Living matter theory

Physics Department, UNIST 2016

Tsvi Tlusty, tsvi@unist.ac.kr

Aim: **Examine concepts and questions of living matter within a physical framework.**

Outline: candidate principles of living matter

1. **What is living matter?** basic processes, molecules, systems and scales.
2. **Information theory** and its application to biology: noise and representation.
3. **Physics of DNA and proteins** (molecular recognition): the physical basis.
4. **Networks in biology** (biochemical, genetic, neural): emergence, collective.
5. **Evolution theory:** memory.

Bibliography will be given for each lecture. There are no comprehensive textbooks.

Two useful introductory texts are: 1. **Physical biology of the cell** (Philips et al.)

2. **Biophysics** – search for principles (Bialek)

Course philosophy and procedure

- **WHY and WHEN?** Now, because biological data becomes as quantitative as in physics High demands for living matter theory (beyond narrative stage)

- **HOW?** Open ended evolution, according to feedback and knowledge. Open discussion – please ask/comment.

- **WHO?** No structured theory and canonic textbook.

- **WHERE?** Here in UNIST we have a unique research center, IBS CSLM, which focuses on living and soft matter.

Lecture 1:

Life and the stuff it is made from

Sources:

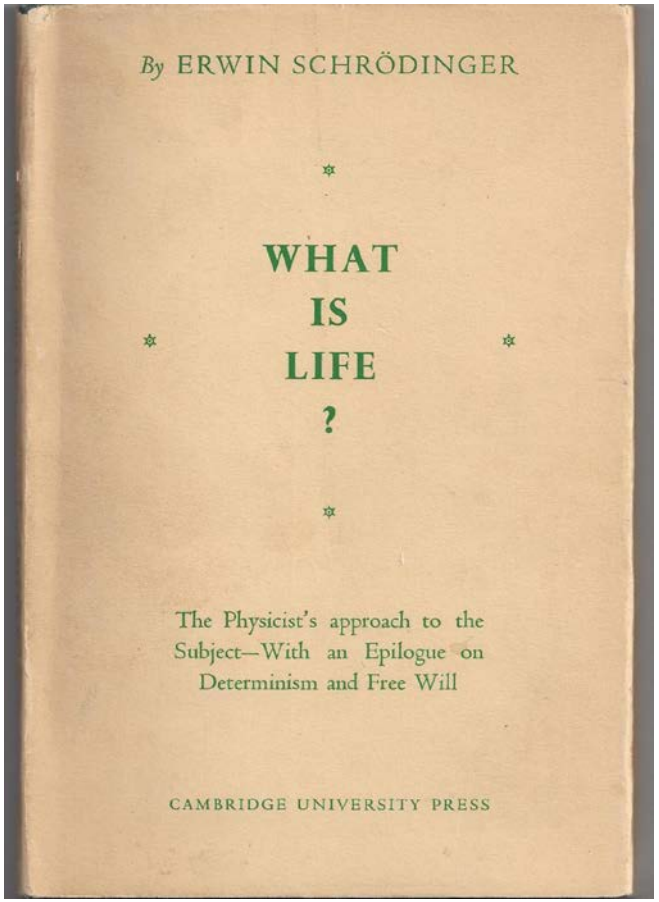
- Physical Biology of the cell (Philips et. al)
- Biology by the numbers (Philips and Milo).
- A few others

Tsvi Tlusty, tsvi@unist.ac.kr

What is Life?

- We know it when we see it,
  at least that's what we think.

- Open question that many philosophers and scientists have considered.



Schrodinger 1944: "The large and important and very much discussed question is: How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?

The preliminary answer can be summarized as follows: The obvious inability of present-day physics and chemistry to account for such events is no reason at all for doubting that they can be accounted for by those sciences."

Can life be defined?

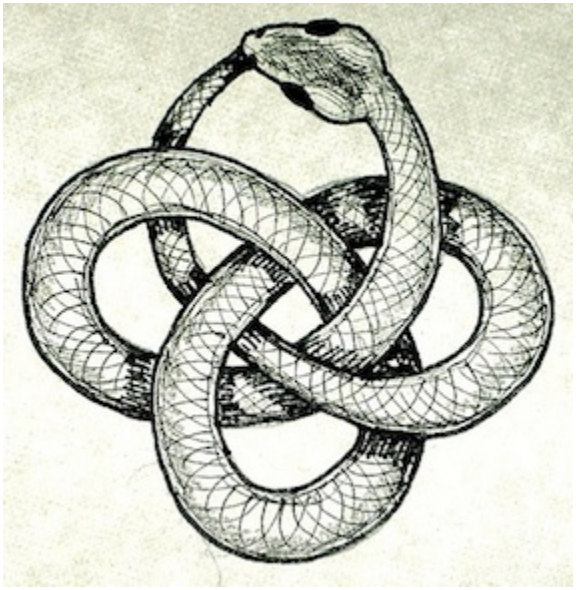
Webster dictionary

LIFE *1a* : the quality that distinguishes a vital and functional being from a dead body.

***b* :** a principle or force that is considered to underlie the distinctive quality of animate beings

***c* :** an organismic state characterized by capacity for metabolism, growth, reaction to stimuli, and reproduction

***2a* :** the sequence of physical and mental experiences that make up the existence of an individual



The ouroboros or oroboros. Ancient Greek: "tail-devouring snake“ is an ancient symbol depicting a serpent or dragon eating its own tail. The ouroboros often symbolizes self-reflexivity, introspection, or cyclicality.

Wikipedia

LIFE is a characteristic distinguishing physical entities having biological processes, such as signaling and self-sustaining processes, from those that do not.

More on definitional loops in *Levari et al.* PRX 2012.

Plan B: define life by its symptoms

- Descriptive definition: A living thing has all or most of the following traits:

- **Homeostasis**: regulation (pH, temperature)
- **Organization**: composed of one or more cells.
- **Metabolism**: chemicals and energy into cellular components (anabolism) and decomposing organic matter (catabolism).
  Living things require energy to live.
- **Growth and development**.
- **Adaptation**: response to environment. This ability is fundamental to the process of evolution.
- **Response to stimuli**: senses of multicellular organisms. Motions toward the sun (phototropism), or nutrient (chemotaxis).
- **Reproduction**: the ability to produce new individual organisms.

The stuff life is made from

- Biological macromolecules are specific to living organisms.

- Small set of elements:

carbon (C), oxygen (O), nitrogen (N), hydrogen (H), phosphorous (P)

- **Four main classes**:

(A) Nucleic acids (DNA): store and process information…

(B) Protein (hemoglobin): catalysis structure, signaling…

(C) Lipids (phosphatidylcholine): membrane, organelles…

(D) Carbohydrates (polysaccharide): energy storage, surface, cell wall…

- Huge diversity in each class.

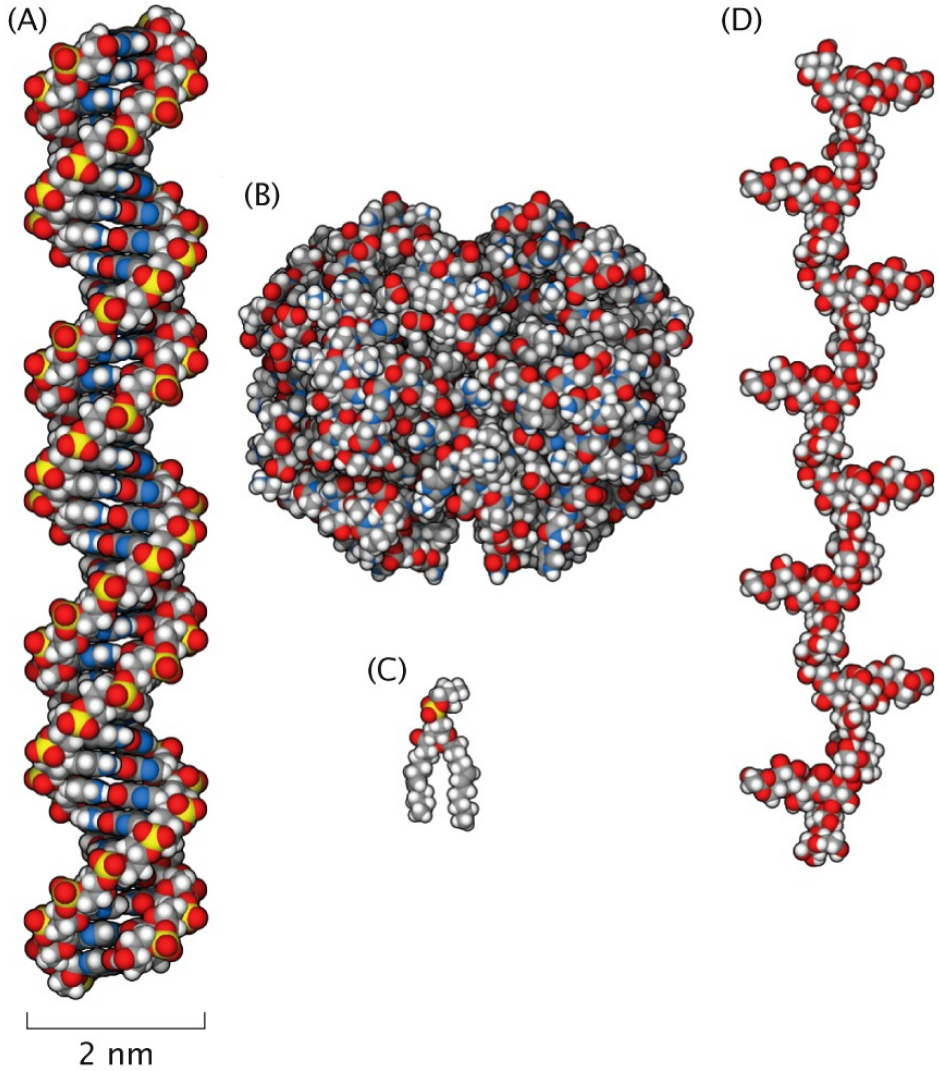


Figure 1.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Why life molecules are so “simple”?

- Only four classes of molecules?
- Why these classes and no others?

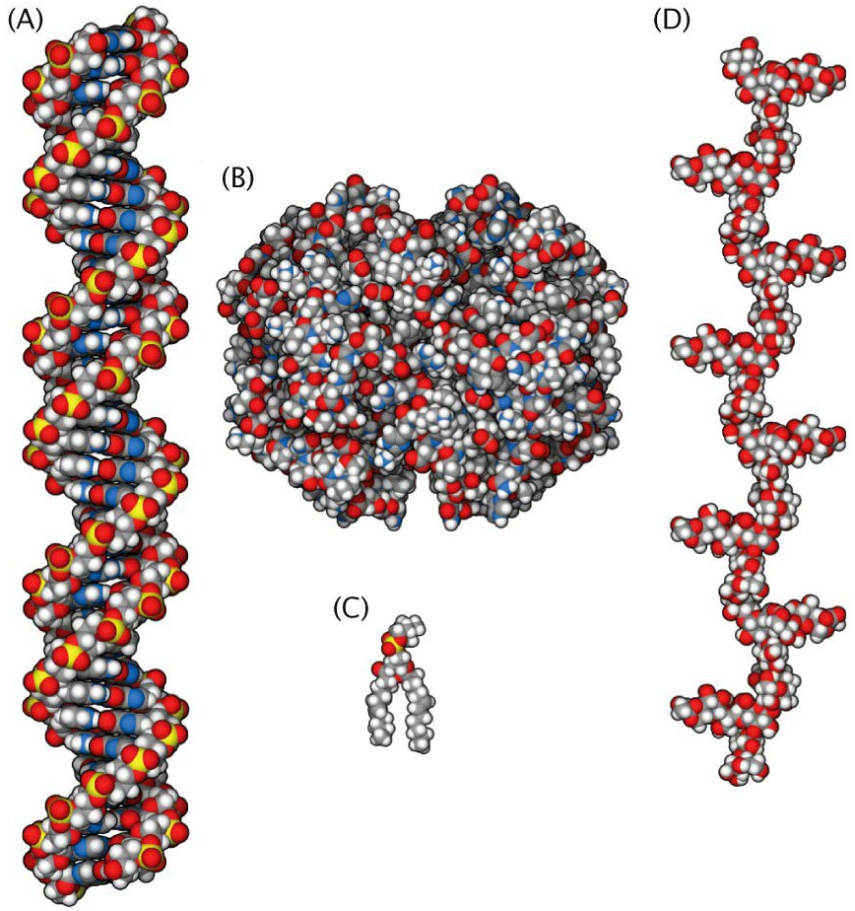
- Origin of Life –
  many theories, none testable (yet?).

Advantages:

- Small number of basic elements and precursors.

- Modular combinatorial chemistry.

- Universal to all organisms that can eat\decompose each other.



Nucleic Acids and Proteins: Two Languages

- Small alphabets:
  - 4 nucleic bases
  - 20 amino acids.
- Translation:
  - 3-letter words (codons)
  - Each encodes 1 AA.
- Hierarchical languages: modular subunits.

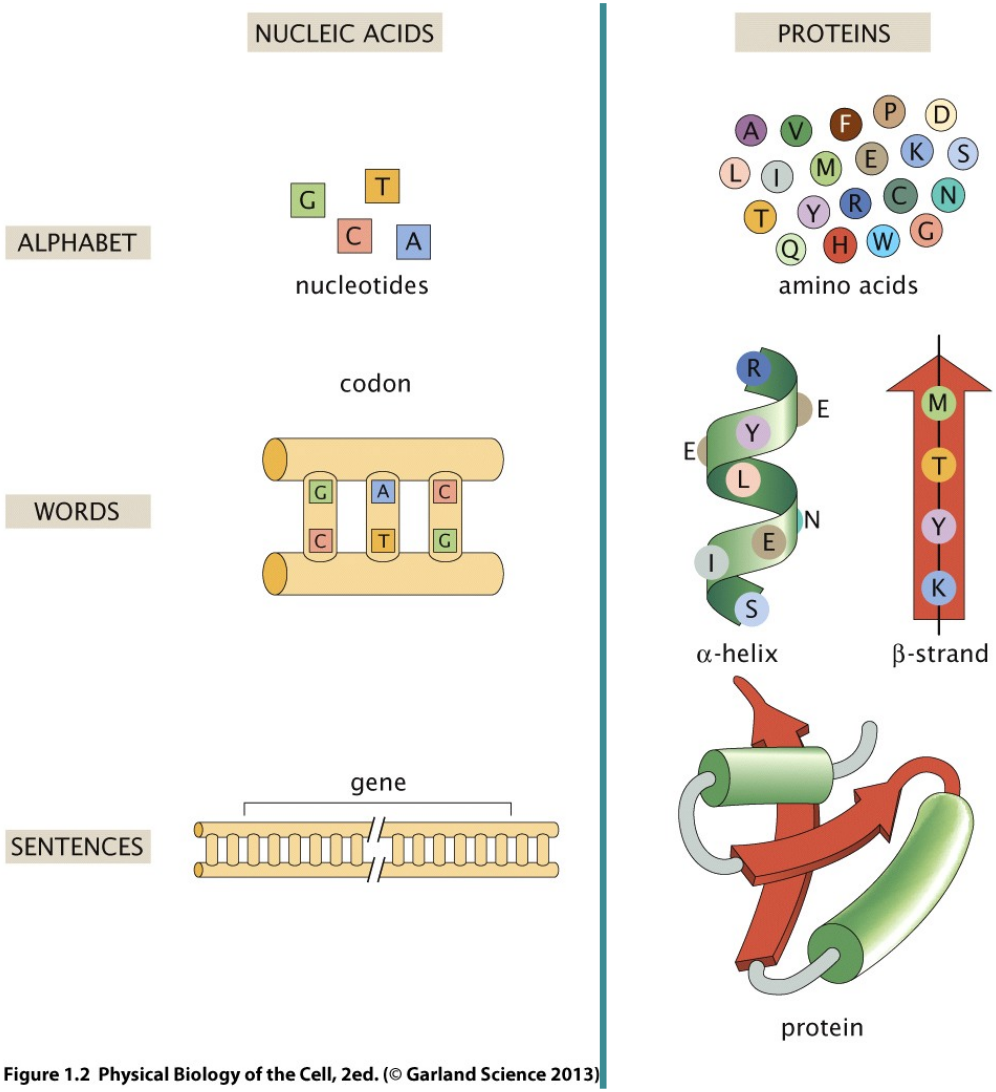


Figure 1.2 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1929).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1955).

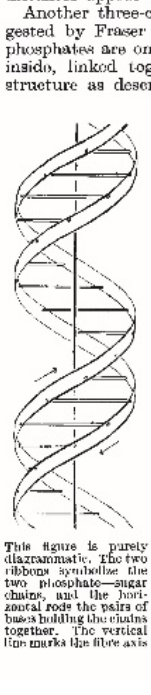
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolise the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

WHILE the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleo-

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[5] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of J_n, the nth order Bessel function. A straight line may be drawn approximately through

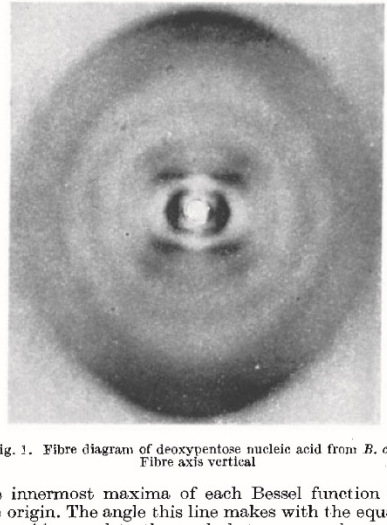


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. c* Fibre axis vertical

the innermost maxima of each Bessel function the origin. The angle this line makes with the equ is roughly equal to the angle between an elemen the helix and the helix axis. If a unit repeats n ti along the helix there will be a meridional reflex (J_n^2) on the nth layer line. The helical configura produces side-bands on this fundamental freque the effect[5] being to reproduce the intensity distribu about the origin around the new origin, on the layer line, corresponding to C in Fig. 2.

We will now briefly analyse in physical terms s of the effects of the shape and size of the repeat or nucleotide on the diffraction pattern. First, if nucleotide consists of a unit having circular symm about an axis parallel to the helix axis, the wh

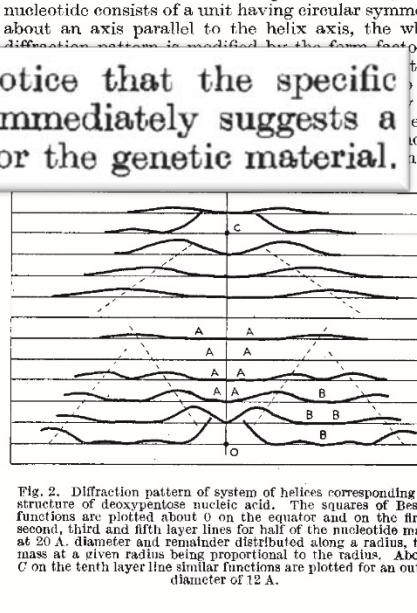


Fig. 2. Diffraction pattern of system of helices corresponding structure of deoxypentose nucleic acid. The squares of Bes functions are plotted about 0 on the equator and on the fir second, third and fifth layer lines for half of the nucleotide m at 20 A. diameter and remainder distributed along a radius, t mass at a given radius being proportional to the radius. Abo C on the tenth layer line similar functions are plotted for an ou diameter of 12 A.

The DNA language is written in four letters

A (adenine), **T** (thymine), **G** (guanine), **C** (cytosine).

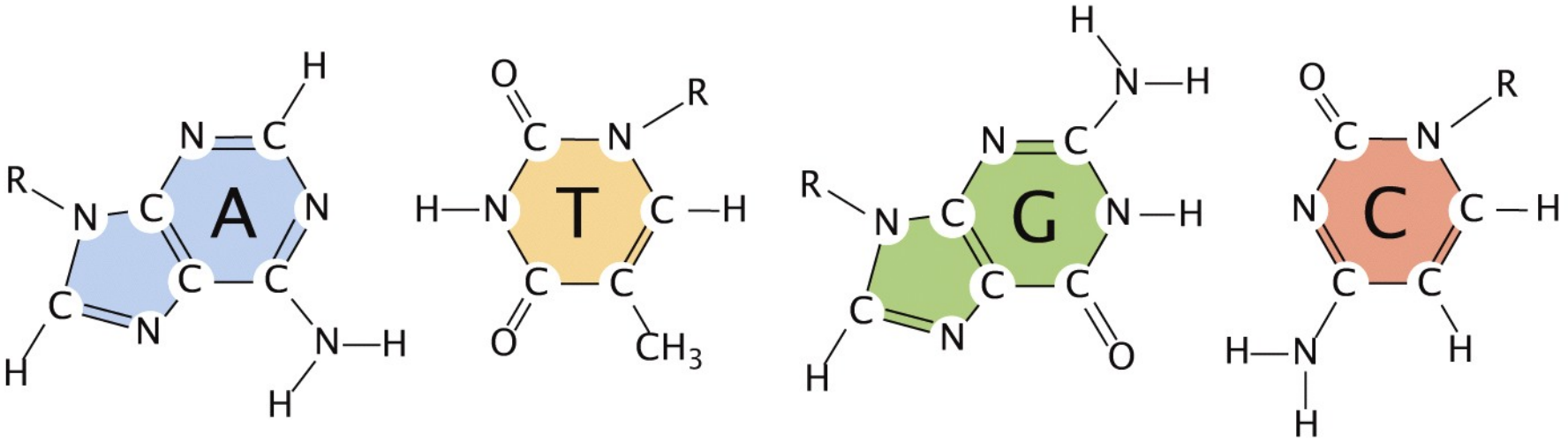
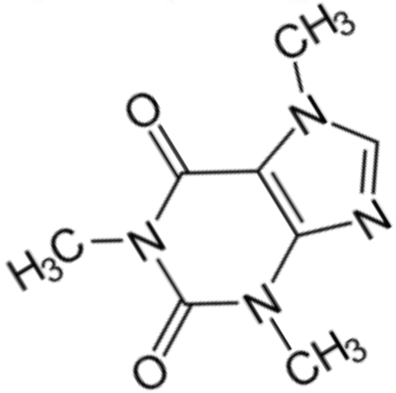
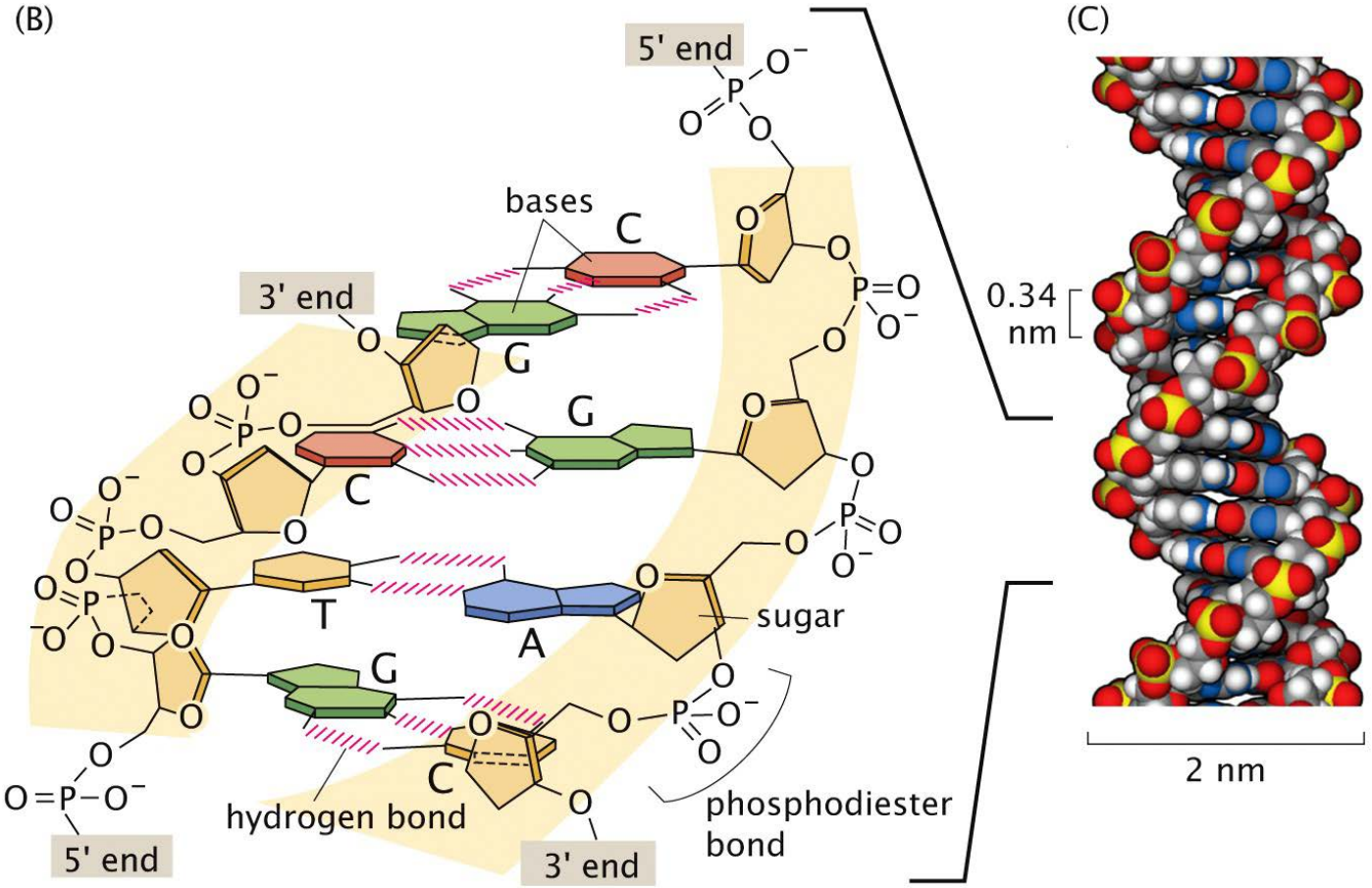


Figure 1.3a Physical Biology of the Cell, 2ed. (© Garland Science 2013)



Base complementarity enables reliable processing and storage of genomic data

Hydrogen bonds: **A—T G—C**



The genetic code is the dictionary that relates the two languages

- The code is redundant:

  20 AA/64 codons.

- The code is smooth: changing one letter often does not change AA.

