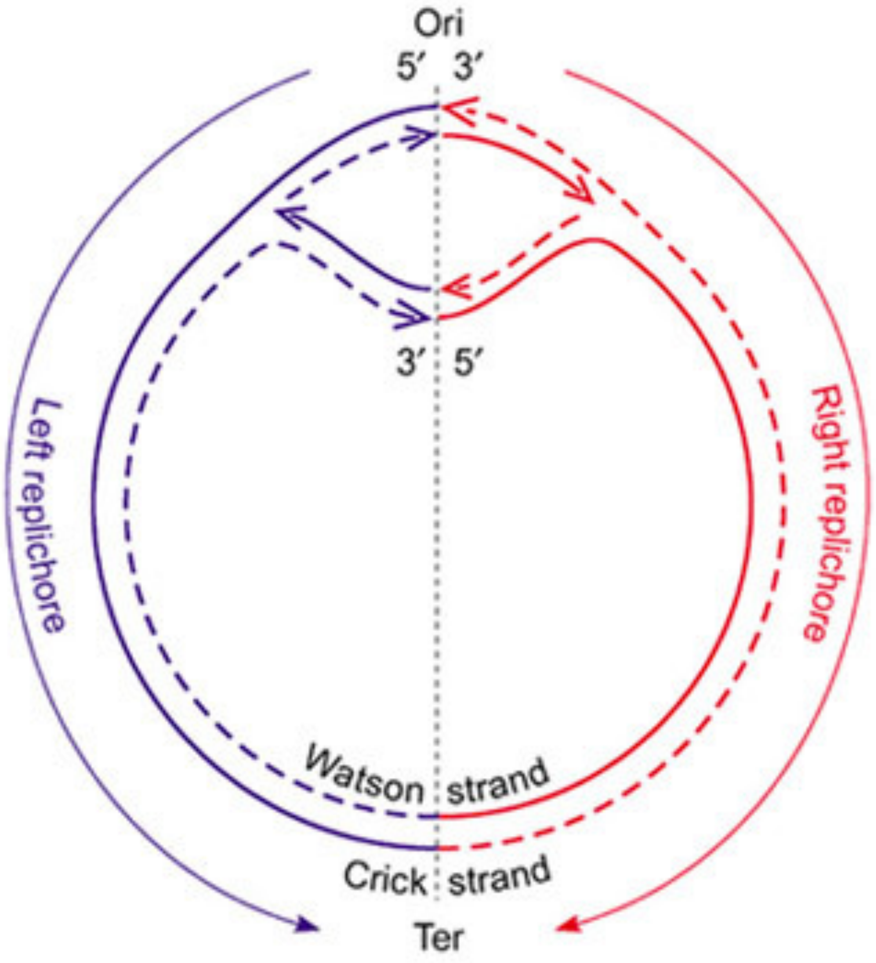


©1998 GARLAND PUBLISHING

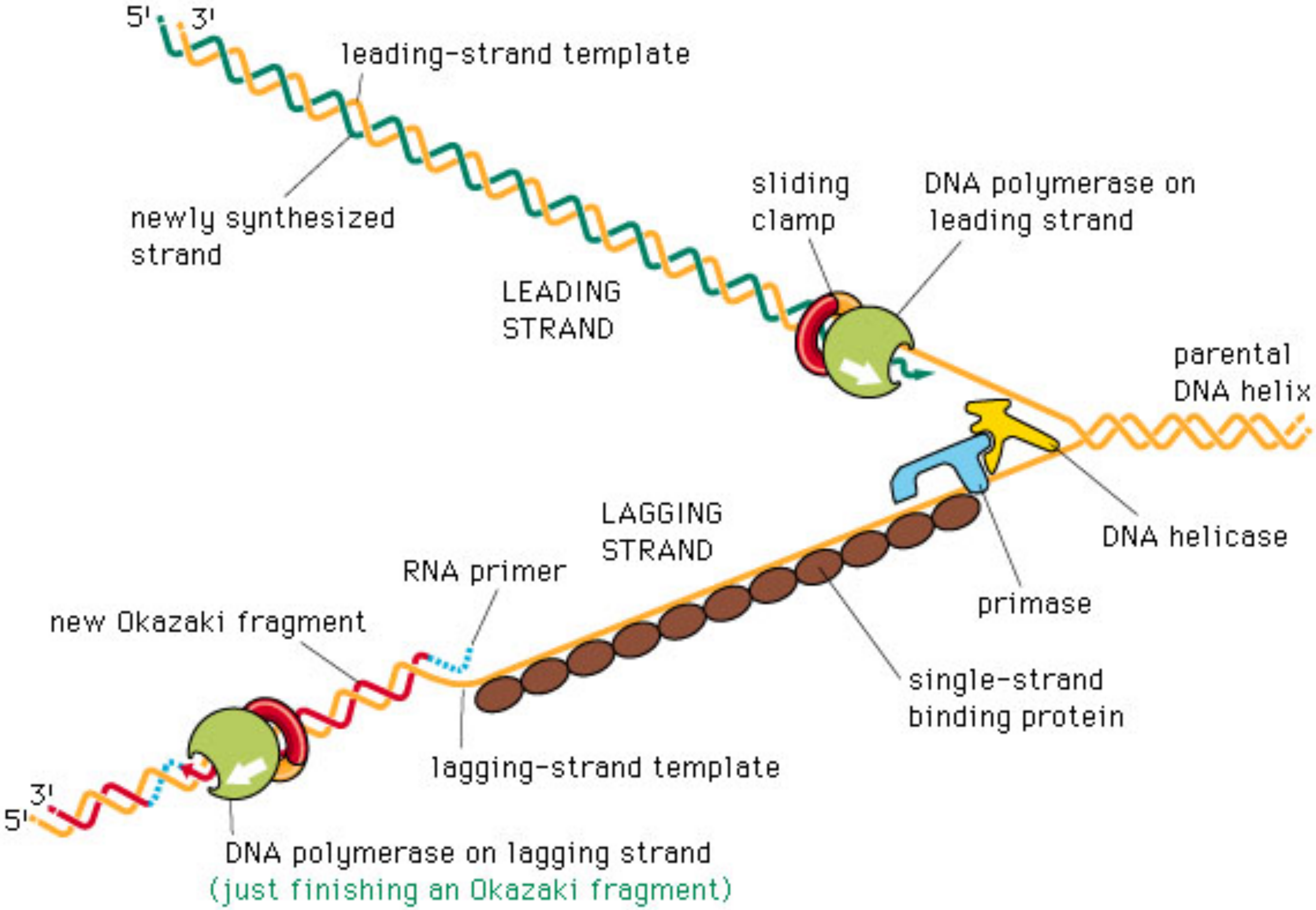


©1998 GARLAND PUBLISHING

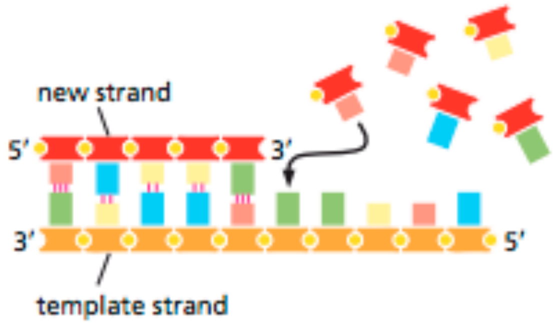
Replication of prokaryotic chromosomes



The replication fork in more detail



©1998 GARLAND PUBLISHING



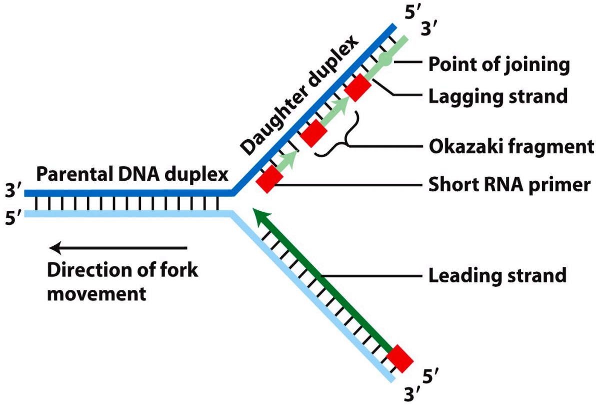
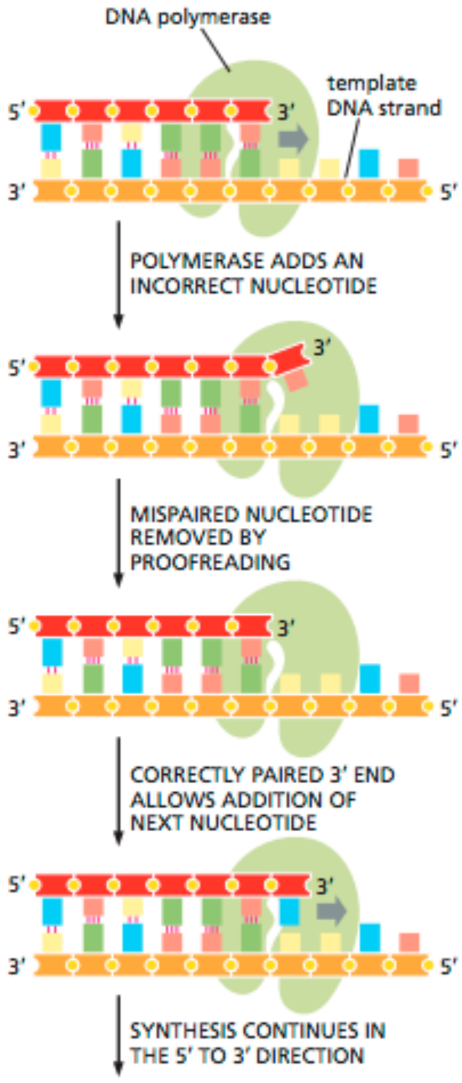


