

代謝ネットワークのシステムバイオロジー

東京大学大学院 新領域創成科学研究科
慶応義塾大学 先端生命科学研究所
産総研 生命情報科学研究センター

有田正規
ARITA Masanori
arita@k.u-tokyo.ac.jp

時の流行: システムバイオロジー

□ 英語では Systems Biology

生命のメカニズムを「システムティックに、システムとして理解する」という発想から生まれた。

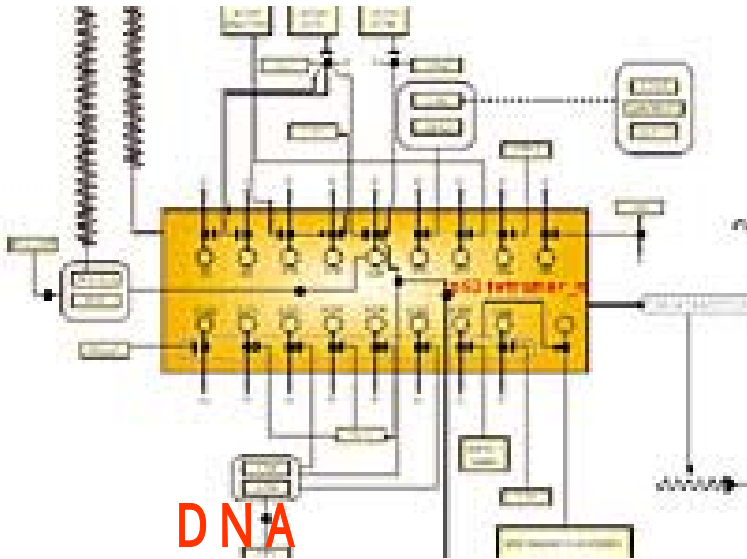
ゲノム、トランスクリプトーム、メタボローム、プロテオーム、など、全ての情報を統合して解析する。当然ながら、代謝の知識も計算機上で統合する。

「システムとして」の意味

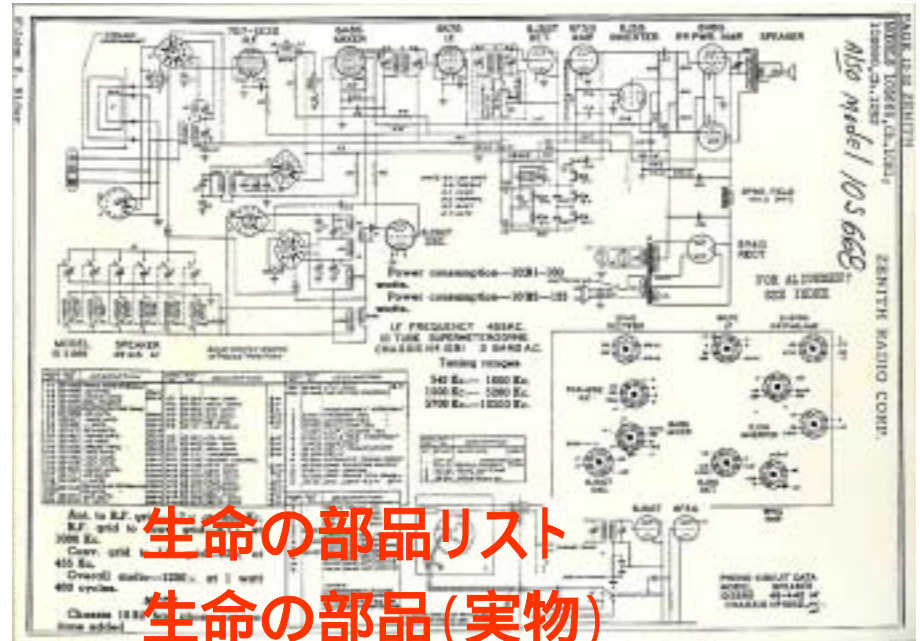
「システム」〔おもに工学で〕きわめて多数の構成要素から成る集合体で、各部分が有機的に連繋(レンケイ)して、全体として一つの目的を持った仕事をするもの。
(新明解国語辞典)

