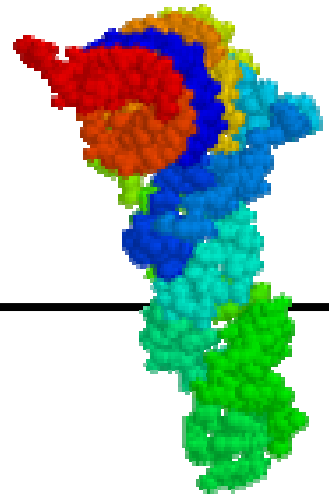
Genomics, Computing, Economics & Society



10 AM Thu 27-Oct 2005
week 6 of 14

MIT-OCW Health Sciences & Technology 508/510

Harvard Biophysics 101

Economics, Public Policy, Business, Health Policy

Class outline

Photo removed due to copyright reasons.

(1) Topic priorities for homework since last class

(2) Quantitative exercises: psycho-statistics, combinatorials, random/compression, exponential/logistic, bits, association & multi-hypotheses, **linear programming optimization**

(3) Project level presentation & discussion

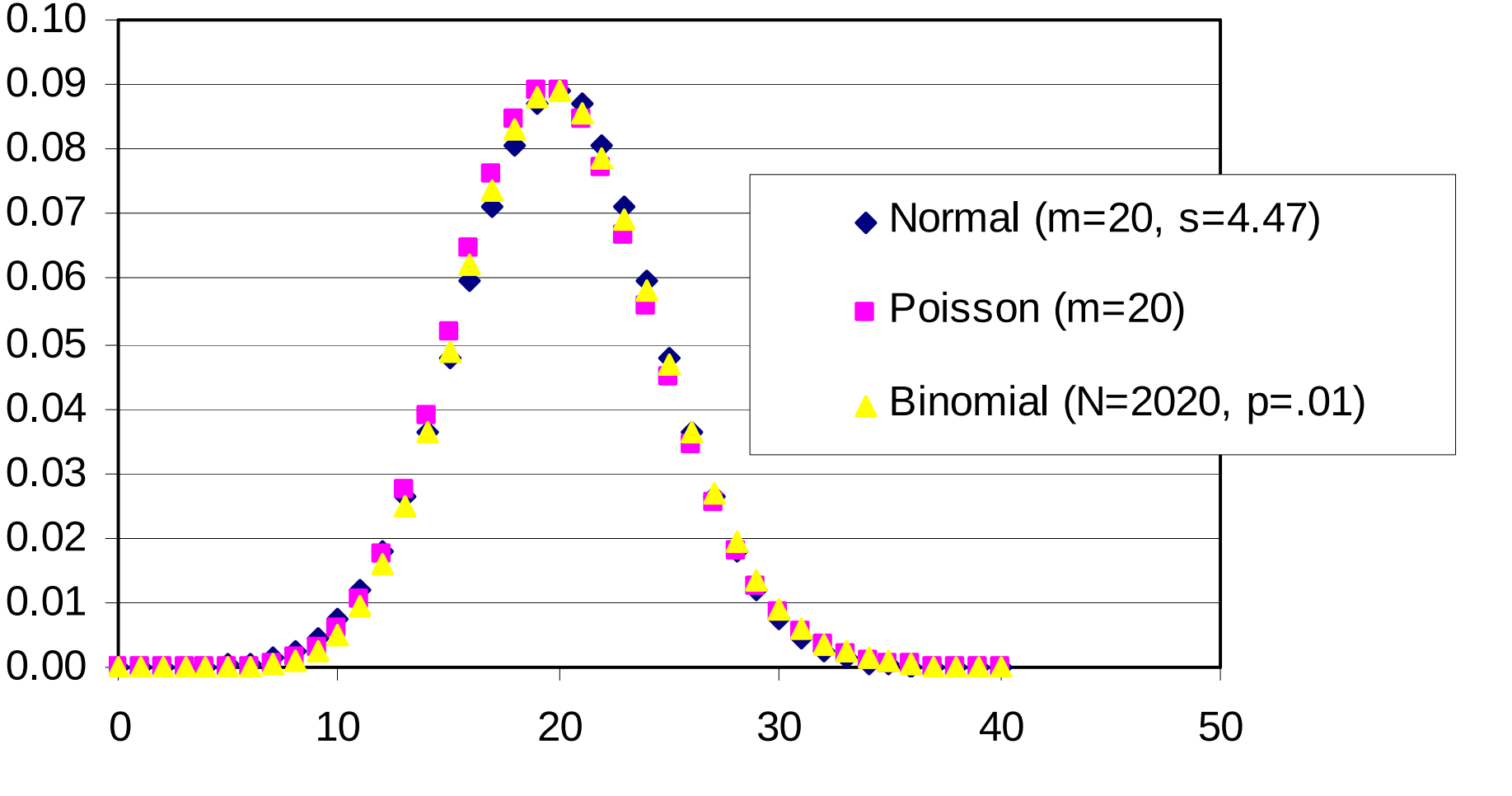
(4) Sub-project reports & discussion:

Personalized Medicine & Energy Metabolism

(5) Discuss communication/presentation tools

(6) Topic priorities for homework for next class

Binomial, Poisson, Normal



Binomial frequency distribution as a function of $X \in \{\text{int } 0 \ldots n\}$

p and q $\quad 0 \le p \le q \le 1 \quad$ $q = 1 - p$ $\quad$ two types of object or event.