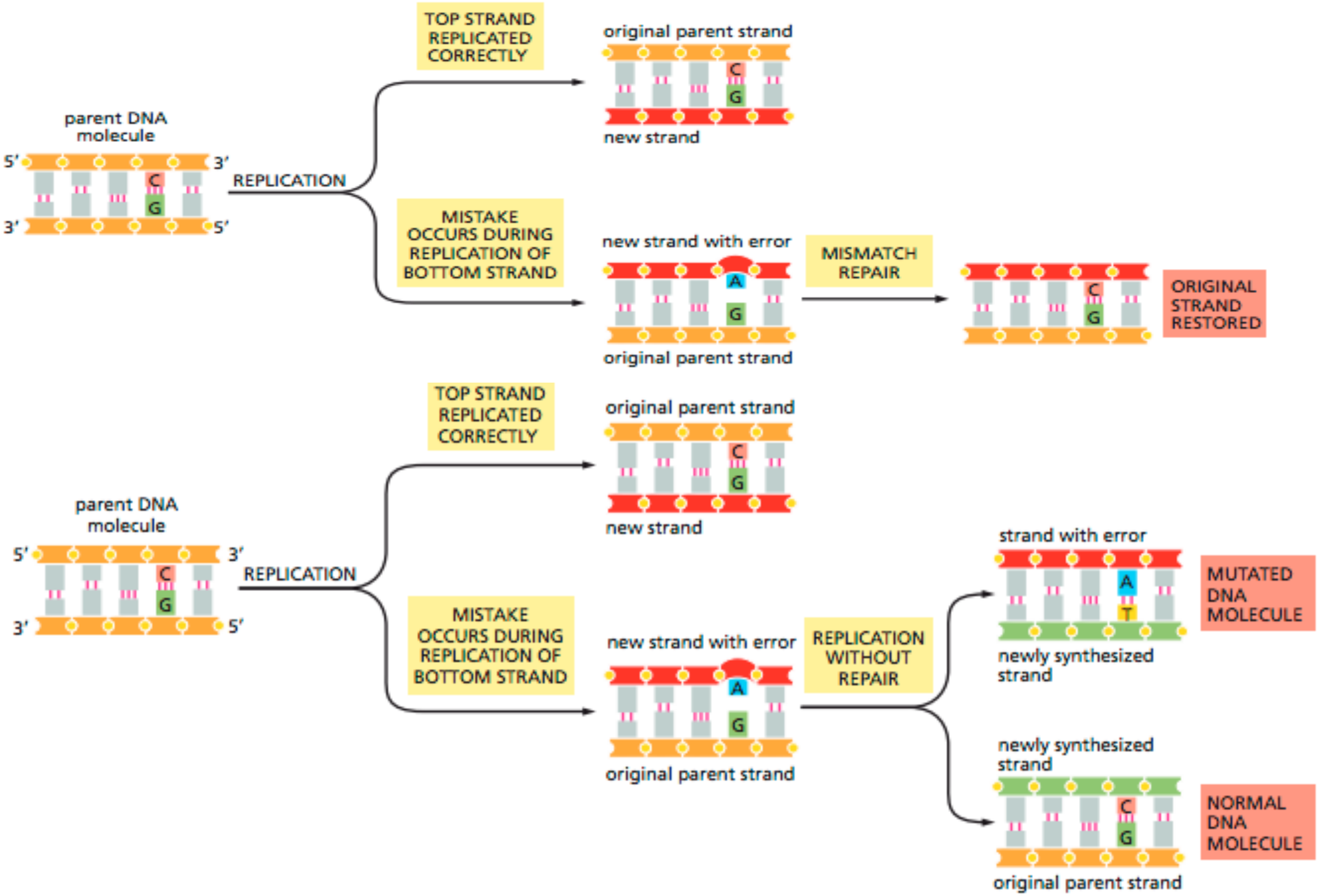
Figure 4-30
Molecular Cell Biology, Sixth Edition
© 2008 W. H. Freeman and Company

There is proof reading during DNA synthesis

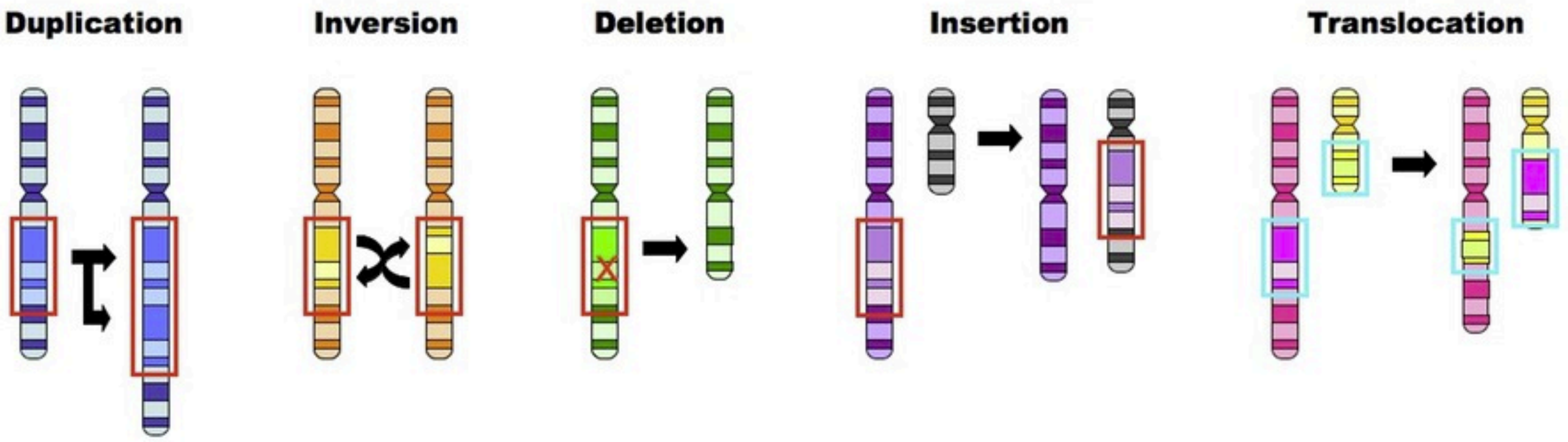

DNA polymerase
template DNA strand
5′
3′
3′
5′
POLYMERASE ADDS AN INCORRECT NUCLEOTIDE
MISPAIRED NUCLEOTIDE REMOVED BY PROOFREADING
CORRECTLY PAIRED 3′ END ALLOWS ADDITION OF NEXT NUCLEOTIDE
SYNTHESIS CONTINUES IN THE 5′ TO 3′ DIRECTION

However mistakes may remain

DNA repair and DNA mutations during replication



Other types of mutations may happen at the chromosomal level



Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

Evolution

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

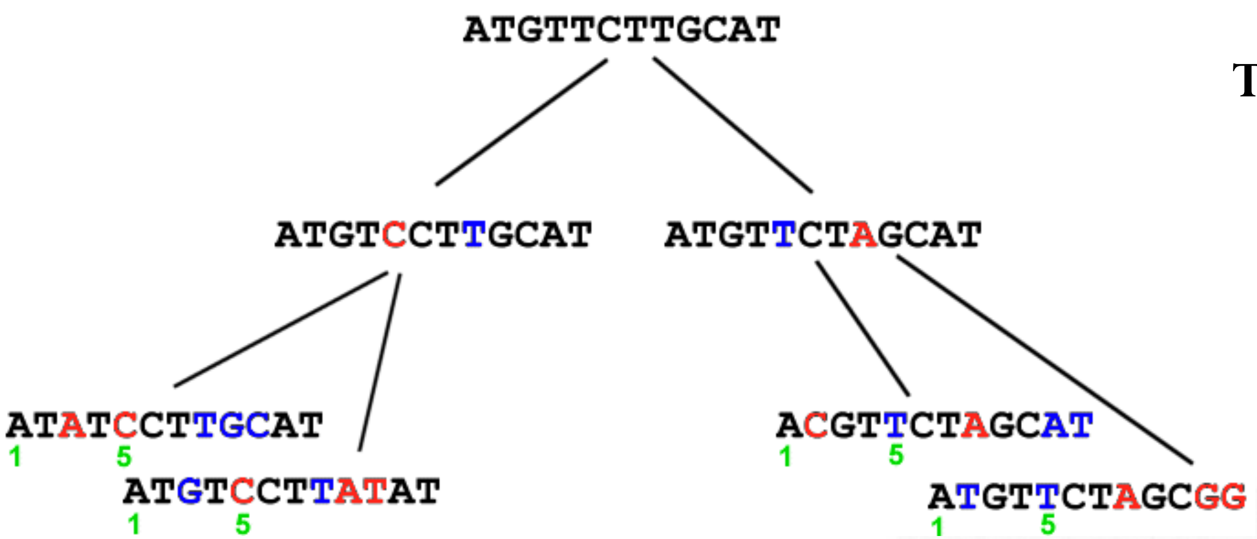
Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT such books are never static! They are in fact continuously changing in a process that may even lead to the creation of new species!

Speciation



ATGTTCTTGCAT

ATGTCCTTGCAT

ATGTTCTAGCAT

ATATCCTTGCAT
1 5

ATGTCCTTATAT
1 5

ACGTTCTAGCAT
1 5

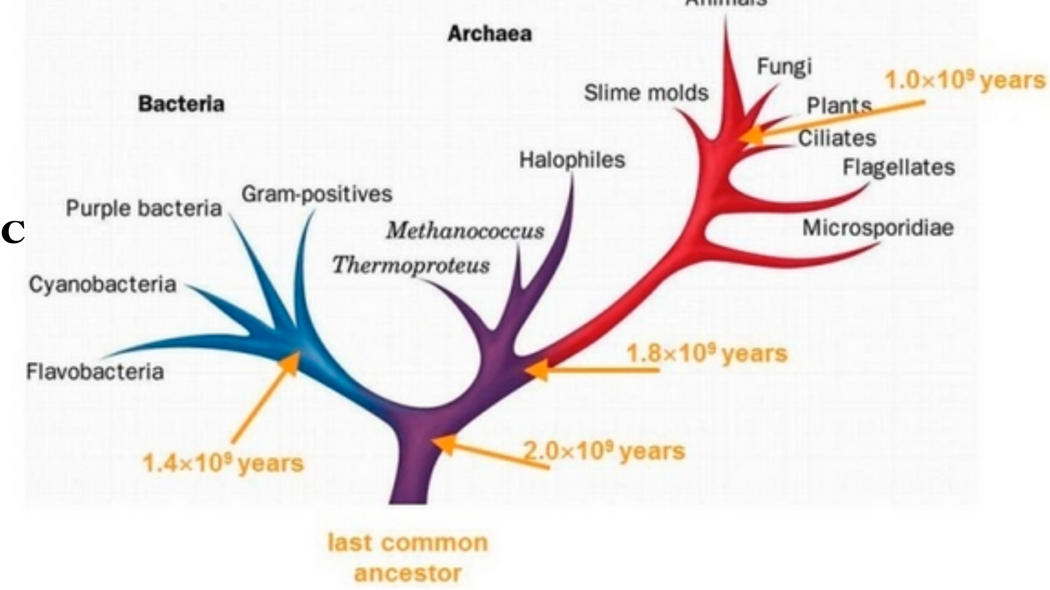
ATGTTCTAGCGG
1 5

Two main cases of speciation

There is geographic separation

There is no geographic separation

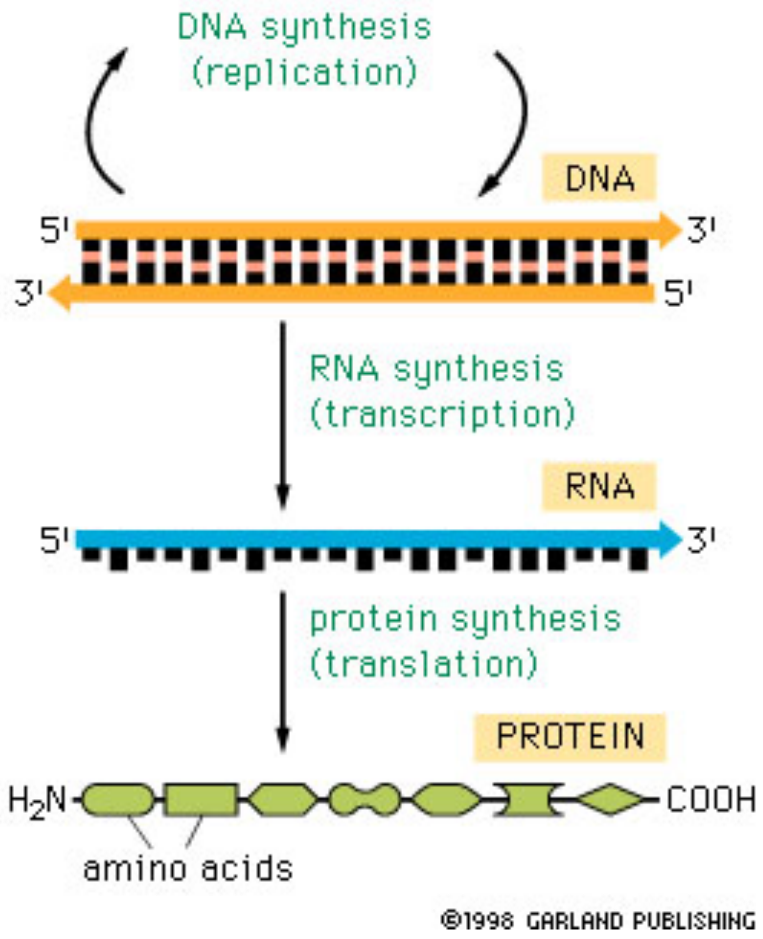
Species evolution usually represented in the form of a co-called phylogenetic tree