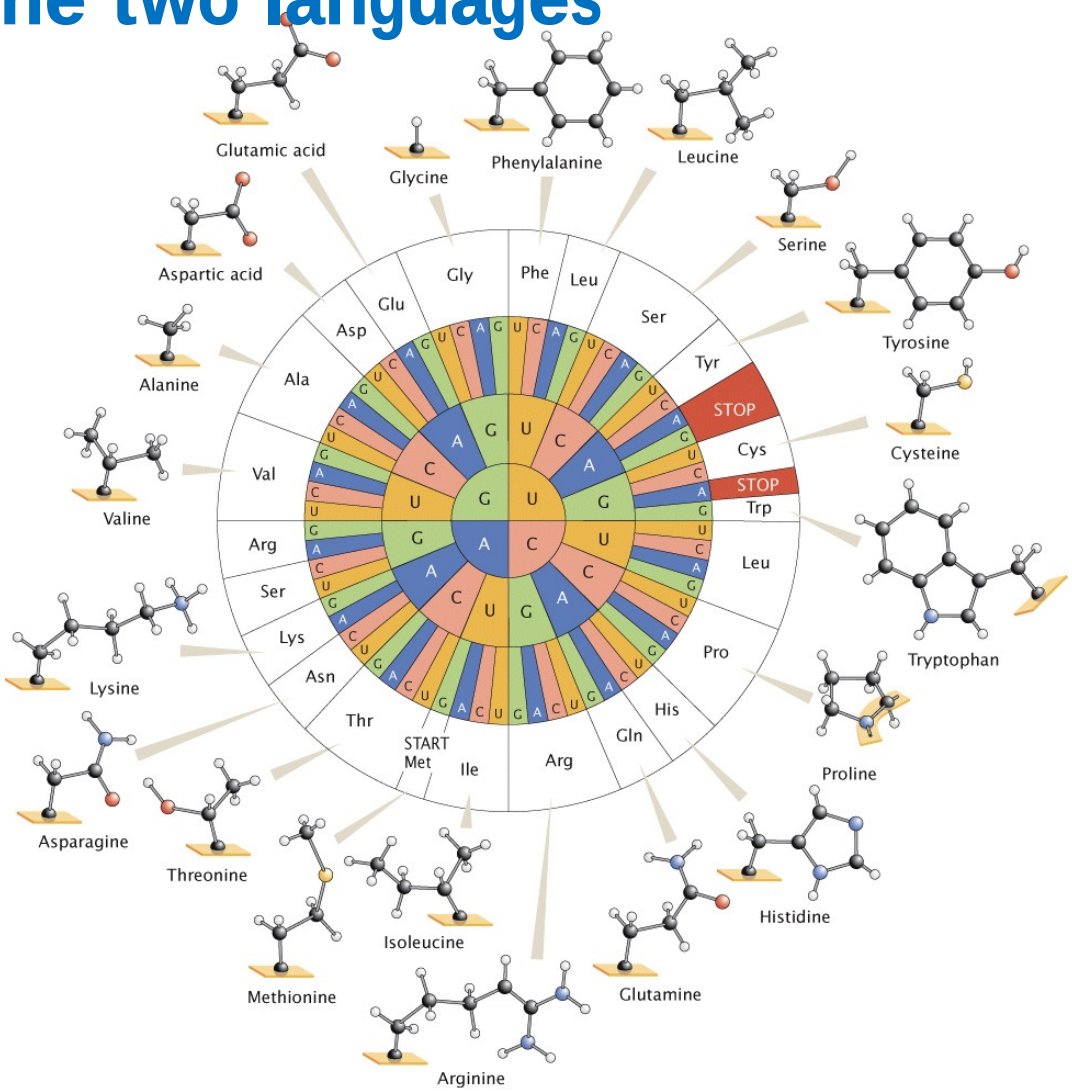


Figure 1.4 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Interlude: the power of scaling in natural science

- Determine “natural variables” and basic equations.
- Gives good estimates of quantities of interest:

How strong is the bomb?

The formation of a blast wave by a very intense explosion.
II. The atomic explosion of 1945

By Sir Geoffrey Taylor, F.R.S.

THE FIRST ATOMIC BOMB

AN ANALYSIS OF LOS ALAMOS PROJECT'S "TRINITY" TEST EXPLOSION

JORNADA DEL MUERTE,
ALAMOGORDO AIR BASE, NEW MEXICO
5:30 A.M. JULY 16, 1945.

Dimensional analysis

Assumptions:

- Energy released in a small space.
- Spherical shockwave.

Therefore the only parameters are:

time t: $[t] = T$

density ρ: $[\rho] = ML^{-3}$

Energy E: $[E] = ML^2T^2$

$$R(E, \rho, t) = E^{\alpha} \rho^{\beta} t^{\gamma}$$

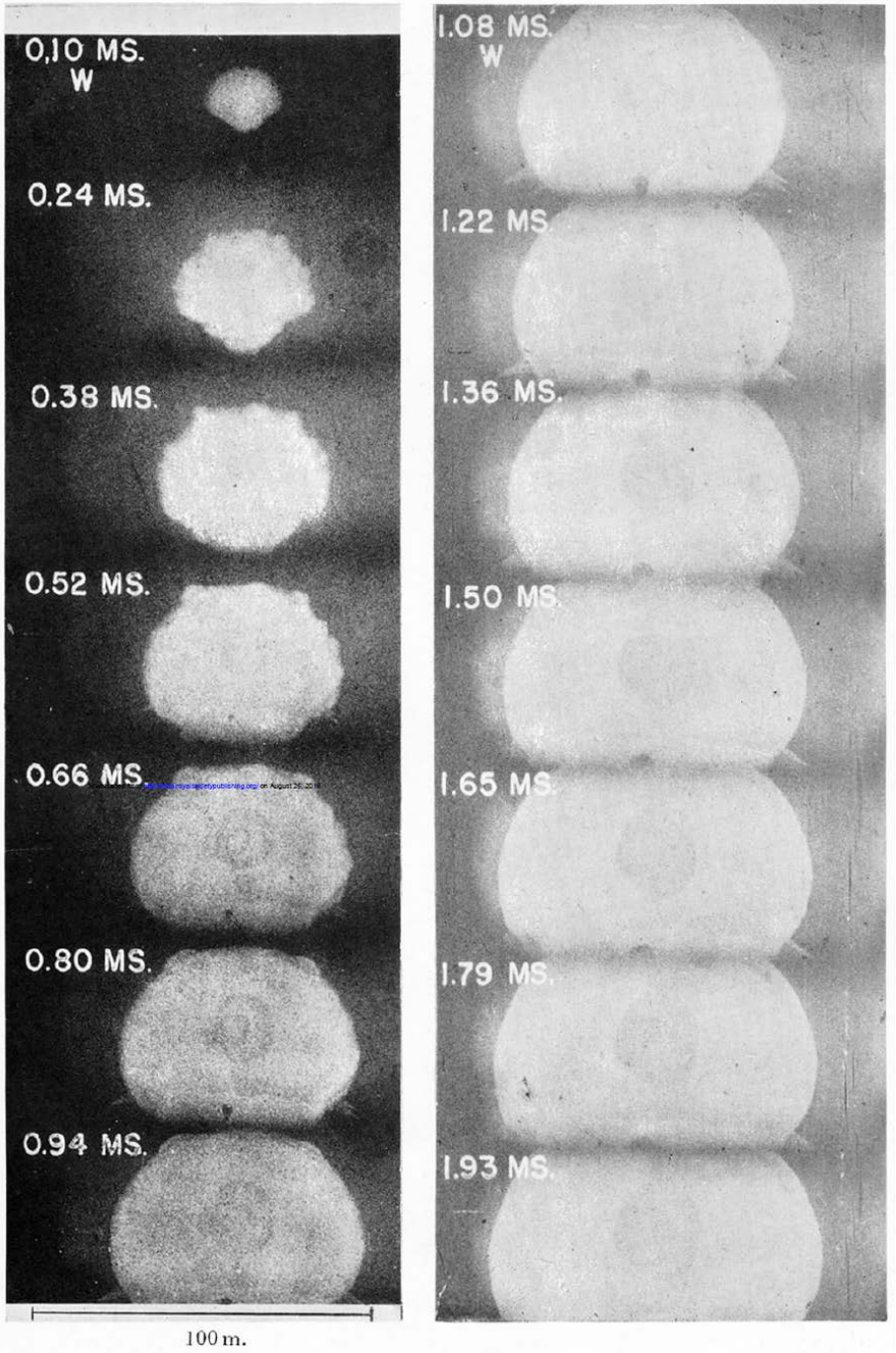
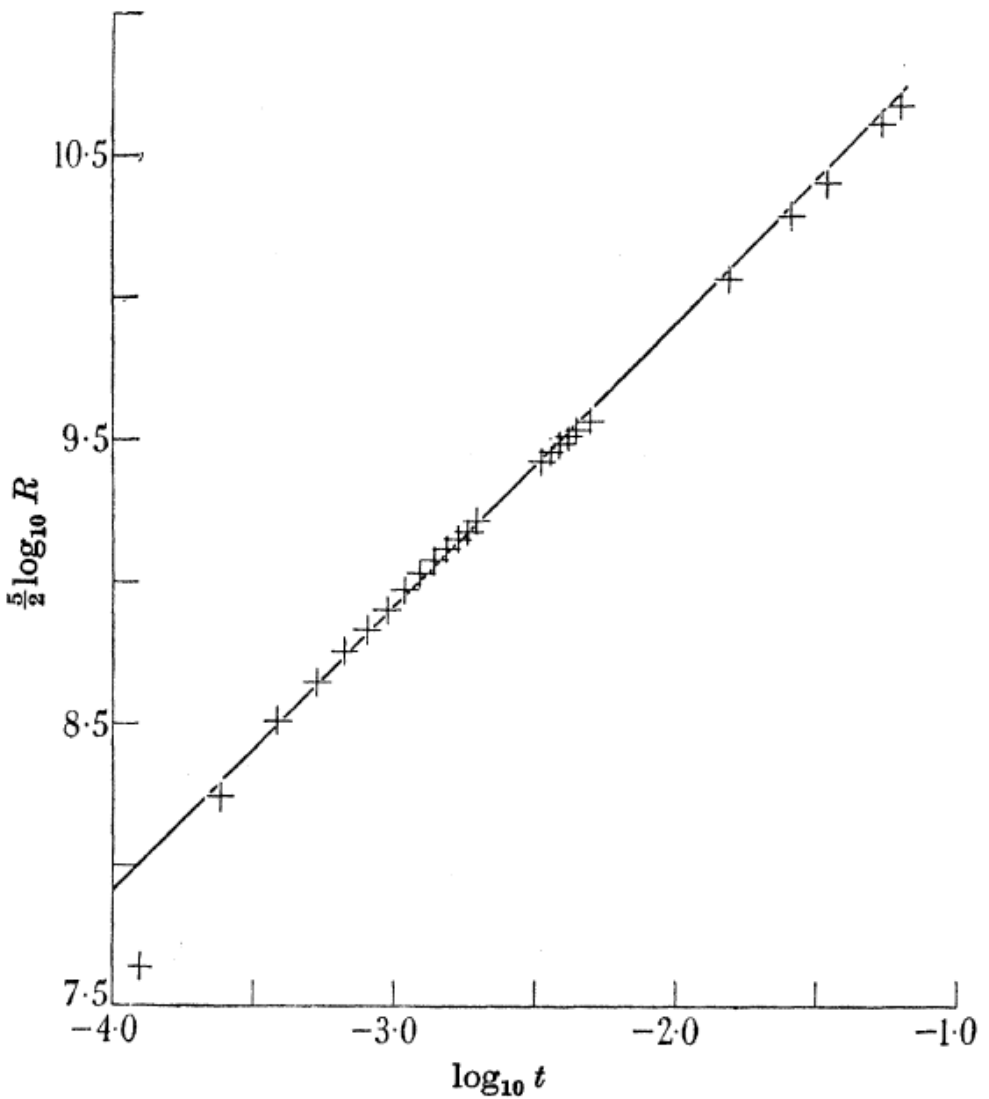


FIGURE 6. Succession of photographs of the 'ball of fire' from $t = 0{\cdot}10$ msec. to $1{\cdot}93$ msec.

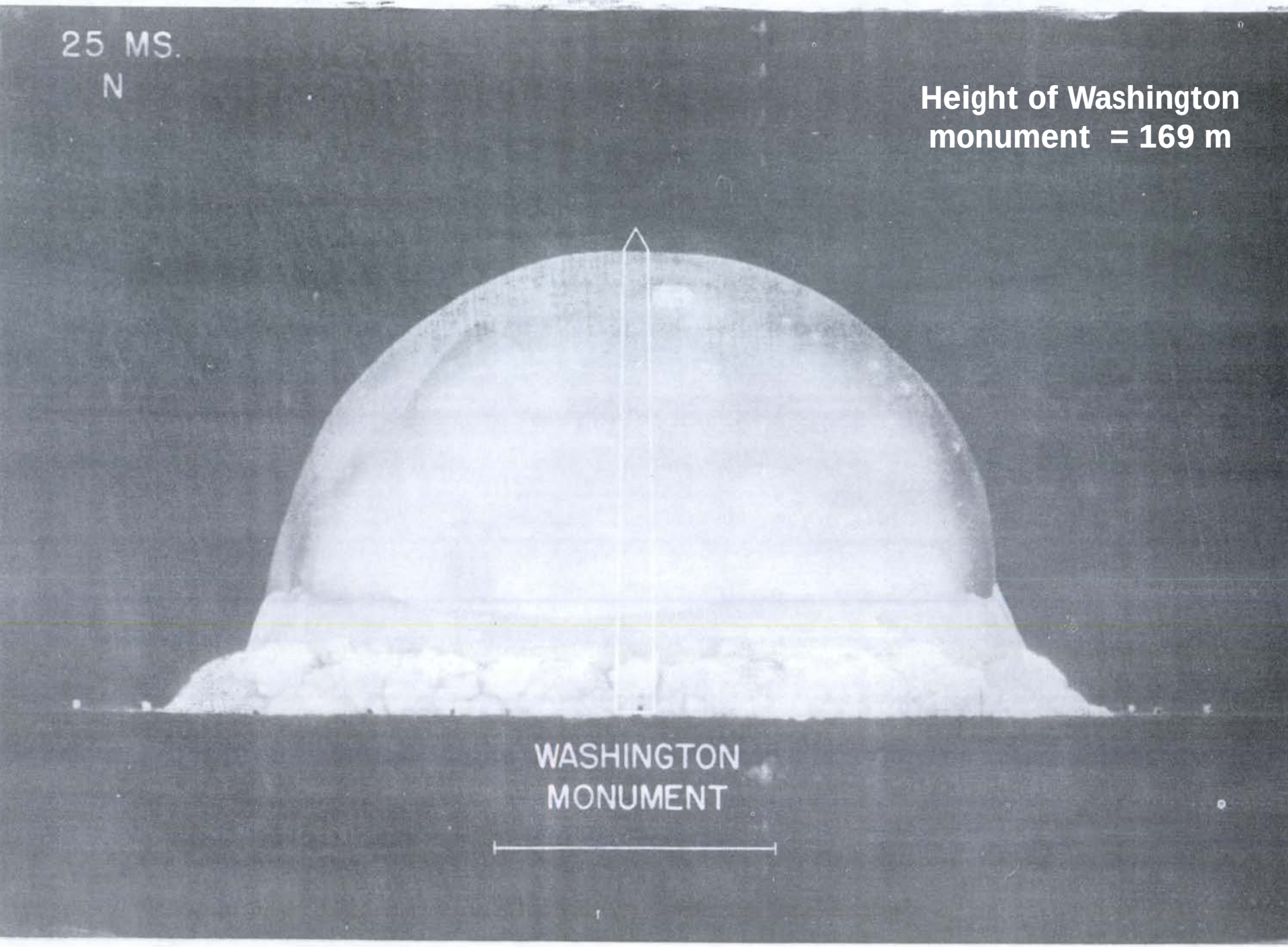
Scaling yields power laws

$$R(E,\rho,t) = \left(\frac{E}{\rho}\right)^{1/5} t^{2/5}$$

$$E = \rho \frac{R^5}{t^2}$$



IGURE 1. Logarithmic plot showing that $R^{\frac{5}{2}}$ is proportional to t.


25 MS.
N
Height of Washington monument = 169 m
WASHINGTON MONUMENT

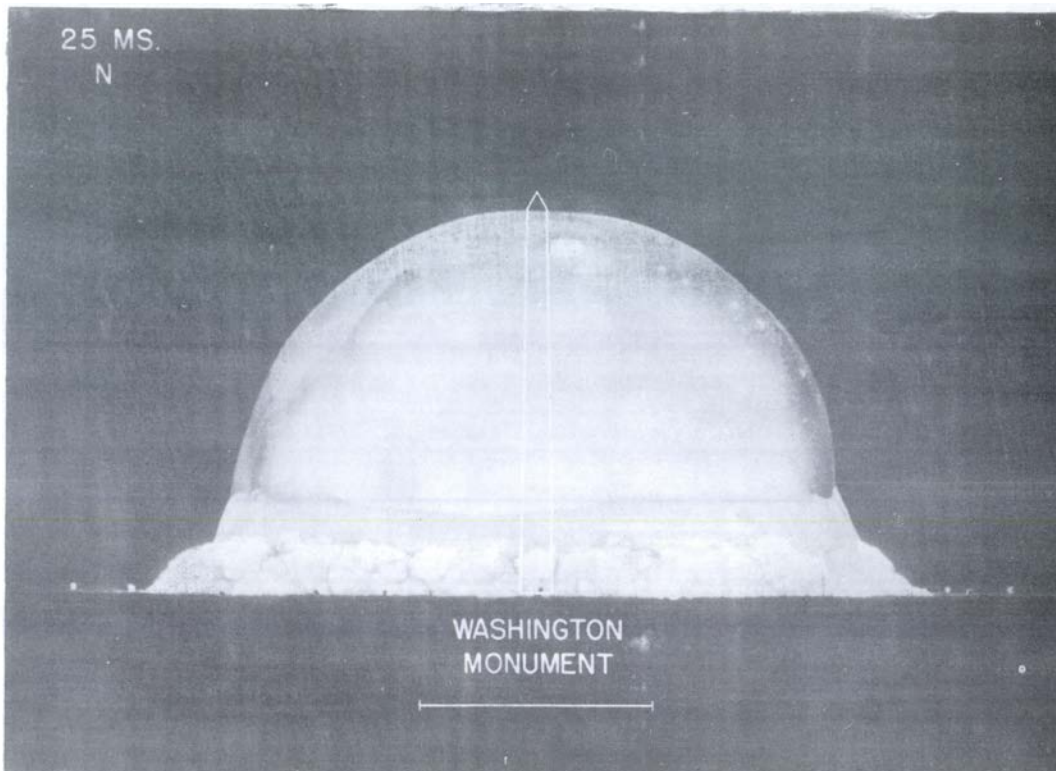
Scaling yields excellent guesstimates

$$E = \rho\left(\frac{R^5}{t^2}\right) = 1.2\,\frac{\text{kg}}{\text{m}^3} \times \frac{\left(1.3\cdot 10^2\,\text{m}\right)^5}{\left(2.5\cdot 10^{-2}\,s\right)^2} = 7\cdot 10^{13}\,\text{J}$$

$1\ \text{ton TNT} = 4.2\cdot 10^9\,\text{J}$

E = 17 kt

Actual yield was about 22 kt



We can derive natural units from physical constants

TABLE I An abbreviated list of the CODATA recommended values of the fundamental constants of physics and chemistry based on the 2014 adjustment.

Quantity	Symbol	Numerical value	Unit	Relative std. uncert. u_r
speed of light in vacuum	c, c_0	299 792 458	$\mathrm{m\ s^{-1}}$	exact
magnetic constant	μ_0	$4\pi \times 10^{-7}$	$\mathrm{N\ A^{-2}}$	
		$= 12.566\,370\,614... \times 10^{-7}$	$\mathrm{N\ A^{-2}}$	exact
electric constant $1/\mu_0 c^2$	ϵ_0	$8.854\,187\,817... \times 10^{-12}$	$\mathrm{F\ m^{-1}}$	exact
Newtonian constant of gravitation	G	$6.674\,08(31) \times 10^{-11}$	$\mathrm{m^3\ kg^{-1}\ s^{-2}}$	4.7×10^{-5}
Planck constant	h	$6.626\,070\,040(81) \times 10^{-34}$	J s	1.2×10^{-8}
$h/2\pi$	$\hbar$	$1.054\,571\,800(13) \times 10^{-34}$	J s	1.2×10^{-8}
elementary charge	e	$1.602\,176\,6208(98) \times 10^{-19}$	C	6.1×10^{-9}
magnetic flux quantum $h/2e$	Φ_0	$2.067\,833\,831(13) \times 10^{-15}$	Wb	6.1×10^{-9}
conductance quantum $2e^2/h$	G_0	$7.748\,091\,7310(18) \times 10^{-5}$	S	2.3×10^{-10}
electron mass	m_e	$9.109\,383\,56(11) \times 10^{-31}$	kg	1.2×10^{-8}
proton mass	m_p	$1.672\,621\,898(21) \times 10^{-27}$	kg	1.2×10^{-8}
proton-electron mass ratio	m_p/m_e	1836.152 673 89(17)		9.5×10^{-11}
fine-structure constant $e^2/4\pi\epsilon_0\hbar c$	α	$7.297\,352\,5664(17) \times 10^{-3}$		2.3×10^{-10}
inverse fine-structure constant	α^{-1}	137.035 999 139(31)		2.3×10^{-10}
Rydberg constant $\alpha^2 m_e c/2h$	R_∞	10 973 731.568 508(65)	$\mathrm{m^{-1}}$	5.9×10^{-12}
Avogadro constant	N_A, L	$6.022\,140\,857(74) \times 10^{23}$	$\mathrm{mol^{-1}}$	1.2×10^{-8}
Faraday constant $N_A e$	F	96 485.332 89(59)	$\mathrm{C\ mol^{-1}}$	6.2×10^{-9}
molar gas constant	R	8.314 4598(48)	$\mathrm{J\ mol^{-1}\ K^{-1}}$	5.7×10^{-7}
Boltzmann constant R/N_A	k	$1.380\,648\,52(79) \times 10^{-23}$	$\mathrm{J\ K^{-1}}$	5.7×10^{-7}
Stefan-Boltzmann constant $(\pi^2/60)k^4/\hbar^3 c^2$	σ	$5.670\,367(13) \times 10^{-8}$	$\mathrm{W\ m^{-2}\ K^{-4}}$	2.3×10^{-6}
Non-SI units accepted for use with the SI				
electron volt (e/C) J	eV	$1.602\,176\,6208(98) \times 10^{-19}$	J	6.1×10^{-9}
(unified) atomic mass unit $\frac{1}{12}m(^{12}\mathrm{C})$	u	$1.660\,539\,040(20) \times 10^{-27}$	kg	1.2×10^{-8}

Example of natural units : Planck units

- Set c = ħ = G = 1 (quantum meets gravity)

Quantity	Expression	Metric value	Name
Length (L)	$l_{\mathrm{P}} = \sqrt{\frac{\hbar G}{c^3}}$	1.616×10^{-35} m	Planck length
Mass (M)	$m_{\mathrm{P}} = \sqrt{\frac{\hbar c}{G}}$	2.176×10^{-8} kg	Planck mass
Time (T)	$t_{\mathrm{P}} = \sqrt{\frac{\hbar G}{c^5}}$	5.3912×10^{-44} s	Planck time

Advantages:

- Simple formulas.
- Physical meaning.
- No prototypes.