したがって、常に各部分との連携(インタラクション)を考えて解析を進めなくてはならない。が、生体ネットワークの概要は従来も実験で得られてきたはずでは？

部品を研究する時代は終わった システムバイオロジー vs ラジオ



DNA
タンパク質,代謝物
相互作用
細胞の振舞い
組織の振舞い



生命の部品リスト
生命の部品(実物)
配線、信号経路
ラジオ
コンピュータ

オーム研究は仮説生成の手段である

従来は分子単位のアプローチにより、細胞内のパスウェイを研究していた。

問題点

- 全体像をつかむのに時間がかかる
- 複数のパスウェイがどう相互作用するか、明らかにならない
- データが散逸している（抜けた部分が多い）

雑然たる知識を普遍妥当な知識に変えるには、大量の分子相互作用に関する（確証がなくても）「仮説」が必要。

オーム：網羅的解析に基づく仮説の生成

Table 1. Two systematic ways to learn about pathways^a

	(1) Directly observe the interactions	(2) Observe states induced by interactions
	Protein–DNA interactions	Gene expression
Methods	Chromatin immunoprecipitation followed by microarray analysis	DNA microarrays; SAGE
Databases	TRANSFAC (http://transfac.gbf.de/TRANSFAC/) BIND (http://www.bind.ca/)	GEO (http://www.ncbi.nlm.nih.gov/geo/) ArrayExpress (http://www.ebi.ac.uk/microarray/ArrayExpress/)
	Protein–protein interactions	Protein levels, locations, modifications
Methods	Two hybrid system Co-immunoprecipitation followed by mass spectrometry	Mass spectrometry; 2D PAGE Protein tagging followed by fluorescence microscopy; Protein arrays
Databases	BIND (http://www.bind.ca/) DIP (http://dip.doe-mbi.ucla.edu/) BRITE (http://www.genome.ad.jp/brite/) MIPS (http://mips.gsf.de/)	SWISS-2DPAGE (http://us.expasy.org/ch2d/) TRIPLES (http://ygac.med.yale.edu/triples/) Scansite (http://scansite.mit.edu/)
	Metabolic interactions and reactions	Metabolite and drug levels
Methods	No truly systematic measurements, although protein arrays show promise	Mass spectrometry; two-dimensional NMR Current challenge is to determine the molecular identities of all distinct compounds detected
Databases	MetaCyc (http://biocyc.org/metacyc/) KEGG (http://www.genome.ad.jp/kegg/) Klotho (http://www.biocheminfo.org/klotho/)	Public repositories of metabolic profiles not widely available, although data exchange standards for expression profiles (e.g. MAGE-ML) might support metabolic data in future.

^aThis table is provided as a representative sample of methods and databases, not as a comprehensive listing. We apologize in advance to those whose work was omitted because of space considerations.

T. Ideker & D. Lauffenburger *Trends Biotechnol.* 21(6), 2003.

仮説の正しさ: トランスクリプトームの場合

精神分裂病患者(米国人口の1%以上)において遺伝子発現量が大きく異なる遺伝子群の比較

- アフィメトリクス geneチップによる前頭葉における発現量 89 genes
- cDNAアレイによる前頭葉 49 genes
- 遺伝学、RT-PCR, ELISAなどによる100研究 97 genes

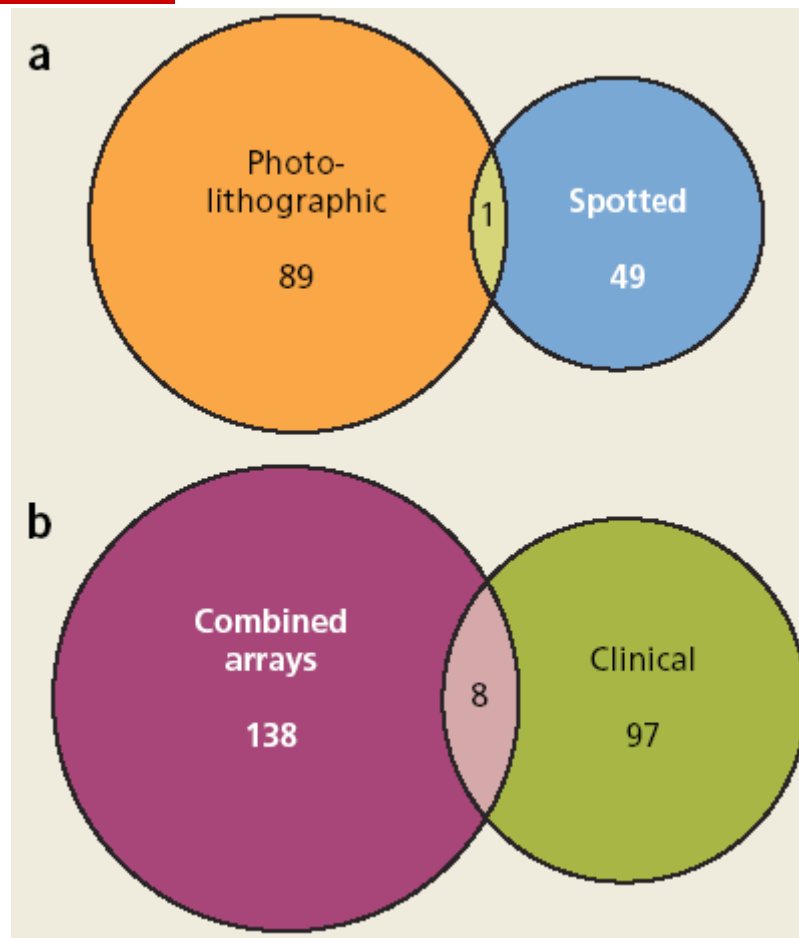
Miklos, GLG & Maleszka, R

Nat Biotechnol, 22(5) 615-621, 2004

一致度

- アレイどうしで共通する遺伝子は8個
 - 発現量の増減まで一致したのは1個 PRKCB1
- 既存の研究と一致したのは8個
 - 3種全てに共通するのは2個 GAD1, MDH1

