Quantitative Molecular Biology

PHYS 176/276

Instructor: Terry Hwa

Winter 2021

1

What is quantitative biology?

➔ quantitative biology ≠ biology + numbers/equations

≠ application of quant tools to bio

➔ use numbers as clutches to gain predictive understanding

Why quantitative biology?

-- because biology *is* quantitative

-- needed to formulate and test falsifiable predictions

-- demanded by synthetic biology

Role of theory

- formulate expectation and predictions (via quantitative model)
- guide the design of new experiments and technology
- power: the generality of (falsifiable) ideas, not necessarily math
  [e.g., Cupernicus, Darwin, Einstein]
- "cost" : the simplifying assumptions, not necessarily forced by math, but required in order to reveal principles

➔ **This course: quantitative molecular biology of bacteria**

2

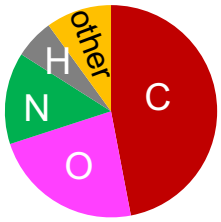
❖ Life of a bacterium:

matter + energy ➔ biomass

TABLE 1. Typical elemental composition of biological specimen

Element	Mass fraction in the following cells	
	Tissue[a]	Bacteria[b]
C	0.50	0.47
N	0.16	0.14
H	0.07	0.06
O	0.25	0.23
P + S + others	0.02	0.10[c]

[Heldal et al, 1985]

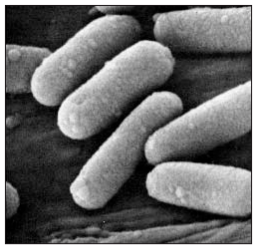


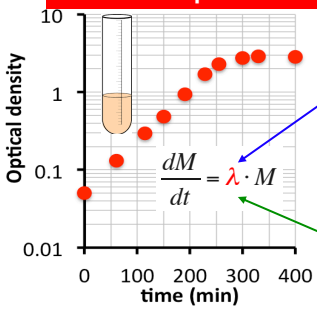
- molar composition: $CH_{1.5}O_{0.35}N_{0.24}$ (+S, P, Mg, Fe, …)
- algae (photosynthesis):
  $CO_2 + H_2O + N_2 + photons$ ➔ $biomass + O_2$
- *E. coli* (minimal medium):
  $glucose + NH_3$ ➔ $biomass + CO_2$

3

growth of E. coli

Can we predict GR & yield?





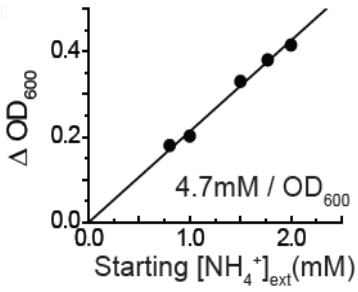
environmental factors:
nutrient types & conc
temperature, pH,
osmolarity, drugs, …

genetic factors:
enzymes & regulation

Learning from the growth curve

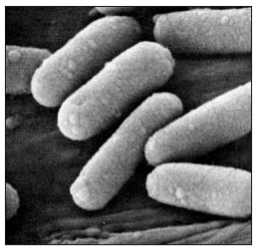
[Monod, Ann Rev Microb. 1949]

- OD_{600} = biomass content
  [1 OD·ml = 0.5mg CDW]
- saturation OD ➔ yield
- (lag: transition from pre-shift phase)

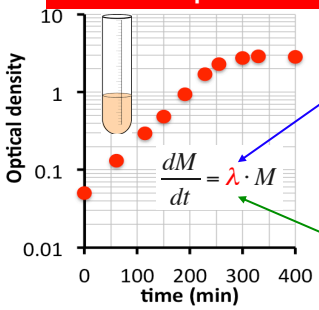


4

growth of E. coli



Can we predict GR & yield?

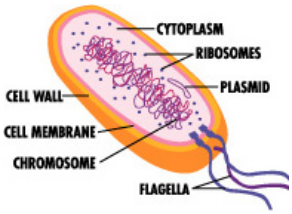


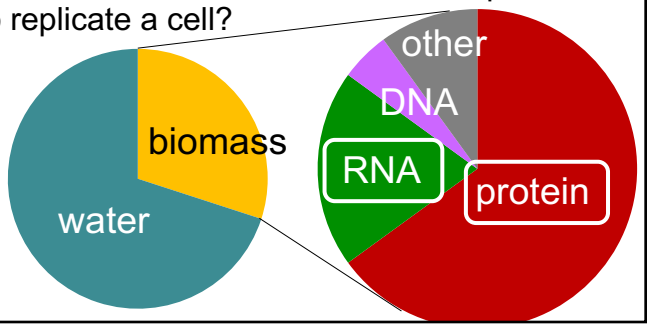
environmental factors:
nutrient types & conc
temperature, pH,
osmolarity, drugs, …

genetic factors:
enzymes & regulation

What does it take to replicate a cell?

