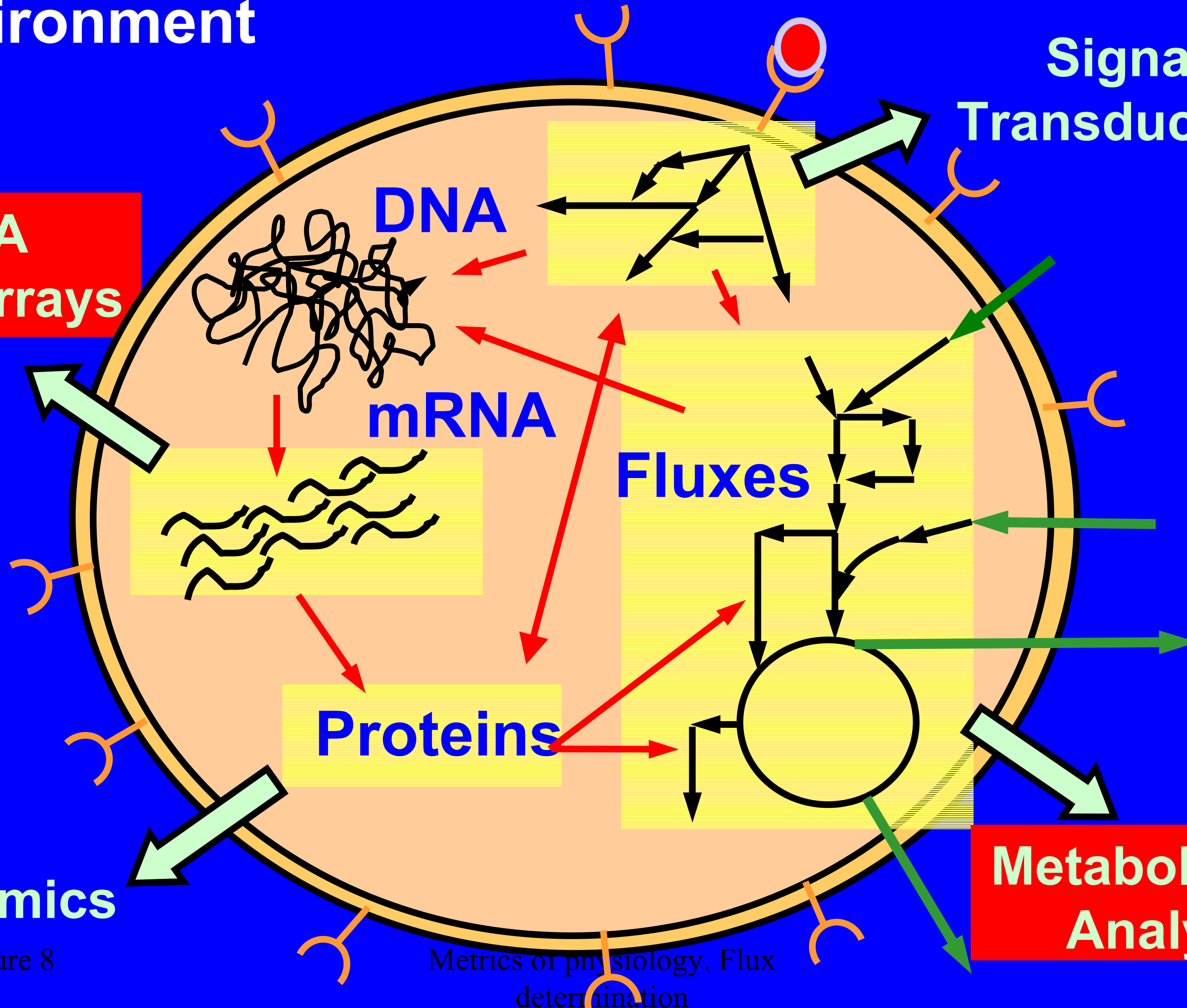
10.555 Bioinformatics

Lecture 8

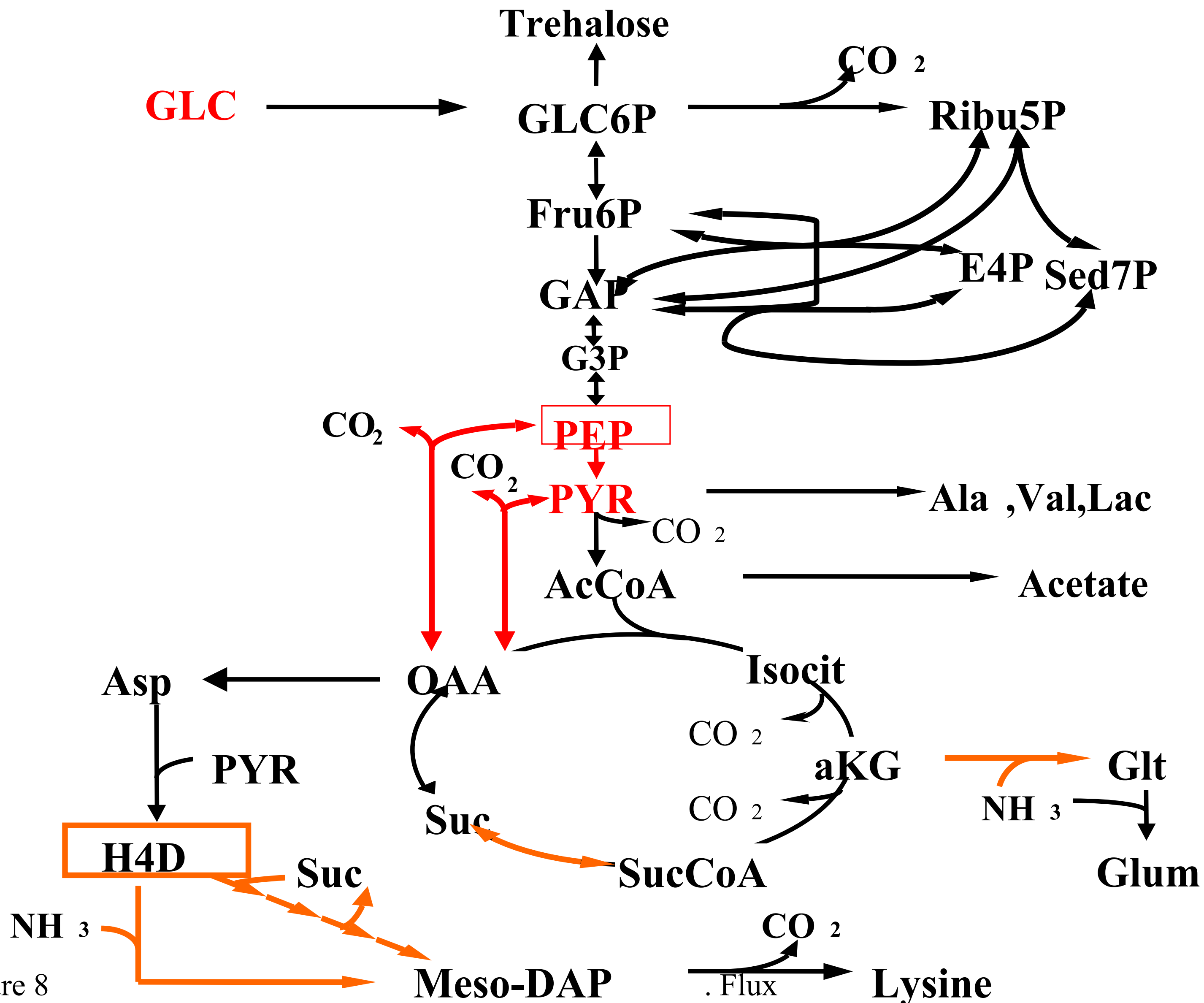
Developing metrics of cell physiology. Flux determination

Gregory Stephanopoulos
MIT

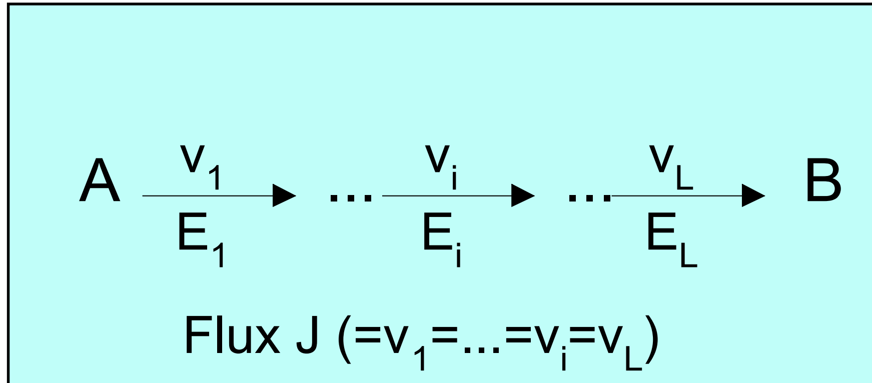
Probing cellular function

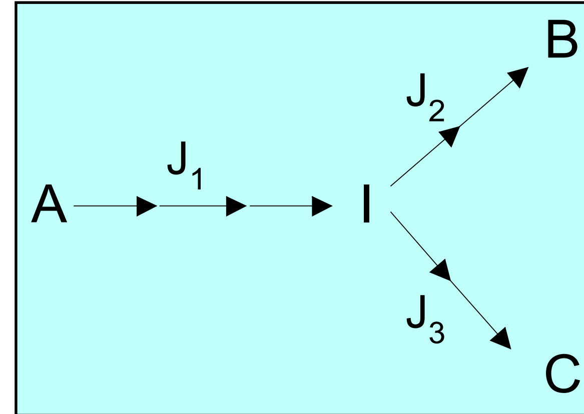


Case study: Aminoacid biosynthesis in *Corynebacterium glutamicum*



Definitions





- Metabolic Pathway: A sequence of *feasible* and *observable* bioreaction steps connecting sets of input and output metabolites.
- Metabolic flux: The rate of material processing through a metabolic pathway

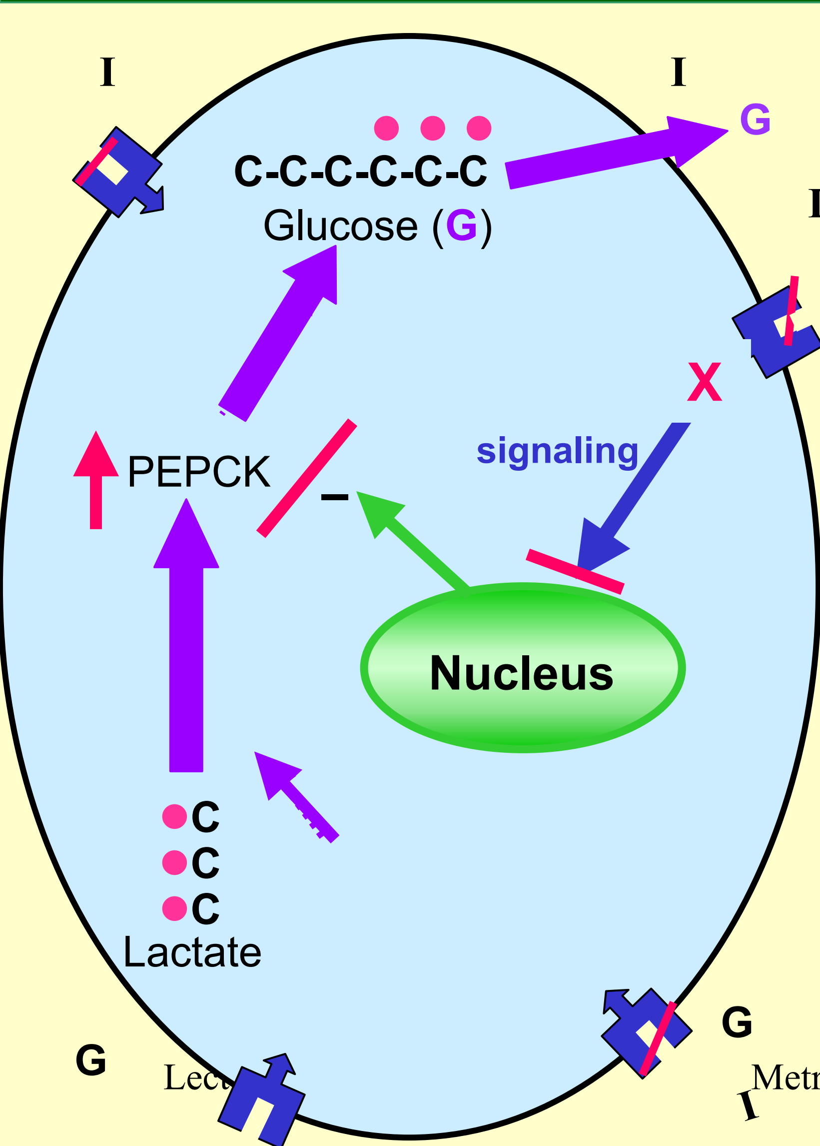
Flux as fundamental determinant of cell physiology

- Along with metabolite concentrations, fluxes define a minimum set of information needed to describe cell physiology
- Fluxes determine the degree of engagement of various enzymes in a conversion pathway
- Fluxes elucidate pathway flux control
- Useful for:
  - Calculating theoretical yields
  - Determining non-measured metabolite rates
  - Observing pathway function *in vivo*