Factorials $\quad 0! = 1 \quad n! = n(n-1)!$

Combinatorics (C= # subsets of size X are possible from a set of total size of n)

$$\frac{n!}{X!(n-X)!} = C(n,X)$$

$$B(X) = C(n, X)\, p^X\, q^{n-X} \qquad \mu = np \qquad \sigma^2 = npq$$

$$(p+q)^n = \sum B(X) = 1$$

$B(X: 350, n: 700, p: 0.1) = 1.53148 \times 10^{-157}$

=PDF[BinomialDistribution[700, 0.1], 350] Mathematica

~= 0.00 =BINOMDIST(350,700,0.1,0) Excel

Poisson frequency distribution as a function of X $\in$ {int 0 ...∞}

$P(X) = P(X\text{-}1)\ \mu/X \quad = \quad \mu^x e^{-\mu}/ X! \quad \sigma^2 = \mu$

n large & p small $\rightarrow$ $P(X) \cong B(X) \quad \mu = np$

For example, estimating the expected number of positives

in a given sized library of cDNAs, genomic clones,

combinatorial chemistry, etc. X= # of hits.

Zero hit term = $e^{-\mu}$

Normal frequency distribution as a function of $X \in \{-\infty ... \infty\}$

$Z = (X-\mu)/\sigma$

Normalized (standardized) variables

$N(X) = \exp(-Z^2/2) / (2\pi\sigma)^{1/2}$
probability density function

npq large $\rightarrow$ $N(X) \cong B(X)$

Mean, variance, & linear correlation coefficient

Expectation E (rth moment) of random variables X for any distribution f(X)

First moment= Mean μ ; variance σ^2 and standard deviation σ

$E(X^r) = \sum X^r f(X)$ $\mu = E(X)$ $\sigma^2 = E[(X-\mu)^2]$

Pearson correlation coefficient $C = \mathrm{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]/(\sigma_X \sigma_Y)$

Independent X,Y implies C = 0,
but C =0 does not imply independent X,Y. (e.g. $Y=X^2$)

P = TDIST(C*sqrt((N-2)/(1-C^2)) with dof= N-2 and two tails.

where N is the sample size.

www.stat.unipg.it/IASC/Misc-stat-soft.html

Under-Determined System

- All real metabolic systems fall into this category, **so far**.
- Systems are moved into the other categories by measurement of fluxes and additional assumptions.
- Infinite feasible flux distributions, however, they fall into a solution space defined by the **convex polyhedral cone.**
- The actual flux distribution is determined by the cell's regulatory mechanisms.
- It absence of kinetic information, we can estimate the metabolic flux distribution by postulating **objective functions(Z)** that underlie the cell’s behavior.
- Within this framework, one can address questions related to the capabilities of metabolic networks to perform functions while constrained by stoichiometry, limited thermodynamic information (reversibility), and physicochemical constraints (ie. uptake rates)

FBA - Linear Program

- For growth, define a growth flux where a linear combination of monomer (M) fluxes reflects the known ratios (d) of the monomers in the final cell polymers.

$$\sum_{allM} d_M \cdot M \xrightarrow{v_{growth}} biomass$$

- A linear programming finds a solution to the equations below, while minimizing an objective function (Z).

  Typically Z= v_{growth} (or production of a key compound).

- *i* reactions

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b}$$

$$v_i \geq 0$$

$$\alpha_i \leq v_i \leq \beta_i$$

$$v_i = X_i$$

Steady-state flux optima

Flux Balance Constraints:

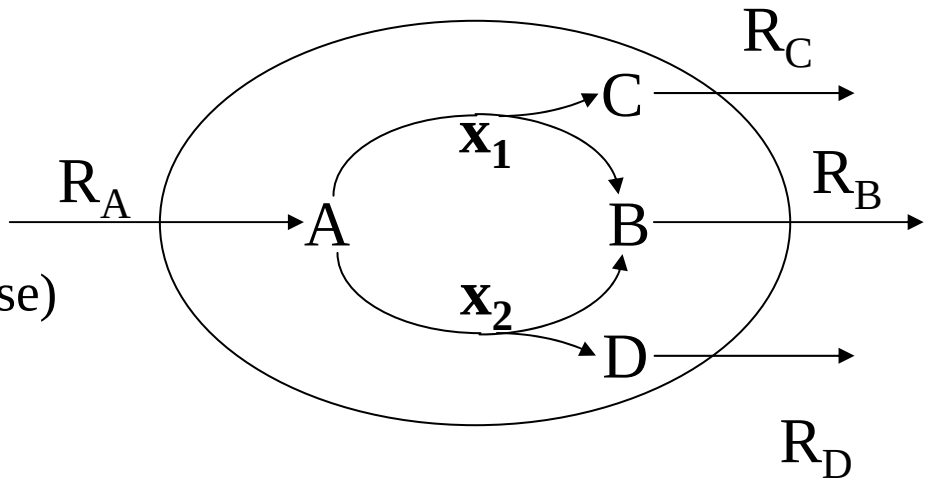
$R_A < 1$ molecule/sec (external)

$R_A = R_B$ (because no net increase)

$x_1 + x_2 < 1$ (mass conservation)

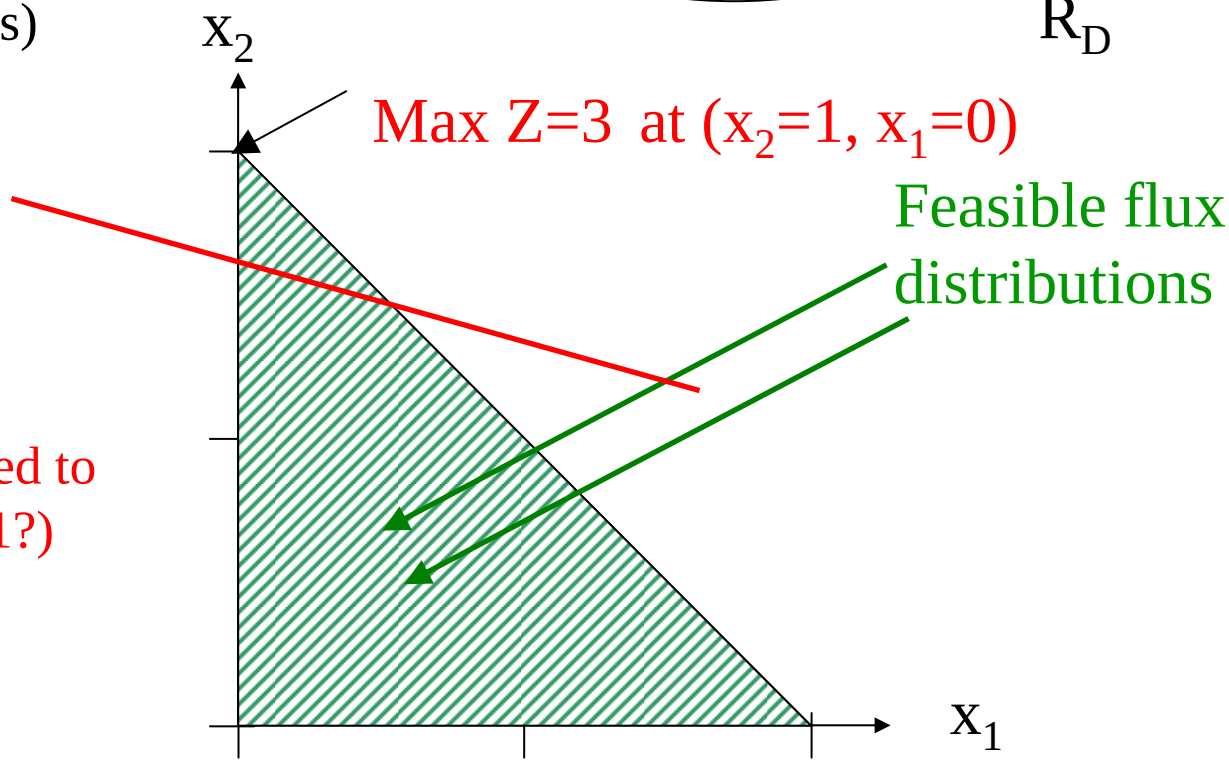
$x_1 > 0$ (positive rates)

$x_2 > 0$



$Z = 3R_D + R_C$

(But what if we really wanted to select for a fixed ratio of 3:1?)