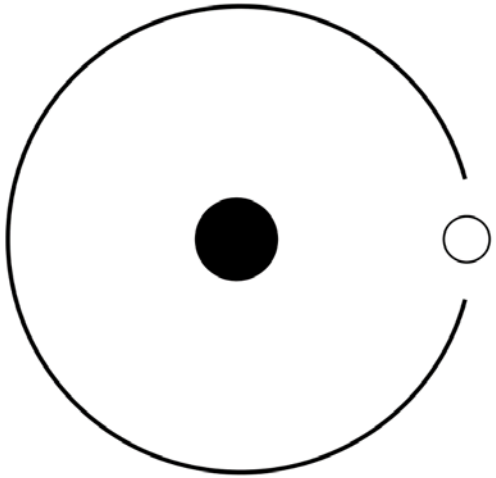
$$G_{\mu\nu} = 8\pi \frac{G}{c^4} T_{\mu\nu} \rightarrow G_{\mu\nu} = 8\pi T_{\mu\nu}$$

$$S_{\mathrm{BH}} = \frac{A_{\mathrm{BH}} c^3}{4G\hbar} \rightarrow S_{\mathrm{BH}} = \frac{A_{\mathrm{BH}}}{4}$$

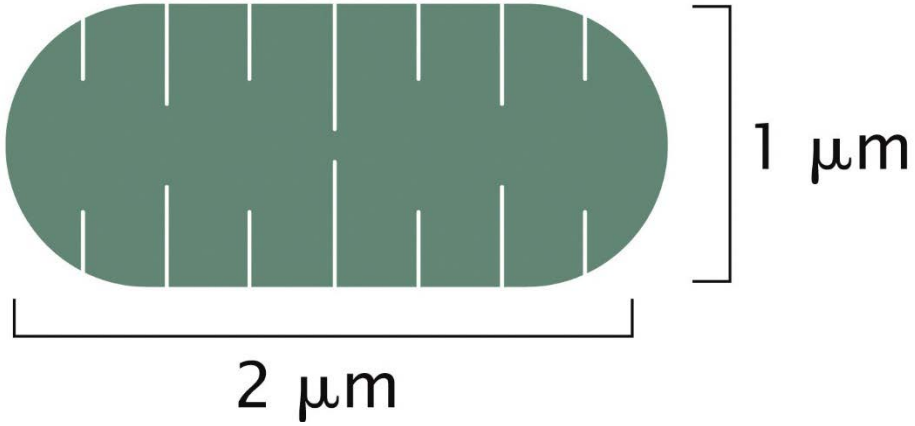
What are the natural units of biology?

- Hydrogen atom

$$R_{Bohr} = \frac{\hbar^2}{me^2} \approx 0.53 \text{ A}$$



- Escherichia coli



- Nothing smaller than a cell is alive

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)

A look into E. Coli – the "hydrogen atom" of biology + a few other atoms

Chapter 2

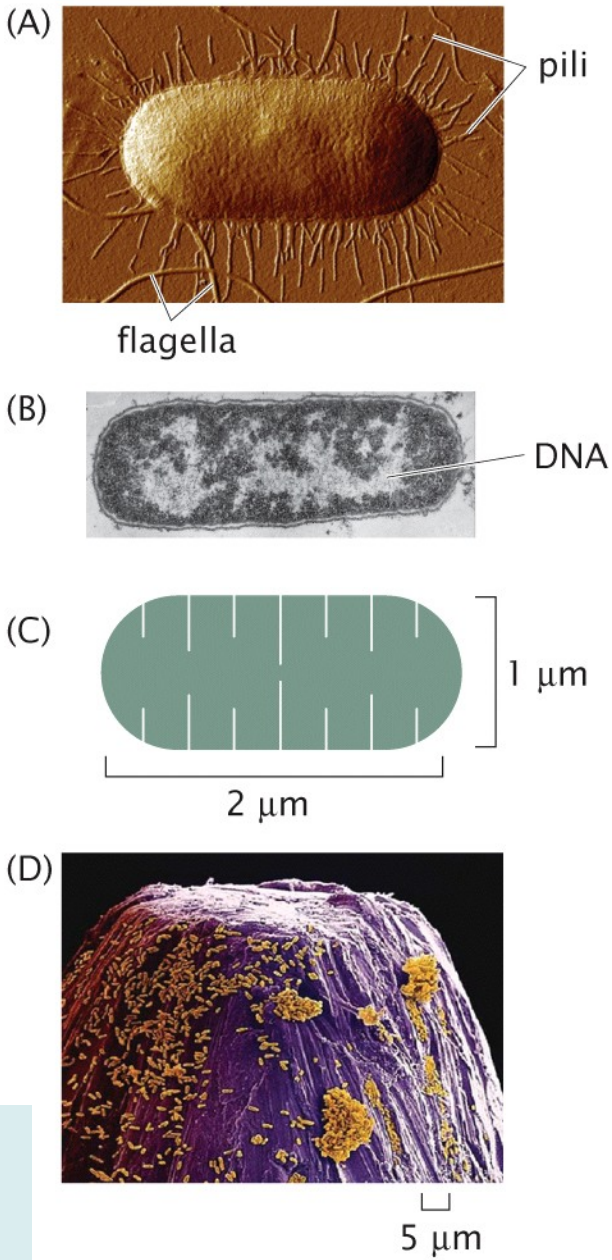
What and Where: Construction Plans for Cells and Organisms

E. coli is a standard prototype cell in biology

• All cells share with the "program" to convert *energy* into *order* and *survival*\reproduction.

•Minimal requirements for life (on Earth)

- Storage: DNA-based genome
- Translation: DNA→ RNA →proteins



(A) Atomic-force microscopy (AFM)
(B) Electron micrograph of a sectioned *E. coli*
(C) the *E. coli* ruler.
(D) Bacteria on the head of a pin.

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

Size of E. coli depends on growth rate

- Cantilever assay: Relation between buoyant mass growth rate (Godin et al., *Nat. Meth* 2010)

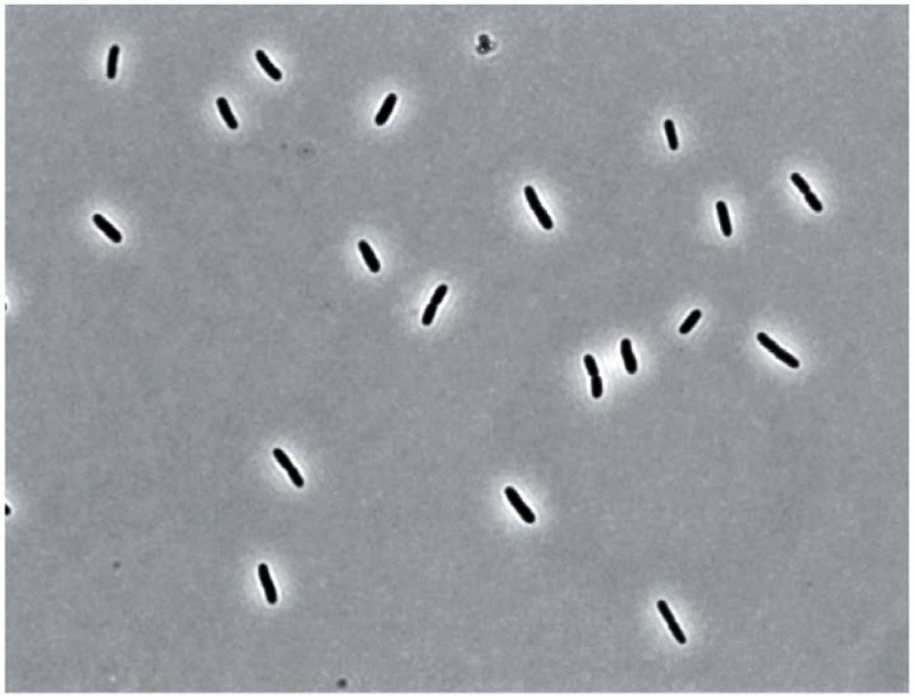


Figure 2.3c Physical Biology of the Cell, 2ed. (© Garland Science 2013)

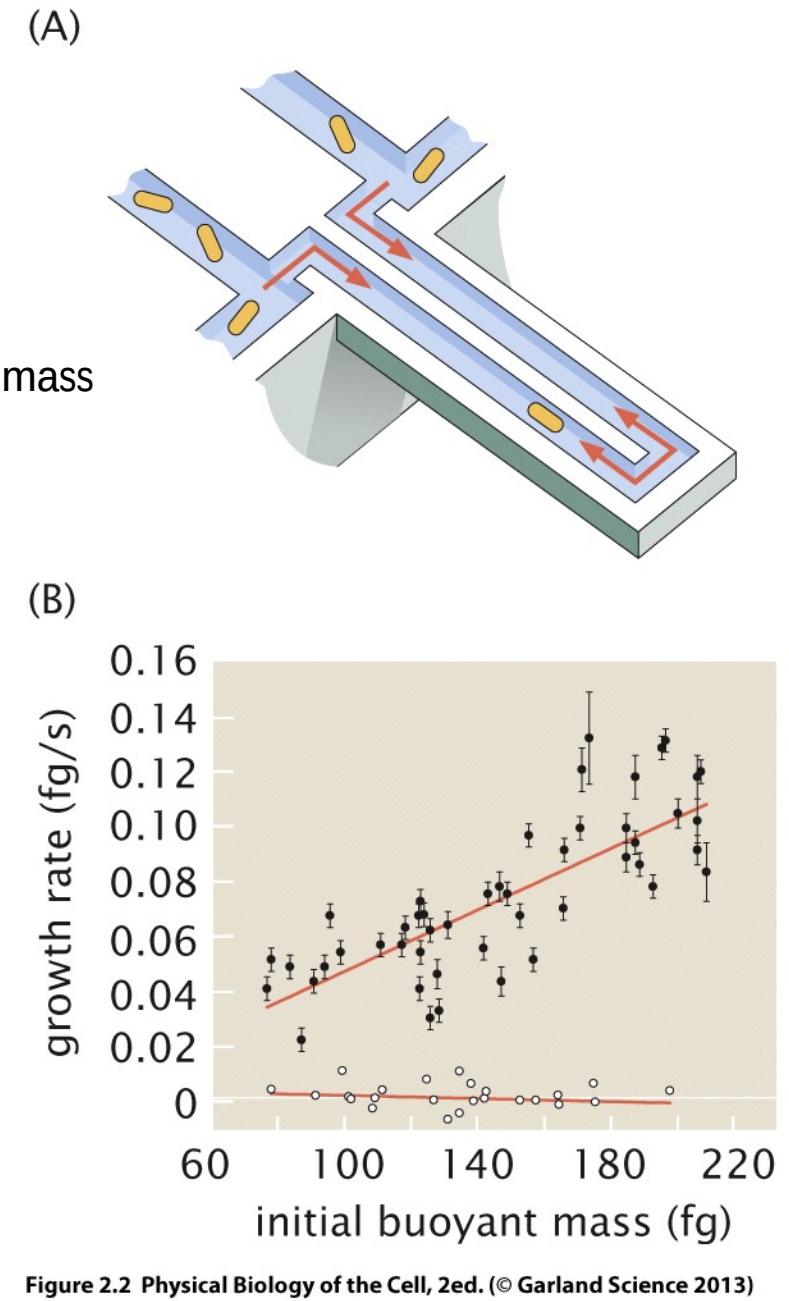


Figure 2.2 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Scaling and counting E. coli: “molecular census”

- Guesstimating how many copies of each molecule.

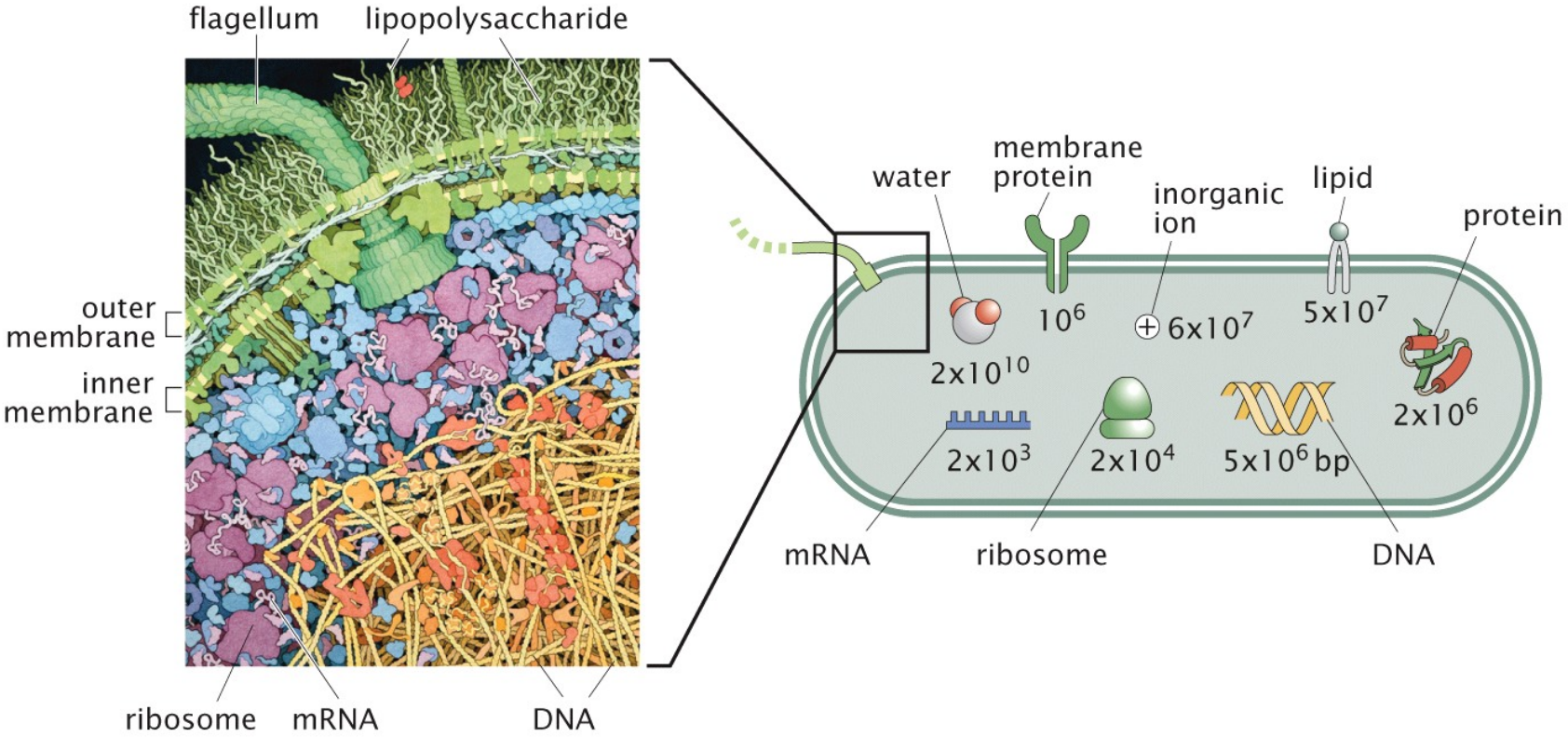


Figure 2.4 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Rule of thumb: concentrations in E. coli units

Conc.	Molec./cell	distance
1 nM	1	1,000 nm
1 μM	10^3	100 nm
1 mM	10^6	10 nm
1 M	10^9	1 nm

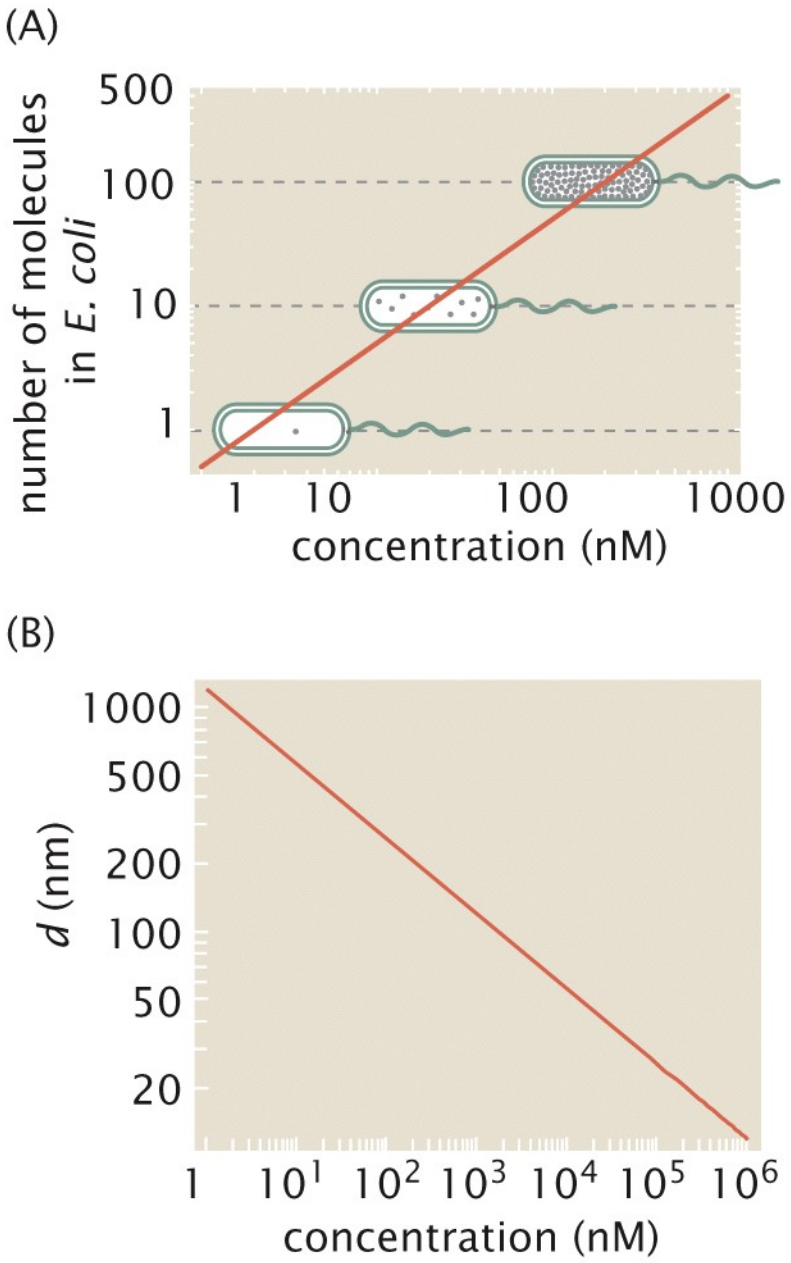
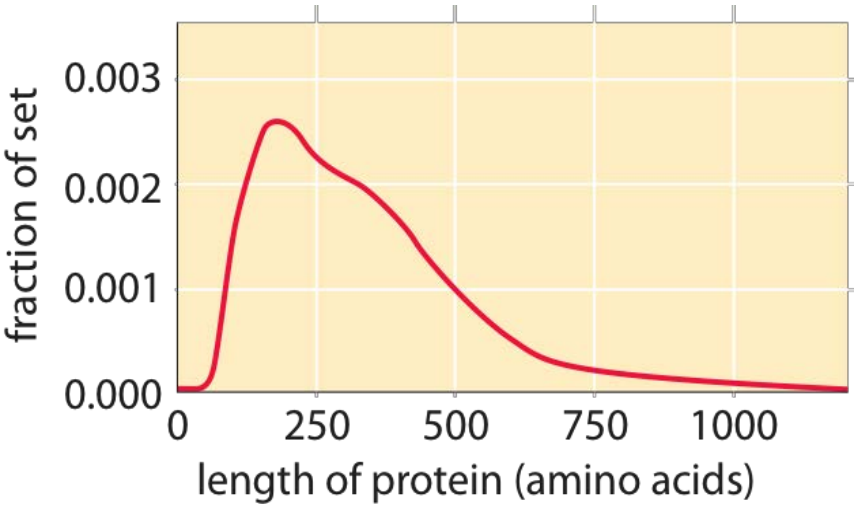
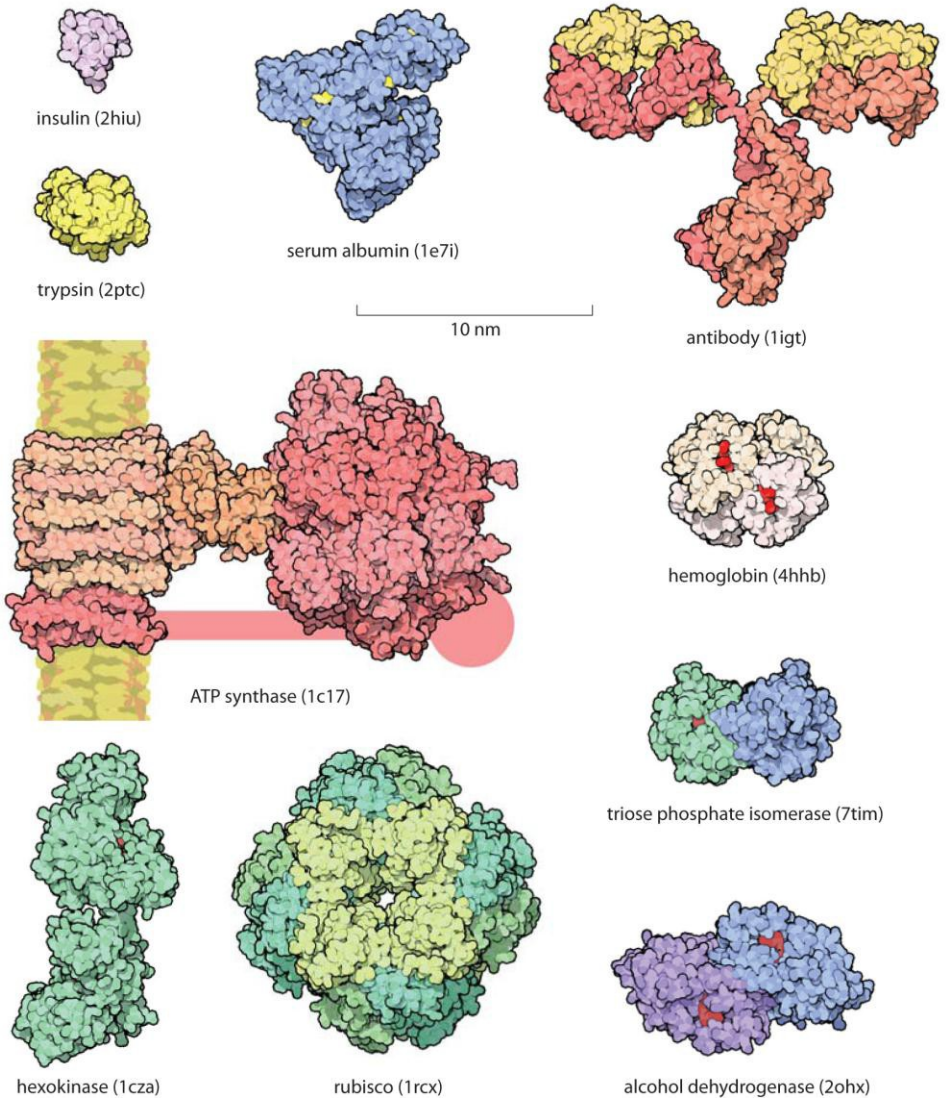


Figure 2.12 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The average protein has 300 amino acids





Can we guesstimate the number of genes from genome size?