HEPATIC INSULIN RESISTANCE, POOR RESPONSE TO INCREASED PLASMA GLUCOSE

HPERGLYCEMIA VIA INCREASED PEPCK GENE EXPRESSION

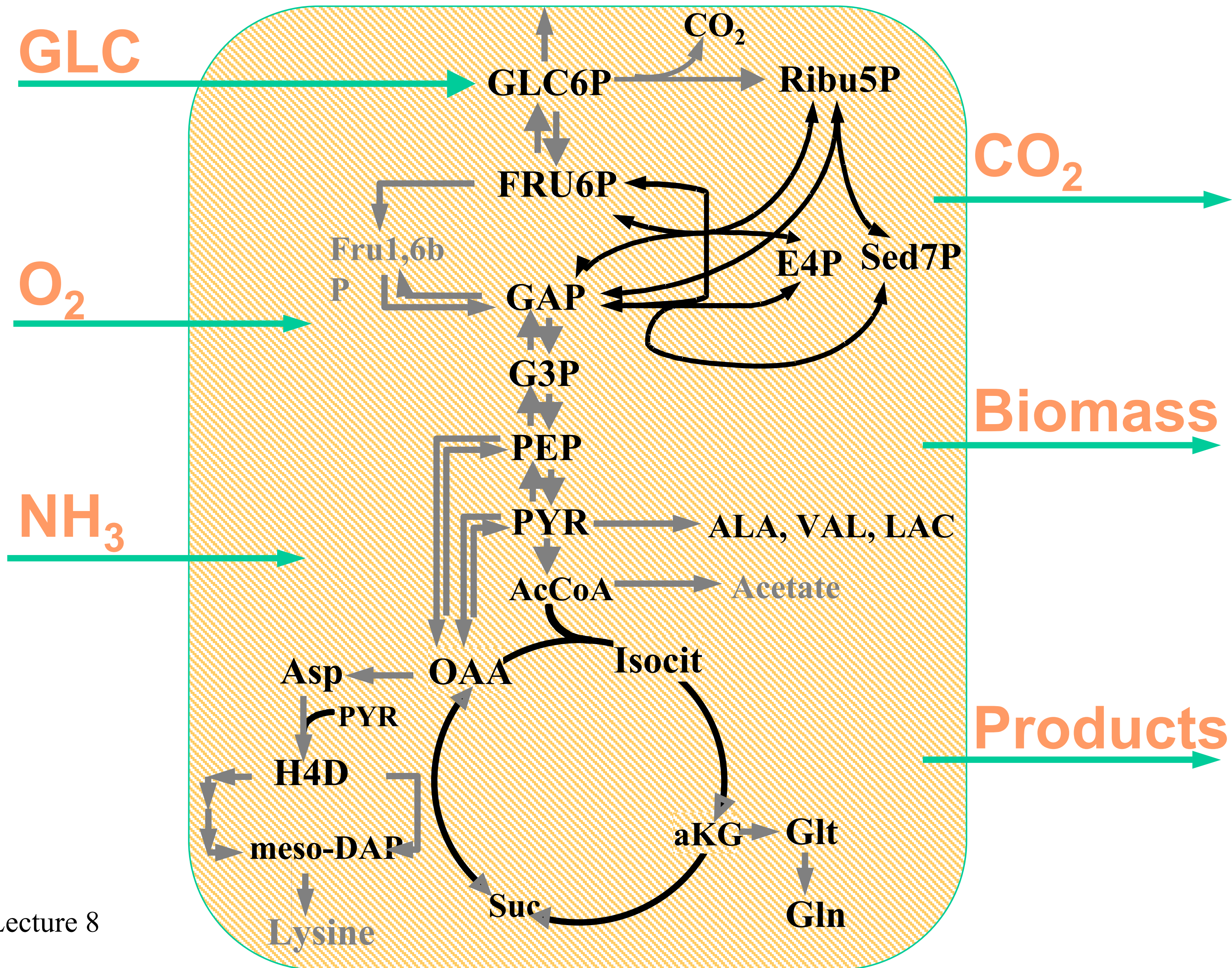
HPERGLYCEMIA VIA DECREASED ACETYL COA CARBOXYLASE & MALONYL COA



In-vivo flux determination

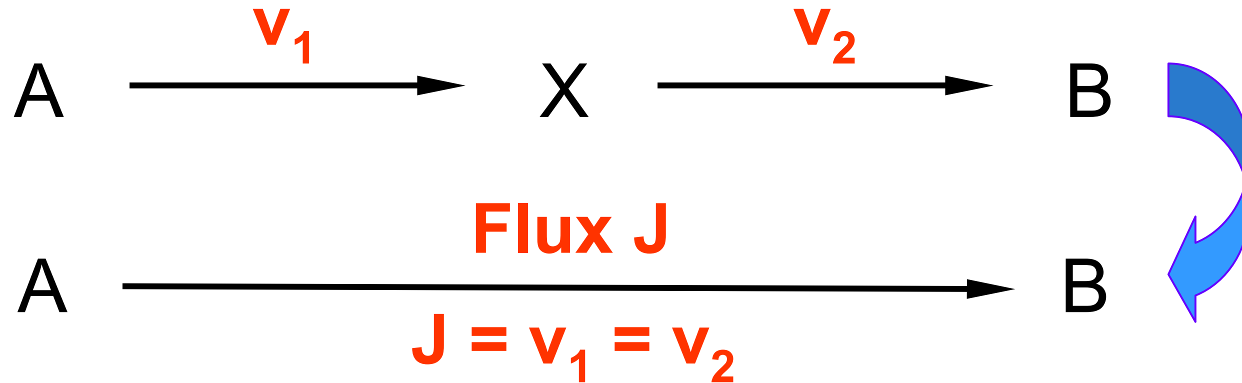
- **Prior research:** Based on isolating pathway of interest from rest of metabolism
  - Limited success. Results of questionable value
- **This research:** Has yielded validated estimates of *in-vivo* metabolic fluxes. Approach is based on:
  - Material balances
  - Measurement of ^{13}C label enrichment in selected (secreted) metabolites
  - Careful analysis of NMR spectra *fine structure*
  - Isotopomer MW distribution measured by GC-MS

How do we measure fluxes?

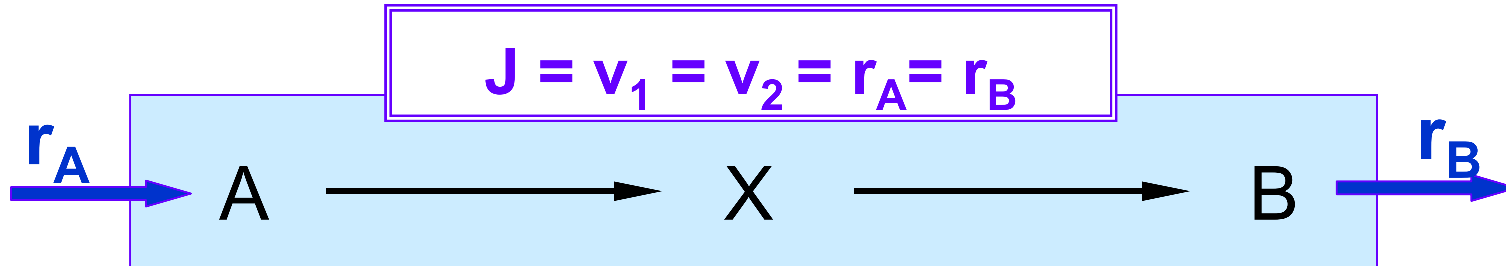


Metabolite Balancing: Linear Systems

- Include only branch points in the network (PSSH)



- Degrees of freedom



Measurement of r_A **or** r_B suffices to determine flux J

Measurement of <u>both</u> r_A and r_B yields **over-determined** system

Metabolite Balancing: Branched Pathways

r_A, v_2, v_3, r_E, v_4, r_G, v_1, v_5, v_6, r_F, r_C

A, D, E, G, B, C, F

7 balances

A : $v_1 + v_2 = r_A$

B : $v_1 - v_5 = 0$

C : $v_5 - v_6 = r_C$

D : $v_2 - v_3 - v_4 = 0$

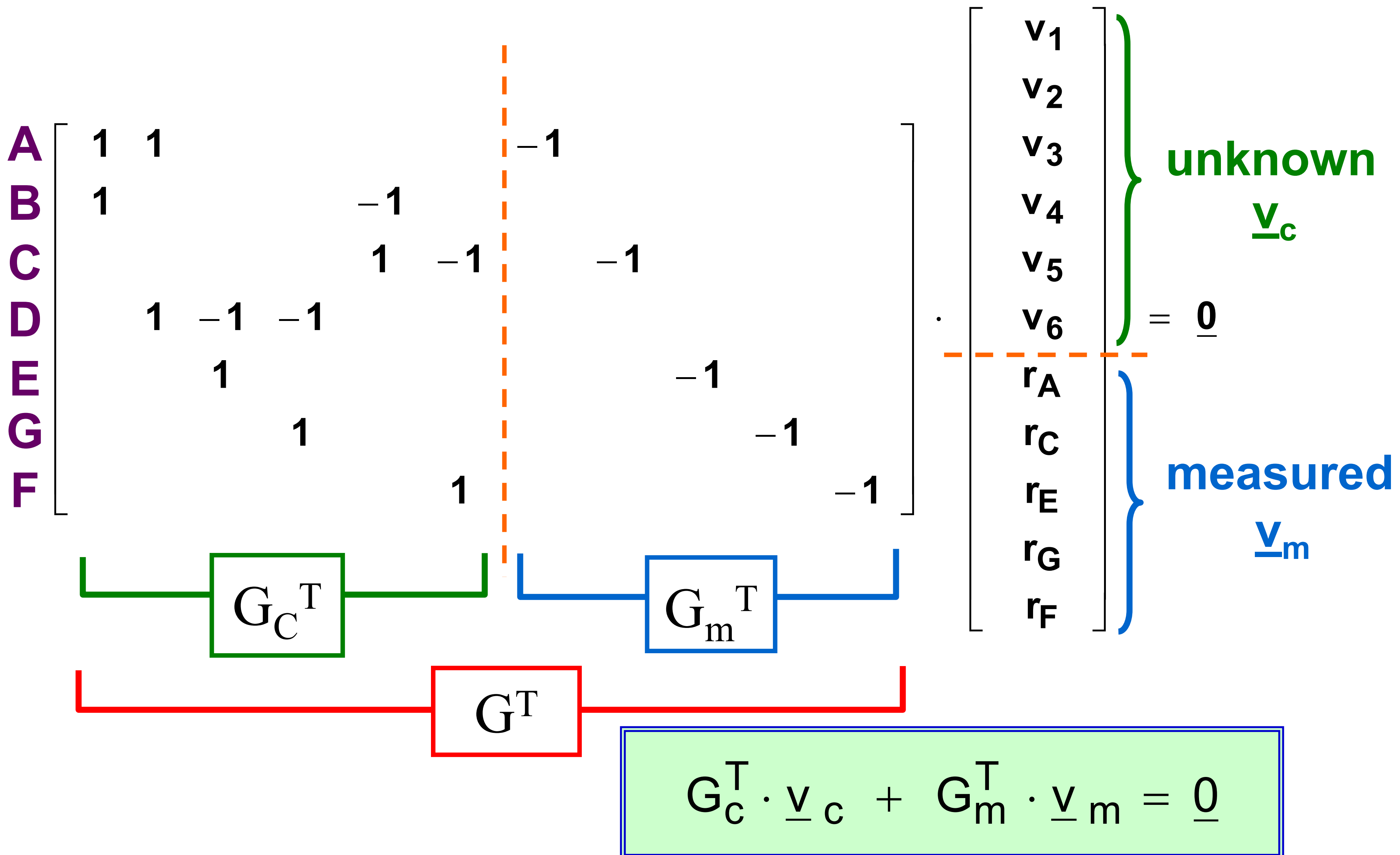
E : $v_3 = r_E$

G : $v_4 = r_G$

F : $v_6 = r_F$

6 unknown fluxes v_1 - v_6

Metabolite Balancing (cont'd)



$$G_c^T \cdot \underline{v}_c + G_m^T \cdot \underline{v}_m = \underline{0}$$

Metabolite Balancing (cont'd)

$$\min \left(\bar{\bar{\underline{v}}}_m - \underline{v}_m \right)^T \cdot \underline{\underline{Cov}}_{v_m}^{-1} \cdot \left(\bar{\bar{\underline{v}}}_m - \underline{v}_m \right)$$

subject to:

$$G_c^T \cdot \underline{v}_c + G_m^T \cdot \underline{v}_m = \underline{0}$$

v_m the 'true' (i.e adjusted to the measurements) value of the vector of measured fluxes

$\bar{\bar{v}}_m$ the measured value of the vector of measured fluxes

$\underline{\underline{Cov}}_{v_m}$ covariance matrix of the measurements (diagonal in the case of independent measurements)

from : measurement distribution / equipment specs / experience

Metabolite Balancing (cont'd)

Ensure that G_C^T is **fully-ranked**

- G_C^T is **singular** (i.e. not fully-ranked), when:

  - \# balances < # of unknown fluxes

  - \# balances ≥ # of unknowns, but

    linear dependencies are present in G_C^T

  because...

- Parallel pathways are present in the network (i.e. linearly dependent reactions called also *structurally singular groups*)

  Examples: anaplerotic pathways, nitrogen assimilation, trans-hydro-dehydrogenase, isoenzymes

- Redundant measurements are present (see example 8.5)

Flux determination by metabolite balancing

- Based on the balance for the conc'n of each metabolite **X:**

$$\mathbf{dX/dt = r_{met} - \mu X_{met} = 0}$$

  - The term μX_{met} (dilution effect by growth) is small relatively to turnover rates (small pools)

  - dX/dt = 0 by Pseudo Steady State Hypothesis

$$\mathbf{r_{met} = G^T v = 0}$$

  (G is rxm matrix of stoichiometry)

- Partition $\mathbf{G^T}$ such that $0 = G^T v = G^T{}_m v_m + G^T{}_c v_c$

$$\mathbf{v_c = -(G^T{}_c)^{-1} G^T{}_m v_m}$$

  Ensure that G_c is of full rank (i.e., do not include linearly dependent reactions or redundant measurements)

Flux determination by metabolite balancing

- Redundancy is important. See notes for general procedure to:

  - Test consistency

  - Reconcile measurements

  - Determine new estimates that satisfy all balances (including the redundant)

Metabolite Balancing (cont'd)

- Degree of observability (D.O) = # of fluxes determinable (observable) from the measurements
- D.O. = rank of G_c^T
- Degree of redundancy (D.R) = # of redundant measurements = # measurements that can also be determined from the balances and the rest of the measurements
- D.R. = # measurements - D.O.

- If # unknown fluxes > D.O. → *underdetermined systems* :
  Linear Programming
- If D.R. > 0 → *overdetermined (<u>for the observable fluxes</u>) systems* :
  Solve by least squares - Gross error determination

Gross Error Determination (1)

Algorithm *(refer to slide 12 on Metabolite Balancing):*

If we have more equations than unknowns, use some of these equations to eliminate all unknown fluxes v_c to get:

$$\underline{\underline{A}} \cdot \underline{v}_m = \underline{0}$$

Note: # rows of A = # of redundant measurements

Due to the presence of measurement noise:

$$\underline{\underline{A}} \cdot \underline{\bar{\bar{v}}}_m = \underline{\varepsilon}$$

$\underline{\varepsilon}$: the residual of the constraints

Gross Error Determination (2)

$$\min \; (\bar{\bar{\underline{v}}}_m - \underline{v}_m)^T \cdot \underline{\underline{Cov}}_{v_m}^{-1} \cdot (\bar{\bar{\underline{v}}}_m - \underline{v}_m)$$

subject to:

$$\underline{\underline{A}} \cdot \underline{v}_m = \underline{0}$$

solution: $\underline{v}_m = (\underline{\underline{I}} - \underline{\underline{Cov}}_{v_m} \cdot \underline{\underline{A}}^T \cdot (\underline{\underline{A}} \cdot \underline{\underline{Cov}}_{v_m} \cdot \underline{\underline{A}}^T)^{-1} \cdot \underline{\underline{A}}) \cdot \bar{\bar{\underline{v}}}_m$

Consistency index *h*:

$$h = (\bar{\bar{\underline{v}}}_m - \underline{v}_m)^T \cdot \underline{\underline{Cov}}_{v_m}^{-1} \cdot (\bar{\bar{\underline{v}}}_m - \underline{v}_m)$$

$$h = (\bar{\bar{\underline{v}}}_m)^T \cdot \underline{\underline{A}}^T \cdot (\underline{\underline{A}} \cdot \underline{\underline{Cov}}_{v_m} \cdot \underline{\underline{A}}^T)^{-1} \cdot \underline{\underline{A}} \cdot \bar{\bar{\underline{v}}}_m$$

Gross Error Determination (3)

If :

$$h \geq \chi^2_{1-\theta}(m)$$

m = degrees of freedom = number of redundant measurements

then:

reject the hypothesis that measurement errors are insignificant with confidence level of 1-θ

Probability Points of the x^2 Distribution with m Degrees of Freedom

	θ (tail area probability)					
Degrees of freedom	0.50	0.25	0.10	0.05	0.025	0.01
m = 1	0.455	1.32	2.71	3.84	5.02	6.63
m = 2	1.39	2.77	4.61	5.99	7.38	9.21
m = 3	2.37	4.11	6.25	7.81	9.35	11.3
m = 4	3.36	5.39	7.78	9.49	11.1	13.3

Gross Error Determination - Example 1

Glucose-Limited Continuous Culture of S. *cerevisiae* CBS 426 with Ethanol and Glycerol Production under Anaerobic Conditions

h [$\chi^2(90\%, 2) = 4.6$]