Applicability of LP & FBA

- Stoichiometry is well-known
- Limited thermodynamic information is required
  - reversibility vs. irreversibility
- Experimental knowledge can be incorporated in to the problem formulation
- Linear optimization allows the identification of the reaction pathways used to fulfil the goals of the cell if it is operating in an optimal manner.
- The relative value of the metabolites can be determined
- Flux distribution for the production of a commercial metabolite can be identified. Genetic Engineering candidates

Precursors to cell growth

- How to define the growth function.
  - The biomass composition has been determined for several cells, *E. coli* and *B. subtilis*.
    - This can be included in a complete metabolic network
  - When only the catabolic network is modeled, the biomass composition can be described as the 12 biosynthetic precursors and the energy and redox cofactors

in silico cells

	E. coli	*H. influenzae*	*H. pylori*
Genes	**695**	**362**	**268**
Reactions	**720**	**488**	**444**
Metabolites	**436**	**343**	**340**
(of total genes	**4300**	**1700**	**1800)**

Edwards, et al 2002. Genome-scale metabolic model of Helicobacter pylori 26695. J Bacteriol. 184(16):4582-93.

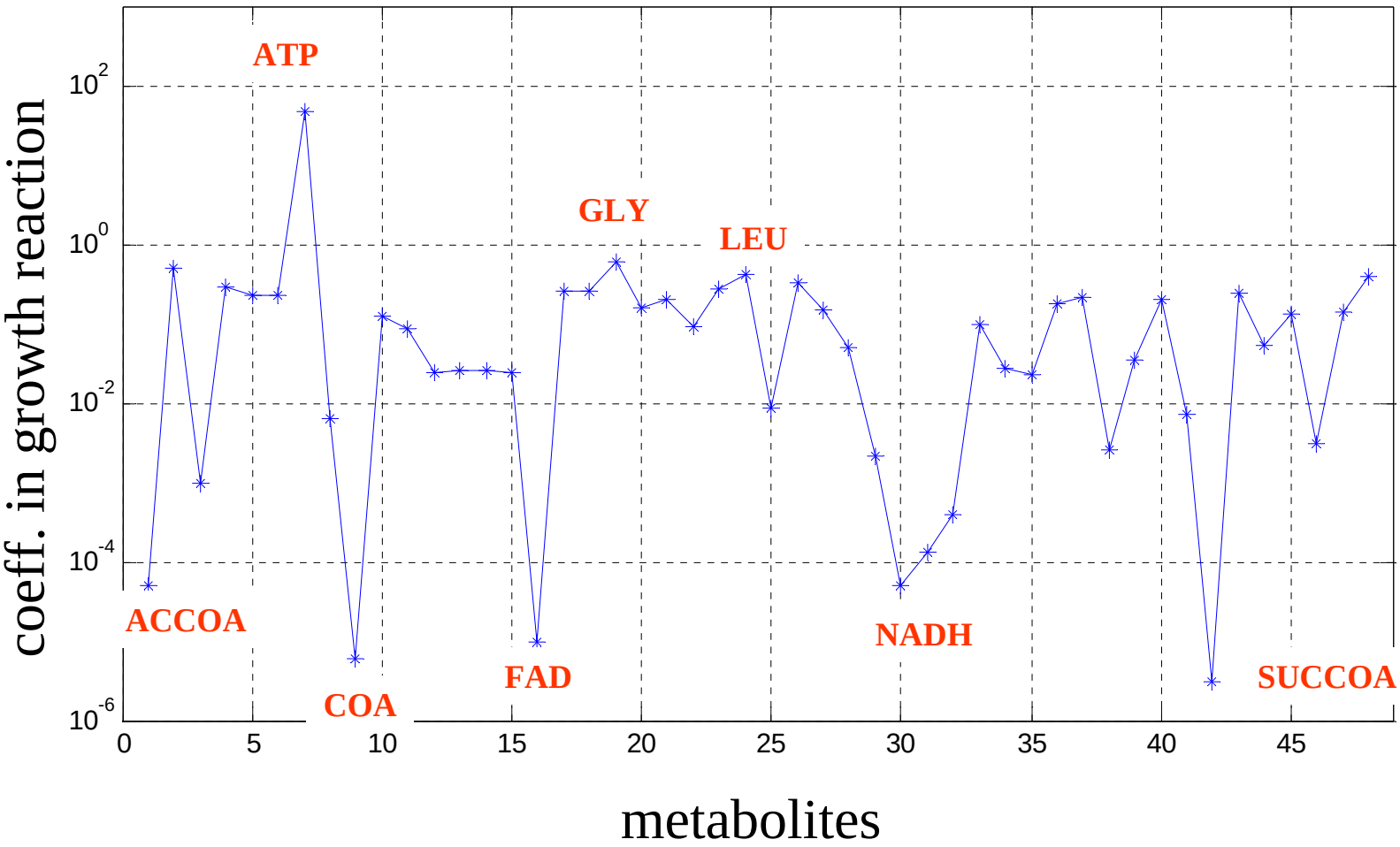
Segre, et al, 2002 Analysis of optimality in natural and perturbed metabolic networks. PNAS 99: 15112-7. (Minimization Of Metabolic Adjustment) http://arep.med.harvard.edu/moma/

Where do the Stochiometric matrices (& kinetic parameters) come from?

