Physics 176/276
Quantitative Molecular Biology

Instructor: Massimo Vergassola

Winter 2014

http://physics.ucsd.edu/students/courses/winter2014/physics176

Historical perspective

18-19th-century: industrial revolution

-- steam boats, railways, assembly line
-- chemical, electrical, steel industries
-- automobiles, tanks, airplanes, …

} controlled conversion of chemical energy for useful mechanical work

➔ key technology: steam engine, piston, turbine & reaction engines
➔ scientific foundation: thermodynamics, physical chemistry

20th-century: information revolution

-- transistors, integrated circuits, microprocessors
-- calculators, personal computers, internet
-- wireless, nano, smart materials, …

} designed manipulation of electronic flow for info processing

➔ key technology: integrated circuits manufacturing & material processing
➔ scientific foundation: quantum mechanics, information theory

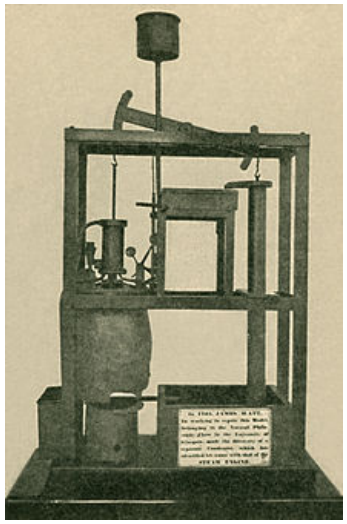
21st-century: bio revolution

-- biofuel, bioremediation
-- rational drug design
-- personalized medicine

} designed manipulation or informed intervention of biological organisms for useful purposes = **synthetic biology**

➔ key technology: DNA sequencing/synthesis, ‘omics ?
➔ scientific foundation: **quantitative & systems biology** ???

Historical perspective on theory/applications

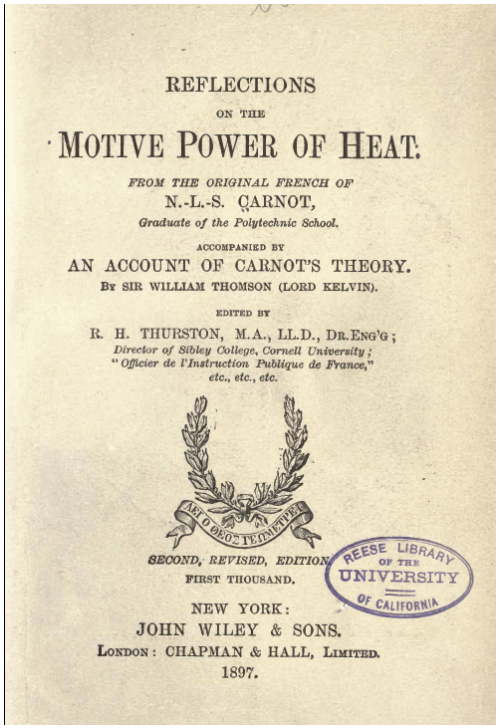


Savery engine ≈ 1700

Newcomen engine ≈1710

James Watt ≈1765

Sadi Carnot ≈1824



REFLECTIONS

ON THE

MOTIVE POWER OF HEAT.

FROM THE ORIGINAL FRENCH OF

N.-L.-S. CARNOT,

Graduate of the Polytechnic School.

ACCOMPANIED BY

AN ACCOUNT OF CARNOT'S THEORY.

BY SIR WILLIAM THOMSON (LORD KELVIN).

EDITED BY

R. H. THURSTON, M.A., LL.D., DR. ENG'G;

Director of Sibley College, Cornell University;

"Officier de l'Instruction Publique de France,"

etc., etc., etc.

SECOND, REVISED, EDITION

FIRST THOUSAND.

NEW YORK:

JOHN WILEY & SONS.

LONDON: CHAPMAN & HALL, LIMITED.

1897.

Steam engines (inefficient) were built on a practical basis well before theoretical understanding.

In Carnot's paper you can count equations on the fingers of your hands

Historical perspective on theory/applications

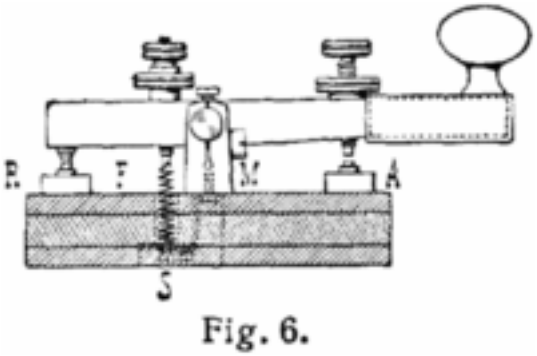


Fig. 6.

Morse ≈1840

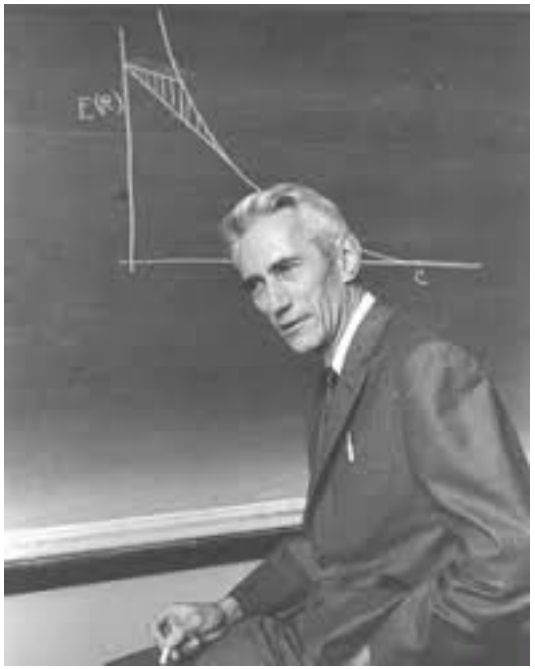
Telegraph lines ≈1891

The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON



Why quantitative biology?

-- because biology *is* quantitative

-- needed to formulate falsifiable predictions

-- demanded by synthetic biology

What is quantitative biology?

➔quantitative biology ≠ biology-inspired physics

≠ application of pre-existing methods to bio problems

Use numbers to gain quantitative and qualitative understanding

Role of theory

- link across different scales, i.e., from components to systems
- formulate expectations and predictions (via quantitative models)
- guide the design of new experiments, e.g. more discriminatory, and technology, e.g. more effective.
- power: the generality of (falsifiable) ideas and principles
- "cost": basic principles usually streamline, simplify the guess-and-try process of technological progress.

New concepts and principles lead to new perspectives

"Heavier than air flying machines are impossible."
-- Lord Kelvin

"Flight by machines heavier than air is impractical and insignificant, if not utterly impossible."
-- Simon Newcomb, Director, U.S. Naval Observatory, 1902

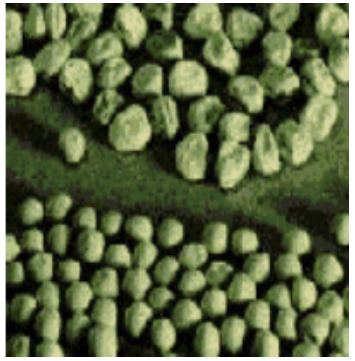
"Aerial flight is one of that class of problems with which man will never be able to cope."
-- Simon Newcomb, 1903

"The popular mind often pictures gigantic flying machines speeding across the Atlantic carrying innumerable passengers in a way analogous to our modern steam ships. . . it seems safe to say that such ideas are wholly visionary and even if the machine could get across with one or two passengers the expense would be prohibitive to any but the capitalist who could use his own yacht."
-- William Henry Pickering, Astronomer, 1910

A few successful paradigmatic examples

-- Mendel's laws of genetics
-- Luria-Delbrück experiment and fluctuation test
-- Hopfield's theory of kinetic proof-reading
-- Ho-Perelson's model of HIV kinetics

Mendel's laws of genetics



First law of Segregation

Parental Cross	F_1 Phenotype	F_2 Phenotypic Ratio	F_2 Ratio
Round x Wrinkled Seed	Round	5474 Round:1850 Wrinkled	2.96:1
Yellow x Green Seeds	Yellow	6022 Yellow:2001 Green	3.01:1
Red x White Flowers	Red	705 Red:224 White	3.15:1
Tall x Dwarf Plants	Tall	I787 Tall:227 Dwarf	2.84:1

Union of Gametes At Random

	D	***d***
D	*DD* (Tall)	*Dd* (Tall)
d	*Dd* (Tall)	*dd* (Short)

Mendel's conclusions

- Hereditary determinants are of a particulate nature (genes).
- Each parent has a gene pair for each trait.
- One member of the gene pair segregates into a gamete, i.e. each gamete carries one member of the gene pair.
- Gametes unite at random, irrespective of the other gene pairs involved.
- For the examples here, one of the two types (alleles) is dominant.

Second law Independent Assortment

		Female Gametes			
		GW	**Gw**	**gW**	**gw**
	GW	**GGWW** (Yellow, round)	**GGWw** (Yellow, round)	**GgWW** (Yellow, round)	**GgWw** (Yellow, round)
Male	**Gw**	**GGWw** (Yellow, round)	**GGww** (Yellow, wrinkled)	**GgWw** (Yellow, round)	**Ggww** (Yellow, wrinkled)
Gametes	**gW**	**GgWW** (Yellow, round)	**GgWw** (Yellow, round)	**ggWW** (Green, round)	**ggWw** (Green,
	gw	**GgWw** (Yellow, round)	**Ggww** (Yellow, wrinkled)	**ggWw** (Green, round)	**ggww** (Green, wrinkled)

Phenotype	**Genotype**
9 Yellow, Round Seed	*G_W_*
3 Yellow, Wrinkled Seed	*G_ww*
3 Green, Round Seed	*ggW_*
1 Green, Wrinkled Seed	*ggww*

Remarkable features

- Quantitative experiments (engineered pure lines and used statistics) and did not just describe what he saw: physicists do not have the monopole….
- Strong abstraction (laws synthesizing data and predicting results of new experiments, e.g. backcrossing or co-dominant traits) with new concepts (gene) going even beyond the existing limits (it will not be clear what a gene is for more than a century….).
- Created the new field of genetics (even though it took some time to rediscover…).
- A bit of luck helps….. Not all traits are Mendelian and segregate independently.

MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE[1,2]

Microbiologist

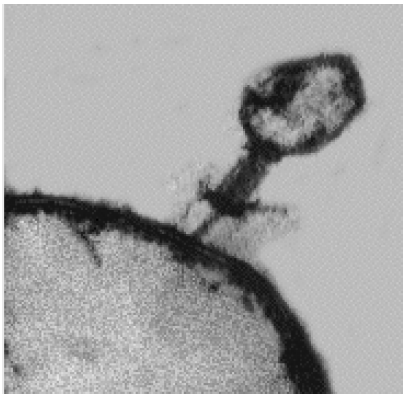
Theoretical physicist

S. E. LURIA[3] AND M. DELBRÜCK

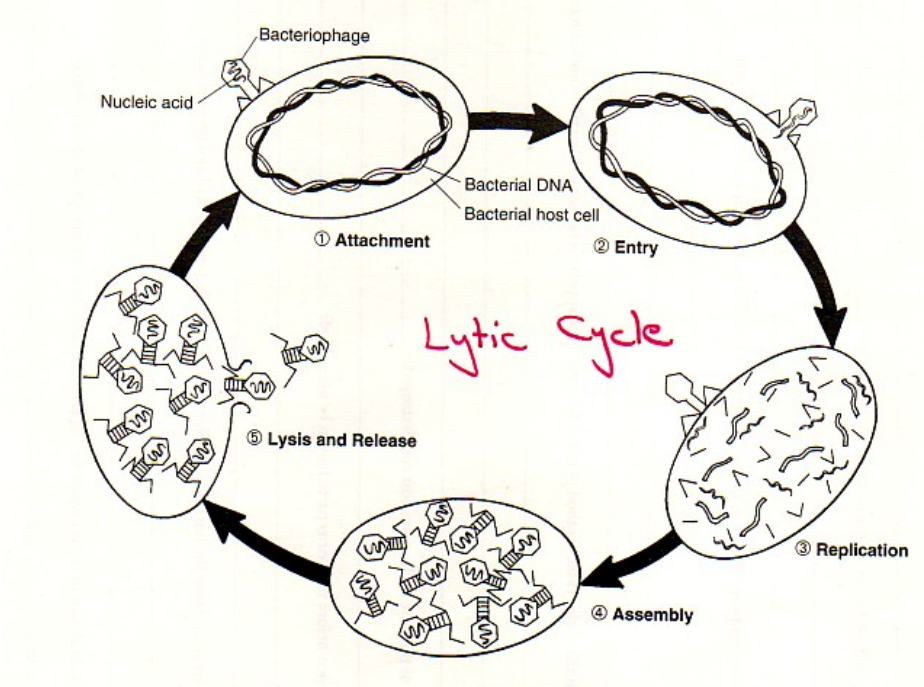
Indiana University, Bloomington, Indiana, and Vanderbilt University, Nashville, Tennessee

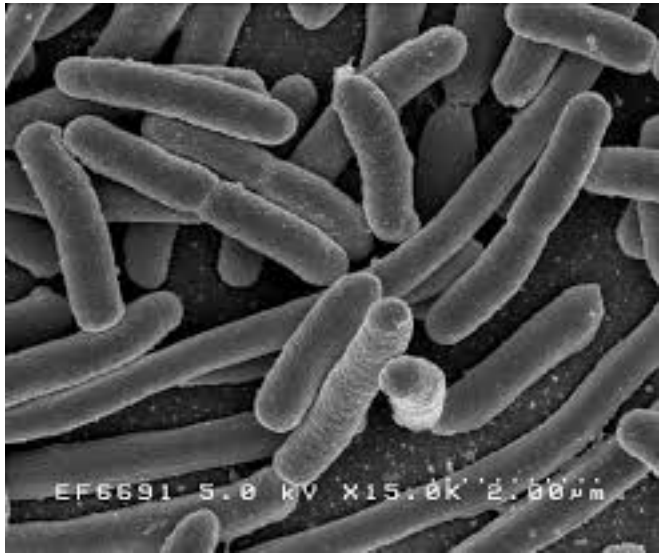
Received May 29, 1943

Are mutations random or induced by natural selection?



(Bacterio)Phages are viruses of bacteria





A small fraction of *E. coli* bacteria in a colony are resistant to infection by phages.

Use the phage-bacteria system for an experiment to discriminate between the two hypotheses: random vs induced mutations ?

Poisson statistics

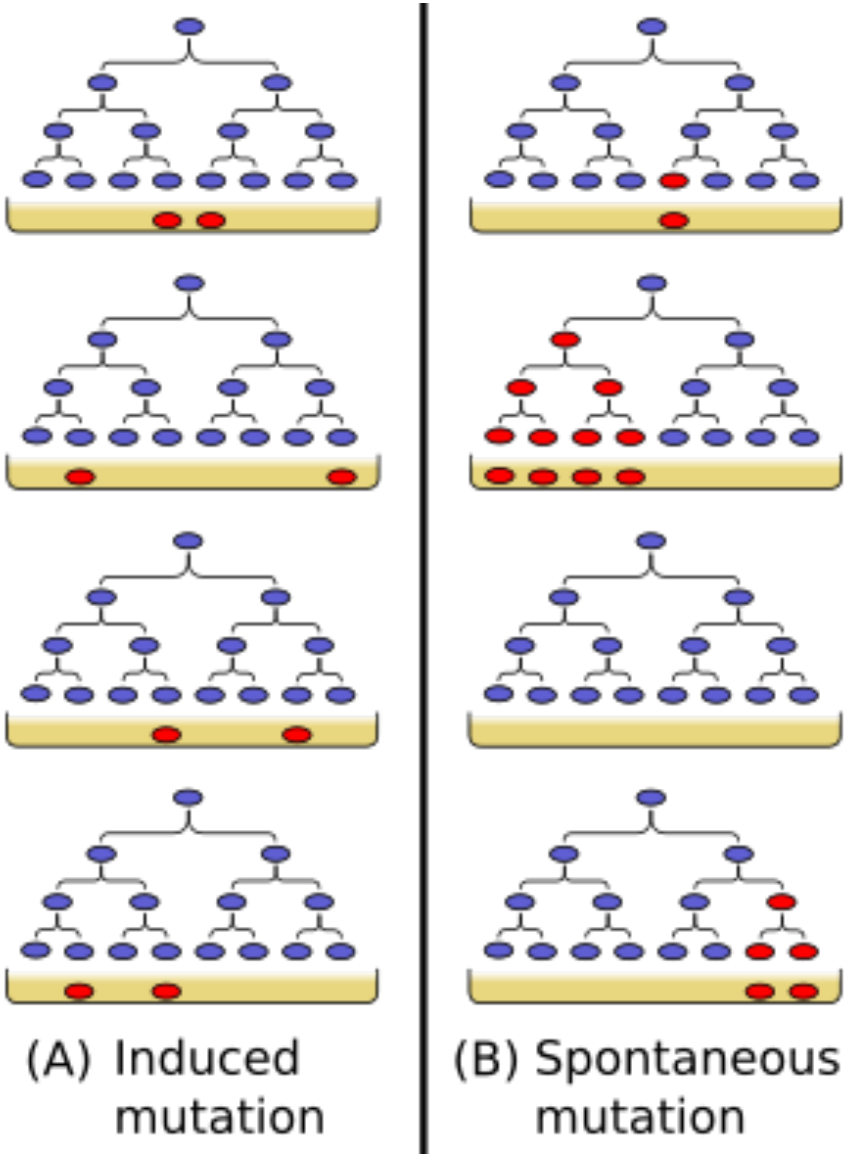


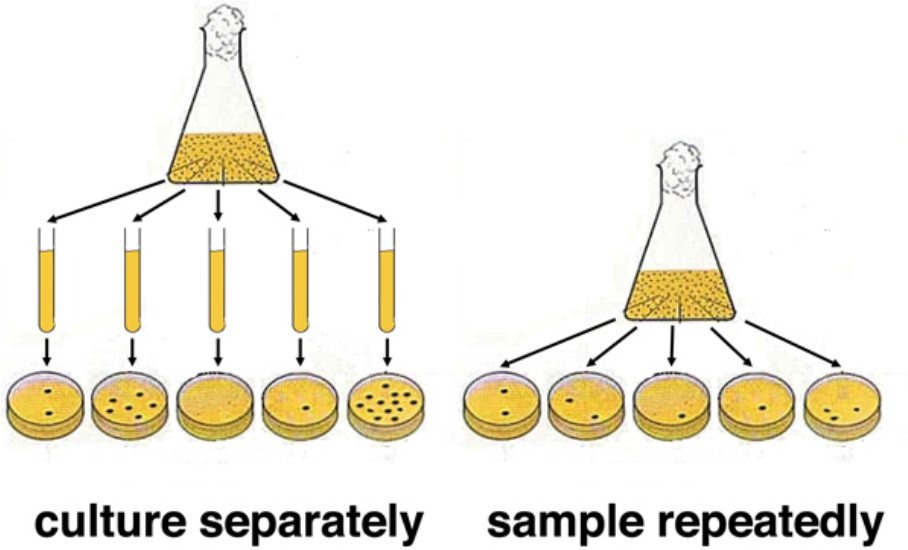
TABLE 1

The number of resistant bacteria in different samples from the same culture.

SAMPLE NO.	EXP. NO. 10a RESISTANT COLONIES	EXP. NO. 11a RESISTANT COLONIES	EXP. NO. 3 RESISTANT COLONIES
1	14	46	4
2	15	56	2
3	13	52	2
4	21	48	1
5	15	65	5
6	14	44	2
7	26	49	4
8	16	51	2
9	20	56	4
10	13	47	7
mean	16.7	51.4	3.3
variance	15	27	3.8
χ^2	9	5.3	12
P	.4	.8	.2

The experiment invalidates the hypothesis of directed mutations

Luria-Delbruck Fluctuation Test



culture separately sample repeatedly

TABLE 2

The number of resistant bacteria in series of similar cultures.

EXPERIMENT NO.	1	10	11	15	16	17	21a	21b
Number of cultures	9	8	10	10	20	12	19	5
Volume of cultures, cc	10.0	10.0	10.0	10.0	.2*	.2*	.2	10.0
Volume of samples, cc	.05	.05	.05	.05	.08	.08	.05	.05
Culture No.								
1	10	29	30	6	1	1	0	38
2	18	41	10	5	0	0	0	28
3	125	17	40	10	3	0	0	35
4	10	20	45	8	0	7	0	107
5	14	31	183	24	0	0	8	13
6	27	30	12	13	5	303	1	
7	3	7	173	165	0	0	0	
8	17	17	23	15	5	0	1	
9	17		57	6	0	3	0	
10			51	10	6	48	15	
11					107	1	0	
12					0	4	0	
13					0		19	
14					0		0	
15					1		0	
16					0		17	
17					0		11	
18					64		0	
19					0		0	
20					35			
Average per sample	26.8	23.8	62	26.2	11.35	30	3.8	48.2
Variance (corrected for sampling)	1217	84	3498	2178	694	6620	40.8	1171

Kinetic proofreading

(Hopfield, PNAS, 1974; Ninio, Biochimie, 1975)

How can differences in affinity of ligands be amplified?

$$C + c \underset{k_C}{\overset{k'_C}{\rightleftarrows}} Cc \xrightarrow{W} \text{correct product} \qquad K_C = k'_C/k_C$$

$$D + c \underset{k_D}{\overset{k'_D}{\rightleftarrows}} Dc \xrightarrow{W} \text{error product} \qquad K_D = k'_D/k_D$$

$$C + c \underset{\underbrace{k_C}_{1}}{\overset{\underbrace{k'_C}_{}}{\rightleftarrows}} Cc \underset{\underbrace{m}_{2}}{\overset{\underbrace{m'}_{}}{\rightleftarrows}} Cc^* \xrightarrow[\underbrace{\quad}_{4}]{\underbrace{W}_{}} \text{product}$$

$$l_C \updownarrow l'_C \Big\}3$$

$$C + c$$

Equilibrium

$$C + c \underset{\underbrace{k_C}_{}}{\overset{\underbrace{k_C}_{}}{\rightleftarrows}} Cc \xrightarrow[\underbrace{m'}_{}]{} Cc^* \xrightarrow[\underbrace{W}_{}]{\underbrace{4}_{}} \text{product}$$

$$\downarrow l_C \Big\}3$$

$$C + c$$

Non-equilibrium: energy expended in the phosphorylation steps

If reactions strongly biased towards dissociation, i.e. they circulate many times before product, then specificity squared

Proofreading costs energy!

Proofreading processes are crucial for the cellular dynamics

Replication: DNA polymerase adds incorrect nucleotides with a rate $\approx 10^{-5}$ which is reduced by proofreading processes to the observed error rate$\approx 10^{-9}$.