$D\ (h^{-1})$	No deletion	After deletion of the measurement of: Glucose	Biomass	CO_2	Ethanol	Glycerol
0.012	2.02	0.04	0.00	1.90	1.37	0.28
0.014	3.32	1.27	1.64	0.63	3.31	2.56
0.014	**10.99**	9.53	10.19	**0.00**	9.05	10.99
0.017	0.27	0.13	0.16	0.04	0.26	0.24
0.027	1.52	n.a.	n.a.	n.a.	n.a.	n.a.
0.030	0.33	0.18	0.13	0.28	0.02	0.04
0.034	1.52	0.32	0.52	0.56	1.52	1.09
0.051	**7.17**	3.37	4.50	**0.74**	5.98	6.79
0.053	0.57	0.26	0.18	0.51	0.06	0.03
0.056	0.88	0.61	0.72	0.01	0.64	0.88
0.058	1.66	1.14	1.34	0.02	1.18	1.65
0.058	**10.20**	8.69	7.81	8.24	**0.00**	4.56
0.062	2.61	2.46	2.58	0.23	0.94	2.43
0.067	0.79	0.78	0.74	0.20	0.15	0.57

Gross Error Determination - Example 2

n-Hexadecane-limited batch growth of *Candida lipolytica* under aerobic conditions

	h $[\chi^2 (90\%, 2) = 4.6]$				
		After deletion of the measurement of			
Time (h)	No deletion	Substrate	O_2	Biomass	CO_2
0	**109.94**	76.04	36.74	46.30	**5.03**
2	**26.08**	24.29	4.06	16.00	**1.90**
4	**21.54**	17.24	7.07	7.69	**0.00**
8	**8.35**	5.63	4.37	1.12	**0.78**
10	4.47	1.72	3.65	0.01	1.68
12	0.92	0.83	0.28	0.40	0.00
14	1.05	1.04	0.03	0.89	0.18
16	3.36	3.19	0.00	3.20	1.23
18	2.68	2.66	0.08	2.36	0.65
19.5	2.60	2.60	0.18	2.14	0.46
21.5	1.97	1.69	0.71	0.85	0.00
23.5	3.68	3.67	0.29	3.07	0.74
25.5	**5.95**	5.21	2.06	3.41	**0.23**

Gross Error Determination - Example 3

Glycerol-limited continuous culture of *Aerobacter aerogens*

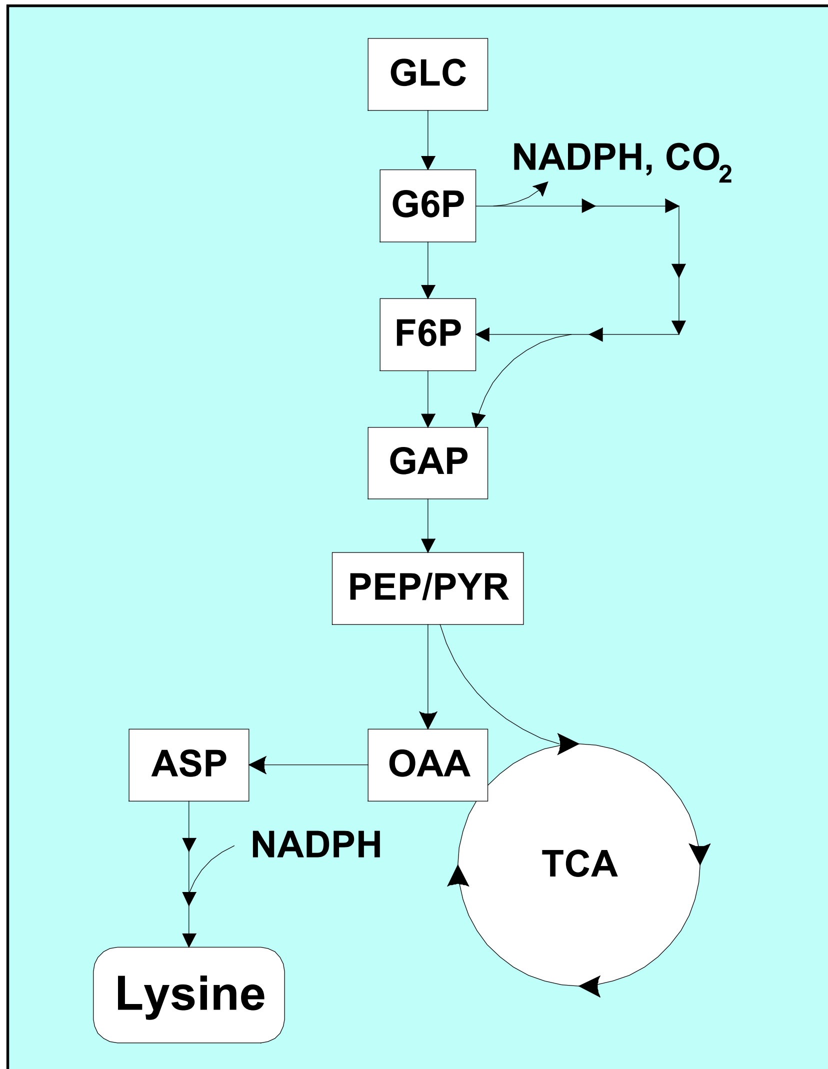
h [χ^2 (90%, 2) = 4.6]

$D(h^{-1})$	No deletion	After deletion of the measurement of: Glycerol	O_2	Biomass	CO_2
0.050	0.06	0.06	0.01	0.06	0.00
0.115	0.20	0.14	0.01	0.16	0.12
0.125	0.24	0.14	0.17	0.11	0.03
0.250	0.26	0.24	0.00	0.26	0.07
0.350	0.32	0.28	0.12	0.24	0.01
0.485	0.40	0.31	0.19	0.25	0.02
0.510	0.33	0.33	0.06	0.30	0.00
0.625	0.39	0.34	0.14	0.29	0.01
0.75	0.50	0.36	0.27	0.28	0.05
0.850	0.40	0.37	0.10	0.33	0.00
0.910	0.38	0.37	0.00	0.38	0.06
0.935	0.46	0.37	0.20	0.30	0.03
0.980	1.48	0.38	0.75	0.64	1.31
1.010	**5.55**	0.37	5.54	**0.07**	4.26

In-vivo flux determination: RESULTS

- **Flux distributions at G6P in *C. glutamicum***
  - Flexible node
  - Trans-hydrogenase activity discovered
- **Unequivocal evidence of PYR carboxylation**
- **Complete kinetic analysis of the H_4D node**
- **Mammalian cell culture:**
  - Glycosylation strongly depends on G6P fluxes
  - Strong correlation between glycolytic fluxes and apoptotic death

Fluxes provide invaluable perspectives of physiological state and metabolic control

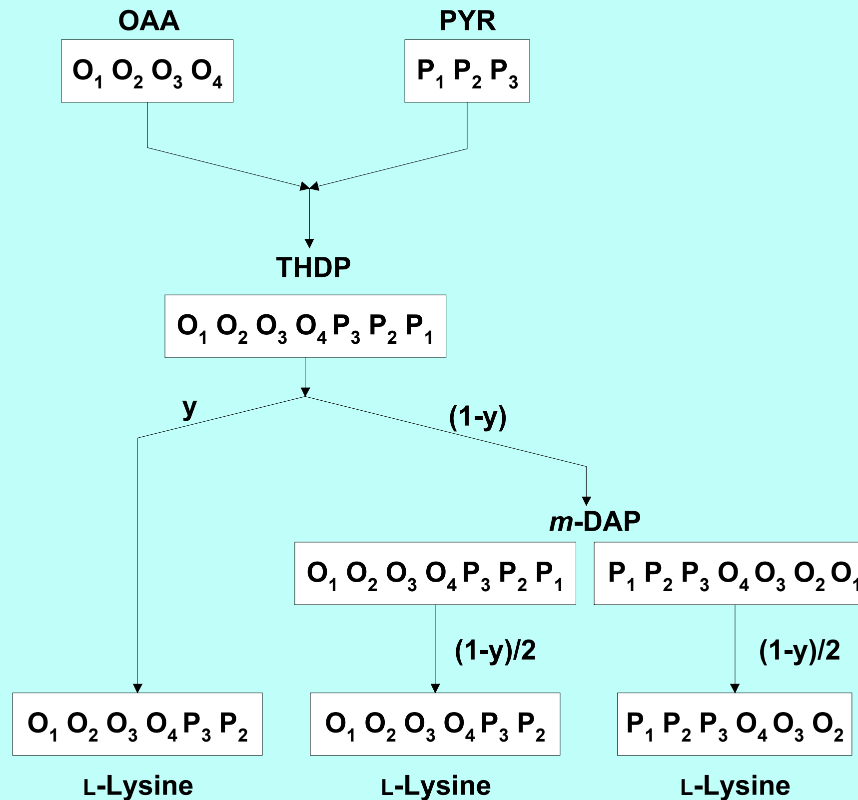


0.02
0.01
0
r_7 (mol/h)
Transient phase
State 1
State 2
0.01
0.02
r_8 (mol/h)

of
ete

Glc+4NADPH+2NAD —— Lys

Flux determination from ^{13}C label enrichment of (secreted) metabolites

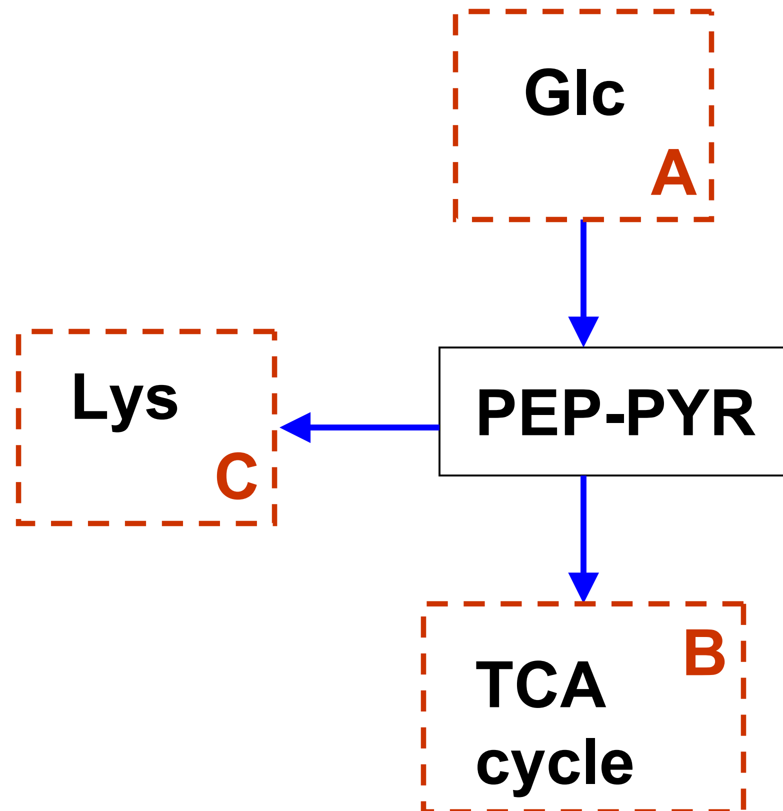


RESULTS (cont'd)

- Determination of *Group Control Coefficients (GCC)* from flux estimates
- GCC are measures of the control exercised by a group of reactions on a flux
- See notes

Group Control Coefficients:

Application to lysine biosynthesis



- Perturbations
  - Glucose concentration
  - GPI mutant
  - Gluconate addition
  - Fluoropyruvate
  - Threonine inhibition

Perturbed branch

Affected branch

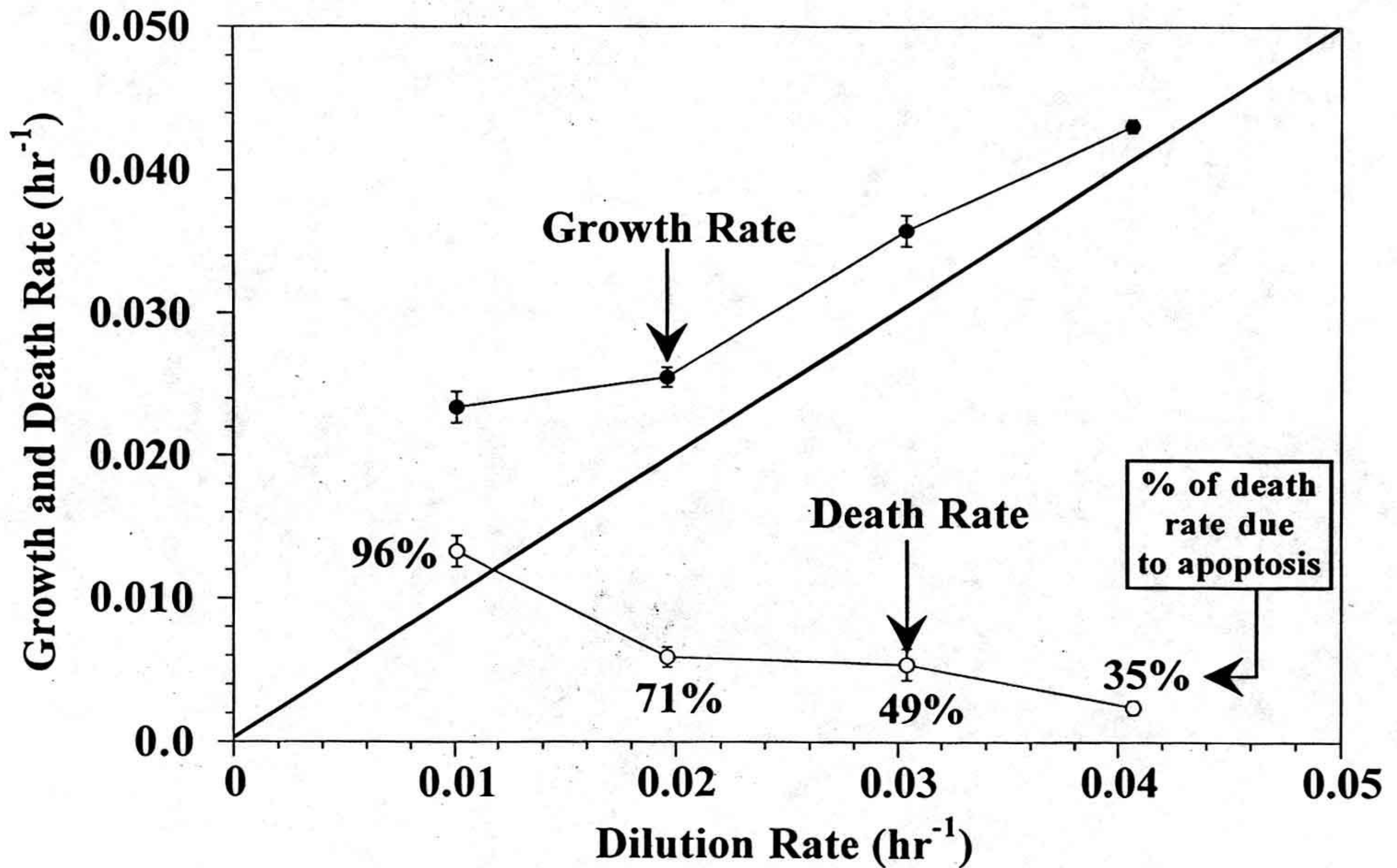
	A	B	C
A	0.07	0.51	0.42
B	0.09	1.22	0.31
C	0.02	-0.34	1.32

Science, **252:** 1675 (1991)
Biotech. Progress, **10:** 320 and 327 (1994)
Biotech.& Bioeng., **58:** 149 (1997)

RESULTS (cont'd)