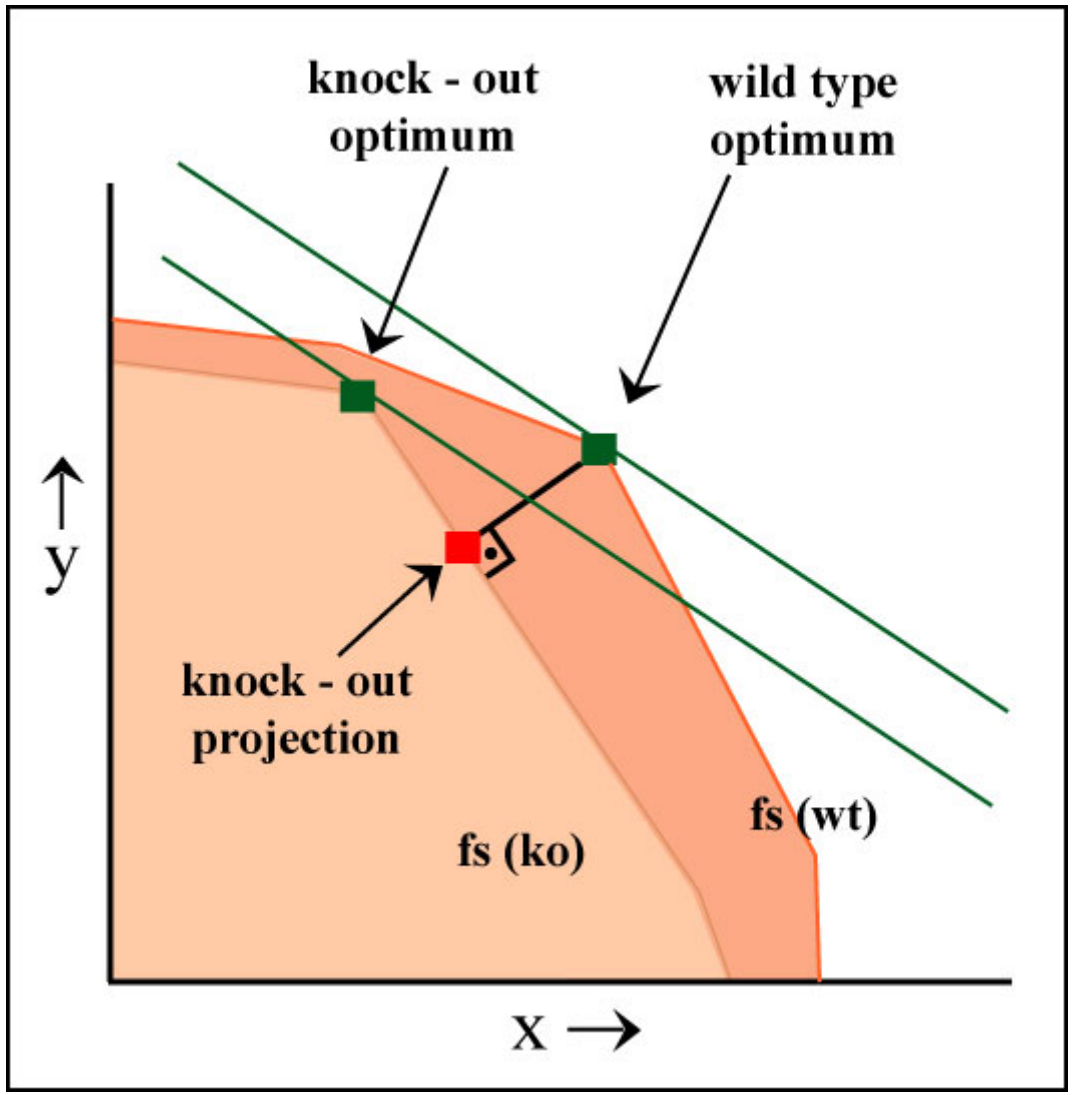
Figures removed due to copyright reasons.

EMP RBC, E.coli
KEGG, Ecocyc

Biomass Composition



Flux ratios at each branch point yields optimal polymer composition for replication



x,y are two of the 100s of flux dimensions

Figure by MIT OCW.