Translation: the fidelity of aminoacids in proteins is improved by proofreading with respect to the bare accuracy ensured by tRNAs

Immune discrimination: first self-non self discrimination by T-cells involves a proofreading cascade (McKeithan PNAS 1995; Altan-Bonnet & Germain, PLoS Comp. Biol. 2005; Francois et al, PNAS 2013)

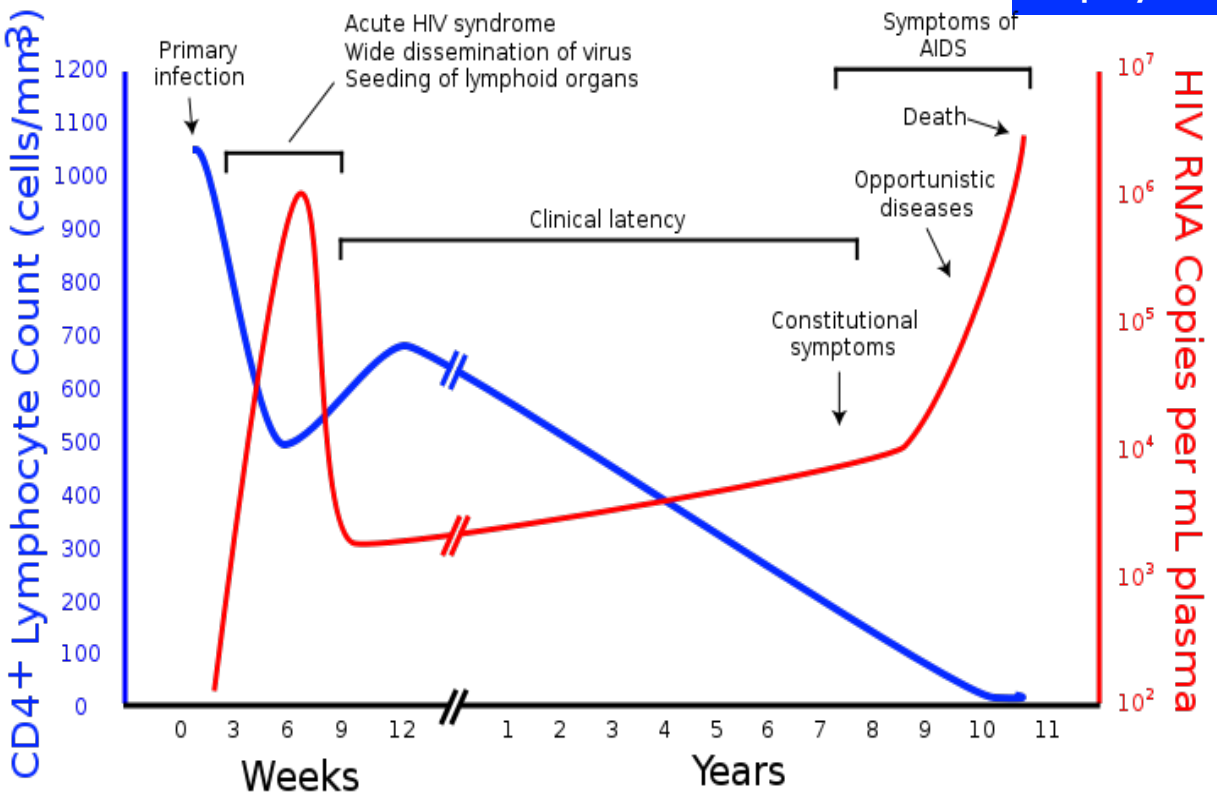
NATURE · VOL 373 · 12 JANUARY 1995

Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection

David D. Ho, Avidan U. Neumann*†, Alan S. Perelson, Wen Chen, John M. Leonard‡ & Martin Markowitz

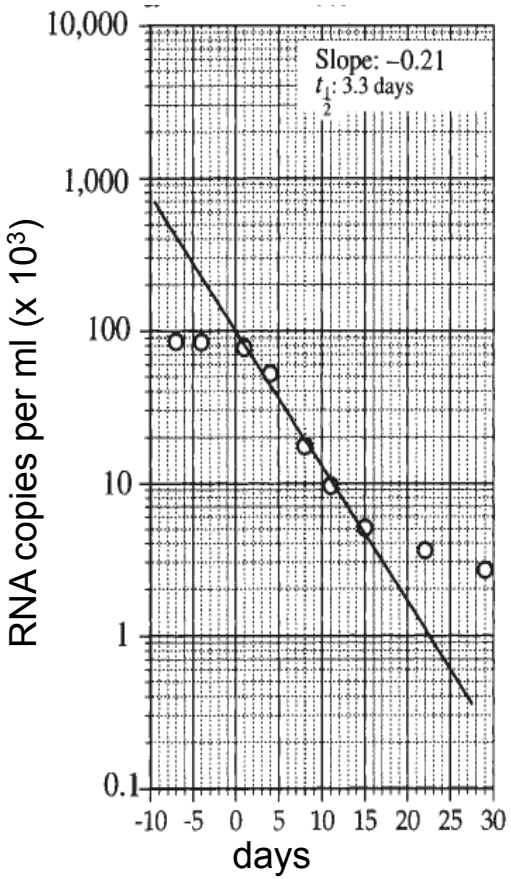
MD with BS in physics

theoretical physicist



AIDS: a disease with long latency (~10 years)

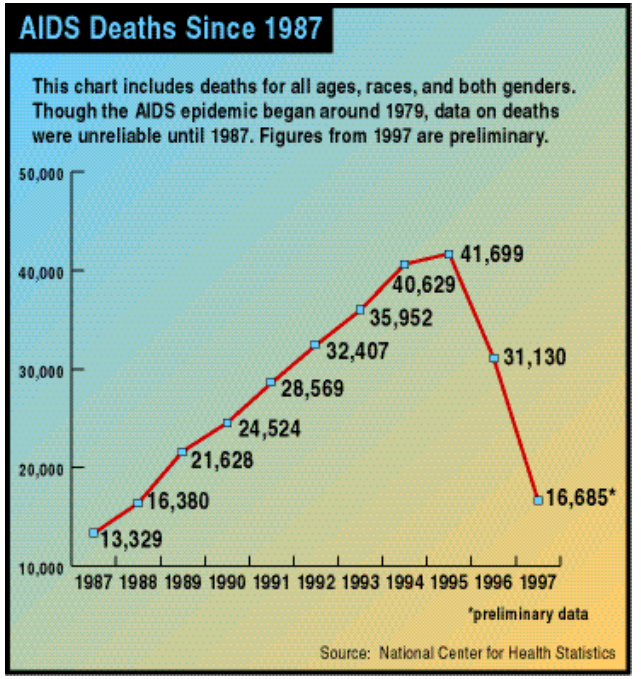
[Ho et al, Nature 1995]



Production-clearance balance perturbed by ABT-538 administered to patients; kinetics assayed (response function of a dynamical system!).

$T_{1/2}$ = 3.3 days

About 10^9 virions/day get cleared and a similar amount of lymphocytes produced!



➔ rapid viral clearance by the immune system: long latency due to balance of fast viral production & clearance

➔ intervention strategy: treat early with multiple drugs

SCIENCE • VOL. 271 • 15 MARCH 1996

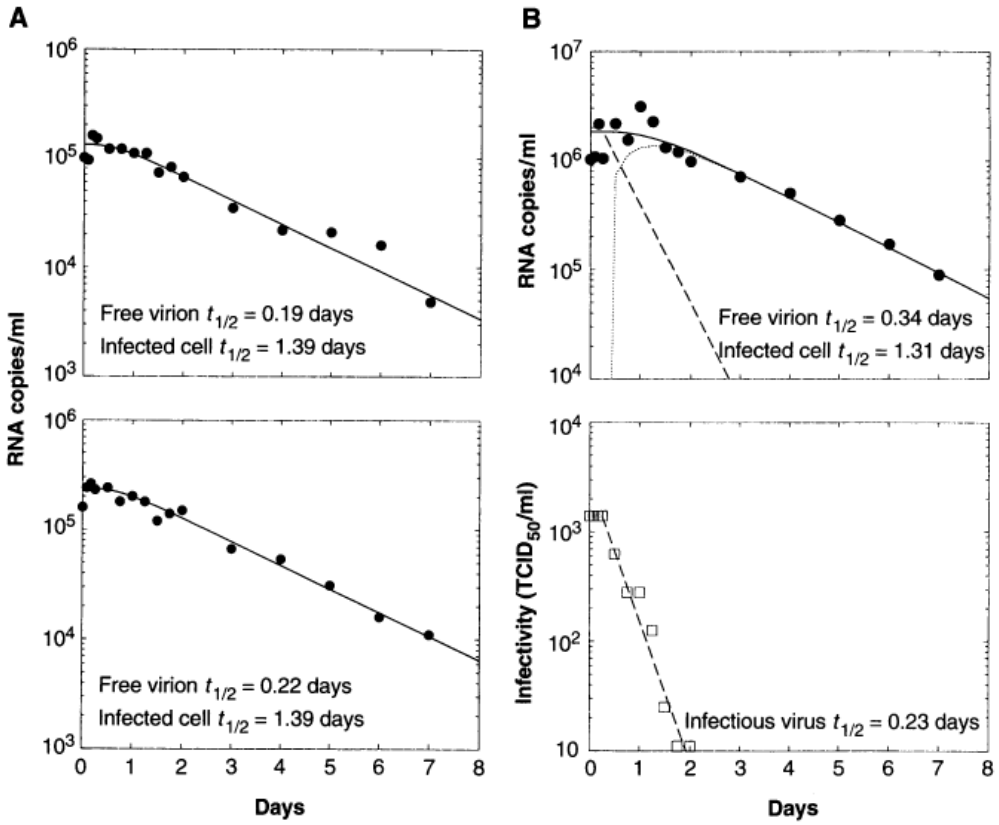
HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time

Alan S. Perelson, Avidan U. Neumann, Martin Markowitz, John M. Leonard, David D. Ho*

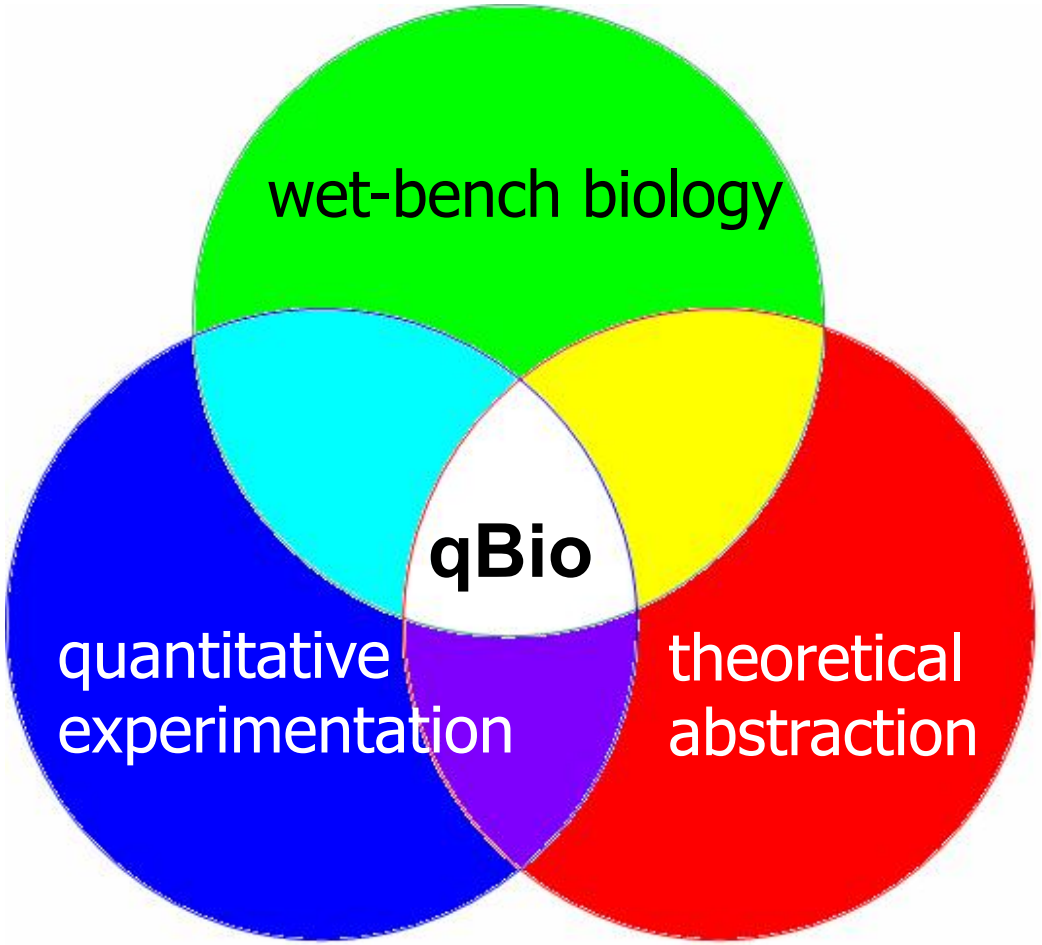
Estimates of the clearance rate *c* of virions and the lifetime *δ* of infected cells

$$V(t) = V_0 \exp(-ct) + \frac{cV_0}{c - \delta}$$

$$\left\{\frac{c}{c-\delta}[\exp(-\delta t) - \exp(-ct)] - \delta t \exp(-ct)\right\}$$



Three key elements of quantitative biology



➔ Collaboration of labs with different expertise

➔ New generation of researchers combining multiple expertise

➔ This course: quantitative molecular biology of bacteria

-- the state of bacterial cells strongly depends on environmental conditions and on a huge number of parameters

-- how can it ever work?