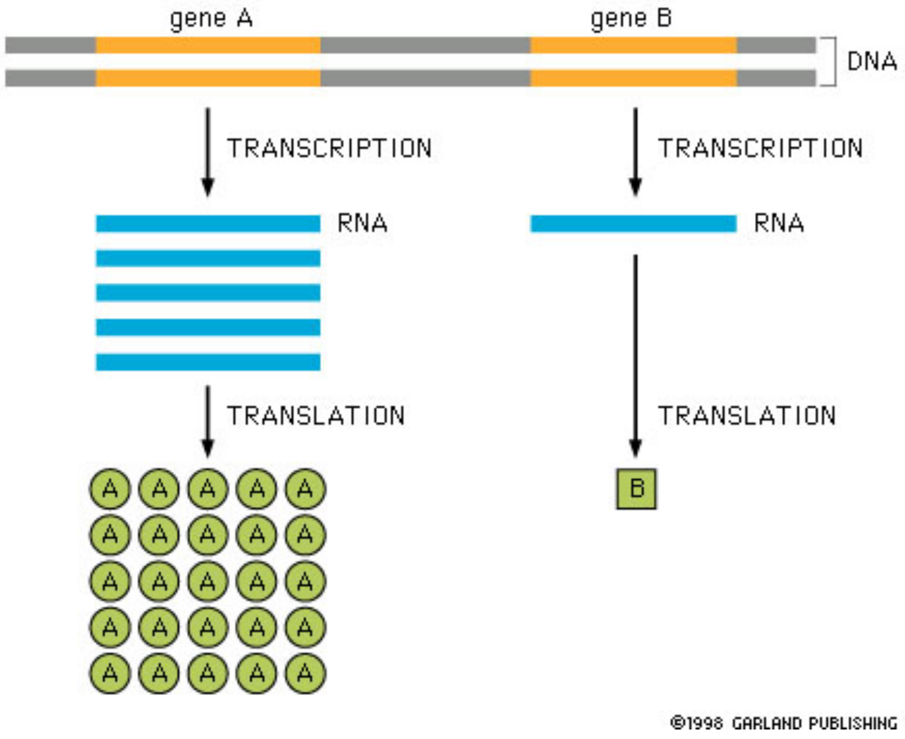
BUT…

More on it (much) later… (not today)

The (so-called) genetic dogma



©1998 GARLAND PUBLISHING



©1998 GARLAND PUBLISHING

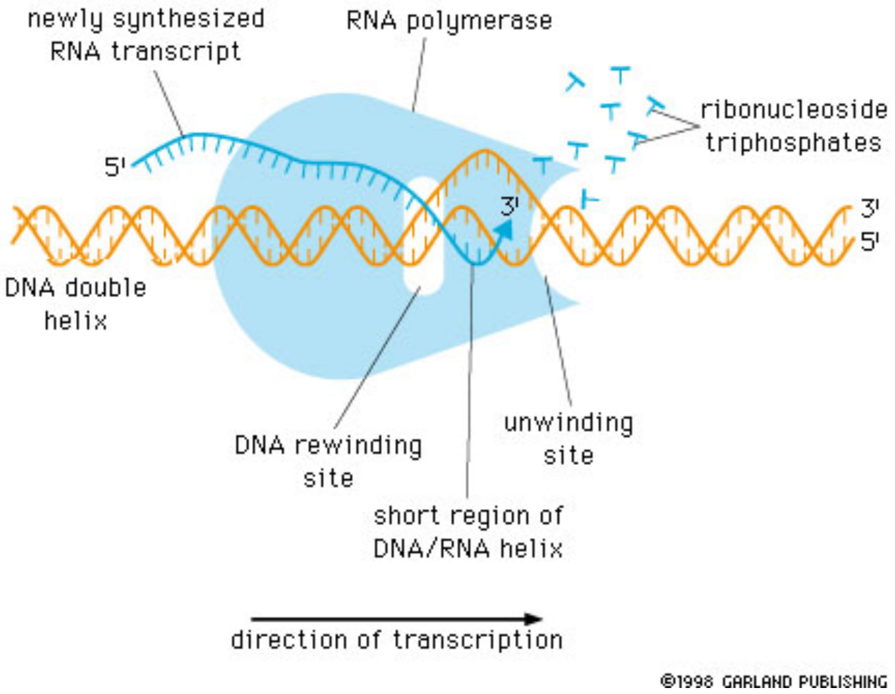
A gene is expressed in two steps:

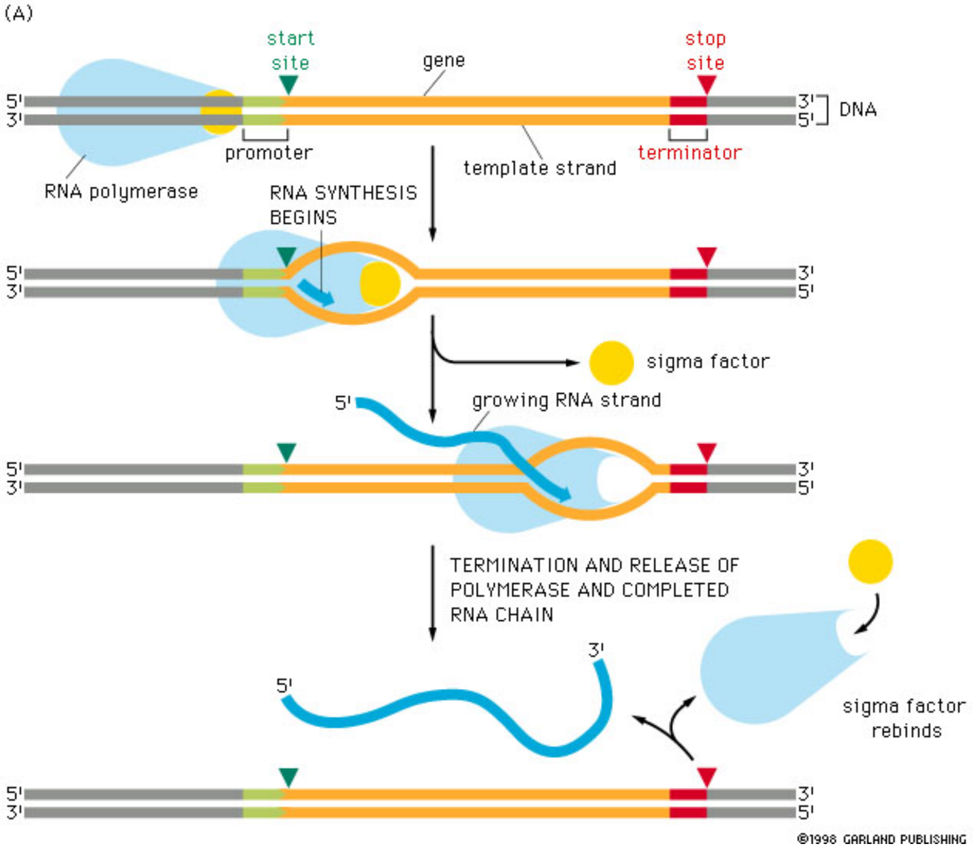
Transcription: RNA synthesis

Translation: Protein synthesis

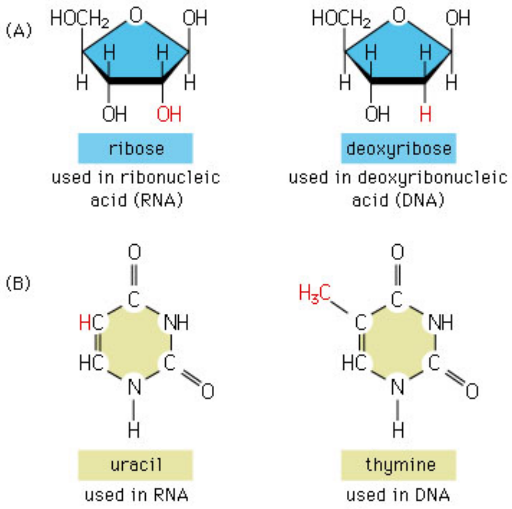
Transcription by RNA polymerase

RNA polymerase = enzyme = protein / Sigma factor = protein

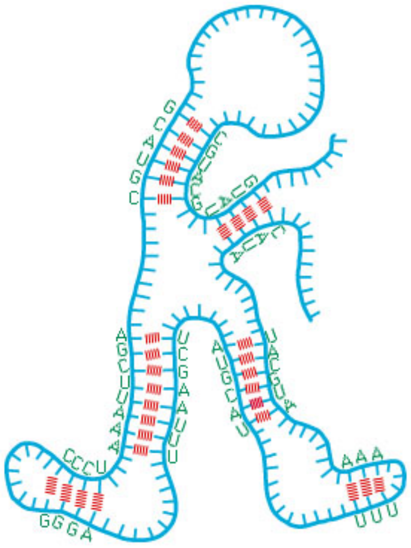

newly synthesized RNA transcript
RNA polymerase
ribonucleoside triphosphates
5'
3'
3'
5'
DNA double helix
DNA rewinding site
unwinding site
short region of DNA/RNA helix
direction of transcription
©1998 GARLAND PUBLISHING



RNA versus DNA



©1998 GARLAND PUBLISHING



©1998 GARLAND PUBLISHING

mRNAs	codes for proteins
rRNAs	forms part of the structure of the ribosome and participates in protein synthesis
tRNAs	used in protein synthesis as an adaptor between mRNA and amino acids
Small RNAs	used in pre-mRNA splicing, transport of proteins to endoplasmic reticulum, and other cellular processes

Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

RNA versus DNA

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT actually, there are other special types of "books" besides the genes