脳の他の部位を調べても結果は類似していた。



仮説の正しさ: プロテオームの場合

Table 2. Estimation of False Positives in Whole-Genome Scans

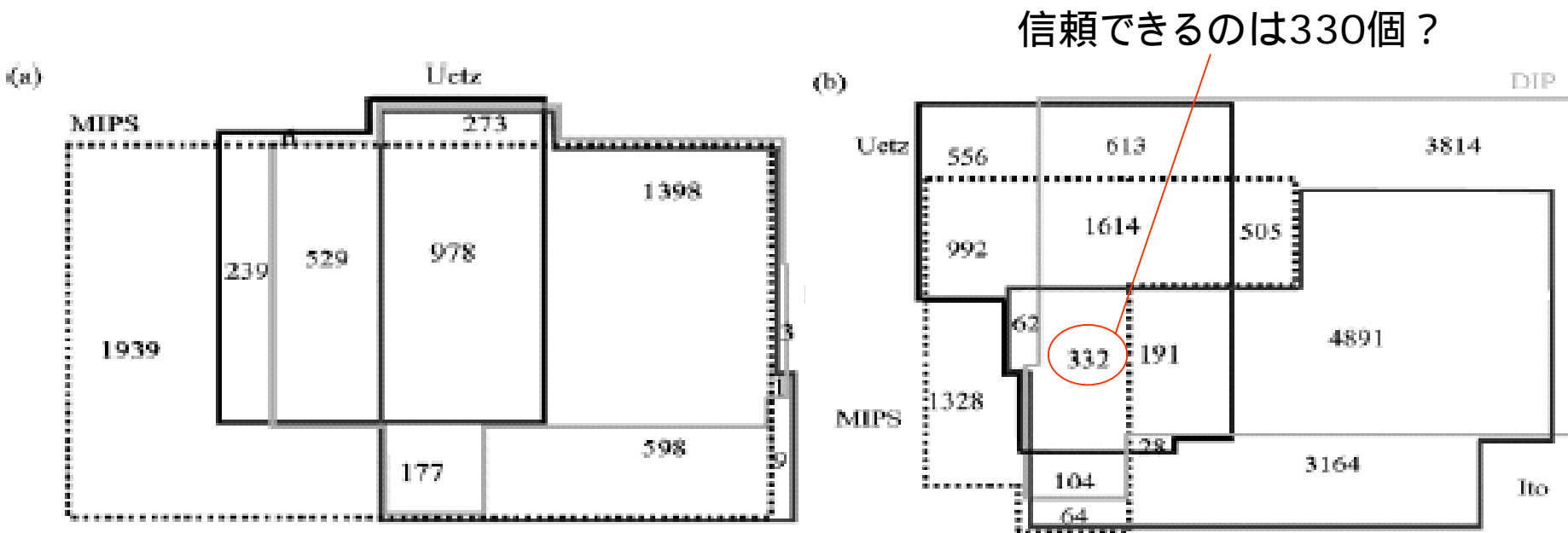
	UETZ	ITO-core	UETZ plus ITO-core	ITO-full
CC	50.1% (5.3%)	45.1% (5.8%)	48.4% (5.5%)	91.8% (0.9%)
PHE	41.8% (7.9%)	39.2% (8.2%)	40.9% (8.0%)	89.0% (1.5%)
SPO	50.1% (6.4%)	48.2% (6.6%)	49.5% (6.5%)	92.1% (1.0%)
average	47.3%	44.2%	46.3%	91.0%

Estimation of false positives in the whole-genome scans, assuming there is no other source for the difference to the MIPS database. Data is given as the mean and standard deviation resulting from the bootstrapping procedure. Abbreviations are the same as in Table 1.

Mrowka et al. *Genome Res.* 11(12)1971-1973, 2001.