condition-dependent

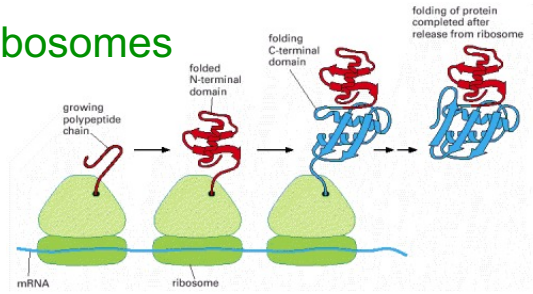




5

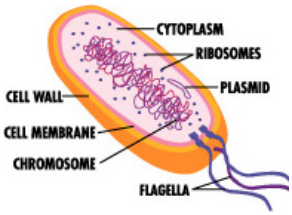
protein = defined sequence of 20 amino acids

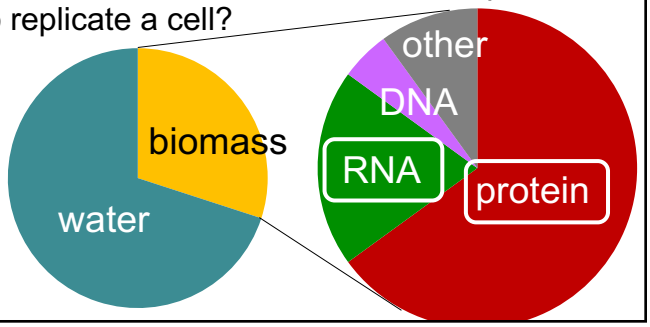
protein synthesis: ribosomes



What does it take to replicate a cell?

condition-dependent

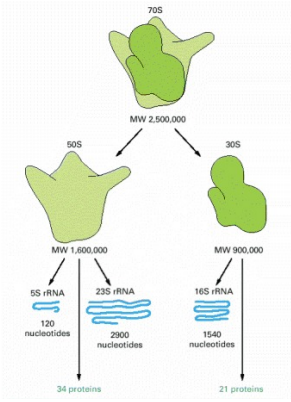




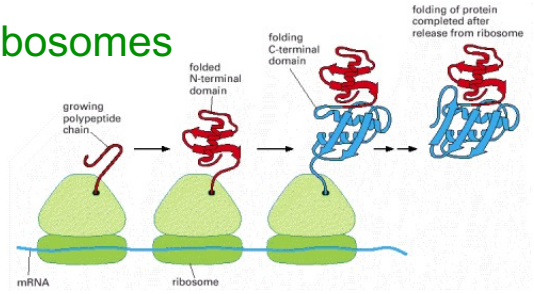
6

protein = defined sequence of 20 amino acids

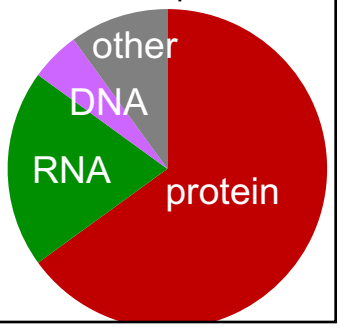
protein synthesis: ribosomes



>85% of all RNA
up to 1/3 of all proteins



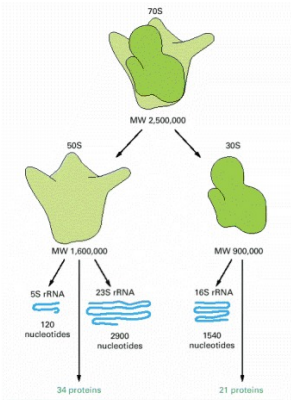
condition-dependent



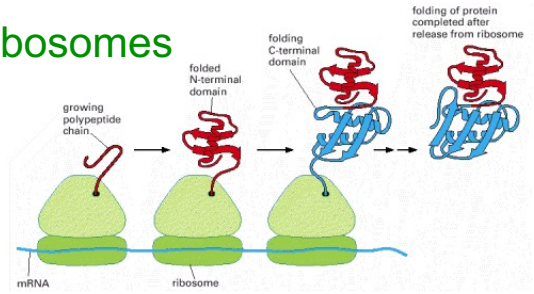
7

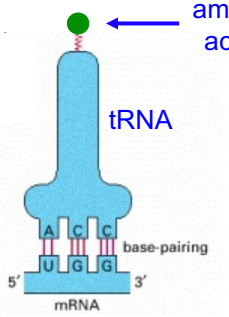
protein = defined sequence of 20 amino acids