Minimization of Metabolic Adjustment (MoMA)

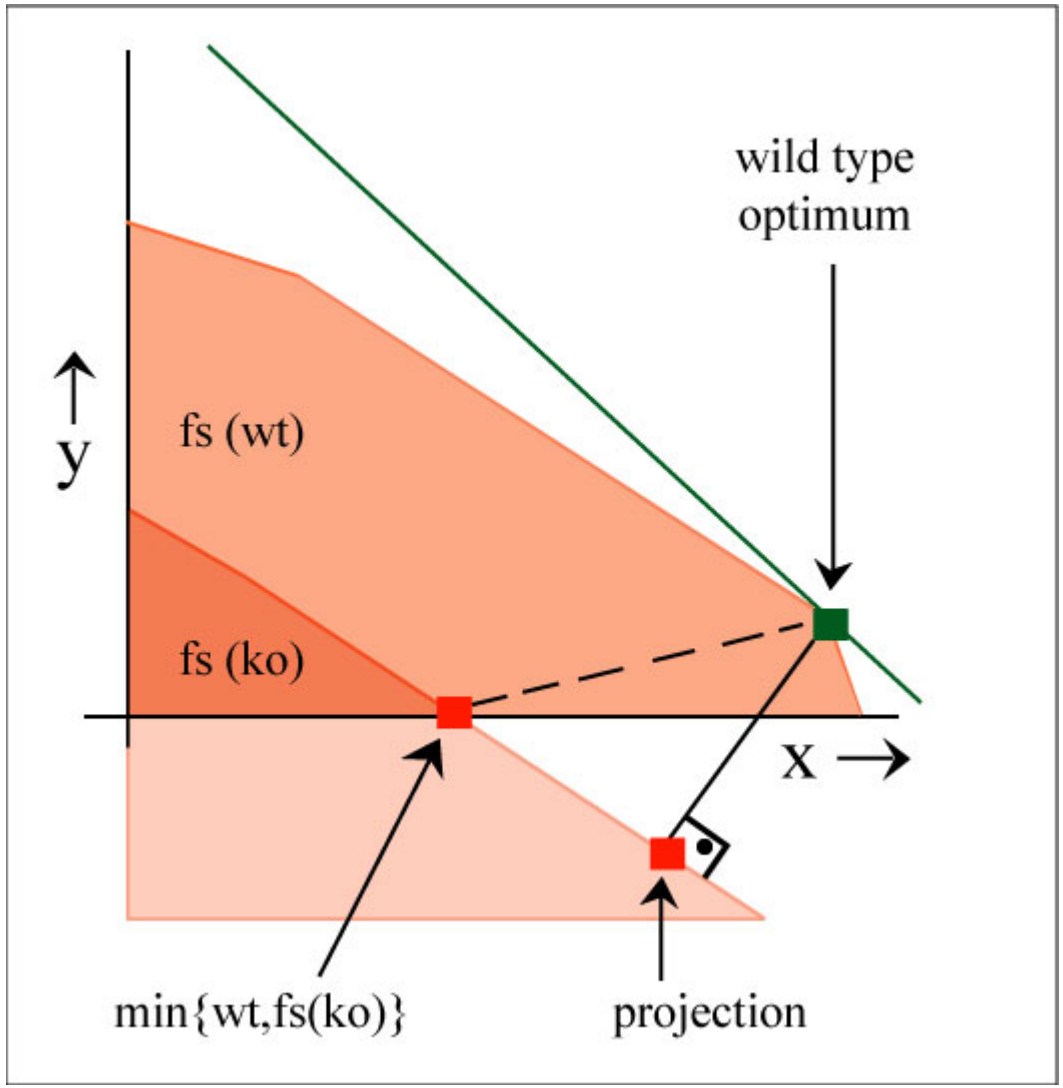


Figure by MIT OCW.

Figure removed due to copyright reasons.

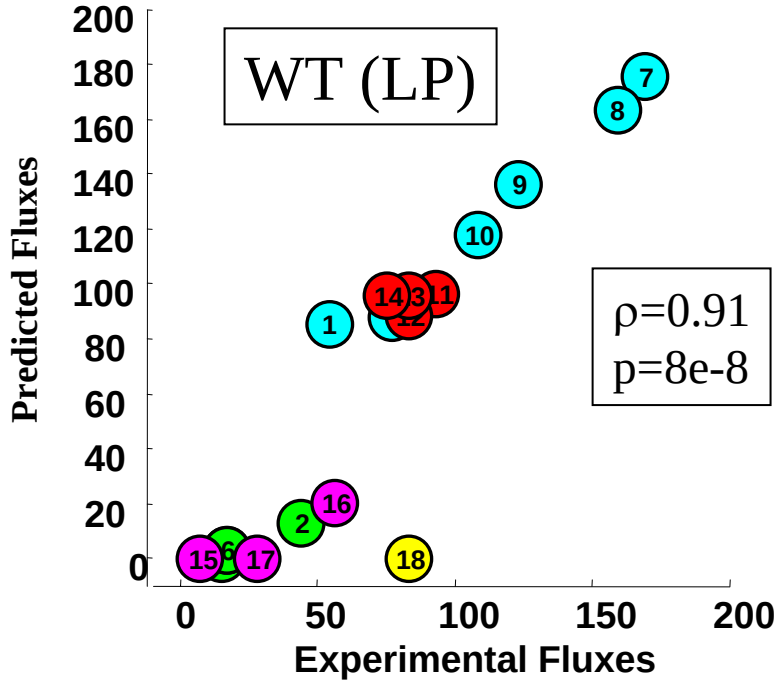
Figure removed due to copyright reasons.

