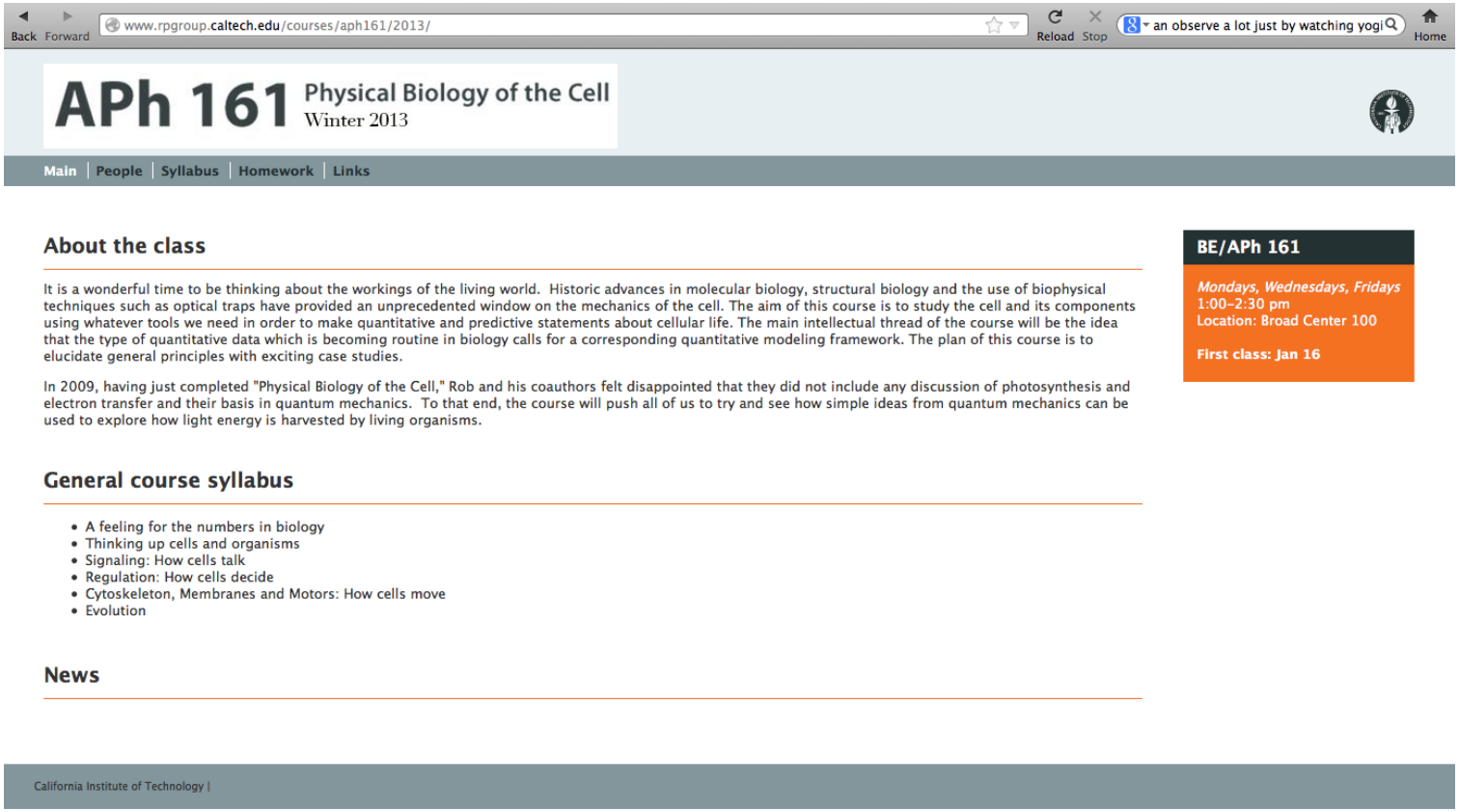
Thinking up a cell

Rob Phillips
Applied Physics and Bioengineering
California Institute of Technology

Procedural issues



- ***Course meeting times and makeups, grading, TAs, …***
- ***Why now and what material?***

Beyond the telescope: new instruments and new science

- ***The spectroscope was an addition to the telescope that made it possible to measure the composition and velocities of stars.***

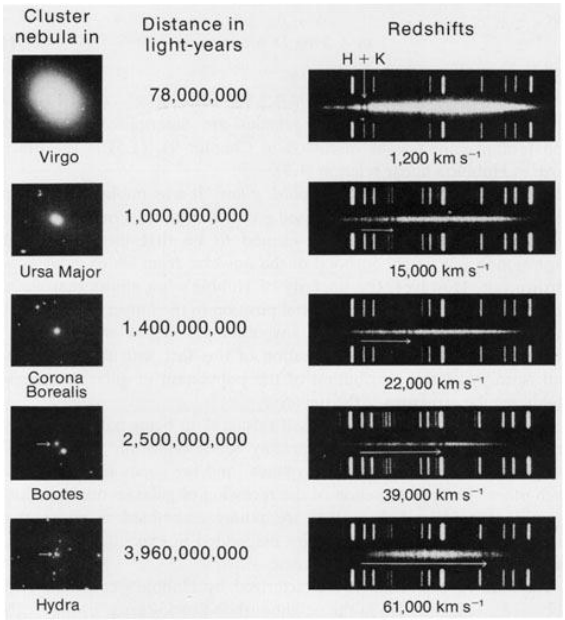
SOME SCIENTIFIC CENTRES.

IV.—THE HEIDELBERG PHYSICAL LABORATORY.

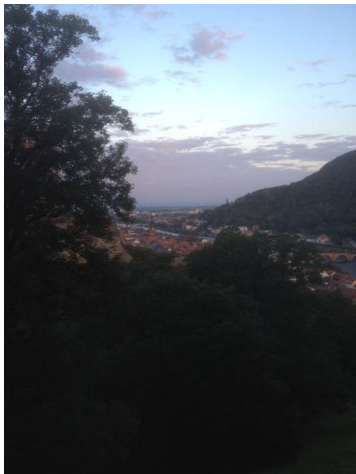
MOST travelled Englishmen are doubtless acquainted with the ancient town of Heidelberg, so famous for the beauty of its situation and the grandeur of its ruined castle. But far fewer know the charms of the long and romantic valley of the Neckar, at the almost sensational exit of which, from the Odenwald into the level plain of the Upper Rhine, Heidelberg stands. So also it is true that while most educated people connect Heidelberg with the great names of Kirchhoff and Bunsen and their epoch-making discoveries in spectrum analysis, it is only the special students who know how large in extent and how important in result and example is the work which has steadily gone on for many years in the physical laboratory in the Friedrichsbau.

Its small beginnings in the middle of the last century are marked by the name of Kirchhoff scratched on the window of what is now the private room of the senior assistant. From this window one may look out over the Rhine plain towards busy Mannheim, as Bunsen and Kirchhoff did one night when a fire was raging there, and they were able by spectroscopic examination of the flames to ascertain that barium and strontium were present in the burning mass. But the same window also looks across the Neckar to the Heiligenberg, along the slopes of which runs the "Philosophers' Walk," the chief of the many paths among the wooded hills around the town, which the two friends were wont to traverse in their daily "constitutionals." Bunsen is known to have said that it was during such walks that his best ideas came to him. One day the thought occurred, "If we could determine the nature of the substances burning at Mannheim, why should we not do the same with regard to the sun? But people would say we must have gone mad to dream of such a thing." All the world knows now what the result was, but it must have been a great moment when Kirchhoff could say, "Bunsen, I *have* gone mad," and Bunsen, grasping what it all meant, replied, "So have I, Kirchhoff!"

Kirchhoff's four-prism spectroscope, together with other apparatus of his, is preserved in the collections of the Laboratory, and well deserves the almost reverential awe with which it was examined by a certain foreign professor, who protested that objects of such historic interest should be kept in a fire-proof safe.

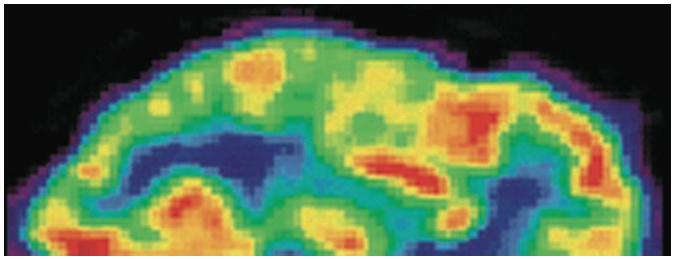


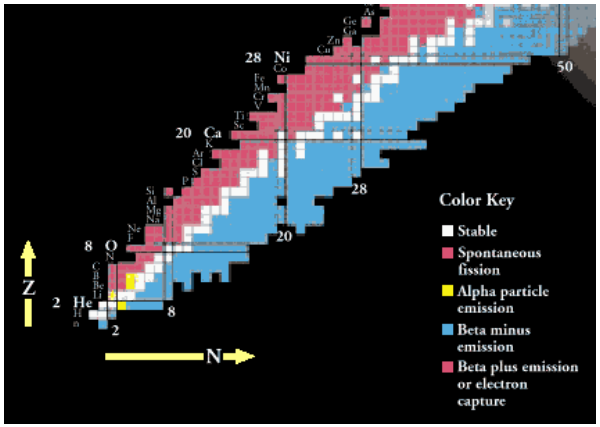
What are biology's equivalent of the spectroscope?

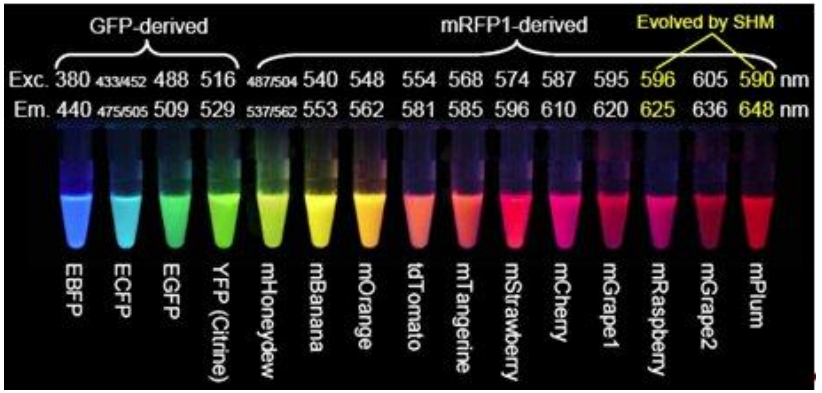


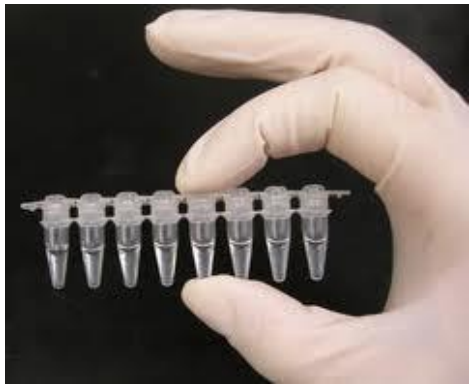
Some day people may look back on the isotope as being as important to medicine as the microscope – archibold Hill

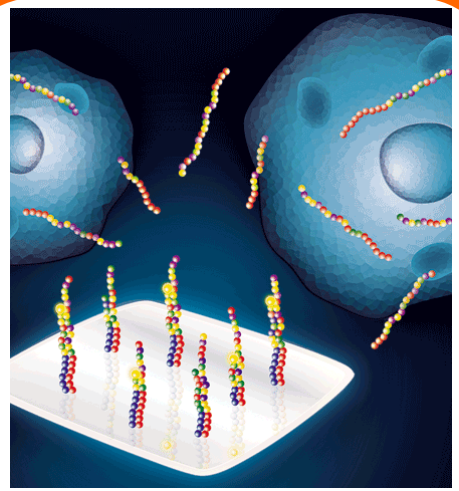
- ***Tools that have revolutionized biology.***

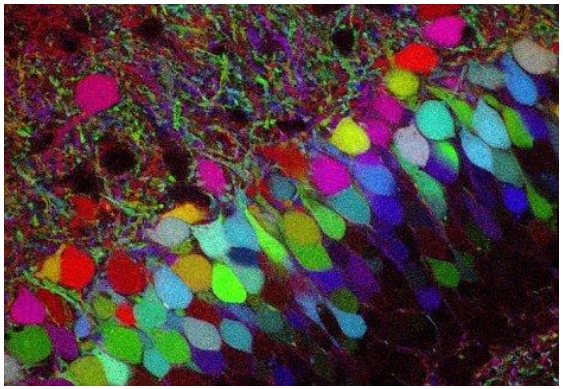








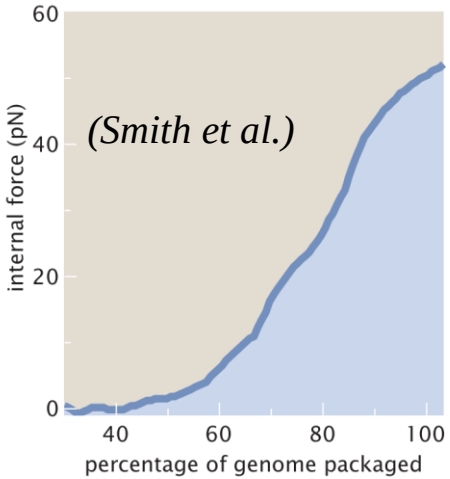




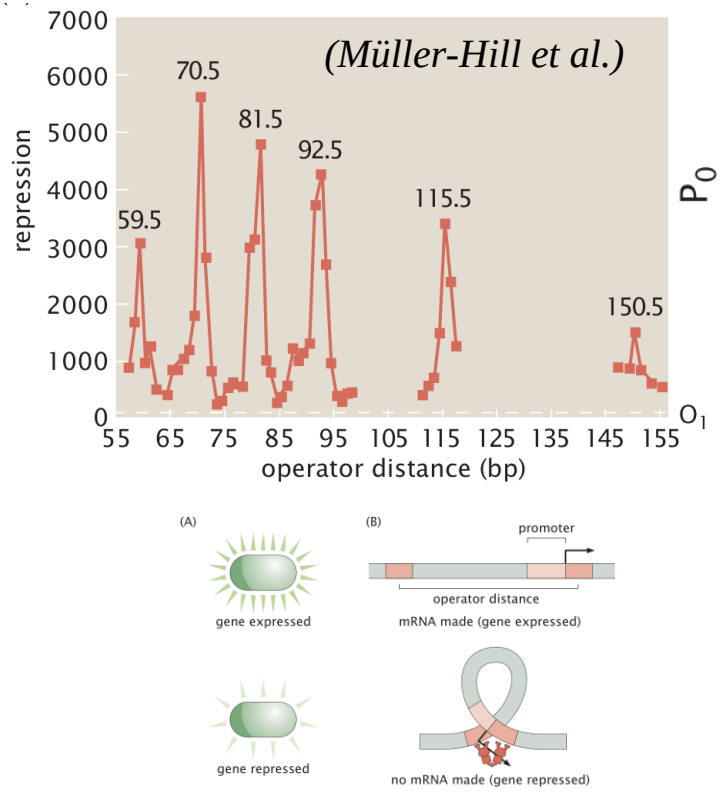
Why now?