protein synthesis: ribosomes

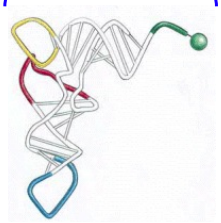


>85% of all RNA
up to 1/3 of all proteins





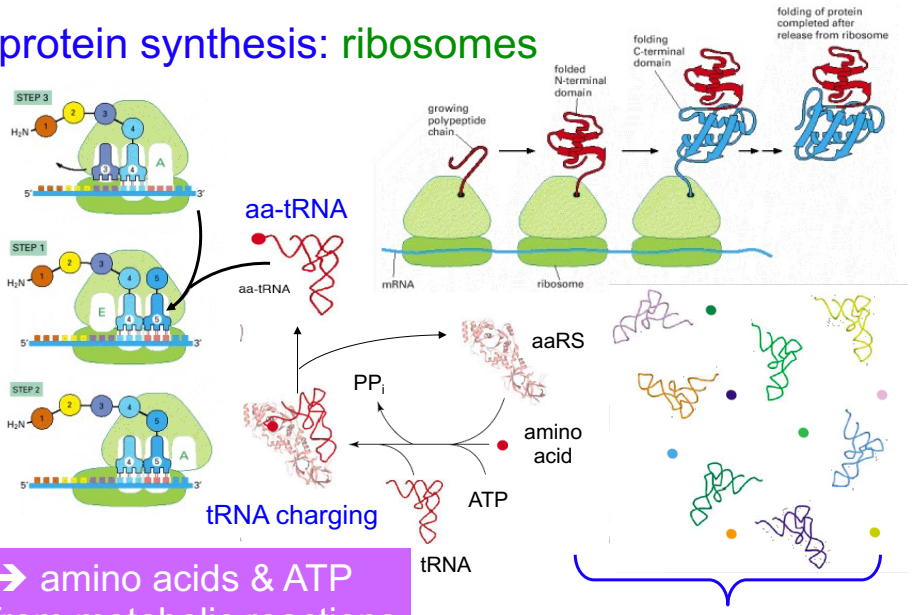
charged tRNA



8

protein = defined sequence of 20 amino acids

protein synthesis: ribosomes

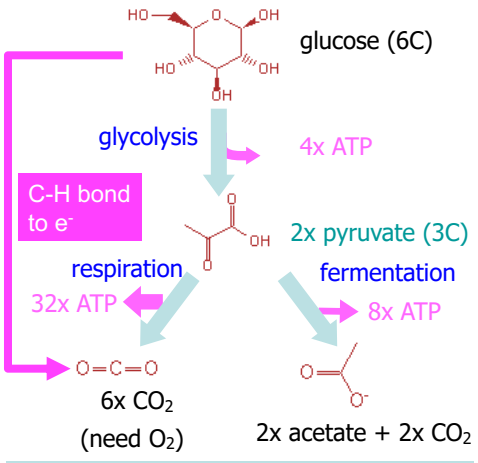


➔ amino acids & ATP from metabolic reactions

9

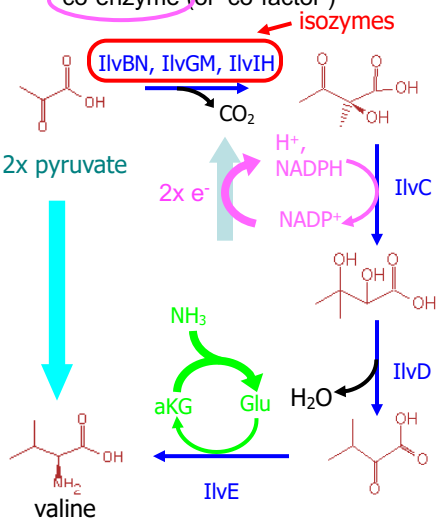
❖ metabolism

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



➔ but many organisms use fermentation even with oxygen (Crabtree effect); why?

- biosynthesis ("precursors" to "building blocks")
  - amino acid
  - nucleic acid
  - lipids
  - co-enzyme (or 'co-factor')



10

❖ metabolism

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

catabolism

- biosynthesis of building blocks
  - amino acid
  - nucleic acid
  - lipids
  - co-enzymes

anabolism

- degradation/recycling (e.g., mRNA)
- typical biochemical reaction:

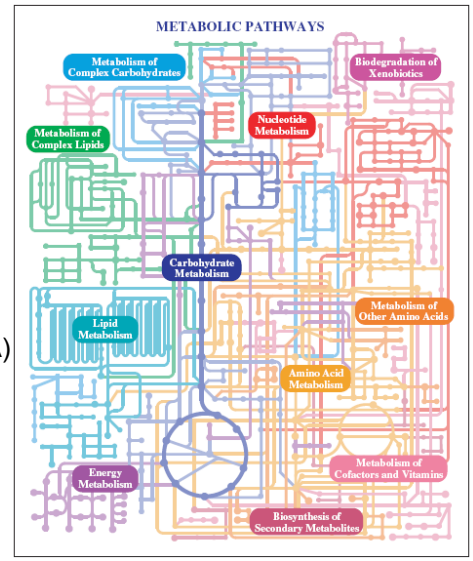
  S + C•b ⇄ S•b + C

  S: substrate

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

  C: co-enzyme

  (needed for difficult reactions)



➔ most reactions catalyzed by enzymes (proteins)

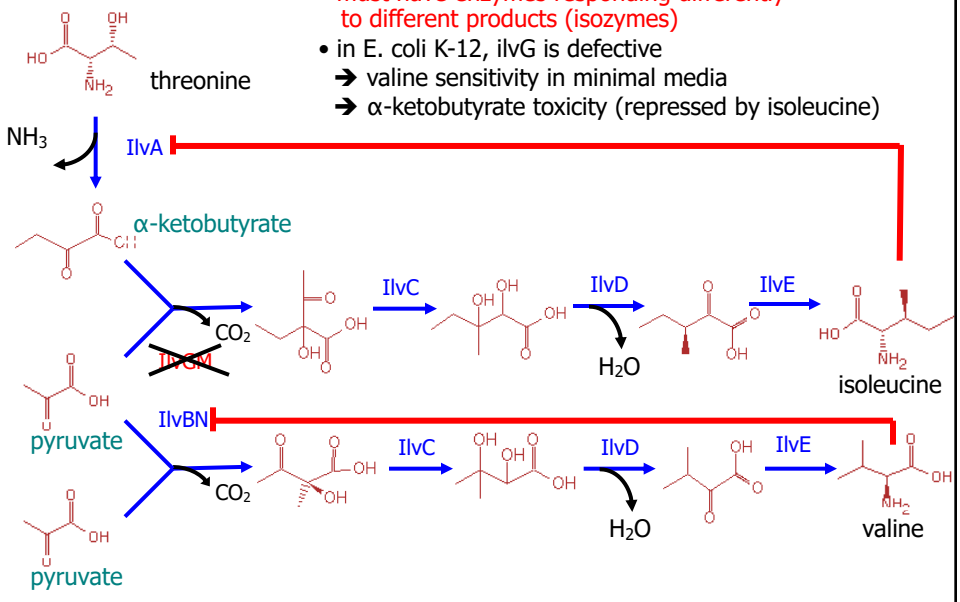
➔ flux of the products and "by-products" need to be balanced