-- how can it be “understood”?

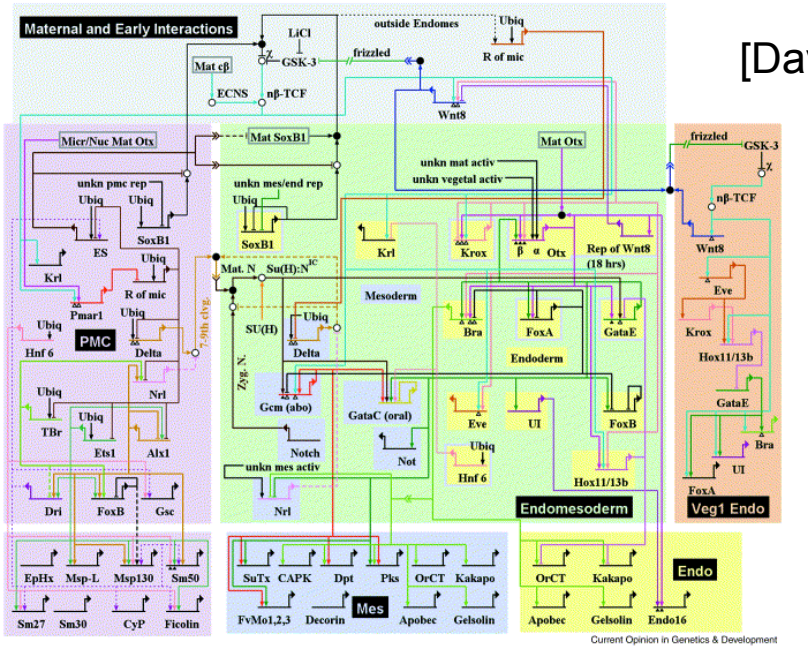
“Now in the further development of science, we want more than just a formula. First we have an observation, then we have numbers that we measure, then we have a law which summarizes all the numbers. But the real glory of science is that we can find a way of thinking such that the law is evident.”

from *The Feynman Lectures on Physics*

Systems biology

1. Scope and focus:

- ➢ biological systems whose <u>functions</u> are derived from the interaction of <u>many sub-components</u>
- ➢ ex: from macromolecluar assemblies to ecological communities
- ➢ current focus: subcellular and cellular processes, e.g., genetic circuits, protein interaction networks



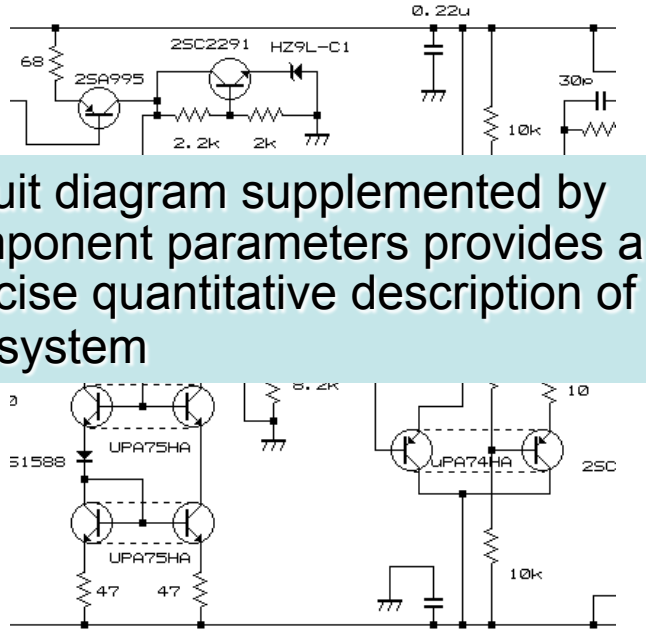
[Davidson et al, 2002]

- ➢ long-term goals:

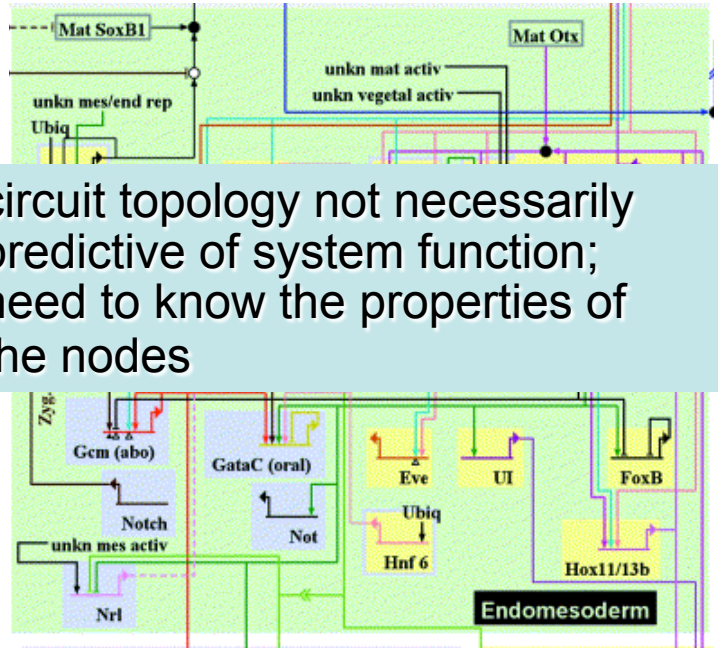
  –mapping out the complete wiring diagram of the cell

  –quantitative, predictive computational model of the cell

2. Circuit diagram as system-level descriptor ?



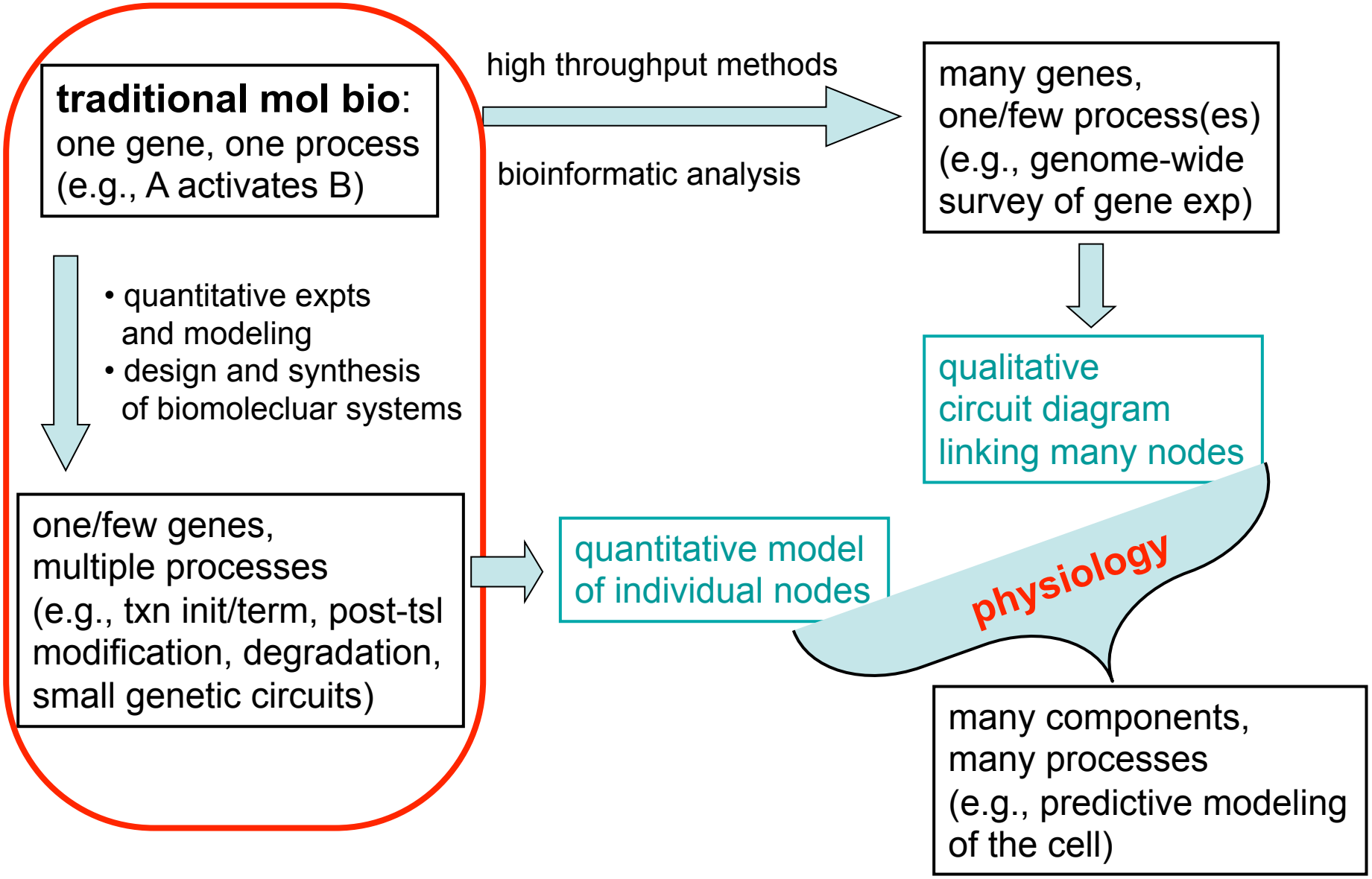
circuit diagram supplemented by component parameters provides a concise quantitative description of the system



circuit topology not necessarily predictive of system function; need to know the properties of the nodes

	electronic circuits	genetic circuits
components	simple & well characterized; many (~10^9); fast (10^{-9} sec)	heterogeneous, most rates unknown; few (~1000); slow (>10 min)
connectivity	physical interconnect between well-insulated components (1~2 inputs per node)	multiply-connected (1~10 inputs per node); regulation at all stages
network complexity	iterated cascades from complex network wiring	combinatorial signal integration from complex molecular control

Experimental & Computational Approaches



Scope of this course

- focus on simple systems (mostly bacteria)
- role of theory, modeling, and computation
- multiple aspects (e.g., tsx initiation, post-tsx control, degradation)
- emphasize **quantitative connections** between molecular and physiological (functional) aspects

❖ course content

– review of molecular microbiology
– molecular interactions: ligand-protein, protein-DNA and protein-protein
– transcriptional initiation control: activation, repression, and combo
– post-transcriptional control: attenuation, termination, degradation
– modeling genetic circuits: bistability and oscillation
– stochastic gene expression and phenotype
– growth physiology and control
– bacterial motility