- ***Often, biological data reports on functional relationships like those that are the lifeblood of physics.***
- ***Data of this variety imposes much stricter demands on biological theory. No amount of words or cartoons suffice to describe such data.***

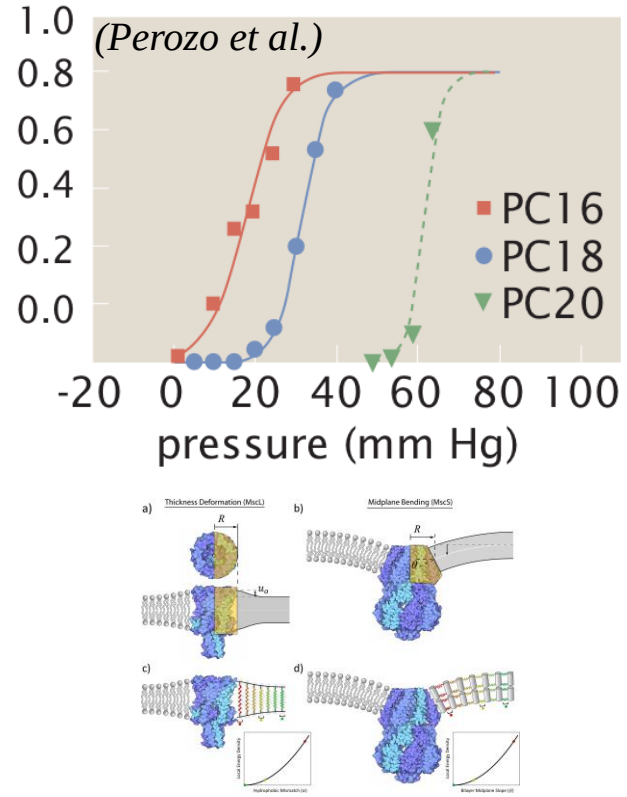
Genome Management



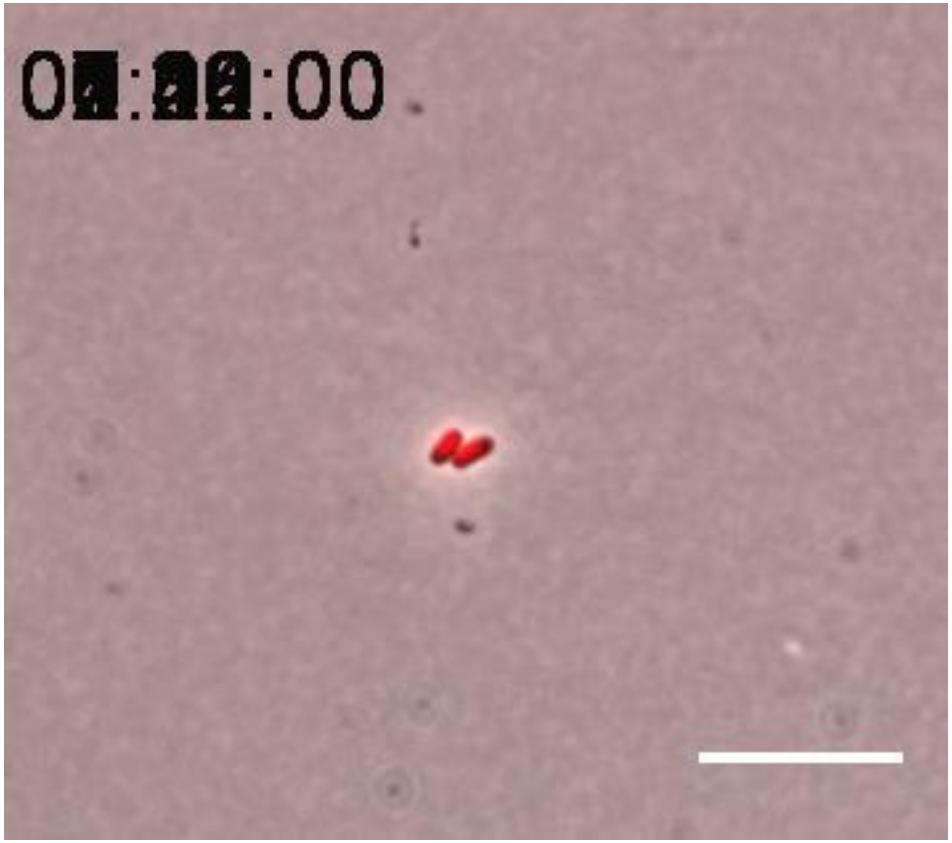
Gene regulation



Ion Channel Gating



“You can observe a lot just by watching” – Yogi Berra



Courtesy of Linda Song and Hernan Garcia

"You can observe a lot just by watching" – Yogi Berra

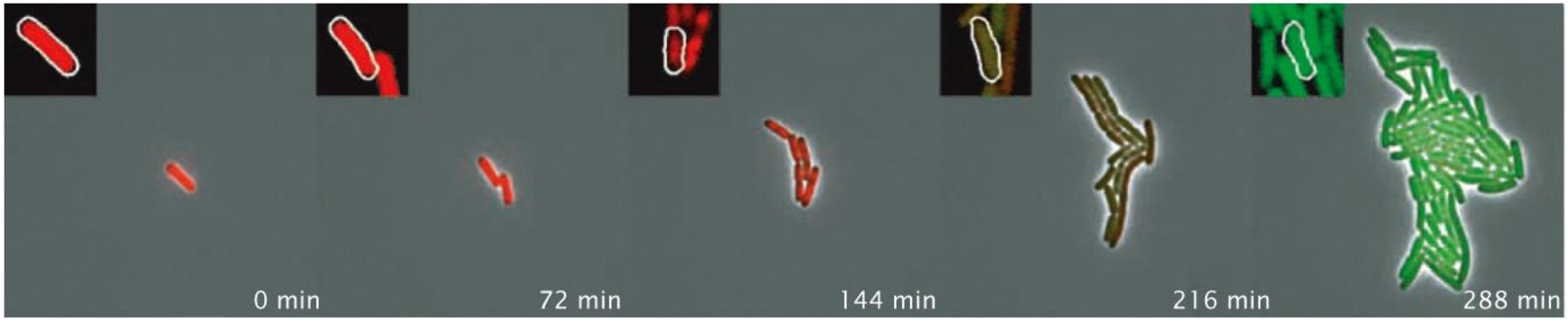
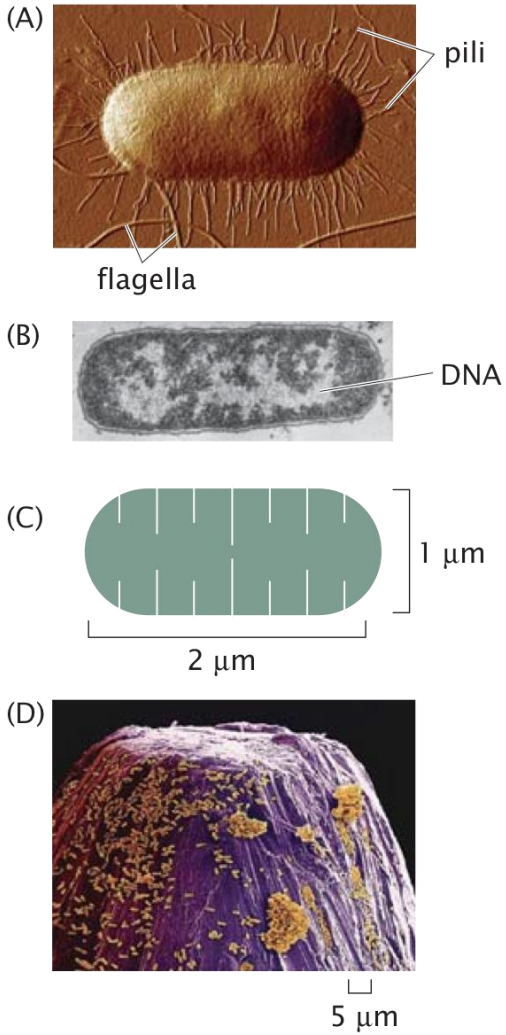


Figure 19.15c Physical Biology of the Cell, 2ed. (© Garland Science 2013)

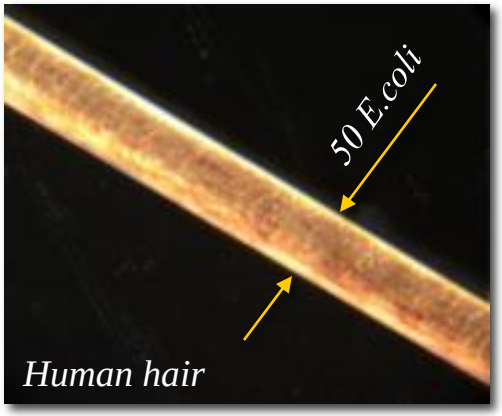
Beautiful paper of Rosenfeld, Young, Swain, Elowitz and Alon, Science 2005

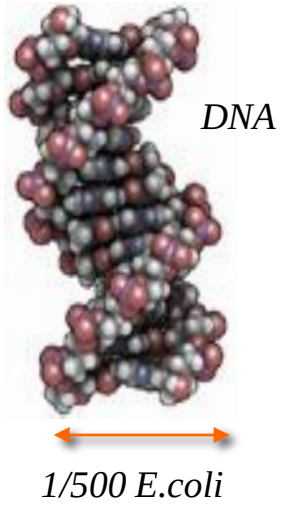
e.coli as The bacterial standard ruler

- All cells share with *E.coli* the fundamental biological directive to convert E_{envir} into structural order and to perpetuate their species.
- Min requirements for the perpetuation of cellular life, as observed on Earth:
  - DNA-based genome
  - mechanisms for DNA → RNA $\xrightarrow{ribosomes}$ proteins



E.coli as a standard ruler:

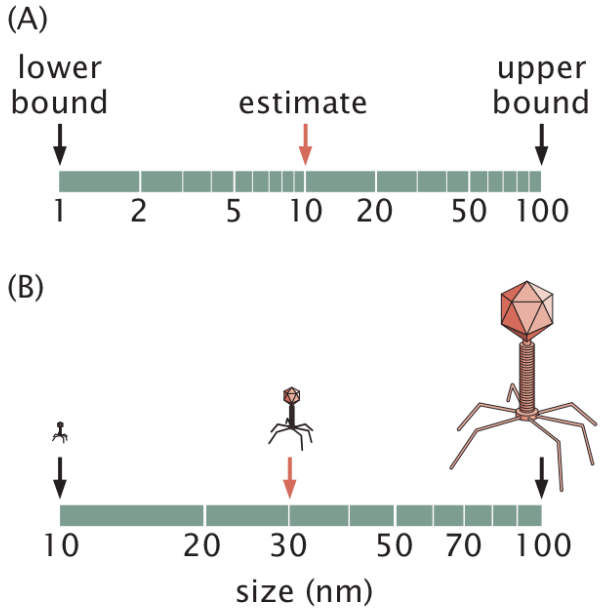


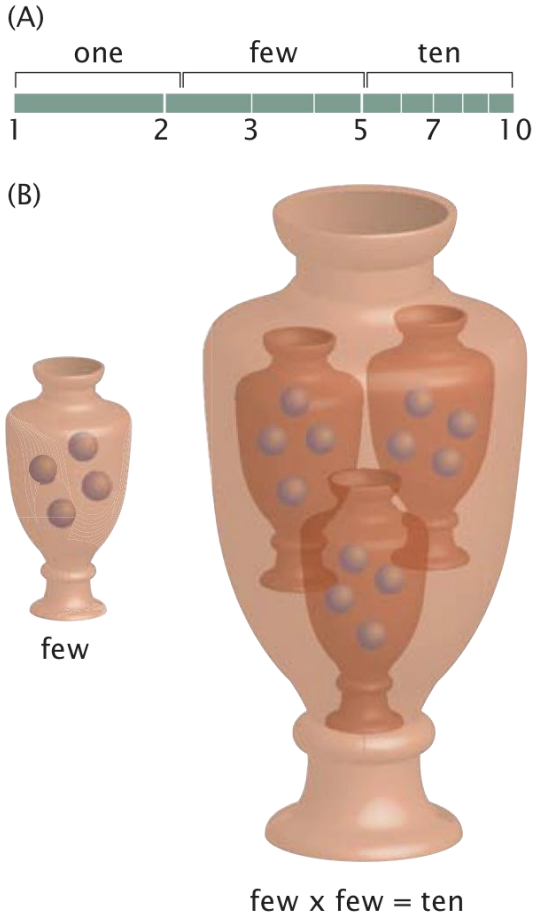


- Note: size of *E.coli* depends on the nutrients provided: richer media => larger size.
- Biochem. studies usually use “minimal medium”: salts+glucose