metabolic control via coordinated regulation of enzyme abundance/activity

11

feedback inhibition

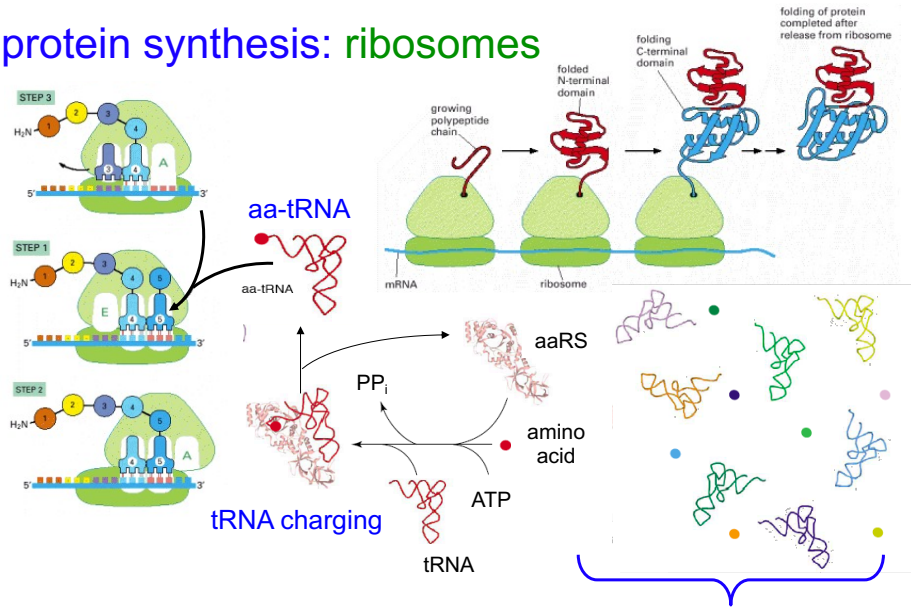
- 1st reaction of pathway often inhibited by product
- same enzymes used for synthesis of valine and isoleucine
- must have enzymes responding differently to different products (isozymes)
- in E. coli K-12, ilvG is defective
  - ➔ valine sensitivity in minimal media
  - ➔ α-ketobutyrate toxicity (repressed by isoleucine)



