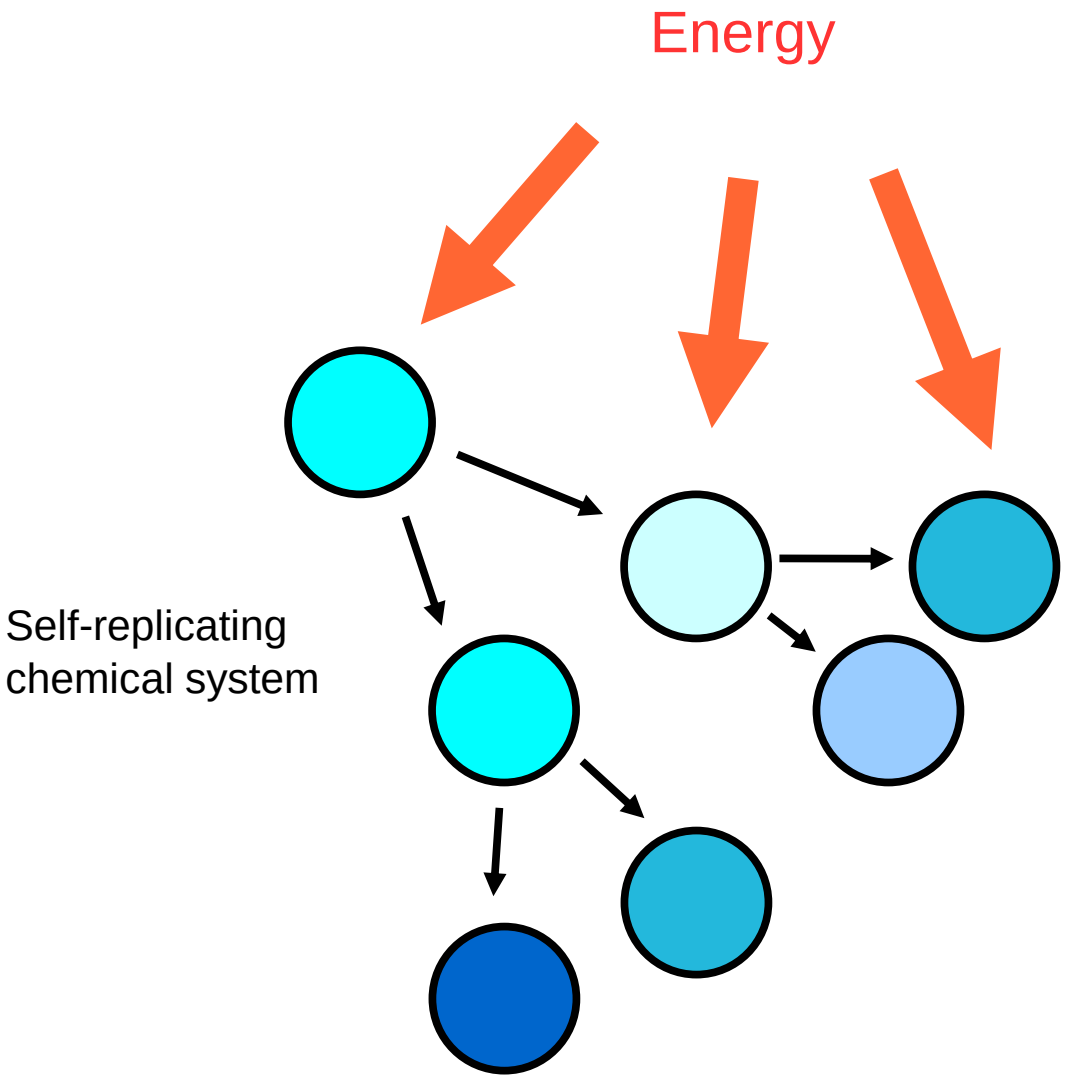
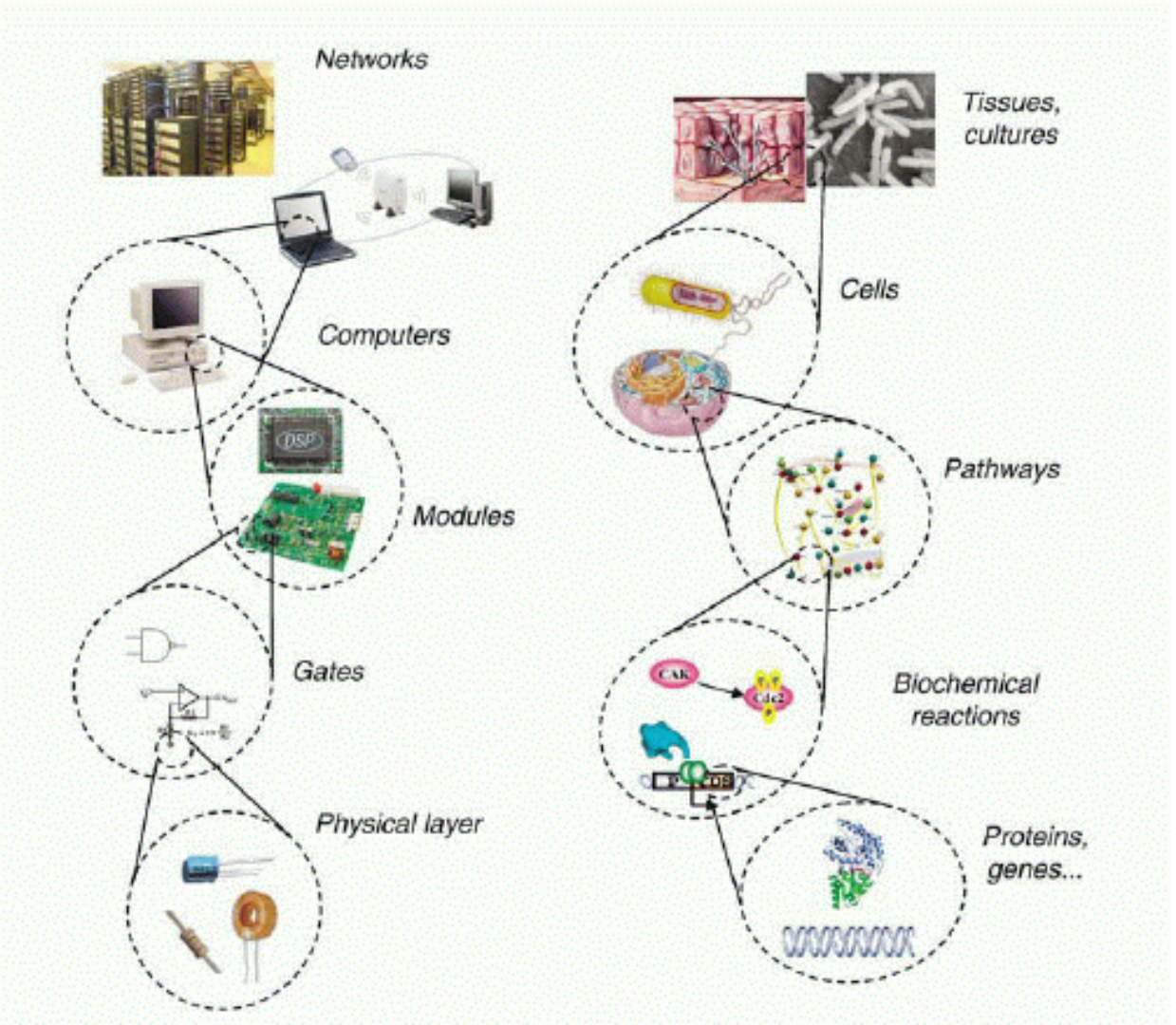
Review of Molecular Biology

What is life?



Cells compute:



(Andrianantoandro, Basu, Karig, Weiss (2006))

Can we decipher the biological hardware and software?

Some Overarching Organizational Rules:

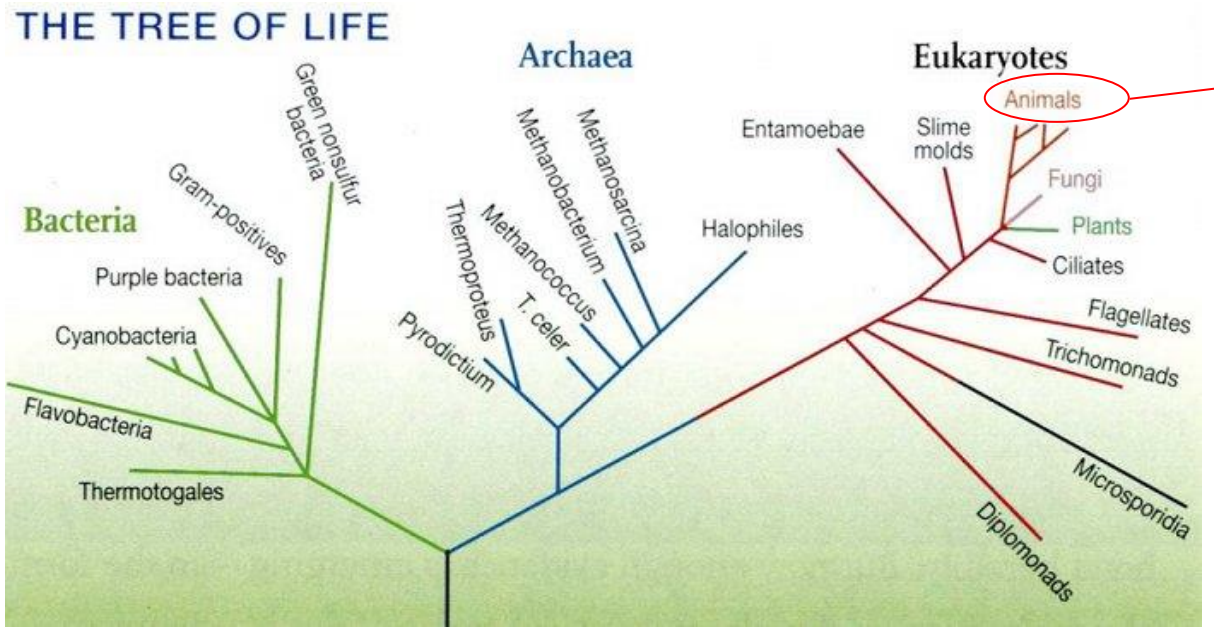
- Biological parts self-assemble
- Biological parts are actively transported
- Biological parts interact specifically

These interactions lead to complex spatial and temporal patterns that produce the variety of life

complex behaviour arising within complex interacting systems
= the language of physics

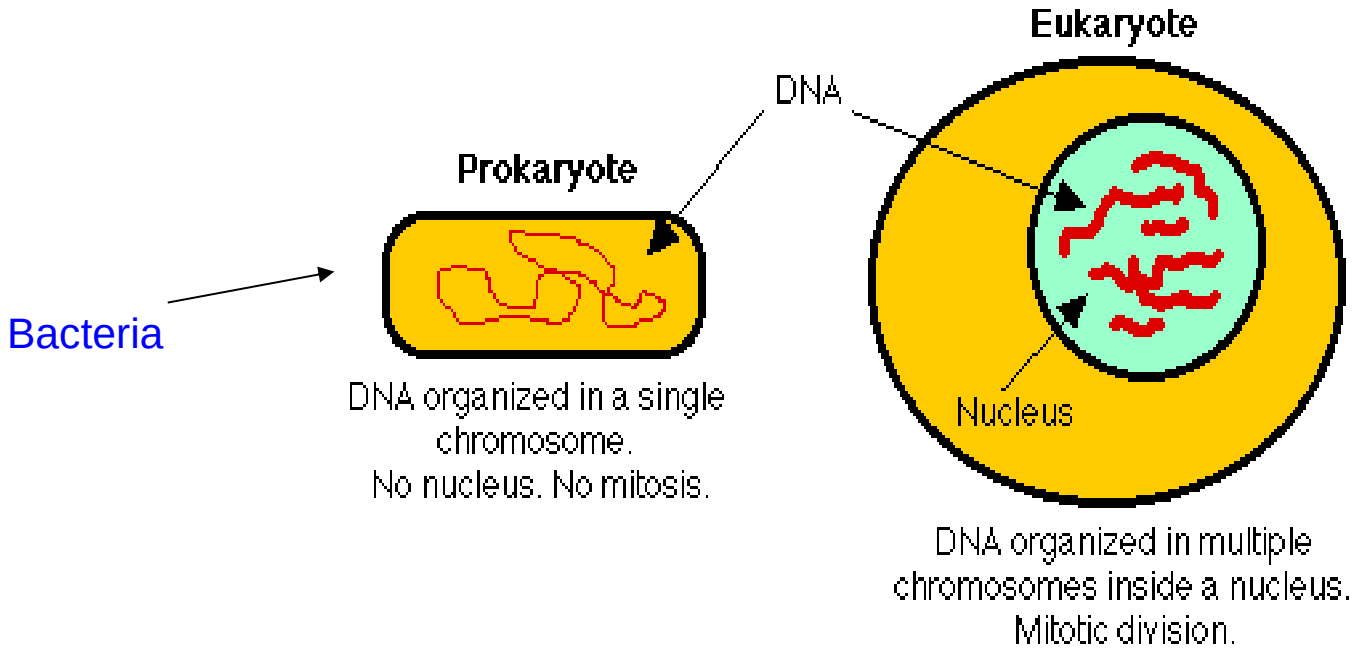
The nature and size of things in biology

Two Generic Cell Types:

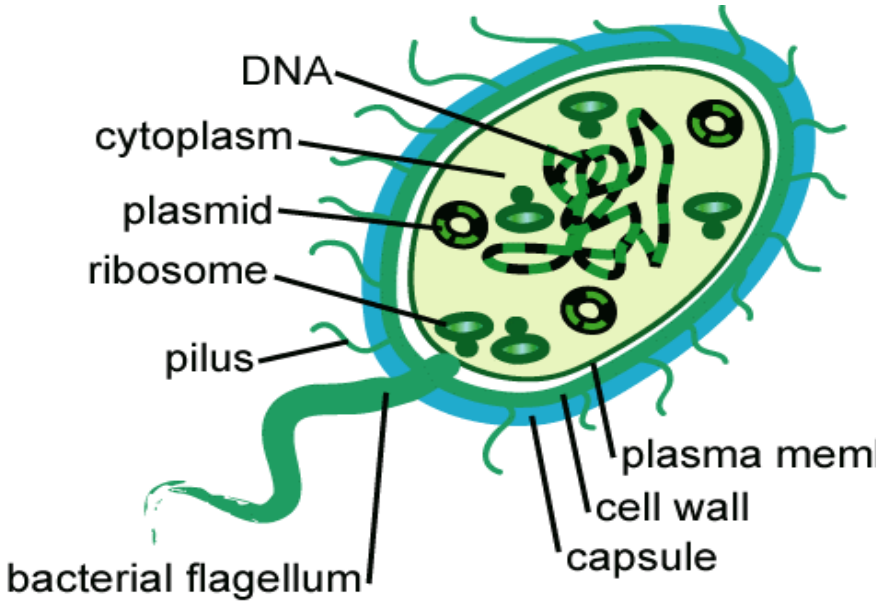


Us

Higher Organisms



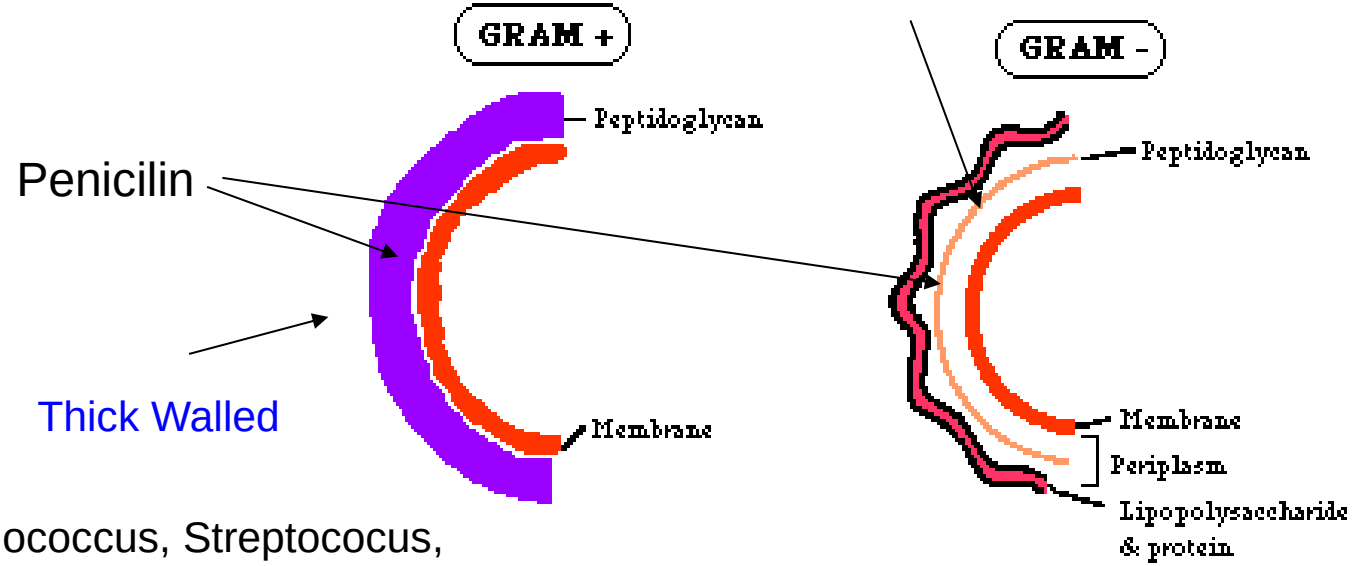
Prokaryotic Cells:



Bad pathogens

(the plague, salmonella, meningitis, E. coli, cholera)

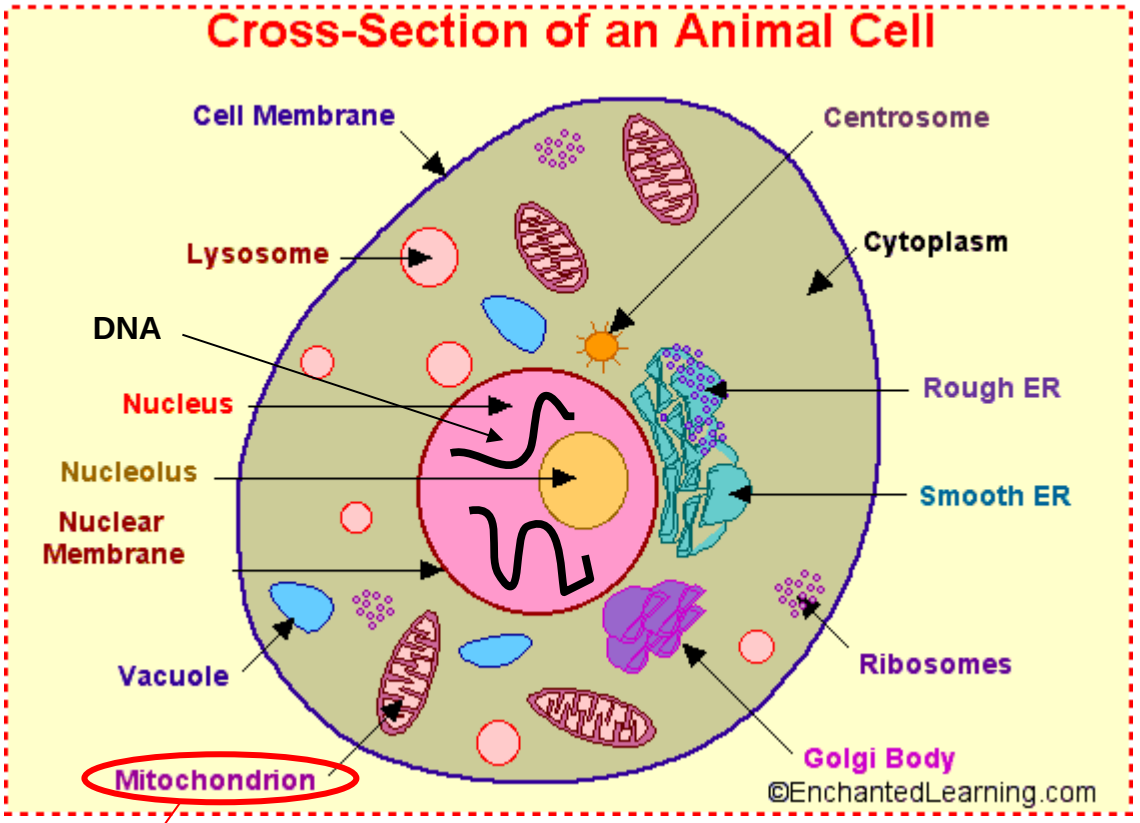
Thin Walled



Thick Walled

(Staphylococcus, Streptococcus, anthrax)

Eukaryotic Cells:



The mitochondria make energy – very important

E. Coli – our biological ruler

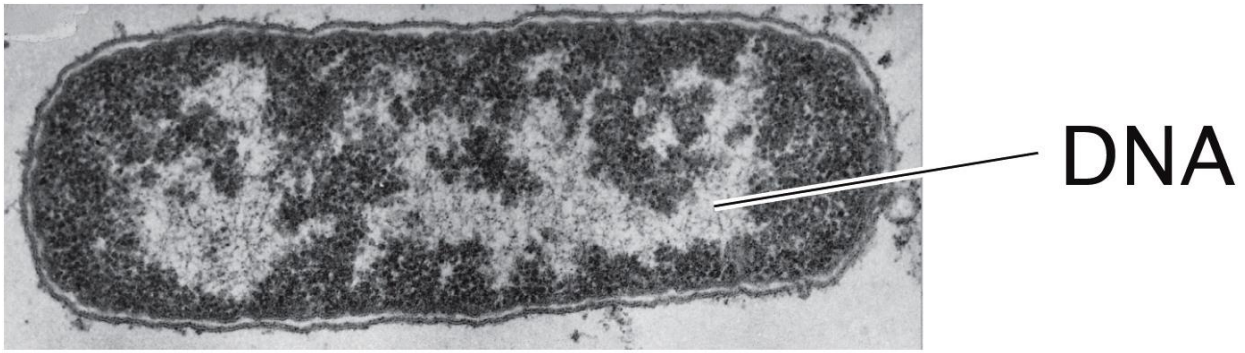


Figure 2.1b Physical Biology of the Cell, 2ed. (© Garland Science 2013)

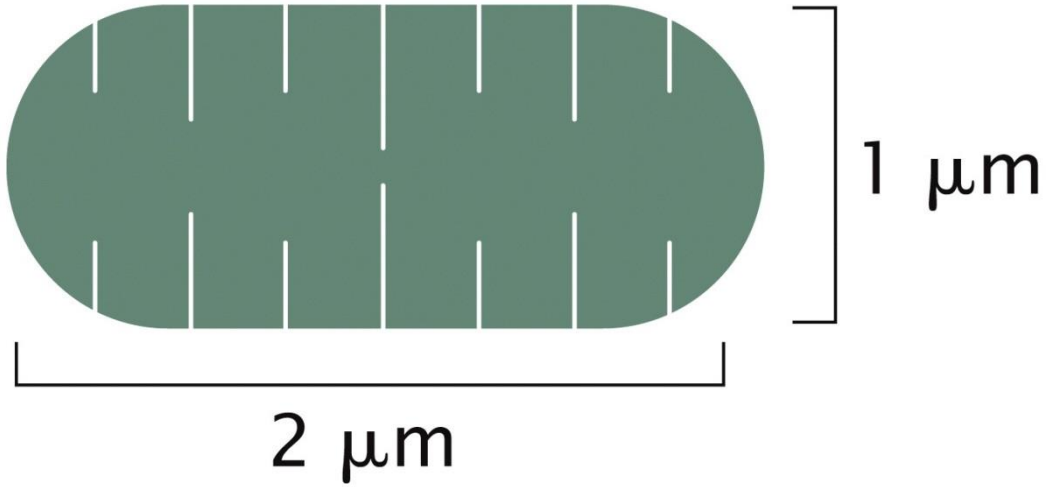


Figure 2.1c Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Huge variety in cells:

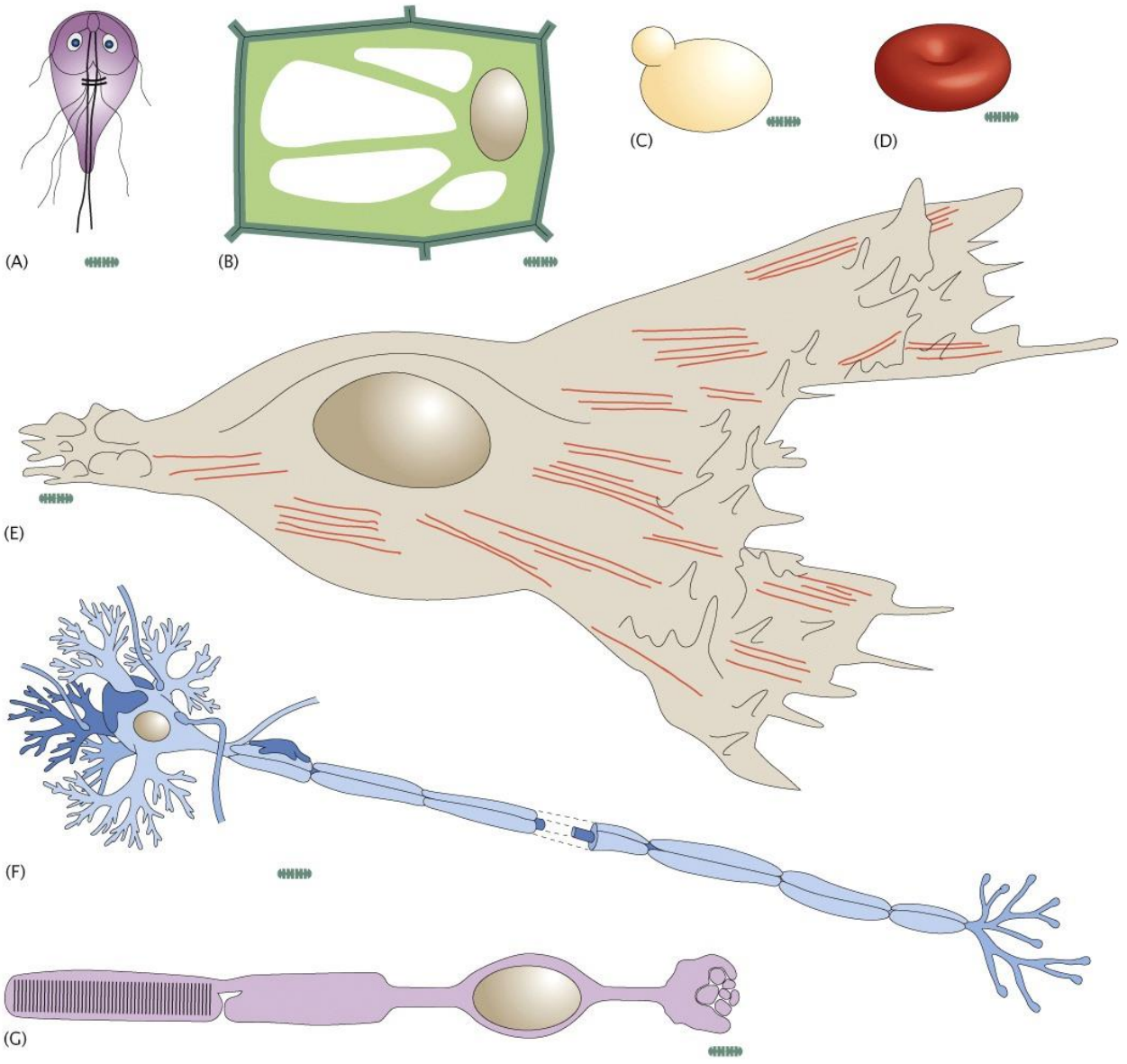


Figure 2.16 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Setting some length scales