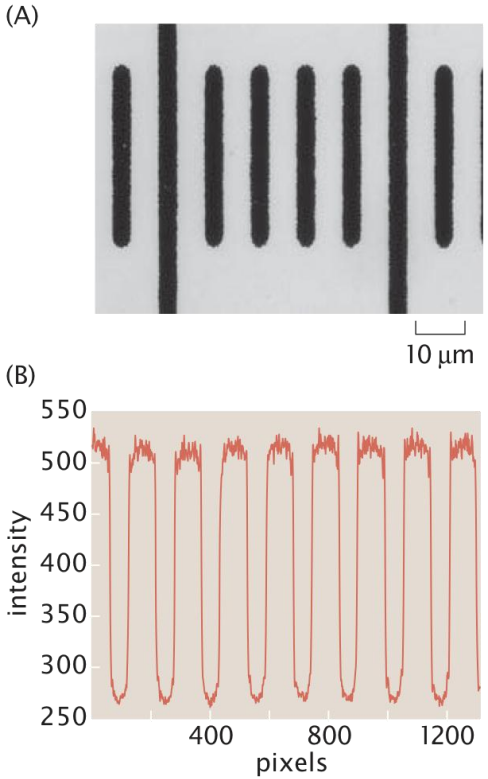
Street fighting mathematics: Rules for estimators

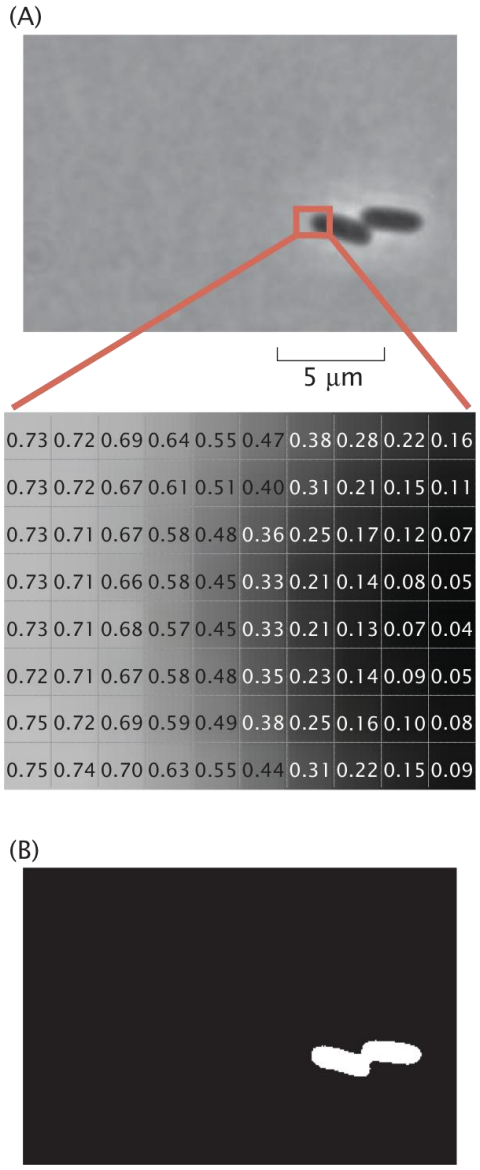
- ***The simple estimates are among the most sophisticated things we will do.***



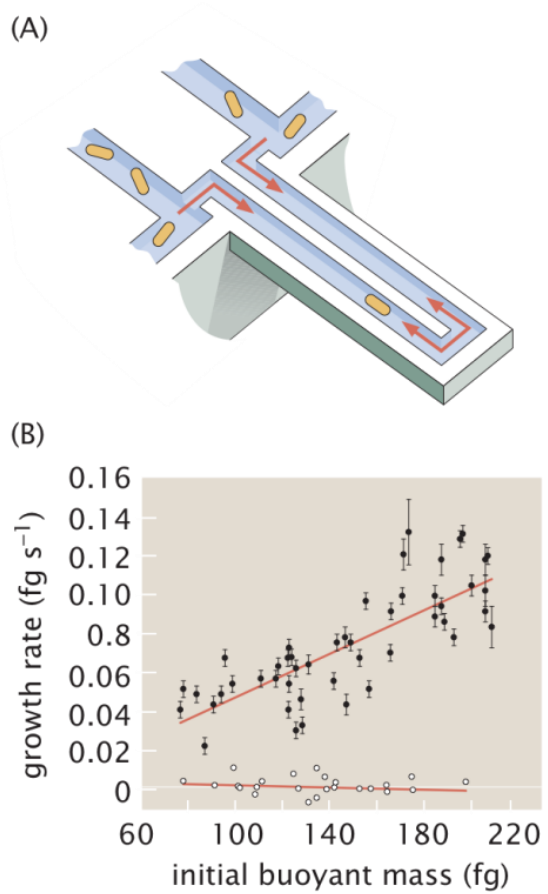


How big is a bacterium – the microscopy answer





How big is a bacterium – dependence upon growth rate



generation time (min)	dry mass per cell (fg=10^{-15}g)
100	148
60	258
40	433
30	641
24	865

(Figures adapted from M. Godin et al., Nat. Meth. 7: 387, 2010.)

Thinking Up the Number of Genes: "What is true for E. coli is true for the elephant"

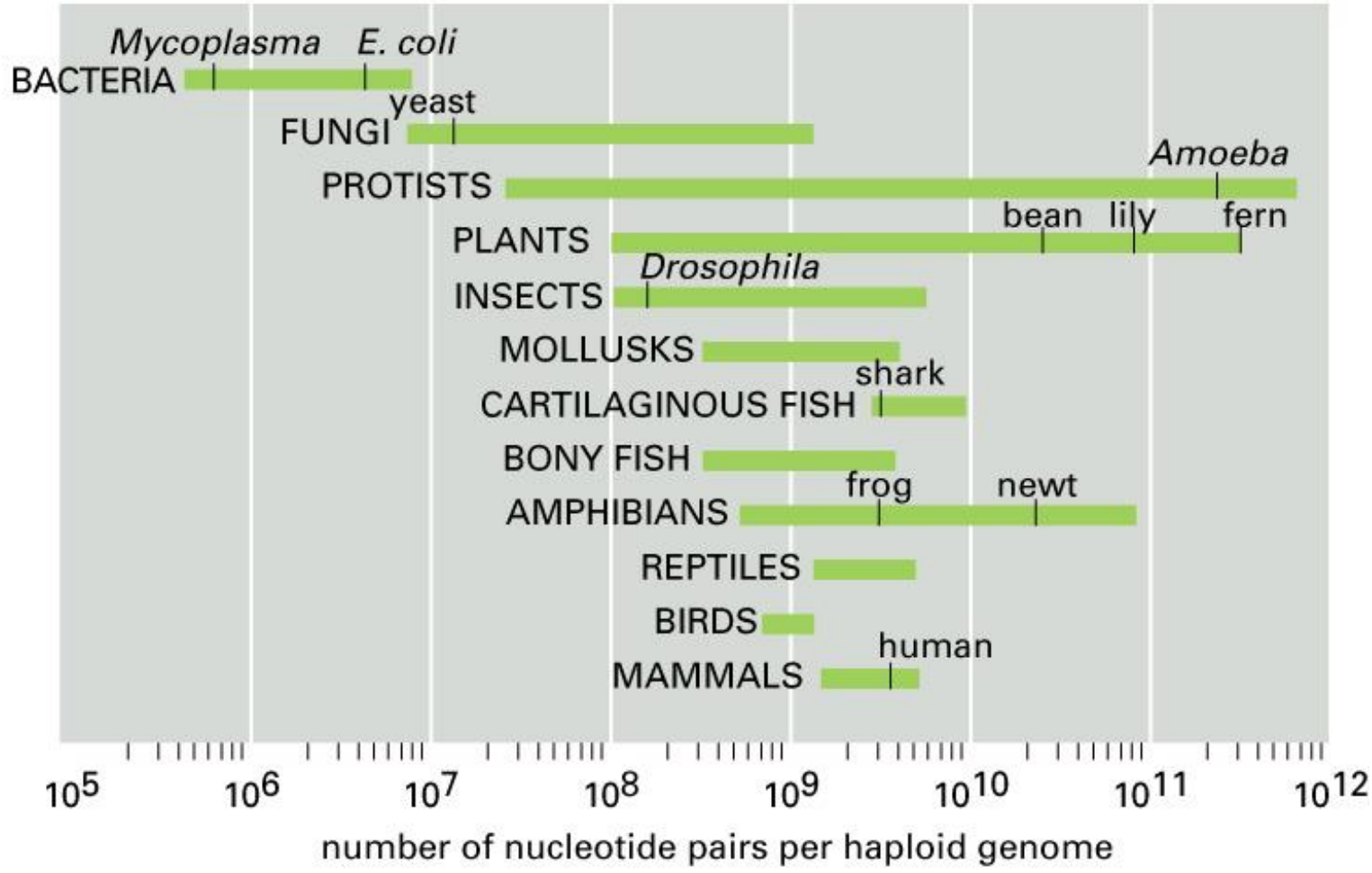
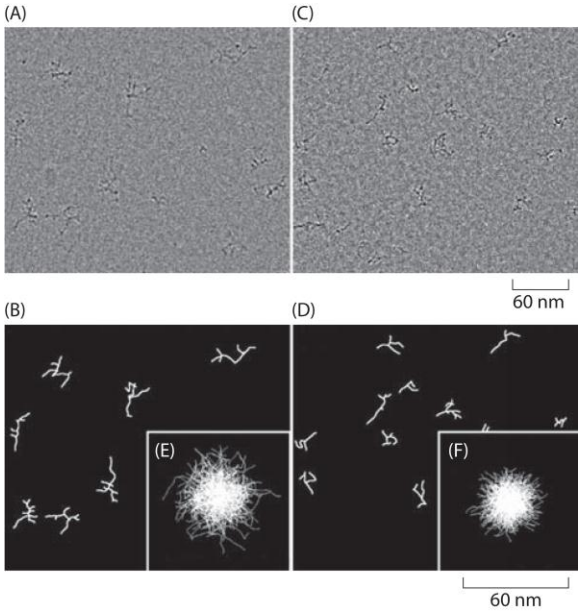
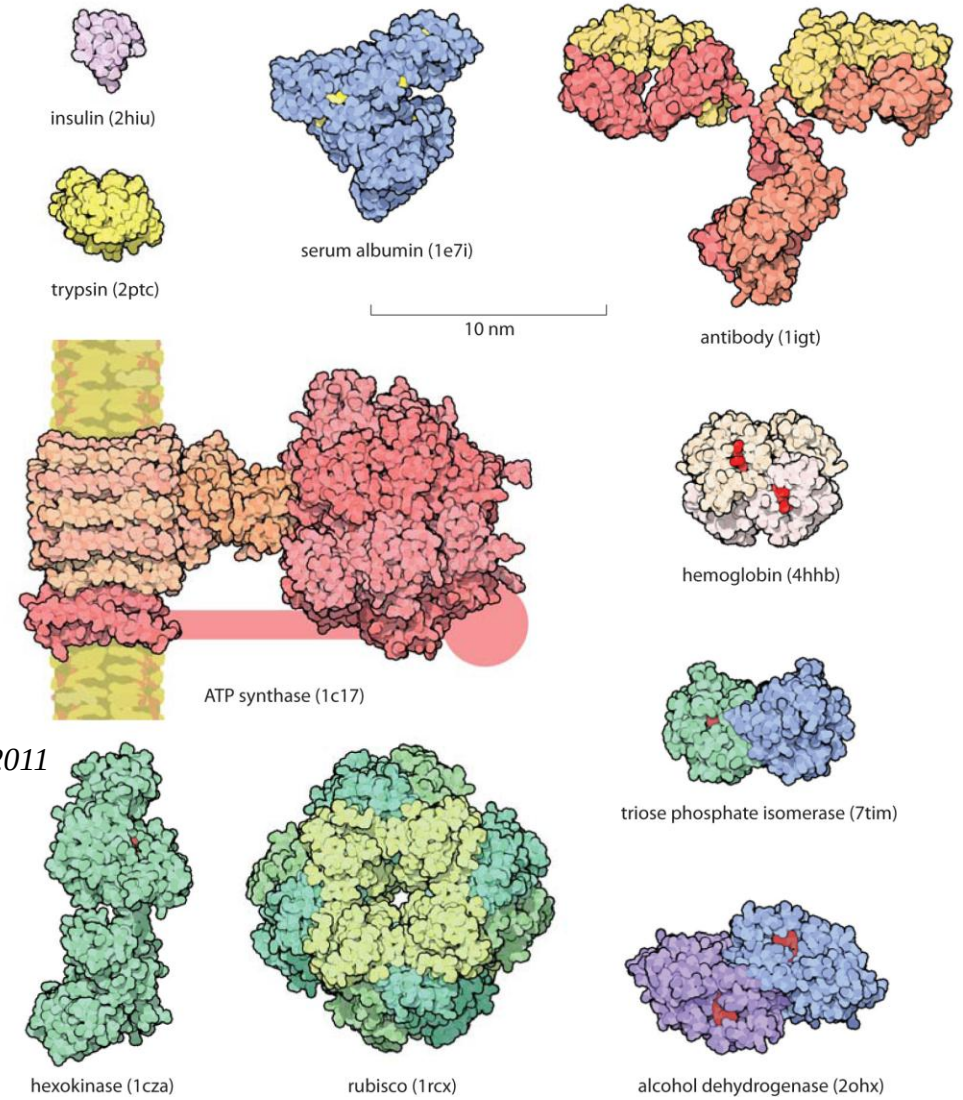


Figure 1–38. Molecular Biology of the Cell, 4th Edition.