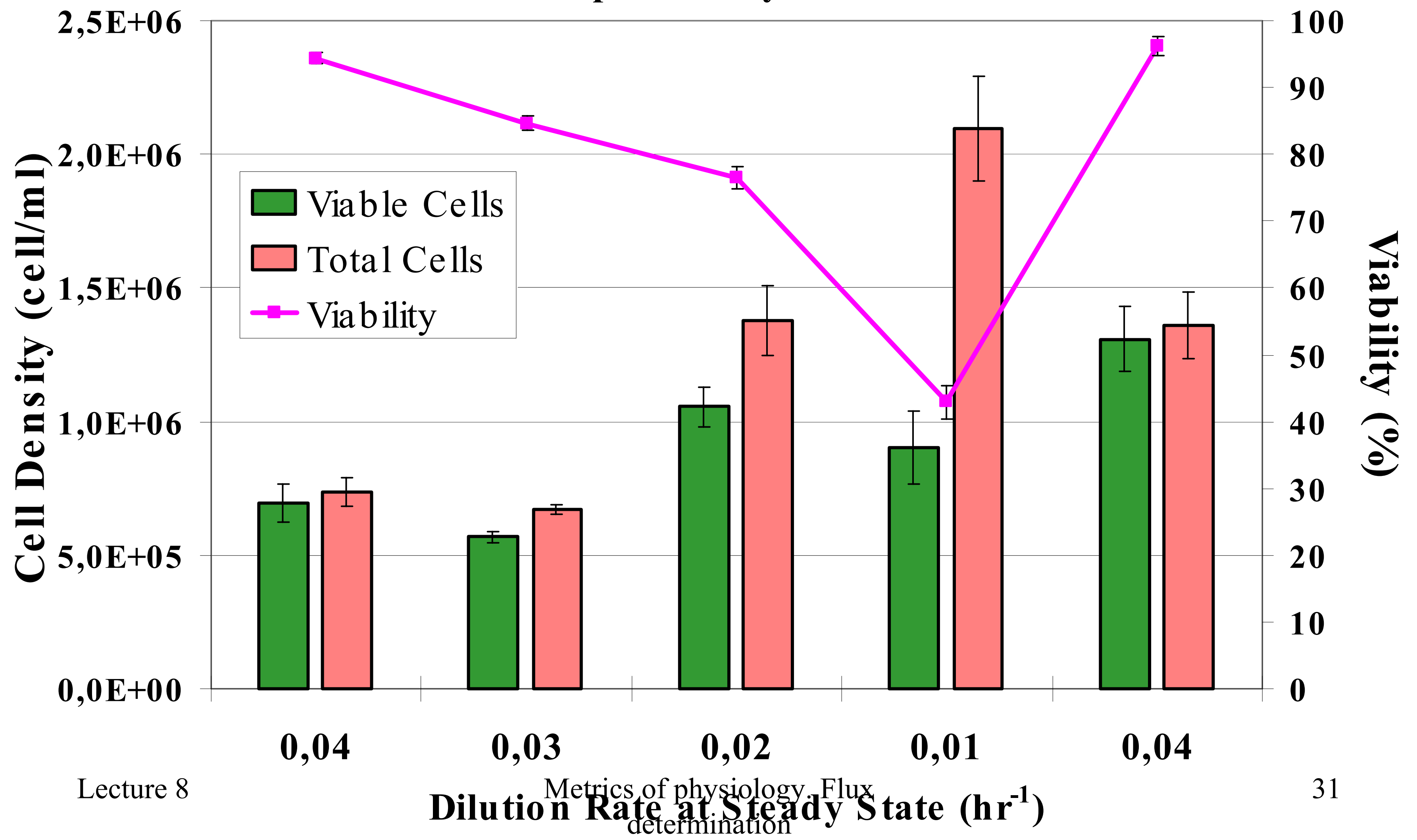
- **<u>Cell death in culture (CHO, Hybridomas)</u>**
  - Strongly correlated to culture energetics
  - Cells with high mitochondrial membrane potential (MMP) are resistant to apoptosis inducers

Growth and death rates vs. dilution rate



determination

Steady state multiplicity



Metabolic flux analysis

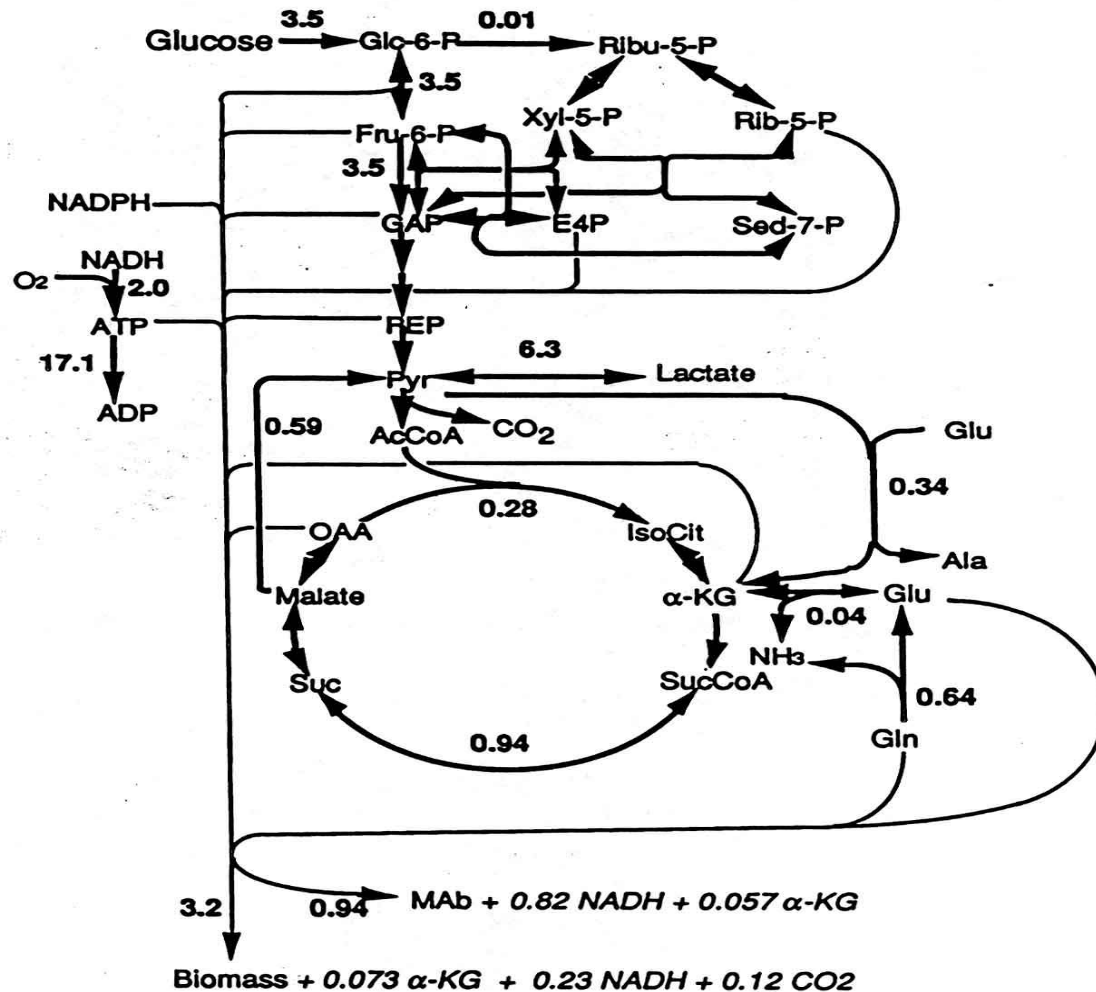


Figure 7.4a Estimated intracellular fluxes at DO=60. The numbers indicate the fluxes in units of mmole/cell/hr * 10^{-10}.

ATP generation and pyruvate flux distribution

1.3
1.2
1.1
1
0.9

$\frac{Flux_{in}}{Flux_{out}}$

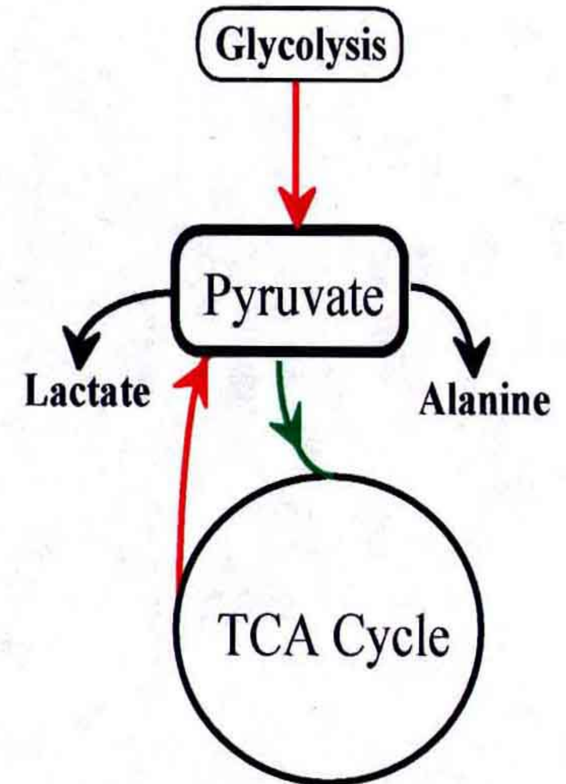
Death rate increases & viability drops

Byproduct Consumption

Byproduct Production

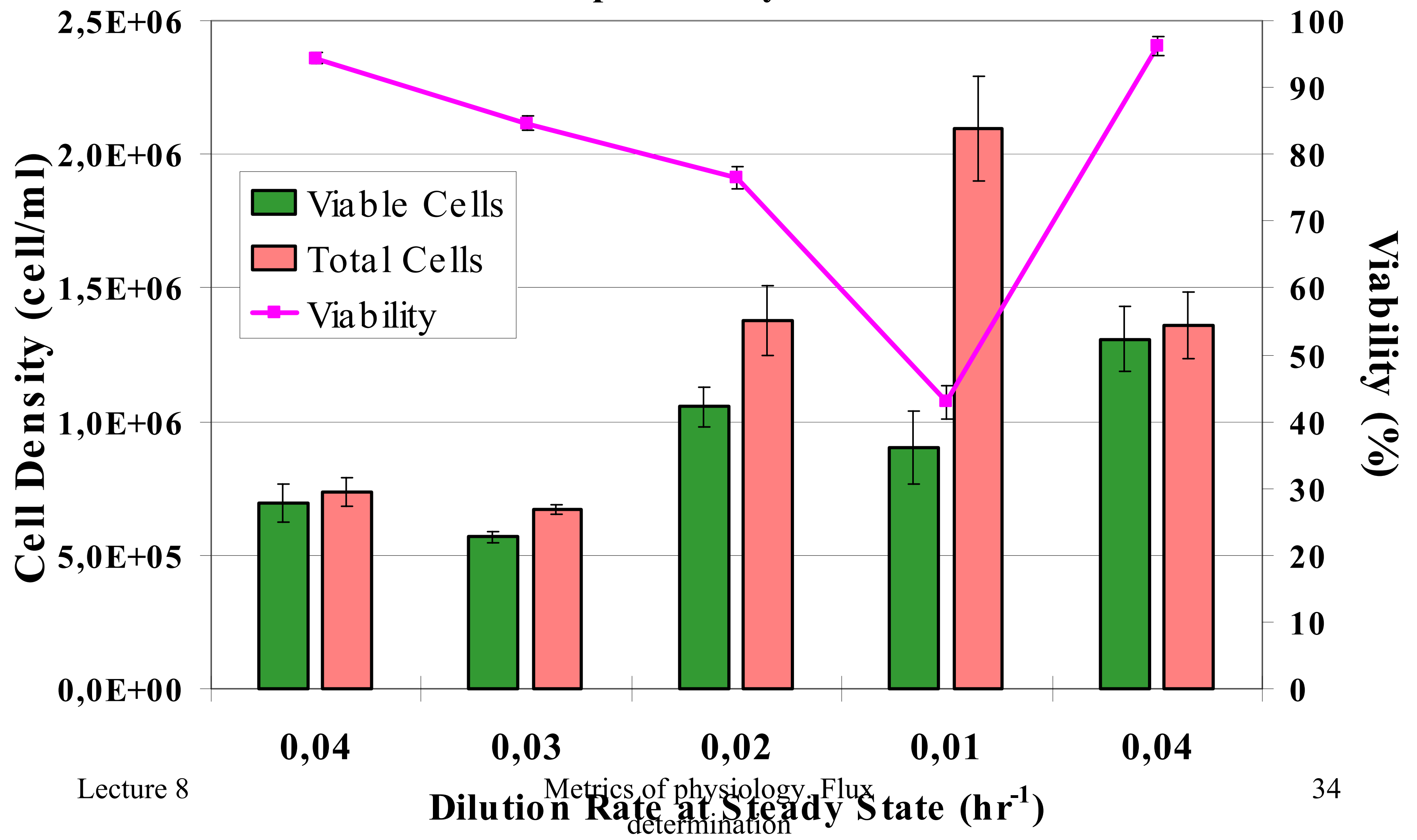
0.016 0.018 0.02 0.022 0.024

Growth Rate (hr^{-1})

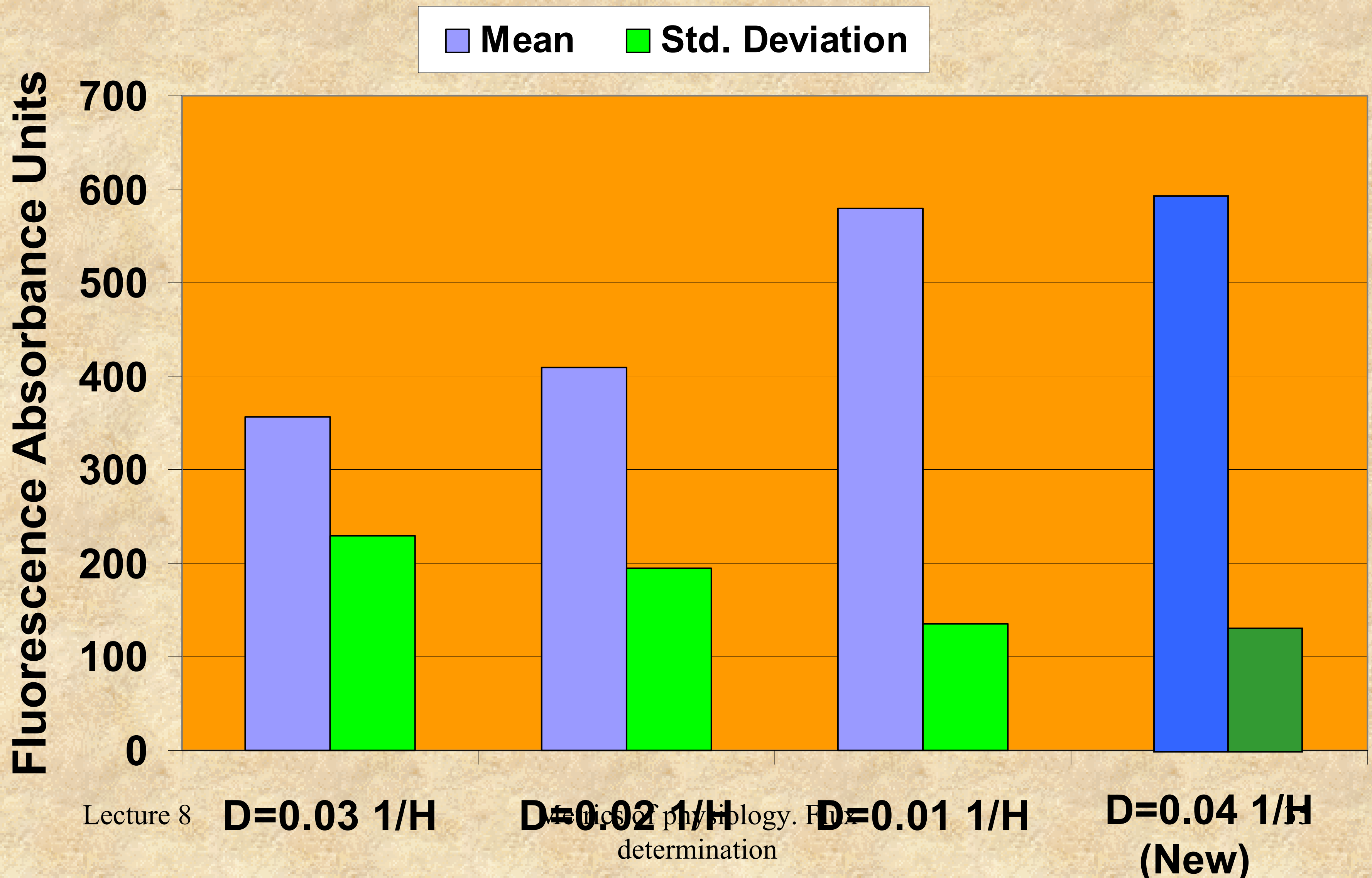


determination

Steady state multiplicity



Hybridoma physiology at steady states



Conclusions

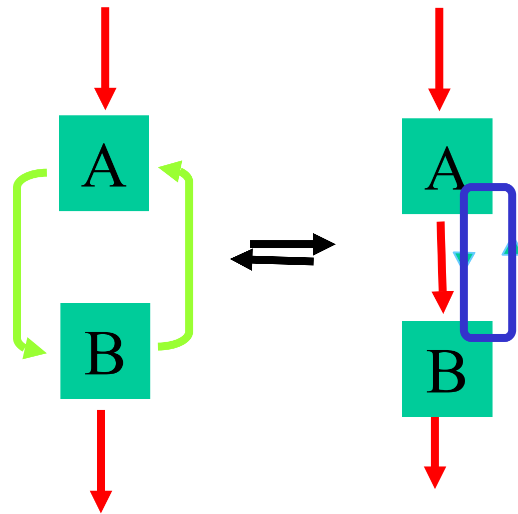
- Cell survival in nutrient-limited environment depends on mitochondria membrane potential
- Hybridomas with high metabolic activity can be selected in chemostats, (or by sorting)
- Hyperactive cells are very efficient in nutrient utilization (minimal lactate production). They retain the hypermetabolic state upon return to nutrient-rich environment