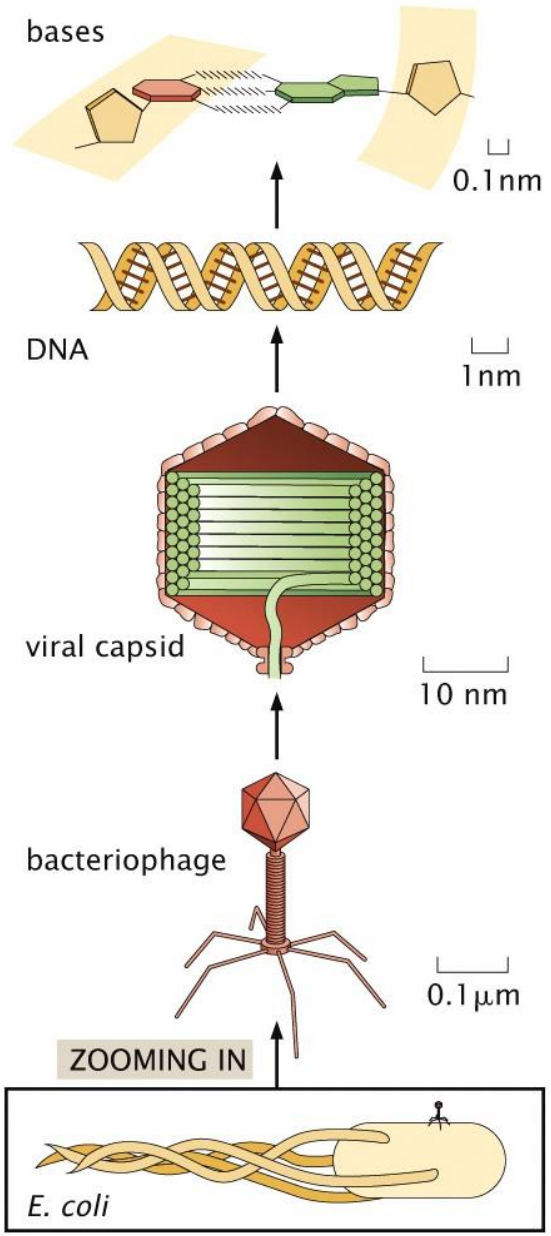


Figure 2.15 (part 1 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)

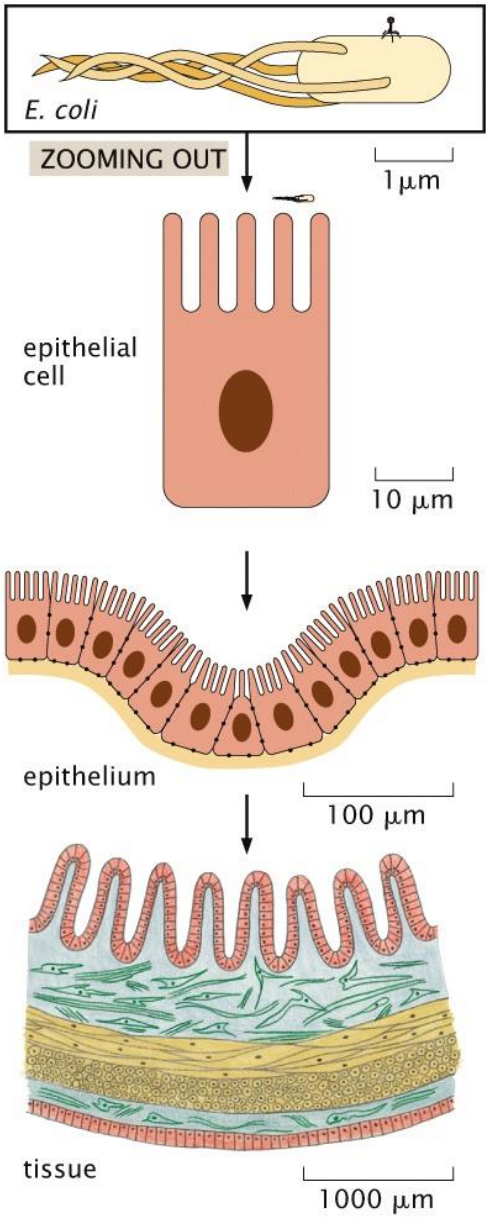


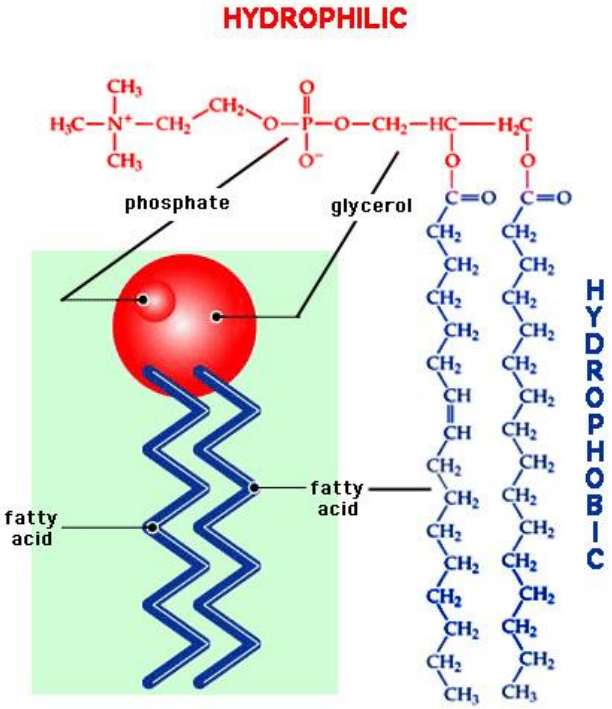
Figure 2.15 (part 2 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)

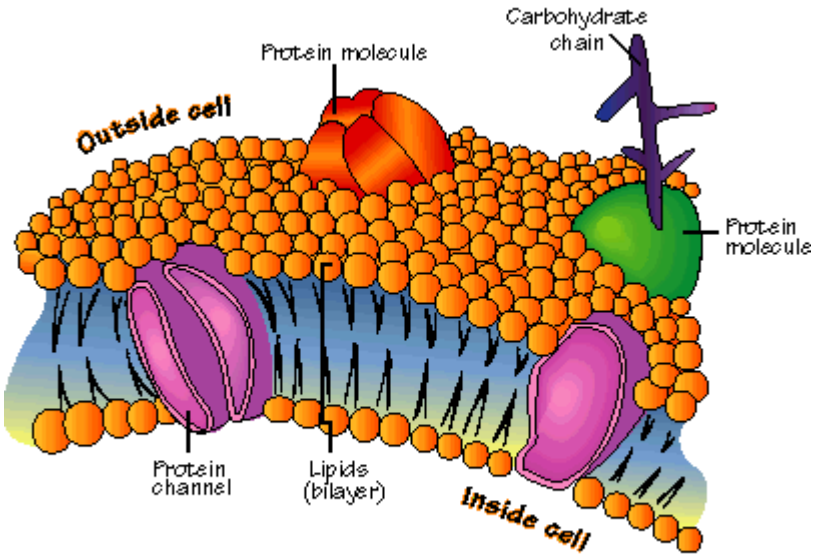
The stuff inside: small molecules

- water – we're 70% H20
- ions: H+, Na+, Ca2+, K+, etc. used to drive transport, pumps
- sugars – glucose, ribose, sucrose – substrates for many metabollic reactions
- nucleotides – form DNA & RNA
  - ATP (adenosine triphosphate – the big energy source of cells)
- amino acids – form proteins
- fatty acids – chain like molecules that are the basis of lipids that make the membrane

Lipids & Membranes: The packaging

- Lipids are molecules which form membrane bi-layers
- consist of two fatty acids joined by a ‘head’ molecule of glycerol
- head group likes water, tail does not like water
- Lipids spontaneously self-assemble into membranes





Lipids & Membranes: a physics view

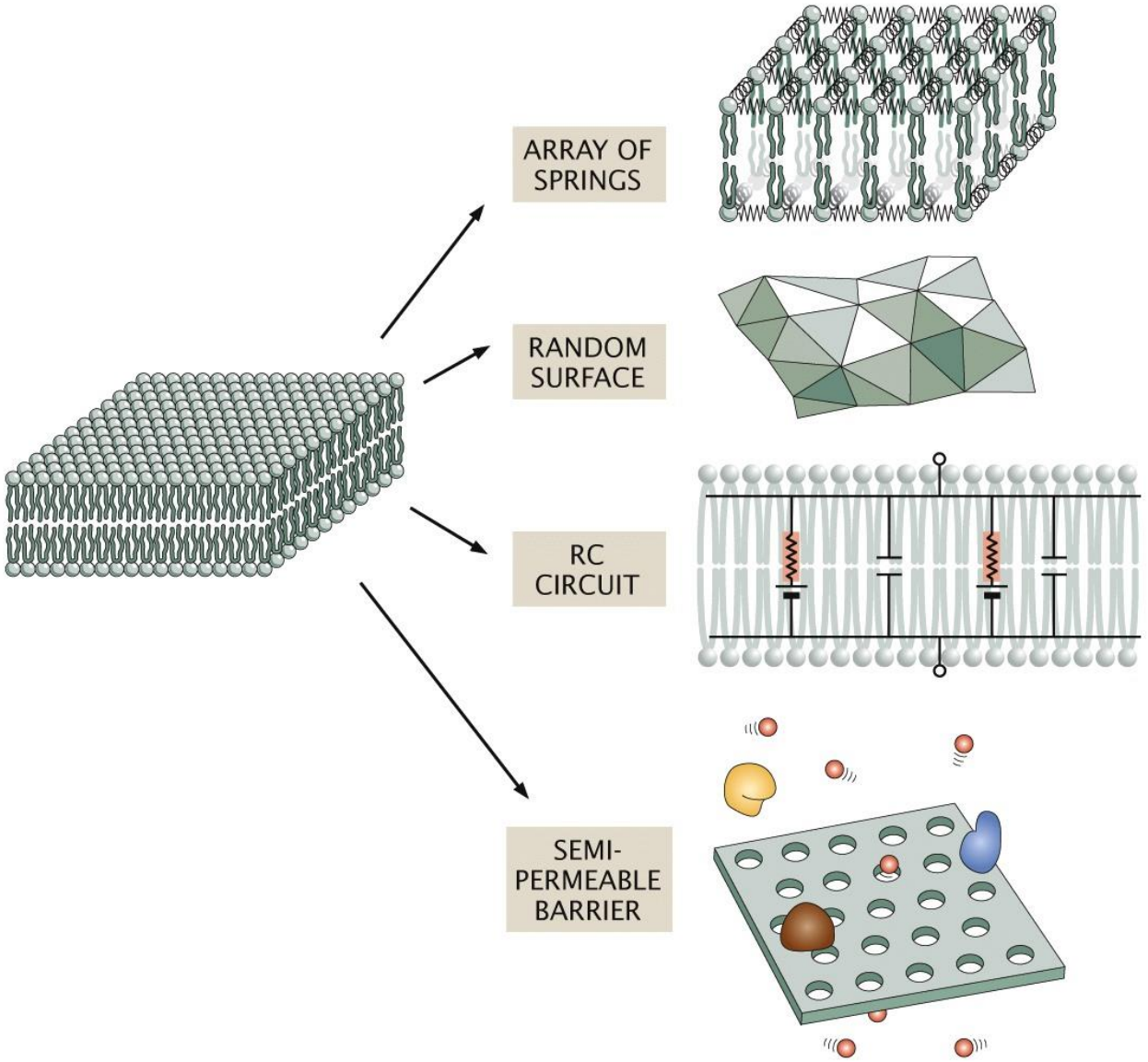
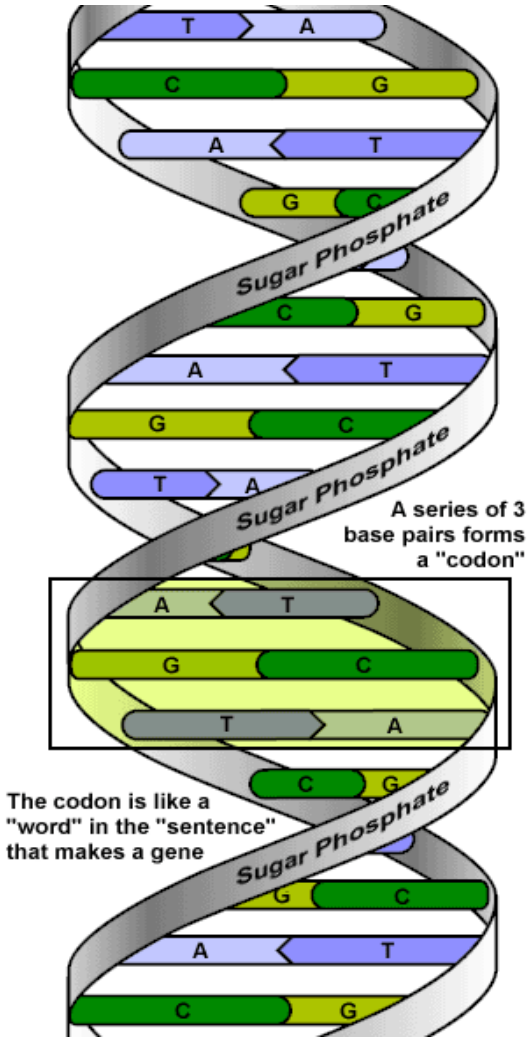
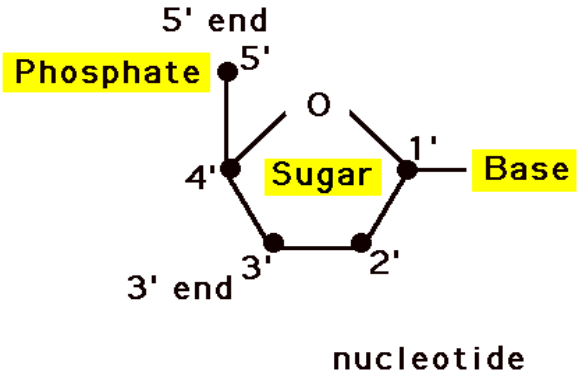


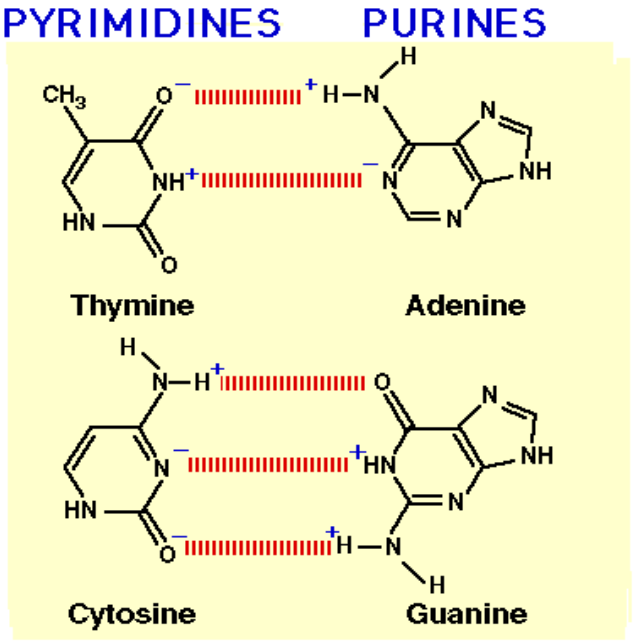
Figure 1.7 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

DNA:

Four different nucleotides: A, C, T, G







A = T

G = C

Watson-Crick Base Pairing

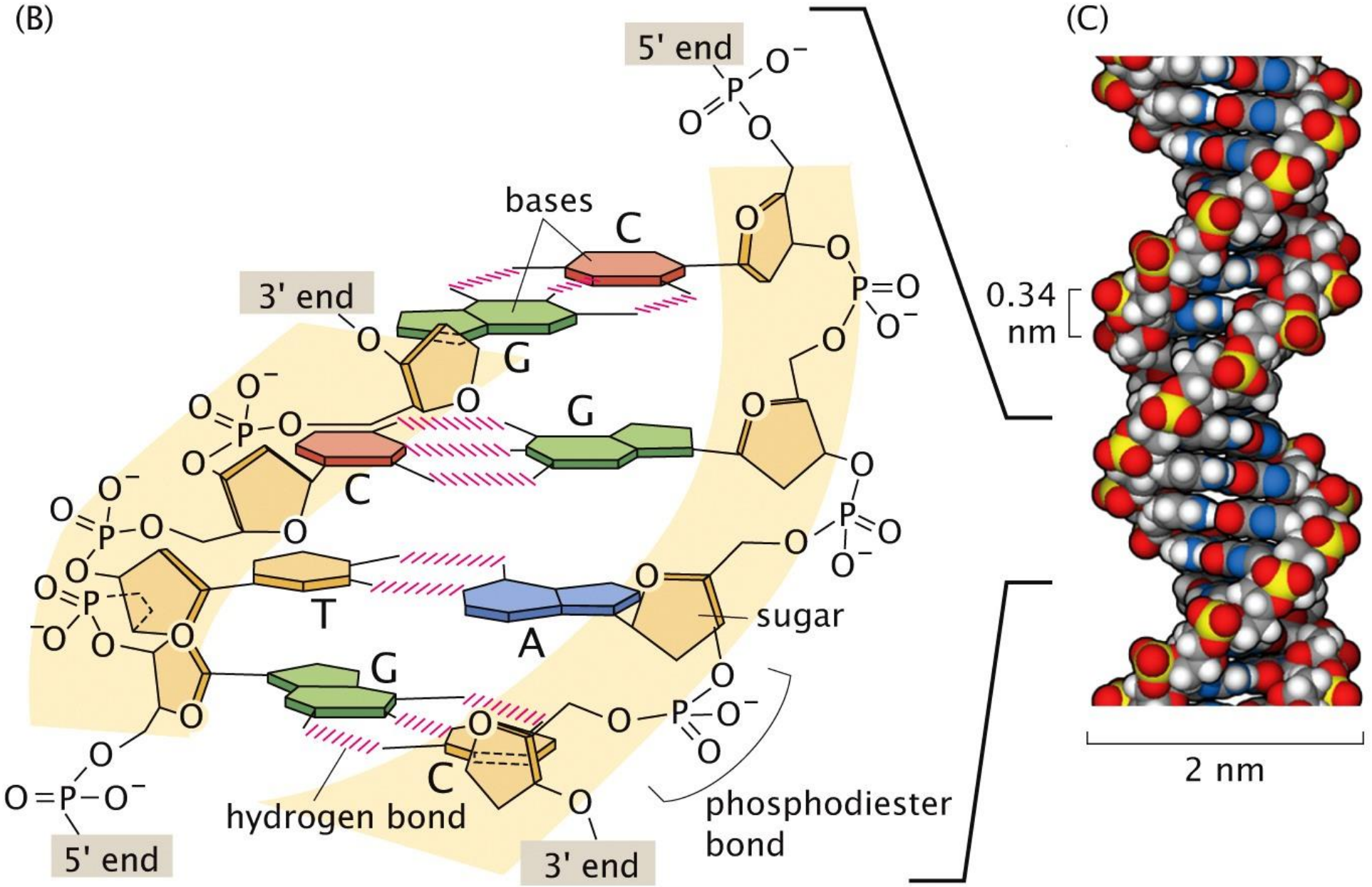
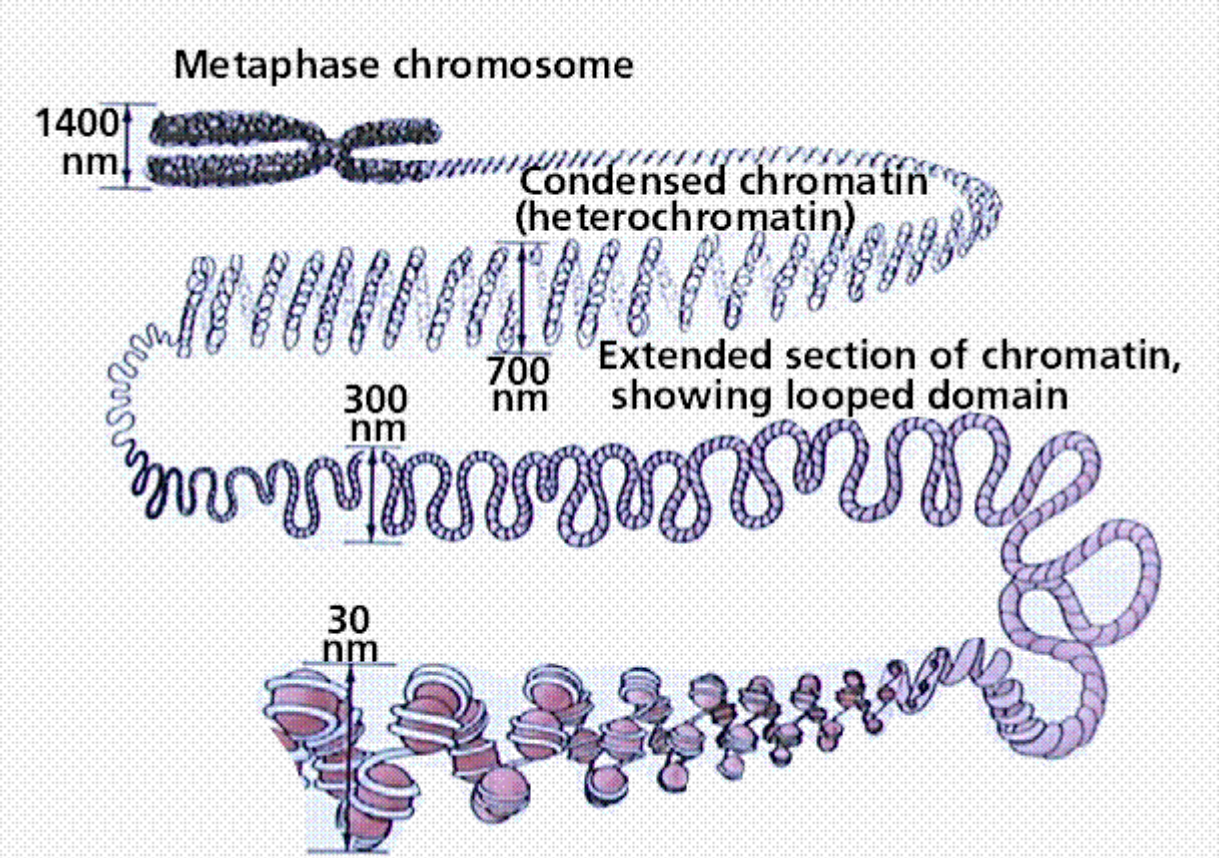


Figure 1.3b Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The Structure of DNA in Cells:

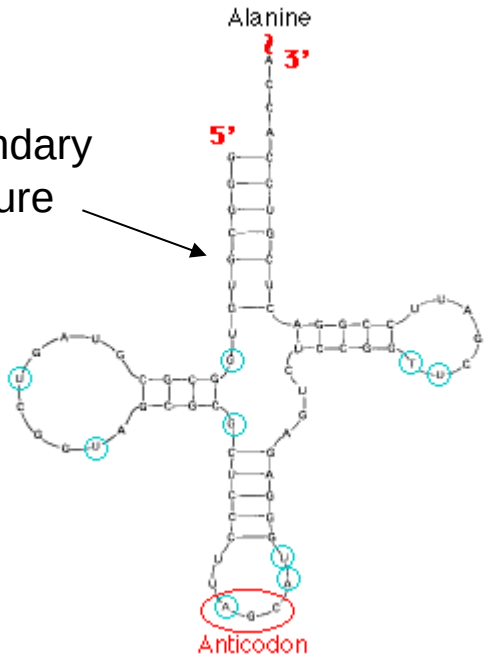
- How do you pack 1 m of DNA into a nucleus that is 2 μm in size?