Class estimate: number of genes in a bacterial genome

Typical mRNA and protein size





RNA 2012 18: 284-299 originally published online December 21, 2011

Ajaykumar Gopal, Z. Hong Zhou, Charles M. Knobler, et al.

Thinking Up the Number of Genes: "What is true for E. coli is true for the elephant"

- ***A betting pool (Las Vegas for estimates) was set up on the number of genes in the human genome and responses varied from 25,000 to 150,000.***
- ***A winner was declared, but the issue remains unsettled.***
- ***Simplest logic: use "typical" protein size of 300 amino acids, which requires roughly 1000 nucleotides to code for them. This naïve estimate says:***

$$N_{genes} \approx N_{bp}/1000$$

- ***Works great for E. coli, fails miserably for elephants (and humans)***

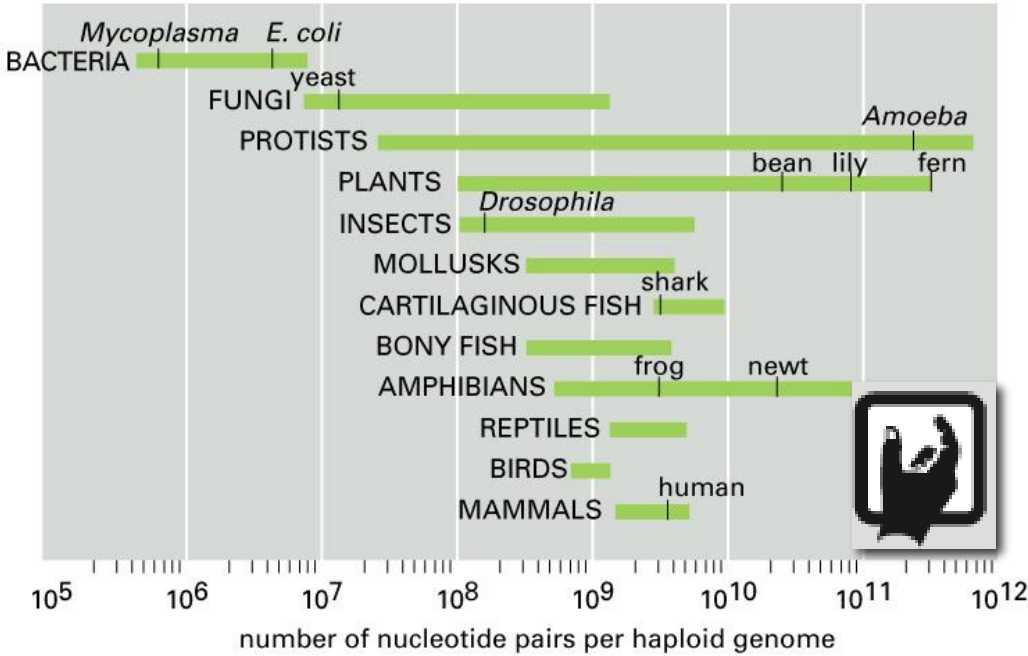


Figure 1–38. Molecular Biology of the Cell, 4th Edition.

The bet: genesweep pool- Science 6 June 2003:

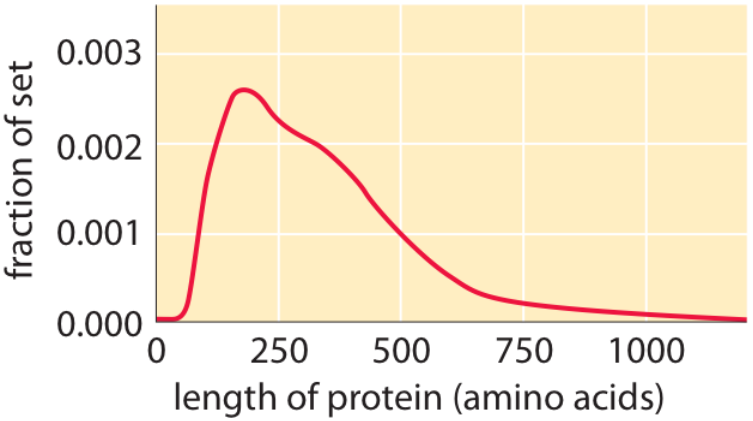
Vol. 300. no. 5625, p. 1484

A Low Number Wins the GeneSweep Pool

Elizabeth Pennisi

COLD SPRING HARBOR, NEW YORK--The human genome has been sequenced, but calculating the number of genes it contains is taking more time. DNA experts have nonetheless decided they know who made the best prediction.

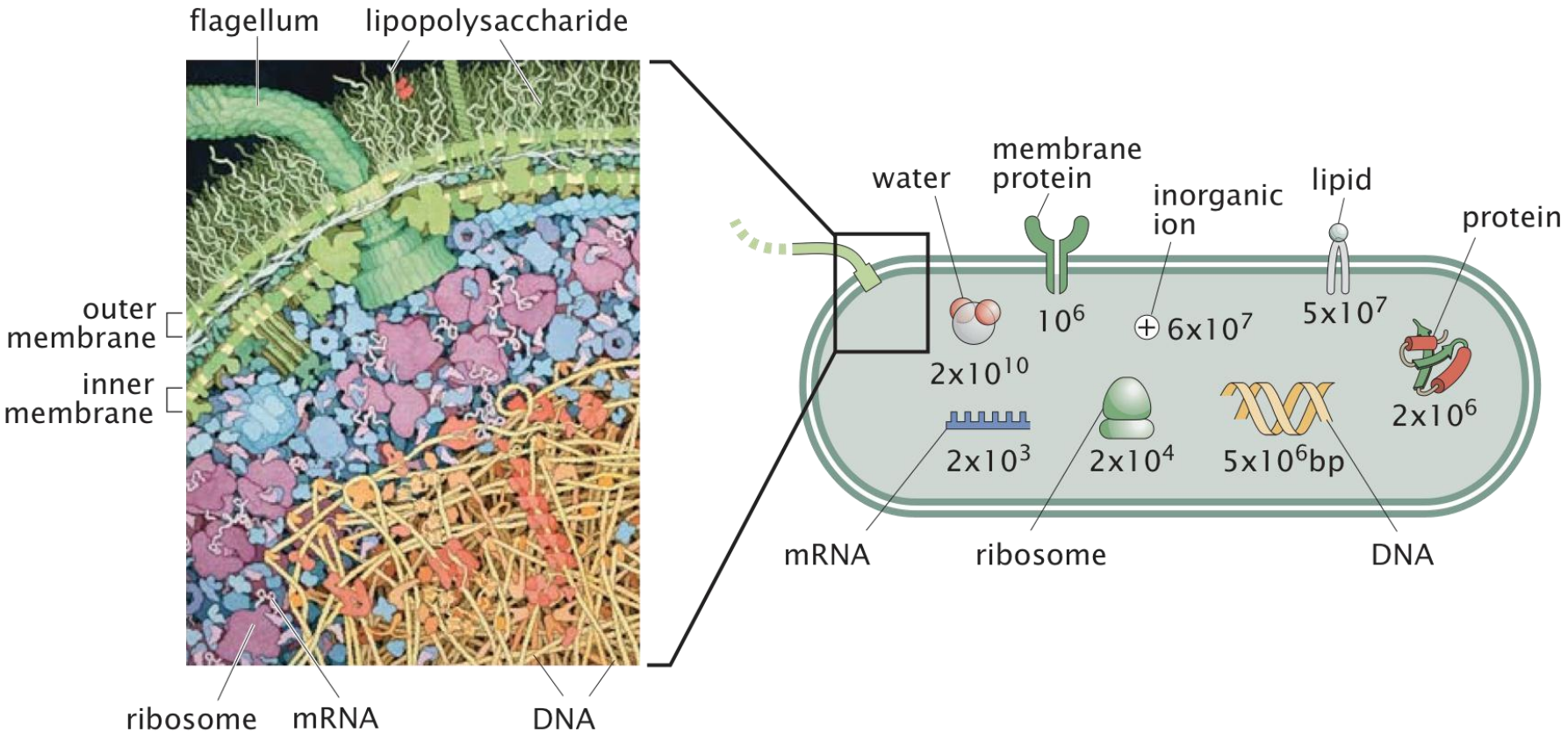
Gene size and number of genes



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

The e. coli census

- What makes up a bacterium and how do we know?



Concentrations in e. coli units – the rule of thumb

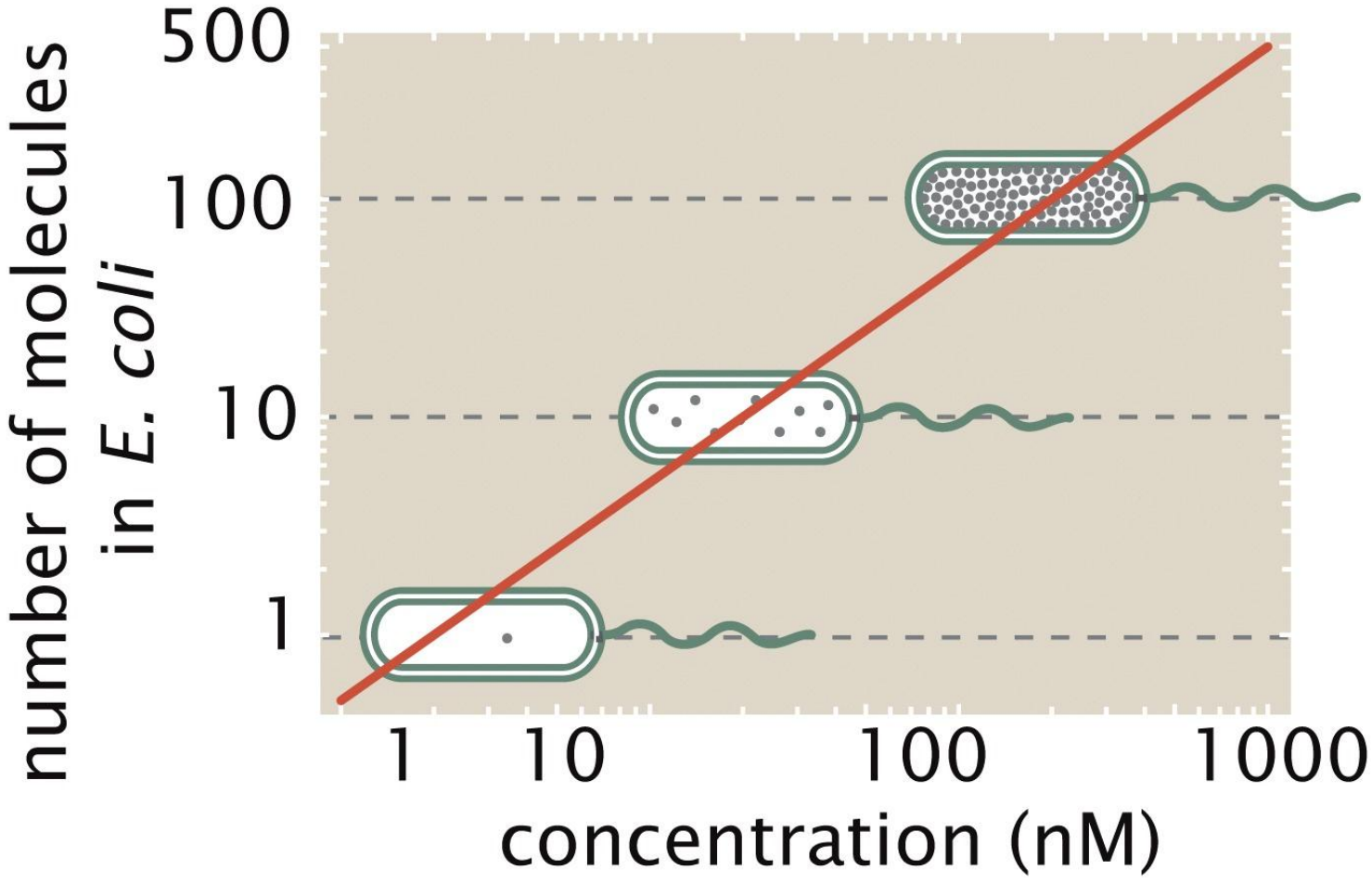
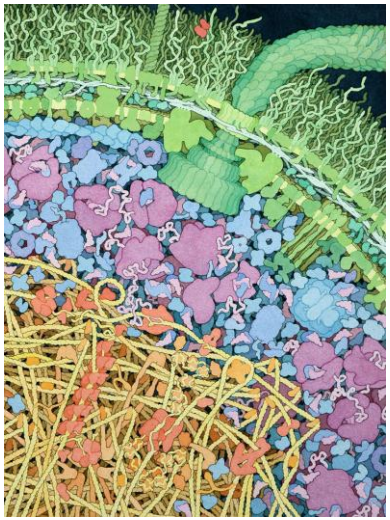


Figure 2.12a Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Sizing up *e.coli*



Estimate $N_{protein}$ in an *E.coli* cell:

- $N_{protein} = m_{total\ protein} / m_{per\ protein}$
- $V_{E.coli} \approx 1$fL, assume $\rho_{E.coli} \approx \rho_{H20} = 1$ g/mL => $m_{E.coli} \approx 1$ pg
- from exp.: cell's dry weight = 30% cell's total weight, protein = 50% of dry weight