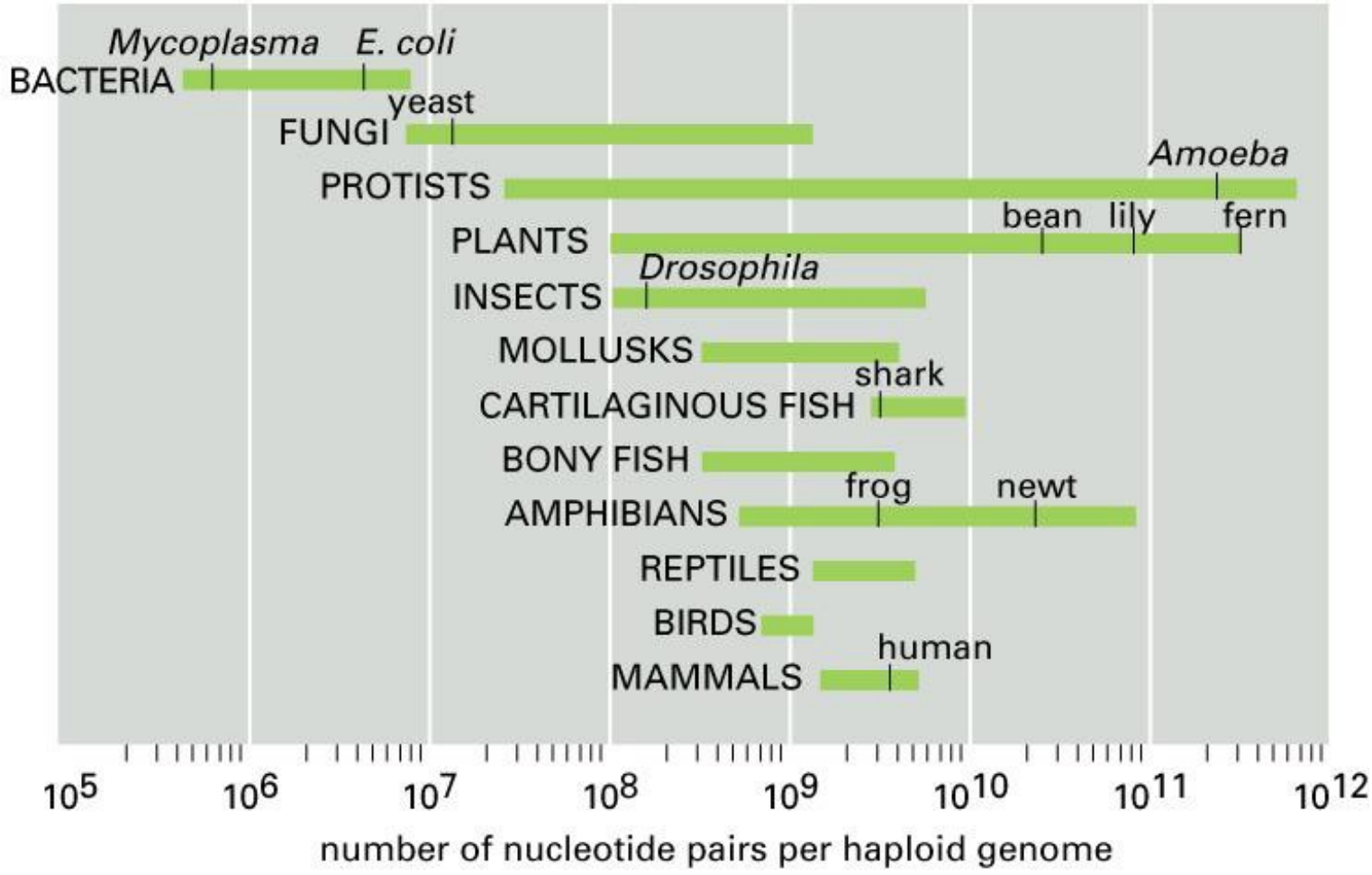


Figure 1–38. Molecular Biology of the Cell, 4th Edition.

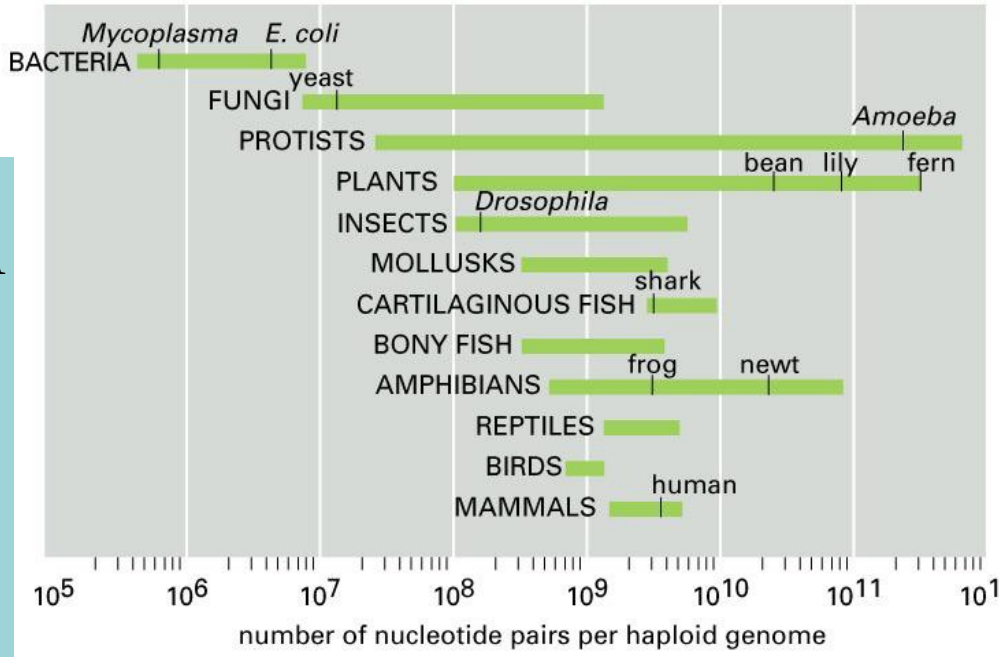
Can we guesstimate the number of genes from genome size?

- 1 protein = 300 amino acids
- 1 amino acid = 1 DNA codon = 3 bp DNA

$\Rightarrow L_{\text{gene}} \approx 10^3 \text{ bp}$

- $L_{\text{genome}} = 5 \cdot 10^6 \text{ bp}$

$\Rightarrow N_{\text{genes}} = \dfrac{L_{\text{genome}}}{L_{\text{gene}}} = \dfrac{5 \cdot 10^6}{10^3} \approx 5{,}000 \text{ genes}$



gure 1–38. Molecular Biology of the Cell, 4th Edition.

- Works great for E. coli.

In eukaryotes most DNA is not coding genes

- Not so great for humans: estimate gives 3 million genes…

Organism	# of protein-coding genes	# of genes naïve estimate: (genome size /1000)	PMID	BNID
HIV 1	9	10	Kuiken et al., 2008	105769
In uenza A virus	10-11	14	19230160	105767
Bacteriophage λ	66	49	Daniels et al., 1983	105770
Epstein Barr virus	80	172	John W. Kimball Biology Pages	103246
Buchnera sp.	610	641	12364605	105757
T. maritima	1,861	1,906	12364605	105766
S. aureus	2,788	2,878	12364605	105763
V. cholerae	3,949	4,033	12364605	105760
E. coli	4,225	4,600	19402753	105443
B. subtilis	4,233	4,215	Rogozin et al., 2002	105753
S. cerevisiae	5,616	12,160	16169922	105444
C. elegans	19,735	100k	16339362	101364
D. melanogaster	13,601	180k	10731132	100200
A. thaliana	20,568	125k	17210932	105446
F. rubripes	23,131	400k	10523524	100280
Z. mays	~49,000	2,400k	16339807	105521
M. musculus	20,210	2,500k	19468303	100310
H. sapiens	19,042	3,080k	19468303	105447
T. aestivum	107,000	16,800k	17010109	105448

E coli census: How many proteins?

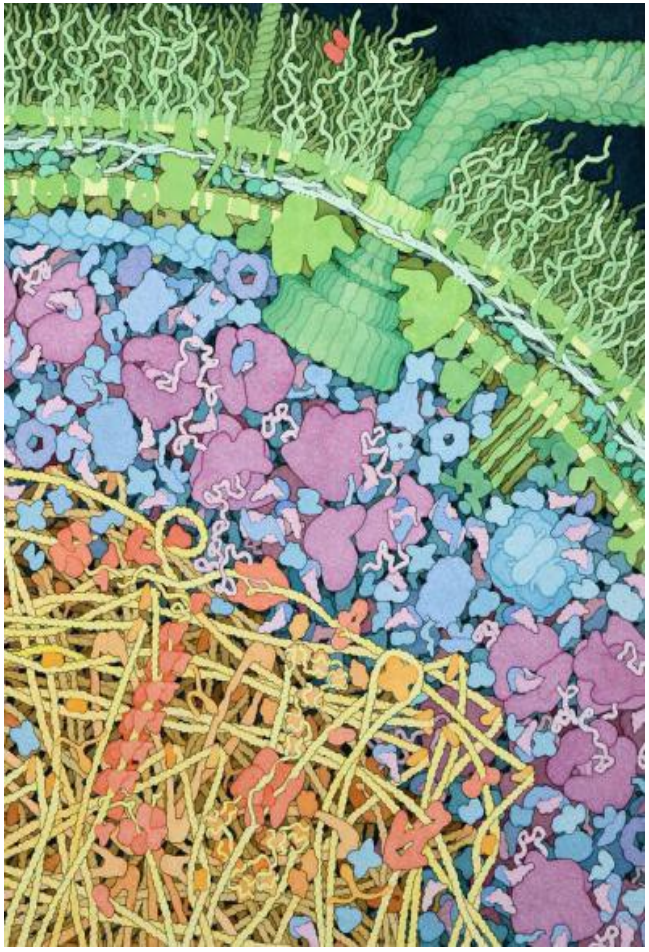
- Dry mass of protein (macromolecules) $\approx 30\%$ of mass
- Proteins are about half $\approx 15\%$ of mass

$$V_{\text{Ecoli}} = 1\mu m^3$$

$$\Rightarrow V_{\text{proteins}} = 0.15 \times 1\ \mu\text{m}^3 = 0.15\ \mu\text{m}^3 = 1.5 \cdot 10^8\ \text{nm}^3$$

$$r_{\text{protein}} = 2\text{nm} \Rightarrow v_{\text{protein}} = \frac{4\pi}{3}\left(r_{\text{protein}}\right)^3 \approx 30\ \text{nm}^3$$

$$N_{\text{proteins}} = \frac{V_{\text{proteins}}}{v_{\text{protein}}} = \frac{1.5 \cdot 10^8\ \text{nm}^3}{30\ \text{nm}^3} = 5 \cdot 10^6$$



First, calculate the mass of E. coli:

- 1 mole H_2O = 18 gr = $N_A \times 18\text{Da} = 18\ \text{cm}^3$ $(N_A = 6 \cdot 10^{23})$
- $\rho_{H_2O} \approx \rho_{\text{Ecoli}} = 1\frac{N_A \text{Da}}{\text{cm}^3} = 10^{-12}\frac{N_A \text{Da}}{\mu\text{m}^3} \approx 1\frac{\text{gr}}{\text{cm}^3}$
- $V_{\text{Ecoli}} = 1\ \mu\text{m}^3$

$$\Rightarrow M_{\text{Ecoli}} = \rho_{\text{Ecoli}} V_{\text{Ecoli}} = 10^{-12}\frac{N_A \text{Da}}{\mu\text{m}^3} \times 1\ \mu\text{m}^3 = 10^{-12} N_A \text{Da}$$

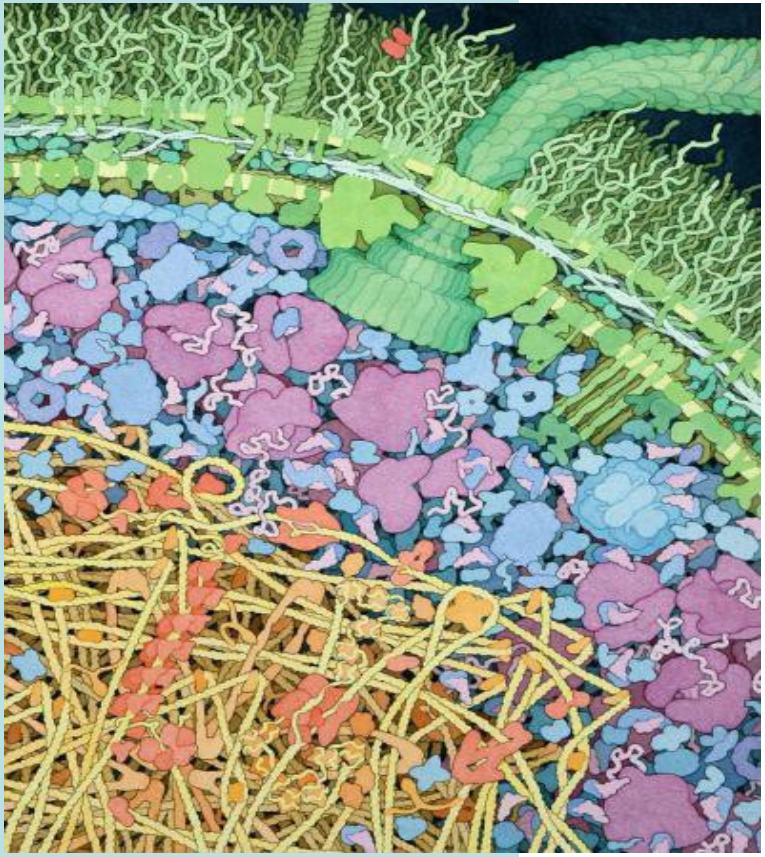
- Dry mass (macromolecules) ≈ 30% of mass
- Proteins are about half ≈15% of mass

$$\Rightarrow M_{\text{proteins}} = 0.15\, M_{\text{Ecoli}} = 1.5 \cdot 10^{-13} N_A \text{Da}$$

- 1 protein = 300 amino acids (AA)
- 1 AA = 100 Da

$$\Rightarrow \quad m_{\text{protein}} = 300\ \text{AA} \times 100\ \text{Da} = 3 \cdot 10^4\ \text{Da}$$

$$N_{\text{proteins}} = \frac{M_{\text{proteins}}}{m_{\text{protein}}} = \frac{1.5 \cdot 10^{-13} N_A \text{Da}}{3 \cdot 10^4\ \text{Da}} = 5 \cdot 10^{-18} N_A = 5 \cdot 10^{-18} \times 6 \cdot 10^{23} = 3 \cdot 10^6\ \text{proteins}$$



E coli census:
How many lipids?

- ribosome mass $m_{\text{ribo}} = 3 \cdot 10^6$ Da
- ribosome mass m_{ribo} = 2/3 rRNA + 1/3 proteins

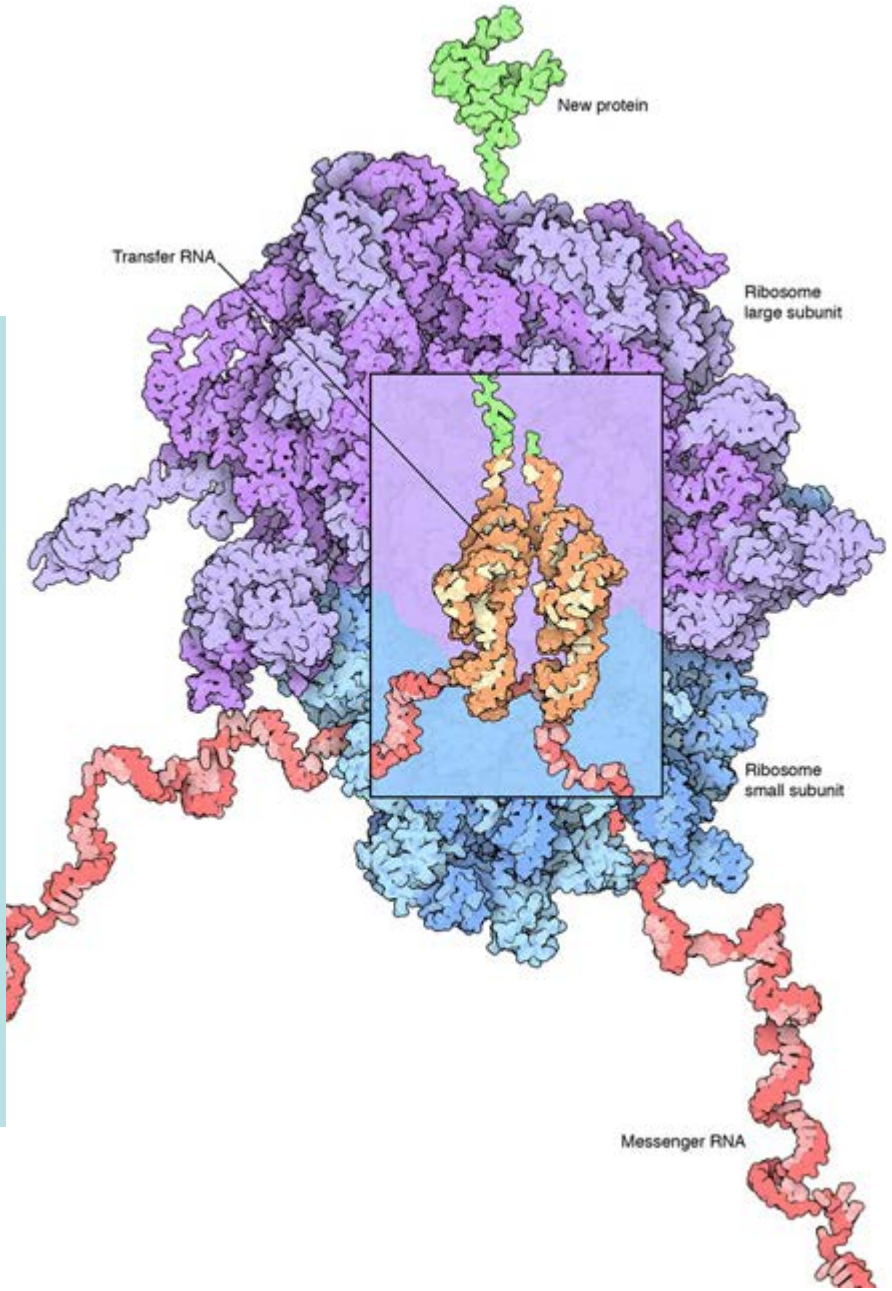
$$\Rightarrow m_{\text{ribo}}^{\text{protein}} = 10^6 \text{ Da}$$

- $M_{\text{ribo}}^{\text{protein}}$ = 20% of M_{proteins}

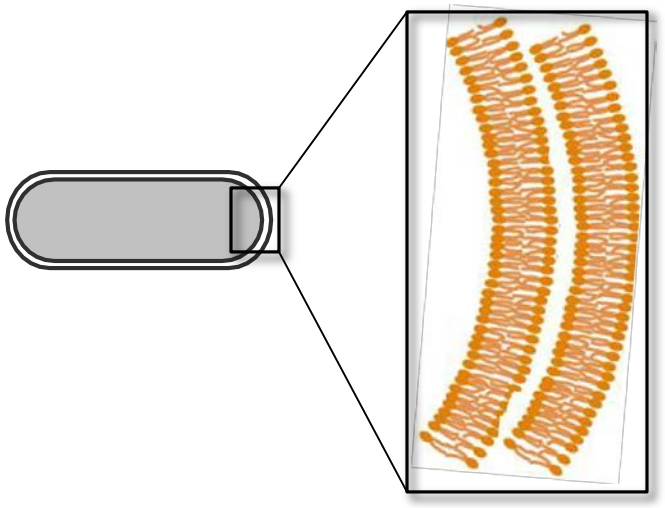
$$= 1.5 \cdot 10^{-13} N_A \text{Da} / 5 = 3 \cdot 10^{-14} N_A \text{Da}$$

$$\Rightarrow N_{\text{ribo}} = \frac{M_{\text{ribo}}^{\text{protein}}}{m_{\text{ribo}}^{\text{protein}}} = \frac{3 \cdot 10^{-14} N_A \text{Da}}{10^6 \text{ Da}}$$

$$= 3 \cdot 10^{-20} N_A = 2 \cdot 10^4 \text{ ribosomes}$$

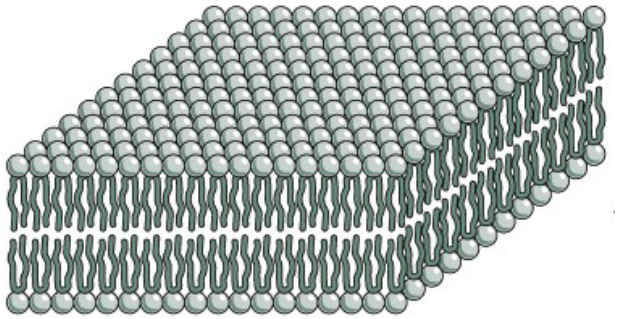


E coli census:
How many lipids?