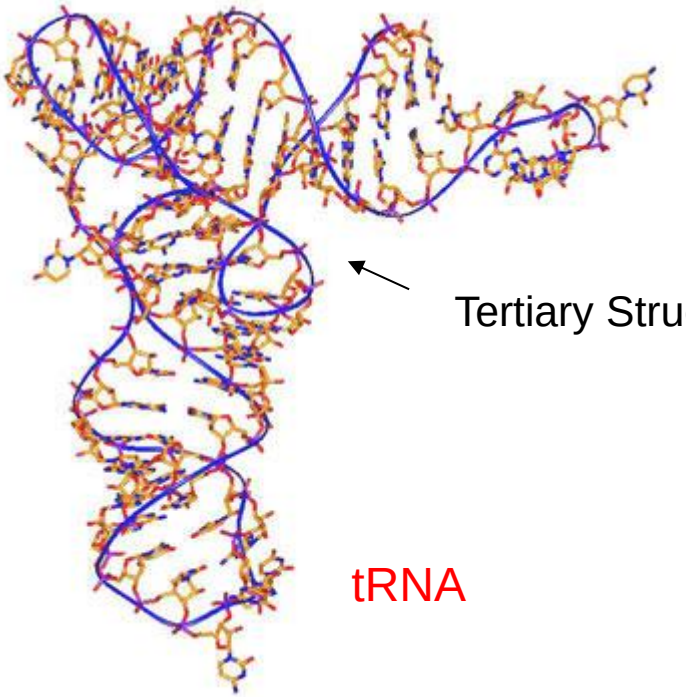
2nm DNA gets wrapped around histone complexes forming what are called nucleosomes (DNA –vely charged & histones are +vely charged)

These then wrap up and pack further and condense

RNA:



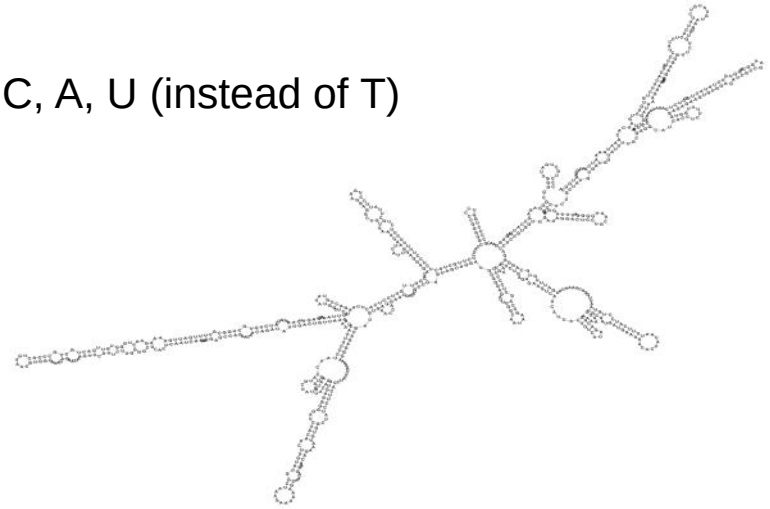
Secondary structure



Tertiary Structure

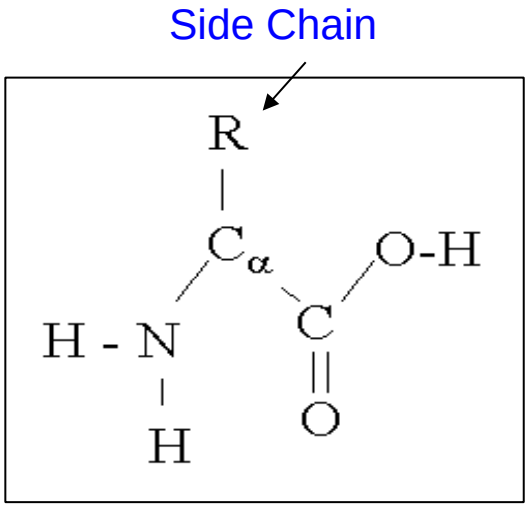
tRNA

Alphabet = G, C, A, U (instead of T)



- carry the genetic information around as mRNA
- can carry out chemical functions

Proteins: Amino Acids



- Proteins are polymers built from 20 amino acids

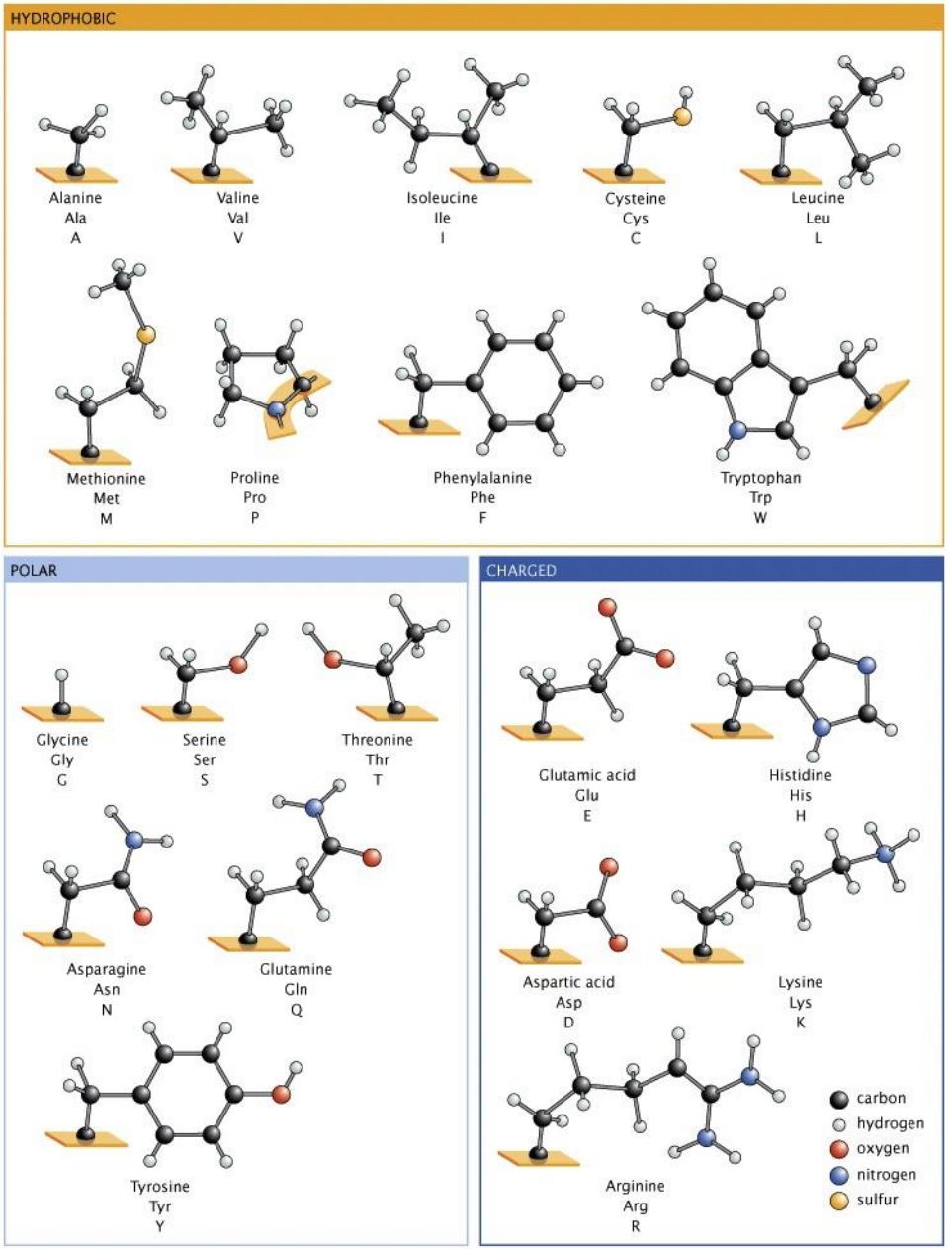
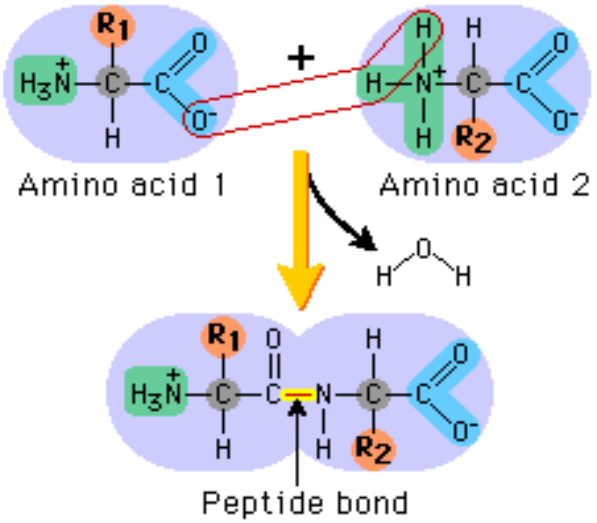
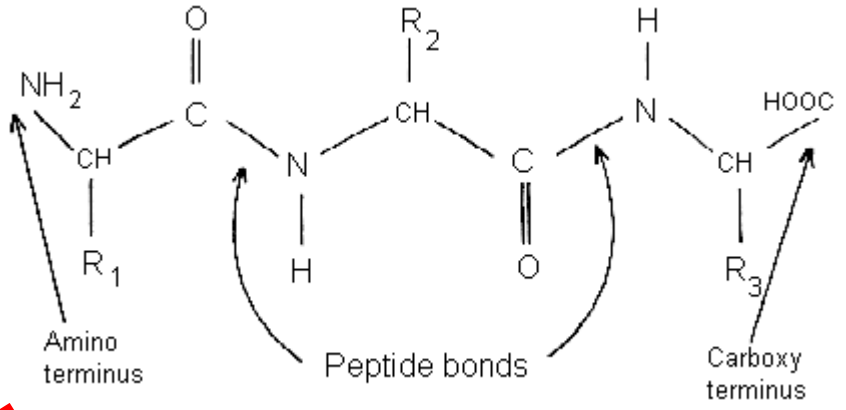
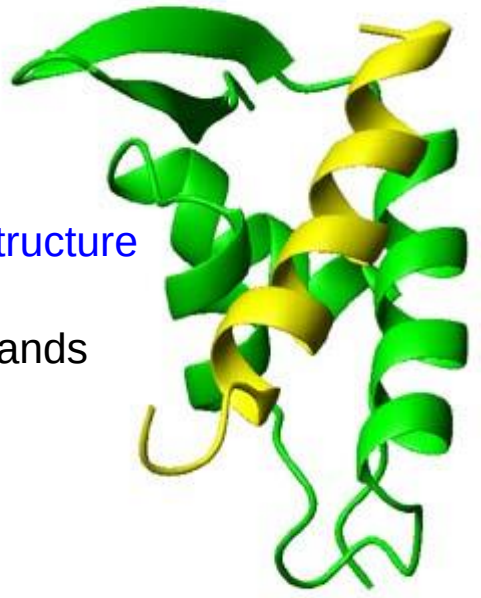


Figure 2.33 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Proteins: Structure

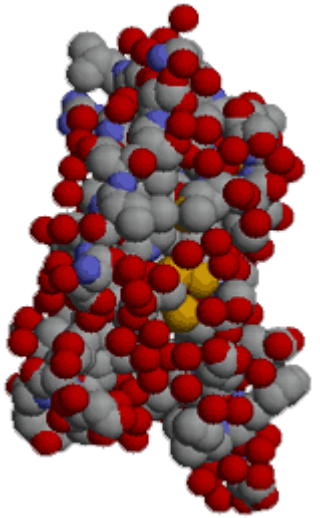






Secondary Structure

Helices & Strands



Tertiary Structure

Densely packed hydrophobic core

Table 1.1: Rules of thumb for biological estimates.

	Quantity of interest	Symbol	Rule of thumb
E. coli			
	Cell volume	$V_{E.\,coli}$	$\approx 1\,\mu m^3$
	Cell mass	$m_{E.\,coli}$	≈ 1 pg
	Cell cycle time	$t_{E.\,coli}$	≈ 3000 s
	Cell surface area	$A_{E.\,coli}$	$\approx 6\,\mu m^2$
	Macromolecule concentration in cytoplasm	$c_{E.\,coli}^{macromol}$	≈ 300 mg/mL
	Genome length	$N_{bp}^{E.\,coli}$	$\approx 5 \times 10^6$ bp
	Swimming speed	$v_{E.\,coli}$	$\approx 20\,\mu m/s$
Yeast			
	Volume of cell	V_{yeast}	$\approx 60\,\mu m^3$
	Mass of cell	m_{yeast}	≈ 60 pg
	Diameter of cell	d_{yeast}	$\approx 5\,\mu m$
	Cell cycle time	t_{yeast}	≈ 200 min
	Genome length	N_{bp}^{yeast}	$\approx 10^7$ bp
Organelles			
	Diameter of nucleus	$d_{nucleus}$	$\approx 5\,\mu m$
	Length of mitochondrion	l_{mito}	$\approx 2\,\mu m$
	Diameter of transport vesicles	$d_{vesicle}$	≈ 50 nm
Water			
	Volume of molecule	V_{H_2O}	$\approx 10^{-2}$ nm^3
	Density of water	ρ	1 g/cm^3
	Viscosity of water	η	≈ 1 centipoise (10^{-2} g/(cm s))
	Hydrophobic embedding energy	$\approx E_{hydr}$	2500 cal/(mol nm^2)

Table 1.1 (part 1 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Table 1.1: Rules of thumb for biological estimates.

	Quantity of interest	Symbol	Rule of thumb
DNA			
	Length per base pair	l_{bp}	$\approx 1/3$ nm
	Volume per base pair	V_{bp}	≈ 1 nm^3
	Charge density	λ_{DNA}	$2\,e/0.34$ nm
	Persistence length	ξ_p	50 nm
Amino acids and proteins			
	Radius of "average" protein	$r_{protein}$	≈ 2 nm
	Volume of "average" protein	$V_{protein}$	≈ 25 nm^3
	Mass of "average" amino acid	M_{aa}	≈ 100 Da
	Mass of "average" protein	$M_{protein}$	$\approx 30{,}000$ Da
	Protein concentration in cytoplasm	$c_{protein}$	≈ 150 mg/mL
	Characteristic force of protein motor	F_{motor}	≈ 5 pN
	Characteristic speed of protein motor	v_{motor}	≈ 200 nm/s
	Diffusion constant of "average" protein in cytoplasm	$D_{protein}$	$\approx 10\,\mu$m^2/s
Lipid bilayers			
	Thickness of lipid bilayer	d	≈ 5 nm
	Area per molecule	A_{lipid}	$\approx \frac{1}{2}$ nm^2
	Mass of lipid molecule	m_{lipid}	≈ 800 Da

Table 1.1 (part 2 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Information Storage, Replication & Processing:

What does the cell use all these molecules/structure for?

1) Storing Chemical Information

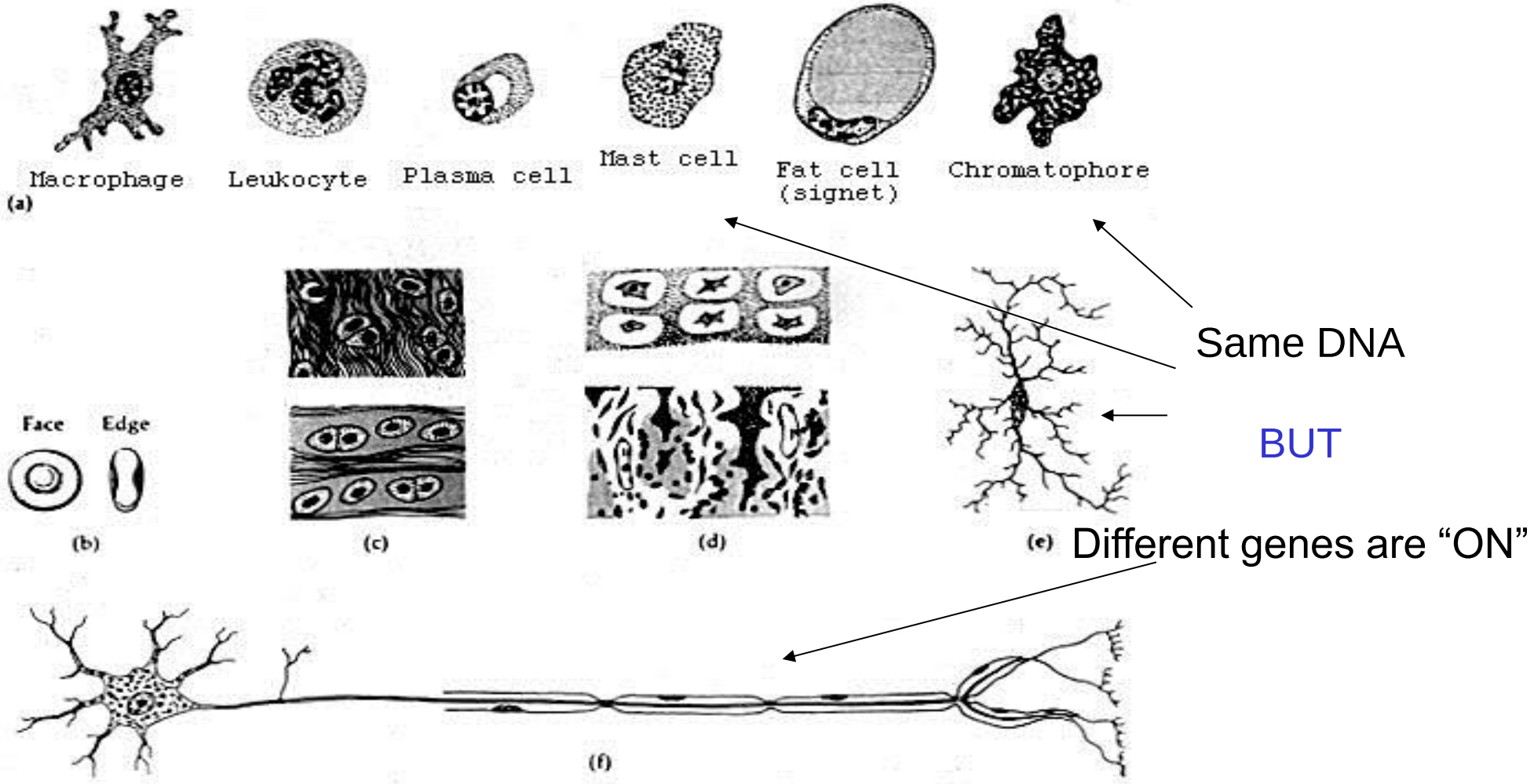
2) Passing On/Replicating this information

3) Processing and calculating with this information

We'll look at each of these steps now

Cellular “states”:

- Cellular state is determined by which genes=proteins are “ON”



- Different cell types = different genetic programs that are being run

Information Storage:

- The information for running the cellular program is stored in the sequence of DNA
- DNA is like the hard drive of a computer
- It stores information, contains programs that get executed at specific times
- Each protein has a specific sequence in the DNA called a ‘gene’, and this DNA is called ‘coding’ DNA, as it codes for messages that make proteins
- The majority of DNA sequence does not code for ‘genes’ and is called non-coding DNA.
- It is the non-coding DNA which contains the ‘programs’ which determine what genes will get made at a given time or place
- different cells are running different programs and thus making different sets of genes

The Sizes of the Hard-drives

Bacteria (usually) have single circular chromosome – have many small plasmids

Eukaryotes have multiple linear chromosomes

- haploid = single copy of every chromosome
- diploid = two copies of every chromosome
- polypoid = multiple copies of every chromosome

organism	#	T	size	genes
virus	1	H	5kb	10
E. coli	1	H	5Mb	4,377
S. Cervisiae	16	H	12Mb	5,570
C. elegans	16	D	100Mb	19,000
A. thaliana	5	D	115Mb	25,498
D. melanogaster	4	D	122Mb	14,000
H. sapiens	23	D	3Gb	25,000

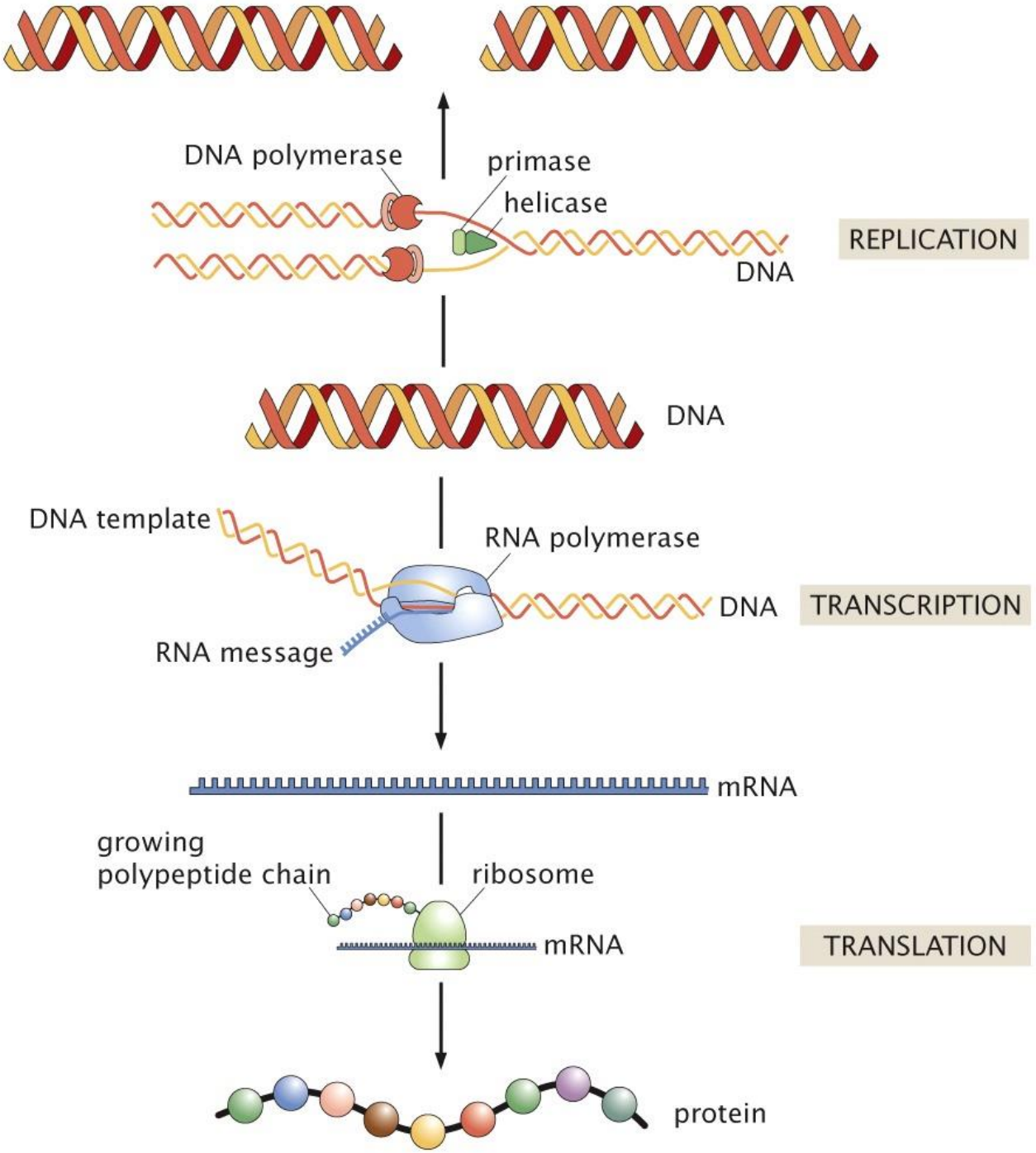
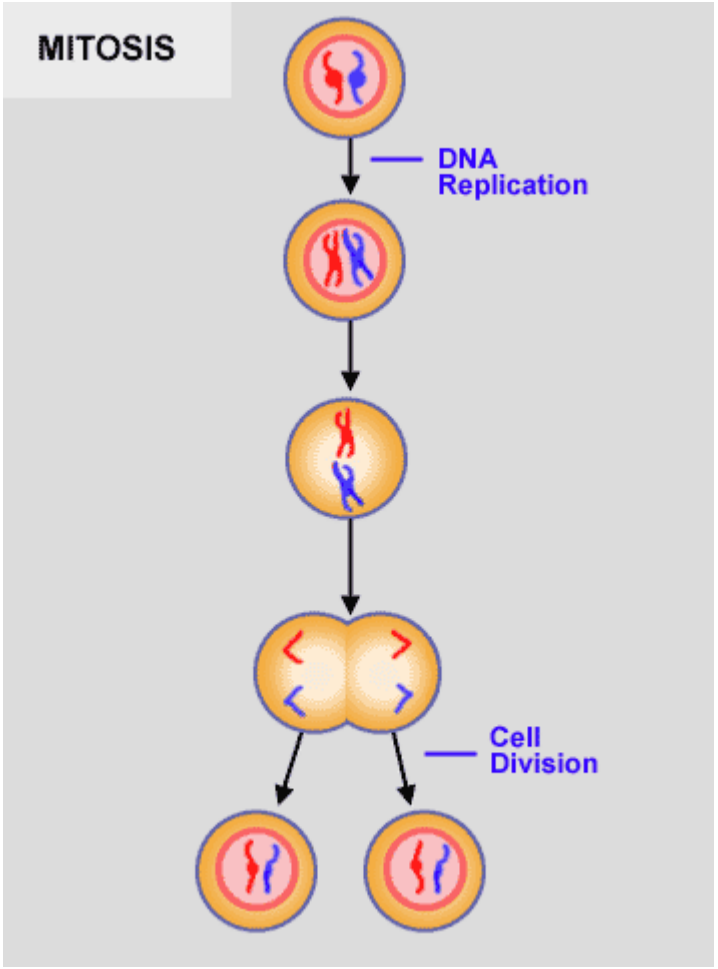


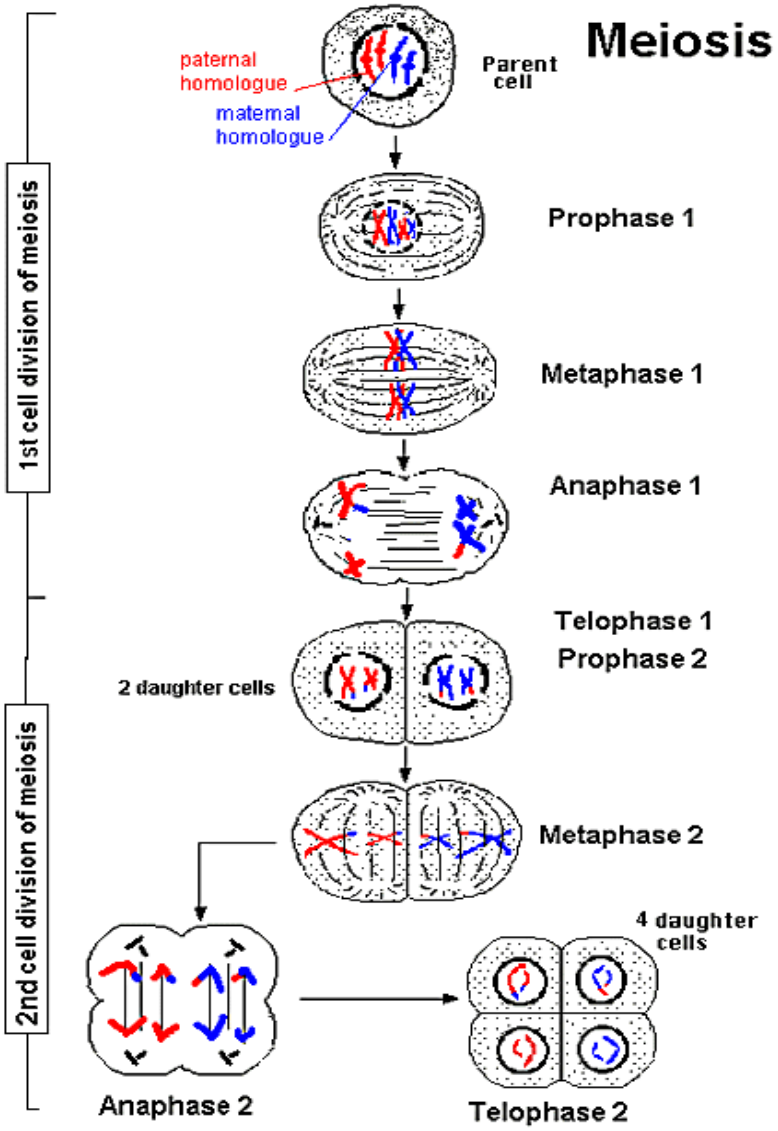
Figure 3.12 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Information Replication: Cell Division:

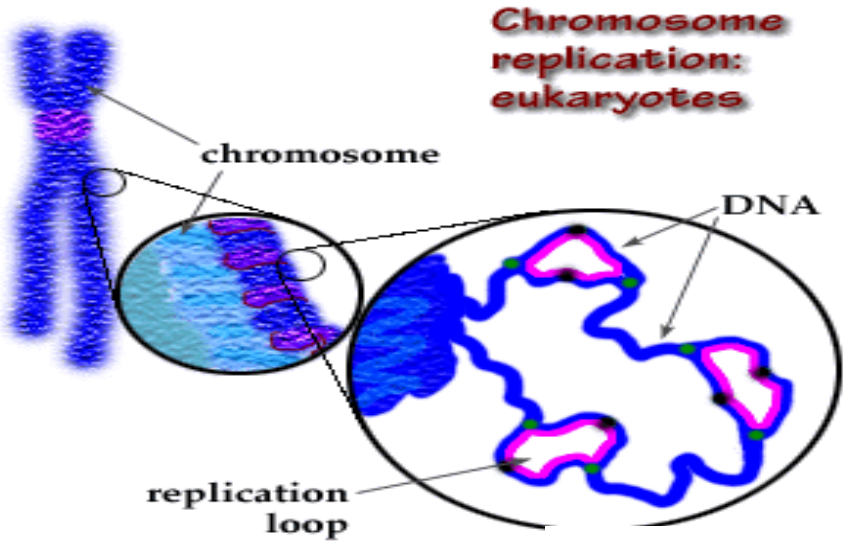
Asexual and normal cell division



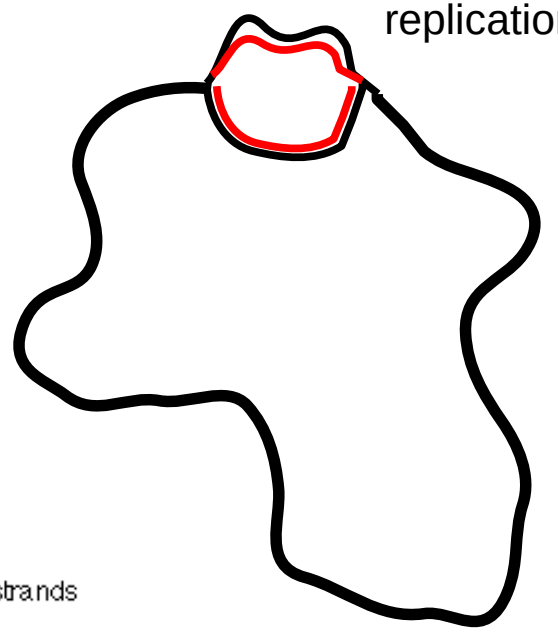
Sex cells = gametes = haploids



DNA replication:



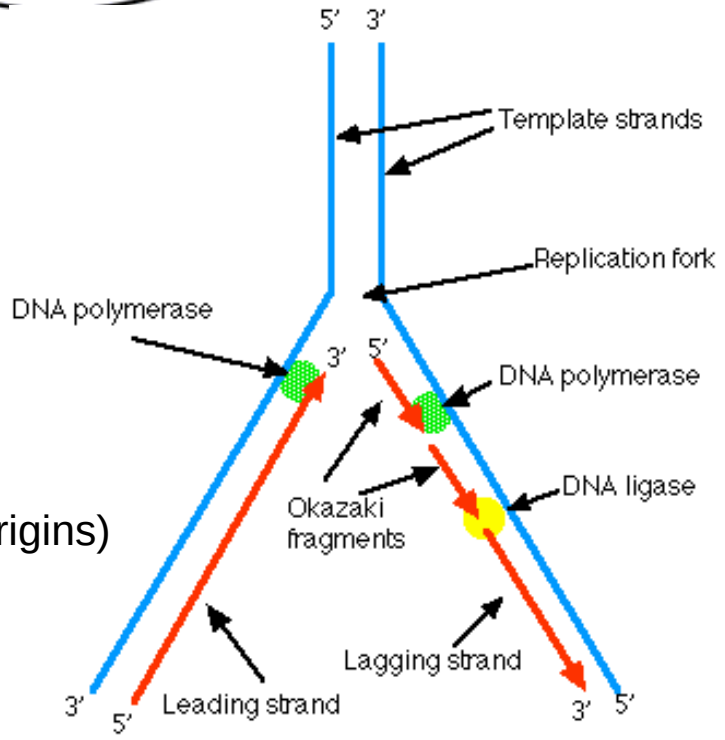
Bacteria have single origin of replication



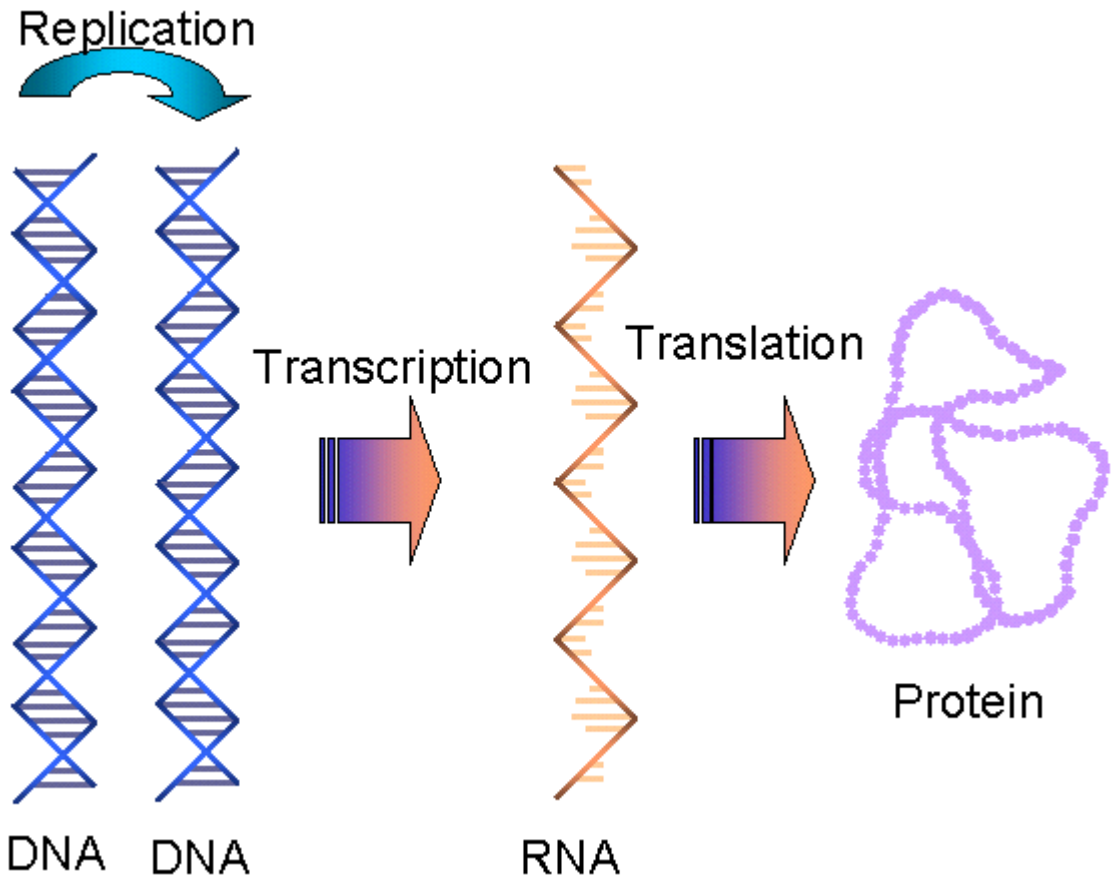
Replication Times:

bacteria ~ 40 mins (1 origin)
1000 nucleotides/s

humans ~ few hours (many origins)
100 nucleotides/s



Information Processing: Central Dogma:



DNA to Proteins? Genetic Code

DNA → mRNA → Protein

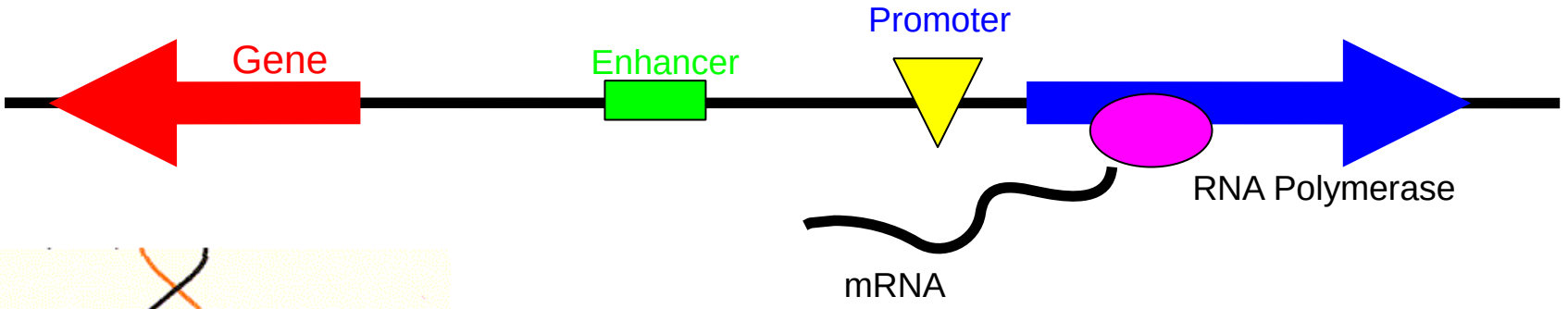
ATGCAGCCGATATAA → AUG CAG CCG AUA UAA → MEPI(STOP)

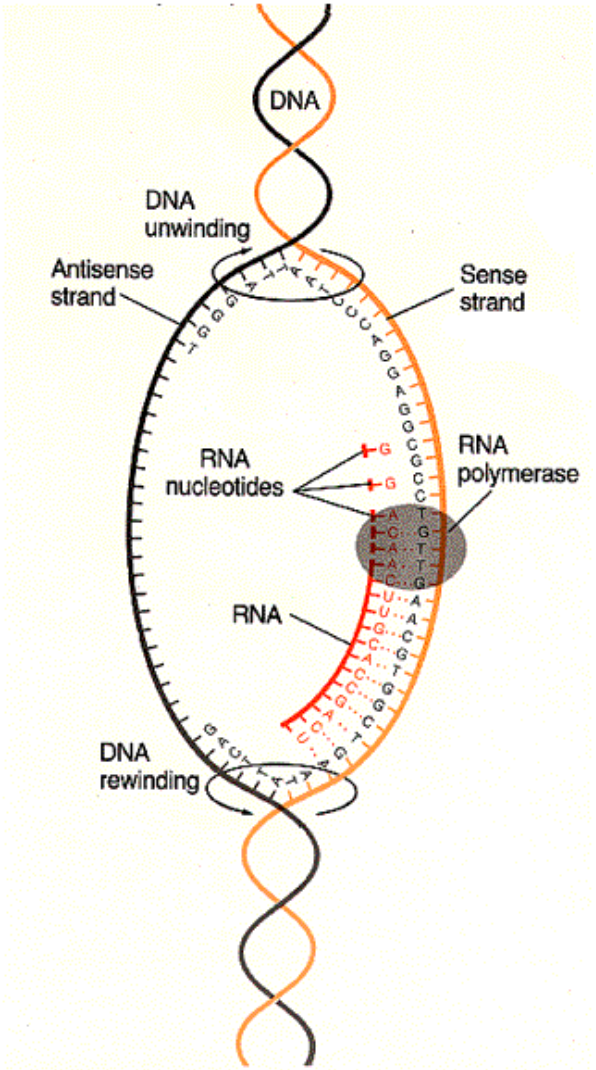
- 64 codons, with highly non-uniform mapping

Start →

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

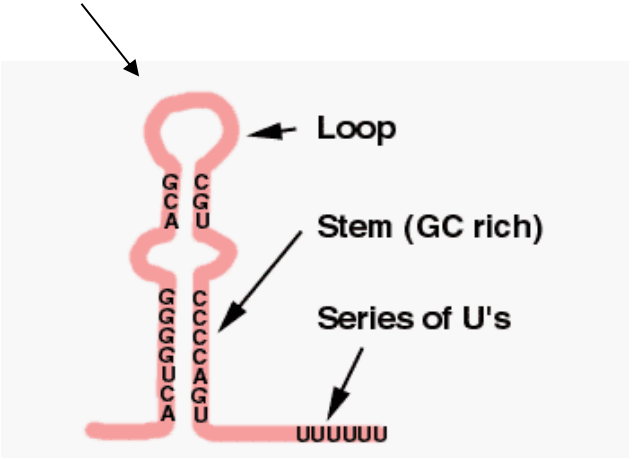
Gene Transcription:





- Bacteria = ON
- Eukaryotes = OFF, requires lots of other help

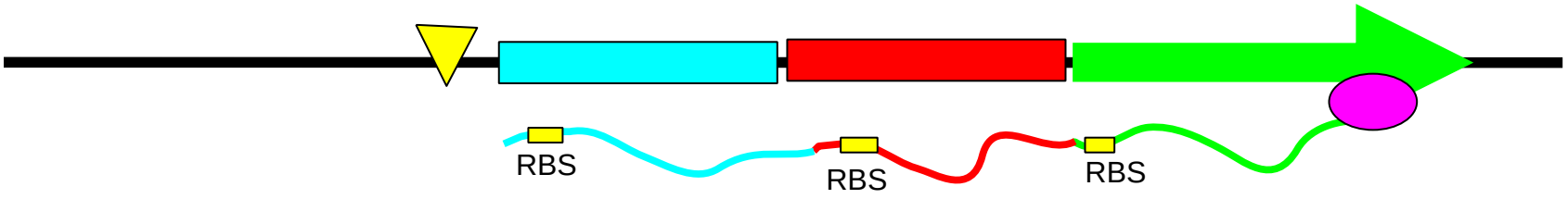
Pushes polymerase off of DNA



Messages are delivered differently:

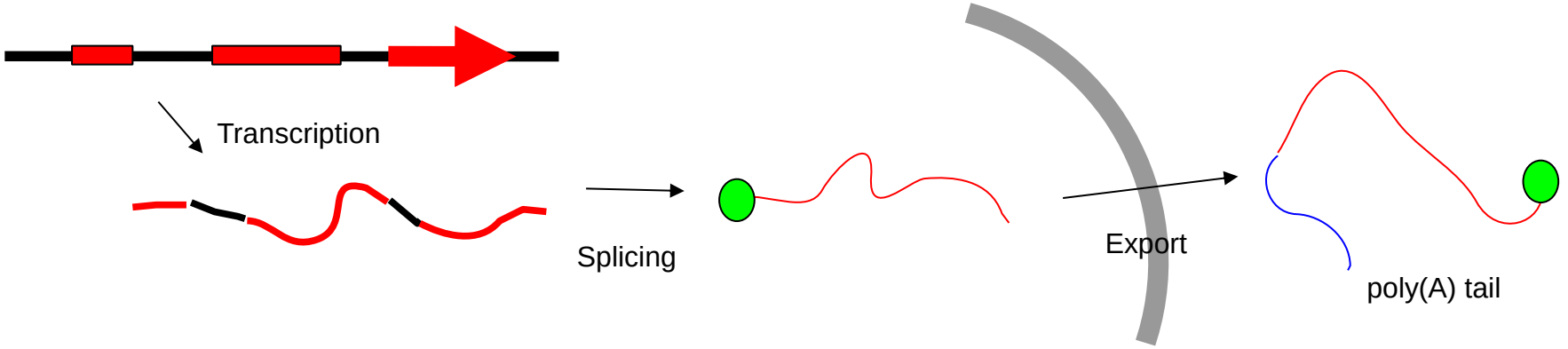
Prokaryotes (Simple):

- Genes in bacteria are organized in operons – one contiguous message RNA is generated

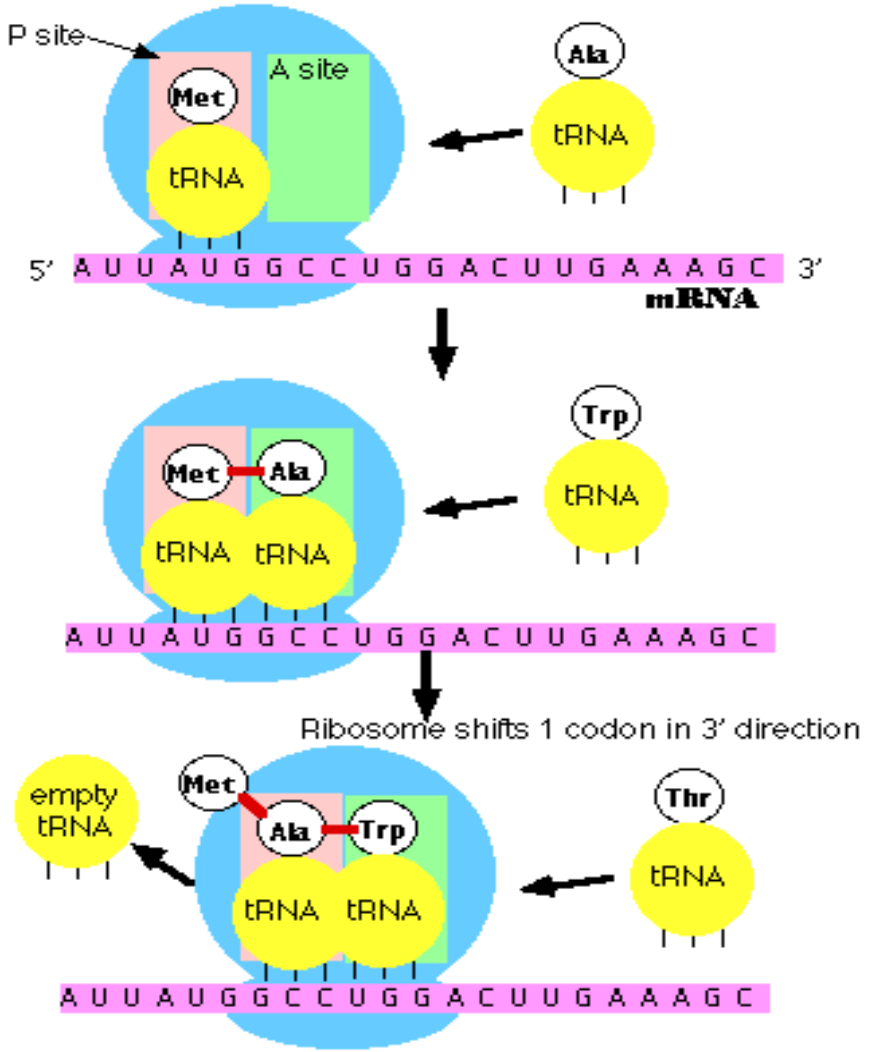


Eukaryotes (Complex):

- Genes contain coding (exons) and non-coding (introns) sequence
- Introns are spliced out and exons are assembled to make final mRNA
- final mRNA is exported out of nucleus and made into protein



Translation:



- Ribosome is big machine that translates mRNA into protein
- In bacteria, Ribosome binds to specific site on mRNA
- In eukaryotes, Ribosome binds to 5' cap
- translation ends when 'STOP' codon is reached

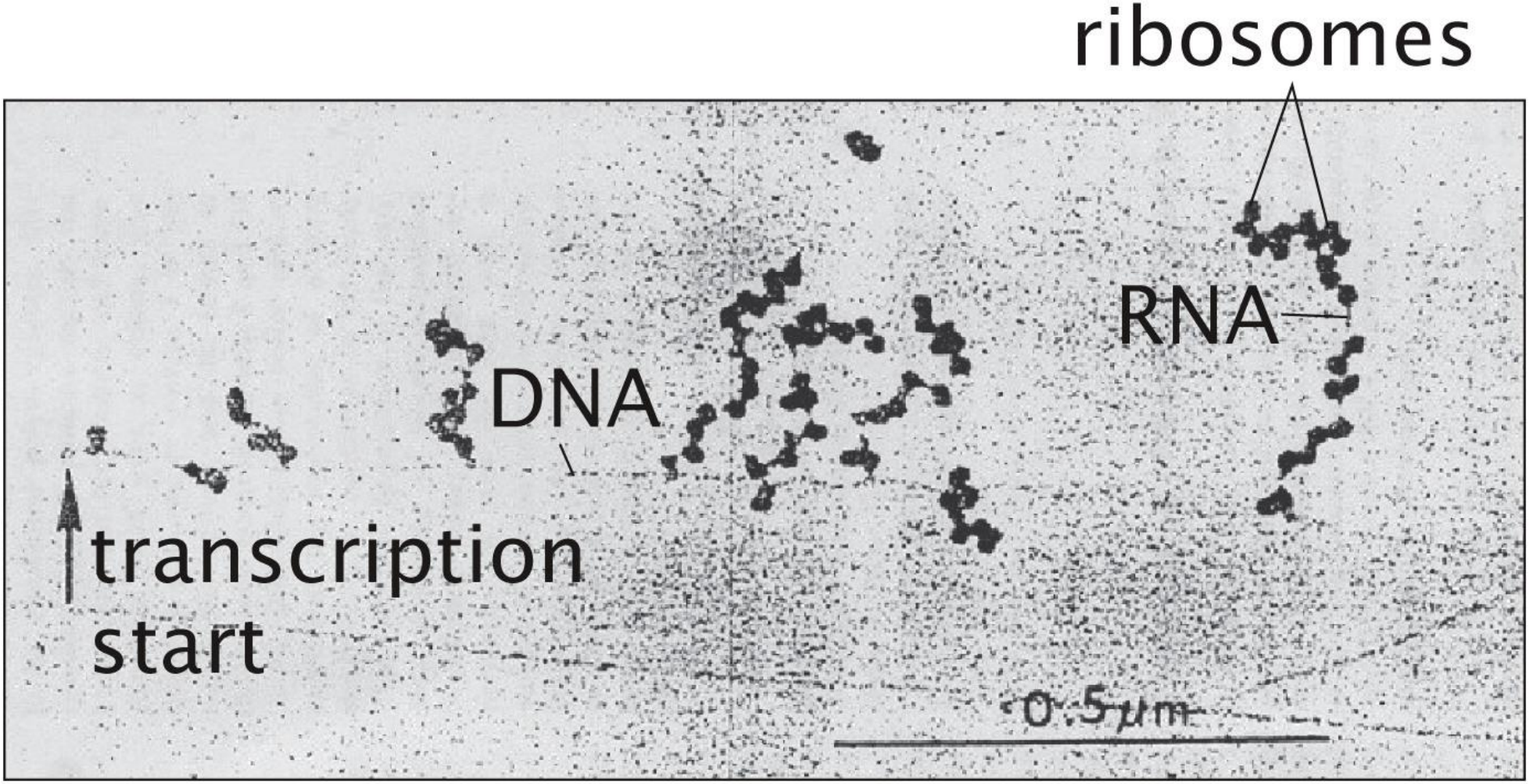
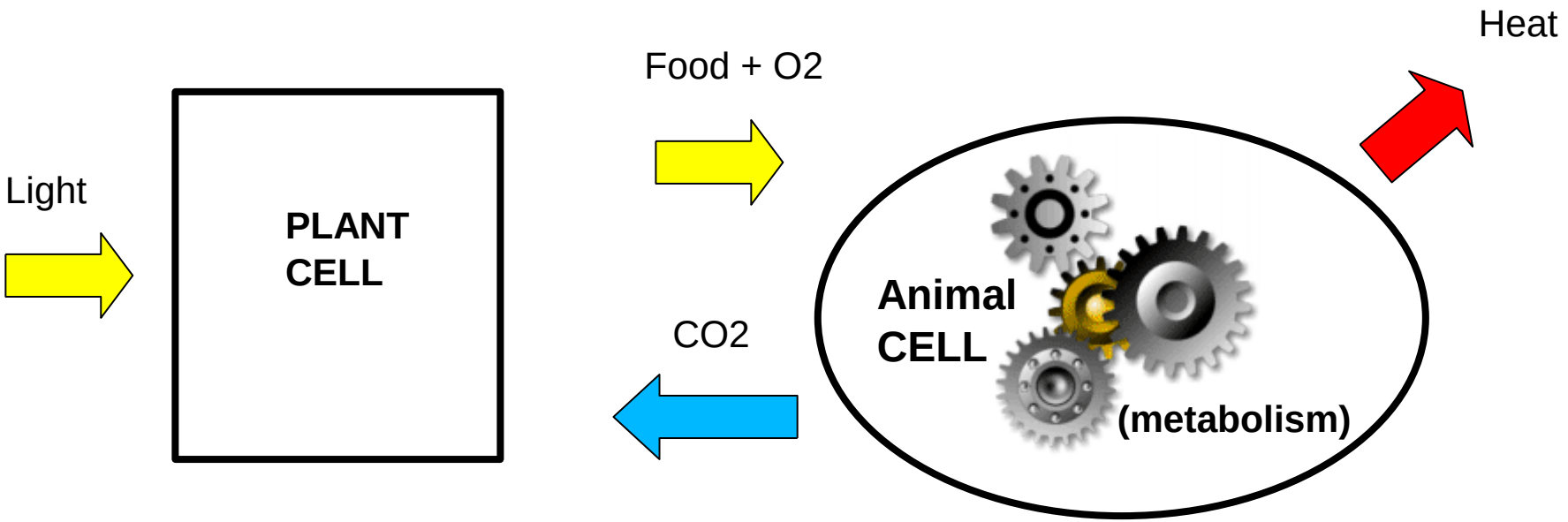


Figure 3.13 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

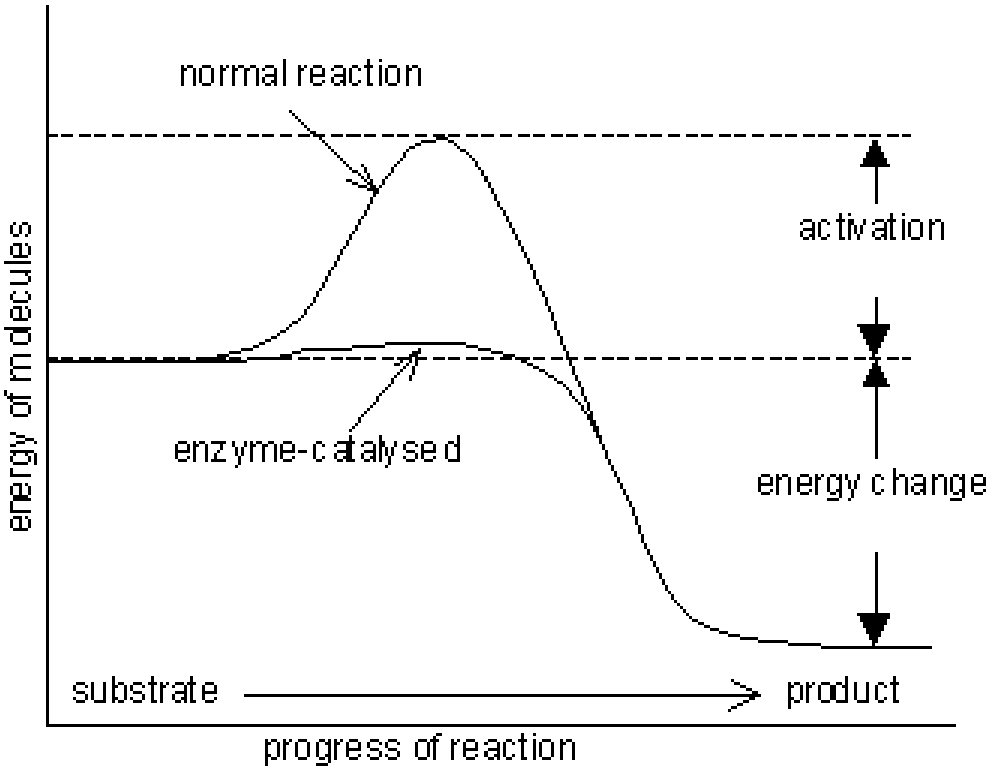
How cells get energy



- Oxidation of food → CO2 and H2O and energy
- Metabolism = processes of converting/synthesizing molecules in cells
  - Catabolism = breaking down of molecules
  - Anabolism = synthesizing of molecules
- Two major energy reserves: sugars, fatty acids

Biochemical reactions

- Most chemical reactions in cells are mediated by proteins called enzymes
- Enzymes act as catalysts, by reducing the activation barrier between two states