12

protein = defined sequence of 20 amino acids

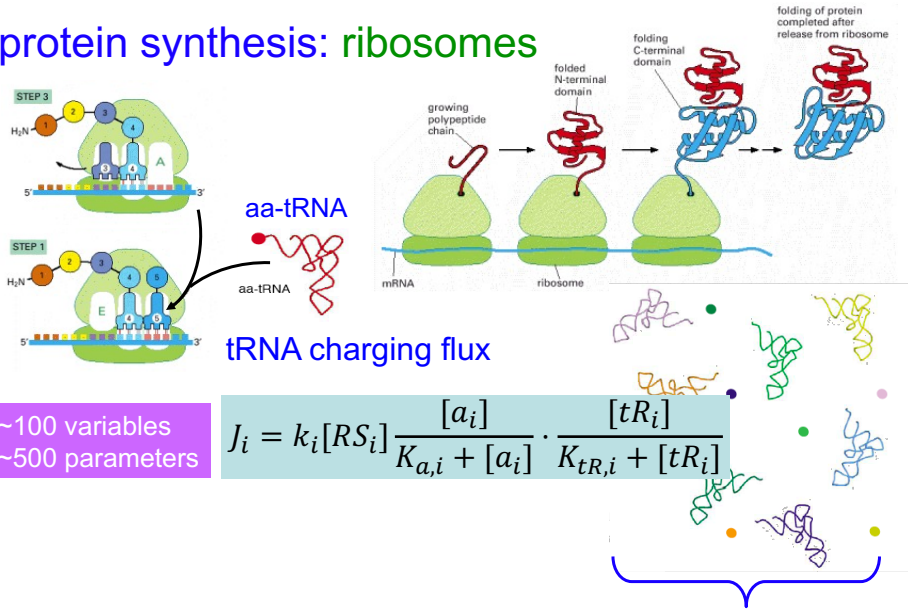
protein synthesis: ribosomes



13

protein = defined sequence of 20 amino acids

protein synthesis: ribosomes



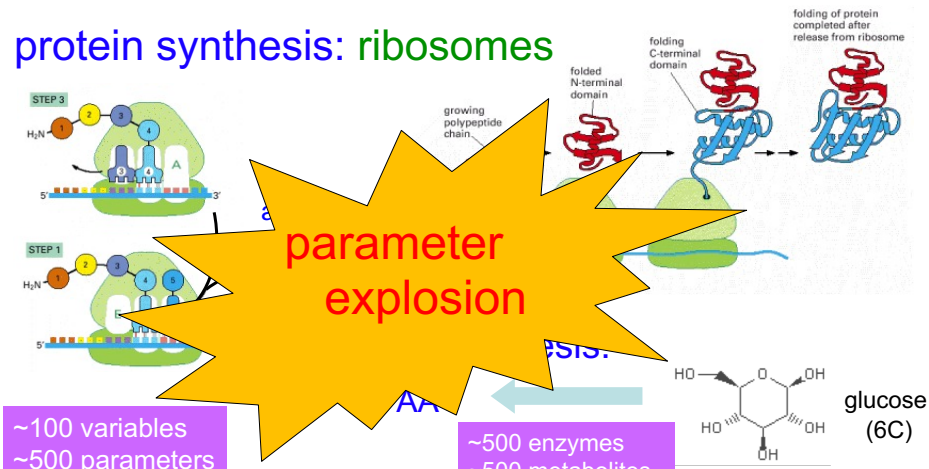
~100 variables
~500 parameters

$$J_i = k_i[RS_i]\frac{[a_i]}{K_{a,i}+[a_i]}\cdot\frac{[tR_i]}{K_{tR,i}+[tR_i]}$$

14

protein = defined sequence of 20 amino acids

protein synthesis: ribosomes

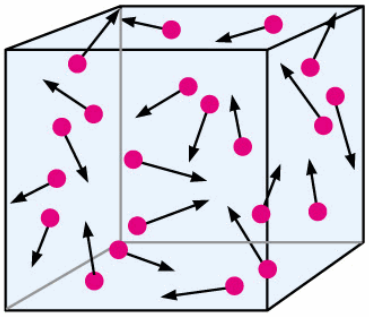


regulation: when and how much proteins to make

dependence on temp, pH, osmolarity, …

15

How to deal with exploding no. of parameters?



Newtonian Mechanics → Thermodynamics

dimension reduction?

$$\frac{d^2\vec{r}_i}{dt^2} = \vec{f}_{ij}(\vec{r}_i - \vec{r}_j)$$

$$PV = nRT$$

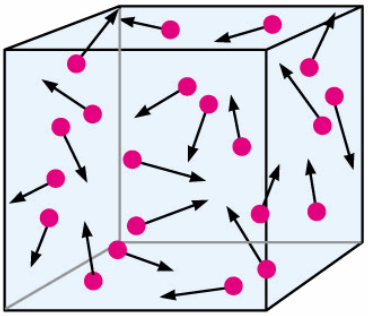
Need moles of parameters:

$\vec{r}_i(t=0) =$ …

$\vec{v}_i(t=0) =$ …

16

How to deal with exploding no. of parameters?



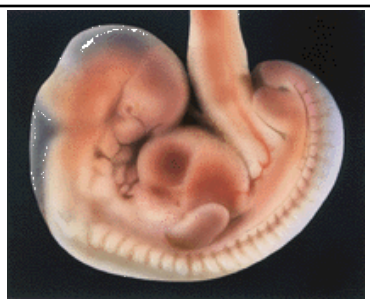
Newtonian Mechanics ⇒ Thermodynamics

dimension reduction?

Statistical Mechanics

$\rho(\vec{q}_i, \vec{p}_i; t) \xrightarrow{t\to\infty} \rho_{eq}(\mathcal{H}(\vec{q}_i, \vec{p}_i))$

17

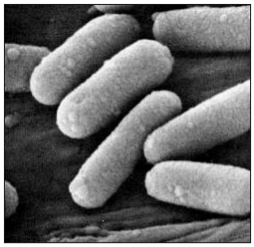


Molecular/Cell Biology ⇒ Physiology

dimension reduction?