- E. coli area $S_{\text{Ecoli}} = 5\ \mu\text{m}^2$
- Lipid area $S_{\text{lipid}} = 0.5\ \text{nm}^2$
- Two, inner and outer, bilayers (4 monolayers)
- 50% of area is covered with lipids

$$\Rightarrow N_{\text{lipids}} = 4 \times 0.5 \times \frac{S_{\text{Ecoli}}}{S_{\text{lipid}}} = 2 \times \frac{5\ \mu\text{m}^2}{0.5\ \text{nm}^2} = 2 \cdot 10^7$$



E coli census:
There are many small molecules

- 70% of E. coli mass is water:

$$M_{\text{Ecoli}}^{\text{H}_2\text{O}} = 70\%\, M_{\text{Ecoli}} = 7\cdot 10^{-13}\ N_{\text{A}}\,\text{Da}$$

- $m_{\text{H}_2\text{O}} = 18$ Da

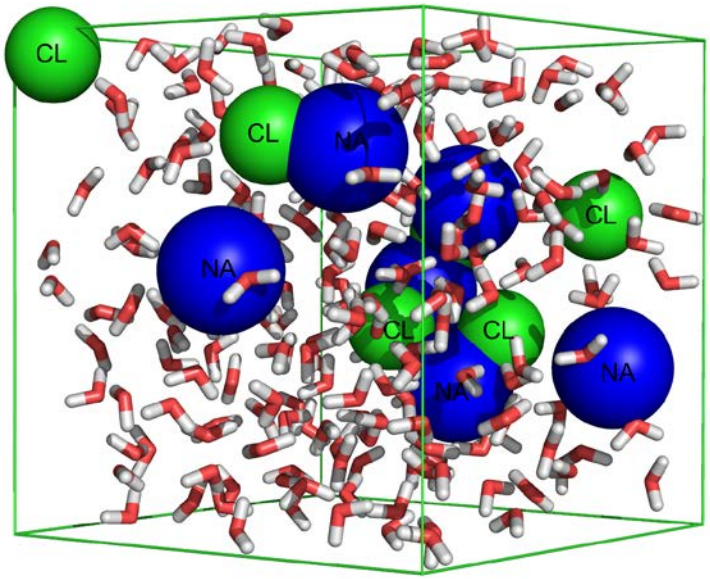
$$\Rightarrow N_{\text{H}_2\text{O}} = \frac{7\cdot 10^{-13}\ N_{\text{A}}\,\text{Da}}{18\ \text{Da}} \approx 4\cdot 10^{-14}\ N_{\text{A}}$$

$$= 4\cdot 10^{-14} \times 6\cdot 10^{23} = 2.4\cdot 10^{10}$$

- 1 nM = 1 molecule/E. coli $\Rightarrow \rho_{\text{H}_2\text{O}} = 24$ M

- *T*ypical concentration of ions (K^+,Na^+...) 100mM

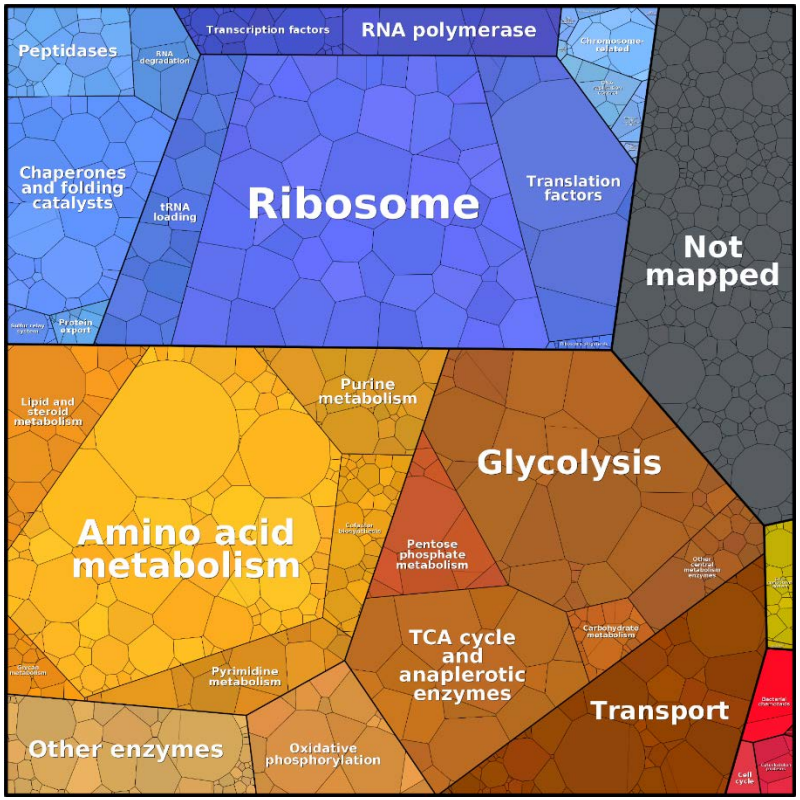
$$\Rightarrow N_{\text{ion}} = \frac{0.1M}{1\ \text{molecule}/1nM} = 10^8$$



And the numbers are

molecules	number	proportion
genes	$5 \cdot 10^3$	3
mRNA	$2 \cdot 10^3$	1
ribosomes	$2 \cdot 10^4$	10
proteins	$2 \cdot 10^6$	10^3
lipids	$2 \cdot 10^7$	10^4
ions	10^8	10^5
water	$2 \cdot 10^{10}$	10^7

Proteins perform diverse functions



www.proteomaps.net

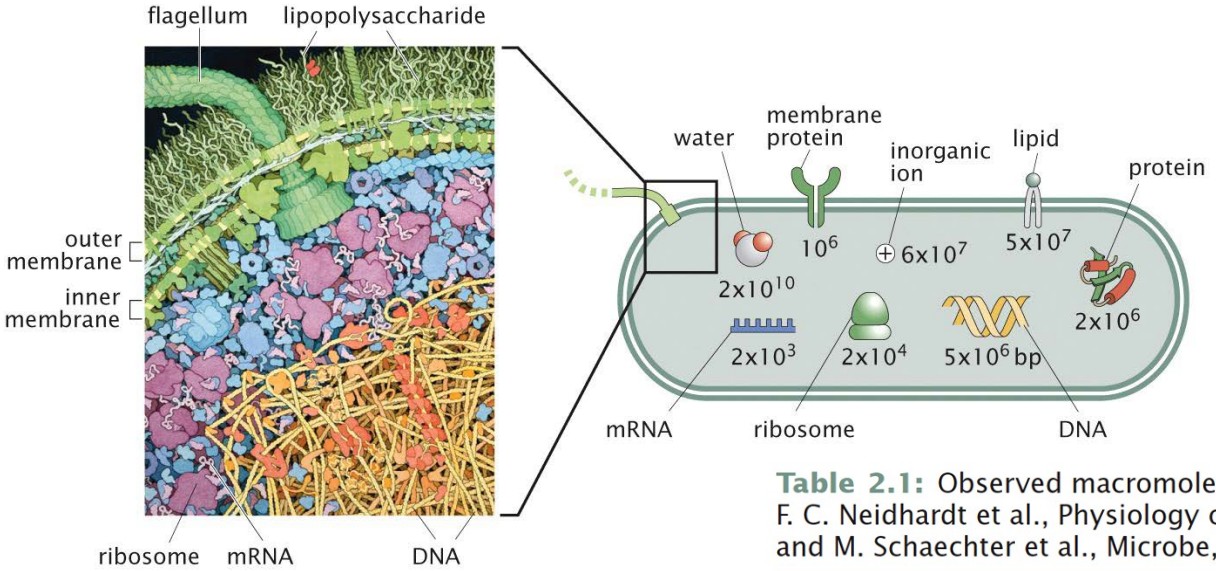


Table 2.1: Observed macromolecular census of an *E. coli* cell. (Data from F. C. Neidhardt et al., Physiology of the Bacterial Cell, Sinauer Associates, 1990 and M. Schaechter et al., Microbe, ASM Press, 2006.)

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**	

How do we count?

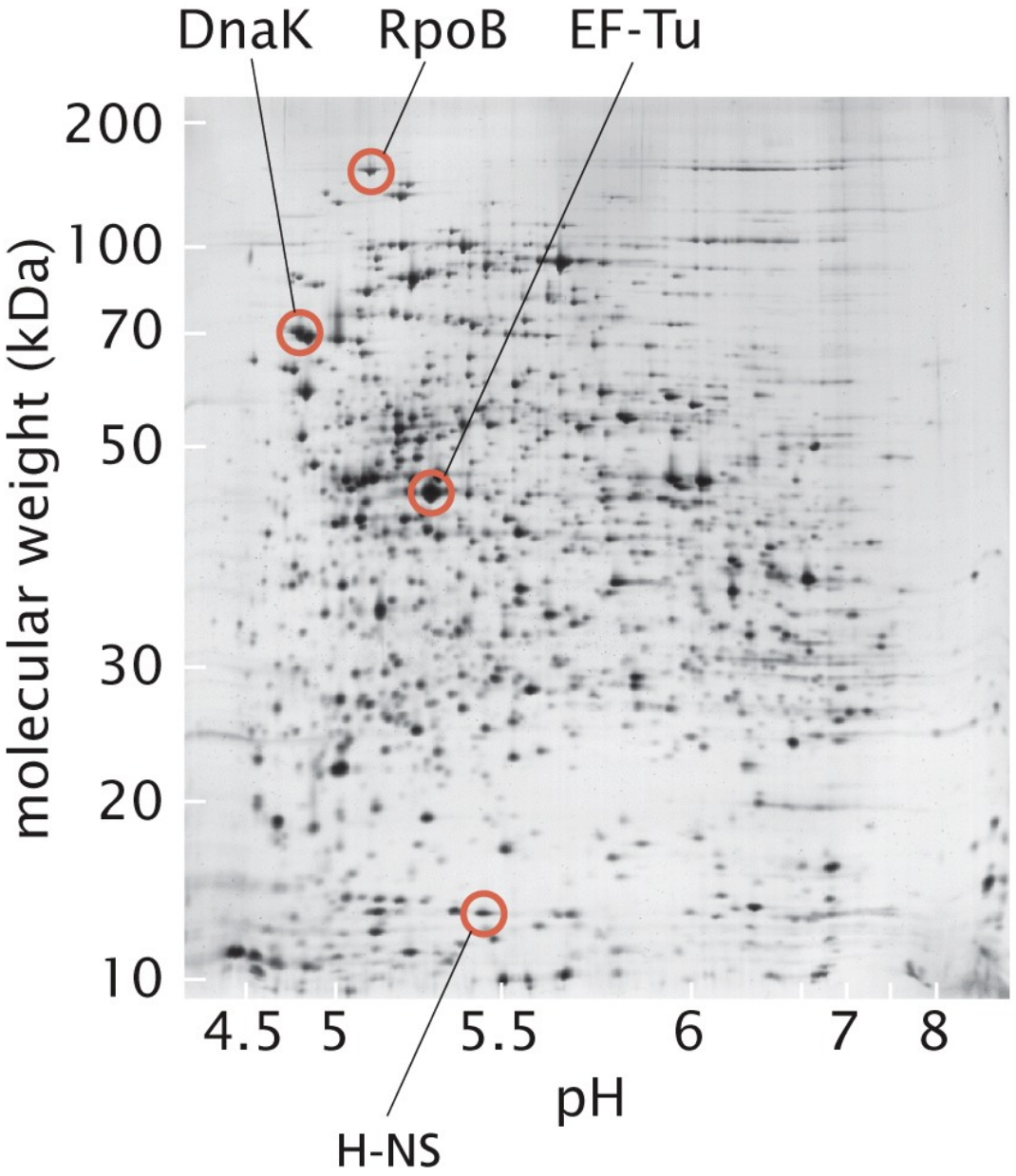


Figure 2.5 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The classical method:
Protein census of *E. coli* in 2D gel.

Hard to count small numbres

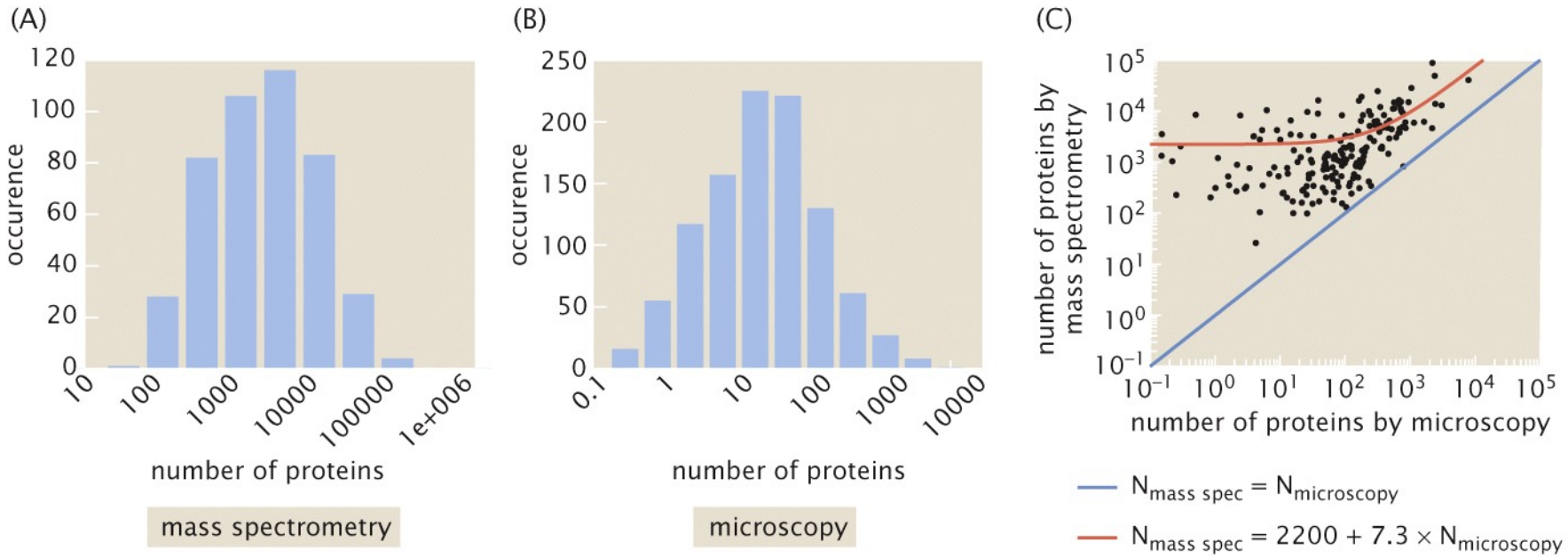


Figure 2.6 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Protein census of *E. coli* using (A) mass spectrometry and (B) fluorescence microscopy with protein fusions to fluorescence proteins. (C) Comparison showing a discrepancy

(A, adapted from P. Lu et al., *Nat. Biotechnol.* 25:117, 2007; B, adapted from Y. Taniguchi et al., *Science* 329:533, 2010.)

Cells and the small numbers limit

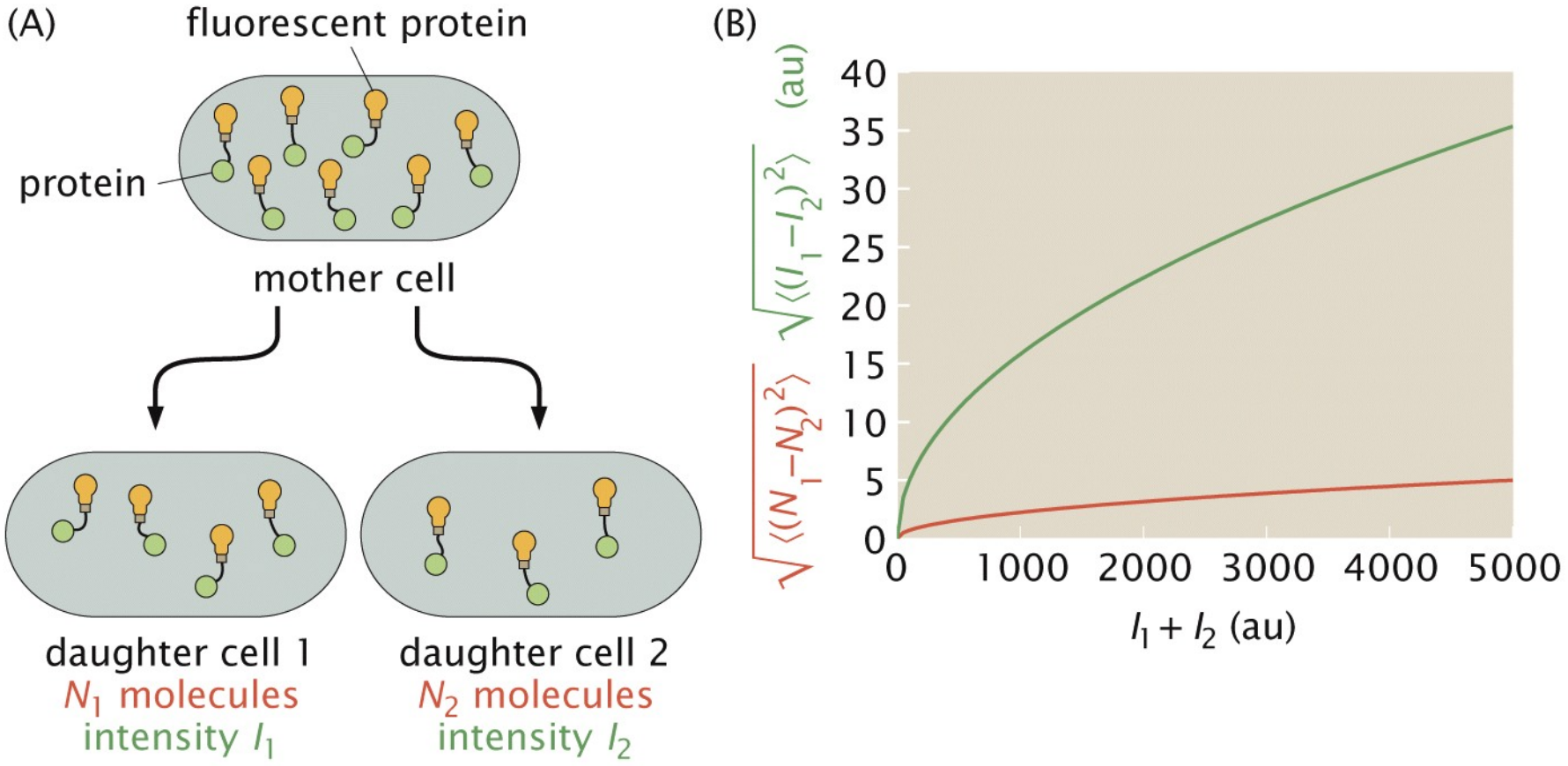


Figure 2.10 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Cells and the small numbers limit

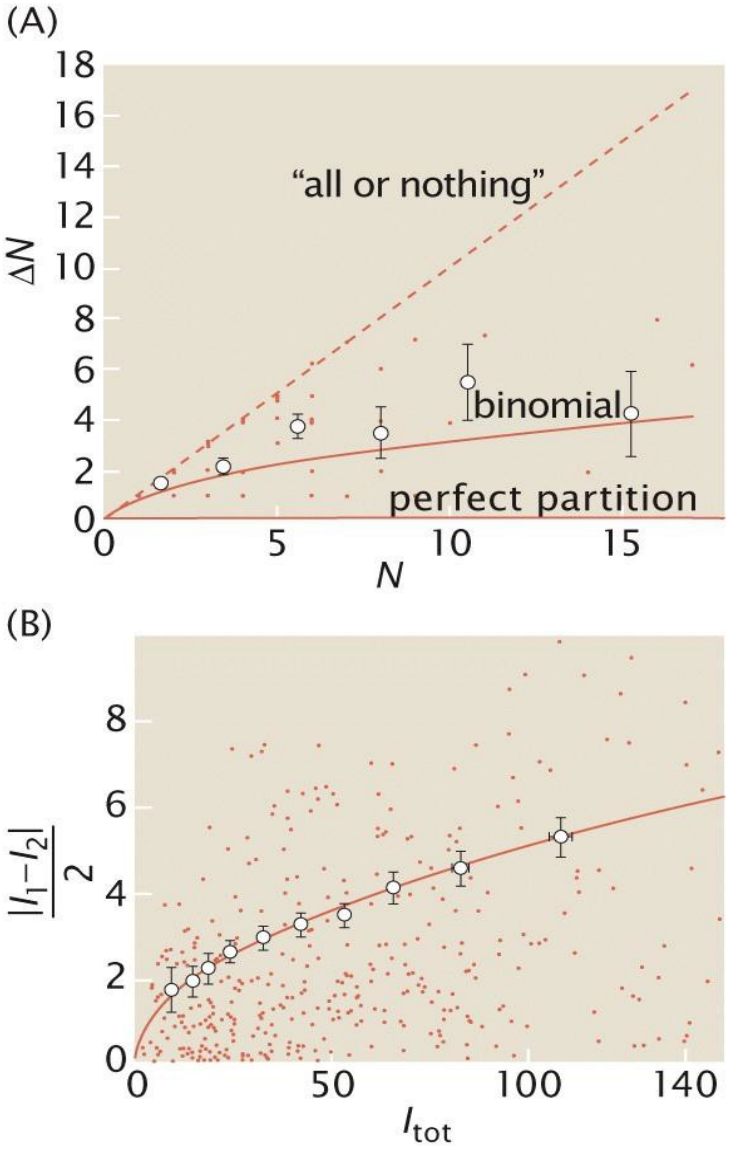
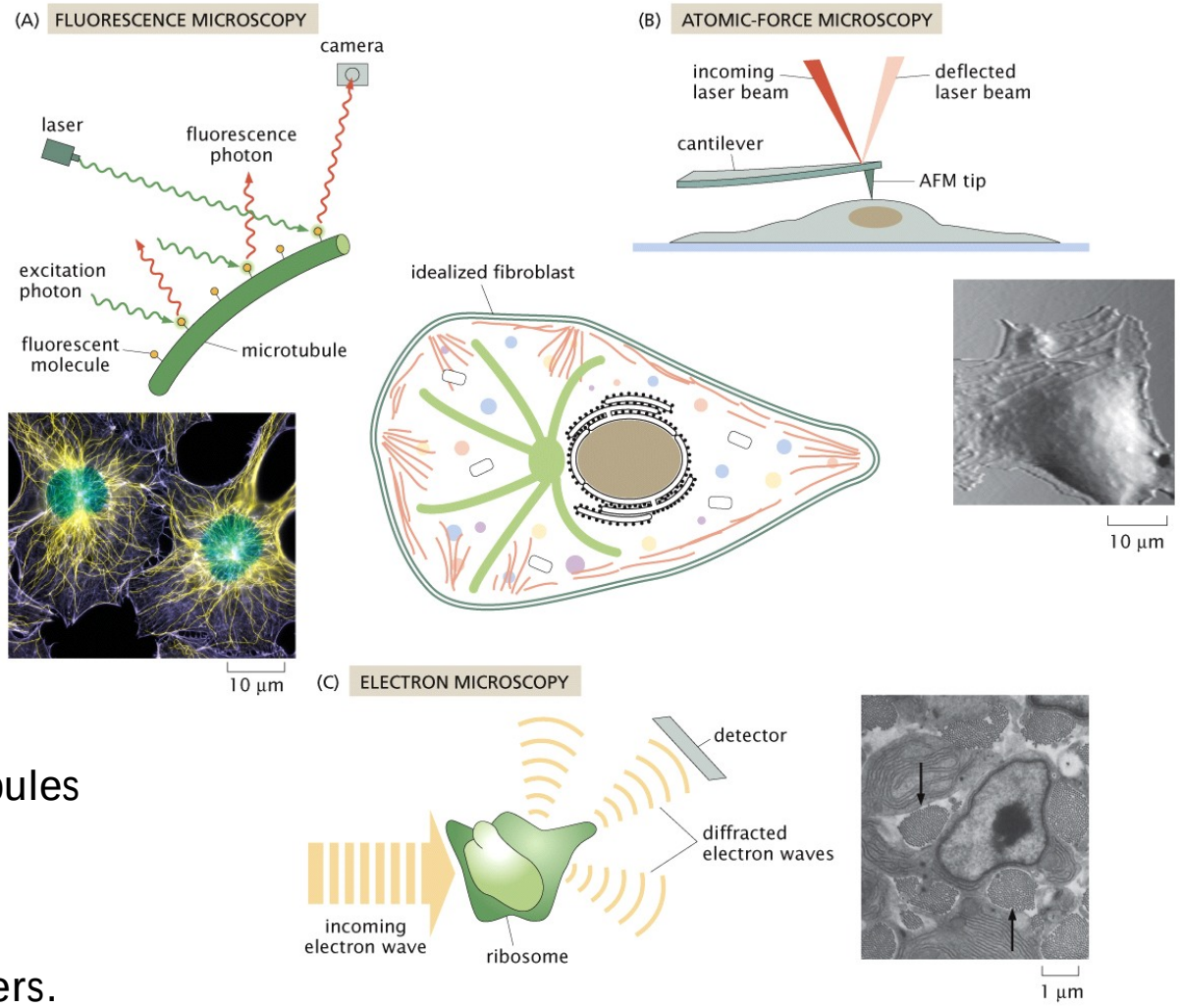


Figure 2.9 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

(A) Difference in mRNA # between two daughters of a mother with *N* mRNAs. (B) Difference in fluorescence of two daughters cells for a fluorescent fusion
to a repressor protein as a function of intensity of the mother cell. The line corresponds to binomial partitioning model.

(A, adapted from I. Golding et al., *Cell* 123:1025, 2005; B, adapted from N. Rosenfeld et al., *Science* 307:1962, 2005.)