2-hybridデータの9割以上がfalse positiveだという人までいる。

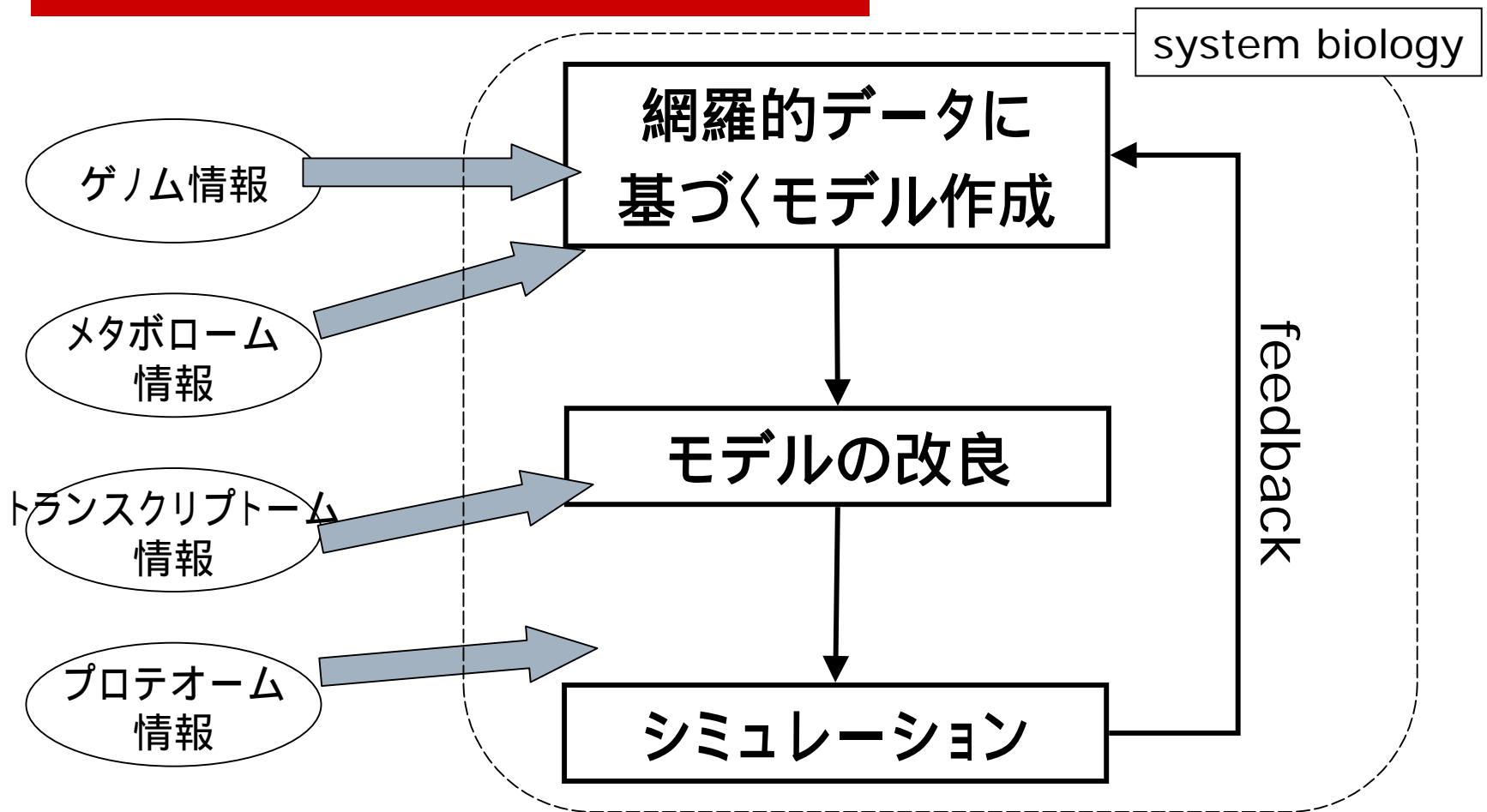
一致度



Relationship between the four studied databases (a) at the protein and at (b) the interaction level.

Yook et al. *Proteomics* 4, 928-942, 2004.

「使えない」データを「使う」には順序がある



ここまでのまとめ

- 「ゲノムがわかれば、全てがわかる。」は嘘。
 - ゲノムは生命の設計図ではなく、部品リスト。
 - インタラクションの部分が配線図にあたる。
- 初期モデルがきちんとしていなければ、トランスクリプトームやプロテオームは役に立たない。
- 発現量の変化ばかり追ったり、クラスタリングをしても役に立たない。

きちんとした初期モデルとは何か？

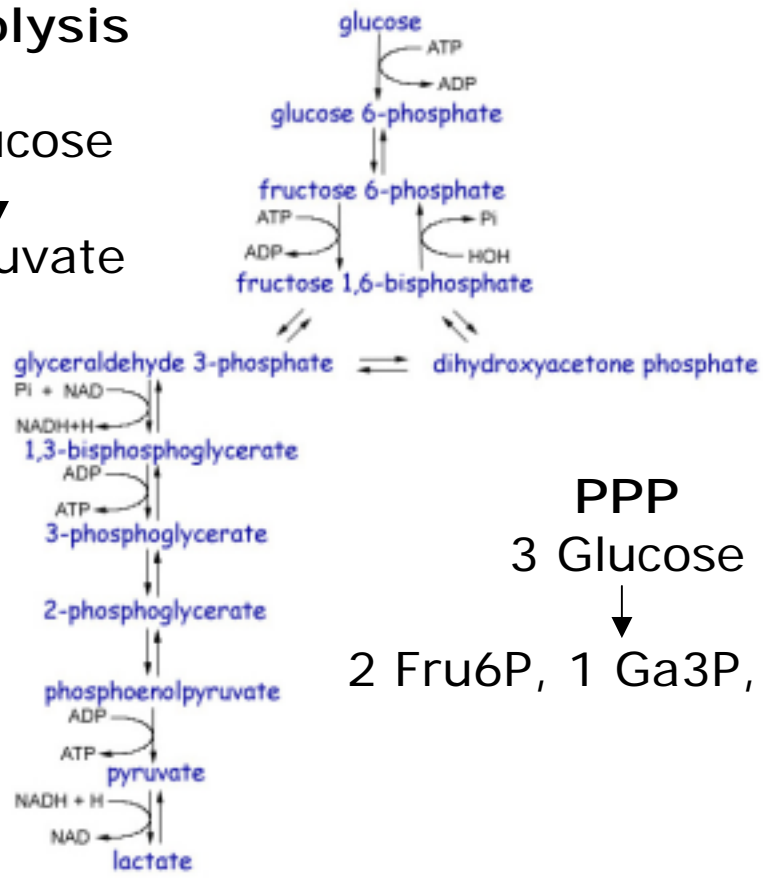
代謝で実現したいこと

- ある状態で、どの経路が動いているのかを知りたい。
 - 個々の生物種に注目した形の代謝マップで、生物種間の比較も視覚的にできるものが欲しい。
 - ある物質を与えられたら、その生合成経路や分解経路をマップの中でハイライトしてくれる機能が欲しい。
 - 遺伝子名や関連する経路を(クリックして別ページに行くのではなく)その場で表示、追加できる機能が欲しい。
-