Almost every cell in an organism contains the same libraries and the same sets of books

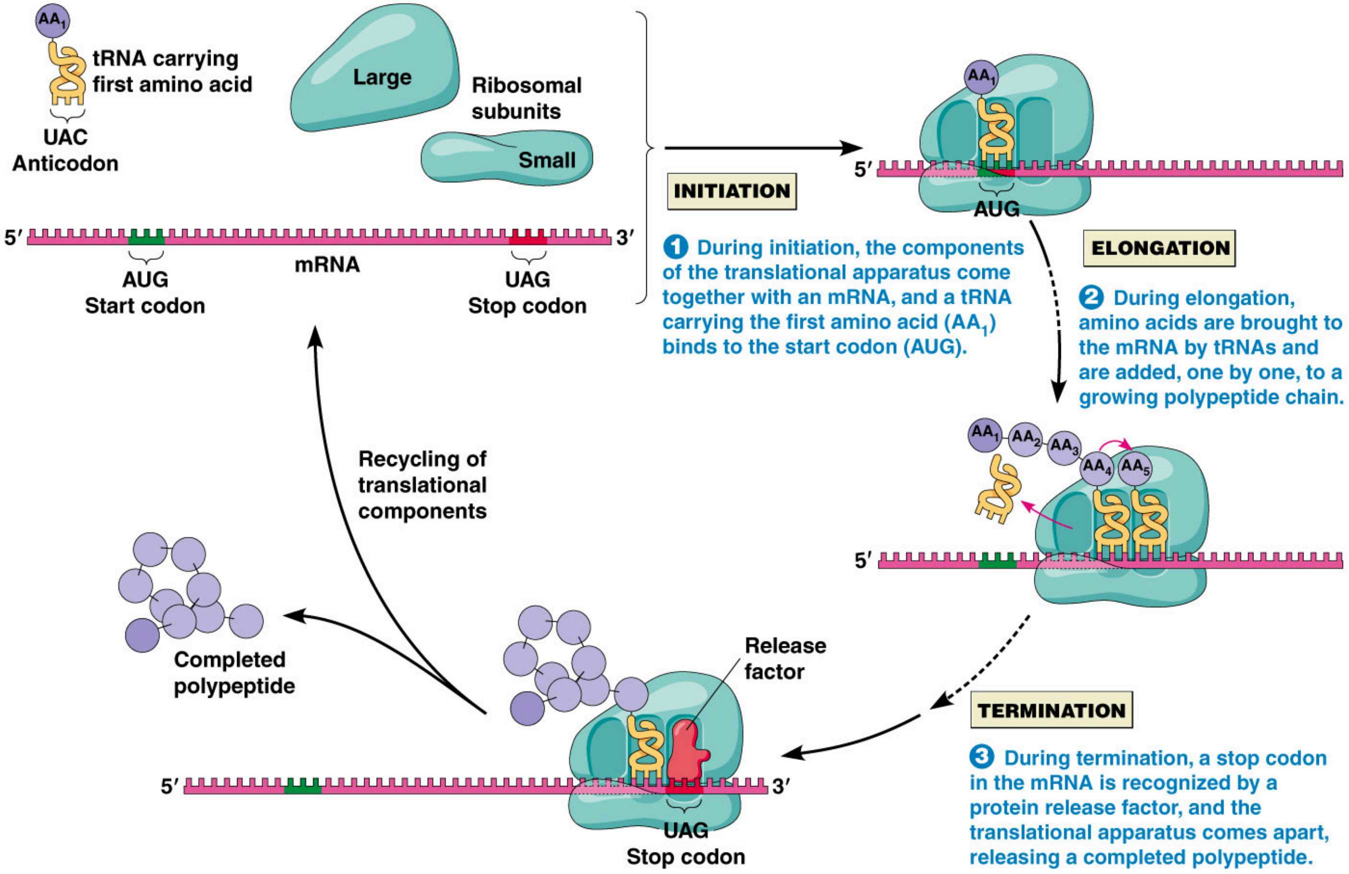
BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

Translation

Ribosome = complex proteins+RNAs



1 During initiation, the components of the translational apparatus come together with an mRNA, and a tRNA carrying the first amino acid (AA_1) binds to the start codon (AUG).

2 During elongation, amino acids are brought to the mRNA by tRNAs and are added, one by one, to a growing polypeptide chain.

3 During termination, a stop codon in the mRNA is recognized by a protein release factor, and the translational apparatus comes apart, releasing a completed polypeptide.

Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

Interactions everywhere

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

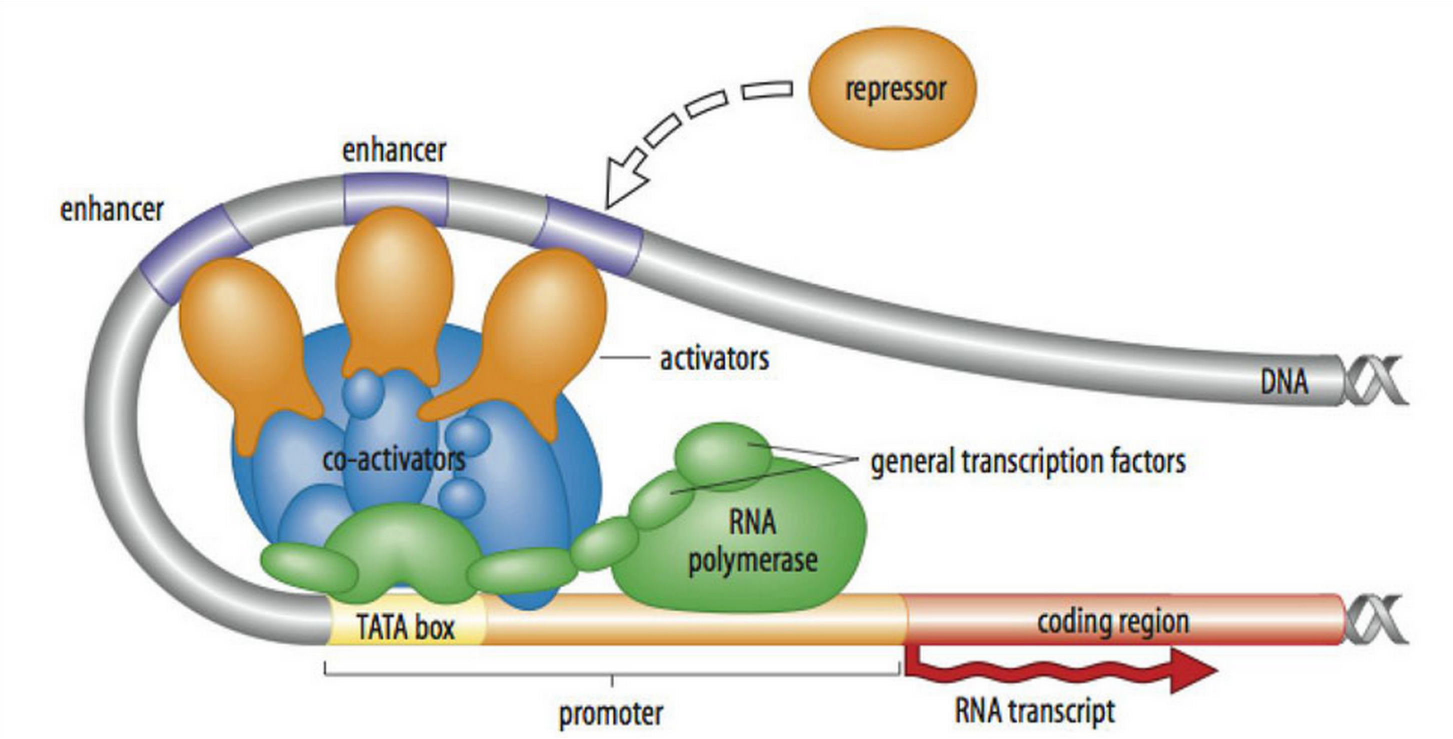
Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

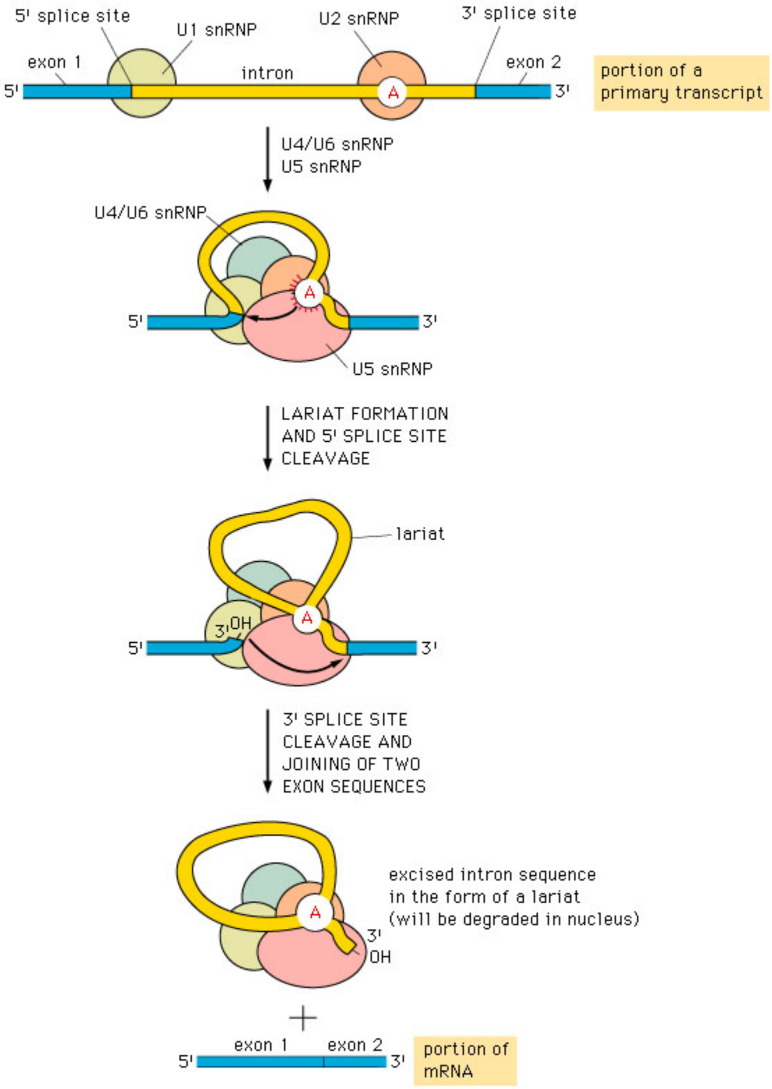
Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT most functions require INTERACTION among different books