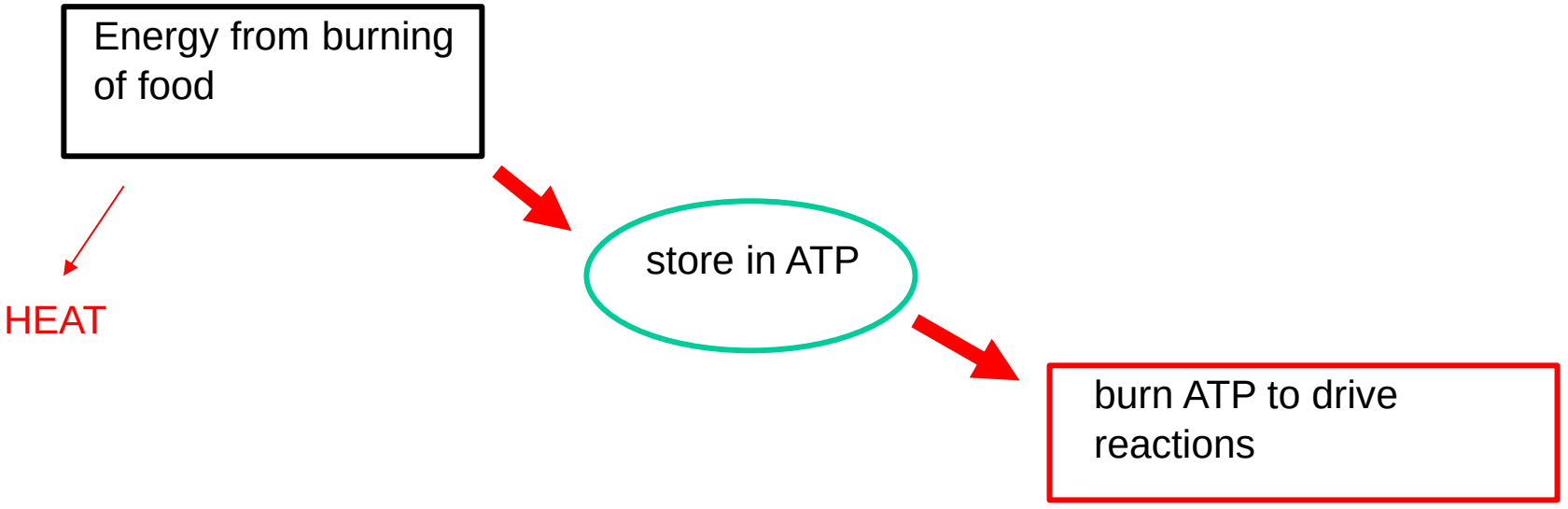


Energy Storage in Cells

- Some chemical reactions in cells are energetically unfavourable (e.g DNA & protein synthesis)
- Couple energetically bad pathway with favourable pathway using enzyme
- favourable pathway is driven by taking energy from energy carriers
- Most pervasive energy carrier is ATP

$$ATP \rightarrow ADP + P + H20 + (13\ kcal/mol = 25\ kT)$$

- DNA synthesis burns 2 ATP



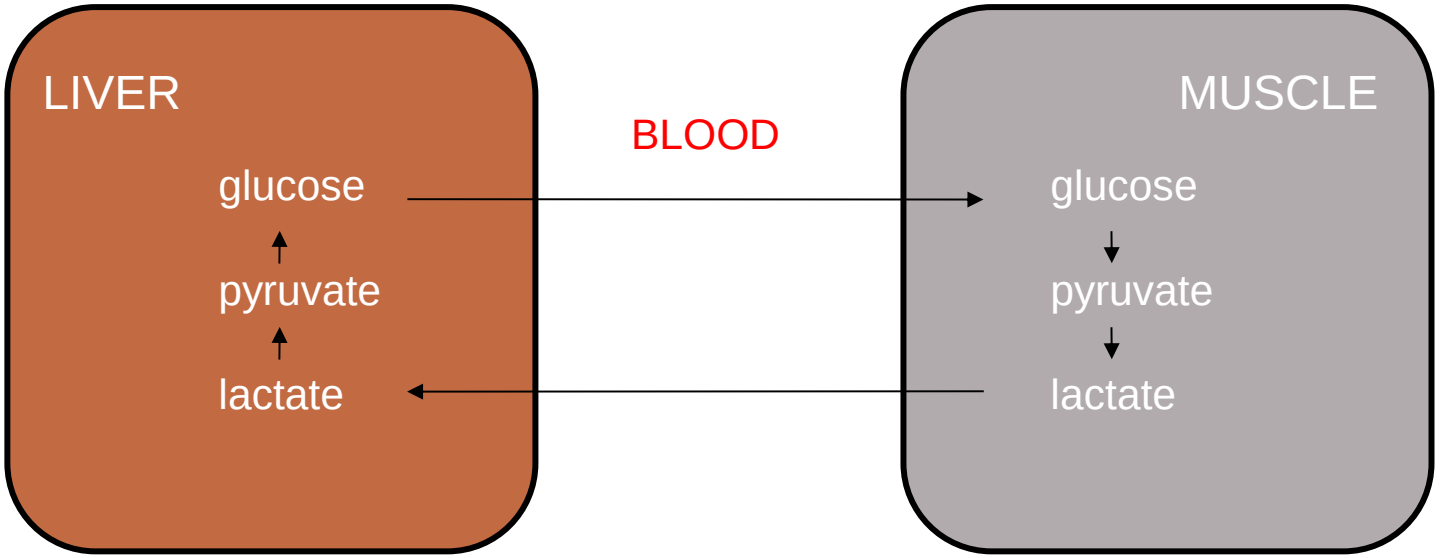
Converting Food to Useful Energy:

Glycolosis

- anaerobic
- converts glucose → 2 ATP

Citric Acid Cycle = Krebs Cycle:

- aerobic
- occurs in the mitochondria
- generates about 30 ATP from a single glucose!!!
- convert about 50% of possible energy (c.f. car ~ 20%)



Numbers of things in cell biology

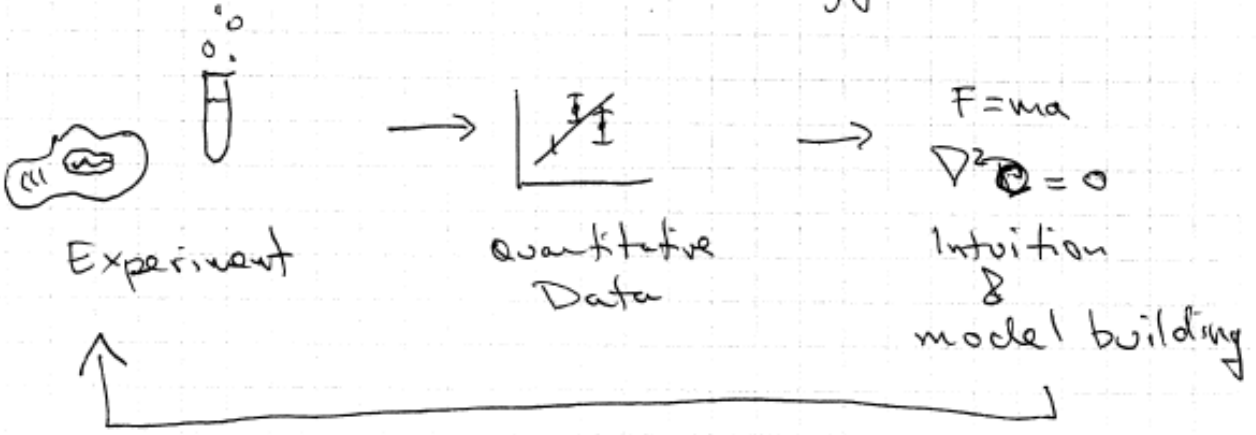
Why knowing numbers in biology matters?

- Given the vast range of scales and the large variety of parts, its important to have a sense of the numbers of things.
- Why? ① being able to estimate numbers of parts etc builds intuition.

  ② allows one to assess what aspects of a given biological process will be important quantitatively

  - in complex systems, the things one ignores are as important as the things that are considered.
  - models are always incomplete

Roadmap to Quantitative Biology:



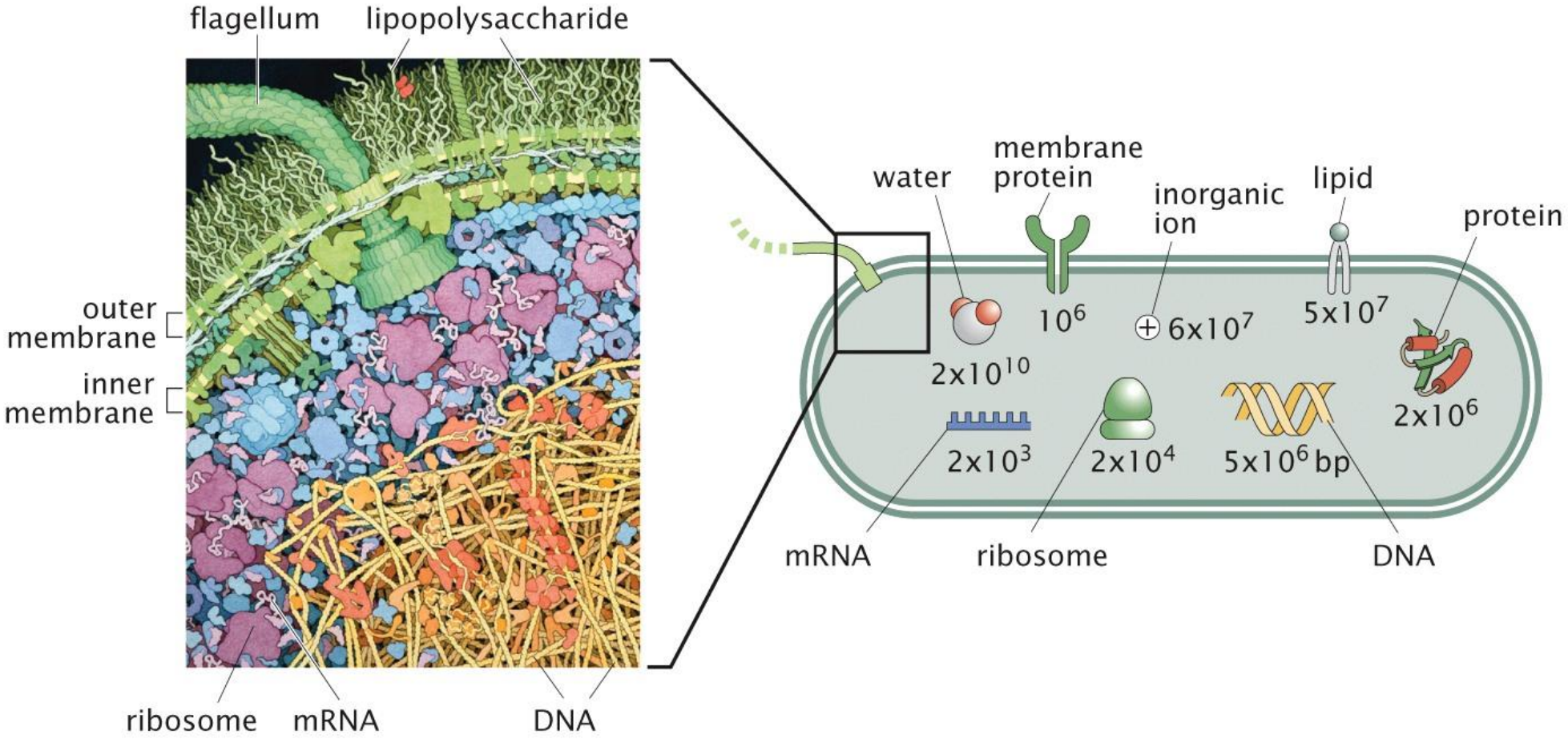


Figure 2.4 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Some estimates of parts:

- Volume: $V_{cell} \approx 1\,\mu m^3 = 1\,fL$
- Area: $A_{cell} \approx 6\,\mu m^2$

Cells are crowded:

Mass: $Mass = density \times volume$

$\approx density_{H_2O} \times V_{cell}$

$\approx 1\,g/mL \times 1\,fL = 1\,pg$

Dry mass: Experiment $\rightarrow$ $M_{dry} \approx 0.3\,pg = 30\%$ Mass

Protein mass: avg protein = 300 amino acids (AA)

$1\,AA = 100\,Da$ & $1\,Da = m_{proton} = 1.6 \times 10^{-24}\,kg$

so average protein mass,

$$m_{protein} = (300) \times (100\,Da) \times (1.6 \times 10^{-24}\,g) \approx 5 \times 10^{-20}\,g$$

Number of proteins in E. coli:

50% of dry mass is protein from experiment

$$N_{protein} = \frac{\text{protein mass}}{\text{mass per protein}} = \frac{0.15\,pg}{5 \times 10^{-20}\,g} \approx 3 \times 10^6$$

Some estimates of parts

Number of Ribosomes:

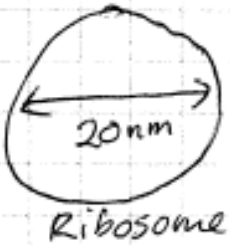
Expt: 20% of protein mass is contributed by ribosomes

$M_{ribosome} = 2.5$ MDa

but 1/3 Ribosome is protein & 2/3 is rRNA

$$N_{Ribosome} = \frac{(0.2)(0.15\ pg)}{830{,}000\ Da} = 20{,}000$$

How much volume do ribosomes occupy?



$$V_{ribosome} = \frac{4}{3}\pi (10\ nm)^3$$

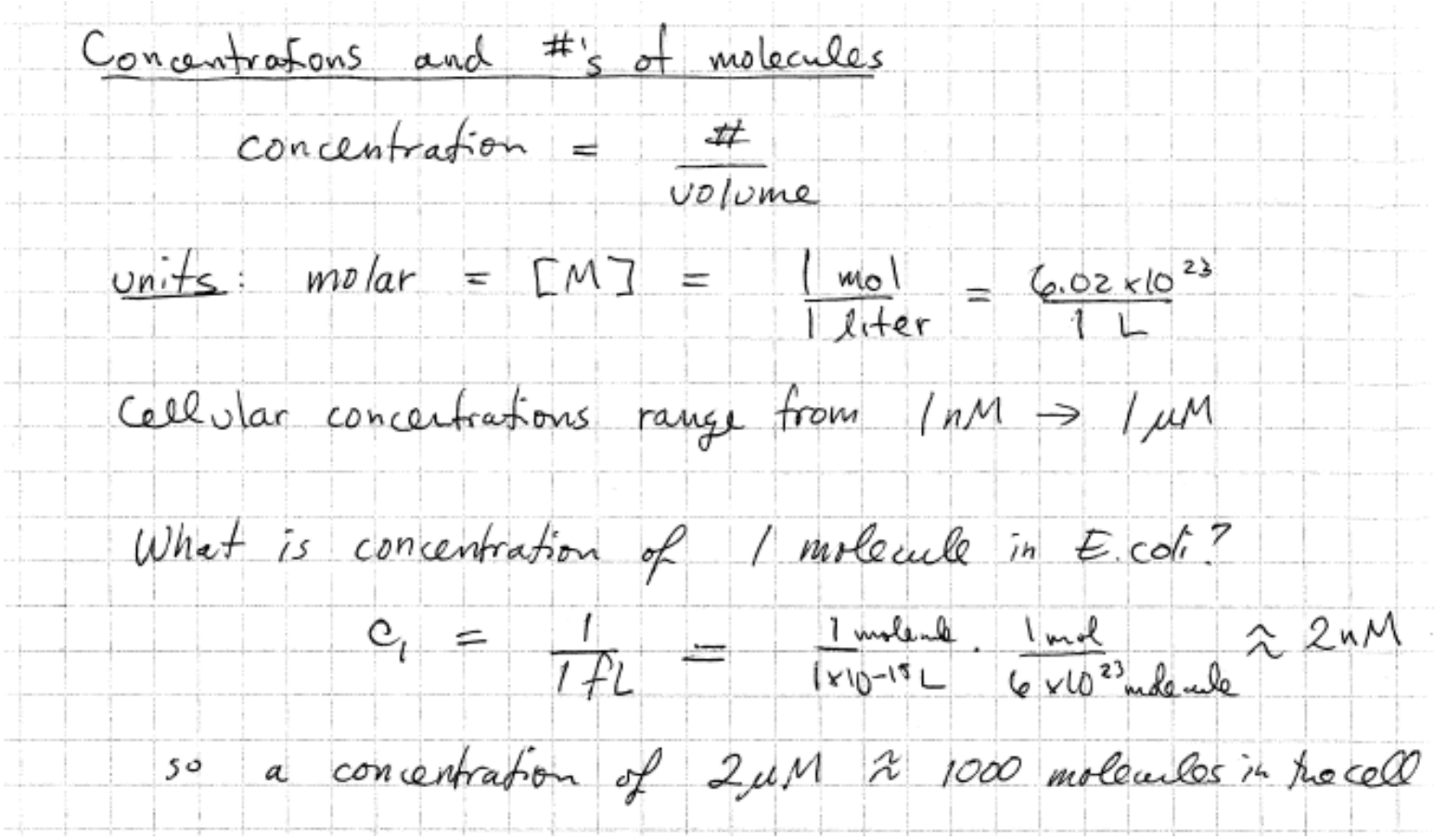
$$V_{tot} = (20{,}000)\frac{4}{3}\pi(10nm)^3 \approx 10^8\ nm^3$$

$$\approx 10\%\ V_{cell}$$

Table 2.1: Observed macromolecular census of an *E. coli* cell. (Data from F. C. Neidhardt et al., Physiology of the Bacterial Cell, Sinauer Associates, 1990 and M. Schaechter et al., Microbe, ASM Press, 2006.)

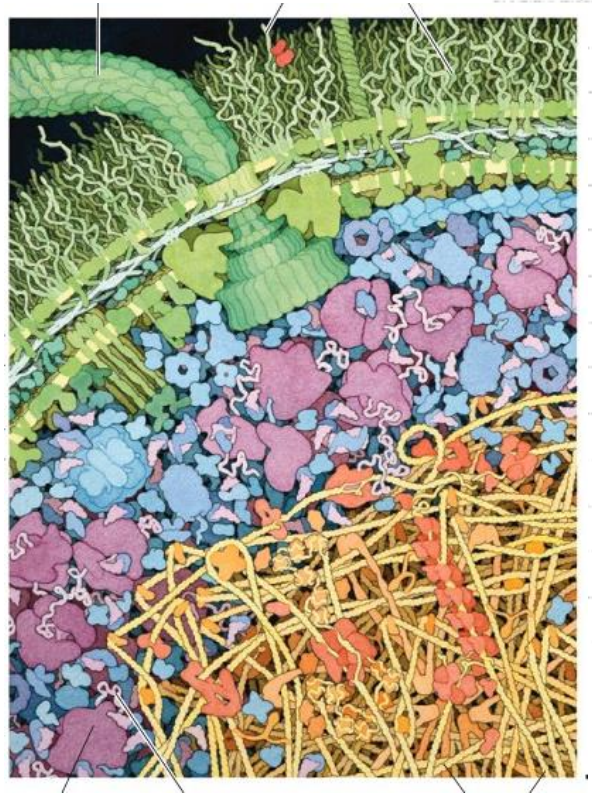
Substance	% of total dry weight	Number of molecules
Macromolecules		
Protein	55.0	2.4×10^6
RNA	20.4	
23S RNA	10.6	19,000
16S RNA	5.5	19,000
5S RNA	0.4	19,000
Transfer RNA (4S)	2.9	200,000
Messenger RNA	0.8	1,400
Phospholipid	9.1	22×10^6
Lipopolysaccharide (outer membrane)	3.4	1.2×10^6
DNA	3.1	2
Murein (cell wall)	2.5	1
Glycogen (sugar storage)	2.5	4,360
Total macromolecules	**96.1**	
Small molecules		
Metabolites, building blocks, etc.	2.9	
Inorganic ions	1.0	
Total small molecules	**3.9**	

What is the concentration of 1 molecule in a E. coli?



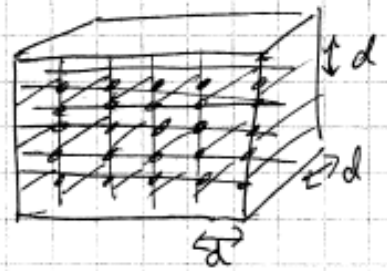
1 nM ~ 1 molecule inside the volume of an E. coli cell

Cells are crowded places



Distance between molecules:

- Assume molecules sit on a square lattice in a square cell of volume V



spacing between molecules is d.

$$c = \frac{N}{V} = \frac{N}{(N \cdot d^3)} = d^{-3}$$

so spacing $= d = c^{-1/3}$

for $c = 2\,\mu M \rightarrow d \sim 150\,nm$

for $c = (1 \times 10^6)\,1\,nM \rightarrow d \sim 1\,nm \leftarrow$ hardly any space between proteins

(1×10^6) ~ # of proteins

The timing of things: timescales in biology

Why timescales matter?

Biology is not static, it's dynamic!

It has dynamics over a range of timescales

Q: How do we treat physical systems that have processes that are operating over many different timescales

A: We choose a timescale of interest and only consider processes that are of the same scale. Faster processes will be considered to be at steady state - we'll define this later.

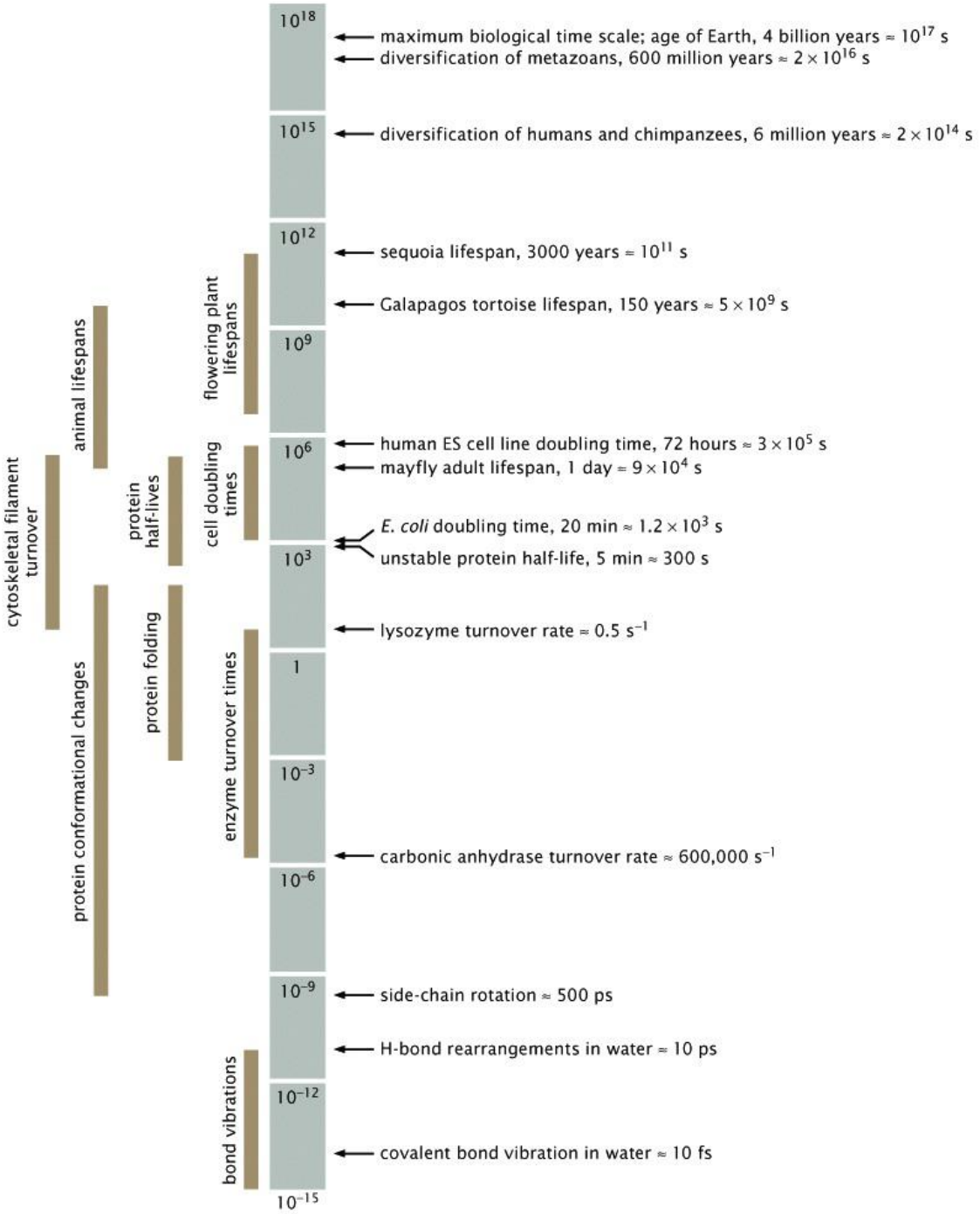


Figure 3.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Development of *Drosophila*

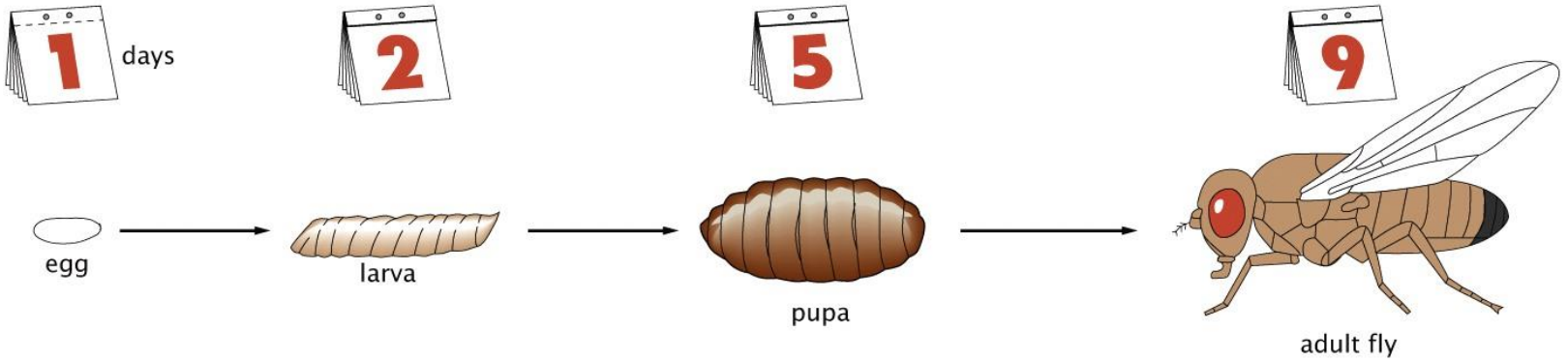


Figure 3.2a Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Early development of *Drosophila*