おさらい： 解糖系とペントースリン酸経路

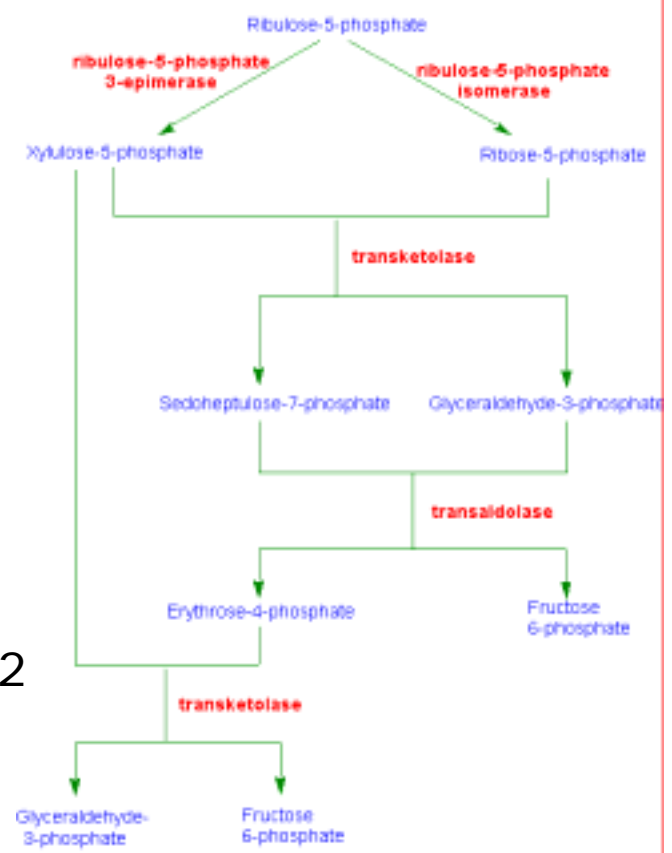
Glycolysis

1 Glucose
↓
2 Pyruvate



PPP
3 Glucose
↓
2 Fru6P, 1 Ga3P, 3CO₂

Non-Oxidative Stage of Pentose Phosphate Pathway



ストイキオメトリー行列

Glycolysis

G6P F6P

F6P 2 G3P

PPP

G6P CO₂, R5P

3 R5P 2 F6P, G3P

G3P Pyr

$$\begin{matrix} & \text{G6P} & \text{F6P} & \text{G3P} & \text{Pyr} & \text{R5P} \\ \left(\begin{array}{cccccc} -1 & 1 & & & & \\ & & -1 & 2 & & \\ -1 & & & & & 1 \\ & & 2 & 1 & & -3 \\ & & & -1 & 1 & \end{array} \right) \end{matrix}$$

各反応における物質の収支を、行列の形で表現したもの。
行： 化学反応、 列： 物質

行列計算による代謝経路の計算

Gaussの消去法(連立一次方程式の解法)を適用。
どの代謝物質からどの物質が合成できるかがわかる。

- 具体的な経路を求めるには、変数(物質)を消去してゆく「順番」を覚えておく必要がある。
- 経路の数は組み合わせ的に大きくなるので、全ての経路を記録することは通常不可能。

Ref: Schuster et al. *Nat. Biotechnol.* 18:326-332, 2000
“Metabolic Engineering”, Stephanopoulos *et al.*,
Academic Press(1998)

行列計算による解法

Table 2. Elementary modes of the combined glycolysis and PPP system^a

Mode	Overall conversion	Participating steps
(1)	$G6P + 3 ADP + 2 Pi + 2 NAD \rightarrow 3 ATP + 2 NADH + 2 Pyr$	{Pgi Pfk Fba TpiA 2Gap 2Pgi 2Gpm 2Eno 2Pyk}
(2)	$G6P + 2 ADP + Pi + NAD + 6 NADP \rightarrow 2 ATP + NADH + 6 NADPH + 3 CO_2 + Pyr$	{-2Pgi Gap 3Zwf 3Pgi 3Gnd Rpi 2Rpe Tktl Tal Tktll Pgi Gpm Eno Pyk}
(3)	$3 G6P + 8 ADP + 5 Pi + 5 NAD + 6 NADP \rightarrow 8 ATP + 5 NADH + 6 NADPH + 3 CO_2 + 5 Pyr$	{2Pfk 2Fba 2TpiA 5Gap 3Zwf 3Pgi 3Gnd Rpi 2Rpe Tktl Tal Tktll 5Pgi 5Gpm 5Eno 5Pyk}
(4)	$G6P + 2 NADP \rightarrow 2 NADPH + CO_2 + R5Pex$	{Zwf Pgi Gnd Rpi Prs_DeoB}
(5)	$5 G6P + ATP \rightarrow ADP + 6 R5Pex$	{5Pgi Pfk Fba TpiA 4Rpi -4Rpe -2Tktl -2Tal -2Tktll 6Prs_DeoB}
(6)	$G6P + 12 NADP \rightarrow 12 NADPH + Pi + 6 CO_2$	{-5Pgi -Fba -TpiA 6Zwf 6Pgi 6Gnd 2Rpi 4Rpe 2Tktl 2Tal 2Tktll Fbp}
(7)	$ATP \rightarrow ADP + Pi$	{Pfk Fbp}
(8)	$2 ADP + Pi + NAD + 3 R5Pex \rightarrow 2 ATP + NADH + 2 G6P + Pyr$	{-2Pgi Gap -2Rpi 2Rpe Tktl Tal Tktll Pgi Gpm Eno Pyk -3Prs_DeoB}
(9)	$2 ADP + Pi + NAD + 4 NADP + R5Pex \rightarrow 2 CO_2 + 2 ATP + NADH + 4 NADPH + Pyr$	{-2Pgi Gap 2Zwf 2Pgi 2Gnd 2Rpe Tktl Tal Tktll Pgi Gpm Eno Pyk - Prs_DeoB}
(10)	$8 ADP + 5 Pi + 5 NAD + 3 R5Pex \rightarrow 8 ATP + 5 NADH + 5 Pyr$	{2Pfk 2Fba 2TpiA 5Gap -2Rpi 2Rpe Tktl Tal Tktll 5Pgi 5Gpm 5Eno 5Pyk - 3Prs_DeoB}
(11)	$8 ADP + 5 Pi + 5 NAD + 4 NADP + 2 G6P + R5Pex \rightarrow 2 CO_2 + 8 ATP + 5 NADH + 4 NADPH + 5 Pyr$	{2Pfk 2Fba 2TpiA 5Gap 2Zwf 2Pgi 2Gnd 2Rpe Tktl Tal Tktll 5Pgi 5Gpm 5Eno 5Pyk -Prs_DeoB}
(12)	$6 R5Pex \rightarrow 5 G6P + Pi$	{-5Pgi Fbp -Fba -TpiA -4Rpi 4Rpe 2Tktl 2Tal 2Tktll -6Prs_DeoB}
(13)	$8 NADP + 2 R5Pex \rightarrow 4 CO_2 + 8 NADPH + G6P + Pi$	{-5Pgi Fbp -Fba -TpiA 4Zwf 4Pgi 4Gnd 4Rpe 2Tktl 2Tal 2Tktll - Prs_DeoB}