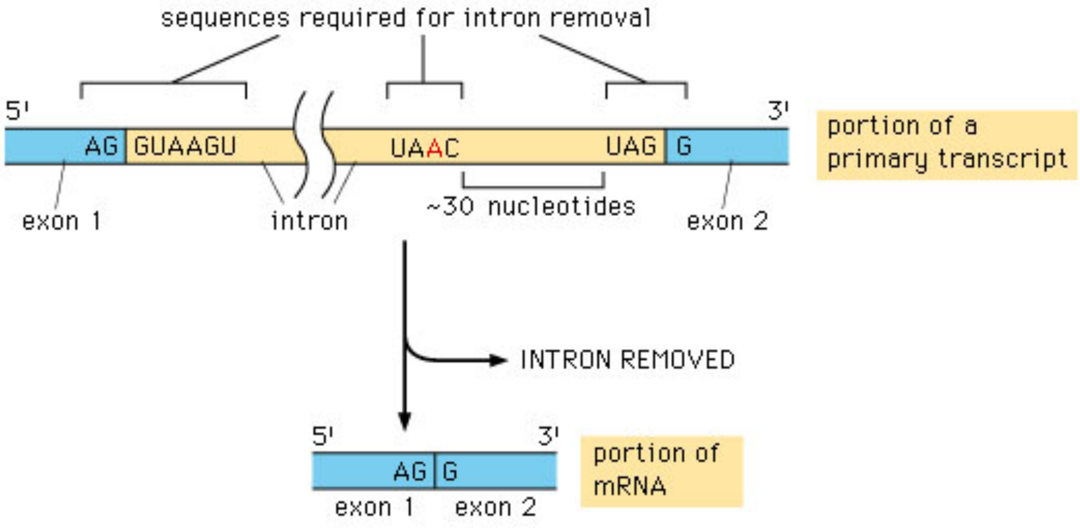
Indeed:
DNA, RNA, and proteins INTERACT among / between them through (sometimes highly specific) binding sites



Eukaryotic genes contain exons and introns

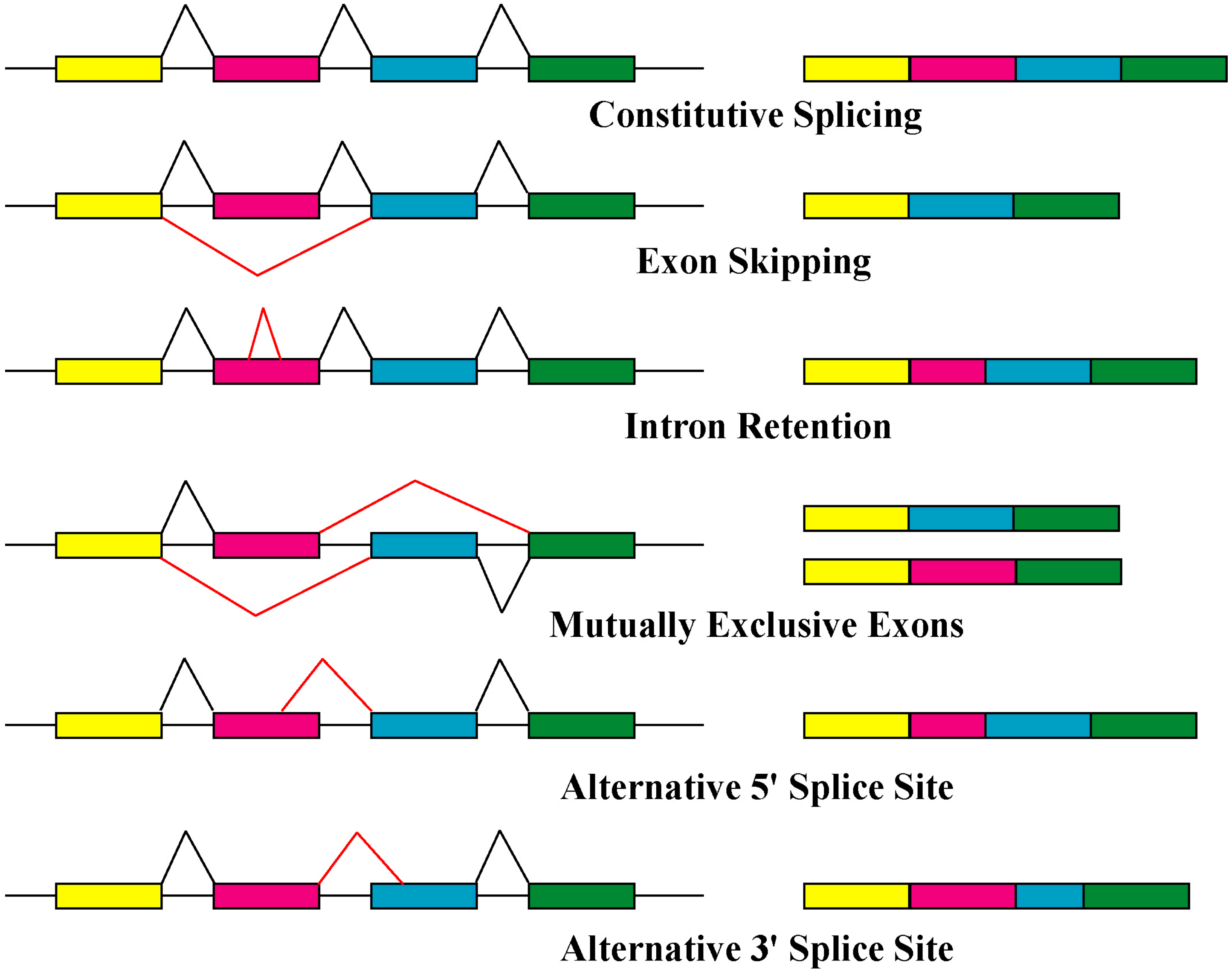

5' splice site
U1 snRNP
U2 snRNP
3' splice site
exon 1
intron
exon 2
5'
A
3'
portion of a primary transcript
U4/U6 snRNP
U5 snRNP
U4/U6 snRNP
5'
A
3'
U5 snRNP
LARIAT FORMATION AND 5' SPLICE SITE CLEAVAGE
lariat
3' OH
5'
A
3'
3' SPLICE SITE CLEAVAGE AND JOINING OF TWO EXON SEQUENCES
excised intron sequence in the form of a lariat (will be degraded in nucleus)
A
3' OH
+
exon 1
exon 2
5'
3'
portion of mRNA

©1998 GARLAND PUBLISHING



©1998 GARLAND PUBLISHING

Splicing and alternative splicing



Constitutive Splicing

Exon Skipping

Intron Retention

Mutually Exclusive Exons

Alternative 5' Splice Site

Alternative 3' Splice Site

Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

Biodiversity of proteins driven by alternative splicing

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

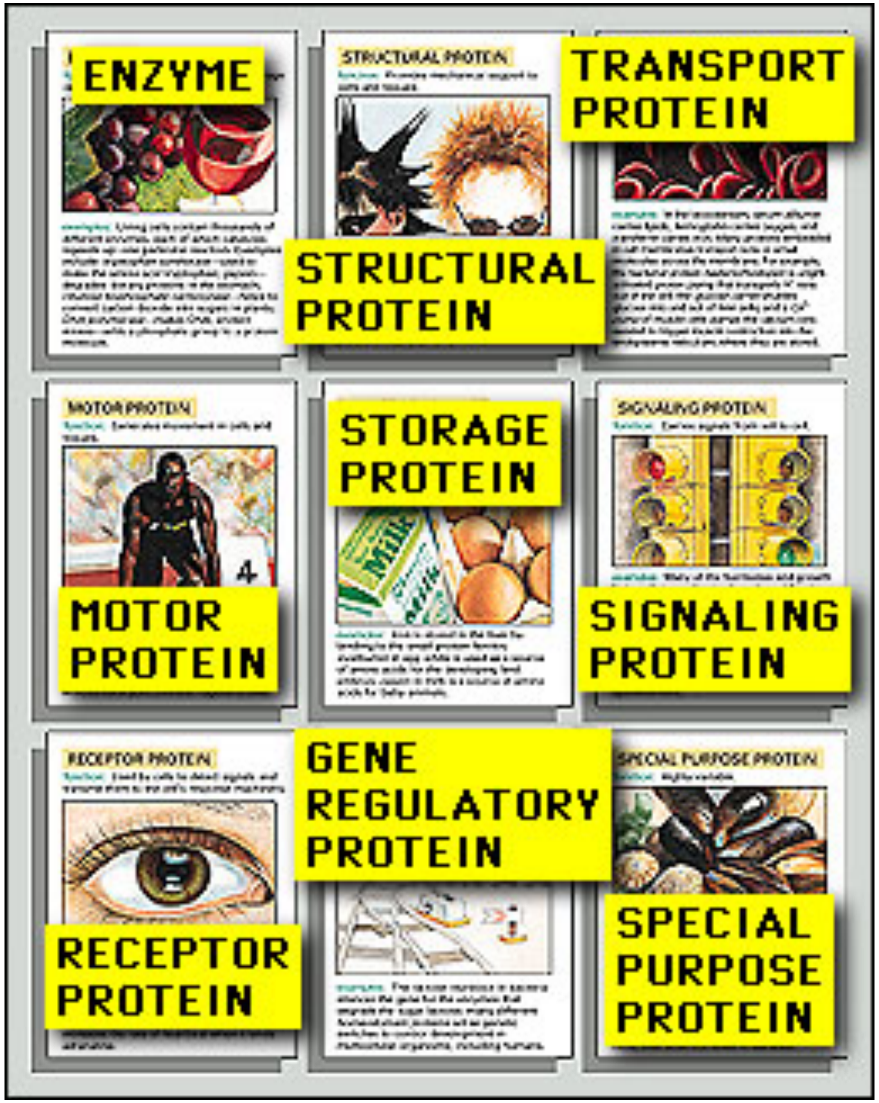
Almost every cell in an organism contains the same libraries and the same sets of books