Important issues in flux determination

- Measurement redundancy allows validation of measurements and pathway biochemistry
- Pathway modification and discovery are part of the flux determination process
- Metabolic and isotopic steady states are strict requirements for isotopic label analysis
- Analogy to material structure determination

Using only extracellular accumulation rates:

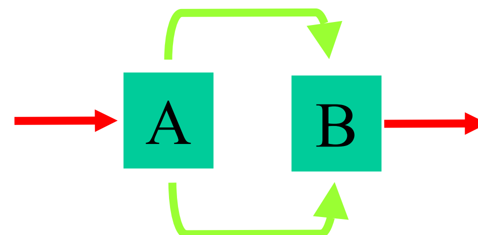
- only net fluxes can be observed



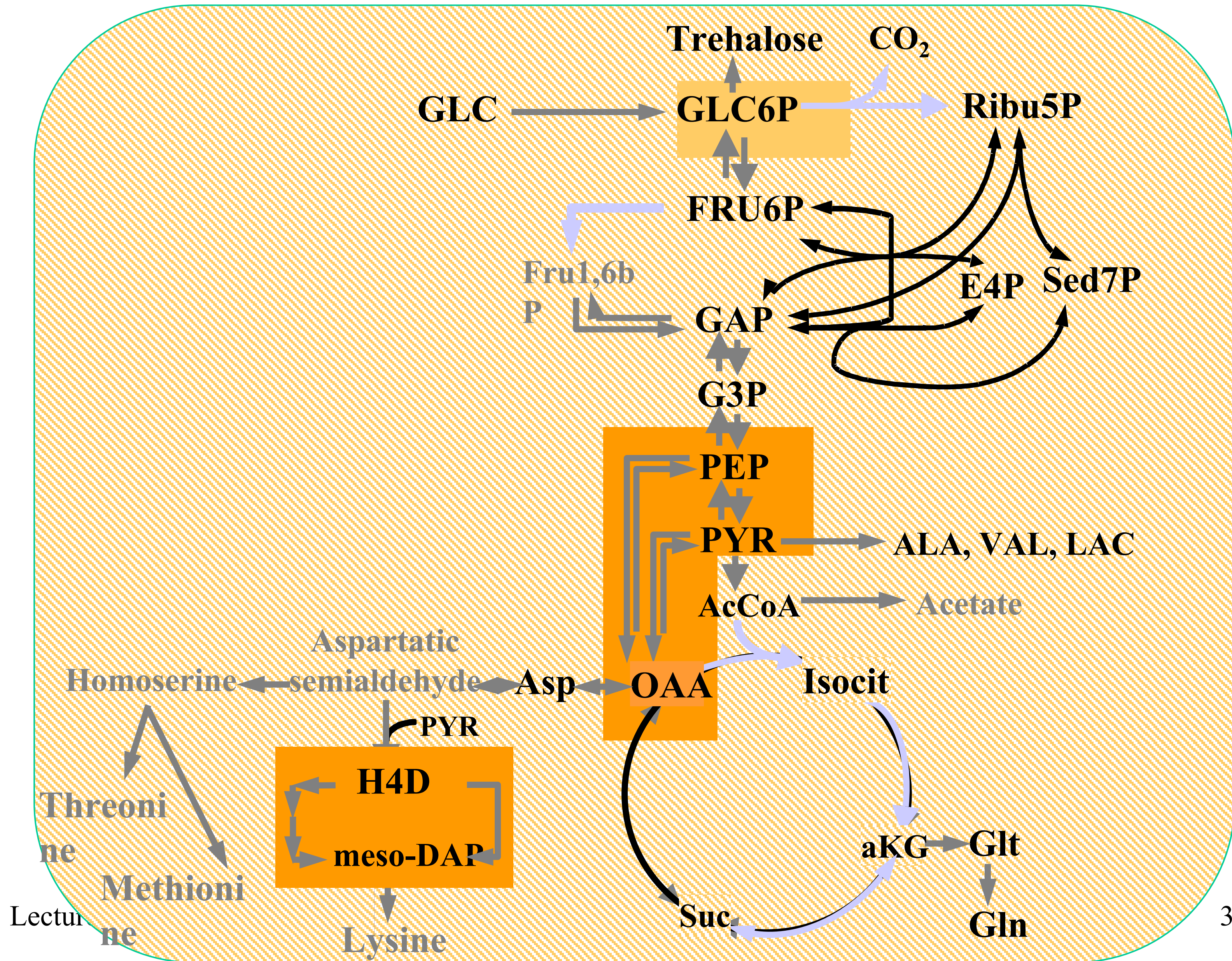
$$v^{net} = v^{f} - v^{b}$$

$$v^{exch} = \min(abs(v^{f}, v^{b}))$$

- net fluxes of pathways that branch apart and rejoin later can not be differentiated



Singular net flux groups in lysine biosynthesis network



Use of tracers in MFA

- Use of radioactive tracers
- Local network results (see example of H_4D branch point, example 9.1)
- Global network analysis. Methods based on balances of:

  - Metabolites

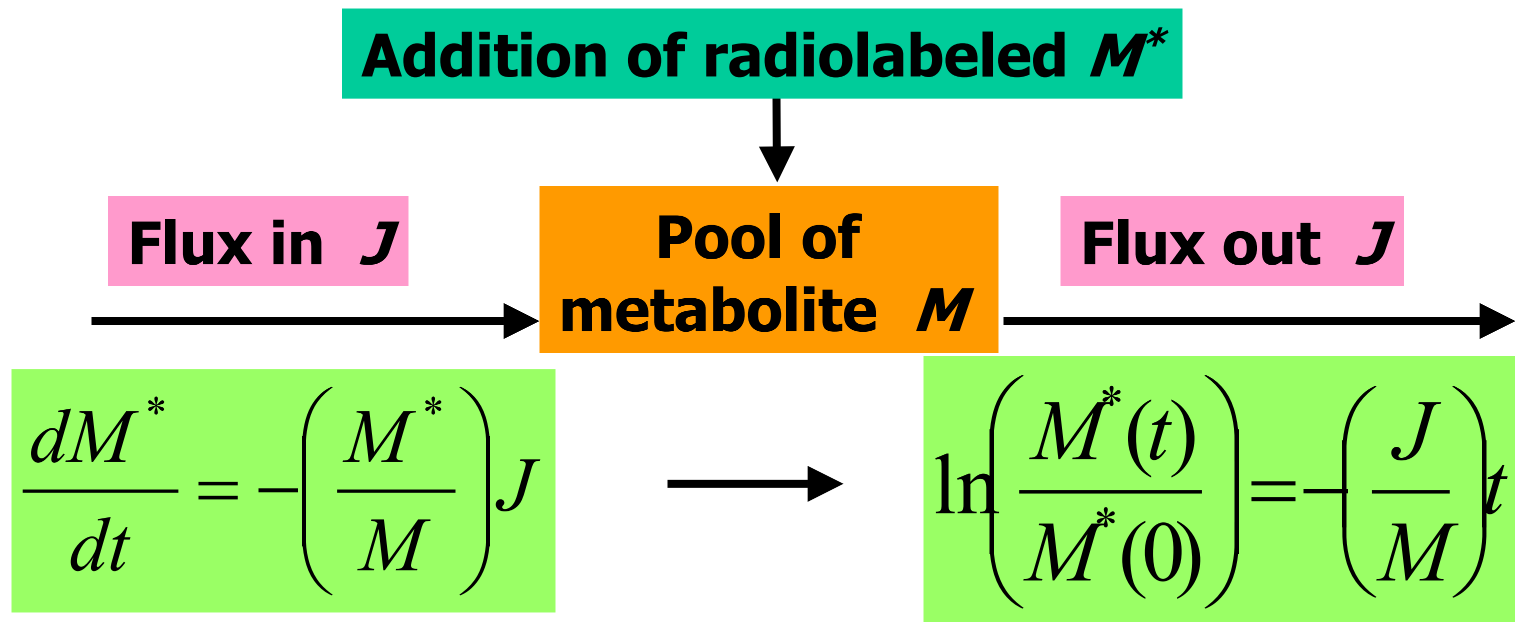
  - Label vs. non-label

  - Enrichment

  - Isotopomers
- Approach depends on measurements available

Use of radiolabeled substrates for flux determination

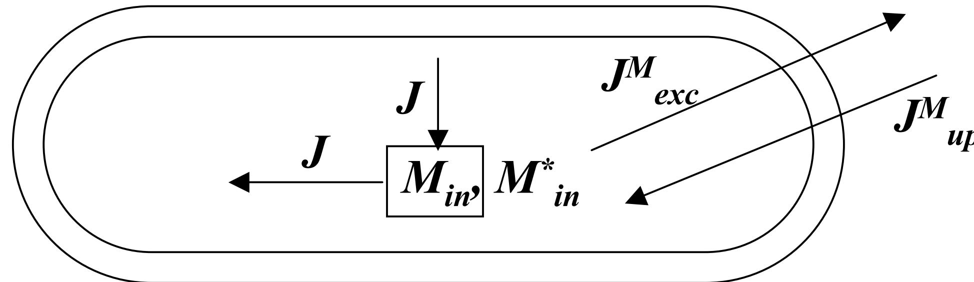
- At steady state, a pulse of a radiolabeled metabolite (M^*) can be used to determine the flux (J) from a metabolite pool (M) (single compartment model)



- Radiolabeled metabolites are used for greater measurement sensitivity and minimal perturbation of the steady state.

Considering transport

Steady state material balance equations on labeled metabolite M (*M**)



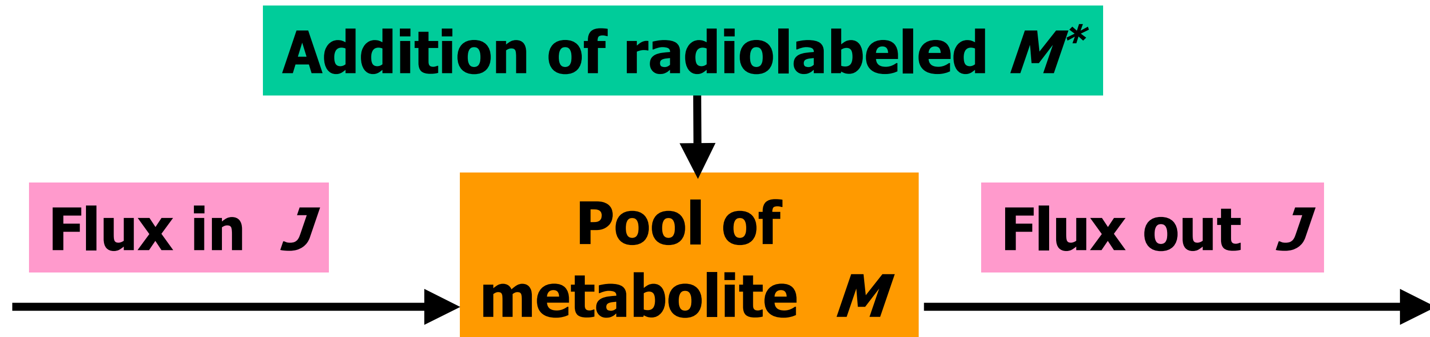
$$\frac{dM^*_{in}}{dt} = -\left(\frac{J}{M_{in}}\right)M^*_{in} - J^M_{exc}\frac{1}{M_{in}}M^*_{in} + J^M_{up}\frac{1}{M_{ex}}M^*_{ex}$$

$$\frac{dM^*_{ex}}{dt} = + J^M_{exc}\frac{1}{M_{in}}M^*_{in} - J^M_{up}\frac{1}{M_{ex}}M^*_{ex}$$