Quantitative Systems Biology

18



Molecular/Cell Biology → Physiology

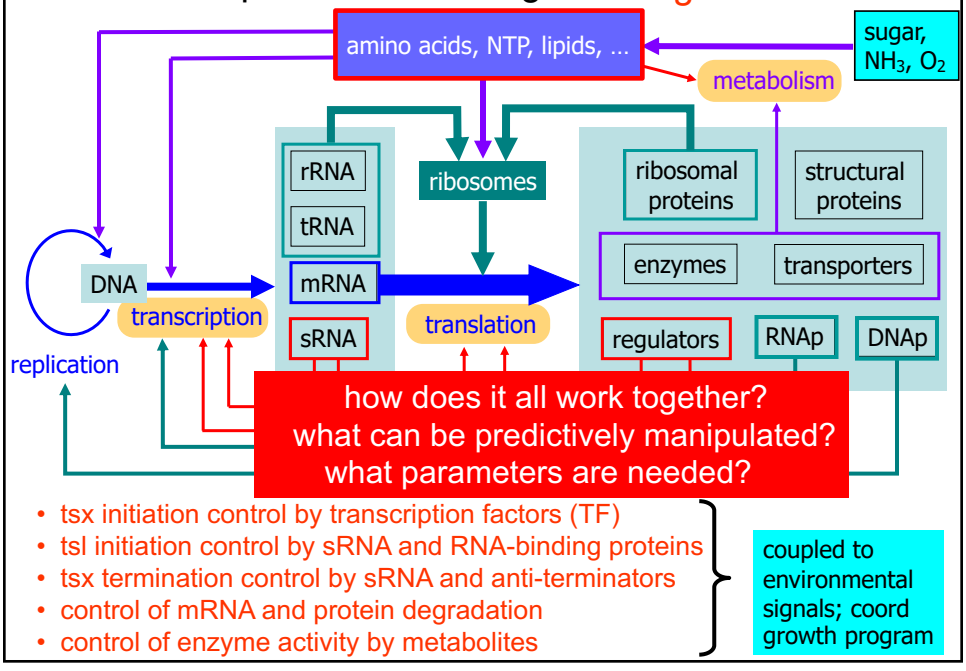
dimension reduction?

Quantitative Systems Biology

This course: explore cellular strategies of dimension reduction using bacterial mechanisms

19

Molecular aspect: Central Dogma + regulation

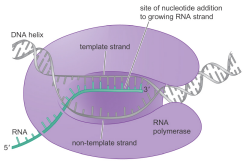


- tsx initiation control by transcription factors (TF)
- tsl initiation control by sRNA and RNA-binding proteins
- tsx termination control by sRNA and anti-terminators
- control of mRNA and protein degradation
- control of enzyme activity by metabolites

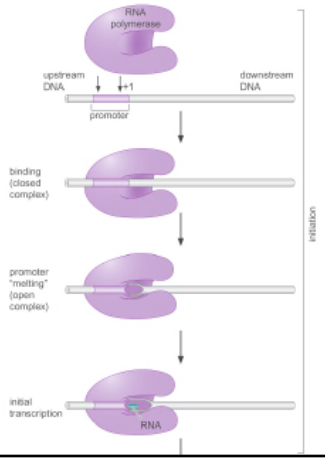
coupled to environmental signals; coord growth program

20

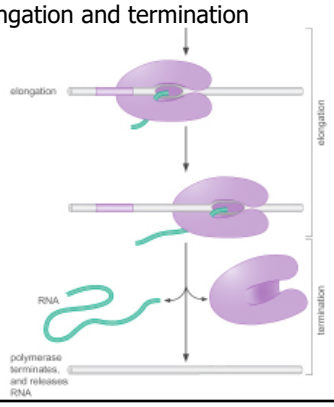
molecular biology of transcription (RNA synthesis)



transcriptional initiation



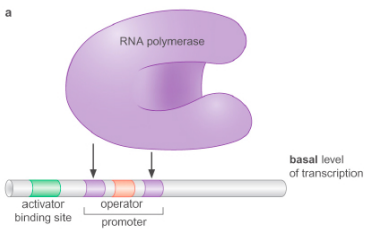
elongation and termination

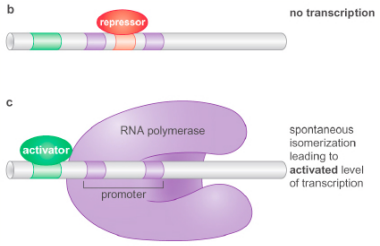


22

transcriptional initiation control

- modulation of RNAp-promoter affinity
  via activators and repressors





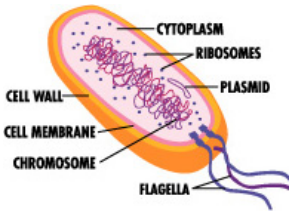
➔ net result: rate of transcriptional initiation dependent on cellular conc of activators and repressors
controlled by metabolites and signaling molecules

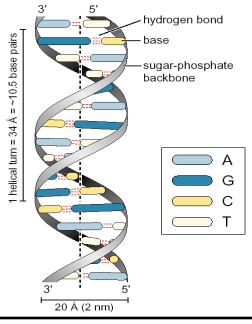
23

Quantitiative aspect

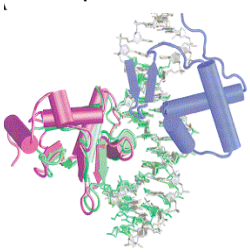
Dimension

- DNA: 2 nm x 2 nm x 3.4 nm/turn
- small proteins: (few nm)^3 or ~10nt
- protein complexes, (10-20 nm)^3 or 30 ~ 60nt
- cell size: 1 μm^2 x 3 μm (condition dependent)
- concentration: 1 molecule/cell ~ 1nM (=0.6/ μm^3)
- intracellular diffusitivity of protein: ~10 μm^2/sec (can usually be regarded as 'well mixed')