  => $m_{total\ protein} \approx 0.15$ pg
- Aver.protein = 300 AA, $m_{AA} \approx 100$ Da => $m_{per\ protein} = 30{,}000$ Da; $1\text{Da} \approx 1.6\times10^{-24}$ g

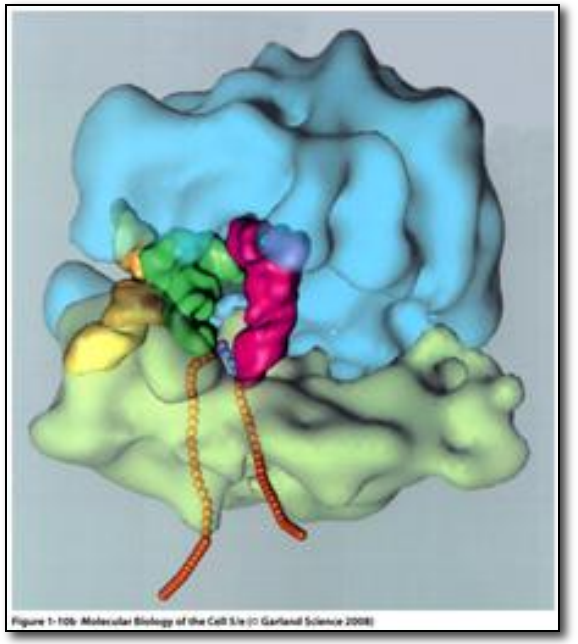
  => $m_{per\ protein} = 5 \times 10^{-20}$ g
- $N_{protein} = m_{total\ protein} / m_{per\ protein} \approx (15 \times 10^{-14}\ \text{g}) / (5 \times 10^{-20}\ \text{g}) \approx 3 \times 10^{6}$

$$N_{protein} \approx 3 \times 10^{6}$$

- 1/3 proteins coded in a typical genome = membrane proteins

  => $N_{cytoplasmic\ protein} \approx 2 \times 10^{6}$, $N_{membrane\ protein} \approx 10^{6}$

Sizing up *e.coli*



Ribosome - cellular machine that synthesizes proteins

Estimate $N_{ribosome}$ in an *E.coli* cell:

- $N_{ribosome} = m_{total\ ribosome} / m_{per\ ribosome}$
- facts: ribosomal protein = 20% cell's total protein,

  $m_{per\ ribosome} \approx 2.5$ MDa,

  $m_{per\ ribosome} \approx$ (1/3 protein + 2/3 RNA)

- $N_{ribosome} = m_{total\ ribosomal\ protein} / m_{protein\ per\ ribosome}$

$$\approx (0.2 \times 0.15 \times 10^{-12}\ \text{g}) / (830{,}000\ \text{Da}) \times (1\ \text{Da}) / (1.6 \times 10^{-24}\text{g}) \approx 20{,}000\ \text{ribosomes}$$

20% of cell's protein is ribosomal

15% of cell's mass is protein

cell mass

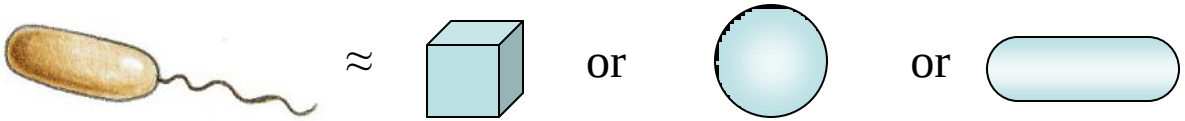
1/3 of the ribosomal mass is protein

$N_{ribosome} \approx 20{,}000$

- $d_{ribosome} \approx 20$ nm => volume taken up by 20,000 ribosomes:

  $V_{total\ ribosome} \approx 10^8\ \text{nm}^3$ <=> 10% of total cell volume

Sizing up *e.coli*



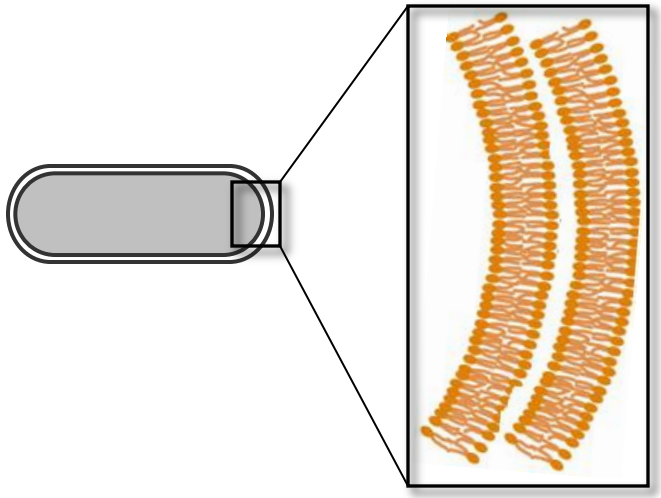
$\approx$ or or

=> surface area $A_{E.coli} \approx 6\ \mu m^2$

Estimate N_{lipid} associated with the inner and outer membranes of an *E.coli* cell:

$$N_{lipid} = 4 \times 0.5 \times A_{E.coli} / A_{lipid}$$

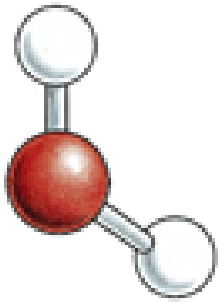
roughly half of the surf. area is covered by membrane proteins rather than lipids



$$\approx 4 \times 0.5 \times (6 \times 10^6\ nm^2) / 0.5\ nm^2 \approx 2 \times 10^7$$

$$N_{lipid} \approx 2 \times 10^7$$

Sizing up *e.coli*



Estimate N_{H_2O} in an *E.coli* cell:

- fact: 70% $m_{E.coli} \approx m_{H_2O}$ => $m_{total\ H_2O} \approx 0.7$ pg

- $N_{H_2O} \approx 0.7 \times 10^{-12}$ g / (18 g/mol) × $(6 \times 10^{23}$ molecules/mol)

  $\approx 2 \times 10^{10}$ water molecules

$$N_{H_2O} \approx 2 \times 10^{10}$$

Estimate $N_{inorganic\ ions}$ in a typical bacterial cell:

- assume a typical concentration of inorganic ions (e.g., K^+) is 100 mM =>

- $N_{inorganic\ ions} \approx (100 \times 10^{-3}$ mol) / (10^{15} μm^3) × $(6 \times 10^{23}$ molecules/mol) × 1 μm^3

  $\approx 6 \times 10^{7}$

  $1L$

  Cell volume

$$N_{ions} \approx 6 \times 10^{7}$$

Taking the molecular census of *e.coli*: **how**?

- An important tool: gel

  - break open cells, keep only protein components

  - separate complex protein mixture into individual molecular species:

1. Load the mixture at one end of the gel, apply el. field across the gel => **diff. proteins migrate at rates ~ net charge;**

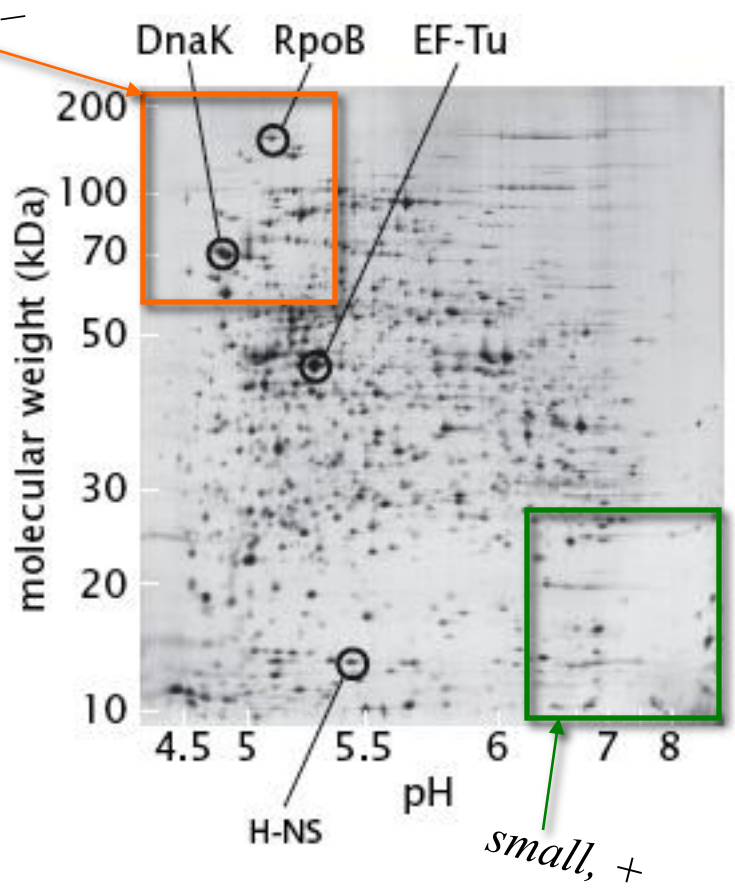
2. Add charged detergent that binds to all proteins => *N* of detergent molecules associated with a protein ~ the protein's size;

3. Apply ⊥ el. field; net charge on the detergent molecules >> original charge of the protein => **diff. proteins migrate at rates ~ protein size;**

  - stain proteins with a non-specific dye to locate

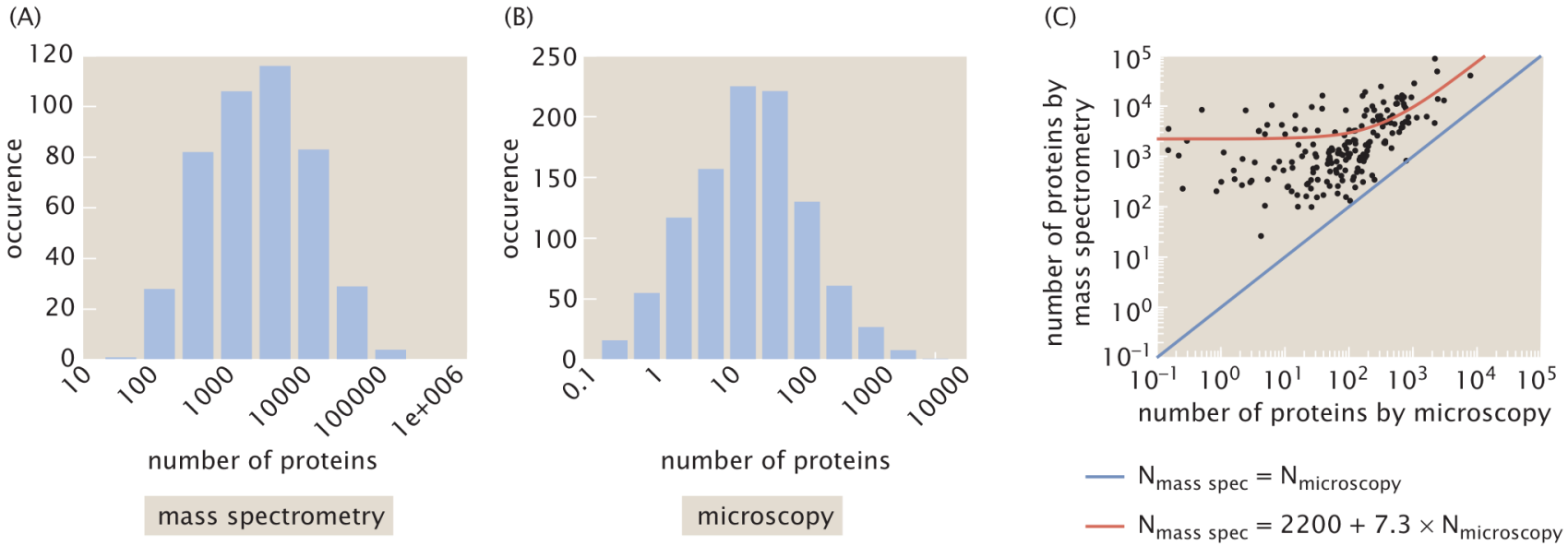
  - cut each spot out, elute protein, determine size and AA-content with mass spectrometry

- Use similar tricks to characterize RNA, lipids, etc.