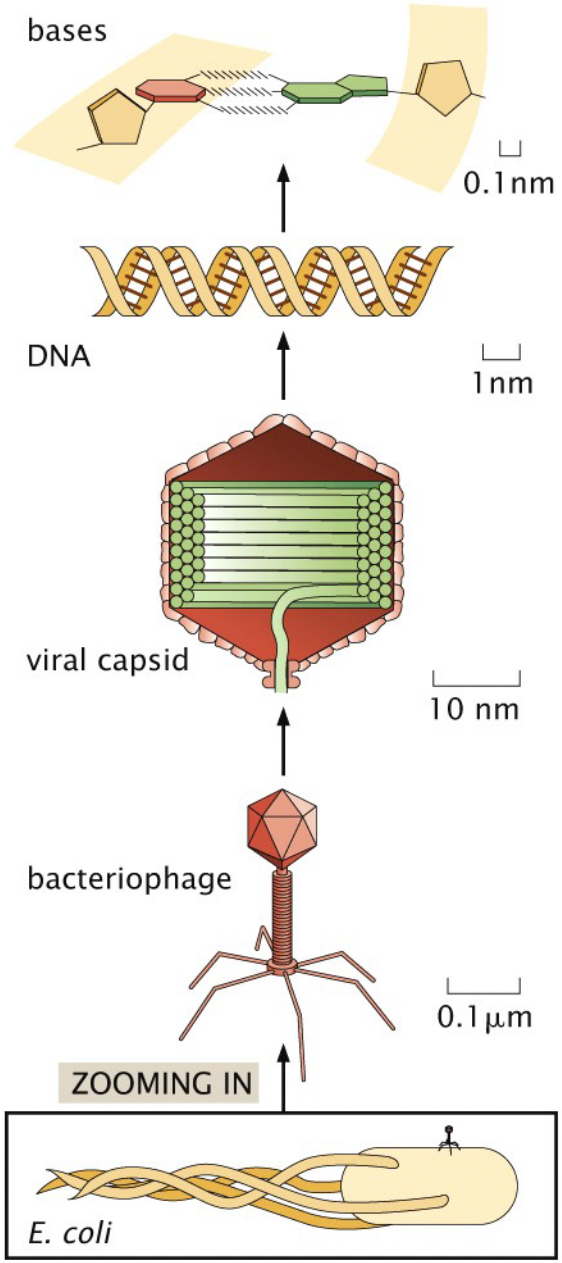
Experimental techniques revealed structure of cells and their organelles

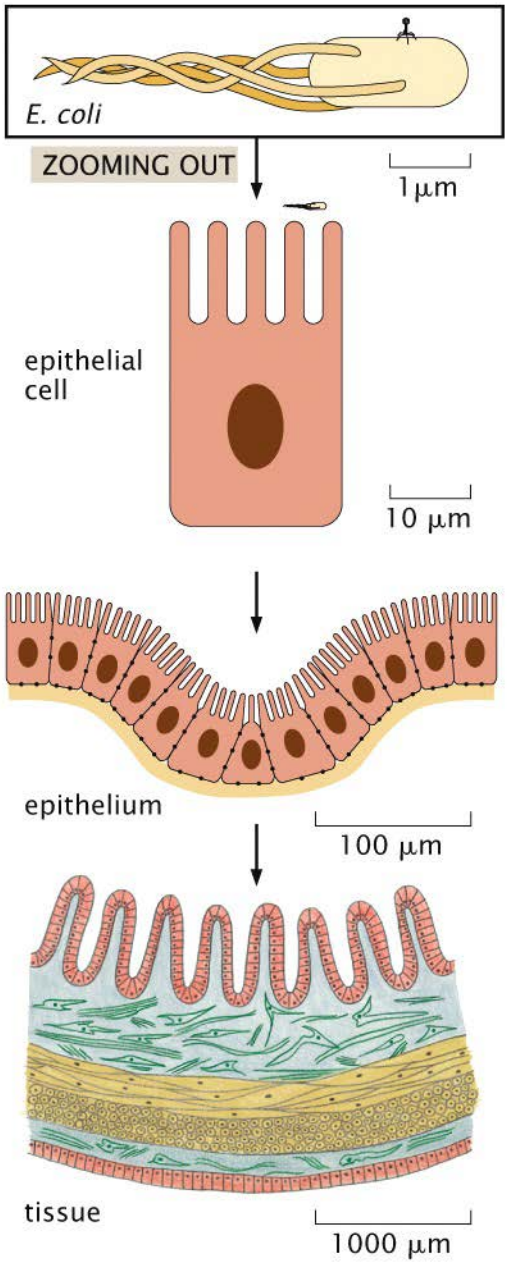


(A) Fluorescence microscopy: fibroblast with labeled microtubules (yellow) and DNA (green).
(B) Atomic-force microscopy.
(C) Electron microscopy: Arrows indicate bundles of collagen fibers.

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

E. Coli in a hierarchy of biological scales





- **Eukaryotic** cell =

DNA genome in a membrane bound nucleus.

Most bacteria and archaea lack nucleus and other elaborate intracellular membrane-bound structures such as the endoplasmic reticulum Golgi apparatus.

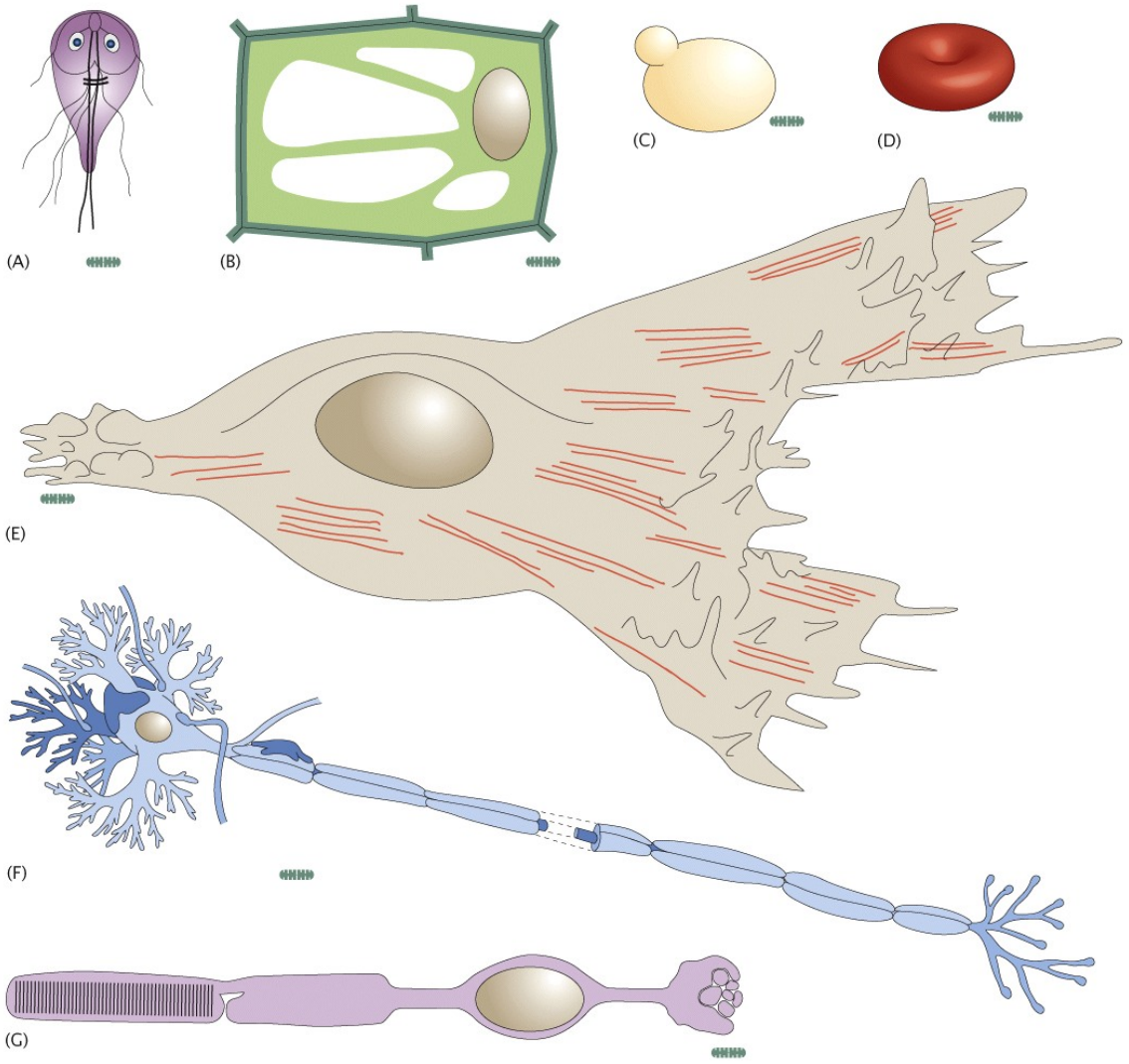


Figure 2.16 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The “helium atom”: the yeast is the simple model eukaryote

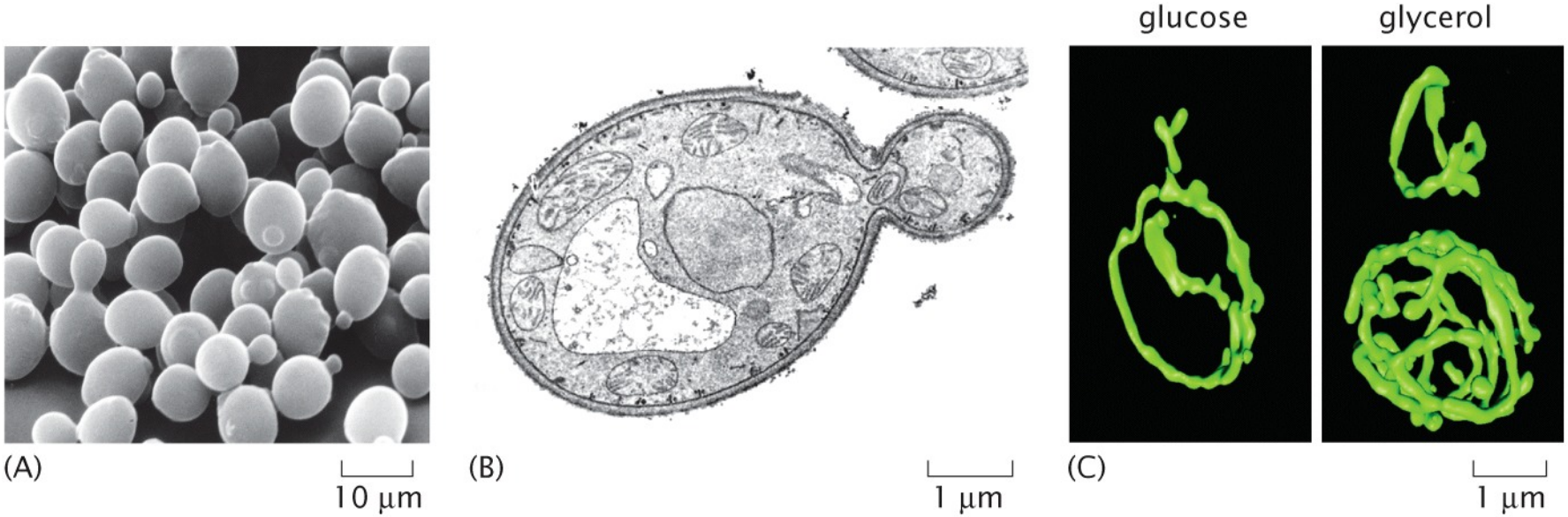


Figure 2.18 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

(A) *Candida albicans* (B) A budding *Candida albicans* yeast cell. (C) Confocal microscopy images of mitochondria in *S. cerevisiae.*

Advantages:

- fungi are close to animals.
- Rapid growth.
- Simple geometry w.r.t. other eukaryotes.

Some yeast numbers

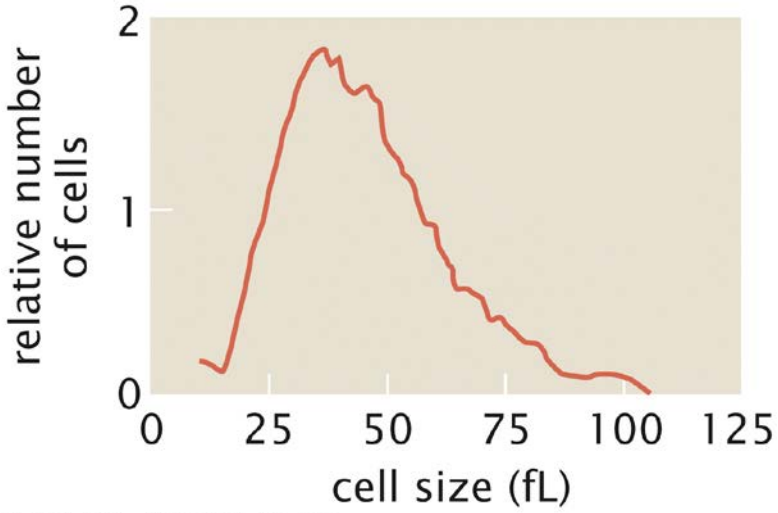


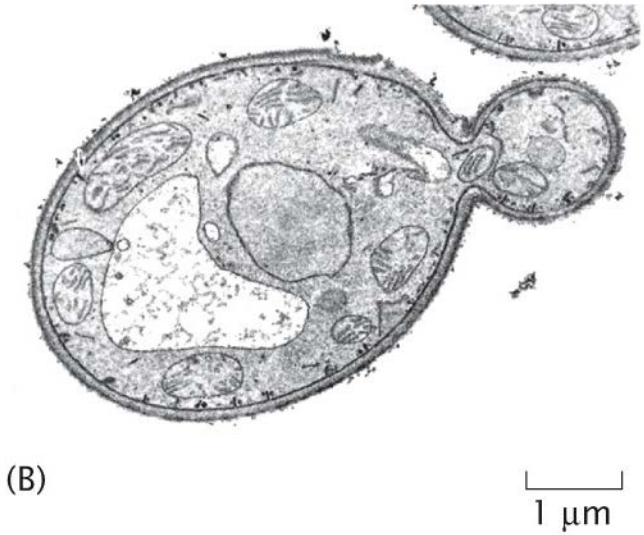
Figure 2.19 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

- Yeast diameter D = 5 μm
- $V_{\text{Yeast}} = \frac{4\pi}{3}\left(\frac{D}{2}\right)^3 \approx 4 \times 2.5^3 = 60\ \mu\text{m}^3 = 60 \times V_{\text{Ecoli}}$
- $S_{\text{Yeast}} = 4\pi\left(\frac{D}{2}\right)^2 \approx 80\ \mu\text{m}^3 = 13 \times S_{\text{Ecoli}}$
- cytoplasmic composition is similar

$\Rightarrow$

$N_{\text{proteins}}^{\text{Yeast}} \approx 60 \times N_{\text{proteins}}^{\text{Ecoli}} \approx 60 \times 3 \cdot 10^6 = 2 \cdot 10^8$ proteins

$N_{\text{lipids}}^{\text{Yeast}} \approx 13 \times N_{\text{lipids}}^{\text{Ecoli}} \approx 13 \times 2 \cdot 10^7 = 3 \cdot 10^8$ lipids



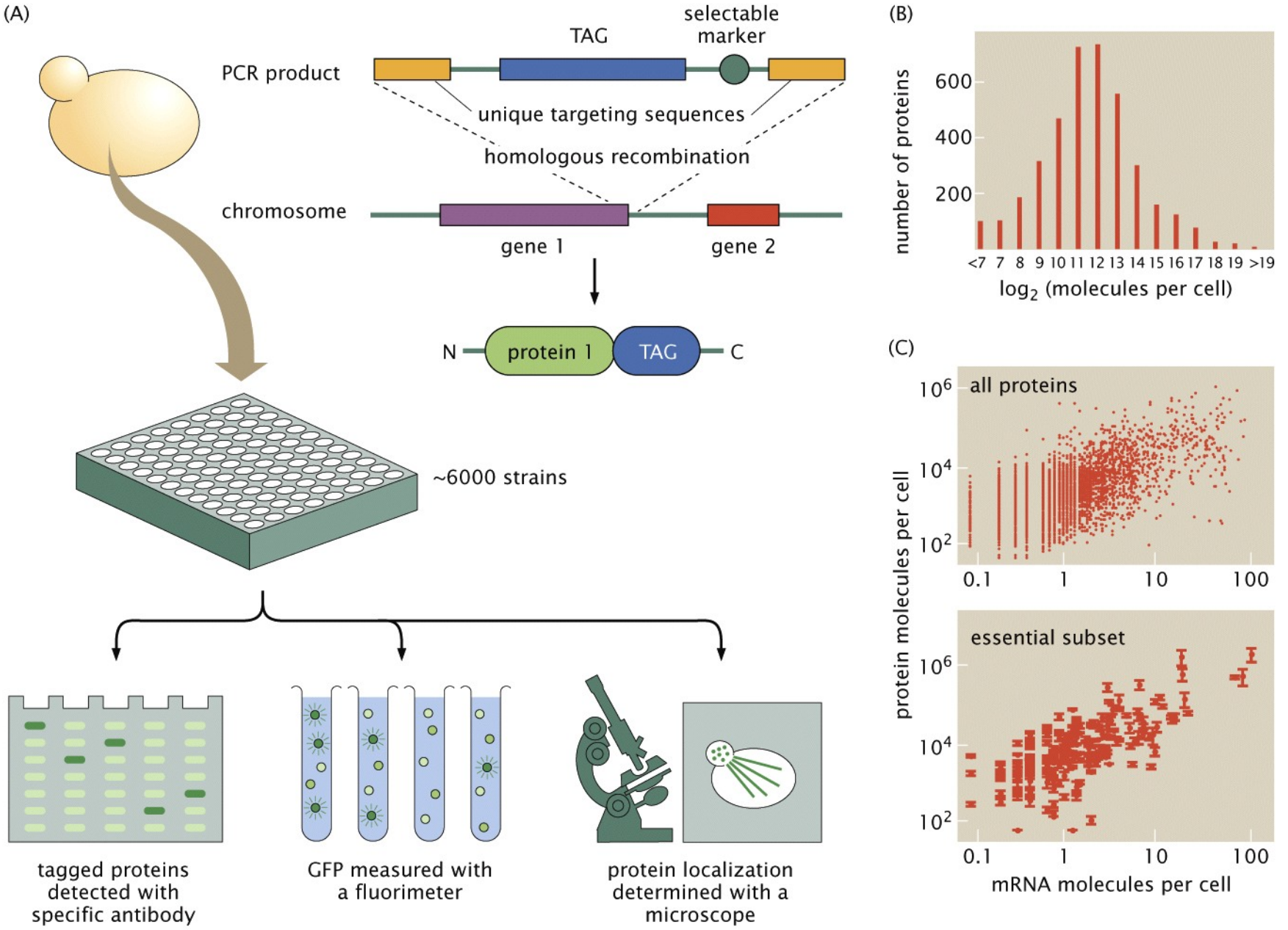


Figure 2.20 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Eukaryotic cell organelles: Nucleus

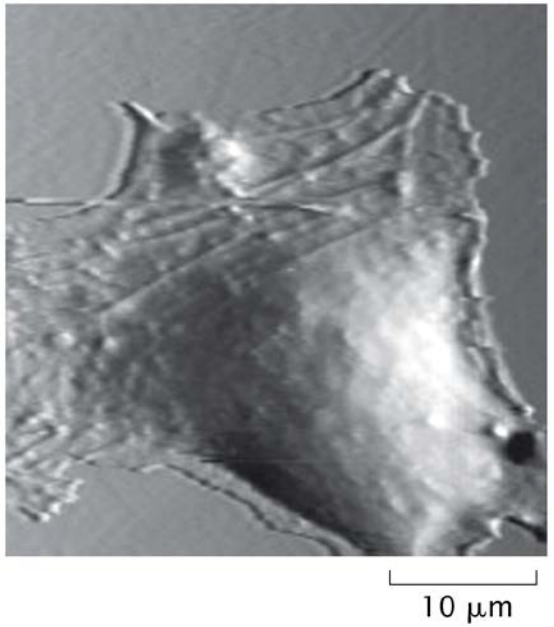
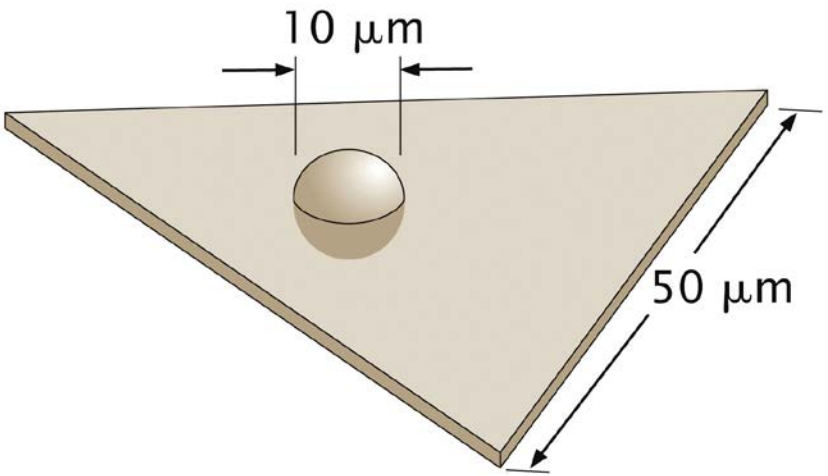
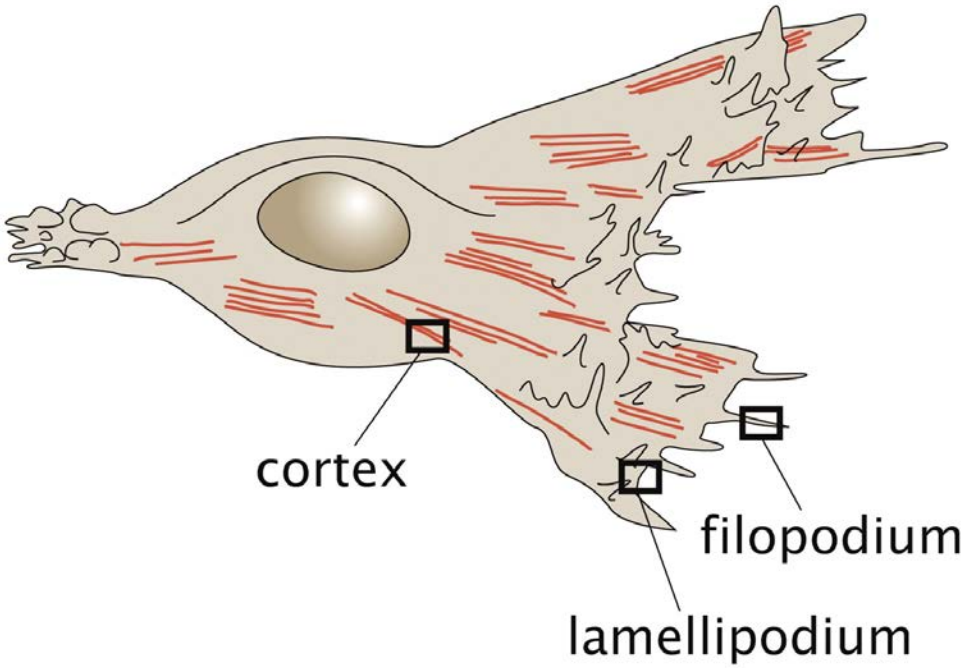


Figure 2.22a Physical Biology of the Cell, 2ed. (© Garland Science 2013)





The Eukaryotic cell and its organelles

Electron microscopy image shows prominent organelles: mitochondria, lysosomes, the rough endoplasmic reticulum, and the Golgi apparatus.