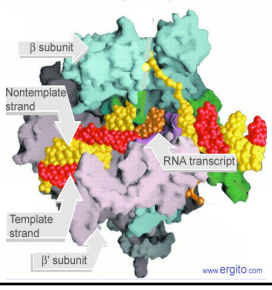




transcription factors



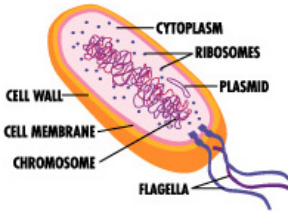
RNA polymerase



24

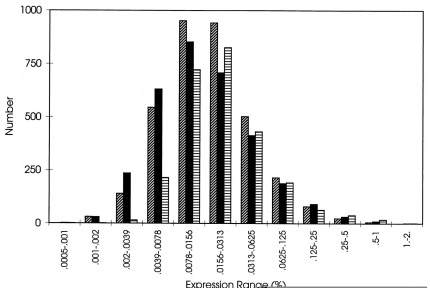
abundance

- ribosomes: ~ 20,000 (52 proteins each)
- RNAp ~ 1,000 (a portion available)
- proteins: 2x10^6 (TF: 10 ~ 1,000 / type)
- mRNA: 3% of RNA
  0.1 ~ 100 copies/cell;
  peaked at 2 ~ 3 copies / cell



25% of bacterial dry mass is concerned with gene expression

Component	Dry Cell Mass (%)	Molecules /cell	Different types	Copies of each type
Wall	10	1	1	1
Membrane	10	2	2	1
DNA	1.5	1	1	1
mRNA	1	1,500	600	2-3
tRNA	3	200,000	60	>3,000
rRNA	16	38,000	2	19,000
Ribosomal proteins	9	10^6	52	19,000
Soluble proteins	46	2.0 x 10^6	1,850	>1,000
Small molecules	3	7.5 x 10^6	800	

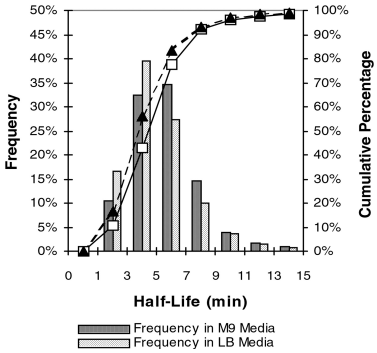


Warning: many of these numbers are dependent on growth condition!

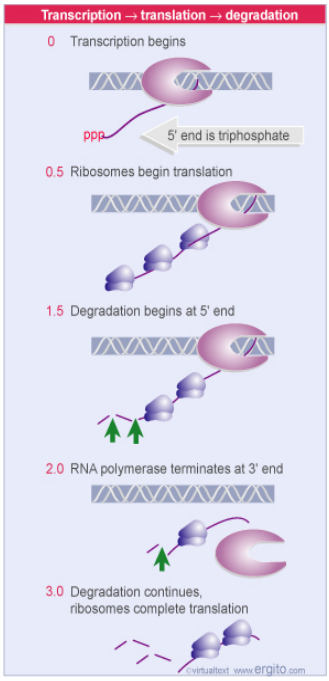
25

rates

- transcription: elongation ~50 nt/s
- translation: ~ 16 aa/s
- transcription-translation coupling:
  infrequently translated mRNA cleaved
- mRNA half-life: < 5 min

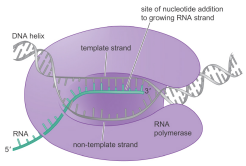


- protein half-life:
  from cell-doubling time (passive decay)
  down to a few min (active proteolysis)

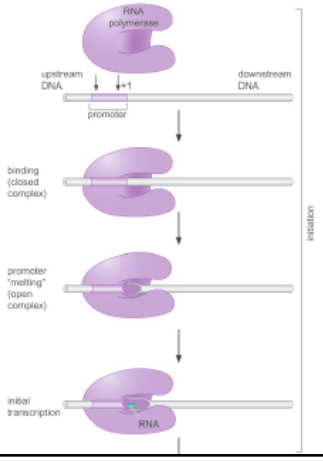


26

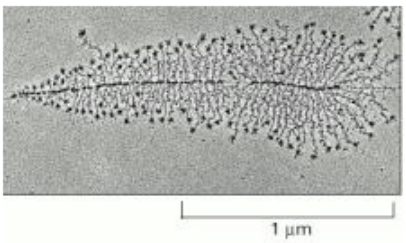
RNA synthesis:



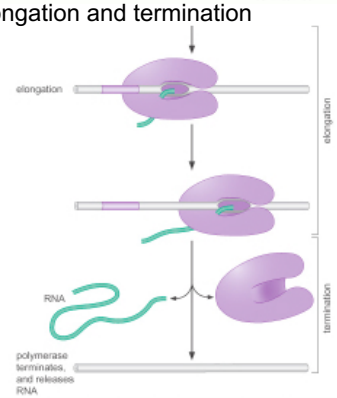
transcriptional initiation



heavily transcribed genes
coding ribosomal RNA



elongation and termination



27

3. **Transcription and translation rates:** In parts (a)-(f), deduce the transcription and translation rates for typical (non-ribosomal) genes in the exponential growth phase. Take the doubling time to be 45 min, and use the average copy number of a gene to be 2. Please report your numerical answers as well as the mathematical expressions.