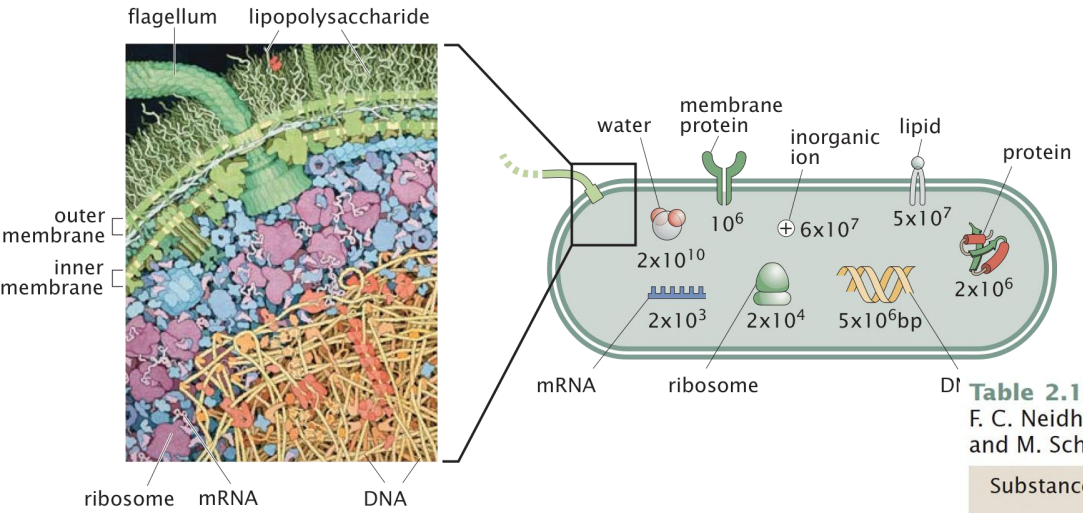
Protein census of E.coli using 2D polyacrylamide gel electrophoresis. Each spot represents an individual protein with a unique size and charge distribution. (Swiss Inst. of Bioinformatics)

e. Coli census



- ***More modern techniques use either mass spectrometry of cells or alternatively, look at living cells with fluorescent fusions.***
- ***We want to think carefully all the time about how things are measured and how well they agree. Remember the measurements on Avogadro's number?***

Sizing up *e.coli*: you should be able to think up this table and figure



Observed molecular census of an E .coli cell. (Data from Neinhardt et al., 1990, and Schaechter et al., 2006)

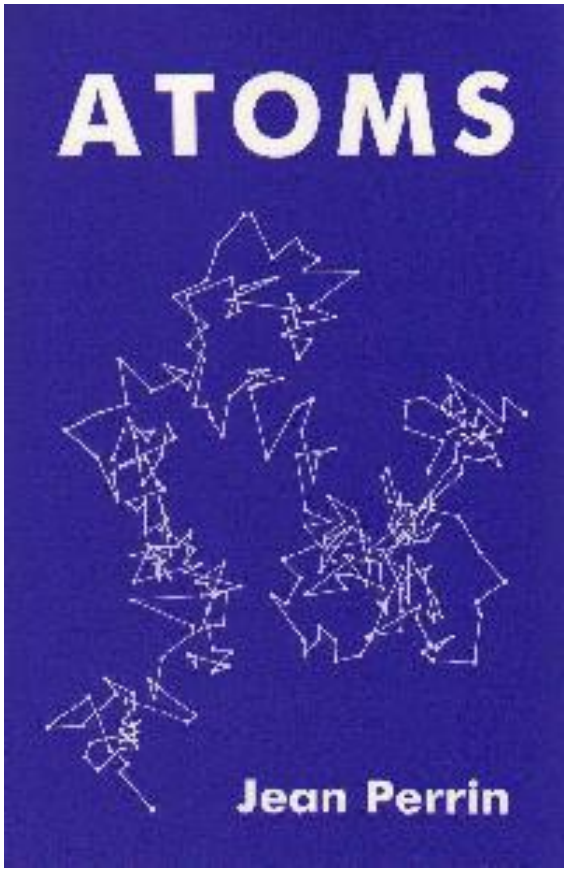
Table 2.1: Observed macromolecular census of an *E. coli* cell. (Data from F. C. Neidhardt et al., Physiology of the Bacterial Cell, Sinauer Associates, 1990 and M. Schaechter et al., Microbe, ASM Press, 2006.)

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**	

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

Coming at Problems Many Ways: Perrin, Avogadro and Fluctuations

- ***Intriguing tradition of using fluctuation-based methods to shed light on the system of interest. After this table, Perrin says: "the existence of the molecule is given a probability bordering on certainty."***
- ***My point: Demanding independent treatments of the same problem is a powerful way to obtain definitive understanding.***



CONCLUSIONS.

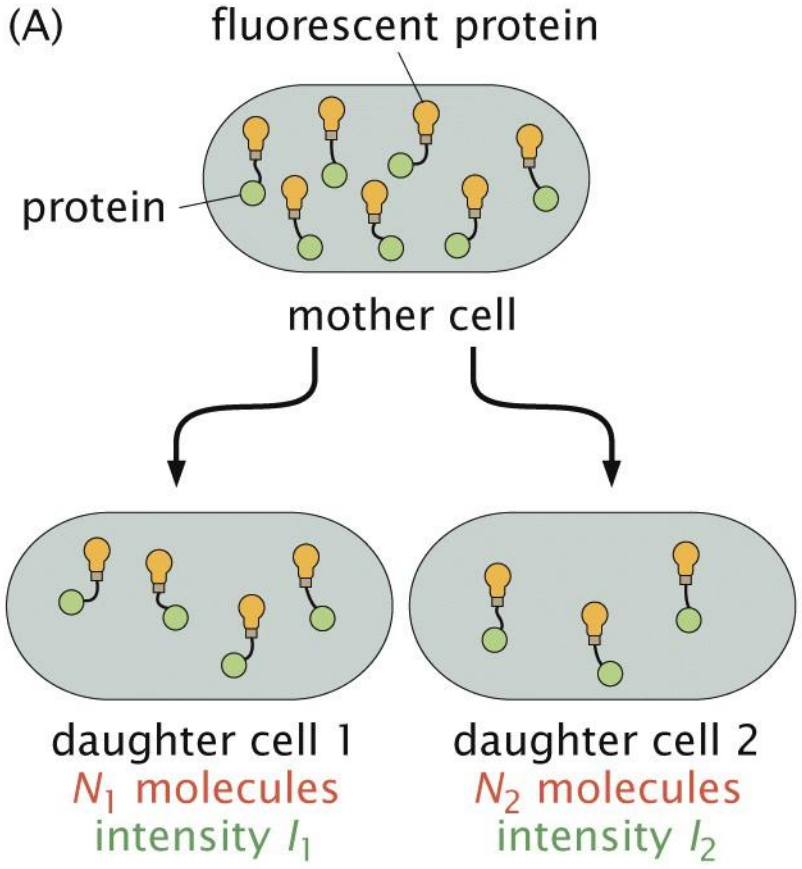
120.—THE AGREEMENT BETWEEN THE VARIOUS DETERMINATIONS.—In concluding this study, a review of various phenomena that have yielded values for the molecular magnitude enables us to draw up the following table :—

Phenomena observed.[1]		$\frac{N}{10^{22}}$.
Viscosity of gases (kinetic theory)		62 (?)
Vertical distribution in dilute emulsions		68
Vertical distribution in concentrated emulsions		60
Brownian movement	*Displacements*	64
	Rotations	65
	Diffusion	69
Density fluctuation in concentrated emulsions		60
Critical opalescence		75
Blueness of the sky		65
Diffusion of light in argon		69
Black body spectrum		61
Charge as microscopic particles		61 (?)
Radioactivity	*Projected charges*	62
	Helium produced	66
	Radium lost	64
	Energy radiated	60

Our wonder is aroused at the very remarkable agreement found between values derived from the consideration of such widely different phenomena. Seeing that not only is the same magnitude obtained by each method when the

[1] Methods by which it may be hoped, in the future, to obtain results of great precision are given in italics.

Cells and the small numbers limit



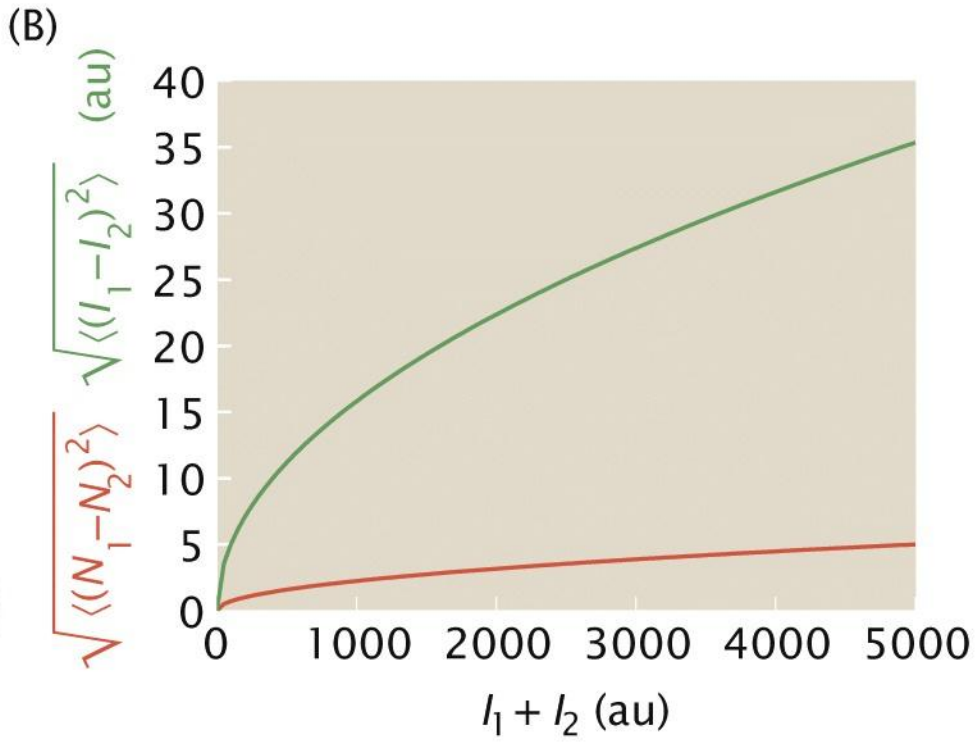


Figure 2.10 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Cells and the small numbers limit

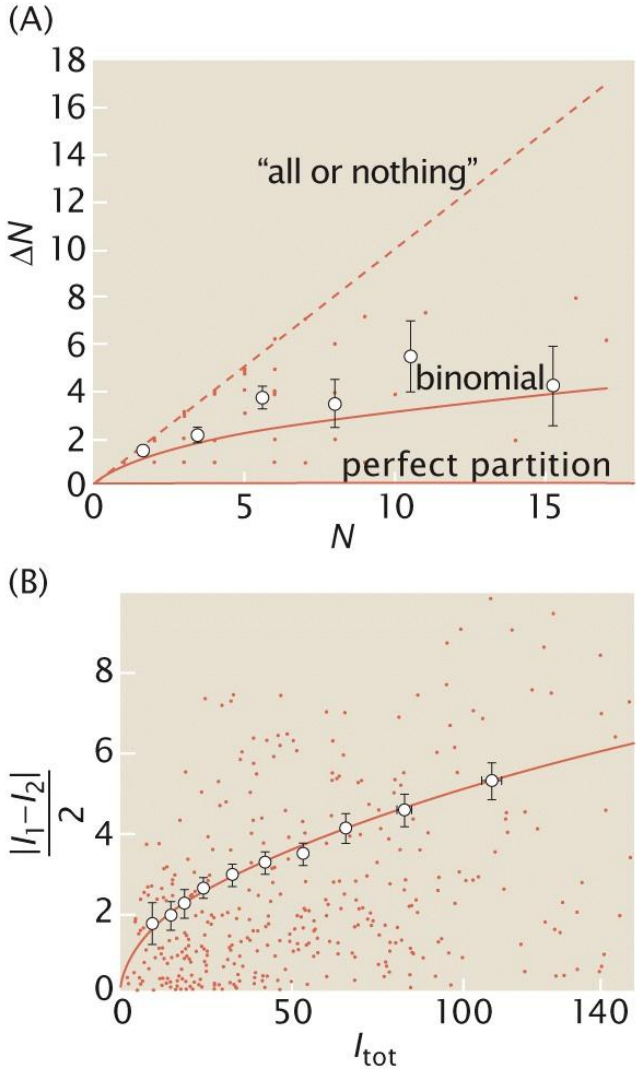


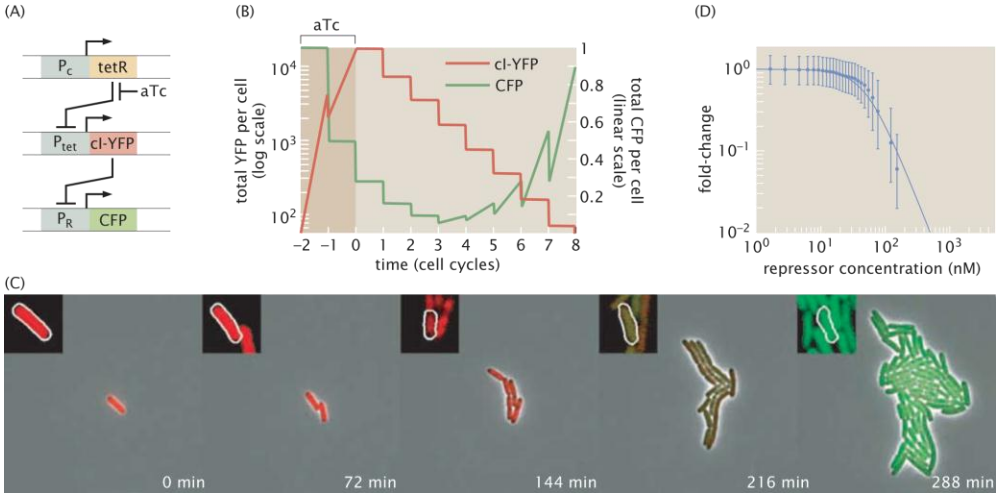
Figure 2.9 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Figure 2.9: Binomial partitioning of mRNA and proteins in *E. coli* during cell division. (A) Difference ΔN in the number of mRNAs between the two daughters given that the mother has N mRNAs. The curves show three possible partitioning mechanisms involving "all or nothing" in which one daughter takes all of the mRNAs, binomial partitioning, and "perfect partitioning" in which each daughter gets exactly half of the proteins from the mother cell. (B) Difference in the fluorescence intensity of the two daughter cells for a fluorescent fusion to a repressor protein as a function of the fluorescence 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.)