^aThe enzymes involved in the elementary modes are given in braces, weighted with their fractional flux. Negative values indicate the reaction is used in the reverse sense. When the reaction Prs_DeoB is considered irreversible, only the upper seven modes are obtained.

解糖系について消去法の全パターンをつくした計算結果。
Schuster et al. *Nat. Biotechnol.* 18:326-332, 2000

線形計画法への発展

- 行列を転置して、流量をかけると収支が出る。

G6P	F6P	G3P	Pyr	R5P		
$\begin{pmatrix} -1 & 1 & & & \\ & -1 & 2 & & \\ -1 & & & & 1 \\ & 2 & 1 & & -3 \\ & & -1 & 1 & \end{pmatrix}$						
	S					
						\Rightarrow
	G6P	F6P	G3P	Pyr	R5P	
						$\begin{pmatrix} -1 & & & & \\ 1 & -1 & & & \\ & 2 & 1 & & -1 \\ & & & 1 & \\ & & & & 1 & -3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix}$
					S^T 流量	

定常状態において、各物質の量は一定だから、 $S^T \cdot v = 0$

制約に基づく流量の解析

- 物質収支を行列表現
(S : 列が物質、 v : 流量)

- 線形計画法を適用
(e は細胞が最適成長するベクトル = oracle であり、 ev により v を e に近づける。)

- 流量には制約を設ける

under $S^T \cdot v = 0$

find $\max \sum e_i v_i$

where

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

求められた流量 v は実測値と一致

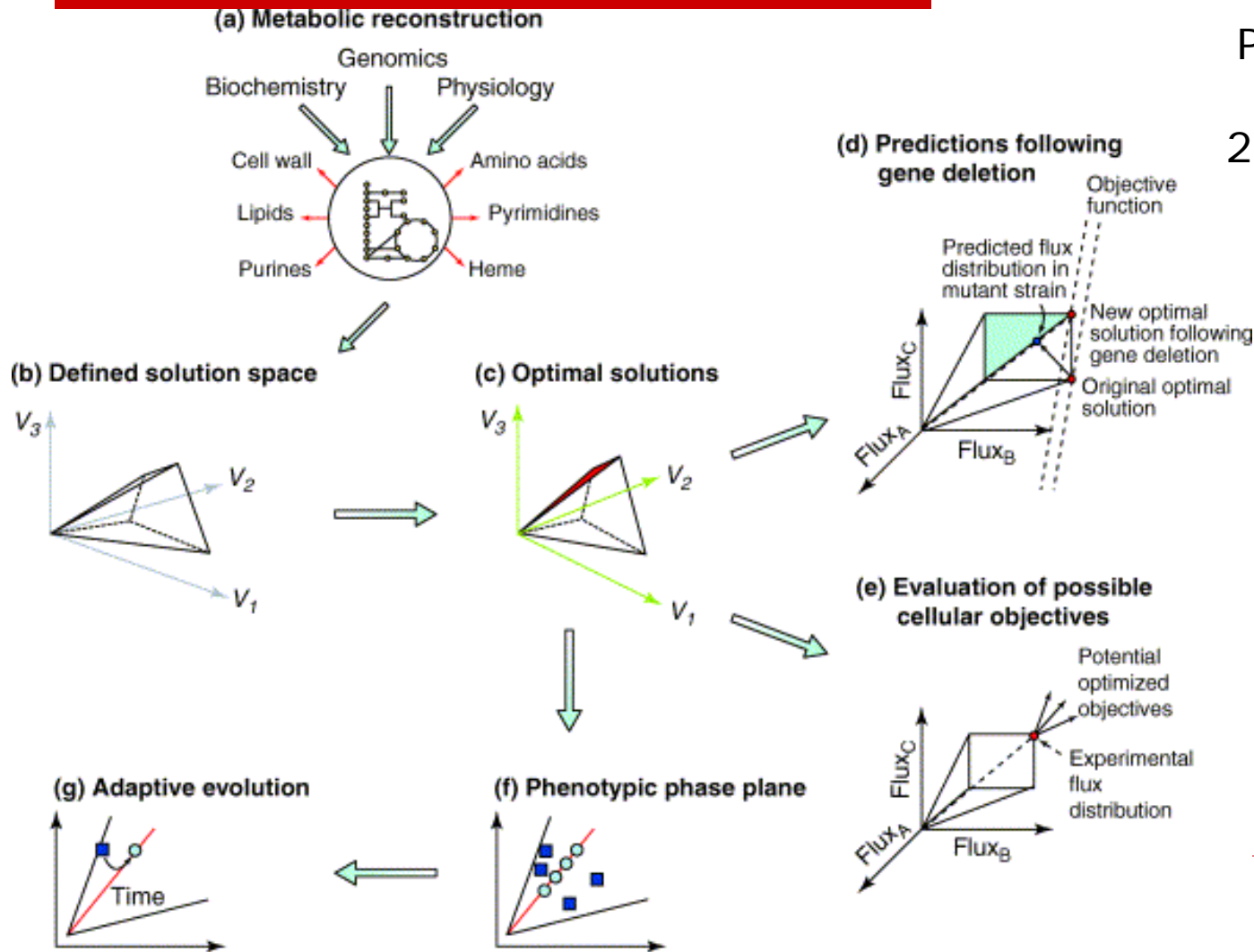
Ref: Edwards JS et al. *Nat. Biotechnol.* 19:125-130, 2001