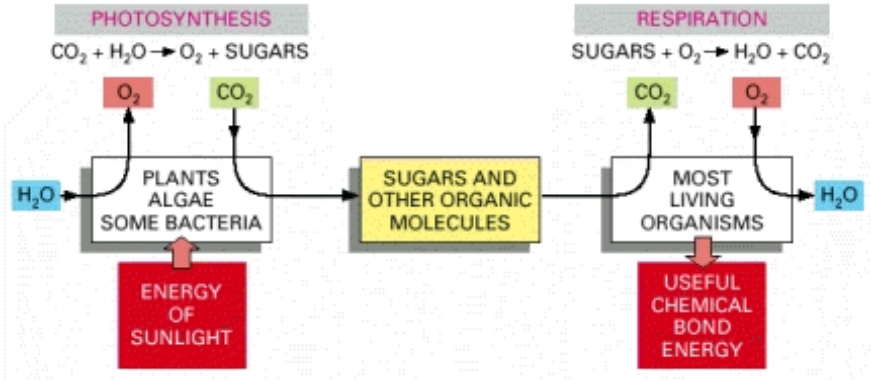
Overview of molecular microbiology

Plan

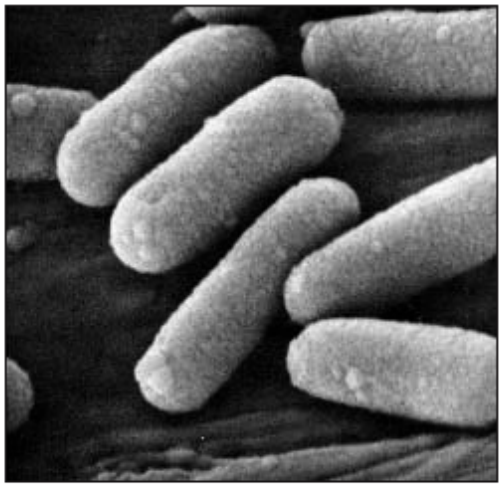
1. biochemical aspects
2. mechanistic aspects
3. regulatory aspects
4. genomic aspects
5. physical aspects
6. comparison to eukaryotes

life of a bacterium: matter + energy ➔ biomass

- chemical composition of biomass: $CH_{1.80}O_{0.43}N_{0.143}$ (+ S, P, Mg, Fe, …)
- molecular composition: [total weight: 10^{-12} g per cell; 70% water]



energy cost of biosynthesis: [minimal medium: ~ 560 J/g dry weight]

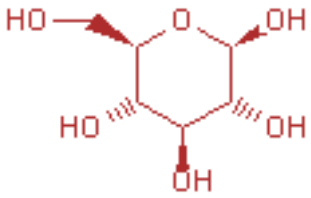


molecular species	%dry weight	energy cost (J/g)
protein	55	220
RNA	20	200
DNA	3	30
small molecules	3	10
lipid membrane	9	80
cell wall	10	20

NB: latent heat of melting H_2O: 334 J/g; cost for proteins is several fold the cost for the peptide bonds holding together the protein: proofreading

❖ metabolism

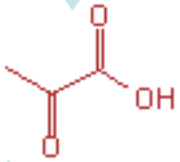
- sequester & breakdown nutrients
  - derive energy
  - generate carbon precursors
  - sequester N, S, P, metals



glucose (6C)

glycolysis

4x ATP+2x NADH



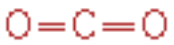
2x pyruvate (3C)

respiration

fermentation

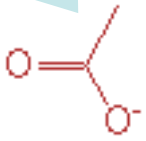
8x ATP

32x ATP



6x CO_2

(need O_2)



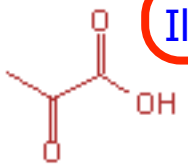
2x acetate + 2x CO_2

➔ Pasteur and Crabtree effects (Sussman et al., 1979)

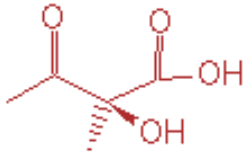
- biosynthesis
  - amino acid
  - nucleic acid
  - lipids
  - co-enzymes

isozymes

IlvBN, IlvGM, IlvIH



CO_2

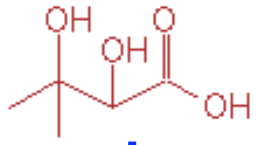


2x pyruvate

H^+, NADPH

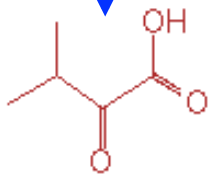
$NADP^+$

IlvC



IlvD

H_2O

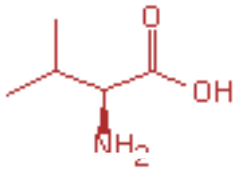


NH_3

aKG

Glu

IlvE



valine

❖ metabolism

- typical biochemical reaction:

  $S + C{\bullet}b \rightleftarrows S{\bullet}b + C$

  S: substrate

  b: component (e.g., CH_3, NH_2, e^-)

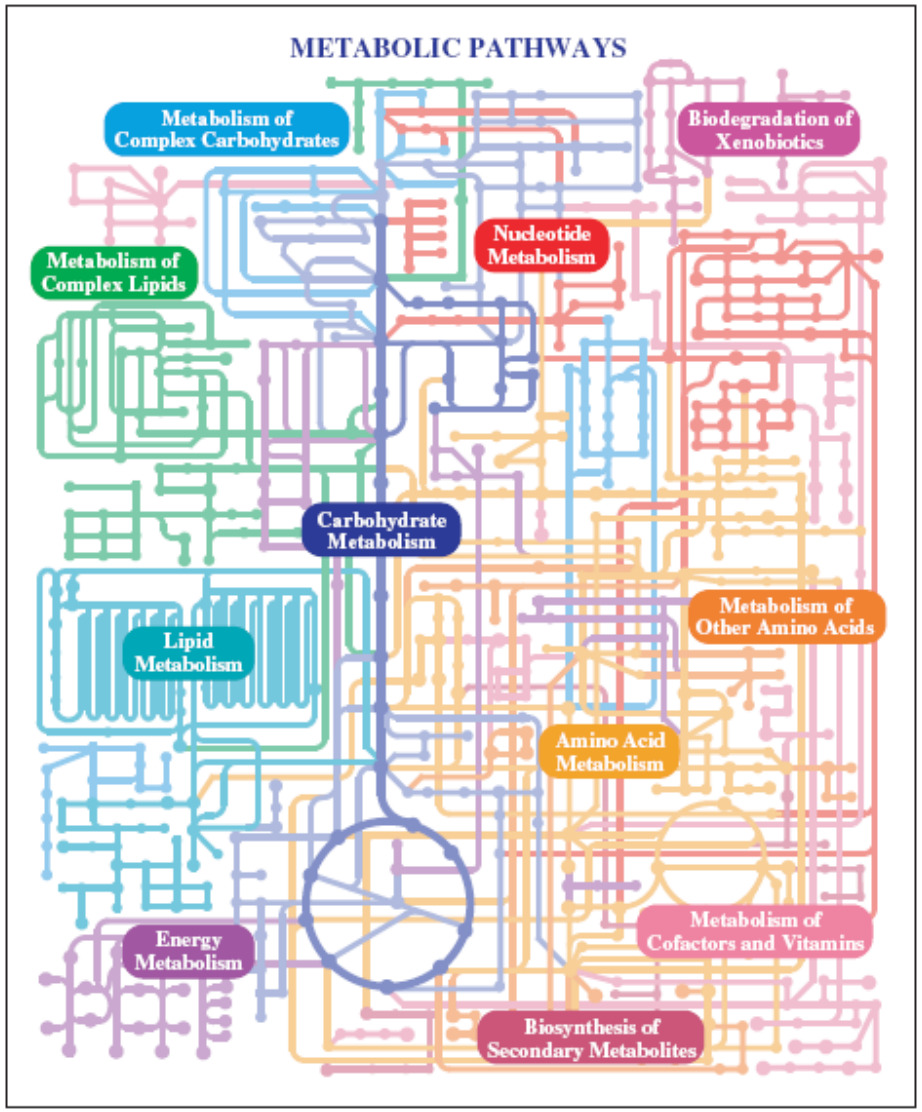
  C: co-enzyme

  (needed for difficult reactions)

➔ most reactions catalyzed by enzymes (proteins)

➔ flux of the products and "by-products" need to be balanced

metabolic control via <u>coordinated regulation</u> of enzyme abundance/ activity



Many enzymes and products are common to different pathways, which branch out of one another. That produces the structure of the graph above. It also leads to interference effects such as growth inhibition.

TABLE 1. *Growth inhibition by amino acids among 356 E. coli strains*[a]

Amino acid tested	No. of strains inhibited
α-Aminobutyric acid	4
Aspartic acid	13
Cystine	40
Histidine	2
Lysine	3
Methionine	4
Norleucine	333
Norvaline	261
Serine	42
Threonine	2
Tyrosine	2
Valine	3

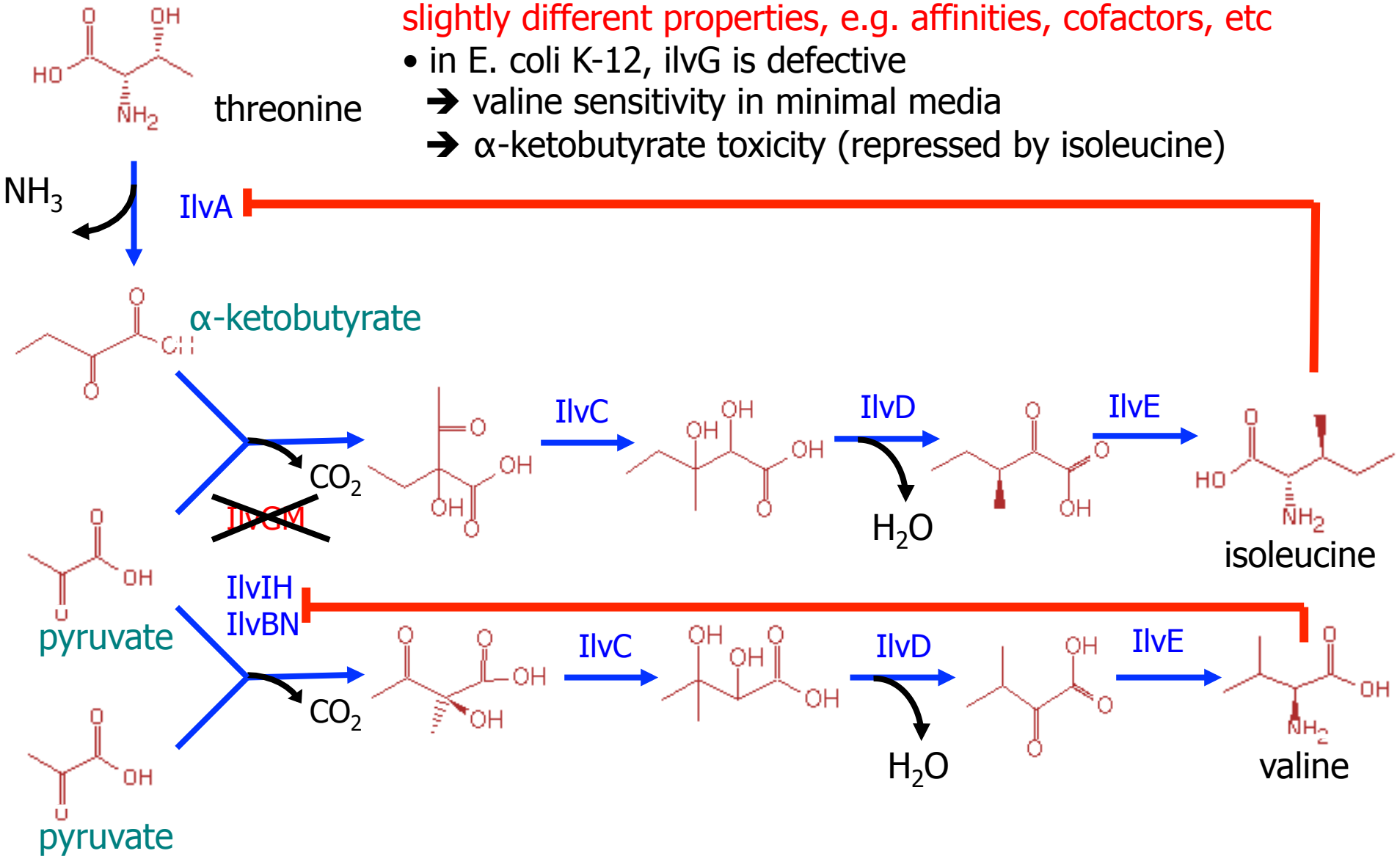
It is very common that growth is inhibited when one amino acid is added to a minimal medium. Growth is often rescued by adding a (specific) second one.