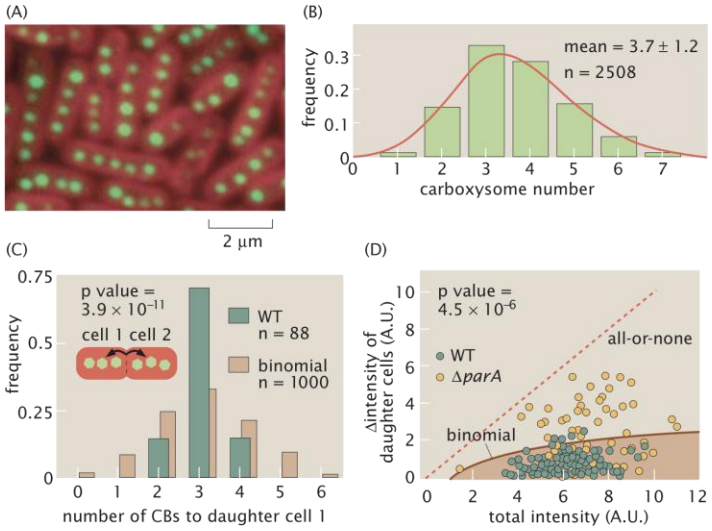
Phylogenetic trees in Concept space: Physical proximity vs biological proximity

- ***Another key feature of the physical approach is that it leads to different phylogenetic trees of relatedness of phenomena.***
- ***Here we talk of expression and photosynthesis in the same breath.***

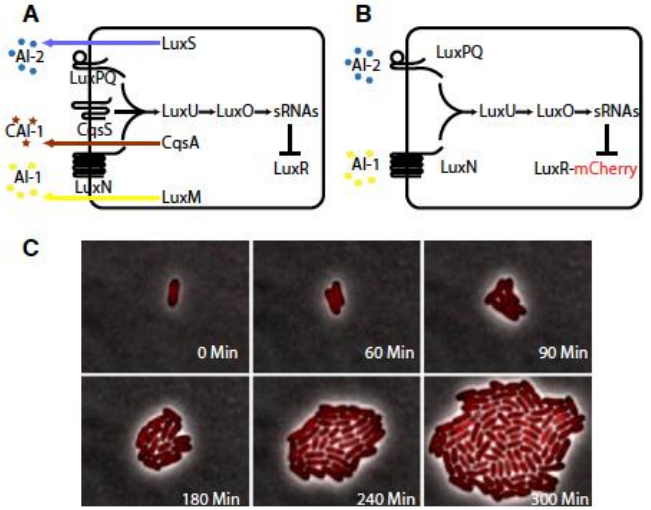
(Rosenfeld, Young, Alon, Swain, Elowitz, Science, 2005)



(Dave Savage, Pamela Silver, et al., Science, 2010)



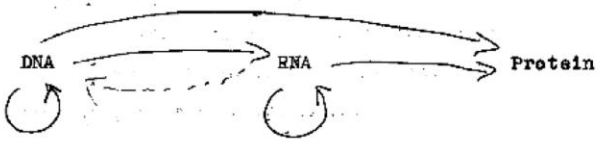
(Teng, Bassler, Wingreen, et al., Biophys. J. 2010)



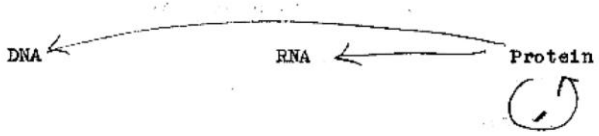
Timing the central dogma

The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it. That is, we may be able to have



but never



where the arrows show the transfer of information.

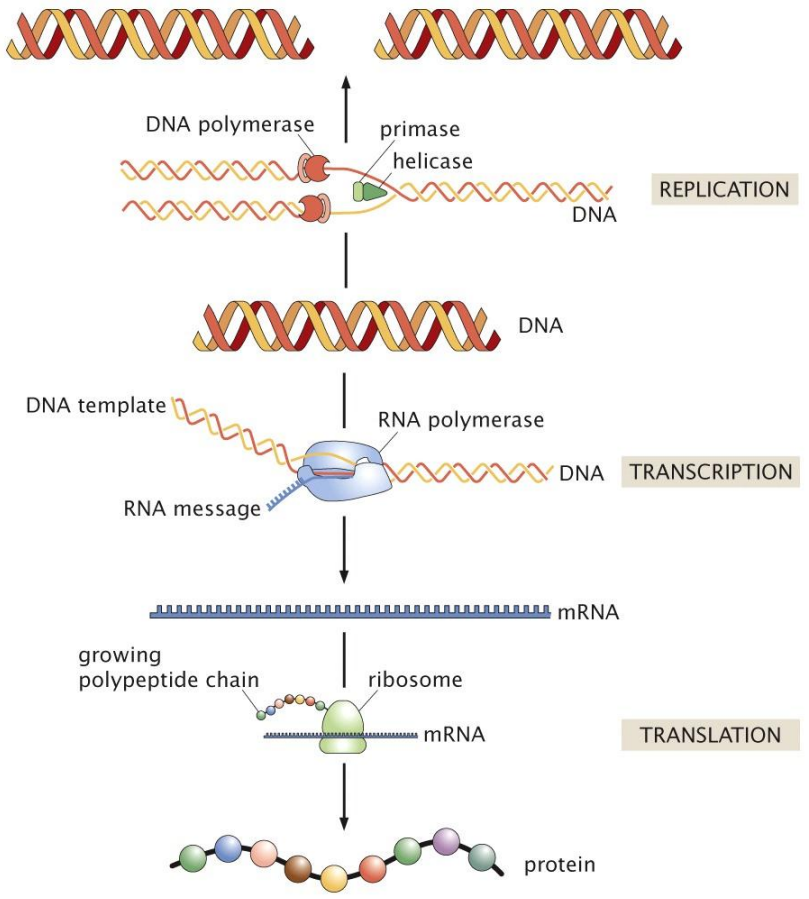
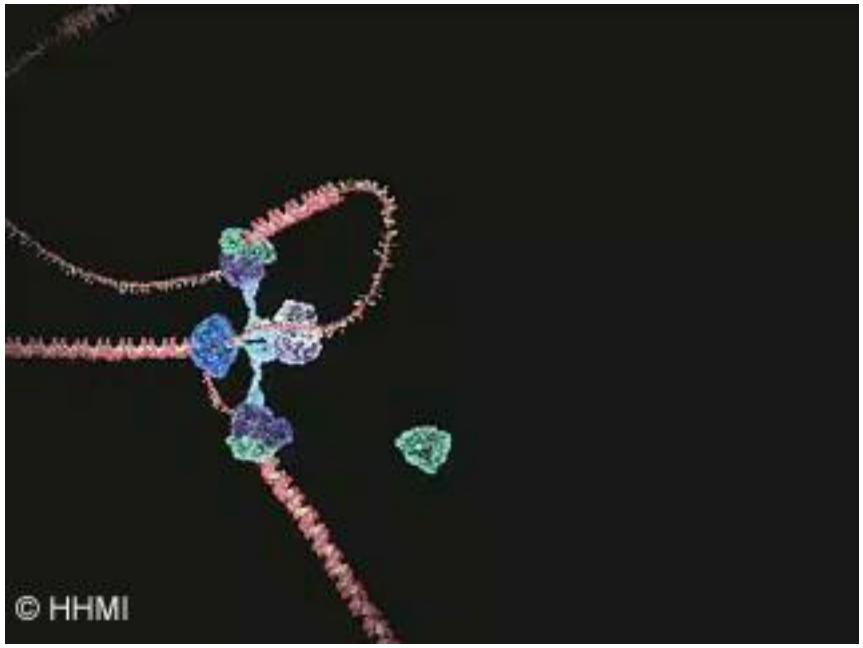
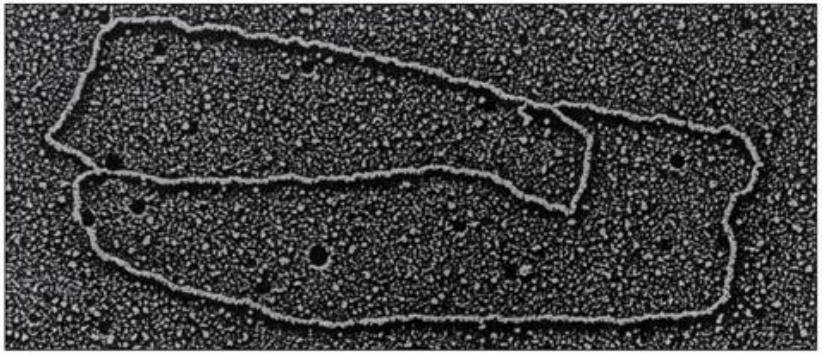


Figure 3.12 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Timing The machines of the central dogma

• *DNA replication in bacteria is undertaken by 2 replication complexes which travel in opposite directions away from the origin of replication on the circular chromosome.*





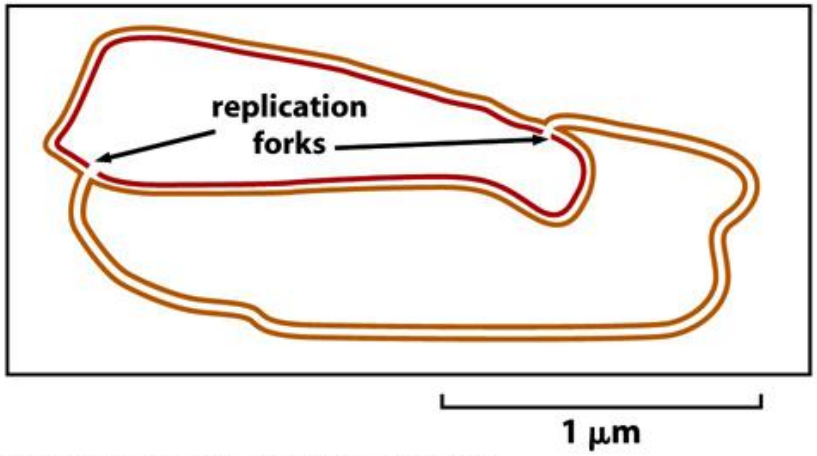
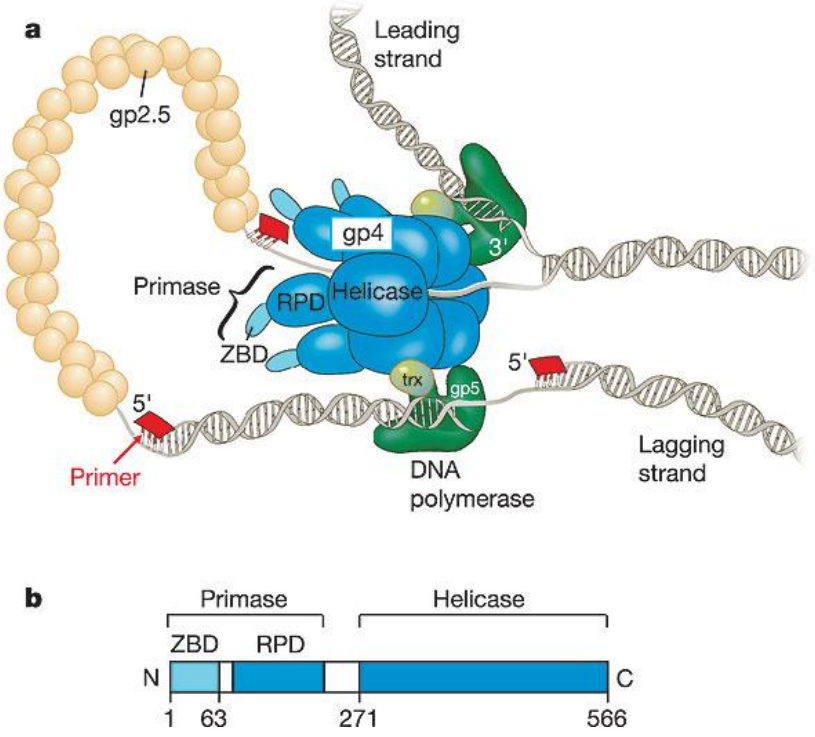


Figure 5-6 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Parental strands: orange, newly synthesized strands: red

Replication at the Single Molecule Level

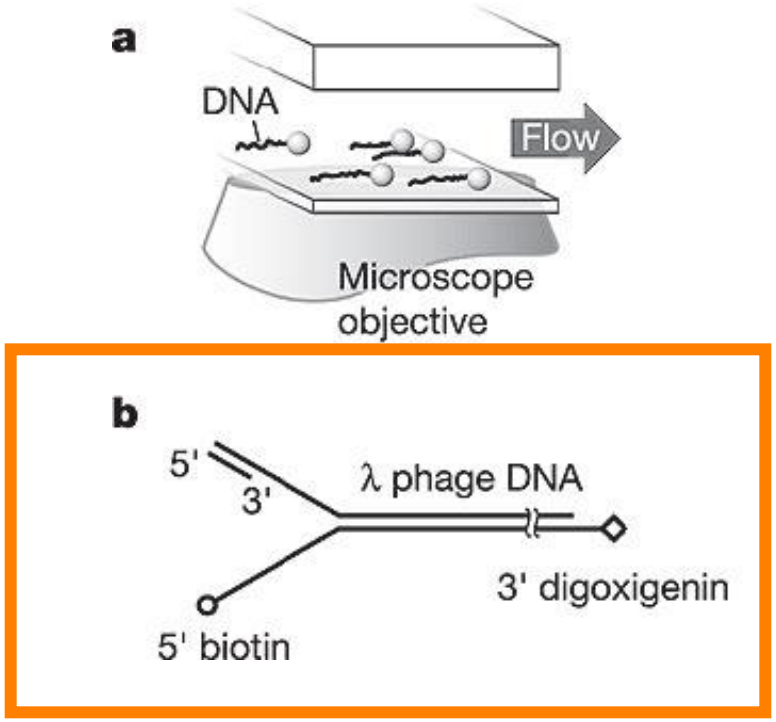
- ***The process of replication involves many tasks: opening up the DNA into single strands, making the copy on both strands, fixing the fragments on the lagging strand.***
- ***This figure shows the replication complex from bacteriophage T7.***
- ***Helicase: opens up the double-stranded DNA double helix.***
- ***Primase: constructs the primers so that the polymerase can make the relevant copy.***