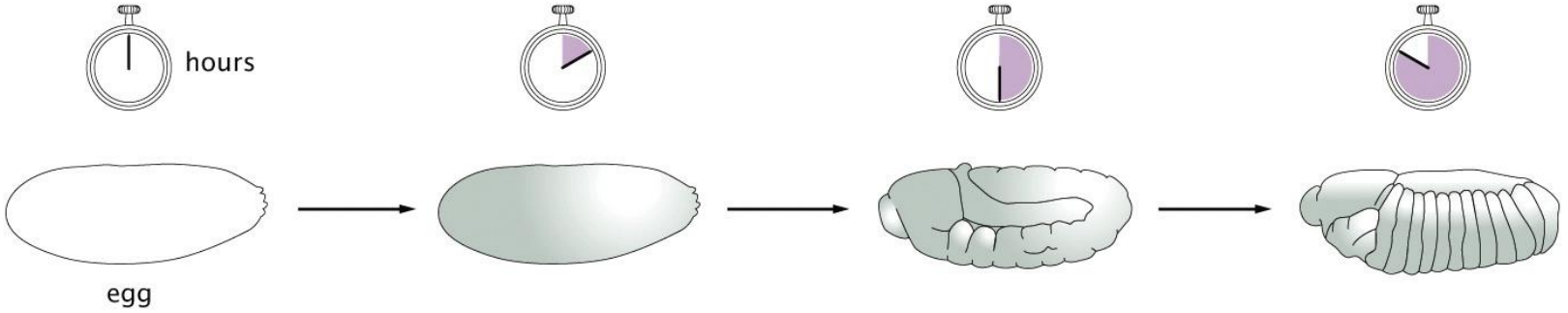


Figure 3.2b Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Bacterial cell division

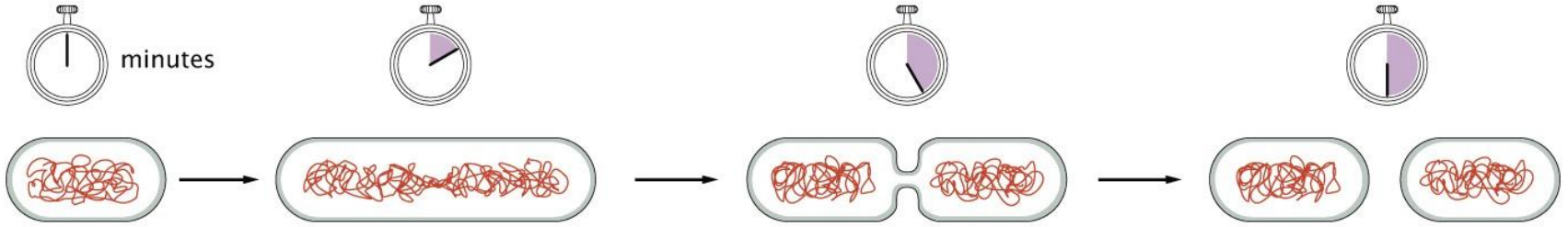


Figure 3.2c Physical Biology of the Cell, 2ed. (© Garland Science 2013)

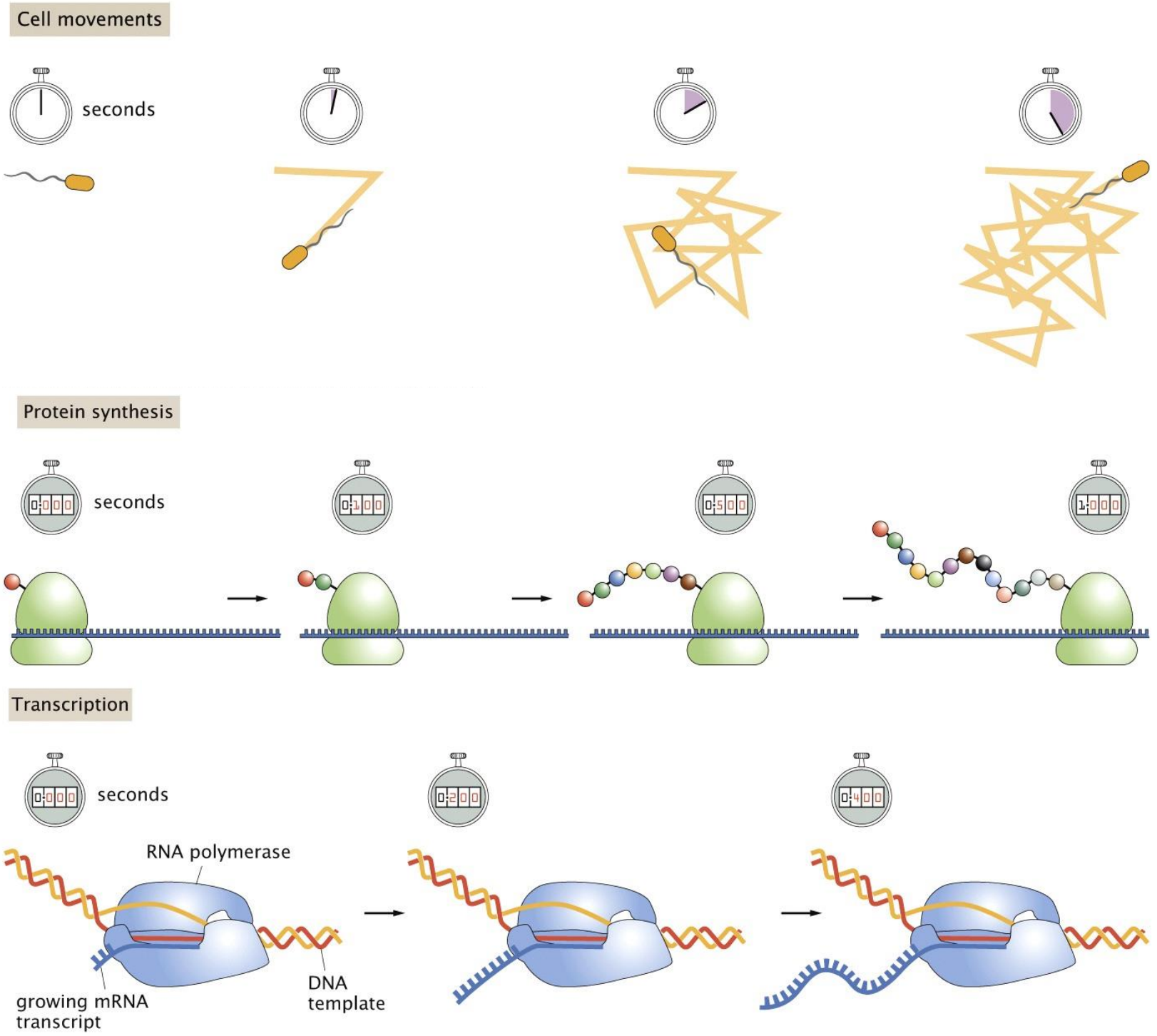


Figure 3.2f Physical Biology of the Cell, 2ed. (© Garland Science 2013)

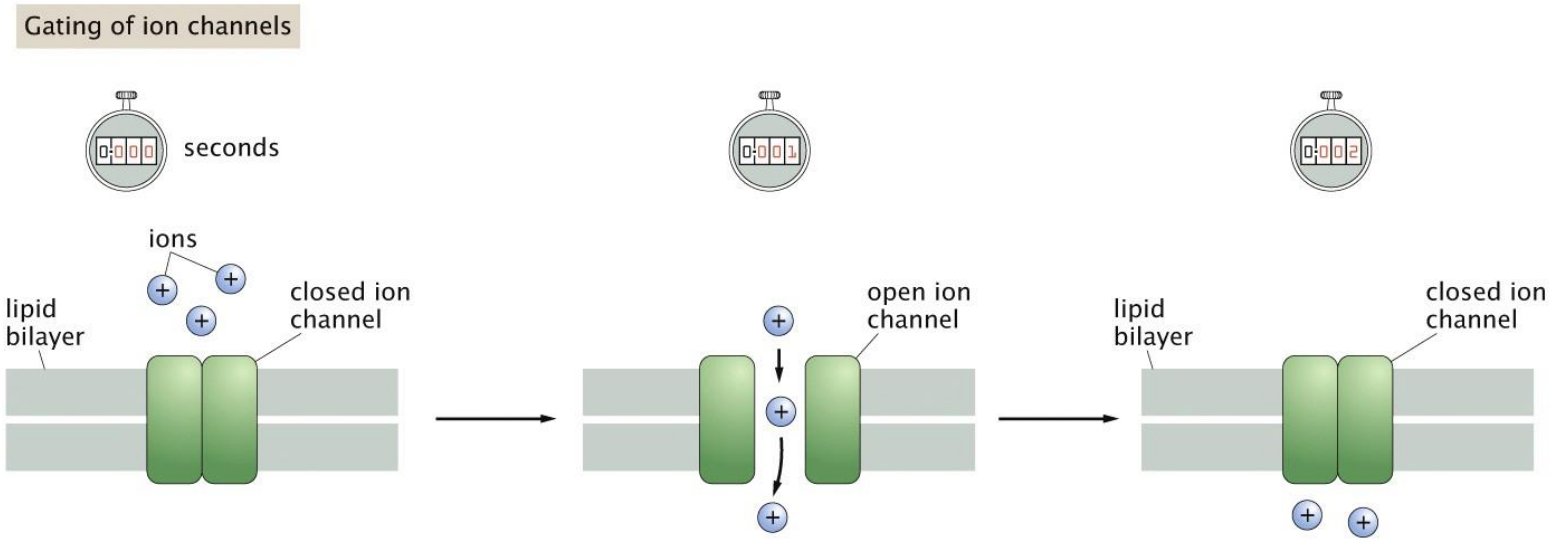


Figure 3.2g Physical Biology of the Cell, 2ed. (© Garland Science 2013)

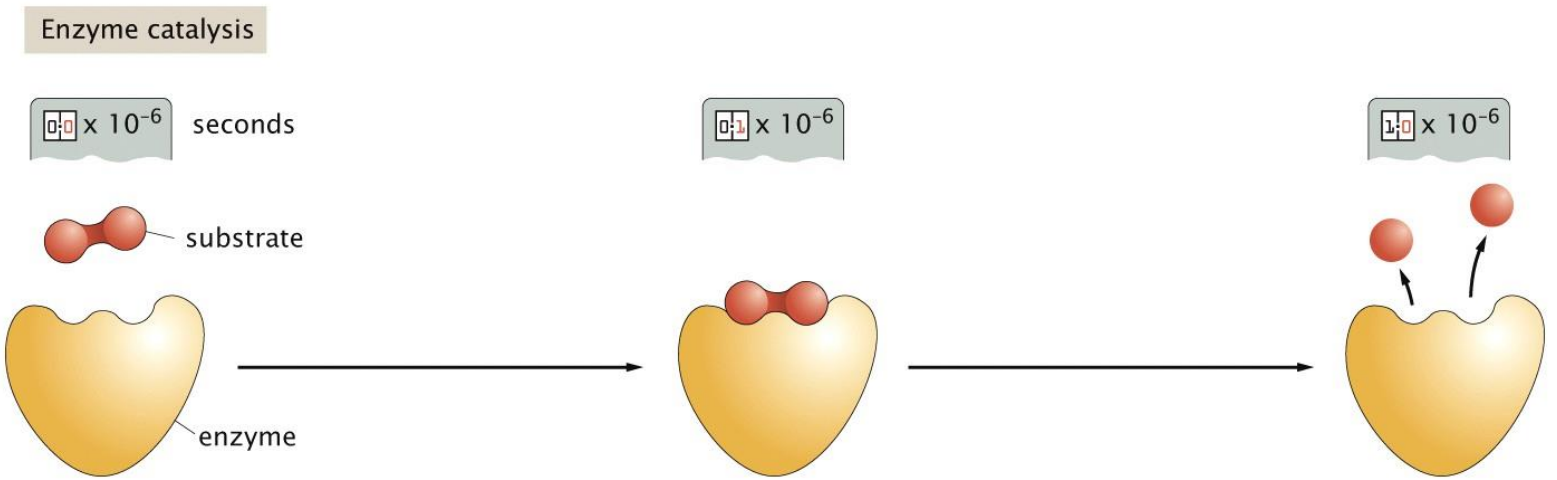


Figure 3.2h Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The lives of molecules: RNA lifetimes

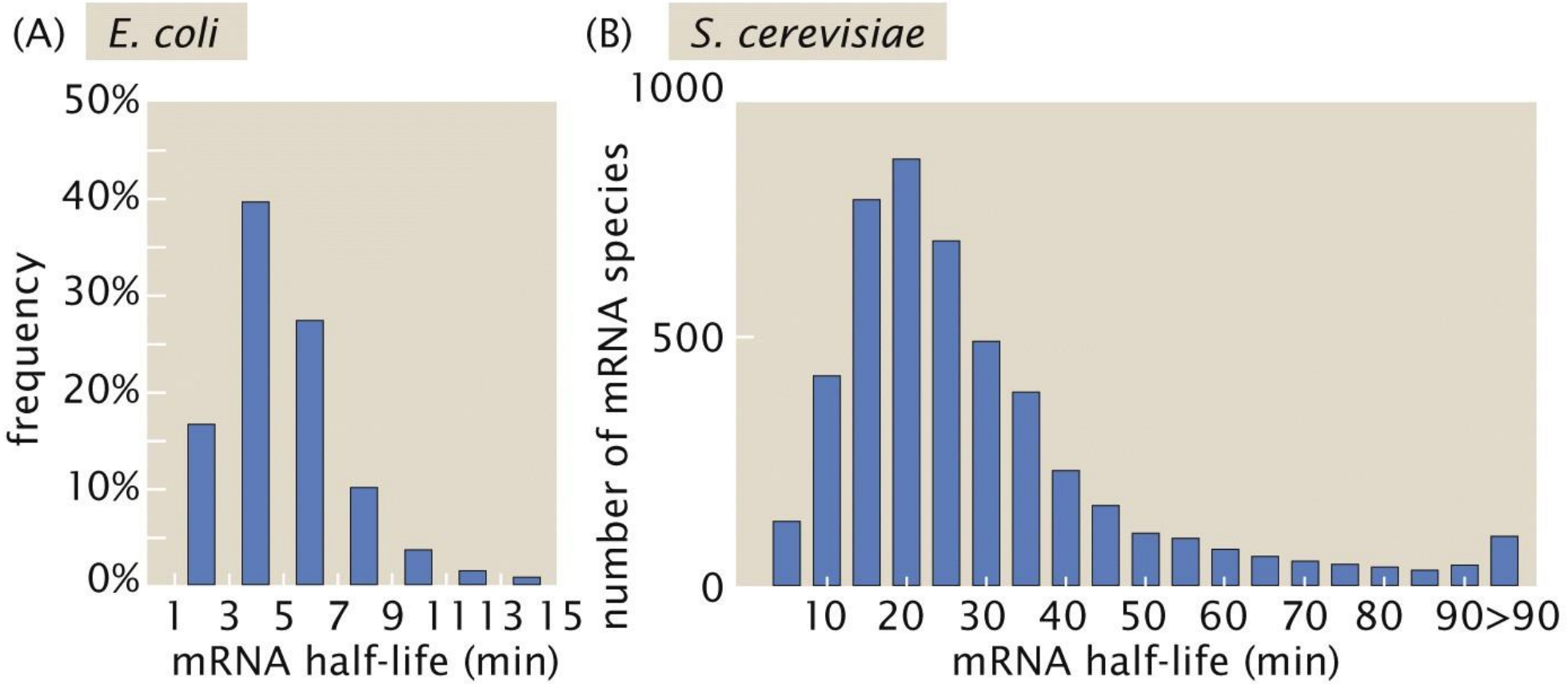


Figure 3.14 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The lives of molecules: Protein lifetimes

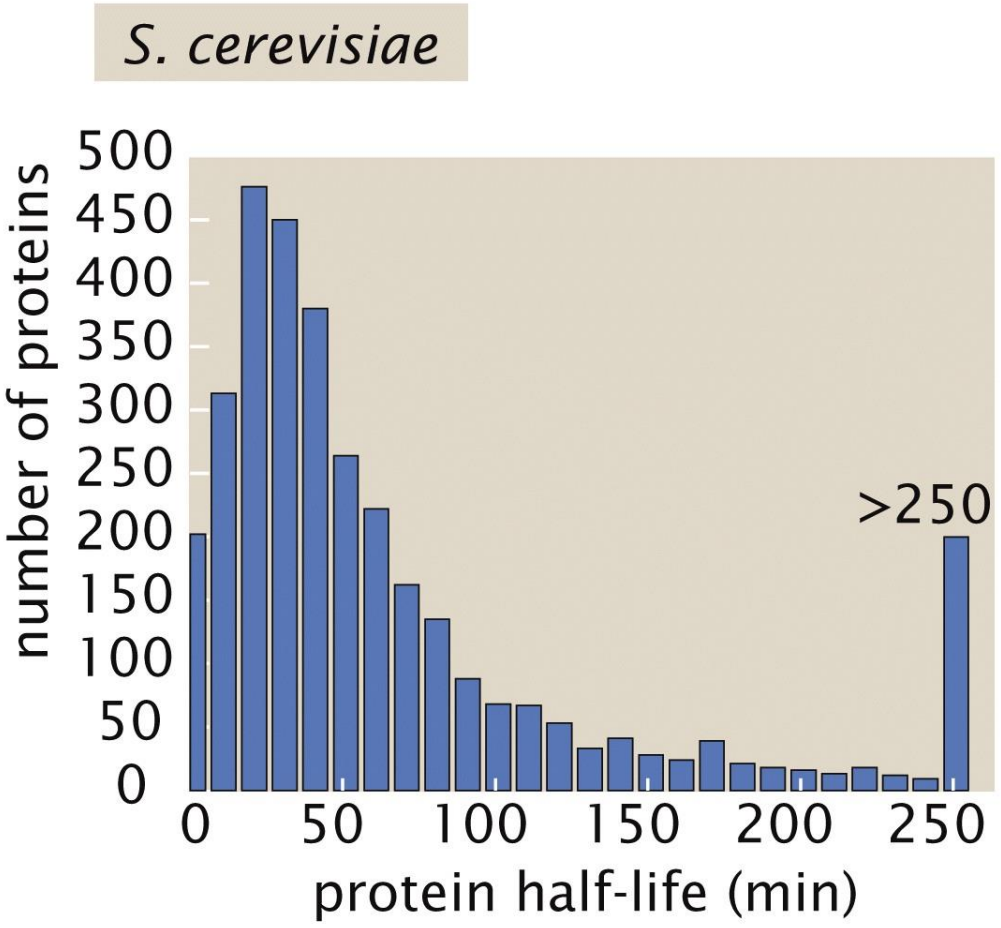


Figure 3.15a Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Timing estimates

How fast is the replication machinery?

- There are 5×10^6 bp in E.coli
- Thus $\text{rate}_{bp} = \frac{5 \times 10^6 \text{ bp}}{3000 \text{ s}} = 2000 \text{ bp/s}$

- In E.coli, 2 DNA polymerases replicate the DNA from 1 origin in opposite directions
- A single DNA polymerase in E coli has a rate of 1000 bp/s

Aside: In higher organisms, DNA polymerase rate ~ 10bp/s and the genomes are much larger

Q: How does the DNA ever get replicated in time?

Timing estimates:

What is the rate of protein synthesis?

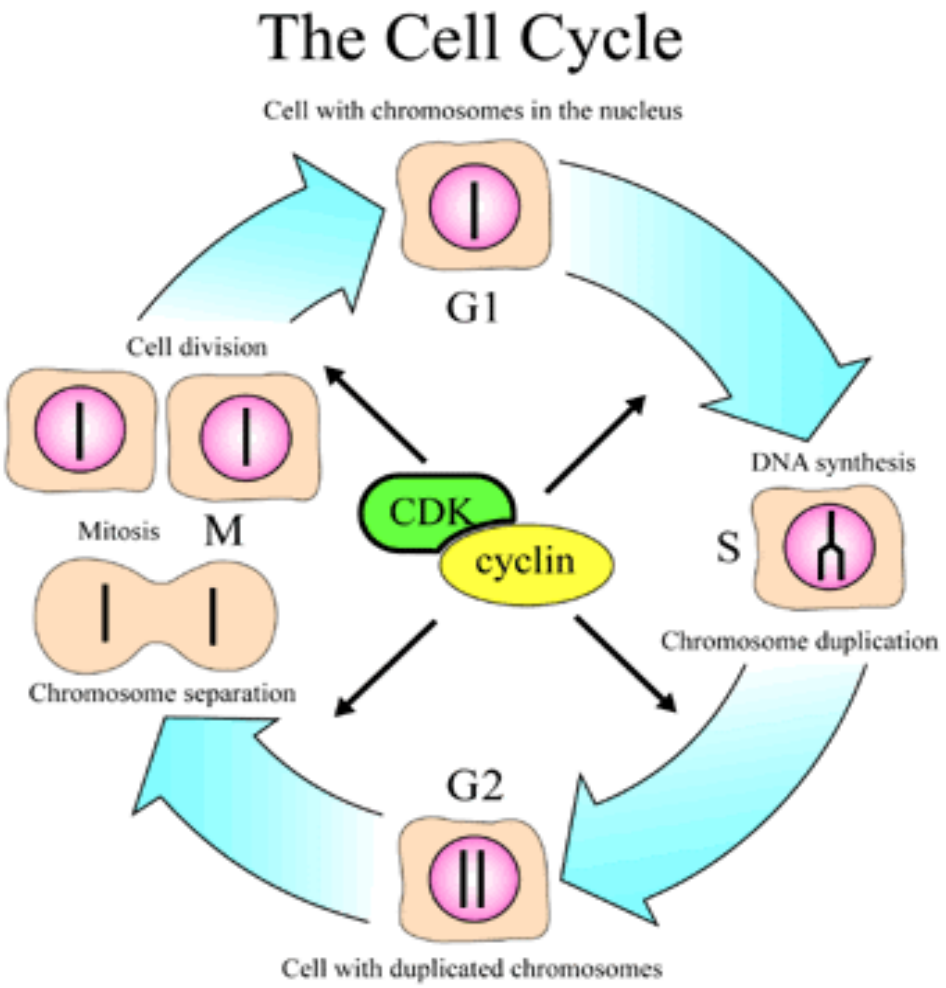
Previously, ~ 3×10^6 proteins in an E. coli cell

$$\text{rate}_{\text{protein}} = \frac{3 \times 10^6}{3000} = 1000 \text{ proteins/s}$$

Q: what are the mechanical properties of these biomachines (DNA polymerase, Ribosomes) that allow them to operate at these rates?

Q: given the # of ribosomes in E. coli, how many proteins per second is ribosome making?

Cell Cycle:



In many biological processes, time is relative.

All that matters is that a task gets completed

So in the cell cycle there are many checkpoints

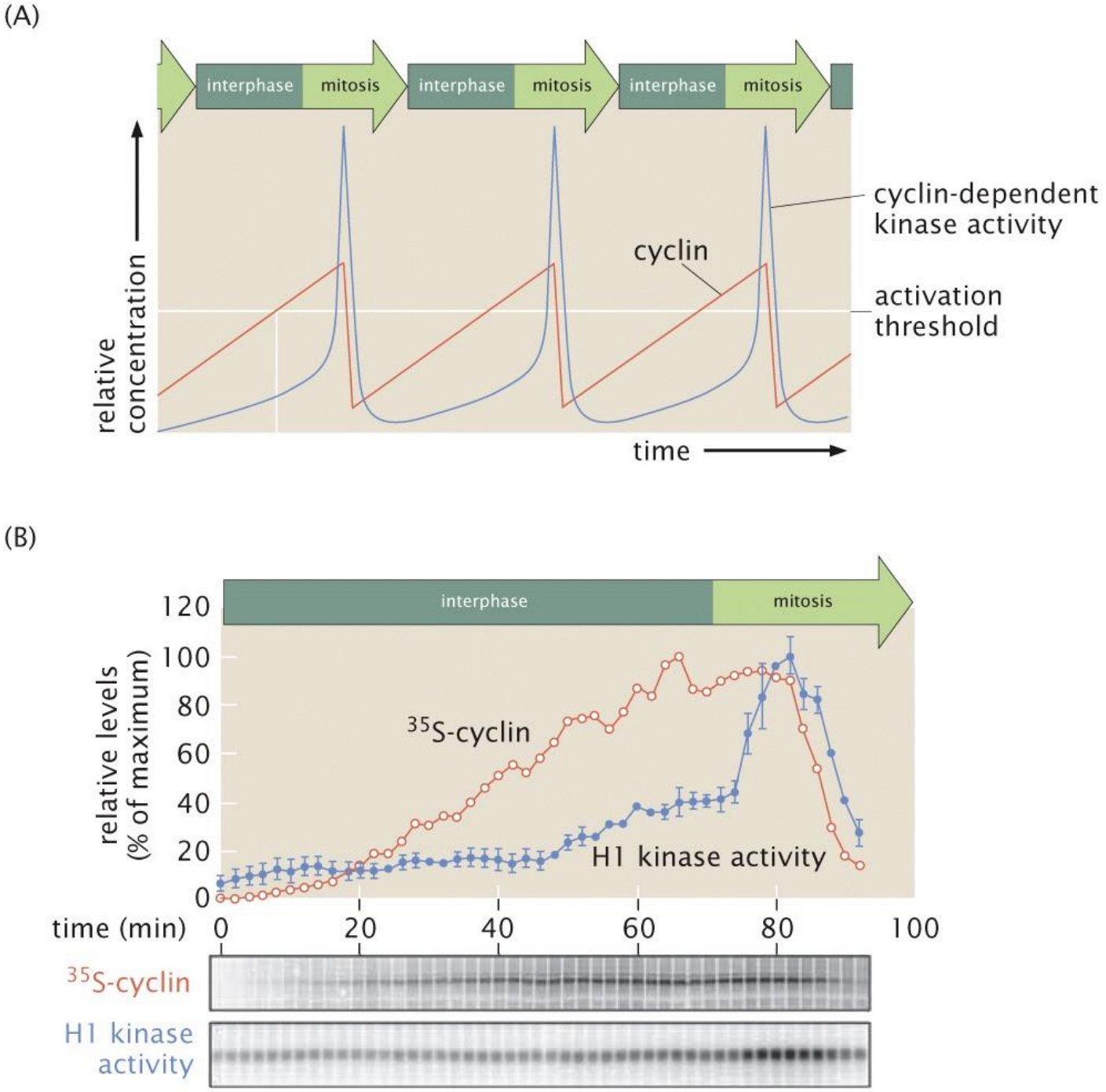
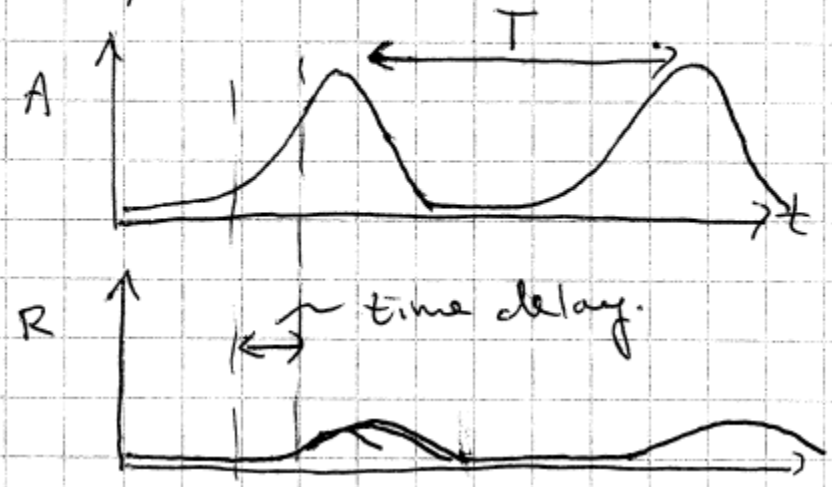


Figure 3.16 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Building an oscillator

General Design of Biological Oscillators

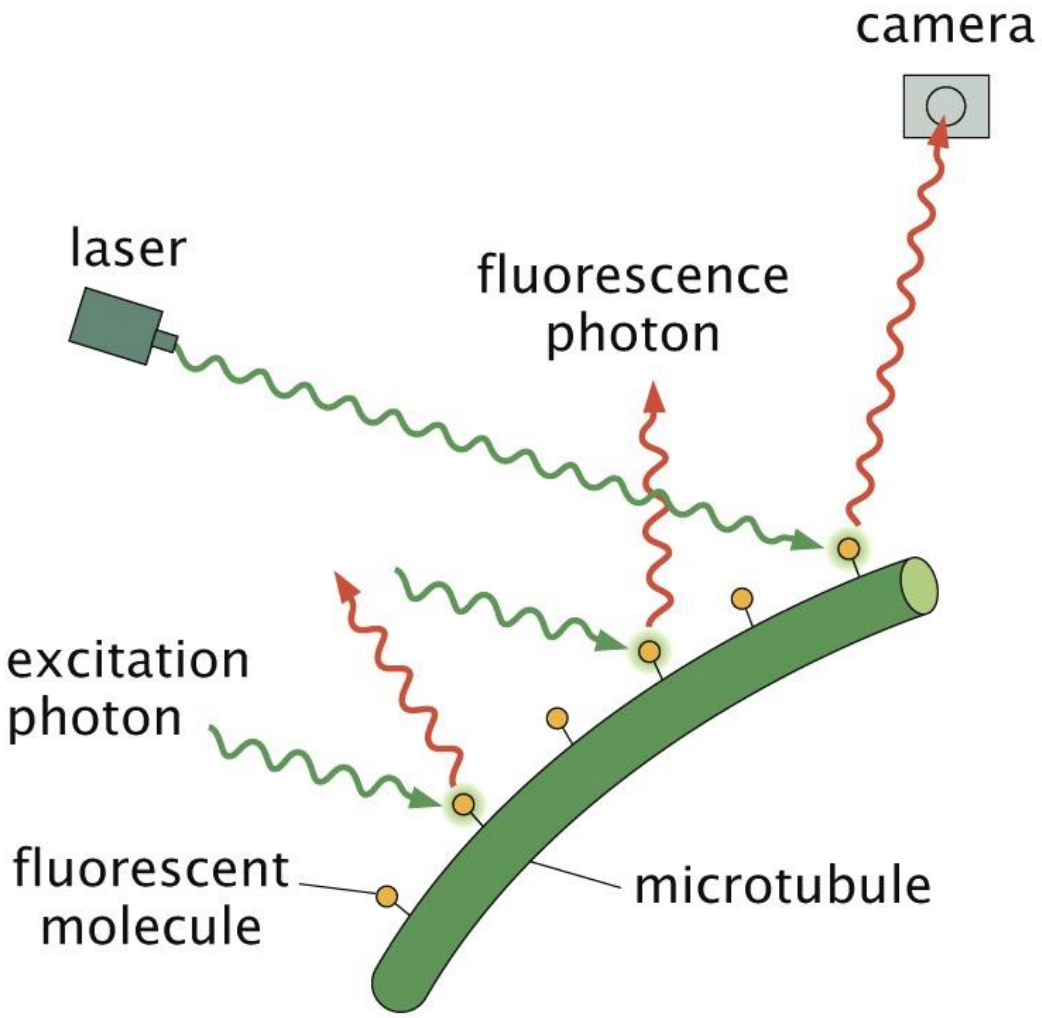


- negative feedback + time delay can produce oscillations.