Table 1: Pertinent values

$\varepsilon_{\mathrm{tsx}}$	maximal transcription speed	$\sim$48 nt/sec
$\varepsilon_{\mathrm{tsl}}$	maximal translation speed	$\sim$16 aa/sec
ℓ_{RNAp}	physical size of RNAp	$\sim$55 nt
ℓ_{Rb}	physical size of ribosome	$\sim$35 nt
τ_{mRNA}	half life of mRNA	$\sim$2 min
δ_{mRNA}	mRNA degradation rate	$\ln 2/\tau_{\mathrm{mRNA}}$
T	cell doubling time	45 min
μ	dilution rate due to growth	$\sim\ln 2/T$
g	average gene copy number	2
$G_{E.coli}$	total genes in *E. coli*	4500
l_{gene}	average gene length	300 aa

(a) Given the maximal speed of transcriptional elongation by RNAp ($\sim$48 nt/sec) and the physical size of RNAp (covers $\sim$55 nt), find the maximal rate at which full-length mRNA transcripts can be synthesized.

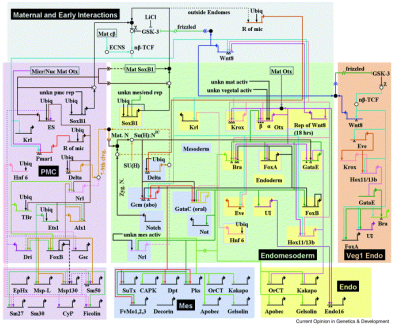
(b) Given the half-life of 2 min for a typical transcript, what is the maximal copy number for each type of such transcripts in the steady state (of balanced exponential growth)?

28

"Systems Biology"

Scope and focus:

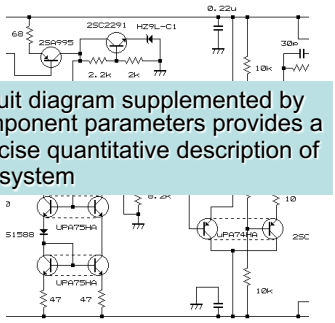
- biological systems whose functions are derived from the interaction of many sub-components
- ex: from macromolecluar assemblies to ecological communities
- central focus: subcellular and cellular processes, e.g., genetic circuits, protein interaction networks



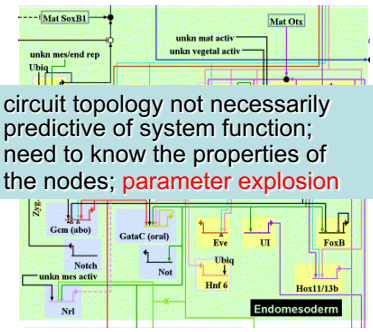
- long-term goals:
  - –mapping out the complete wiring diagram of the cell
  - **–quantitative, predictive computational model of the cell**

29

Circuit diagram as system-level descriptor ?



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

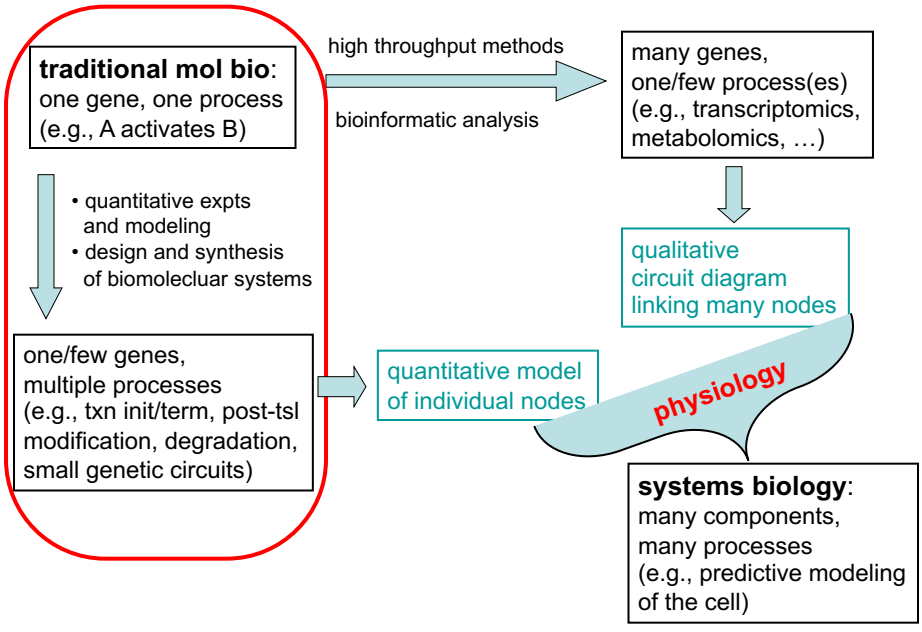


circuit topology not necessarily predictive of system function; need to know the properties of the nodes; parameter explosion

	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

30

Experimental & Computational Approaches



31

scope of this course

- focus on simple systems (bacterial gene regulation)
- role of theory, modeling, and computation
- coarse-grained description at multiple scales (telescoped description)
- **quantitative connections between molecular mechanisms and physiological (functional or behavioral) characteristics**
- **power of functional and mechanistic constraints**

❖ course content

- molecular interactions: ligand-protein, protein-DNA, and protein-protein
- transcriptional initiation control: activation, repression, and combinatorial
- post-transcriptional control: attenuation, termination, degradation
- modeling genetic circuits: bistability and oscillation
- stochastic gene expression and phenotype
- growth physiology and metabolic control

32