TABLE 2. *Prevention of amino acid growth inhibition by other amino acids in various E. coli strains*[a]

No. of strains tested	Inhibition by:	Prevented by:
1	Aspartic acid	Lysine
1	Aspartic acid	Valine
8	Cystine	Methionine
3	Lysine	Methionine
1	Methionine	Leucine or lysine
12	Norleucine	Methionine
5	Norvaline	Leucine or methionine
4	Serine	Glycine
1	Valine	Isoleucine or leucine
1	Valine	Isoleucine or leucine or methionine

feedback inhibition

- 1st reaction of pathway often inhibited by product
- same enzymes used for synthesis of leucine, isoleucine
- Isozymes: enzymes catalyzing the same reaction but with slightly different properties, e.g. affinities, cofactors, etc
- in E. coli K-12, ilvG is defective
  - ➔ valine sensitivity in minimal media
  - ➔ α-ketobutyrate toxicity (repressed by isoleucine)

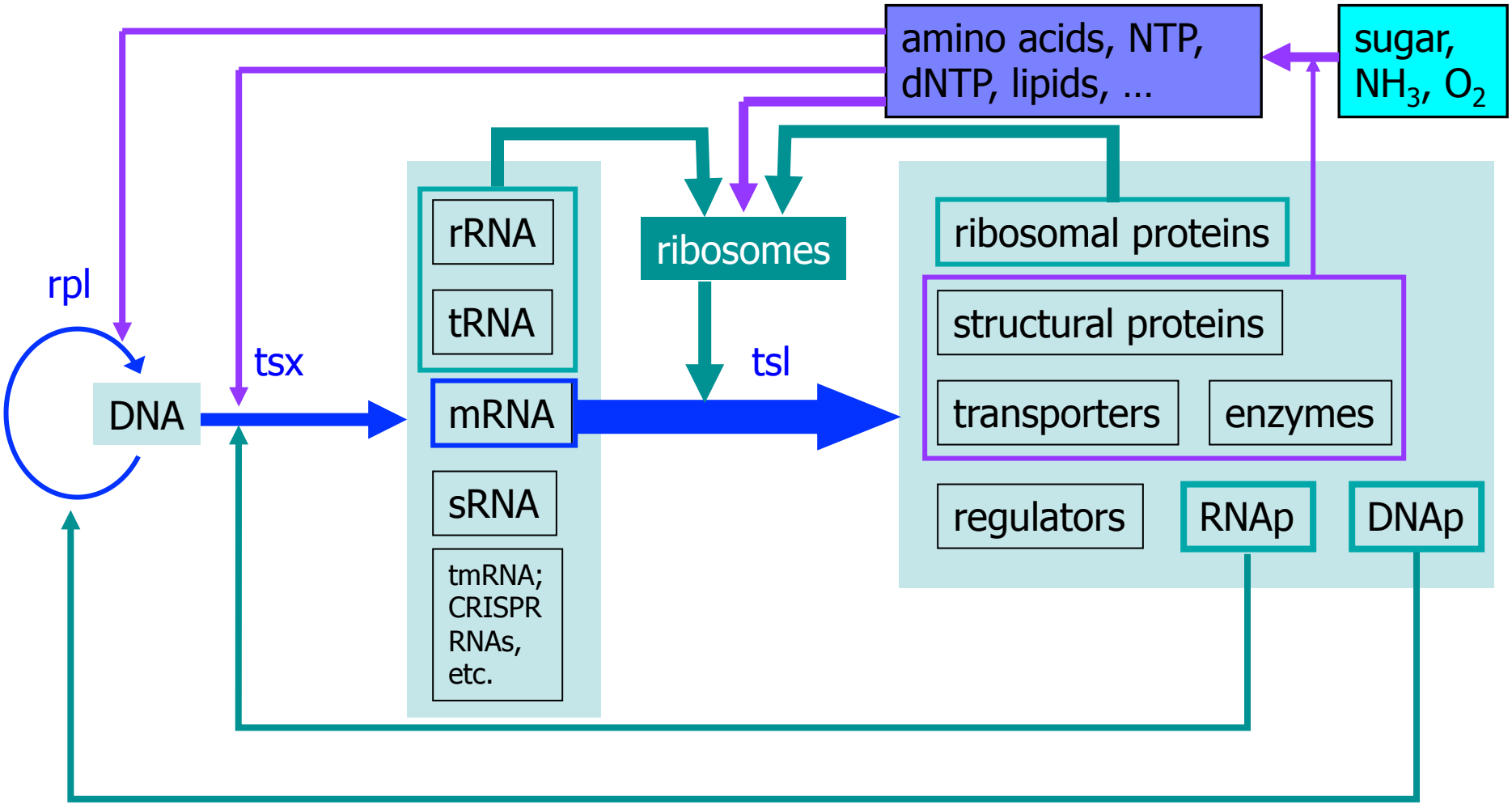


Overview of molecular microbiology

Plan

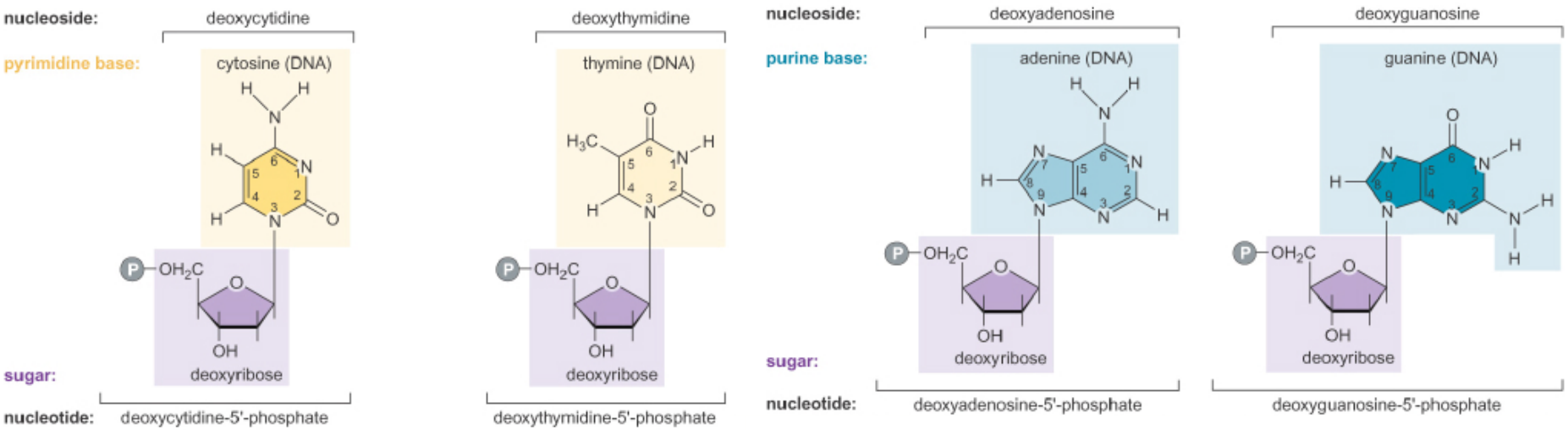
1. biochemical aspects
2. mechanistic aspects
3. regulatory aspects
4. genomic aspects
5. physical aspects
6. comparison to eukaryotes

❖ Central Dogma

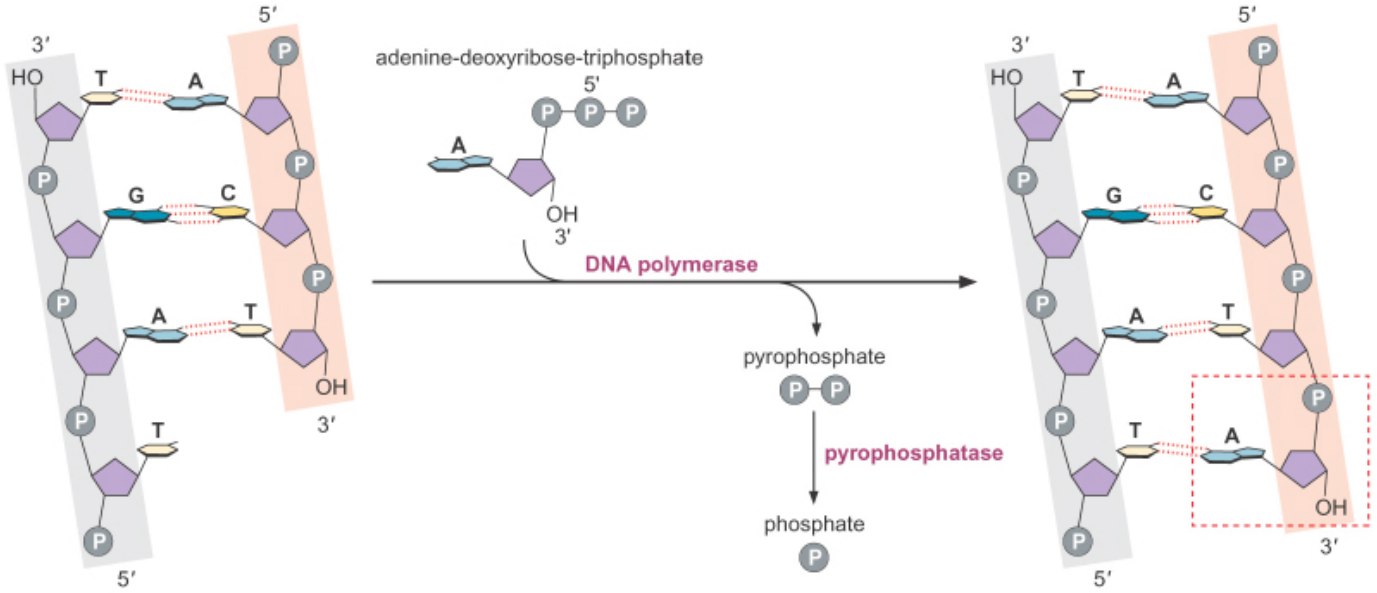


❖ DNA replication

- the four “bases” of DNA: pyrimidines (C, T) and purines (A, G)