- The production and degradation of cyclins, depends upon the abundance of cellular resources ⟹ the absolute time of the period, T, will vary depending on the environment.

Biophysical Experimental Techniques:

FLUORESCENCE MICROSCOPY



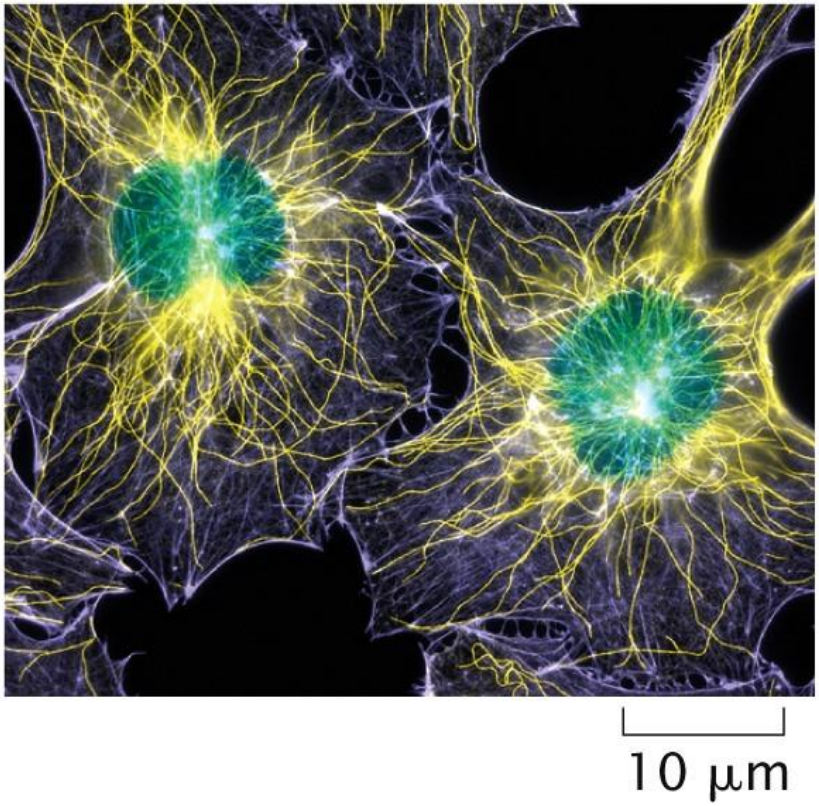


Figure 2.13a Physical Biology of the Cell, 2ed. (© Garland Science 2013)

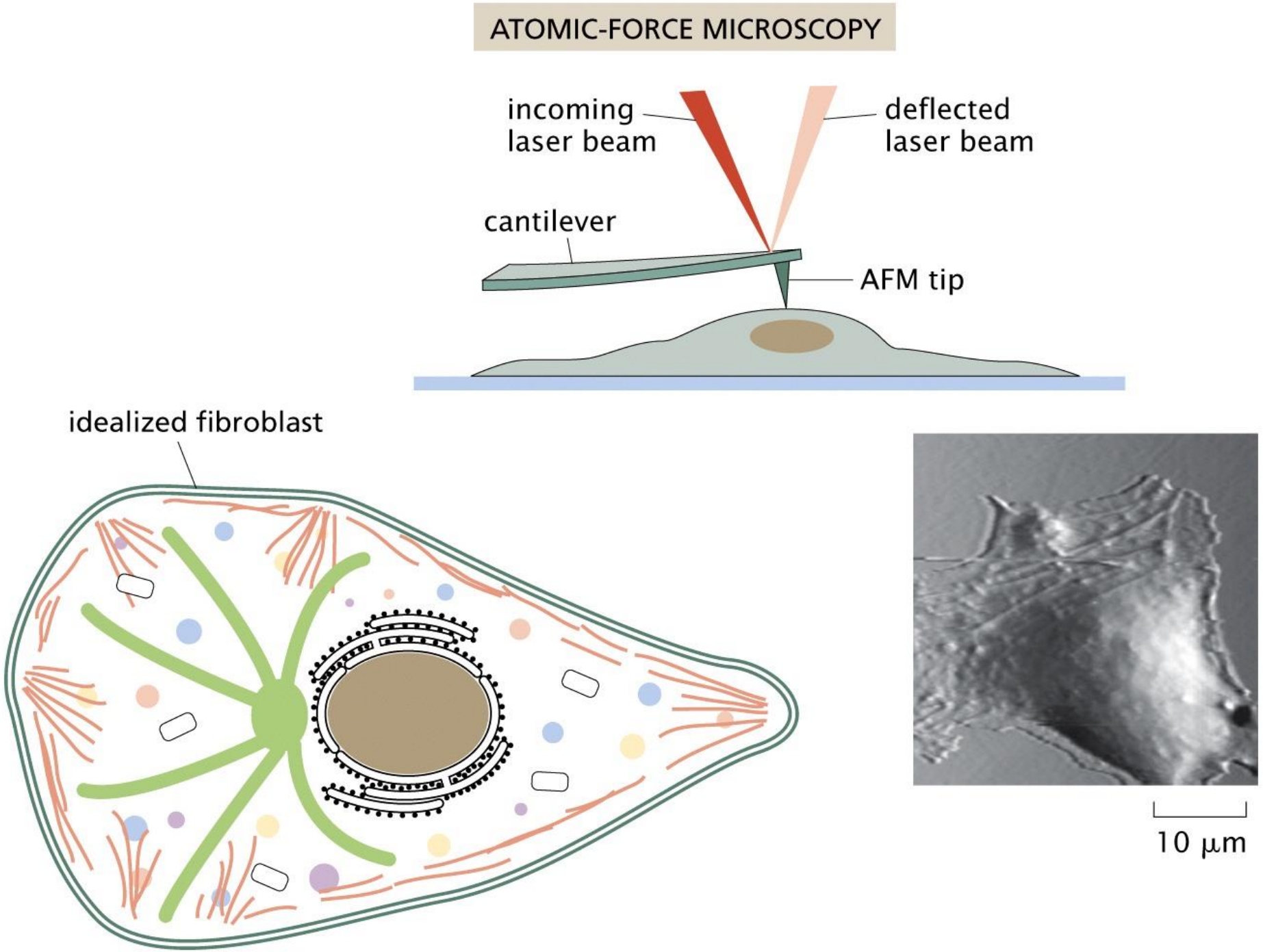


Figure 2.13b Physical Biology of the Cell, 2ed. (© Garland Science 2013)

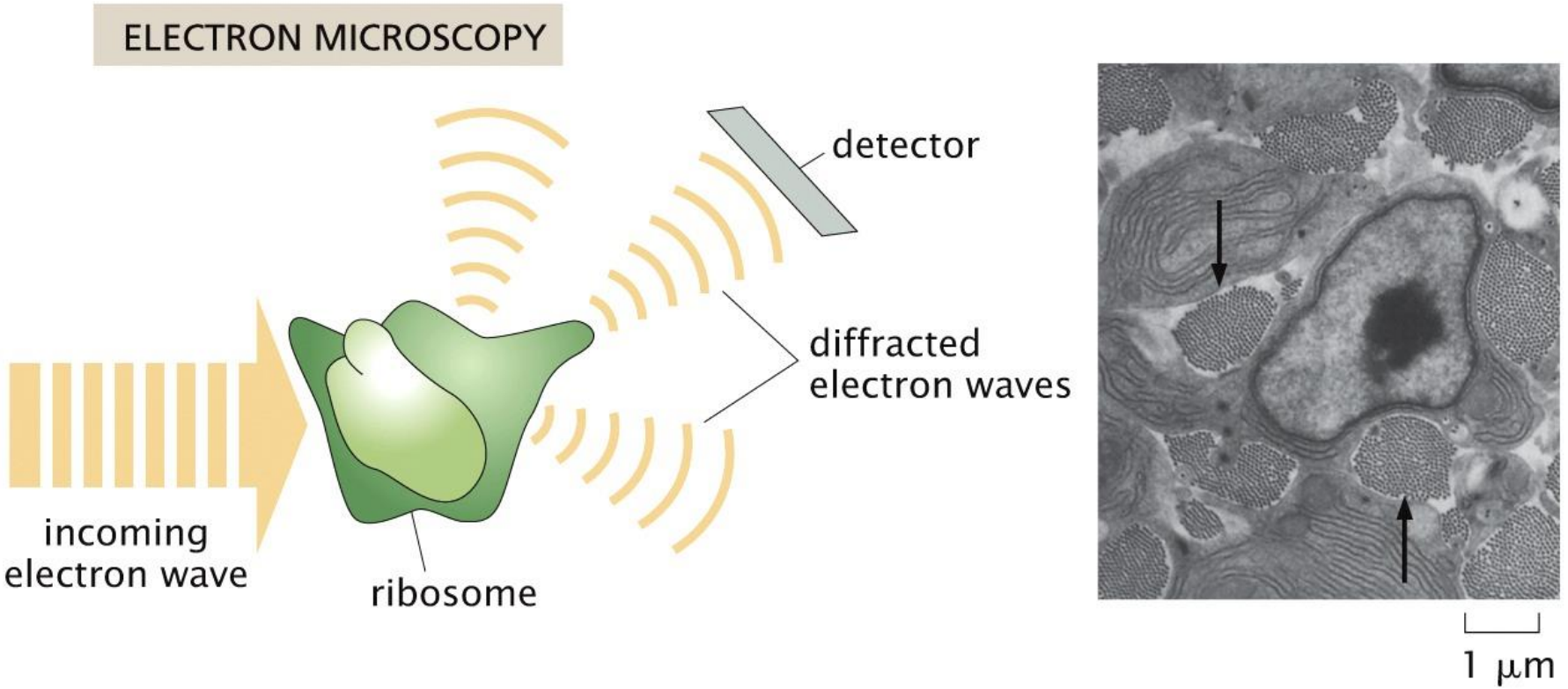


Figure 2.13c Physical Biology of the Cell, 2ed. (© Garland Science 2013)

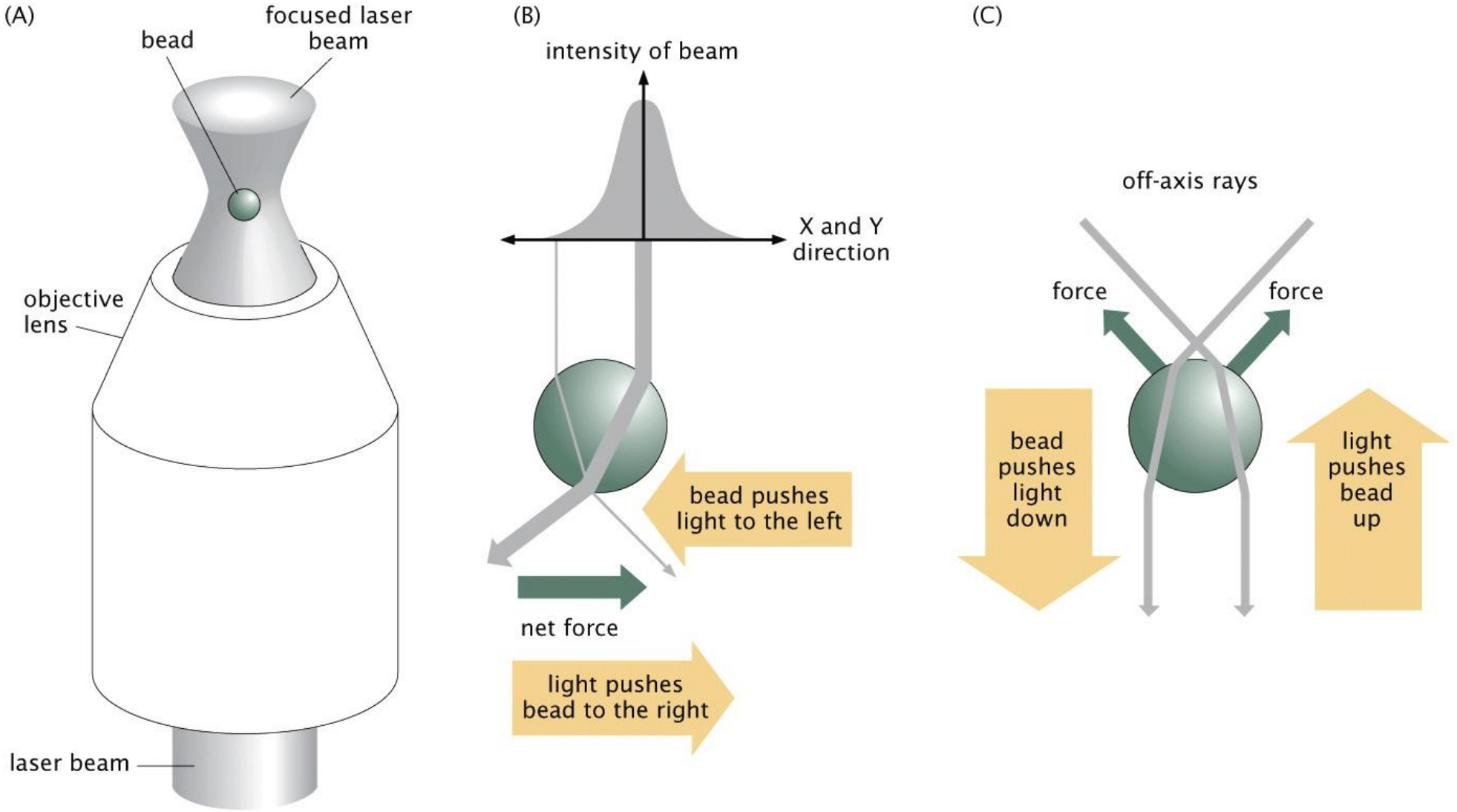


Figure 4.11 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

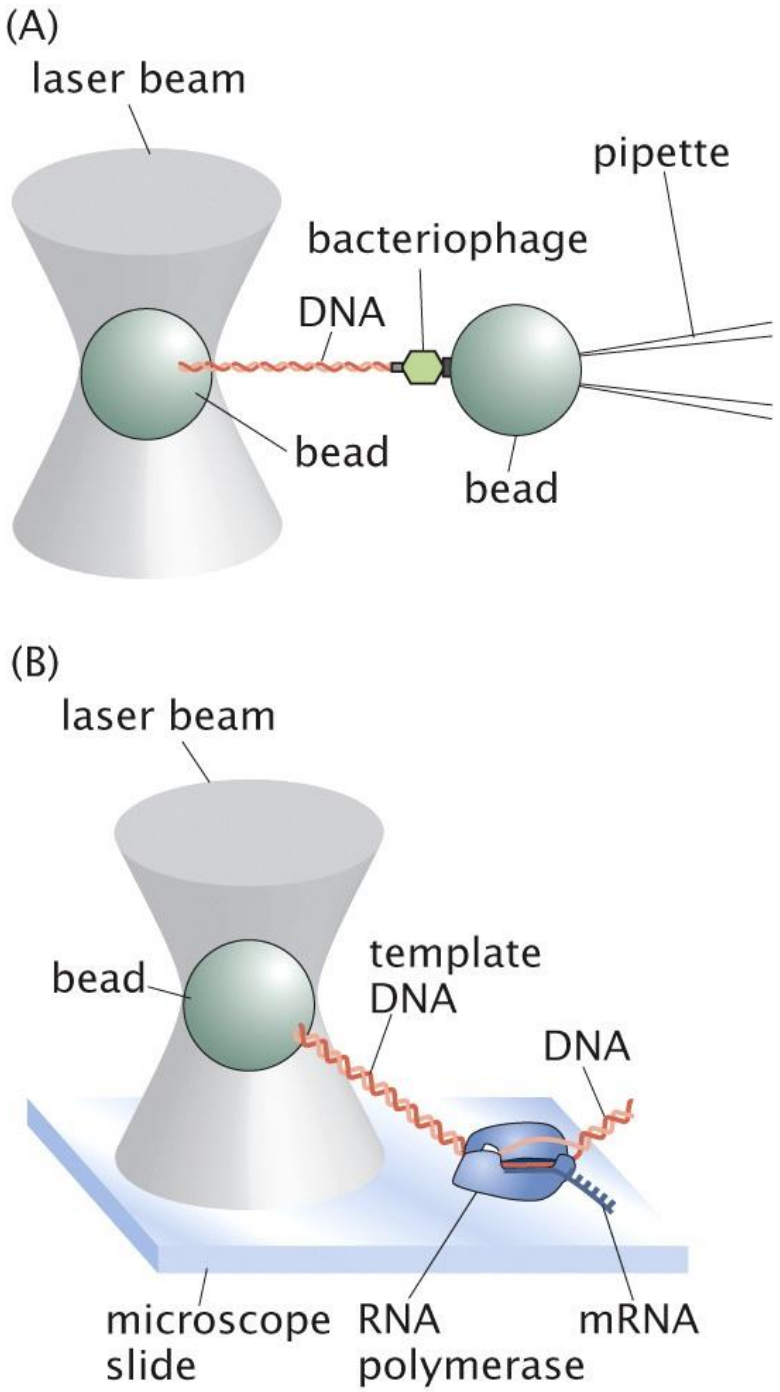
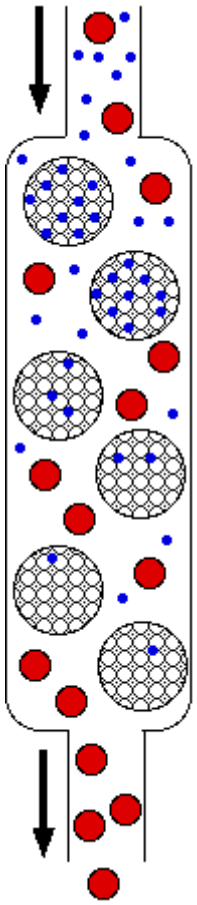


Figure 4.12 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

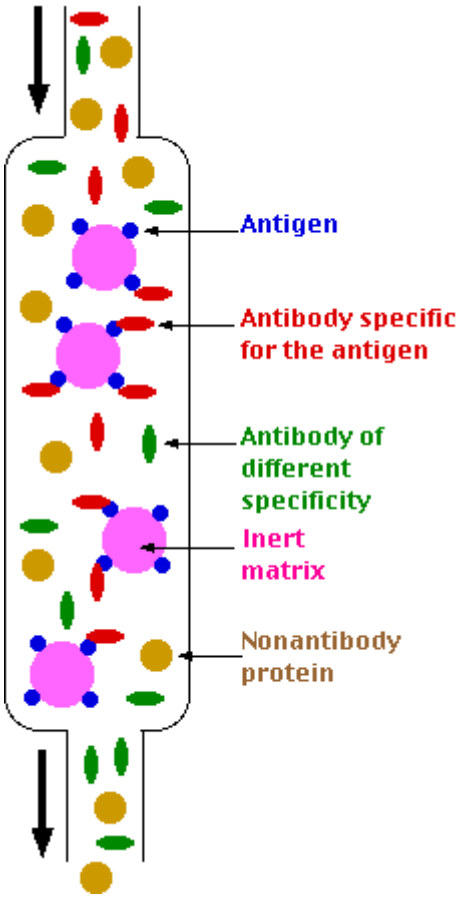
Molecular Biology Experimental Methods:

Chromatography:

- Chromatography is used to separate and purify molecules from a complex mixture



Gel-filtration



Affinity

Types:

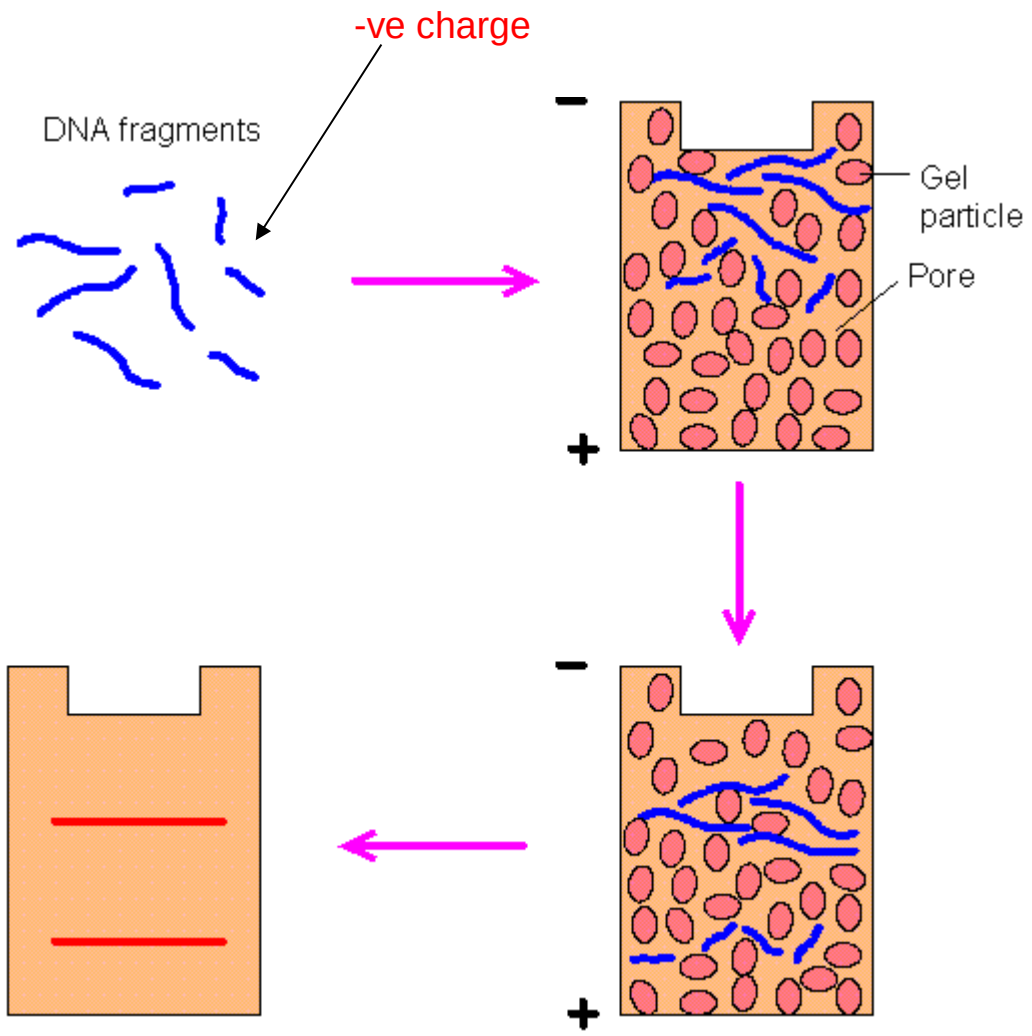
- Charge = ion-exchange chrom.
- Hydrophobicity = hydrophobic chrom.
- Size = gel-filtration chrom.
- Binding = affinity chrom.