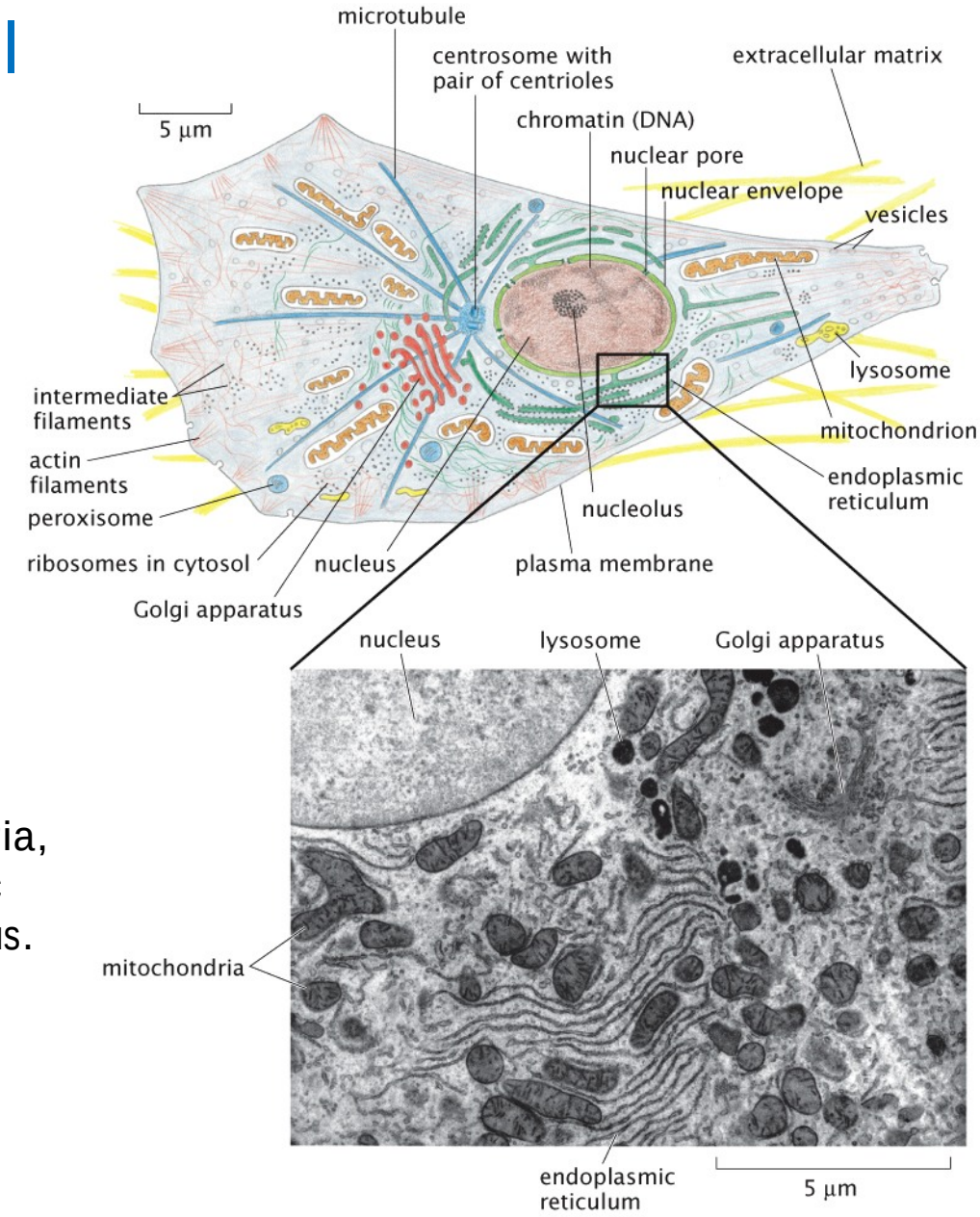


Figure 2.23 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Macromolecules form Assemblies

(A) **Nucleosome** = complex of 8 protein units and DNA.

(B) **Nuclear pore complex** mediates transport to/from nucleus.

(C) **Replisome** is mediates the copying of the genetic material.

(D) **Ribosome** reads mRNA and synthesizes polypeptide chain.

(E) **Proteasome** degrades proteins.

(F) **ATP synthase** synthesizes ATP.

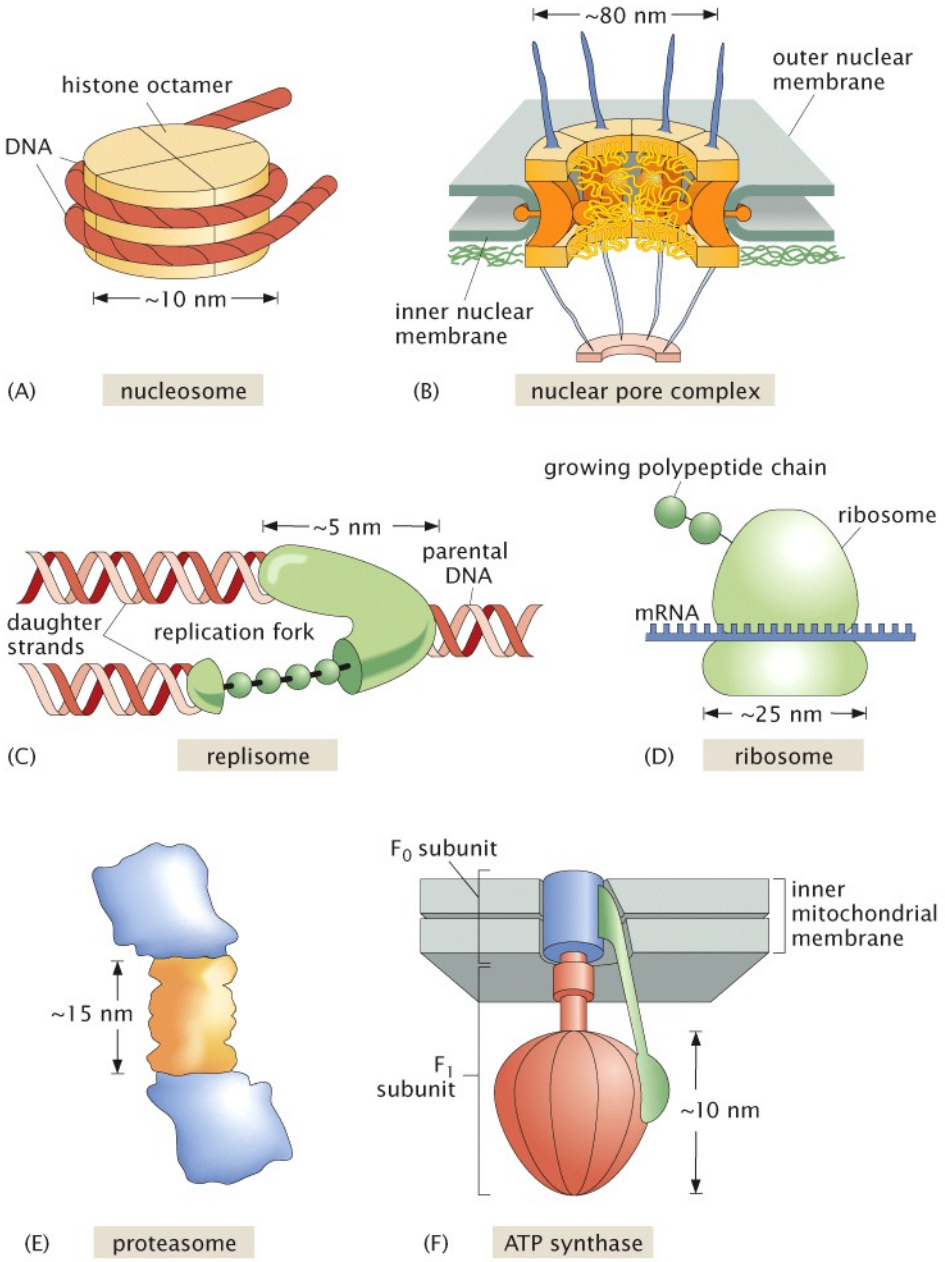


Figure 2.26 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Helical motifs are abundant in molecular assemblies

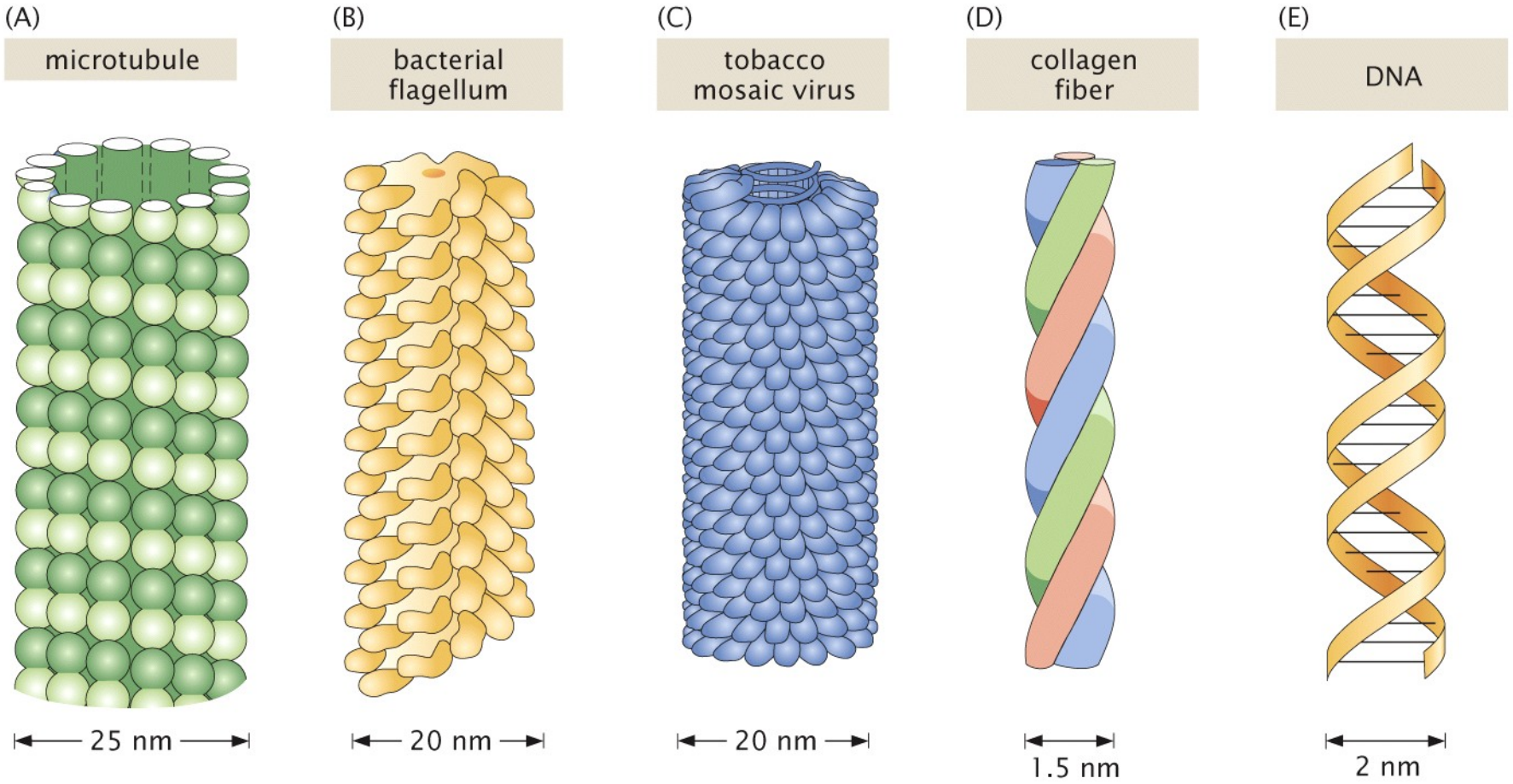


Figure 2.27 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Macromolecular assemblies in superstructures

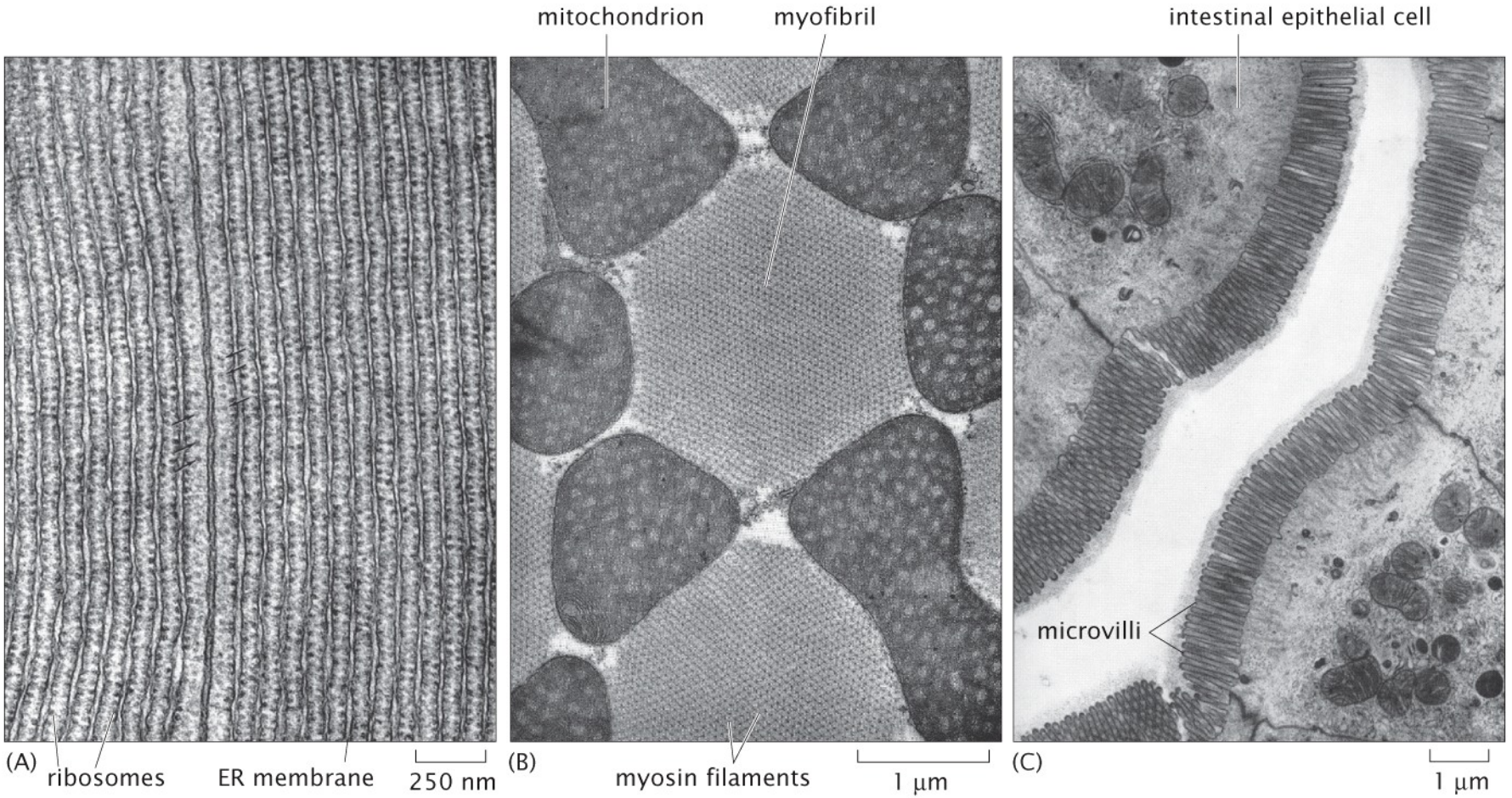


Figure 2.28 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Macromolecular assemblies in superstructures of viruses

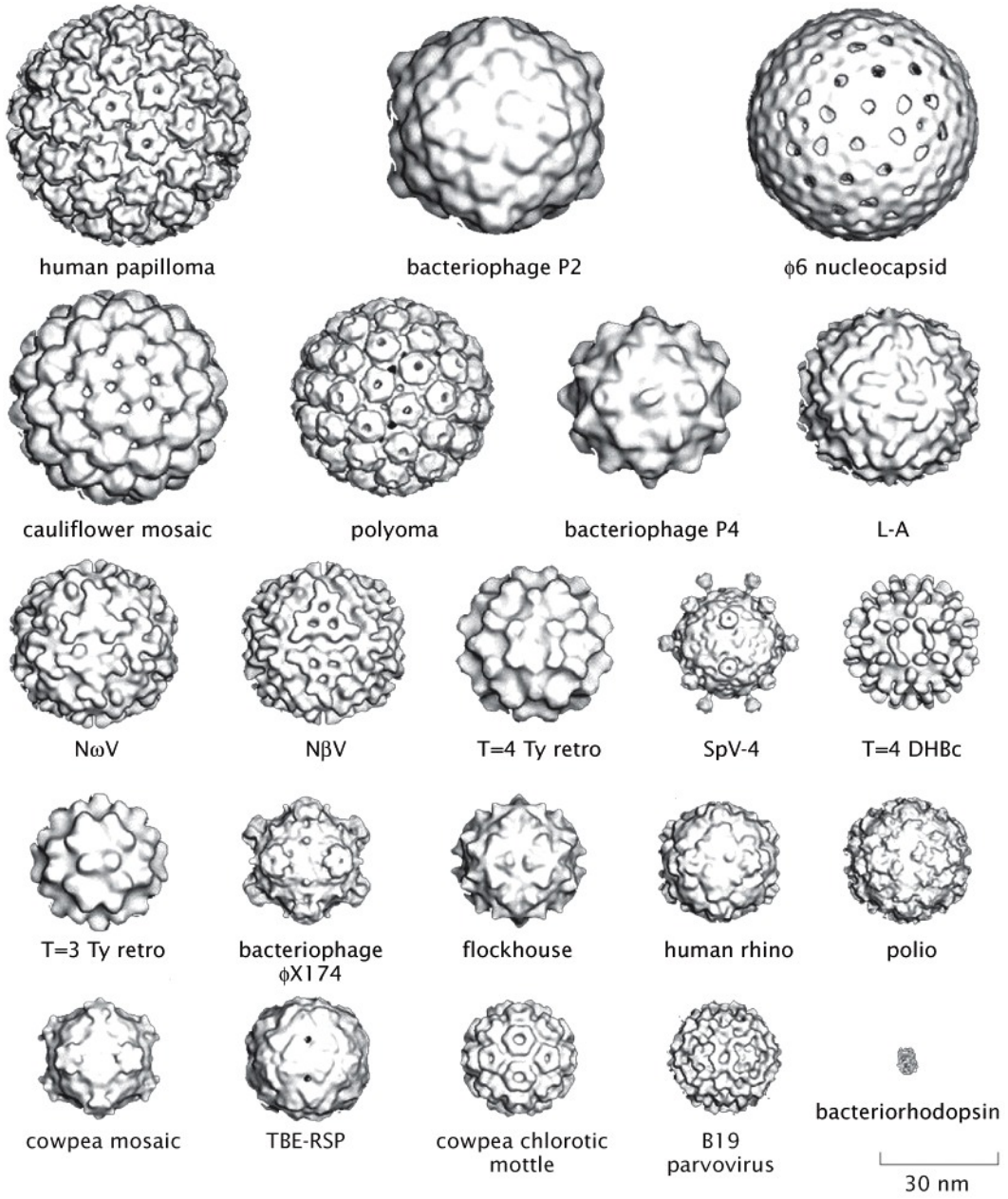
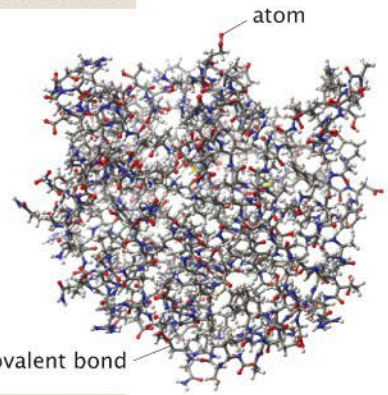


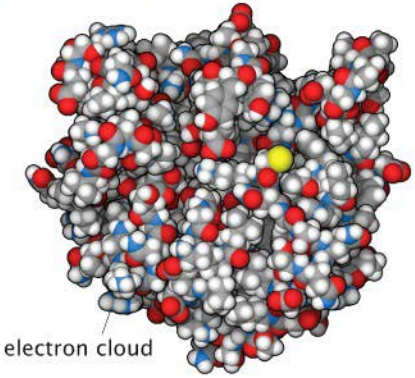
Figure 2.29 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Proteins are superstructures of amino acids

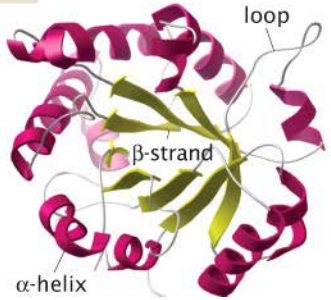
ball and stick

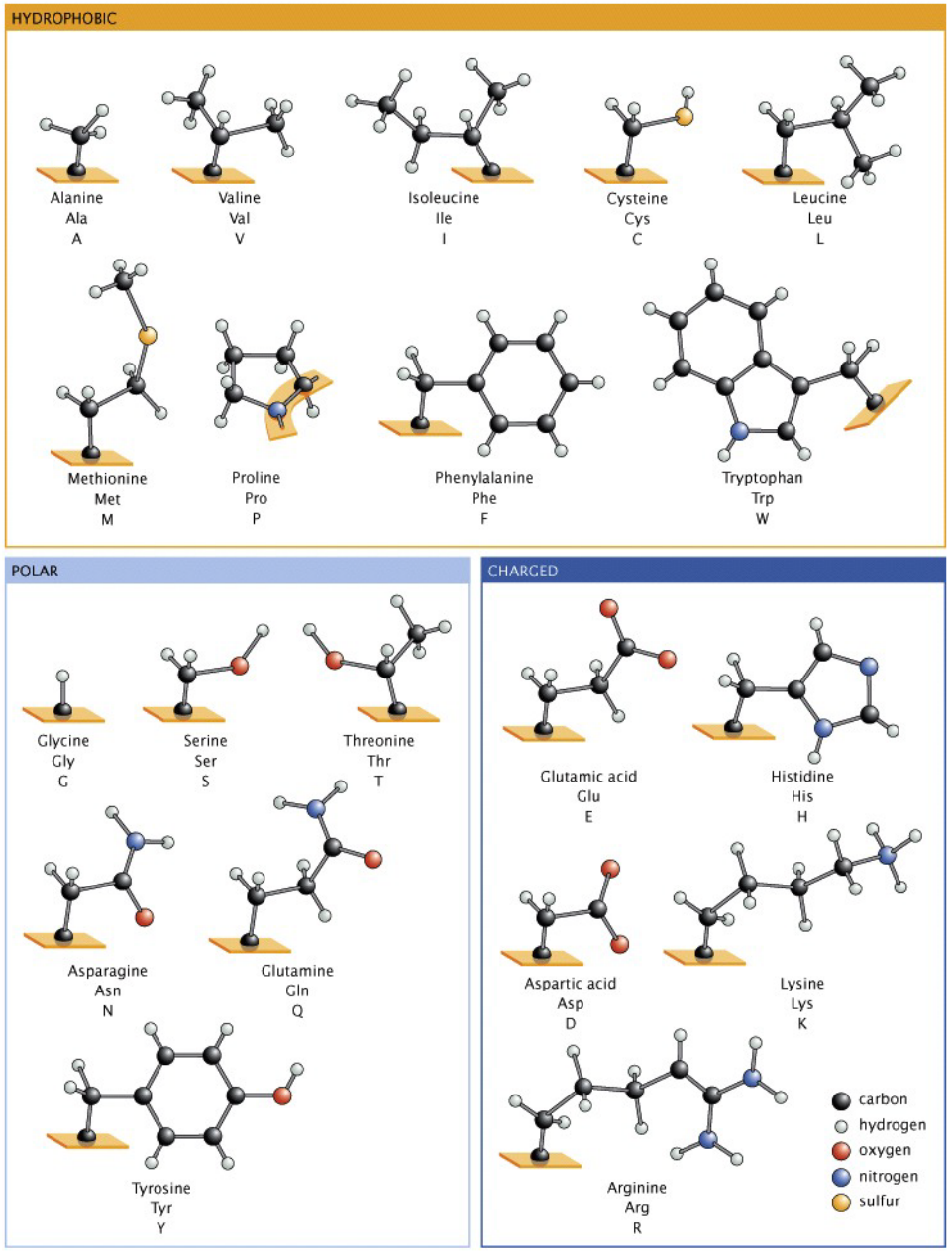


space-filling



ribbon





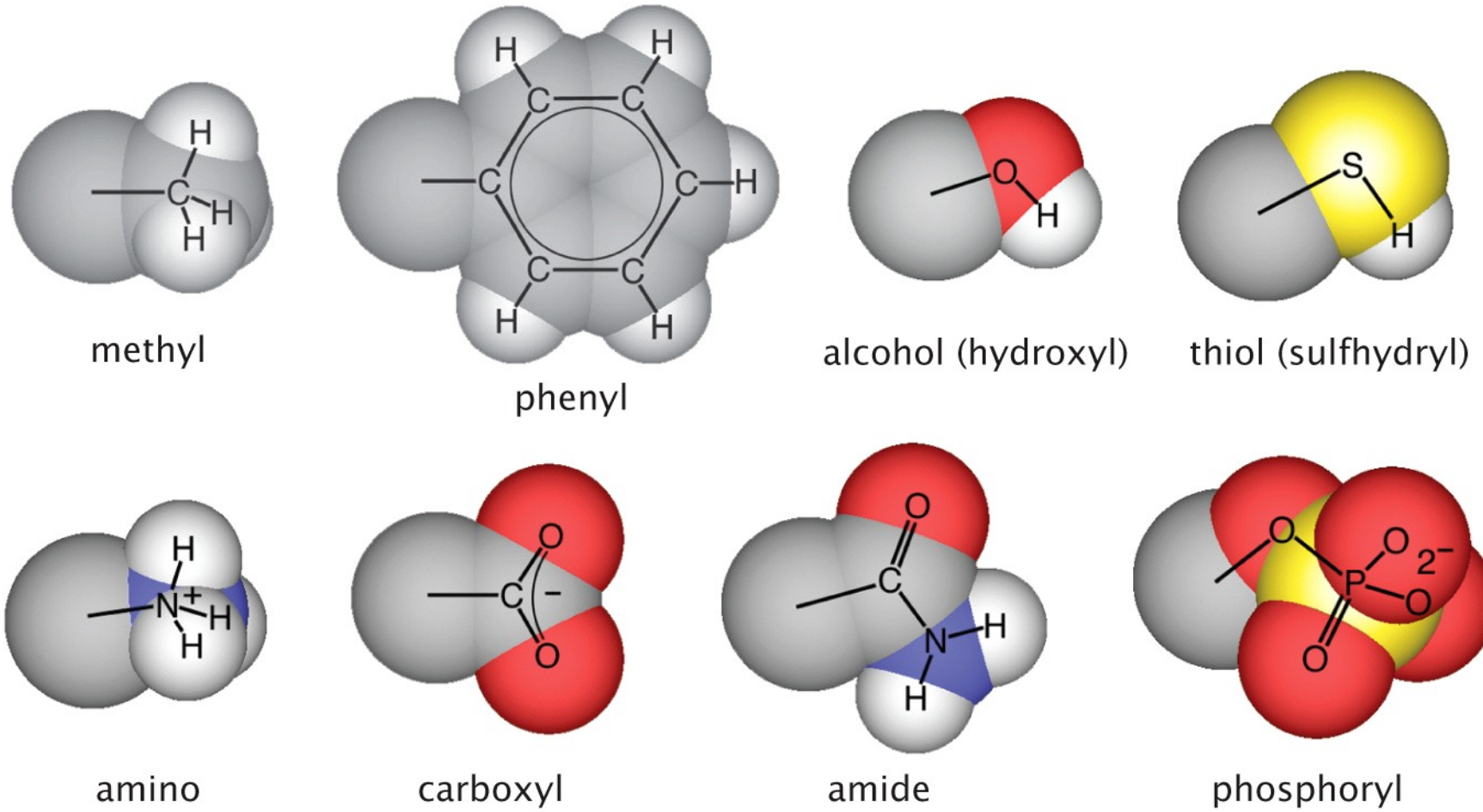


Figure 2.34 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Scaling up to multicellular organization