$$\frac{dM^*_{tot}}{dt} = -\left(\frac{J}{M_{in}}\right)M^*_{in}$$

Thus, we need to measure M_{in} , M^*_{in} and M^*_{tot}

Flux determination with radioactive tracers



$$\frac{dM^*_{tot}}{dt} = -\left(\frac{J}{M_{in}}X\right)M^*_{in}$$

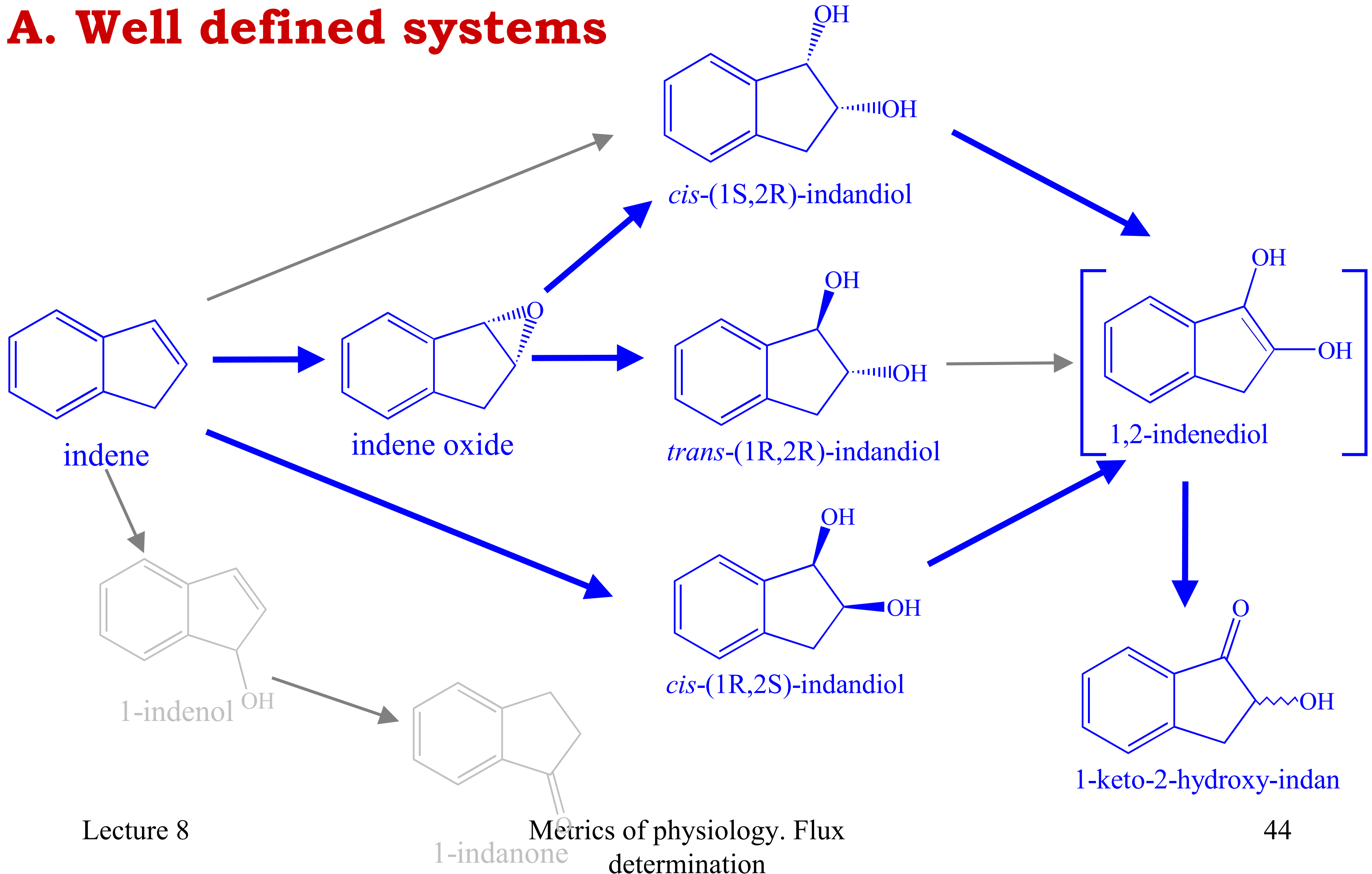
Assuming rapid eqlm. between intracellular and extracellular M:

$$\frac{dM^*_{tot}}{dt} = -\left(\frac{J}{M_{tot}}X\right)M^*_{tot}$$

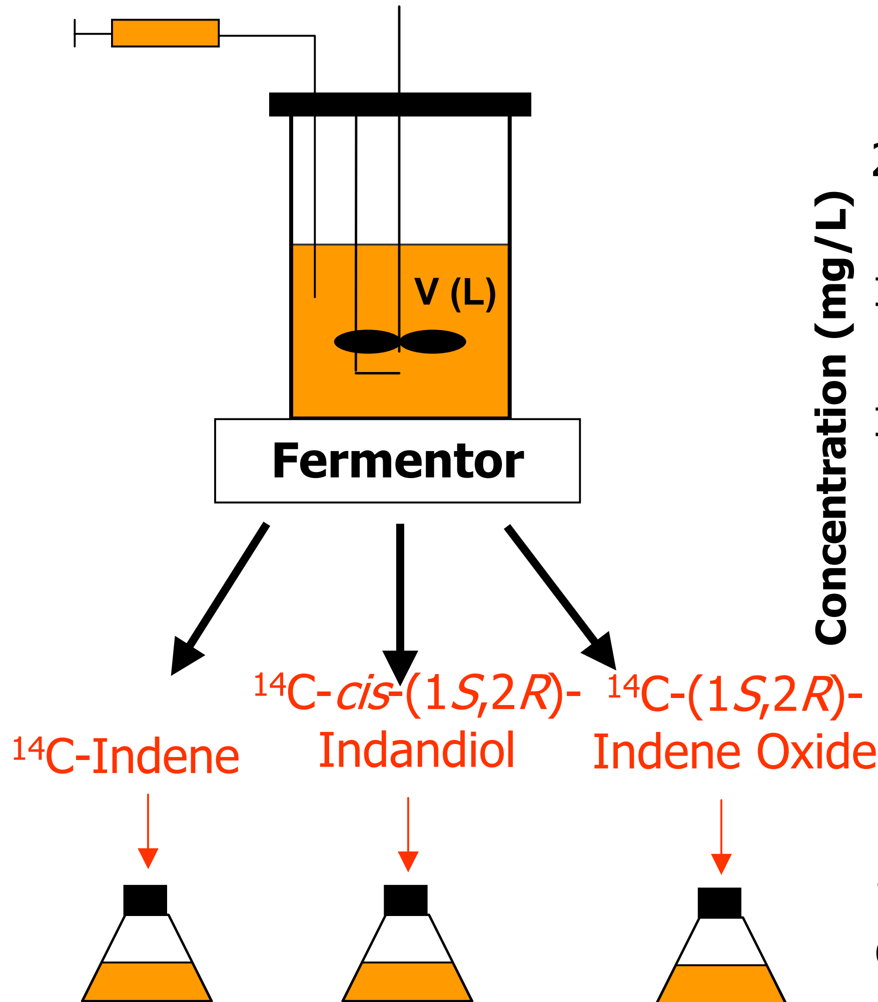
$$\ln\left(\frac{M^*_{tot}(t)}{M^*_{tot}(0)}\right) = -\left(\frac{J}{M_{tot}}X\right)t$$

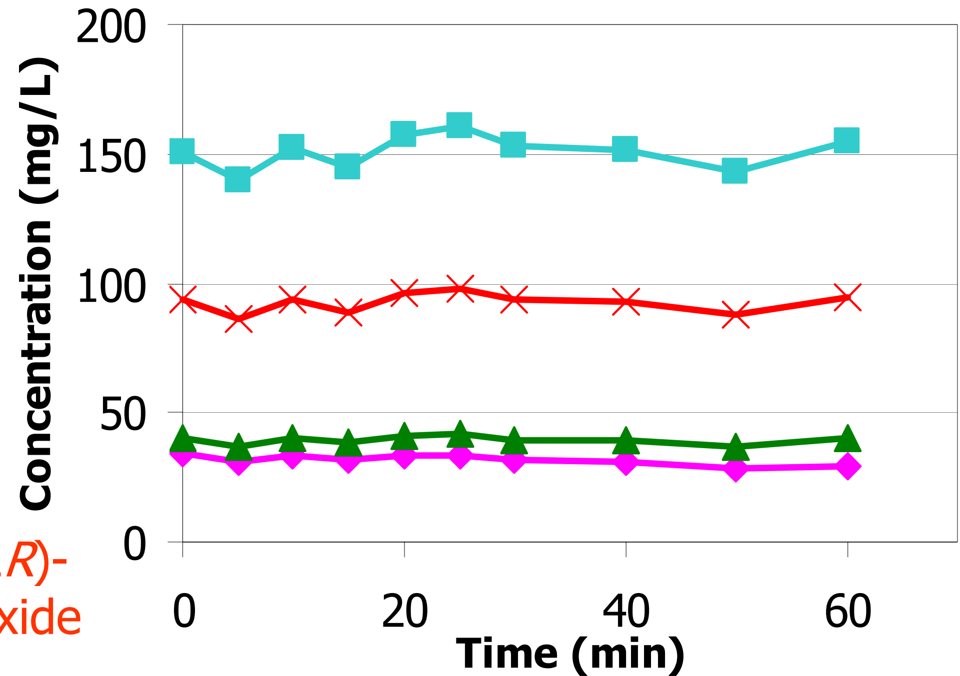
Methods depend on available measurements

A. Well defined systems



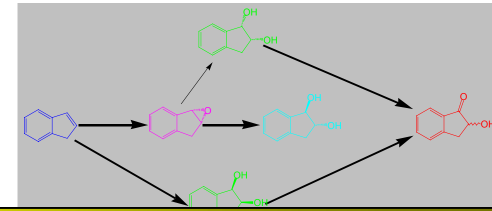
Use of the chemostat for flux measurements