Affinity chromatography can be used to elute very pure protein samples

e.g. - DNA for DNA binding proteins
- antibodies for specific protein
- protein binding partners
- engineered tags (HIS, GST)
= Co-immune precipitation = CoIP

Gel Electrophoresis

- Used to determine size of biomolecules
- -vely charged molecules move in applied field through gel



- For proteins, SDS(detergent) is used to denature proteins and give them net negative charge = SDS-PAGE
- 2D SDS-PAGE can be used to sort proteins by both size and native charge
-

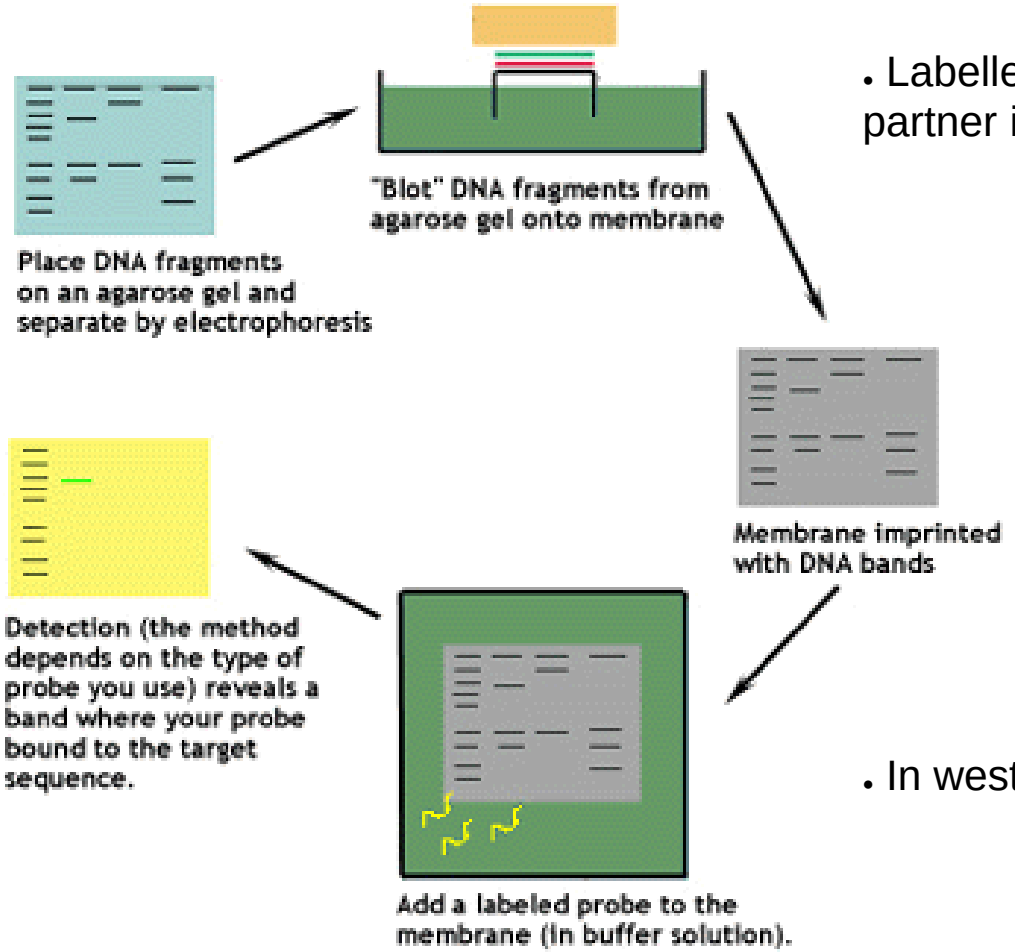
DNA manipulation

- Use restriction enzymes to cut specific nucleotide sequences (e.g. HpaI = GTTAAC; EcoRI = GAATTC)
- Some make blunt ends, some make dangling 'sticky' ends

- Used to chop up large DNA segments (i.e. Chromosome)
- Used to ligate (join) different DNA fragments -> genetic engineering

Northerns, Southerns and Westerns

- Used to detect quantitative amounts of specific DNA (southern), RNA (northern) or proteins (Western) from a complex mixture (i.e. cellular extract).

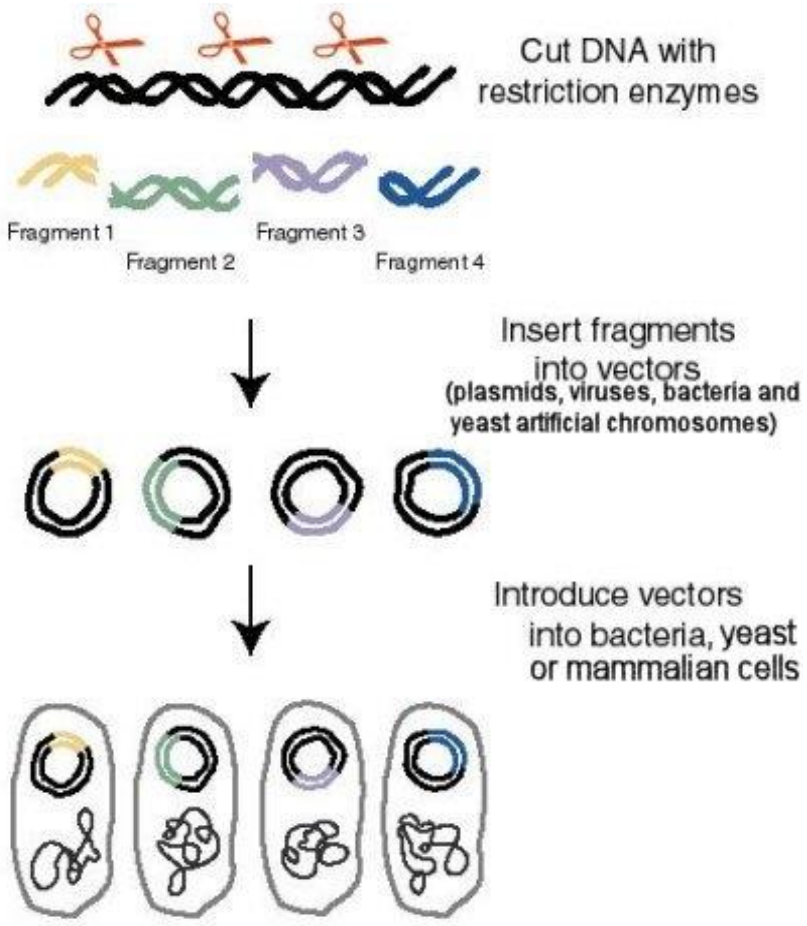


- Labelled DNA probe 'hybridizes' with matching partner in sample

- In westerns, a labelled antibody is used for probe

DNA Cloning

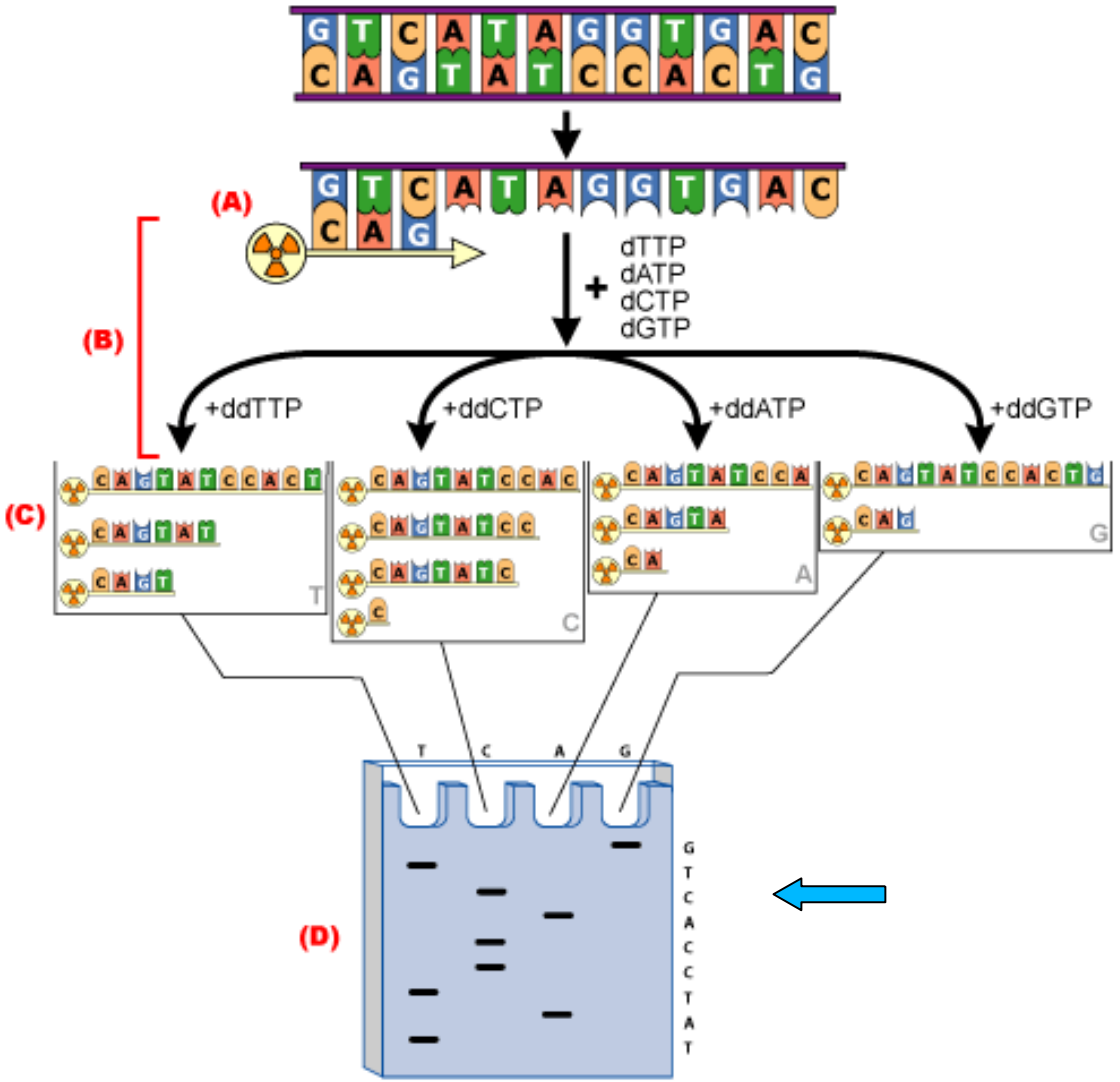
- Replicate specific DNA fragment in large quantities using bacteria
- Fragment is inserted into circular plasmid DNA and transfected into bacteria
- Large collection of bacteria colonies each with different fragmet = library



- Types of libraries = genomic, cDNA
- cDNA library comes only from mRNA
- Use cDNA plasmid to express lots of specific protein = expression vector

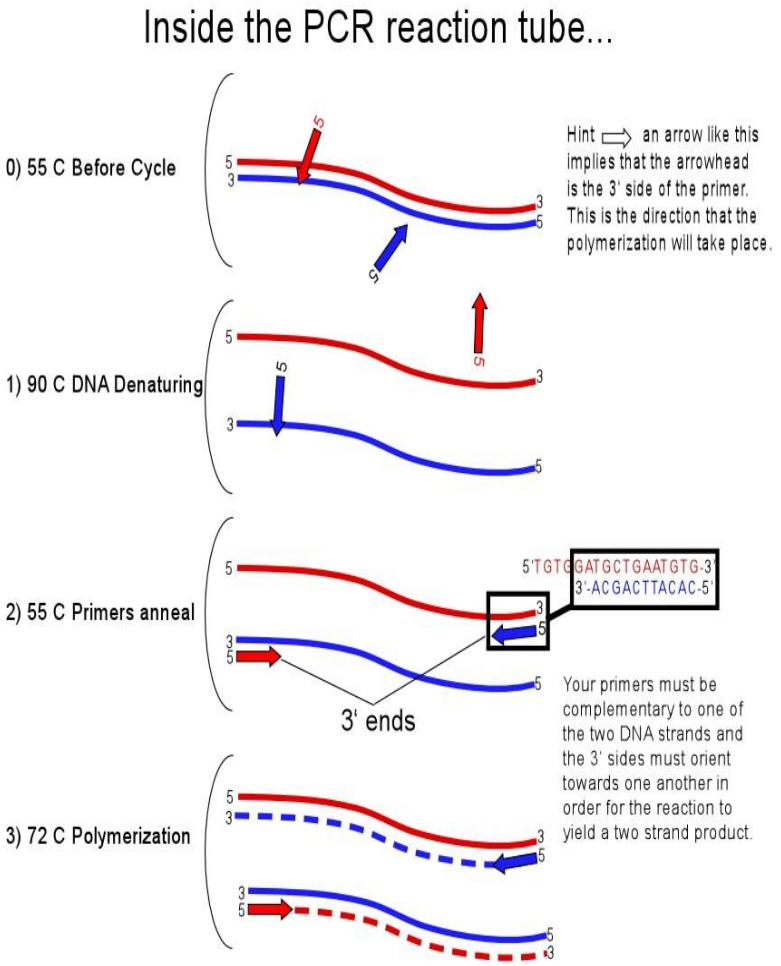
DNA Sequencing

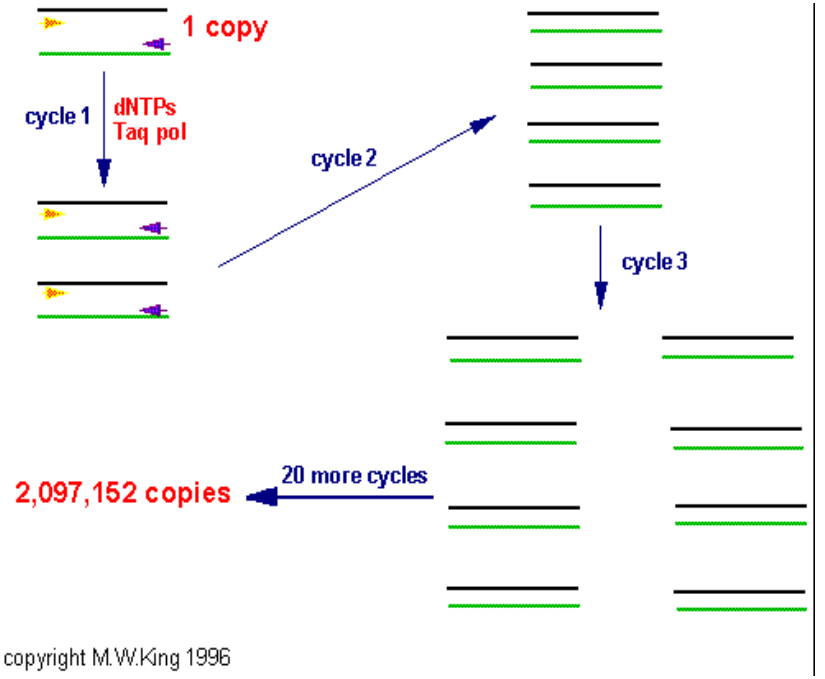
Dideoxy method: use special nucleotides that stop growth of DNA
--> produces DNA that stops at specific letter positions



Polymerase Chain Reaction (PCR):

- Use 2 DNA primers to pull out and amplify specific region of DNA in a sample

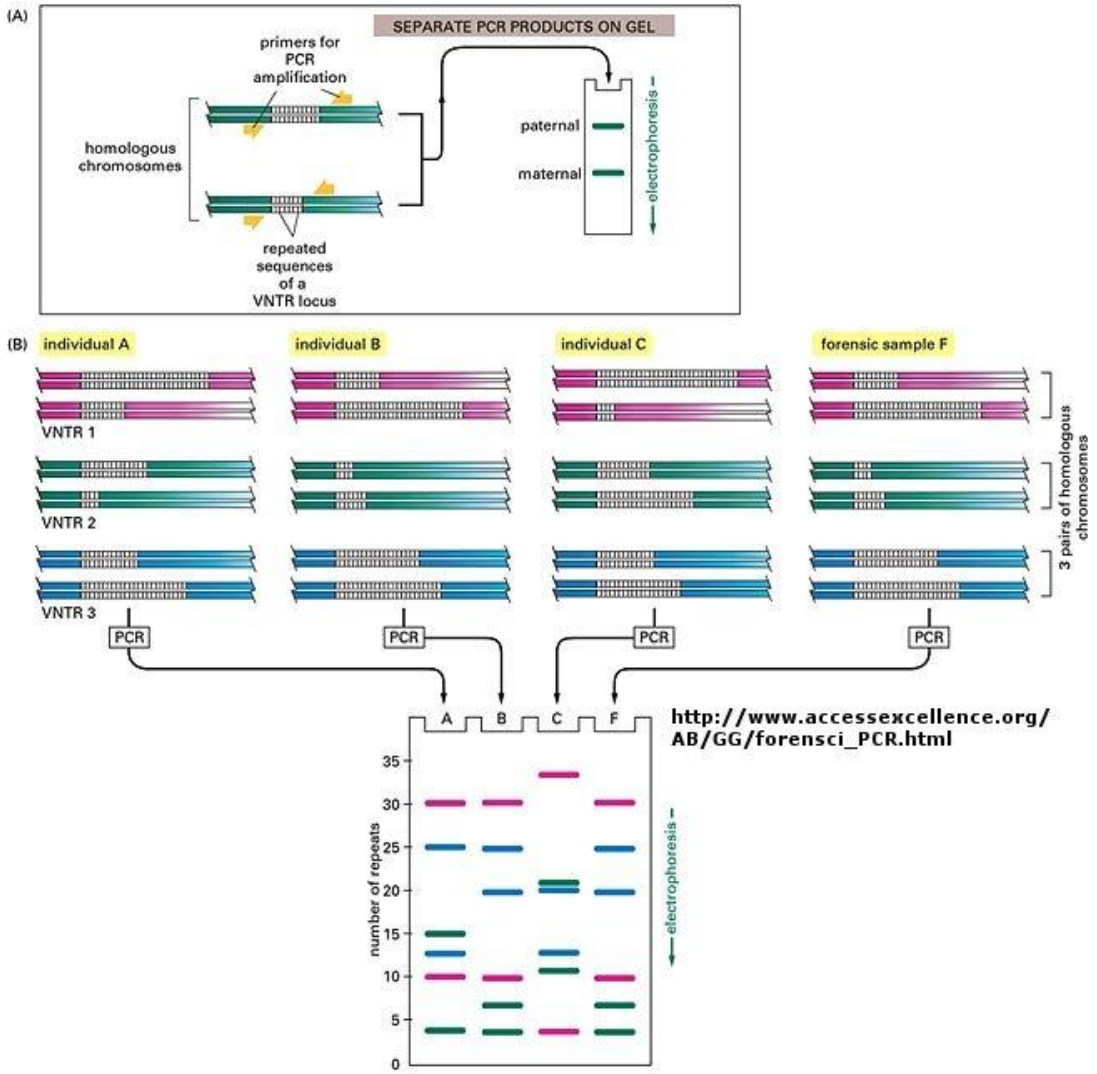




PCR and Forensics

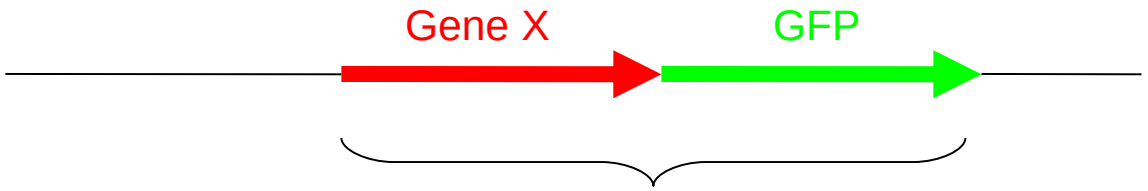
- Our chromosomes contain regions of 'satelite' DNA that vary in length – each individual has specific length = fingerprint

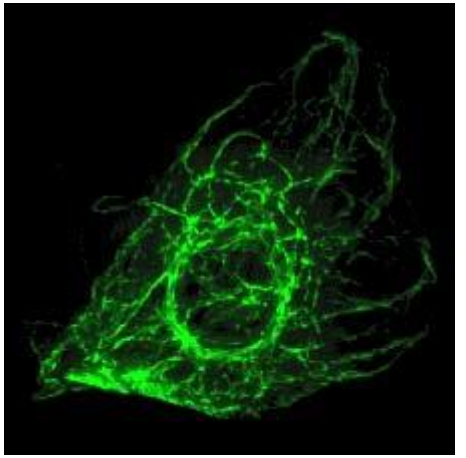
- PCR amplify regions and then use gels to measure the lengths -> yields fingerprint

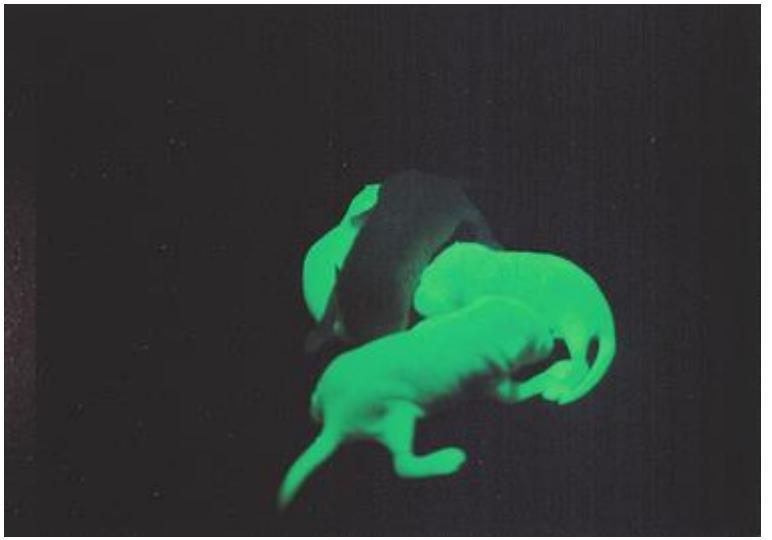


Imaging the action

- Reporter
- Green flourescent proteint GFP lacZ, labelled antibodies

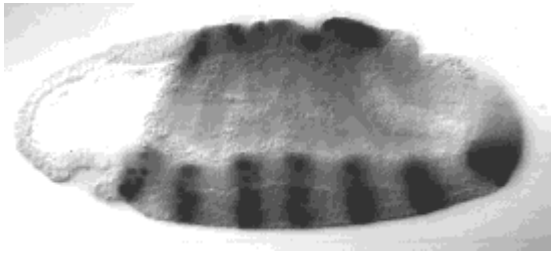


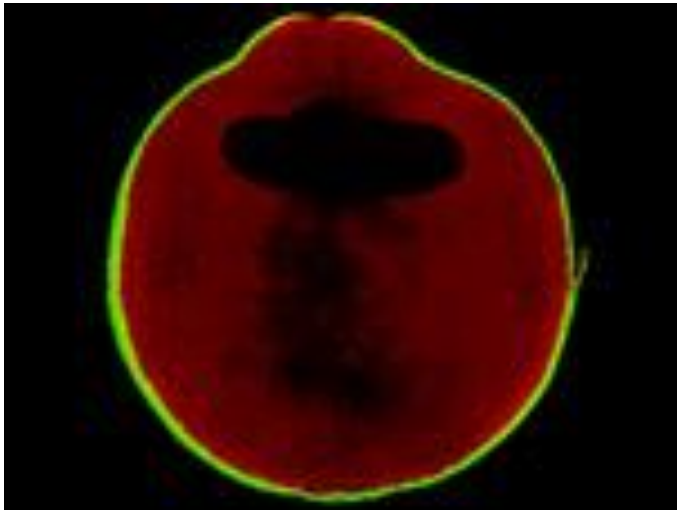


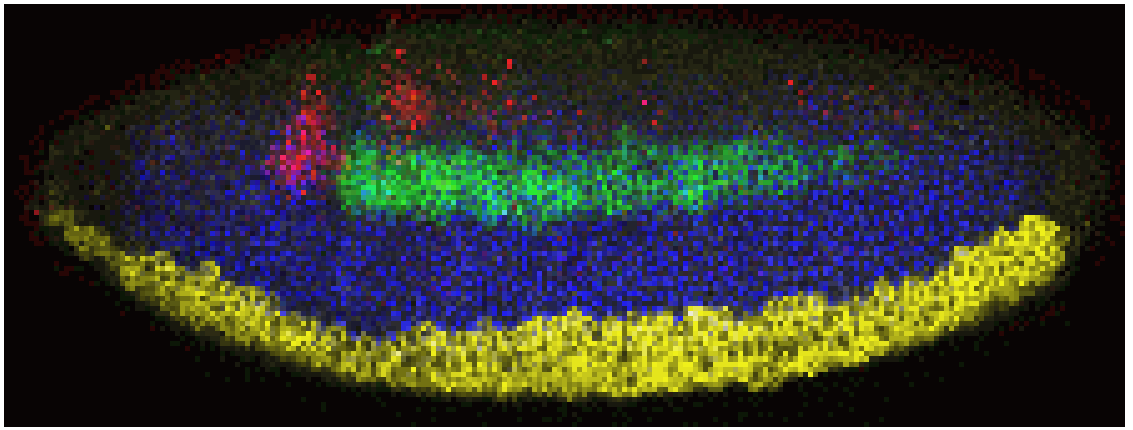


Imaging II

- In-situ hybridization measure the presence of mRNA of specific gene
- takes a snapshot of gene pattern – use flourescently labelled RNA probe
- Use confocal microscope to image different 2D layers build 3D image

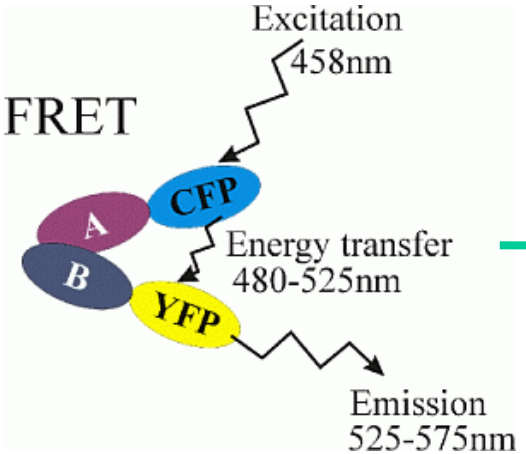






Detecting biomolecular interactions

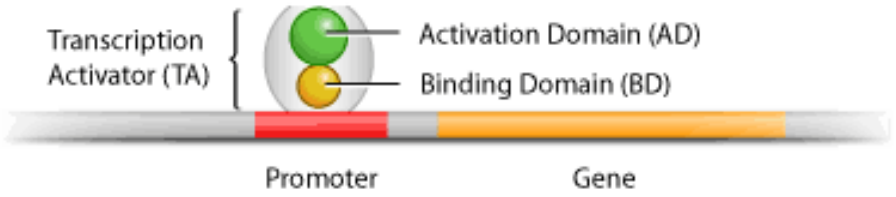
FRET (Flouresence Resonance Energy Transfer)



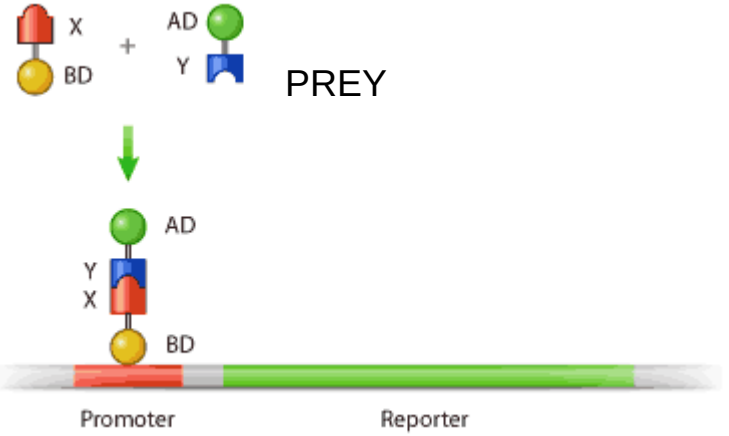
NO interaction

A & B interact

Yeast Two Hybrid



BAIT



Interactions II:

Phage Display:

- Phage = virus that infects bacteria. Has an external coat of protein
- Can fuse foreign proteins to coat proteins of phage
- Used to screen libraries of proteins for specific interaction drug design