BUT even inside a same organism, the "final" books may vary greatly

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

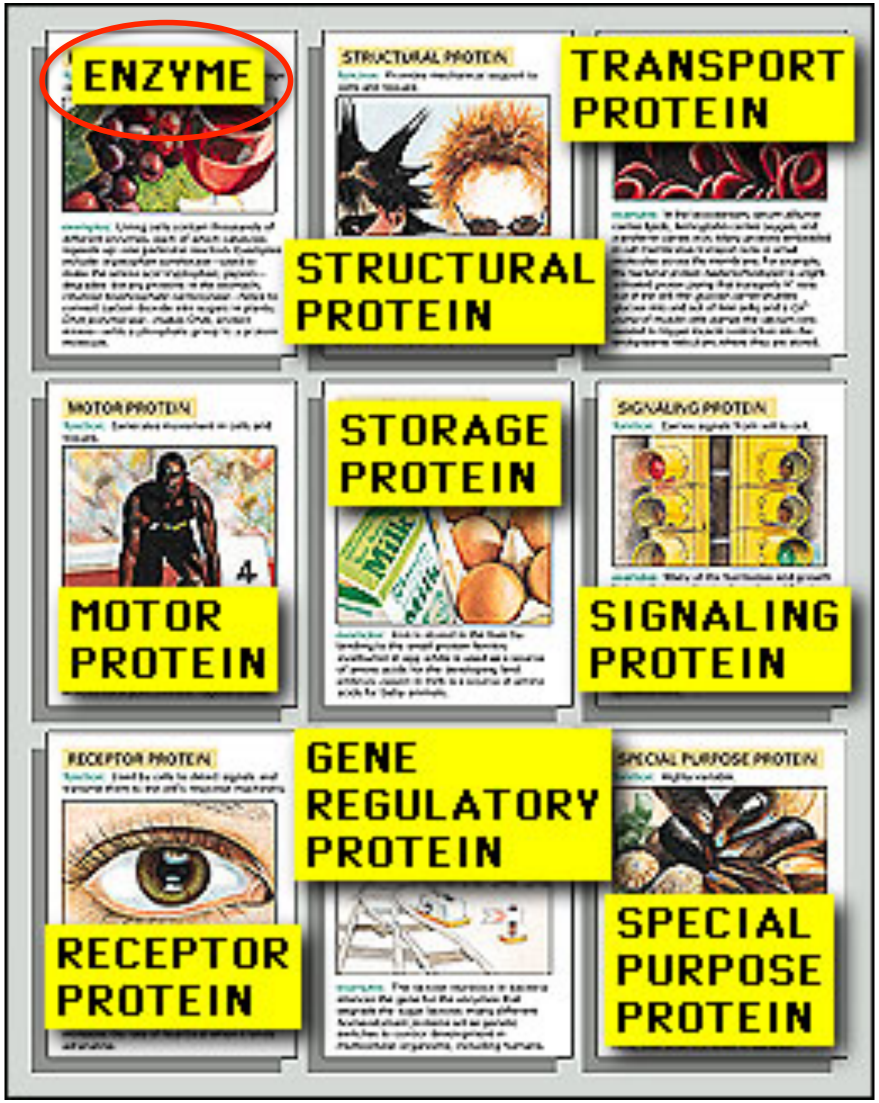
Various functions of proteins



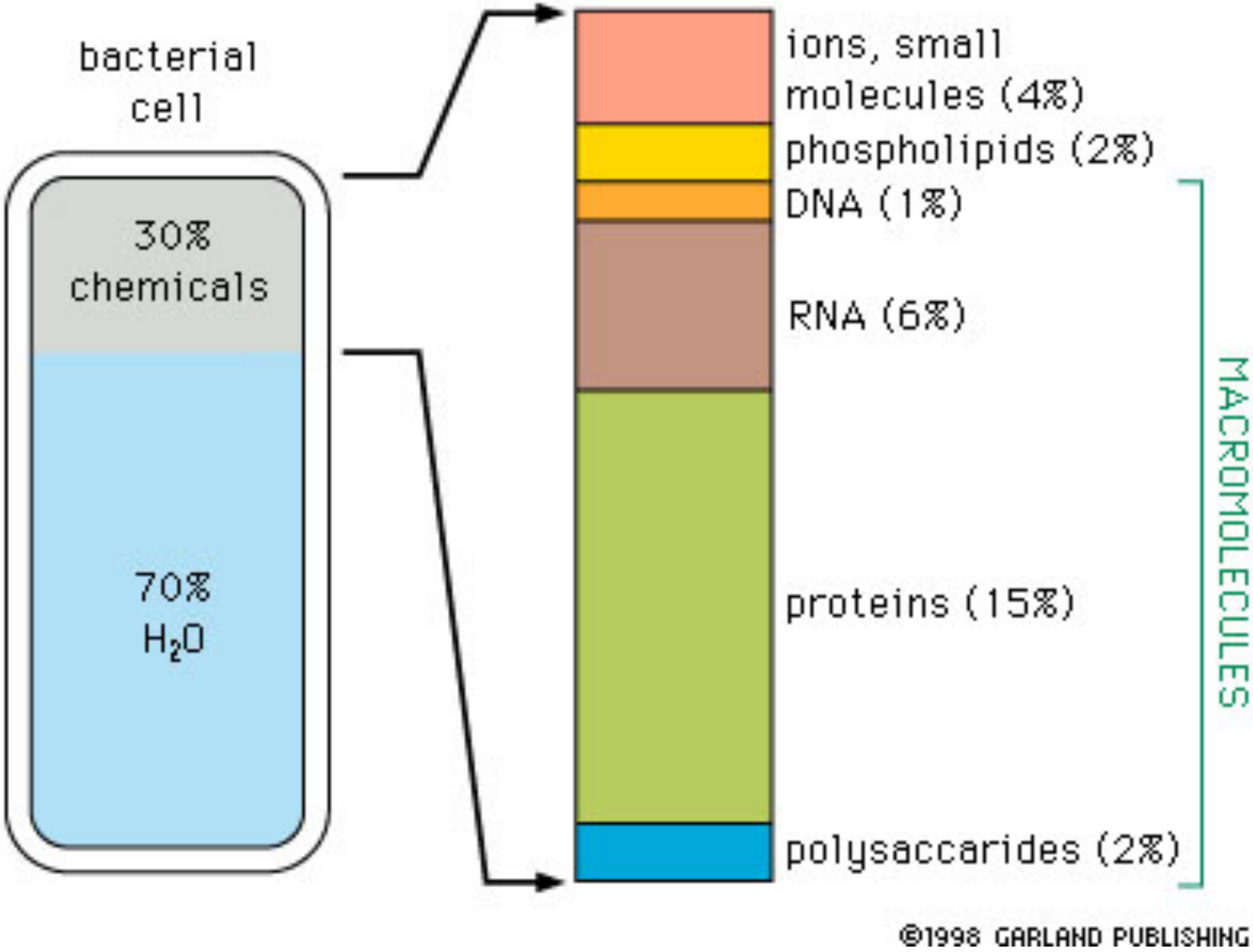
Various functions of proteins

Crucial in metabolism

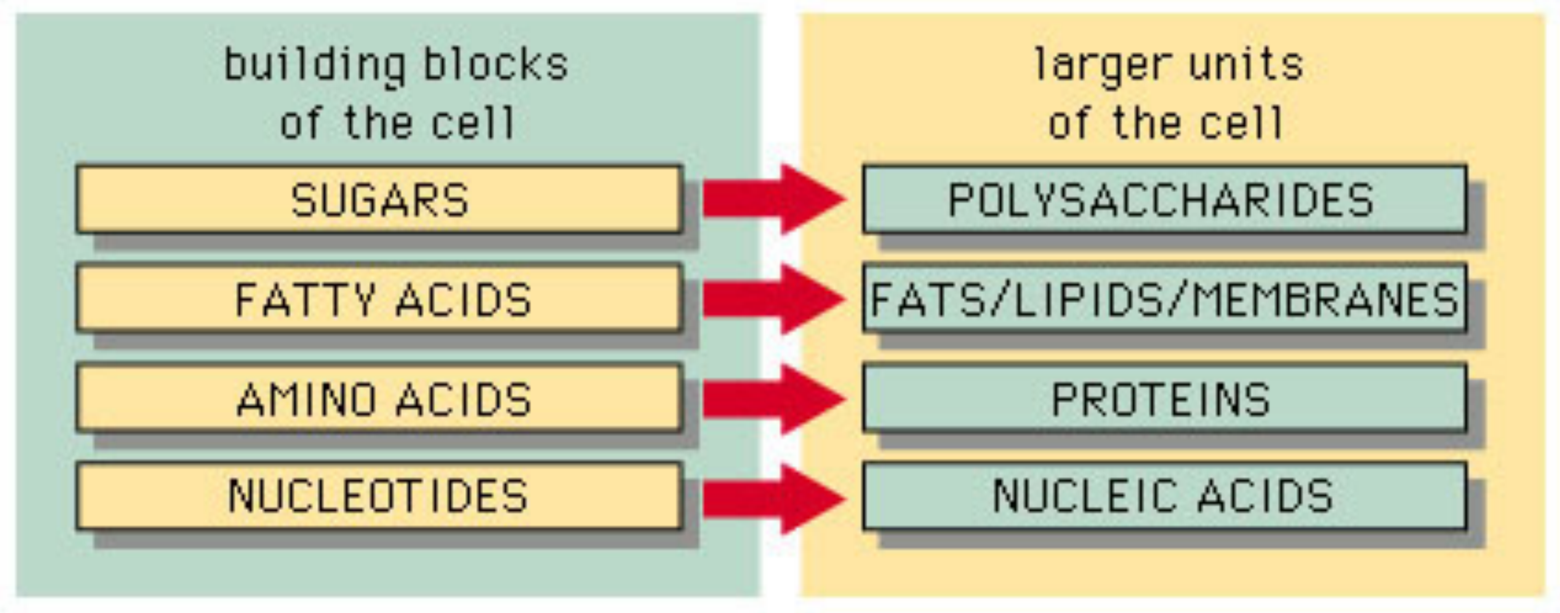
Metabolism = set of life-sustaining chemical transformations within the cells of organisms



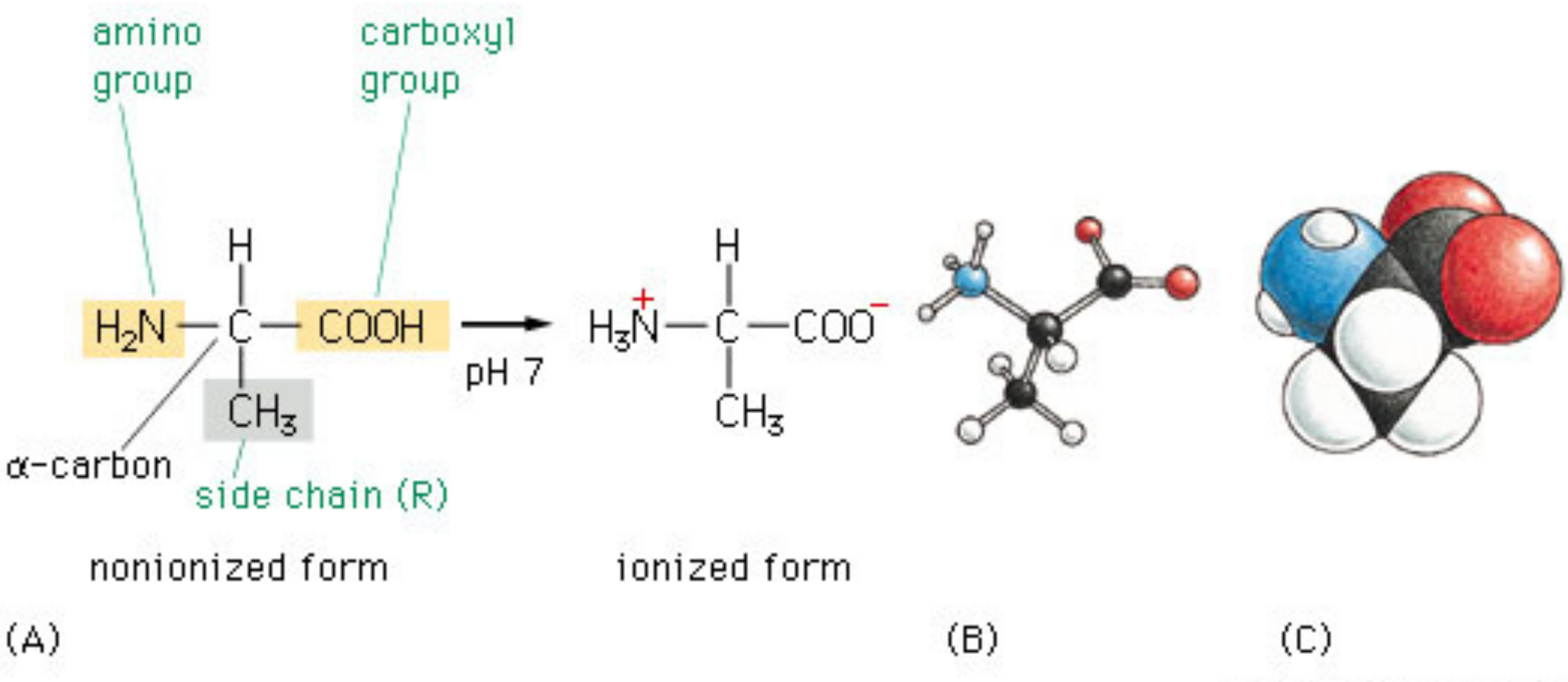
Remember what was said before: macromolecules such as DNA, RNA, proteins and etc are abundant in cells