- DNA synthesis



- the replication fork

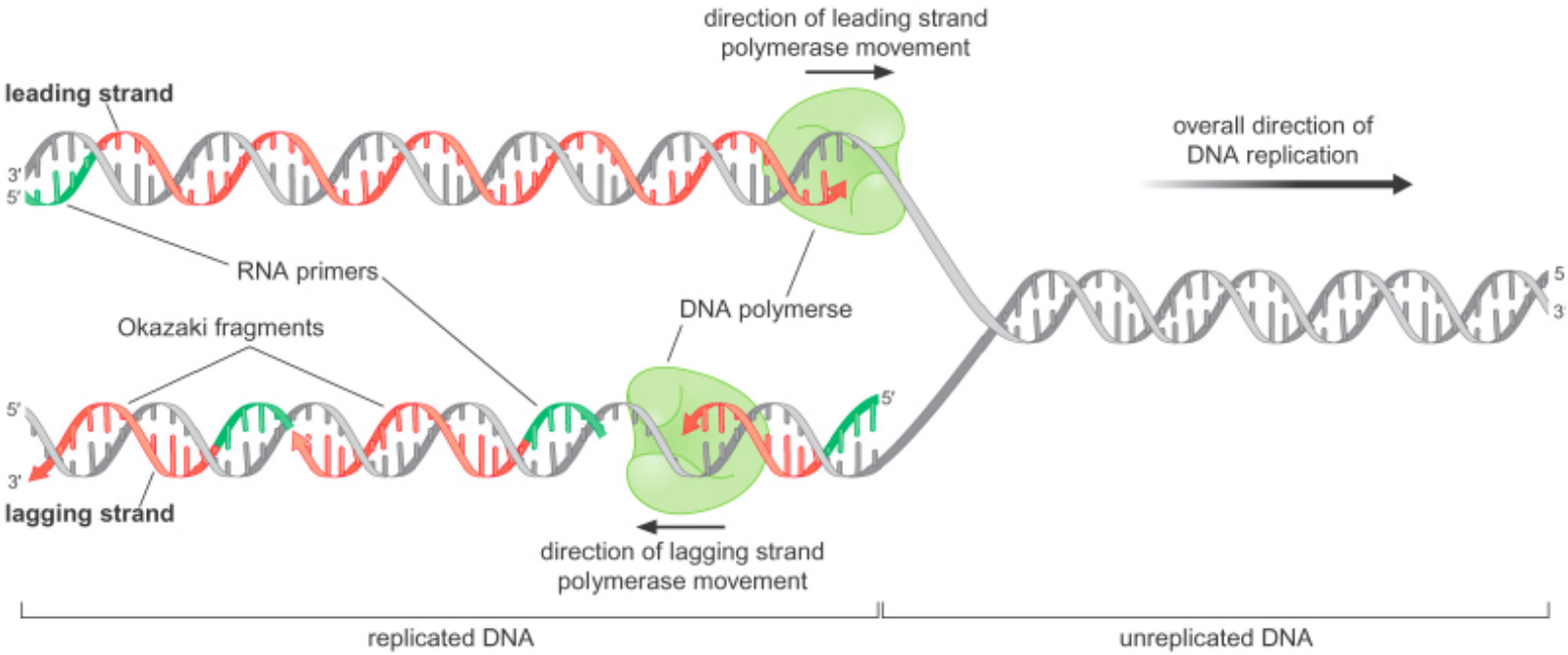


TABLE 8-2 Activities and Functions of DNA Polymerases

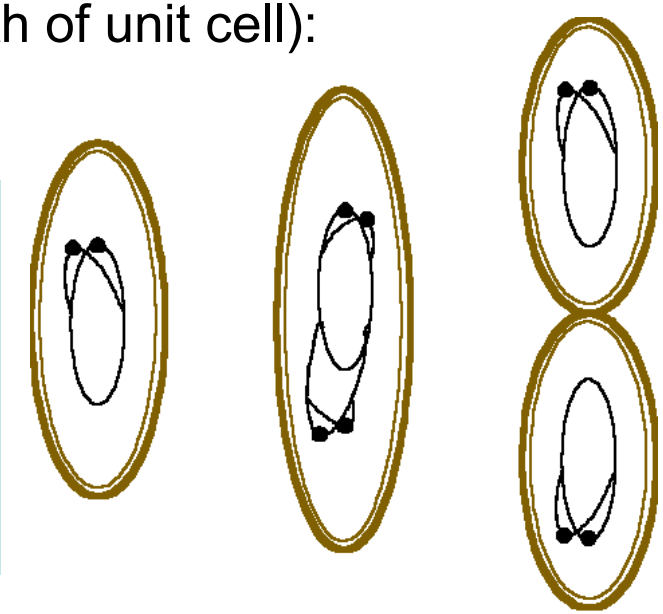
Prokaryotic (*E. coli*)	Number of subunits	Function
Pol I	1	RNA primer removal, DNA repair
Pol II (Din A)	1	DNA repair
Pol III core	3	Chromosome replication
Pol III holoenzyme	9	Chromosome replication
Pol IV (Din B)	1	DNA repair, Trans Lesion Synthesis (TLS)
Pol V (UmuC, UmuD′$_2$C)	3	TLS

- initiation of DNA replication
  - doubling time of *E. coli* can vary over 10x
    [fastest doubling time: ~20 min]
  - 40 min required to replicate chromosome
  - fixed time of 20 min between completion
    of one round of replication and cell division

➔ doubling time > 60 min: waiting time between division & replication

➔ doubling time < 60 min: multiple replication forks

➔ one replication origin every 1.7 μm (length of unit cell):
fast growing cells are larger!

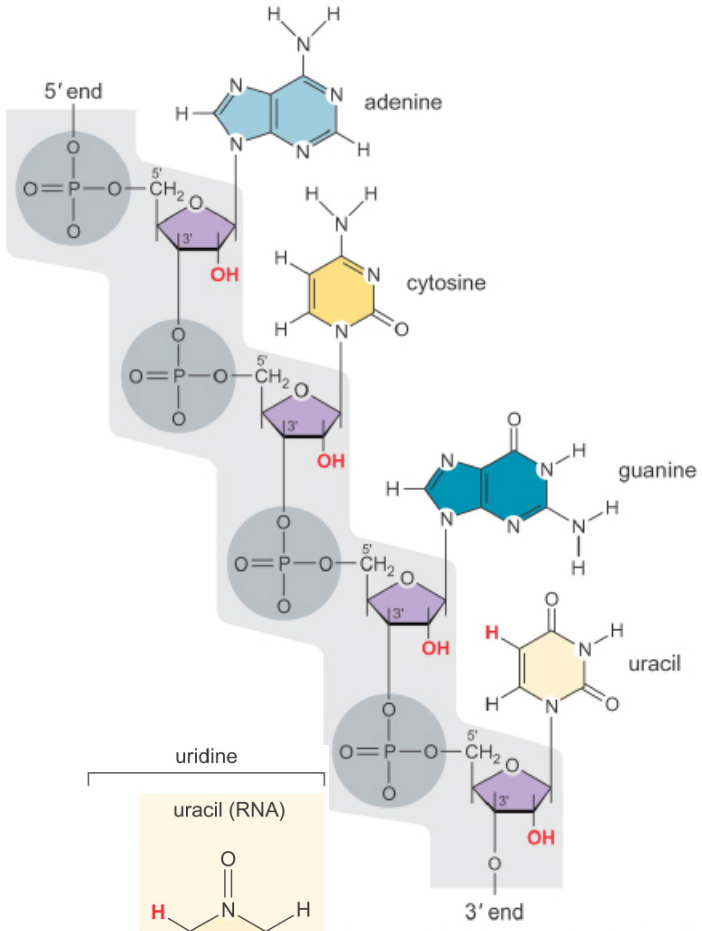


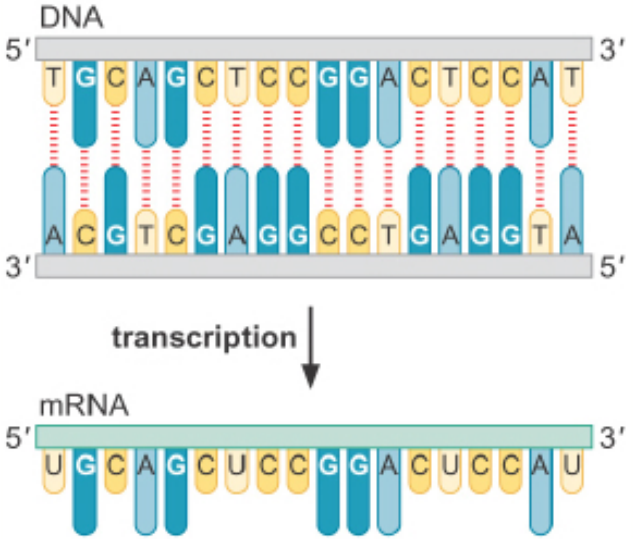
Questions:

- how does the cell "measure" and controls its size?

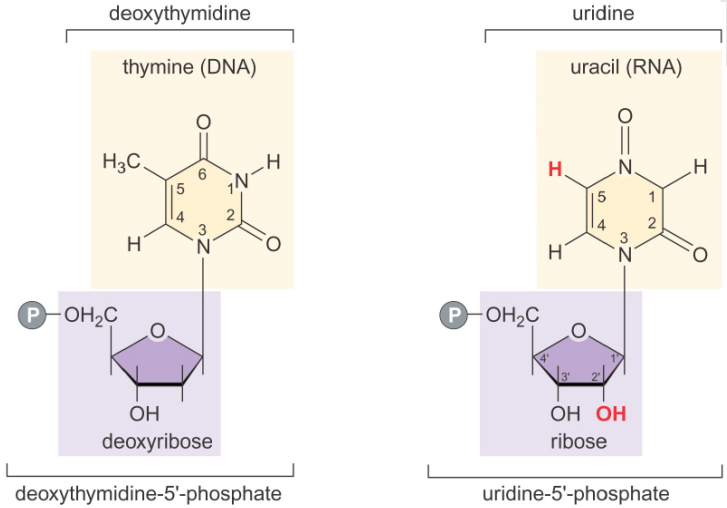
- Position on the chromosome of genes and its relation to expression levels?

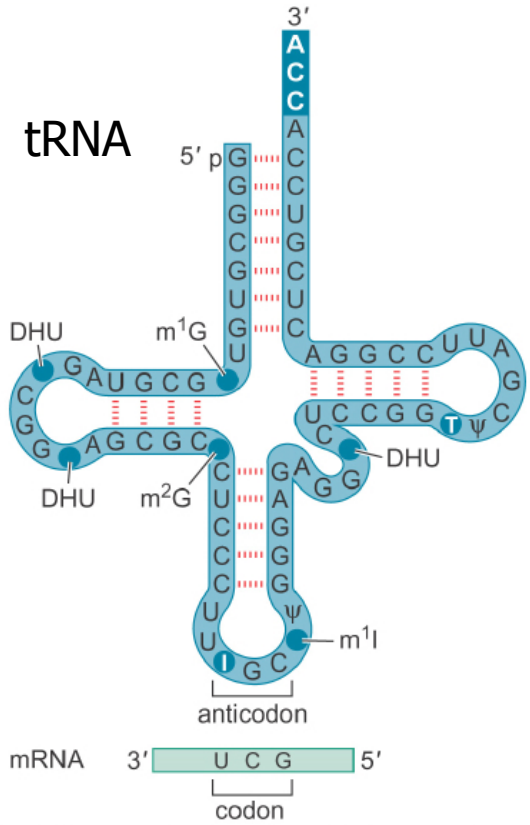
- transcription: DNA to RNA
  - RNA



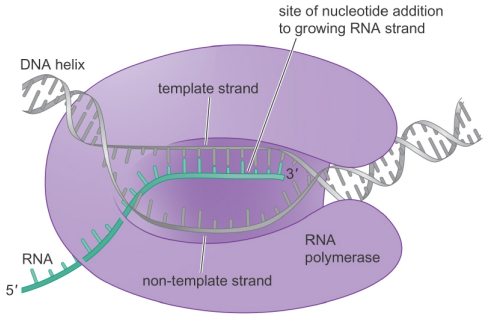


DNA vs RNA:

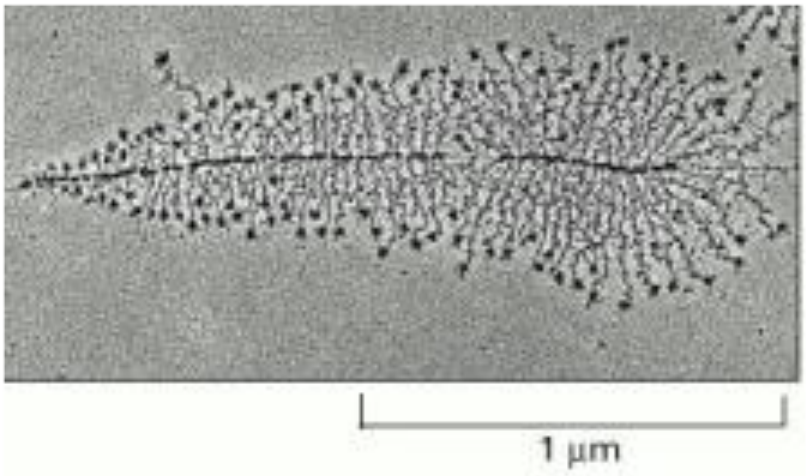




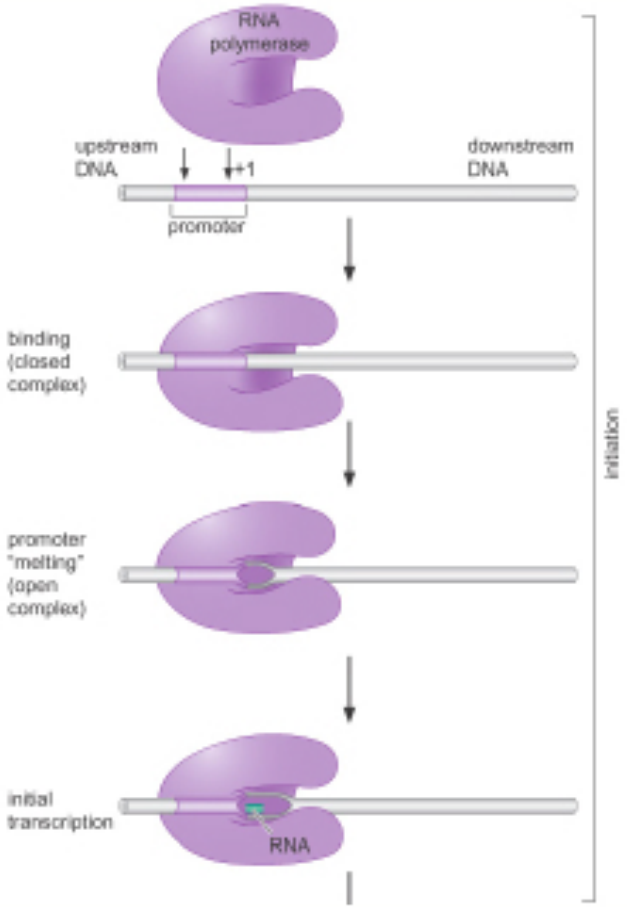
- RNA synthesis:



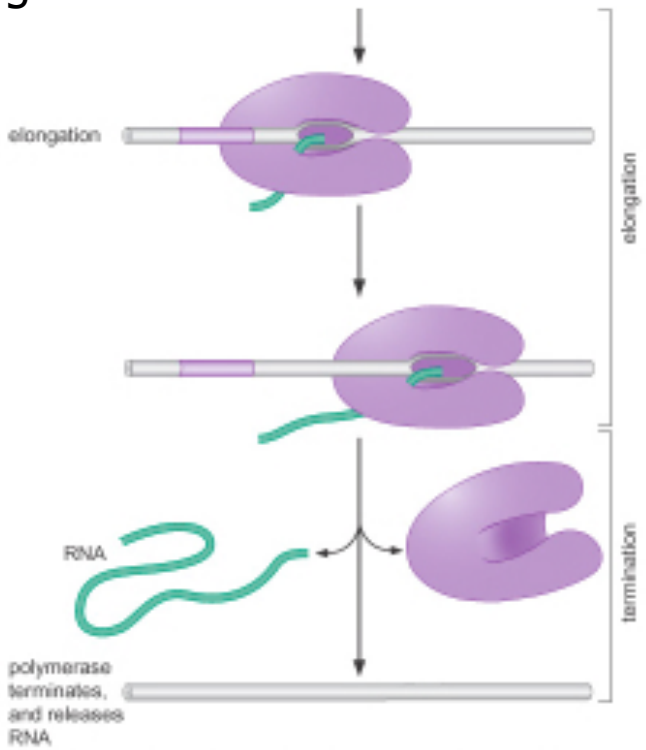
heavily transcribed genes
coding ribosomal RNA



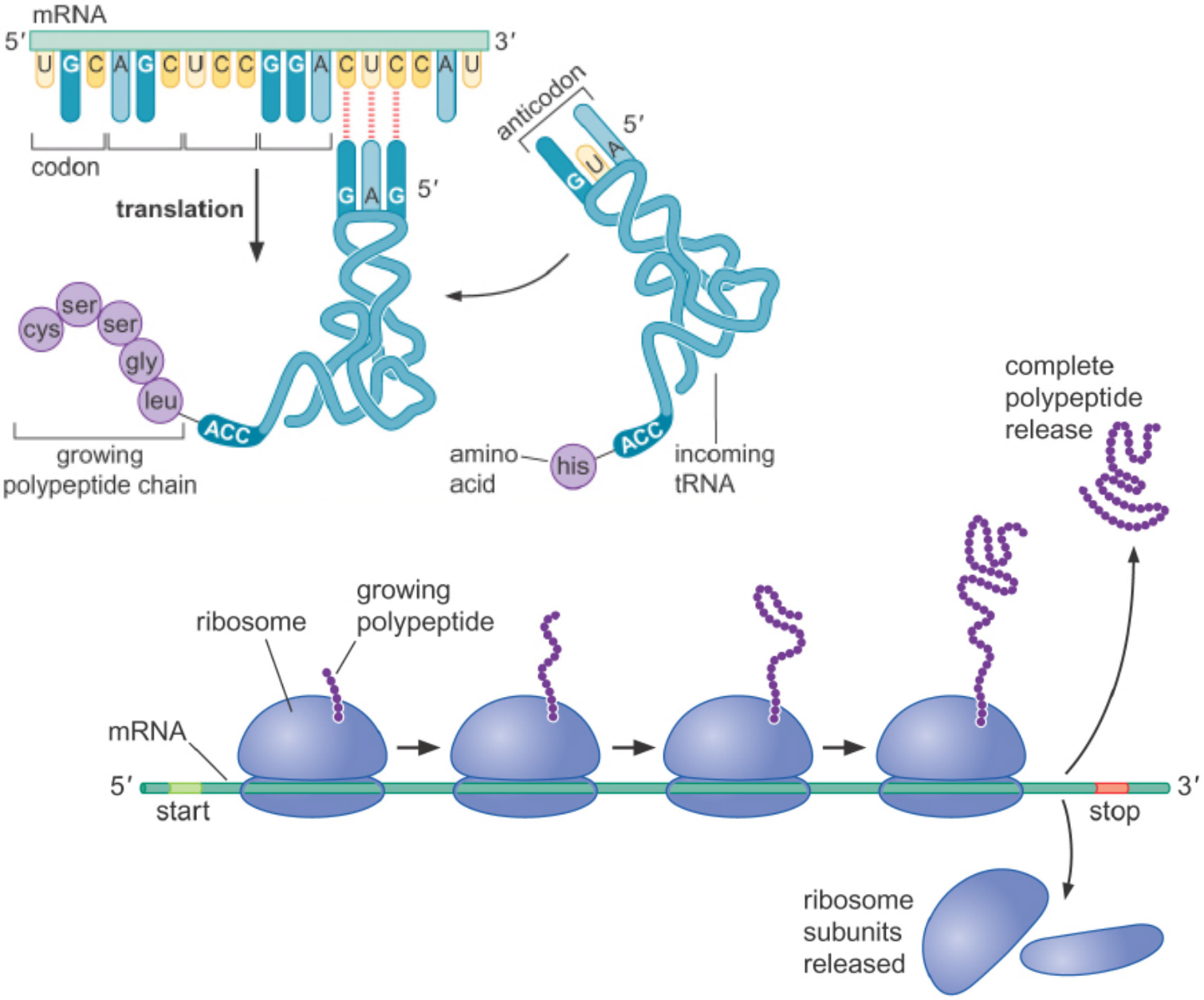
transcriptional initiation



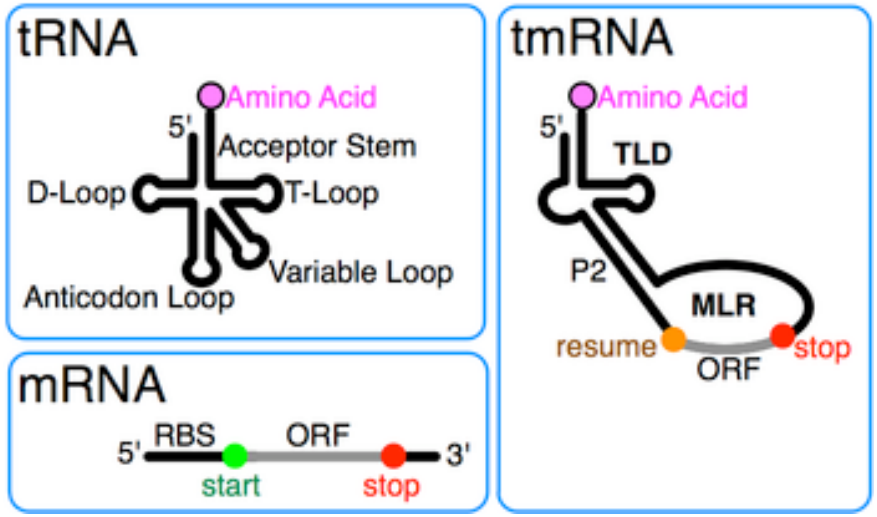
elongation and termination



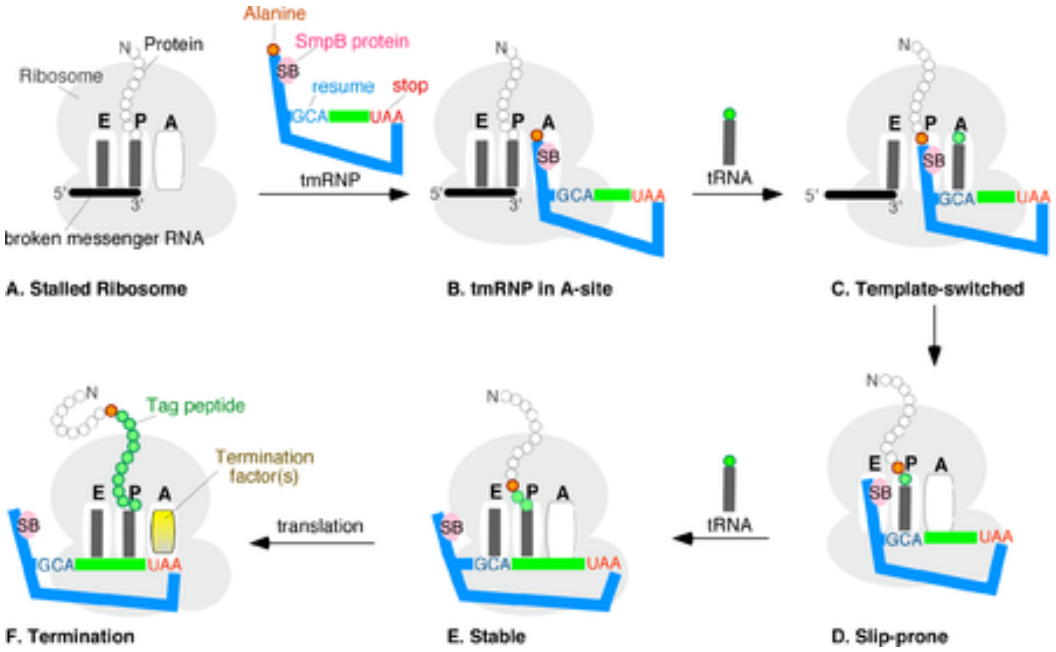
❖ translation: mRNA to protein



❖ translation: tmRNAs



tmRNAs are a structural hybrid of tRNA and mRNA



In *trans*-translation, tmRNA binds to bacterial ribosomes which have stalled, it recycles them and adds a proteolysis-inducing tag to the unfinished polypeptide, which gets degraded.