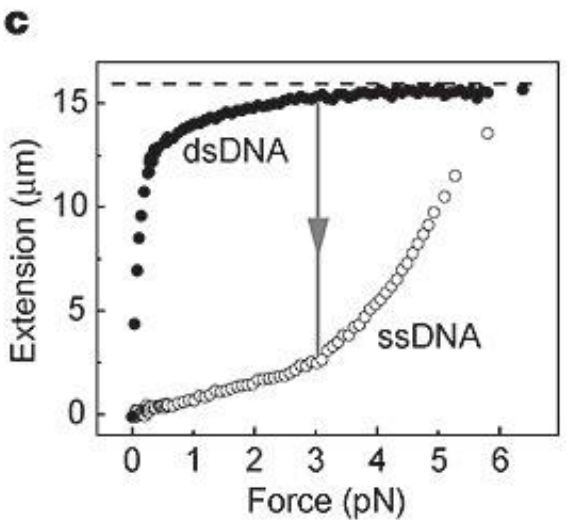


Van Oijen et al., Nature

The Concept of the Replication Experiment

- ***Experimental concept: Tether the DNA to the surface using molecular velcro (routine). DNA has a bead on its end, now apply a weak flow to stretch the DNA out and then monitor the position of the bead as the replication process proceeds.***

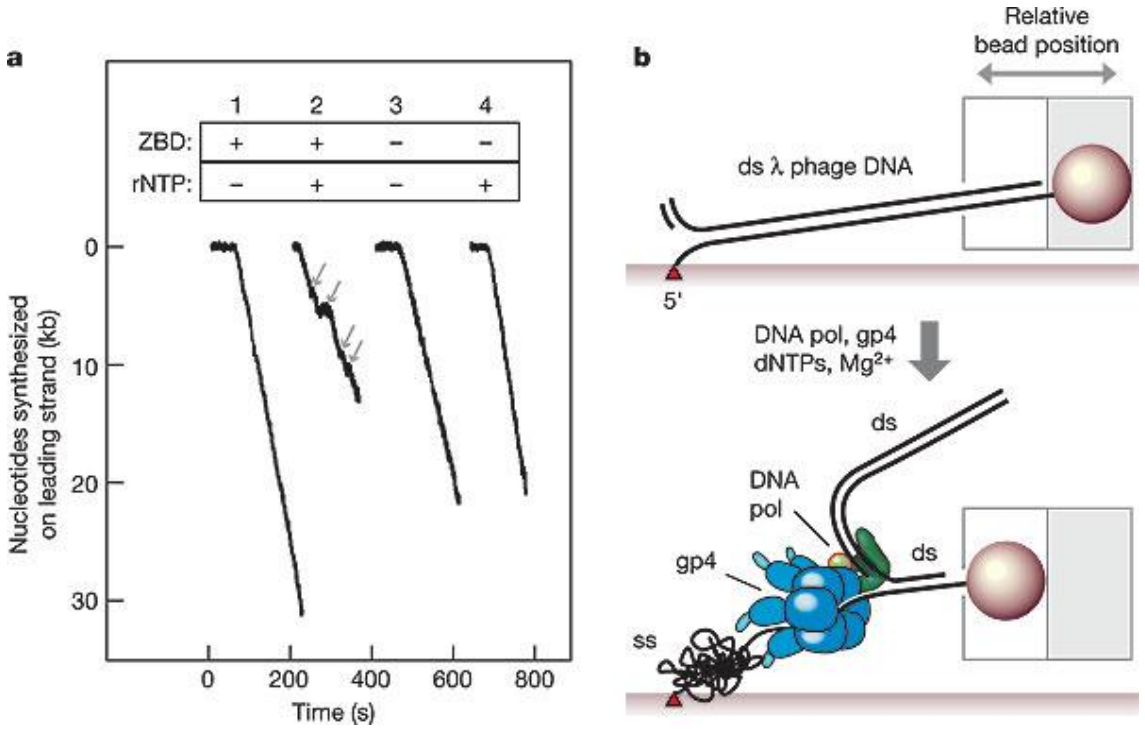




Van Oijen et al., Nature

Data on Leading Strand Synthesis

- ***As replication complex moves along, it replaces dsDNA on tether with ssDNA. This shortens the tether and hence by examining tether shortening (as reported by bead motion) one can measure kinetics of the replication complex.***
- ***Remember, we estimated this rate by thinking about cell division.***



Van Oijen et al., Nature

Measuring the rate of Transcription in cells

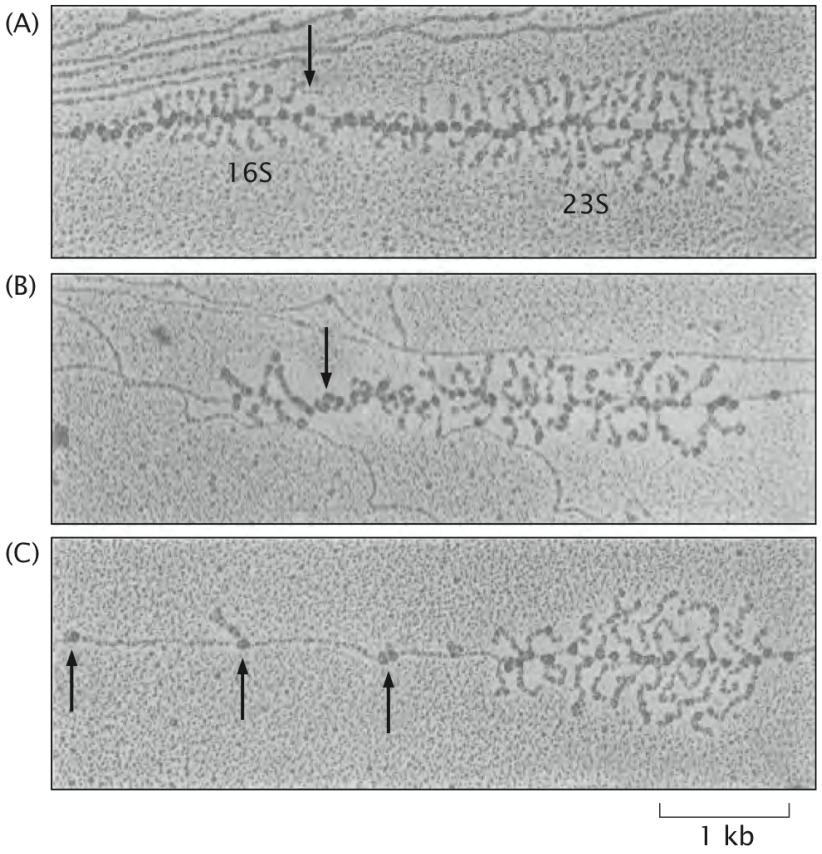
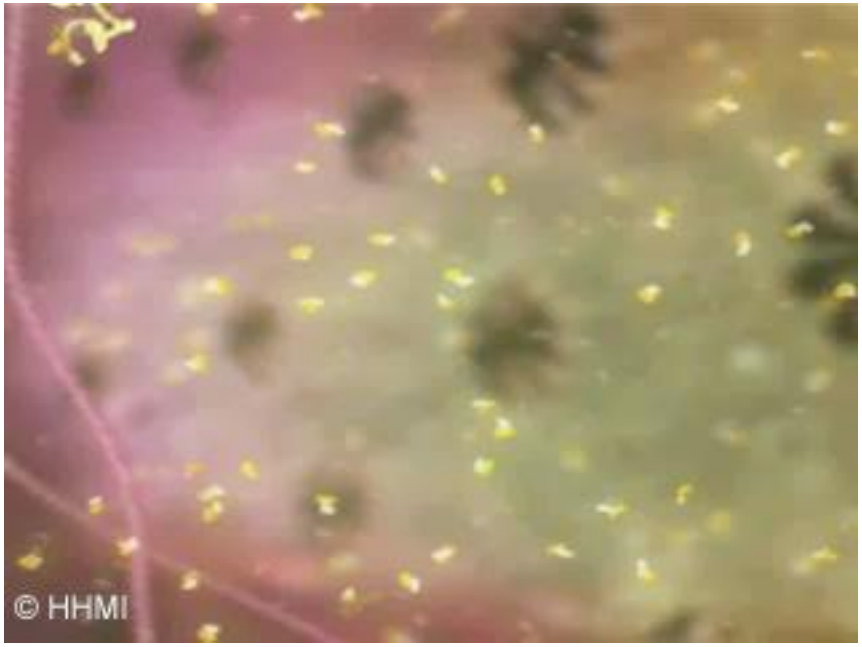


Figure 3.36: Effect of rifampin on transcription initiation. Electron micrographs of *E. coli* rRNA operons: (A) before adding rifampin, (B) 40 s after addition of rifampin, and (C) 70 s after exposure. No new transcripts have been initiated, but those already initiated are carrying on elongation. In (A) and (B) the arrow is used as a fiducial marker and signifies the site where RNaseIII cleaves the nascent RNA molecule producing 16S and 23S ribosomal subunits. RNA polymerase molecules that have not been affected by the antibiotic are marked by the arrows in (C). (Adapted from L. S. Gotta et al., *J. Bacteriol.* 20:6647, 1991.)

Timing The machines of the central dogma

Estimate the mean time to synthesize a typical protein.



Translation in Eukaryotes

- *E.coli*: $N_{protein} \approx 3 \times 10^6$, each protein: 300 AA,

$\tau_{cell} \approx 3000$ s

=> Rate of AA incorporation by the ribosome:

$dN_{AA}/dt \approx (9 \times 10^8 \text{ AA}) / 3000 \text{ s} \approx 3 \times 10^5 \text{ AA/s}$

- $N_{ribosome} \approx 20{,}000$ => rate per ribosome: 15 AA/s (exp.: 25 AA/s)

=> $t_{protein} \approx 20$ s

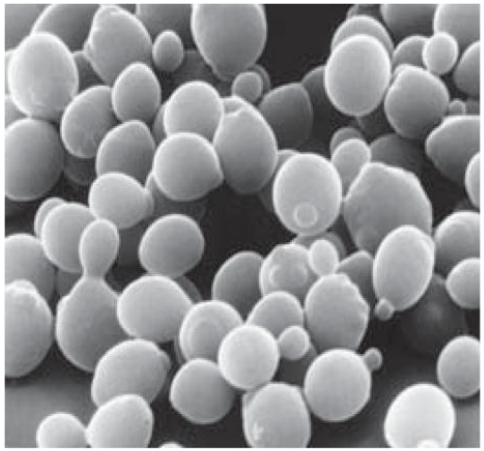
- Rate of protein synth. by the ribosome < rate of mRNA synthesis by RNA polymerase.
- Simultaneous translation of a single mRNA by multiple ribosomes
- => When considering the net rates of processes in cells: *N* of mol. players + the intrinsic rate.

Ribosomes and growth rate

Table 1: Number and fraction of ribosomes as a function of the doubling time. * from E. coli & Salmonella, Chapter 97, Table 3. ** from E. coli & Salmonella, Chapter 97, Table 2. Ribosome dry mass fraction is calculated based on ribosome mass of 2.7MDa (BNID 100118). [Supporting calculations in excel table in vignette directory]

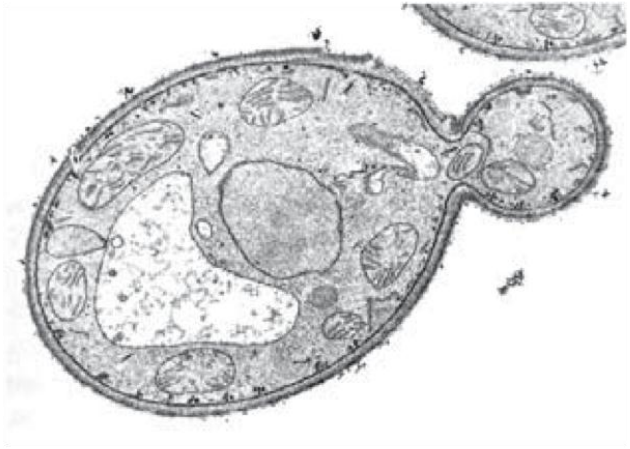
Doubling time [min]	Ribosomes per cell *	Dry mass per cell** (pg)	Ribosome dry mass fraction (%)	Ribosome fraction * doubling time [min]
24	72000	865	37	9.0
30	45100	641	32	9.5
40	26300	433	27	10.9
60	13500	258	24	14.1
100	6800	148	21	20.7

Anatomy of the classic model eukaryote



(A)

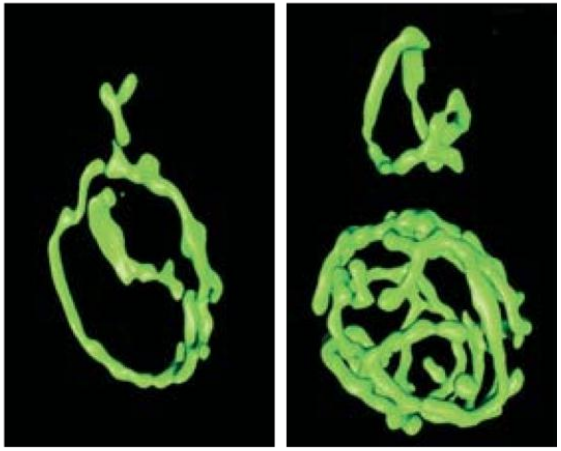
10 μm



(B)

1 μm

glucose glycerol



(C)

1 μm

Census in Yeast