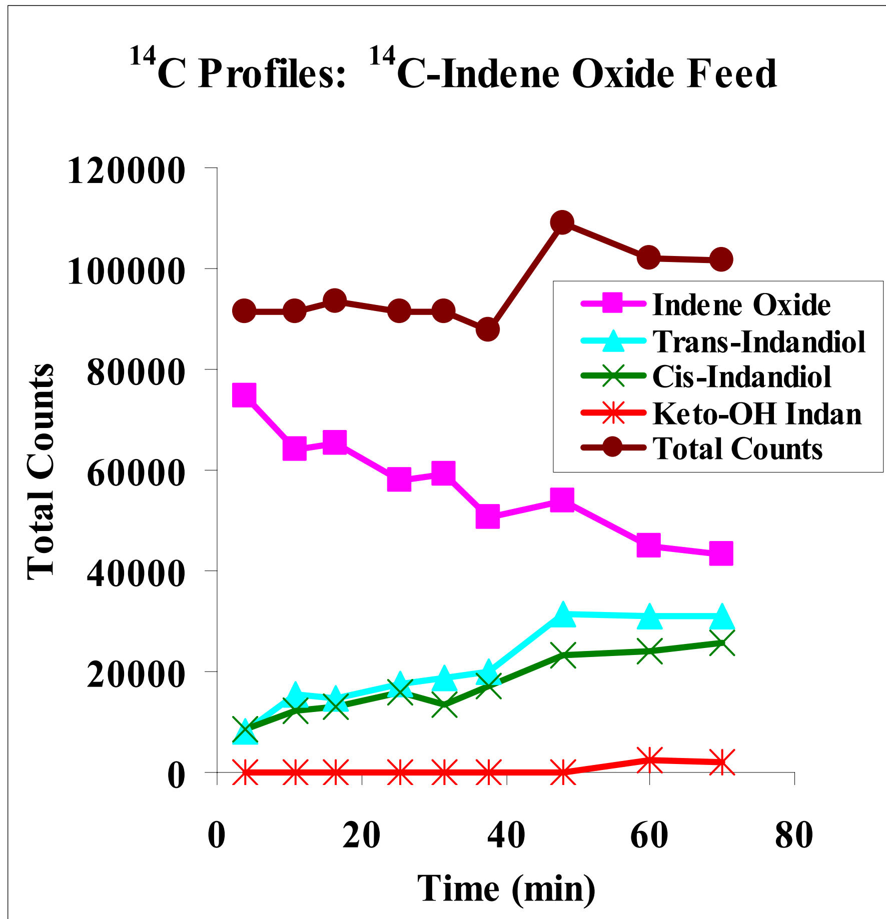


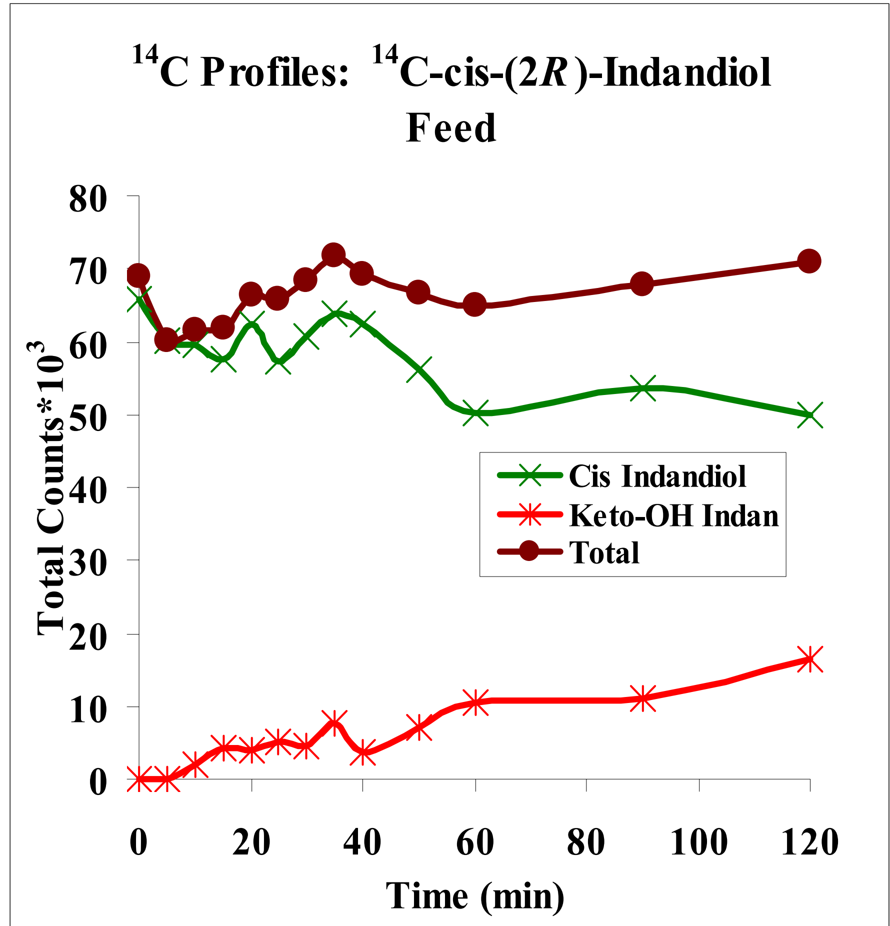


- Assume pseudo-steady state conditions maintained in shake flask

^{14}C-Tracer Experiments

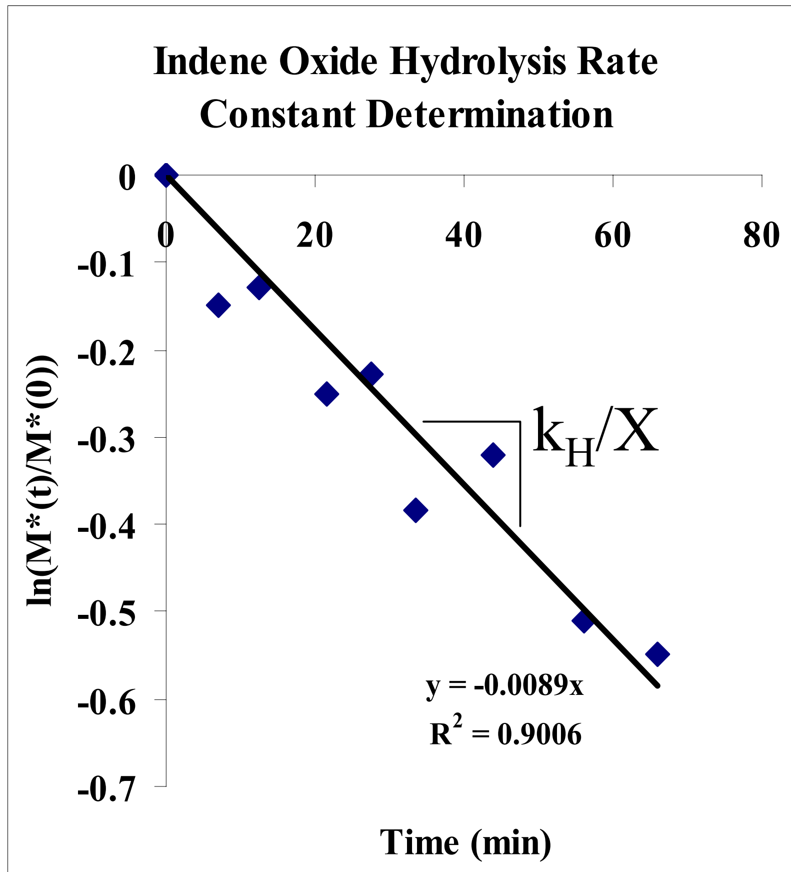


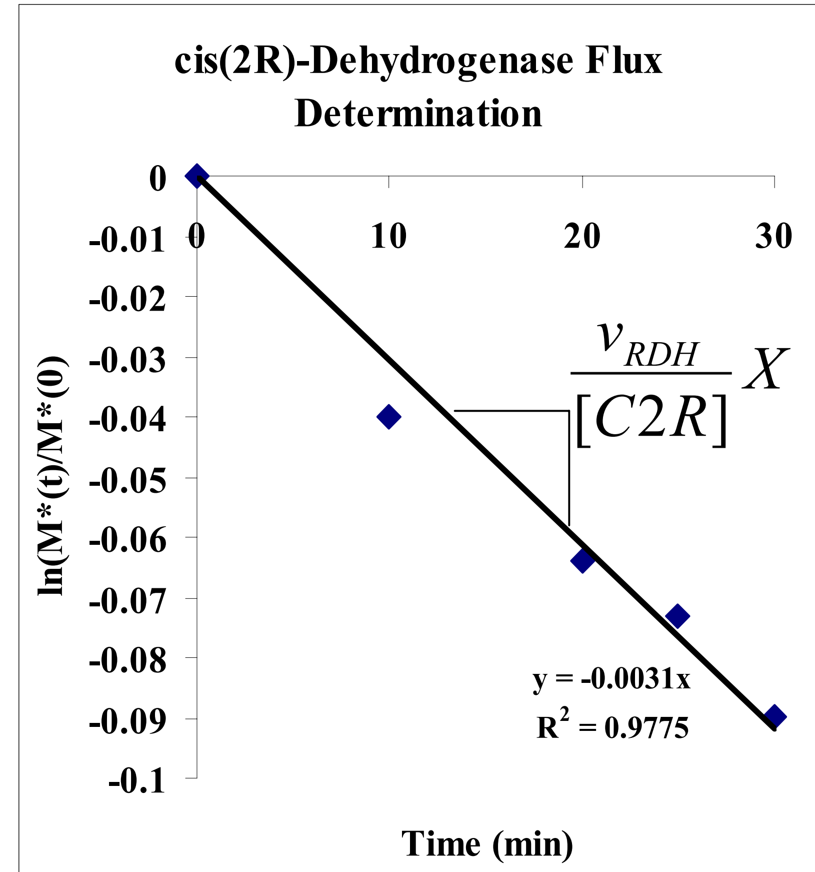




Flux Determination

$$\ln\left(\frac{M^*_{tot}(t)}{M^*_{tot}(0)}\right) = -\left(\frac{J}{M_{tot}}X\right)t = -kt$$





$k_H/X = 0.54 \pm 0.06\ h^{-1}$

$v_{RDH} = 18 \pm 2$ μmol/h/g DCW

Use of tracers in MFA

- Use of radioactive tracers
- **Local network results (see example of H_4D branch point, example 9.1)**
- Global network analysis. Methods based on balances of:

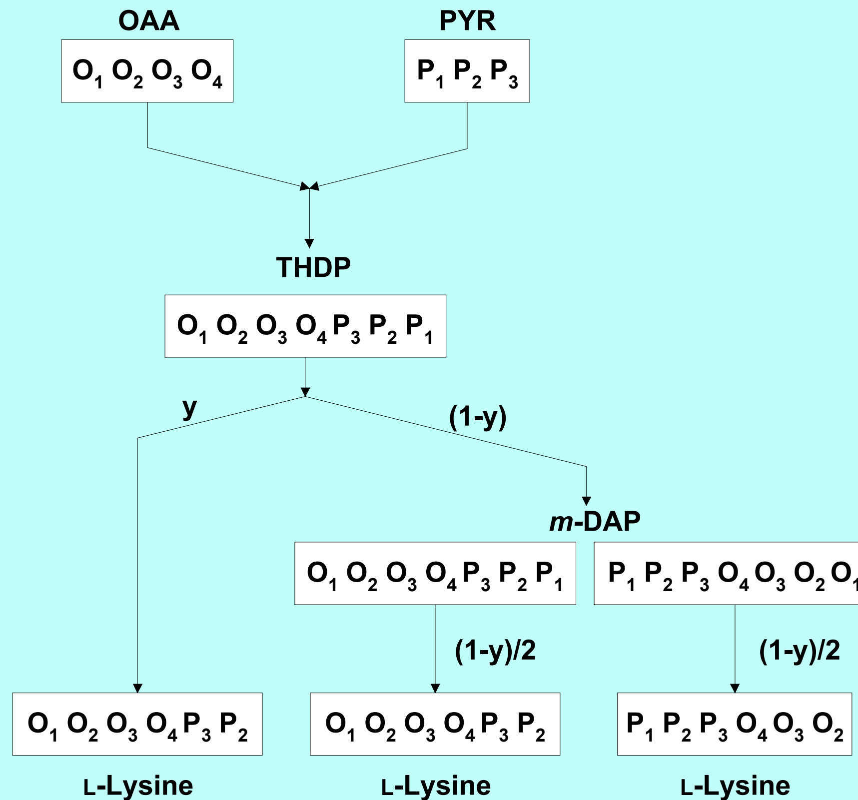
  - Metabolites

  - Label vs. non-label

  - Enrichment

  - Isotopomers
- Atom mapping matrices

Flux determination from ^{13}C label enrichment of (secreted) metabolites



Use of tracers in MFA

- Use of radioactive tracers
- Local network results (see example of H_4D branch point, example 9.1)
- **Global network analysis. Methods based on balances of**:

  - Metabolites

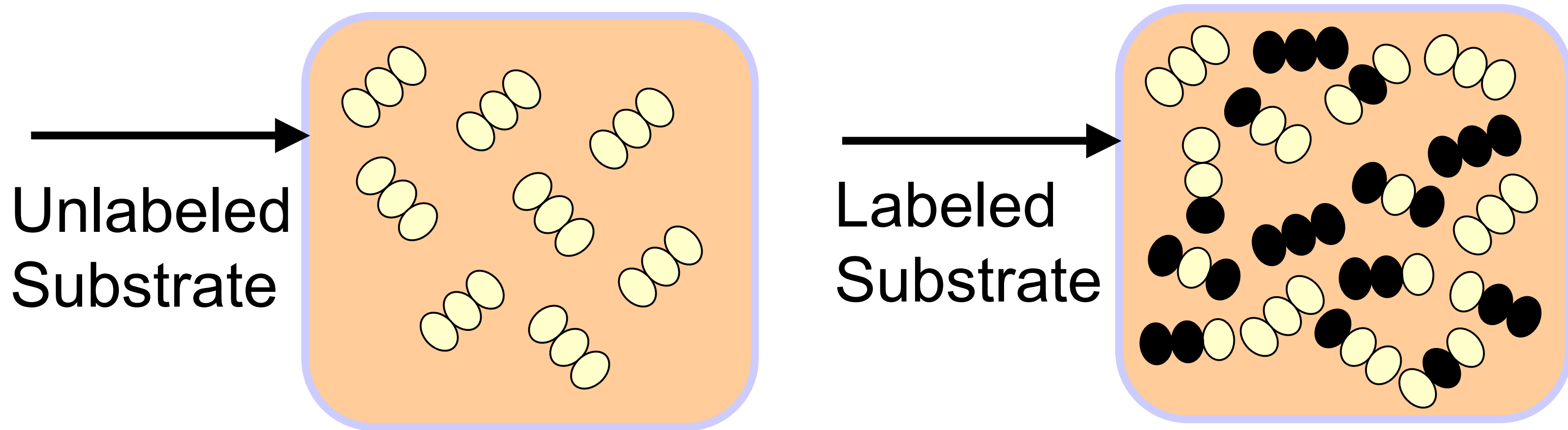
  - Label vs. non-label

  - Enrichment

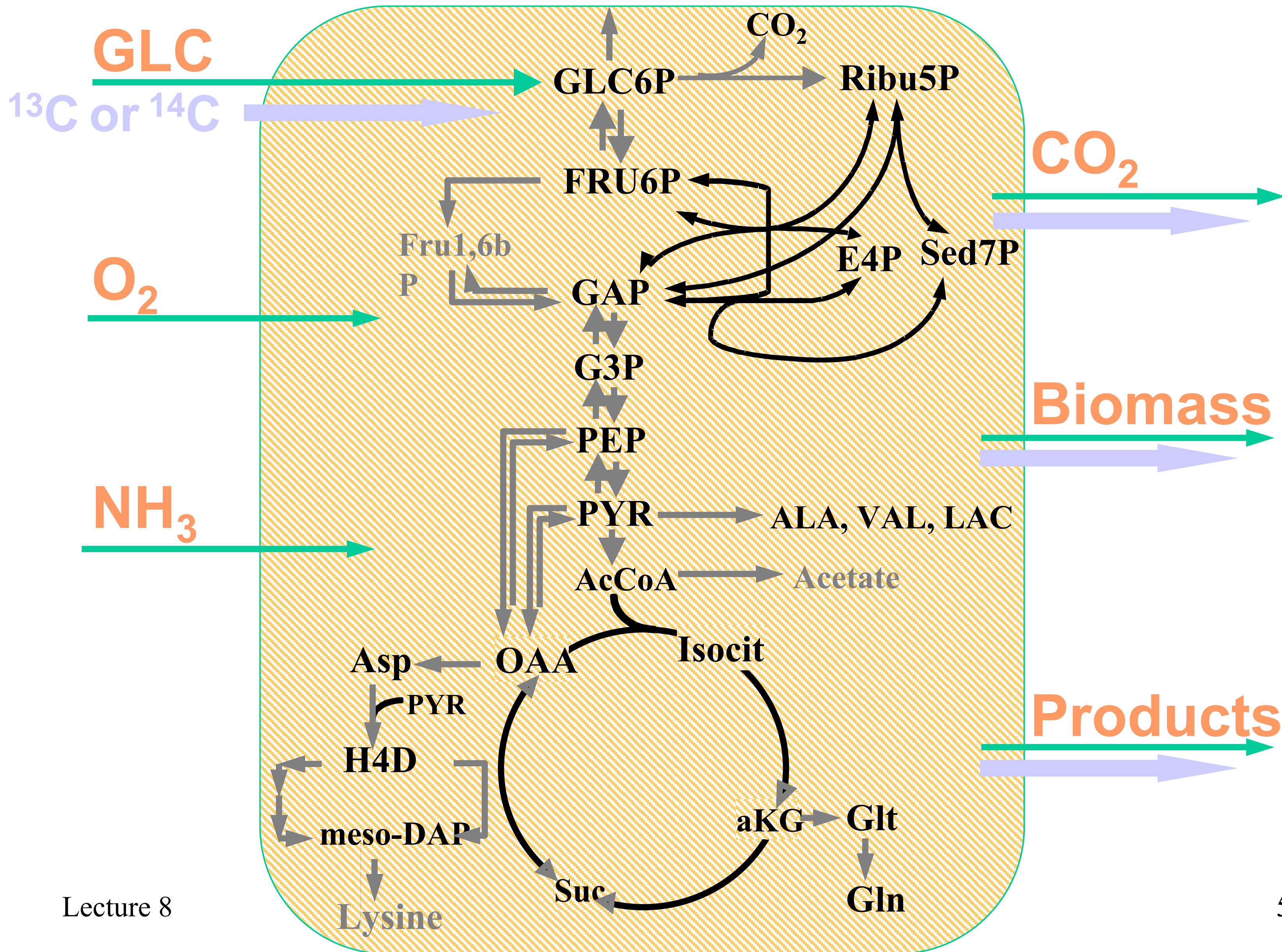
  - Isotopomers
- Atom mapping matrices

Isotopomers

- **Molecules of the same metabolite labeled differently upon introduction of a labeled substrate**

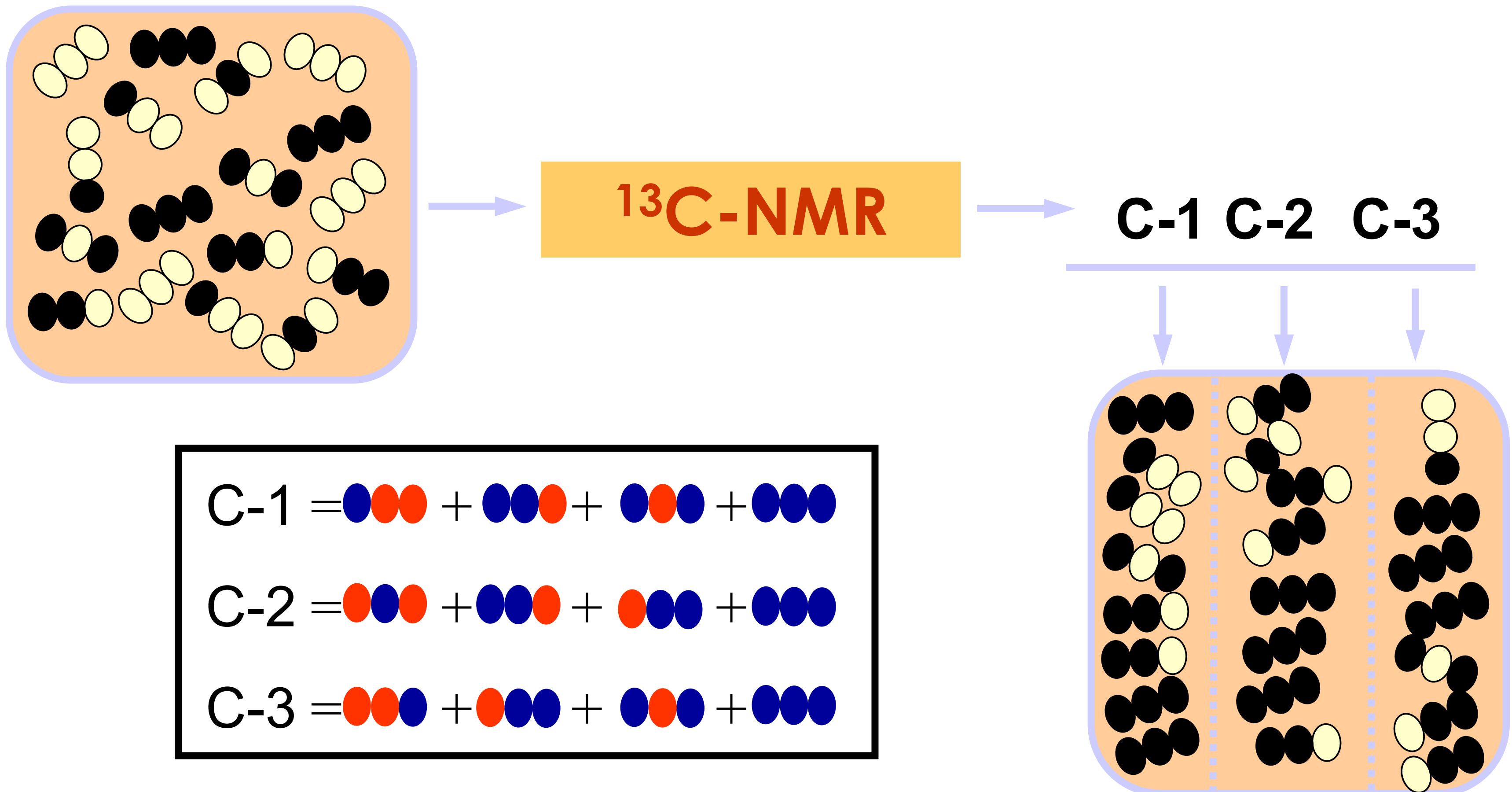


Use of isotopic tracers



^{13}C-NMR Spectroscopy

- Determines the label enrichment of a particular carbon atom in a molecule



Mass Spectrometry

- Differentiates ions based on mass/charge ratio

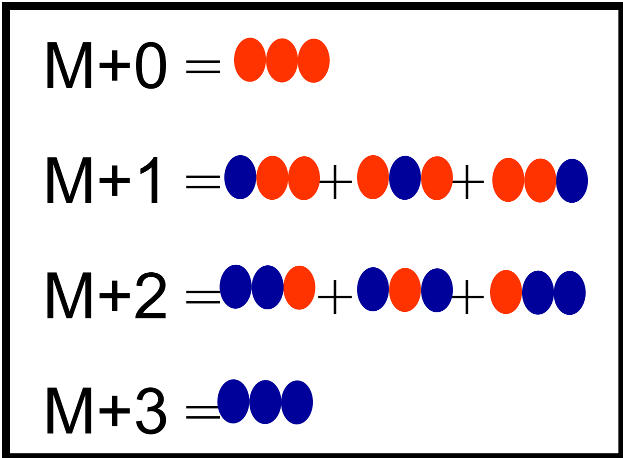
Mass Spectrometer

M+2

M+1

M+0

M+3



Detailed accounting of metabolite isotopomers allows:

- Accurate prediction of ^{13}C enrichment:

$$^{13}C_1 = L_{15} + L_{16} + L_{146} + L_{136}$$

$$^{13}C_2 = L_2 + L_{24} + L_{25} + L_{26} + L_{246}$$

$$^{13}C_3 = L_3 + L_{136}$$

- Prediction of iso-topomer MW distr'n

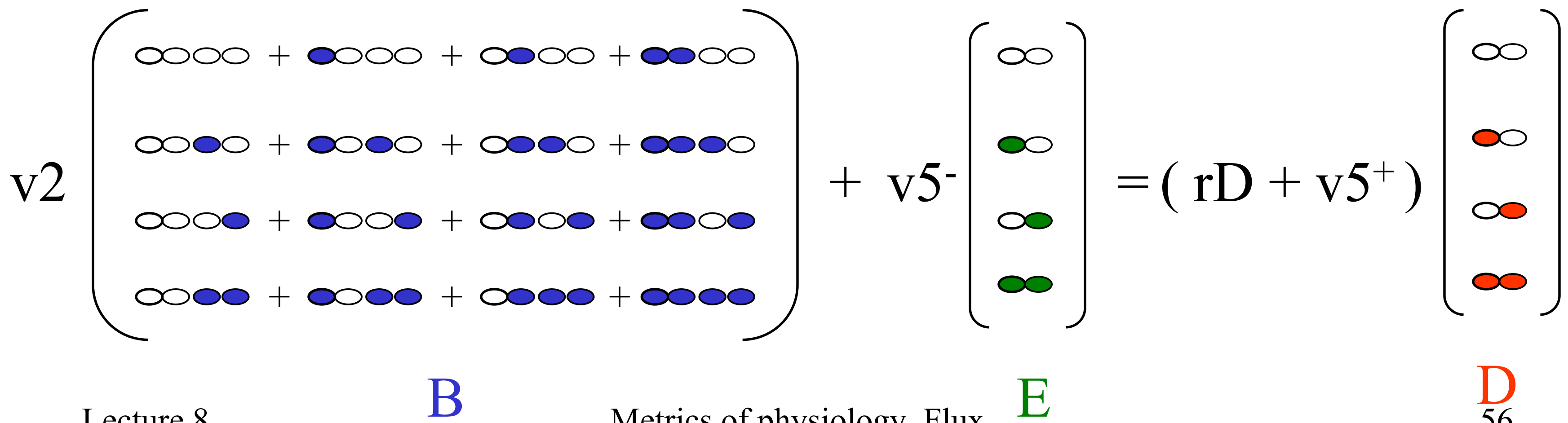
$$L_{M+1} = L_2 + L_3$$

$$L_{M+2} = L_{15} + L_{16} + L_{24} + L_{25} + L_{26} + L_{46}$$

$$L_{M+3} = L_{136} + L_{146} + L_{246}$$

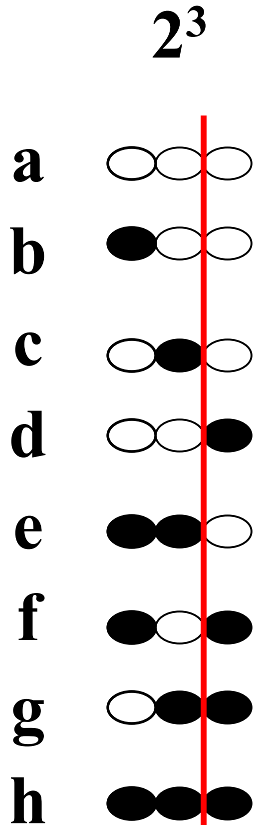
rF
F
rD
D
V2
rA
A
v1
B
v5-
v5+
v3+
v3-
rE
2C
v4
E
rC

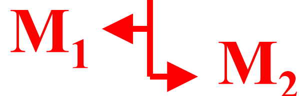
Isotopomer balance around D:



ISOTOPOMERS

- **Isotopomers:** metabolite molecules with different labeling patterns





Carbon enrichment distribution

C-1 = [b] + [e] + [f] + [h]

C-2 = [c] + [e] + [g] + [h]

C-3 = [d] + [f] + [g] + [h]

Molecular weight distribution

M = [a]

M+1 = [b] + [c] + [d]

M+2 = [e] + [f] + [g]

M+3 = [d] + [f] + [g] + [h]

M_1 = [a] + [d]

$M_1 + 1$ = [b] + [c] + [f] + [g]

$M_1 + 2$ = [e] + [h]

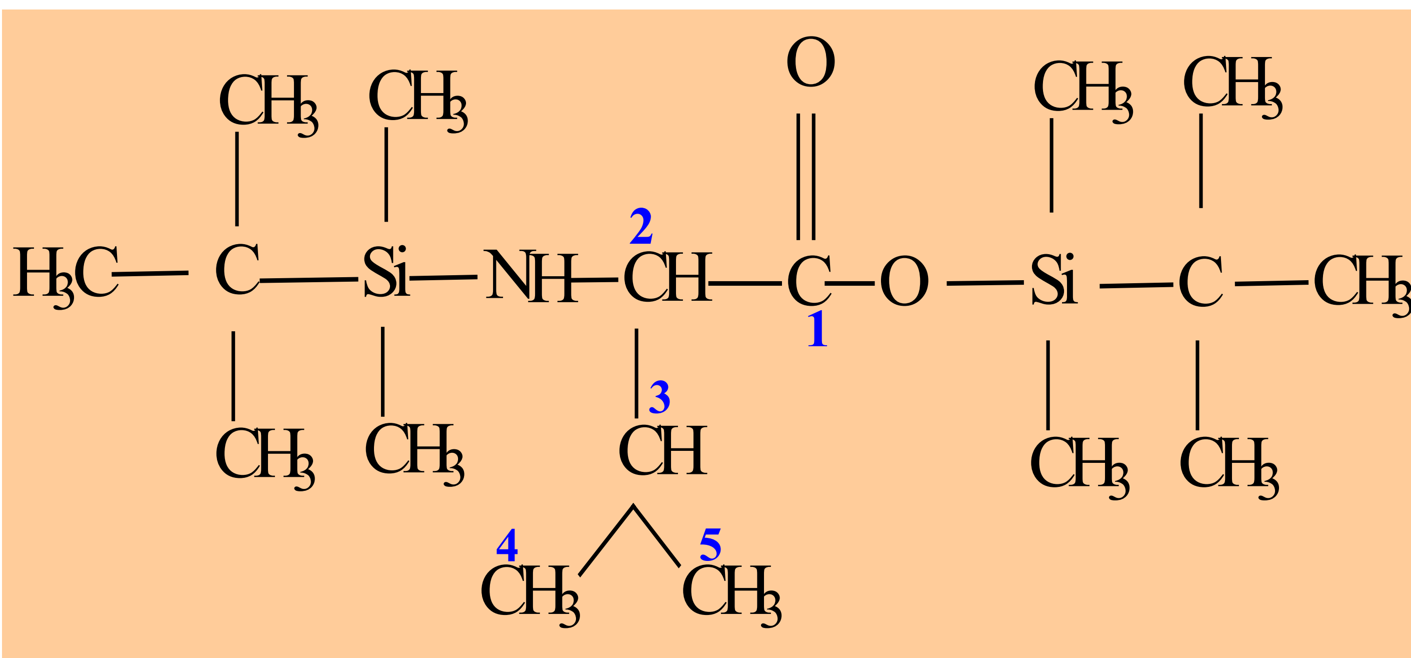
M_2 = [a] + [b] + [c] + [e]

$M_2 + 1$ = [d] + [f] + [g] + [h]

Derivatization of Amino Acids with TBDMS

- Derivatives can be volatilized for GC separation

TBDMS-derivative of Valine:

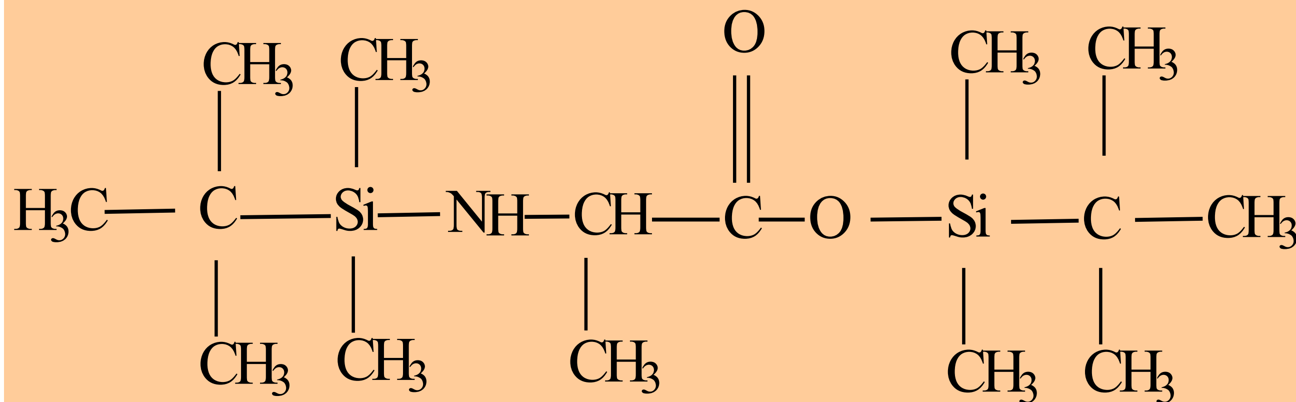


- To obtain the mass isotopomer distribution of the aminoacid, we need to correct the measurements for the natural abundance of the rest of the atoms in the derivatives

Fragmentation of TBDMS-derivatives

- Ionization prior to MS generates fragments of the derivatives

  Positional information about the isotopic tracer can be obtained



260+0 =

260+1 = + +

260+2 = + +

260+3 =

158 + 0 = +

158 + 1 = + + +

158 + 2 = +

ISOTOPOMER DISTRIBUTION ANALYSIS

$$g_i : \underline{v} * \underline{\underline{A}}_i * \underline{X}_{iso} + \underline{X}_{iso}^T * \underline{\underline{Q}}_i(\underline{v}) * \underline{X}_{iso} = c$$

i = 1,2,…,N (number of isotopomers)

$\underline{X}_{iso}$: vector of relative populations of isotopomers

$\underline{v}$: vector of intracellular fluxes

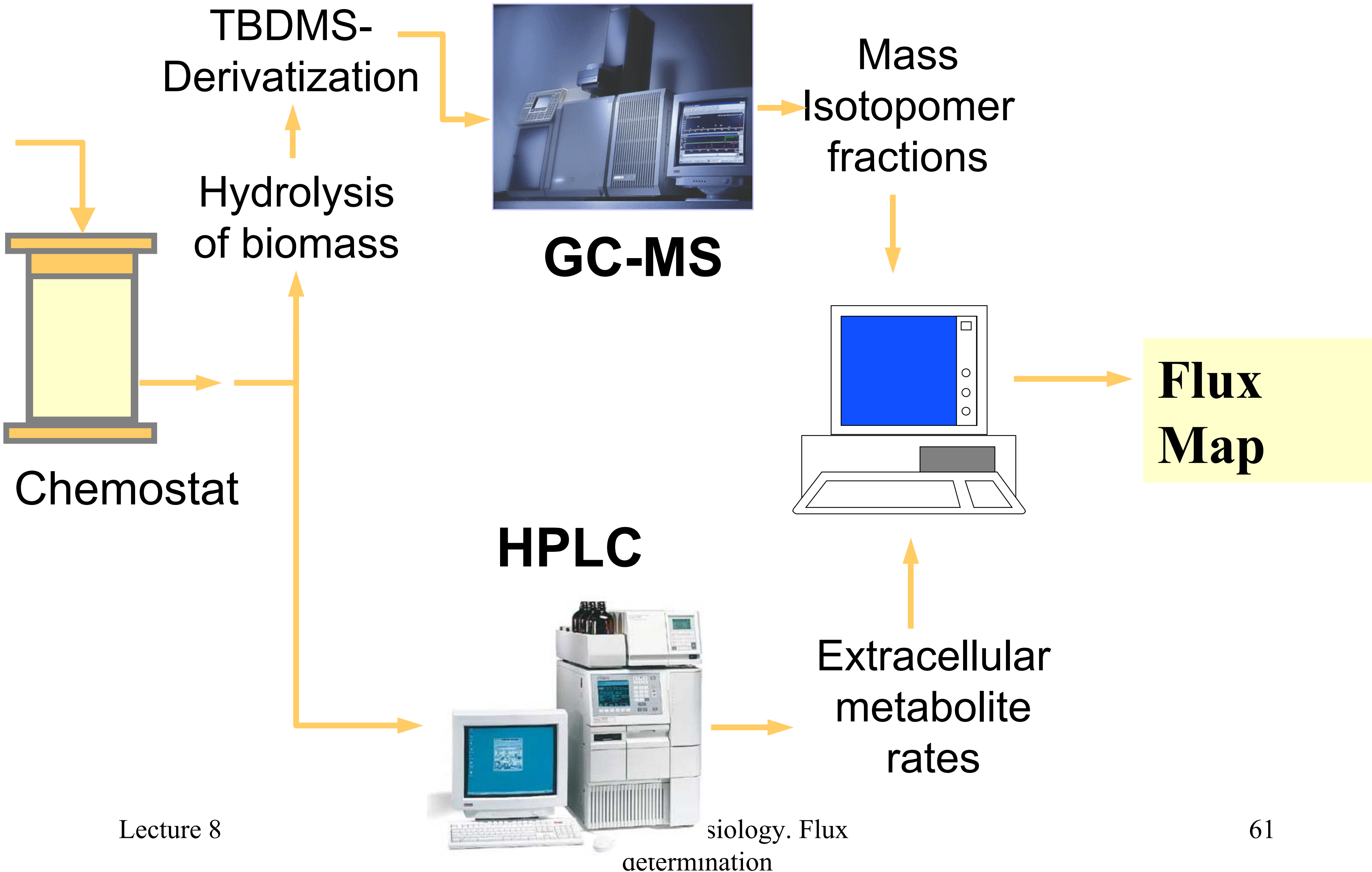
Carbon enrichments $\underline{M}_{enr} = \underline{\underline{K}} * \underline{X}_{iso}$

Molecular weight distribution $\underline{M}_{MW} = \underline{\underline{L}} * \underline{X}_{iso}$

Fine structure of NMR spectra $\underline{M}_{NMR,fine} = \underline{\underline{U}} * \underline{X}_{iso}$

known

Metabolic Flux Determination



Mass spectrometric measurements (1)