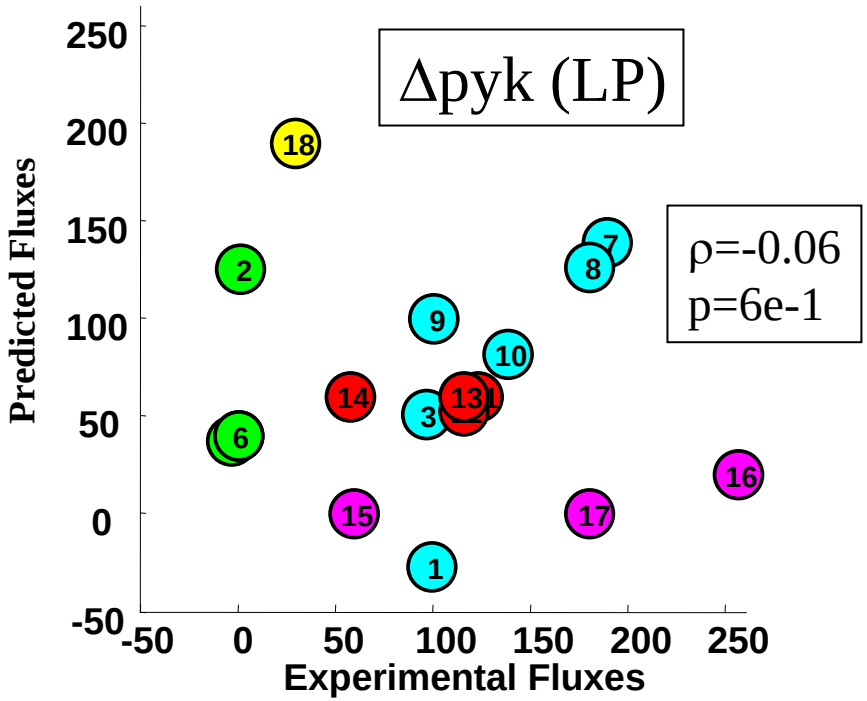
Flux

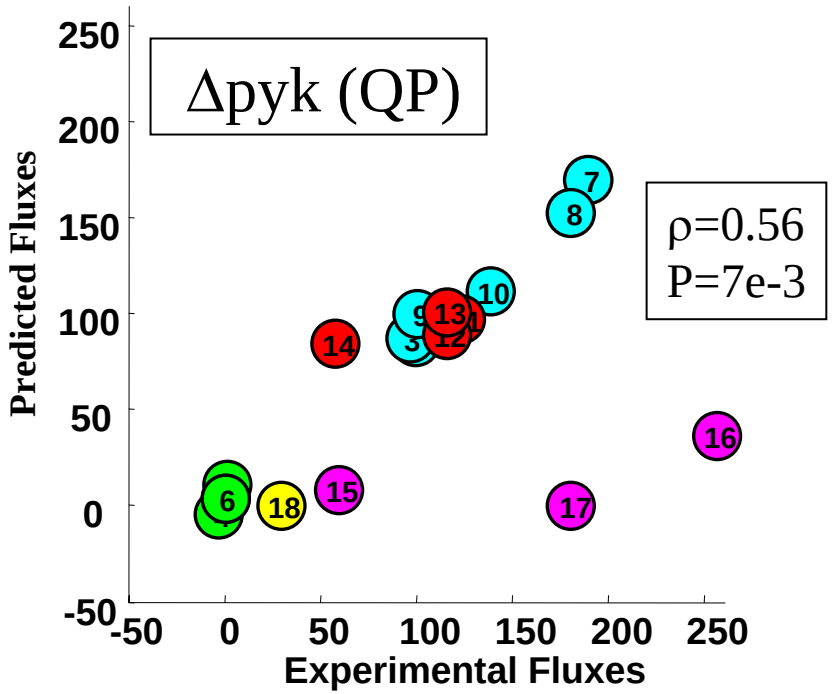
Data

C009-limited

Figure removed due to copyright reasons.