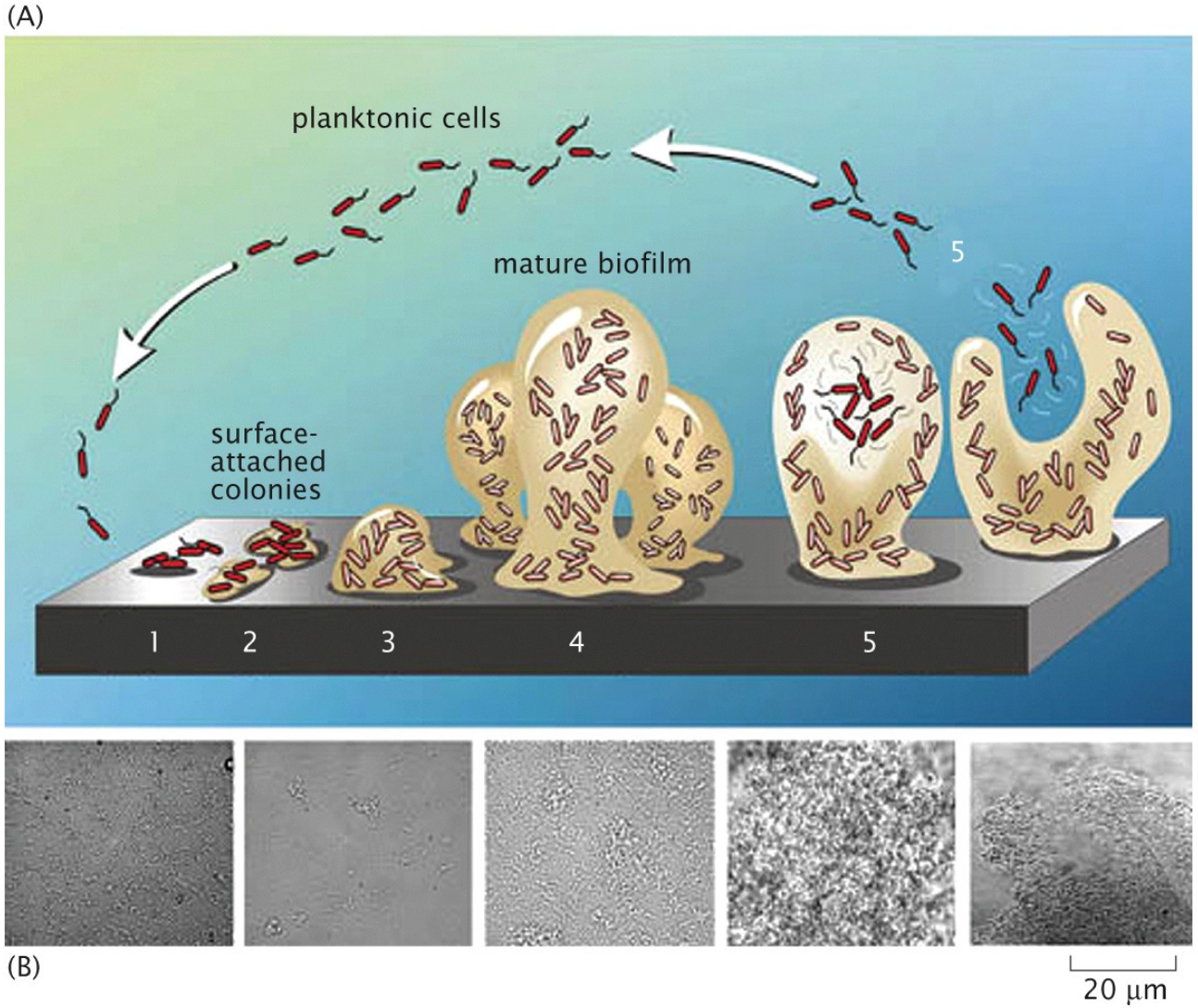


Figure 2.36 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Tissues Are Collections of Cells and Extracellular Matrix

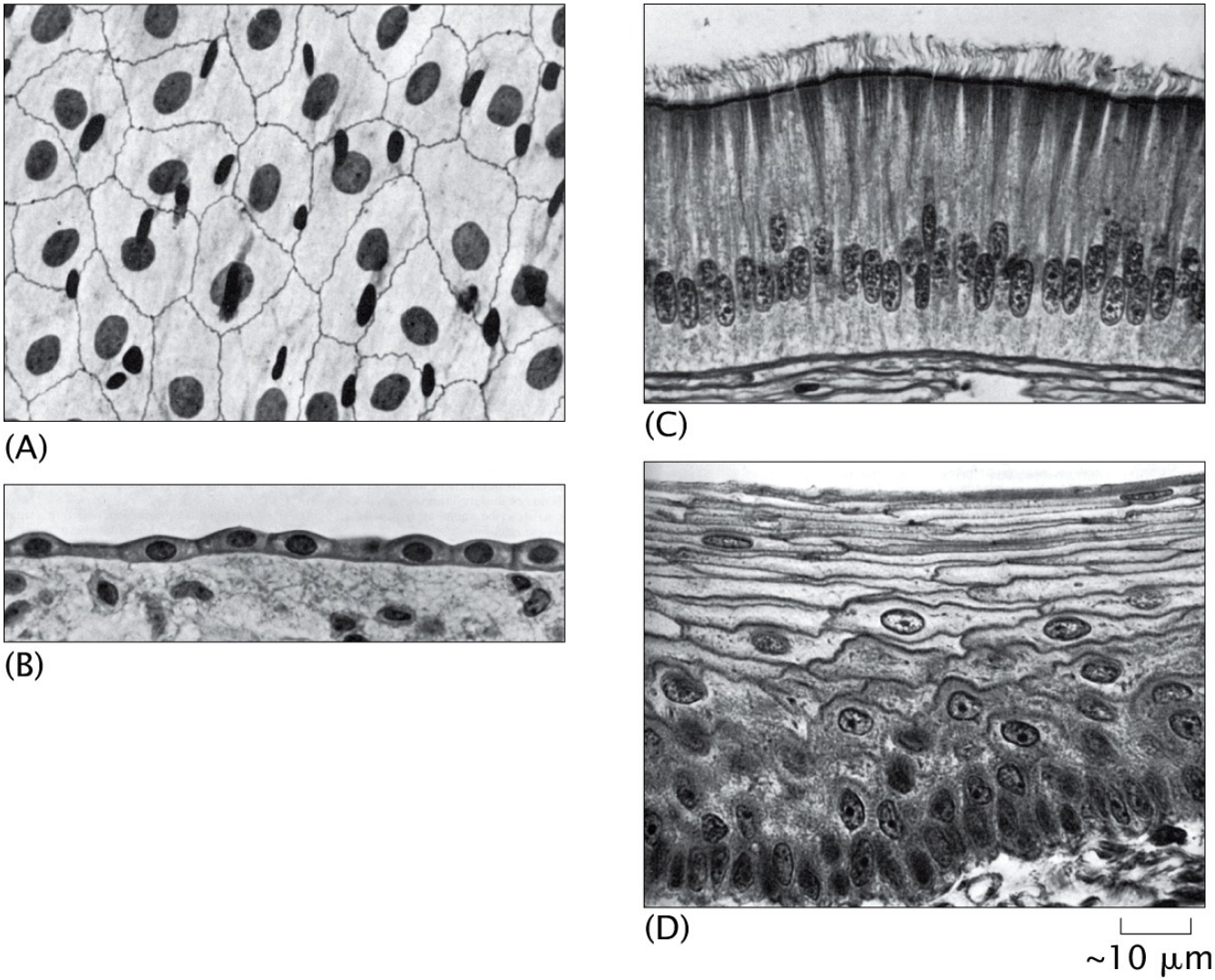
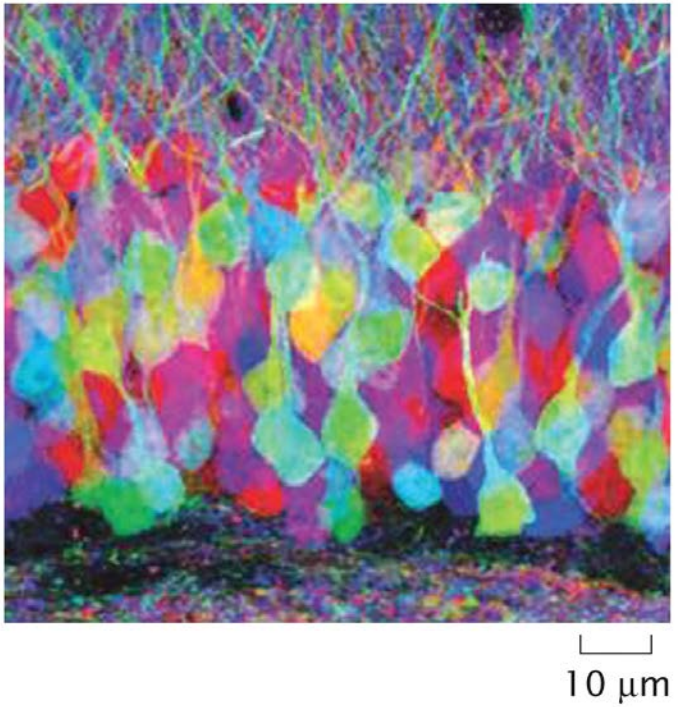


Figure 2.40 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Nerve cells form complex, multicellular structures



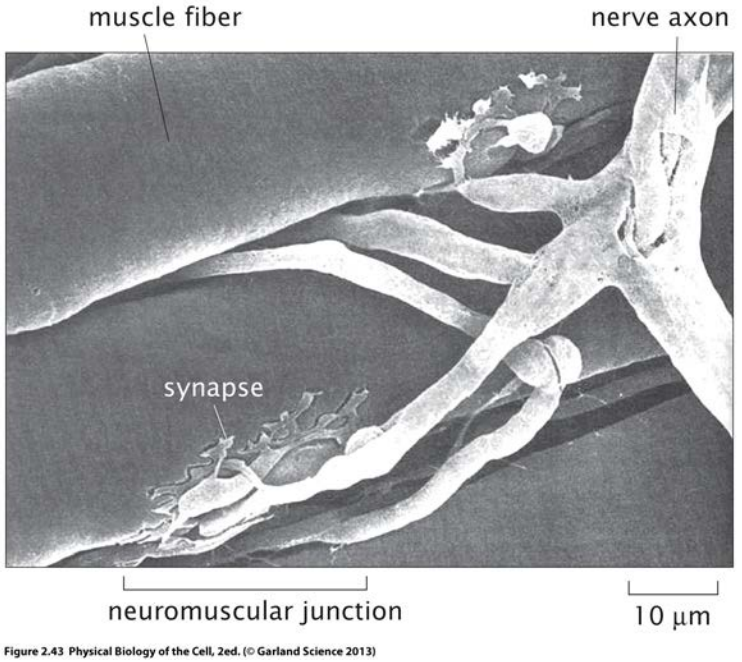
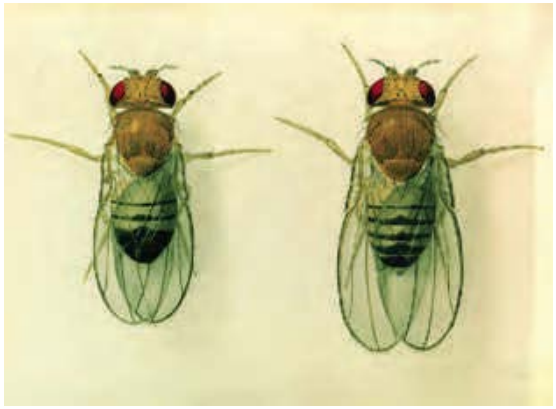


Figure 2.43 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Cells differentiate during development leading to entire organisms



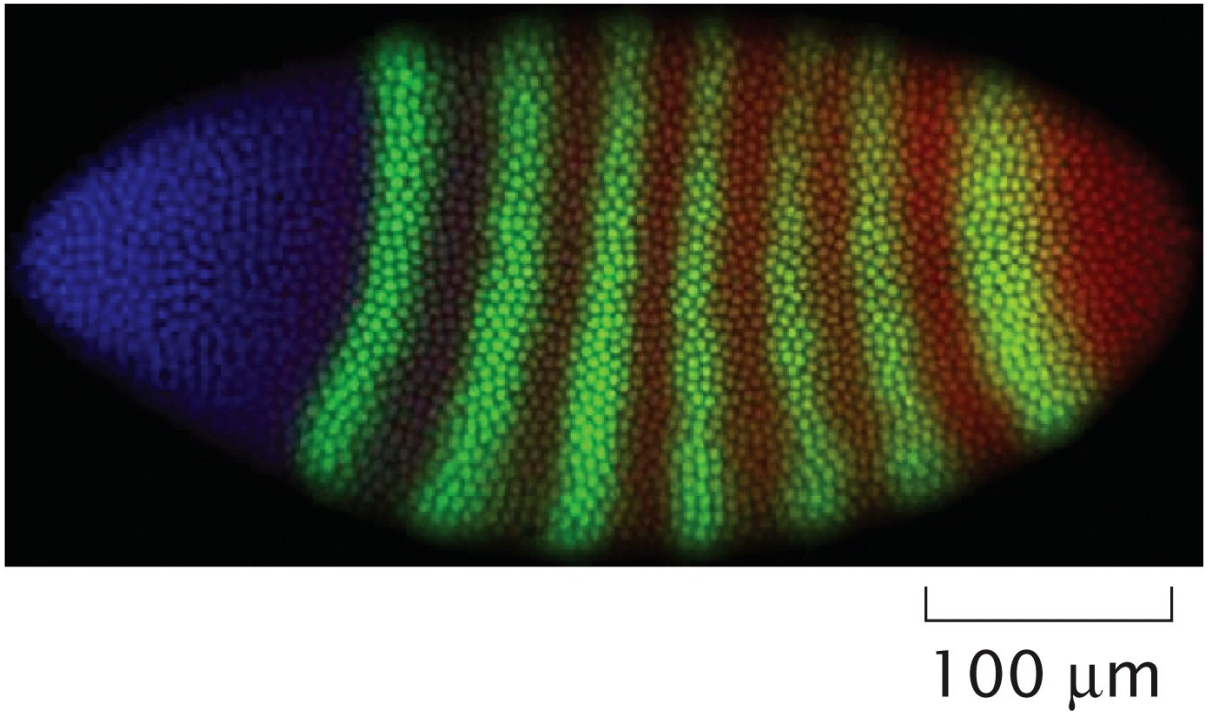


Figure 2.45 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Cells of the nematode worm, *Caenorhabditis Elegans*, fully charted, give a cell-by-cell picture of the organism

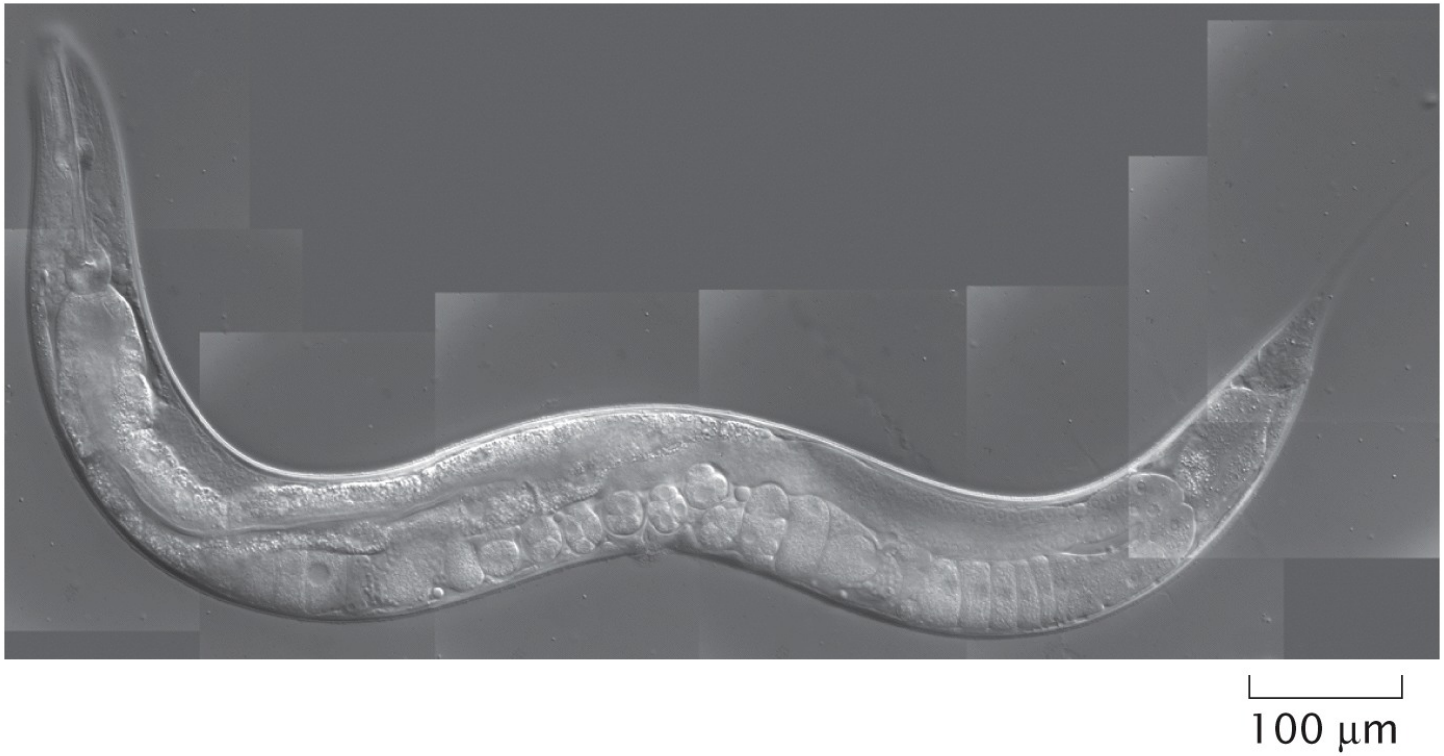


Figure 2.46 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

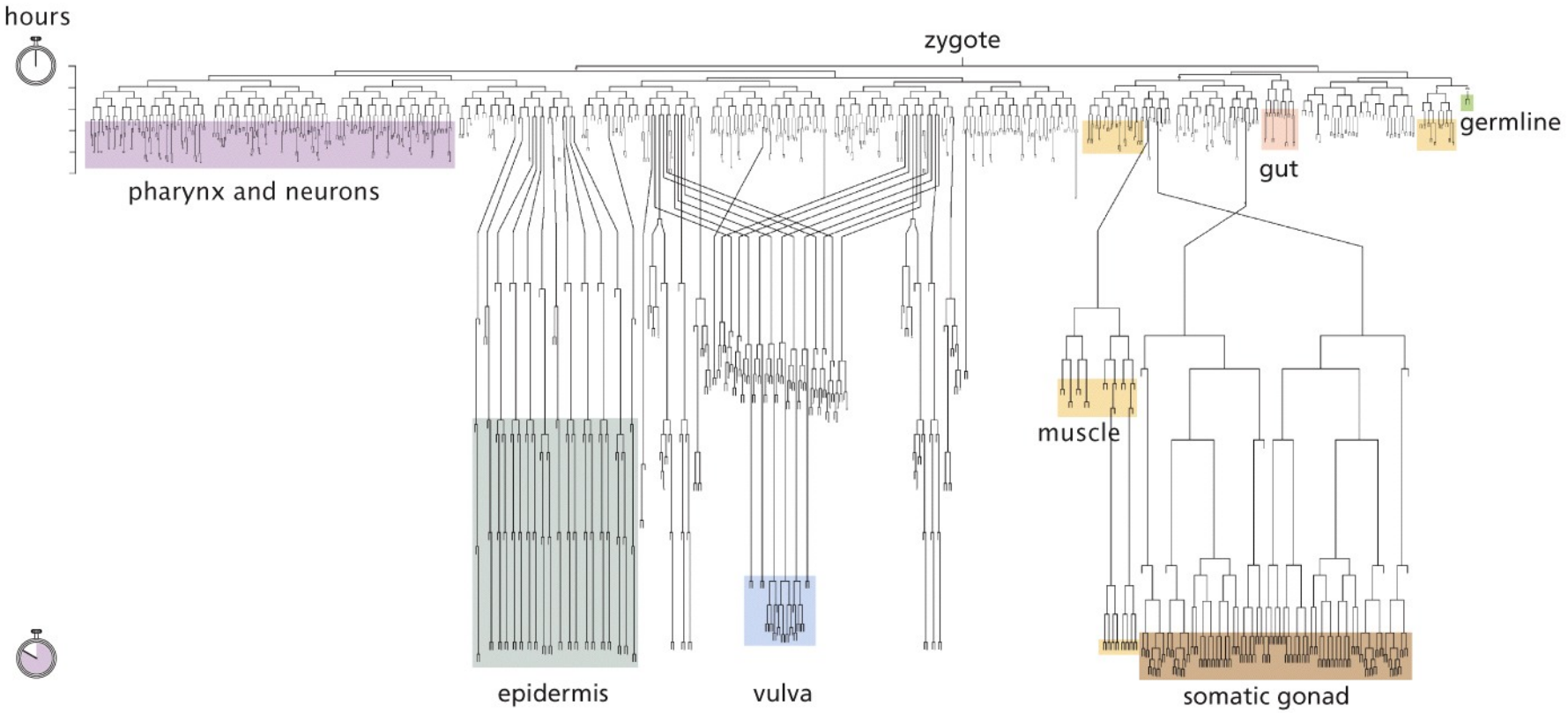


Figure 2.47 Physical Biology of the Cell, 2ed. (© Garland Science 2013)