However small molecules also have an important role: Four main families of small organic molecules in cells

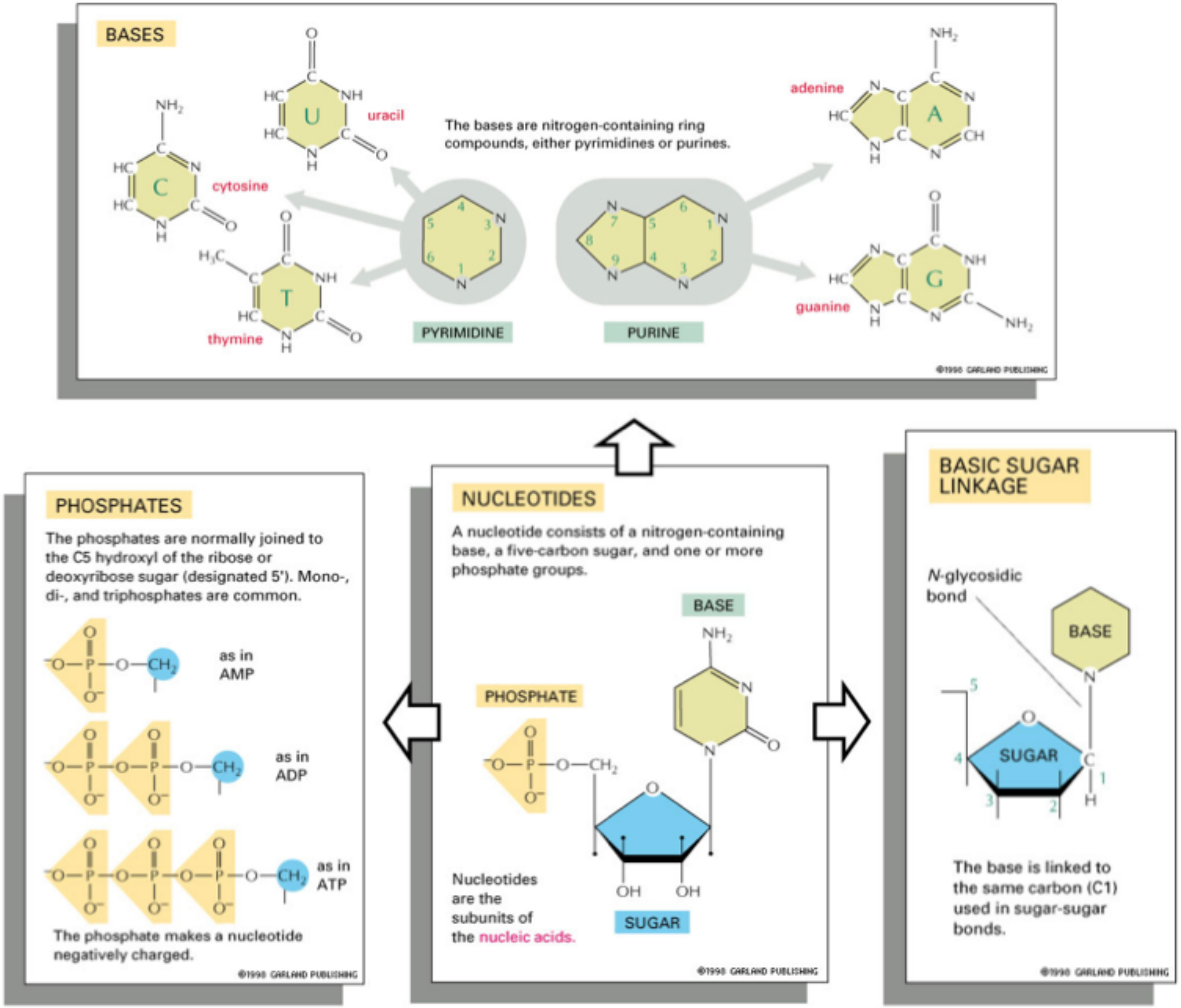


Looking at two small molecules more in particular: Amino acids of which proteins are made



©1998 GARLAND PUBLISHING

Looking at two small molecules more in particular: Nucleotides of which DNA is made



BASES

The bases are nitrogen-containing ring compounds, either pyrimidines or purines.

PHOSPHATES

The phosphates are normally joined to the C5 hydroxyl of the ribose or deoxyribose sugar (designated 5'). Mono-, di-, and triphosphates are common.

The phosphate makes a nucleotide negatively charged.

NUCLEOTIDES

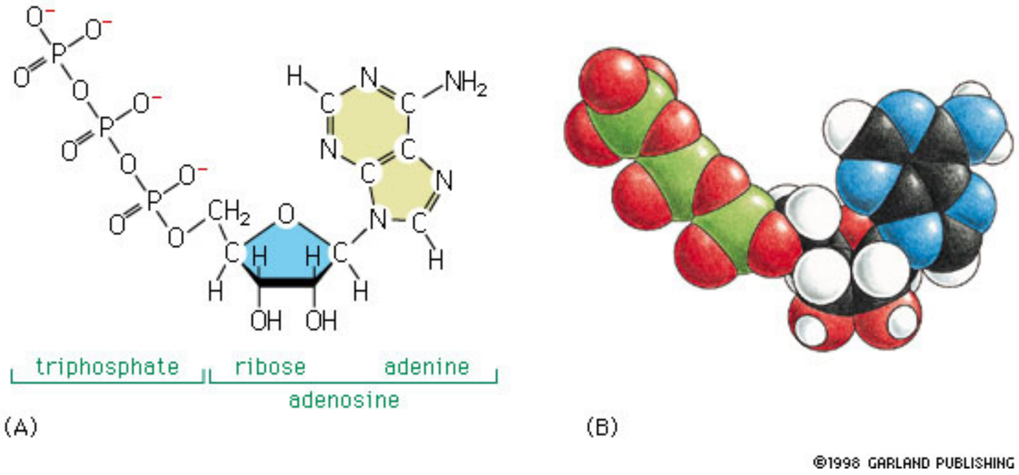
A nucleotide consists of a nitrogen-containing base, a five-carbon sugar, and one or more phosphate groups.

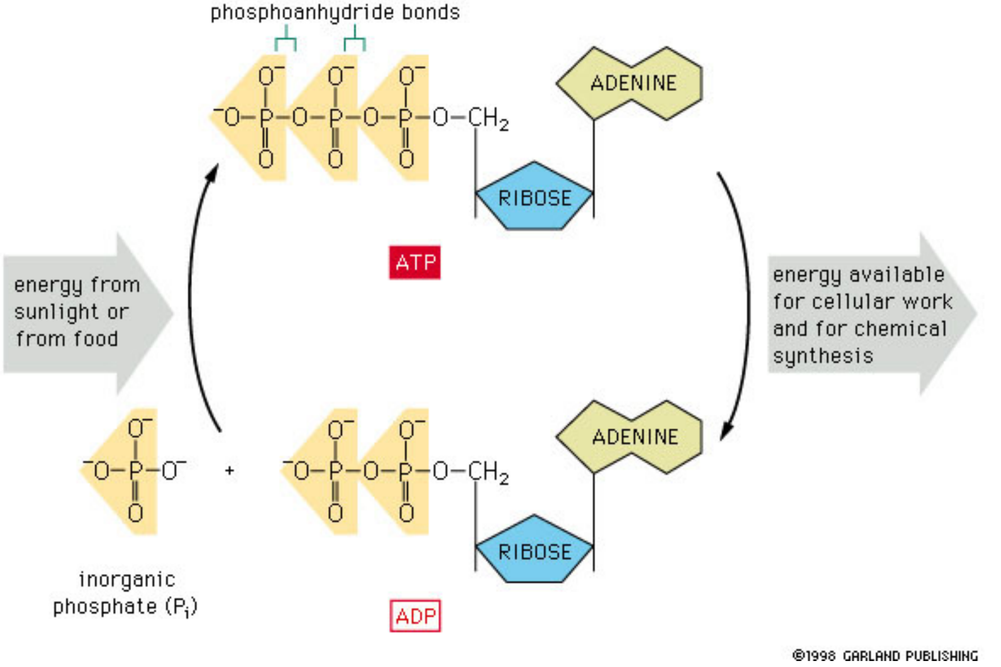
Nucleotides are the subunits of the nucleic acids.

BASIC SUGAR LINKAGE

The base is linked to the same carbon (C1) used in sugar-sugar bonds.