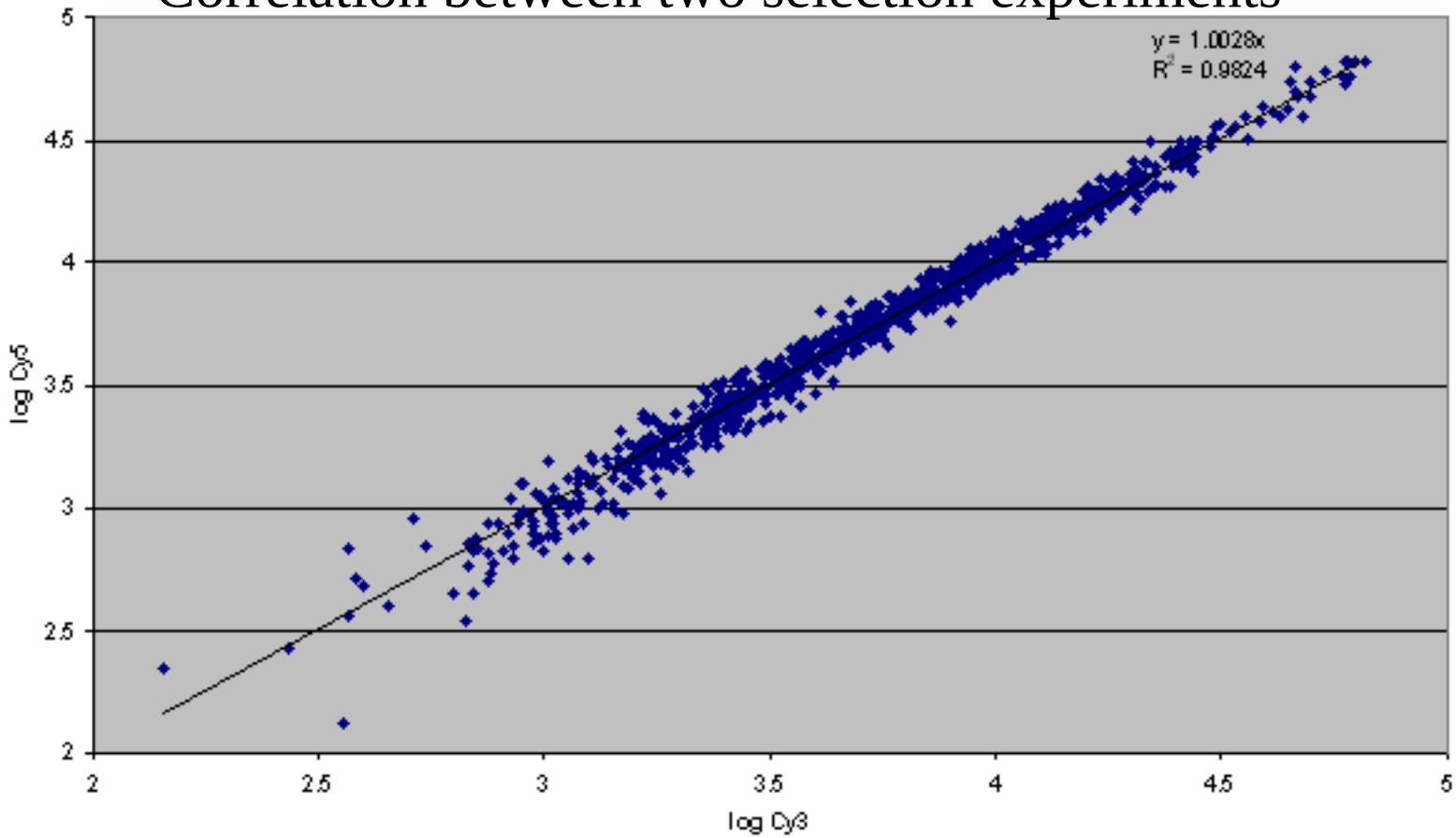






Competitive growth data: reproducibility

Correlation between two selection experiments



Badarinarayana, et al. Nature Biotech.19: 1060

Competitive growth data

On minimal media

		negative selection	small effect	X^2 p-values
LP	**Essential**	142	80	62	
	Reduced growth	46	24	22	
	Non essential	299	119	180	**$4x10^{-3}$**
QP MOMA	**Essential**	162	96	66	
	Reduced growth	44	19	25	
	Non essential	281	108	173	**$1x10^{-5}$**

Position effects

Novel redundancies

Hypothesis: next optima are achieved by regulation of activities.

Non-optimal evolves to optimal

Figures removed
due to copyright
reasons.

Ibarra et al. Nature. 2002 Nov 14;420(6912):186-9. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth.

Non-linear constraints

Desai RP, Nielsen LK, Papoutsakis ET. Stoichiometric modeling of Clostridium acetobutylicum fermentations with non-linear constraints. J Biotechnol. 1999 May 28;71(1-3):191-205.

Class outline

(1) Topic priorities for homework since last class

(2) Quantitative exercise

(3) Project level presentation & discussion

(4) Sub-project reports & discussion

(5) Discuss communication/presentation tools

(6) Topic priorities, homework for next class