But also many more small molecules among which, *e.g.*, one of special interest: ATP: the energy carrier in the cell





Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT life is also chemistry

Almost every cell in an organism contains the same libraries and the same sets of books

BUT life is also chemistry

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT life is also chemistry

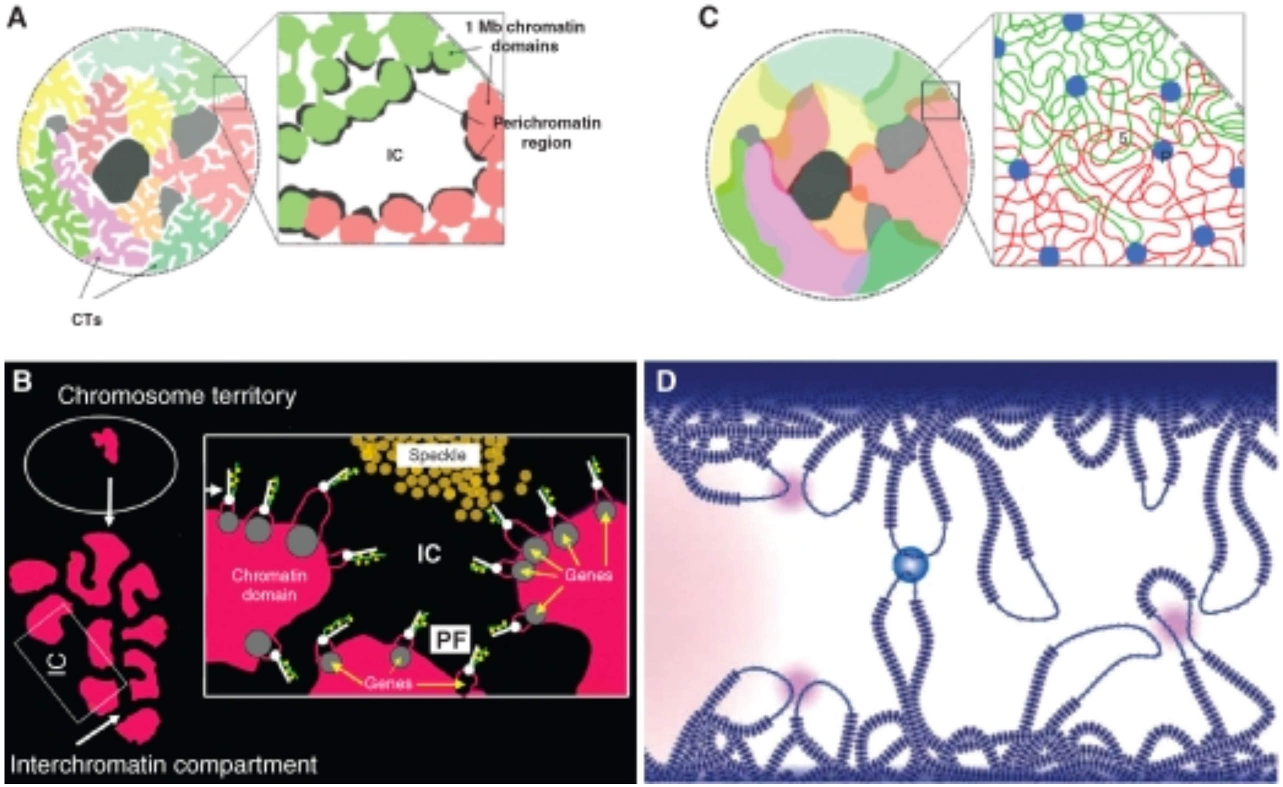
To conclude this (brief) introduction
First, one more important information

DNA in a living cell is in a highly compacted and structured state
Transcription is dependent on such structural state!

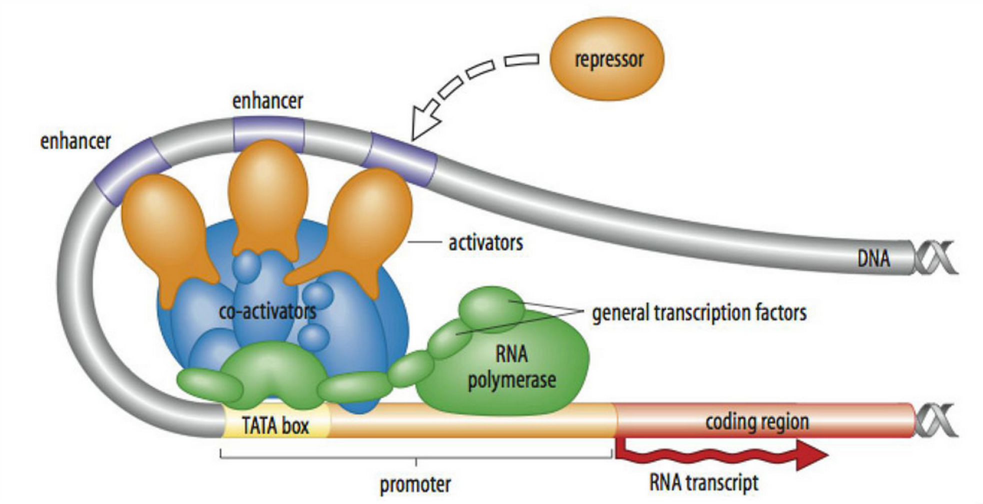
To conclude this (brief) introduction
First, one more important information

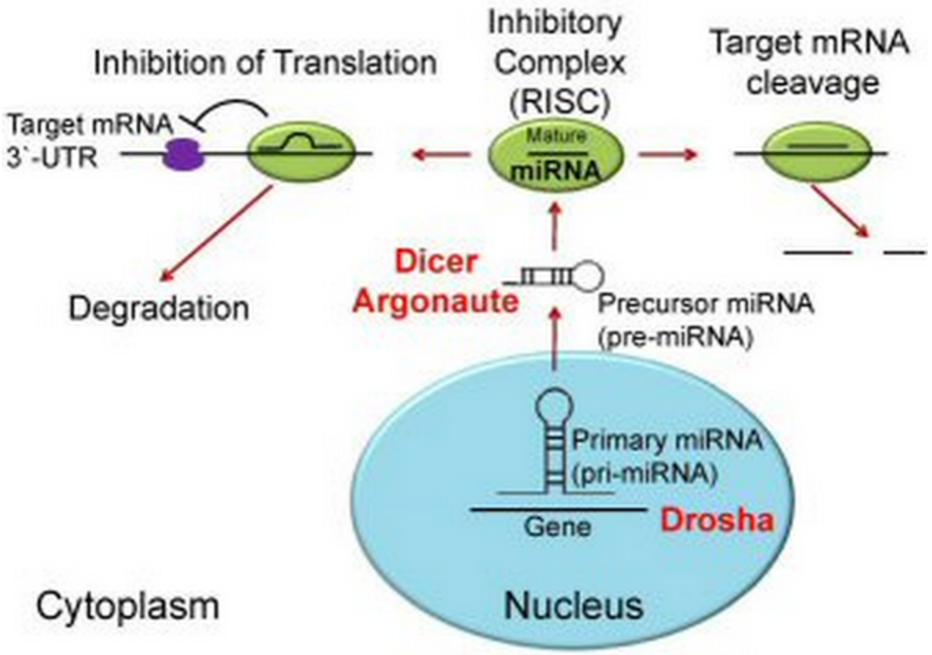
DNA in a living cell is in a highly compacted and structured state
Transcription is dependent on such structural state!

Chromosomes are not like spaghetti inside the nucleus!



And finally (perhaps the most important):
Transcription and translation are REGULATED




Inhibitory Complex (RISC)
Inhibition of Translation
Target mRNA cleavage
Target mRNA 3`-UTR
Mature miRNA
Degradation
Dicer
Argonaute
Precursor miRNA (pre-miRNA)
Primary miRNA (pri-miRNA)
Gene
Drosha
Cytoplasm
Nucleus

Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

BUT (again)… (see later…)

Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

Remember when I said

Nucleus / cytoplasm = library
Chromosome(s) = bookshelves
Genes = books

BUT… (see later)

Almost every cell in an organism contains the same libraries and the same sets of books

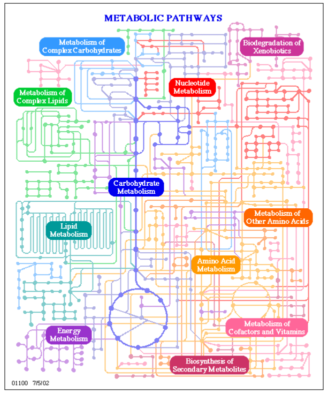
BUT actually, every cell contains the same set of books (genes) indeed, but expressed in highly different ways!

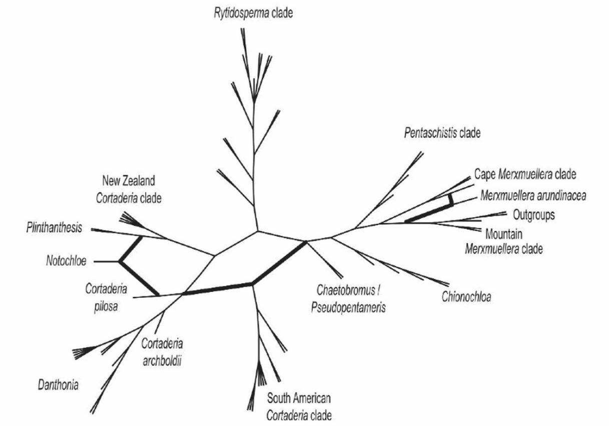
Books represent all the information (DNA) that each cell in the body needs so it can grow and carry out its various functions

BUT (once more time)… (see later…)

The key abstract idea to retain for now however is:
Interactions! Interactions everywhere!
And so networks / graphs, as models or tools

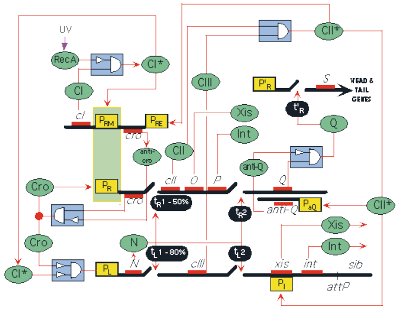
Biochemical networks …but also

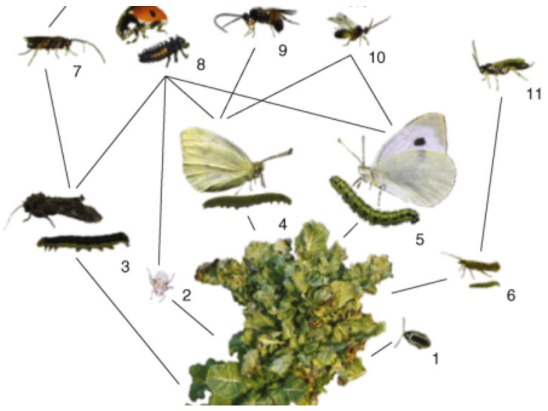

METABOLIC PATHWAYS
Metabolism of Complex Carbohydrates
Biodegradation of Xenobiotics
Nucleotide Metabolism
Metabolism of Complex Lipids
Carbohydrate Metabolism
Lipid Metabolism
Metabolism of Other Amino Acids
Amino Acid Metabolism
Energy Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Secondary Metabolites
01100 7/5/02

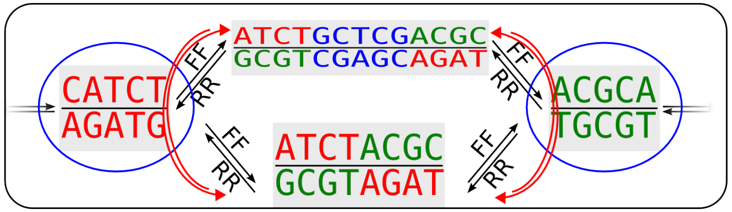


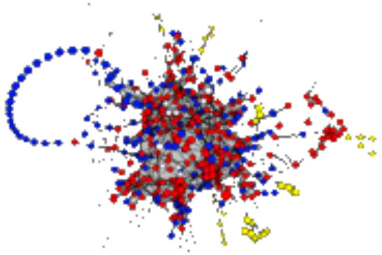
Evolutionary network

"Symbiotic" network









Ecological network

Besides graphs as ways of inferring information related to interactions

A few references for those curious to know more

Molecular biology of the cell, Bruce Alberts & Alexander Johnson

What is life? Erwin Schrödinger
See also: http://whatislife.stanford.edu/LoCo_files/What-is-Life.pdf

The chemistry of life, Steve Rose

In French: La biologie buissonière, Jacques Ninio

And many, many more
If interested in having more references, contact us!