線形モデルの特徴

- 代謝の定常状態を仮定したモデルである。
- 物質の収支を制約として、各酵素反応がどのくらい進んでいるかを最適化する。
- 変数の数 = 酵素反応の数であるため、多次元
(ネットワークの簡略化により次元はだいぶ下がる)

多次元の実最適化をおこなう場合、「ボトルネック」となる変数は少ないことが多い。

Phenotype Phase Plane analysis Extreme Pathway analysis



Price et al. *Trends Biotechnol.*
21:162-169, 2003.

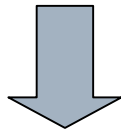
ここまでのまとめ

- 線形計画法により、定常状態における代謝流量を最適化できる。
- 実測できる値は、酸素の消費量、グルコースの消費量など、数が少ない。
- 代謝にはよく使われるスーパーパスウェイが存在する。

では代謝の流量から、個々のパスウェイを取り出すにはどうするか。

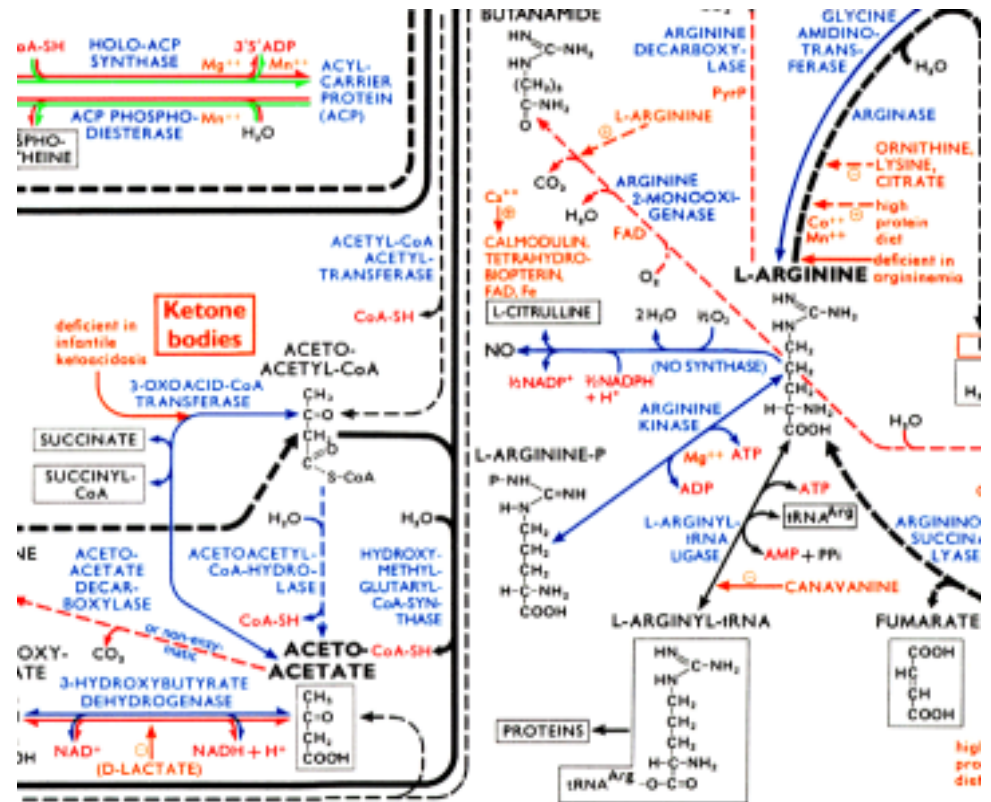
代謝と電車の違い

代謝は、各反応ステップで構造が組み換わる。

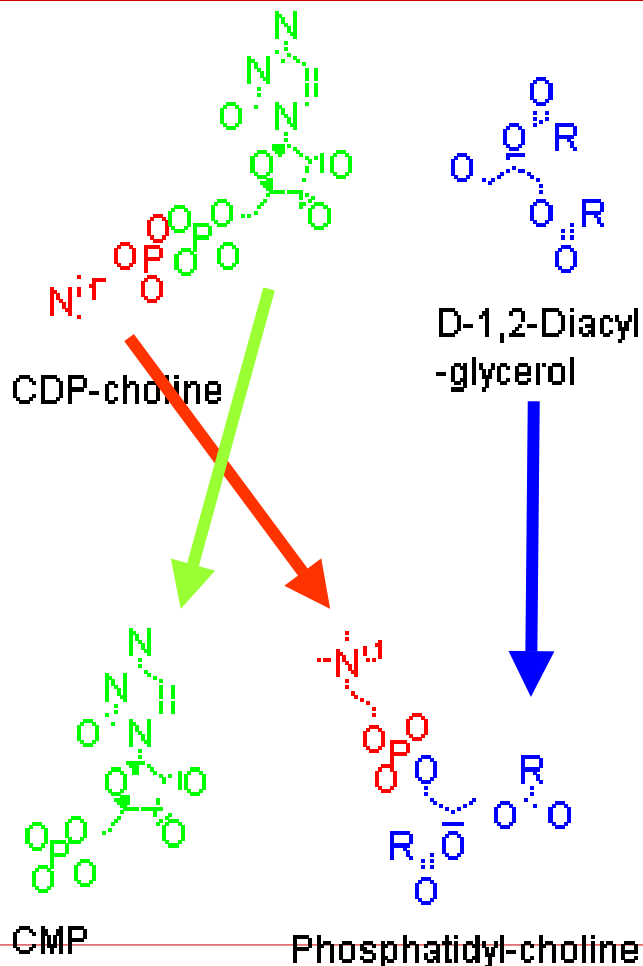


代謝を鉄道にたとえると、駅に停まるたびに隣のホームの列車と車両を組み換えてしまうシステム。

これを表現するには？



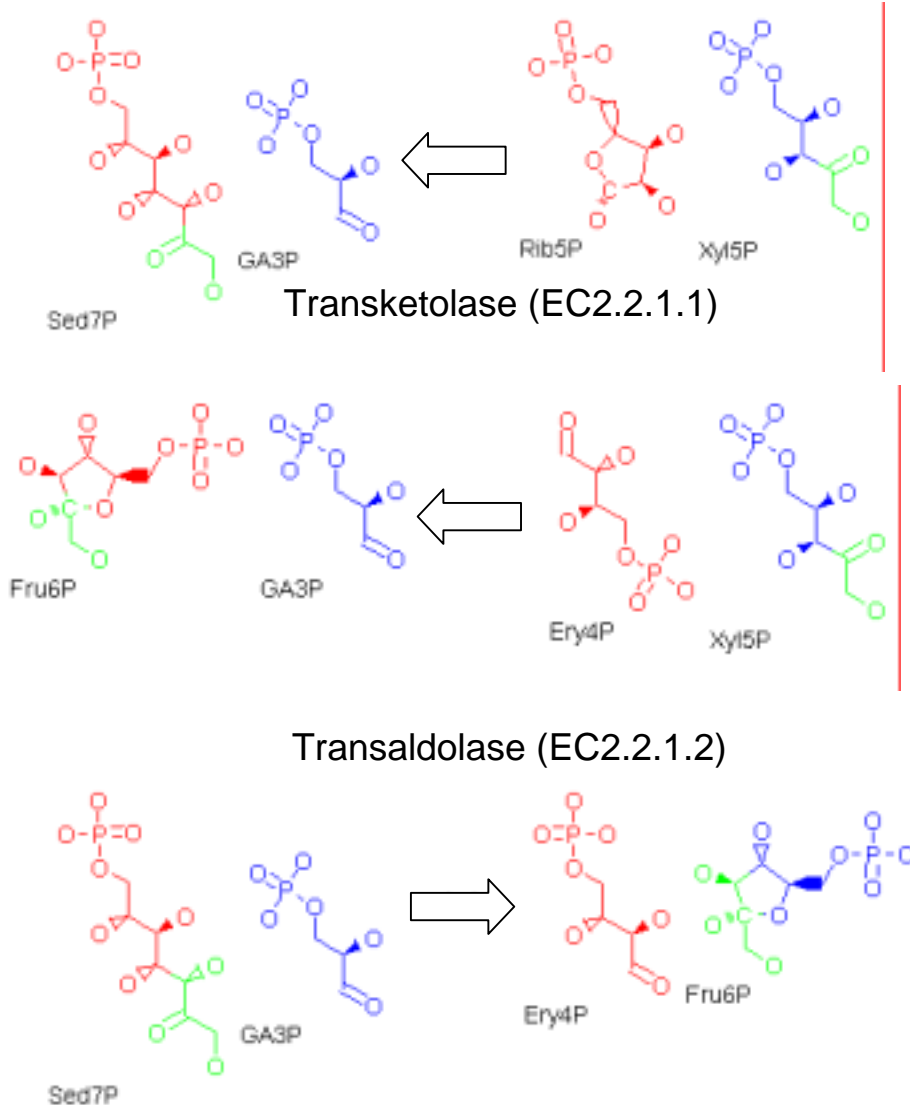
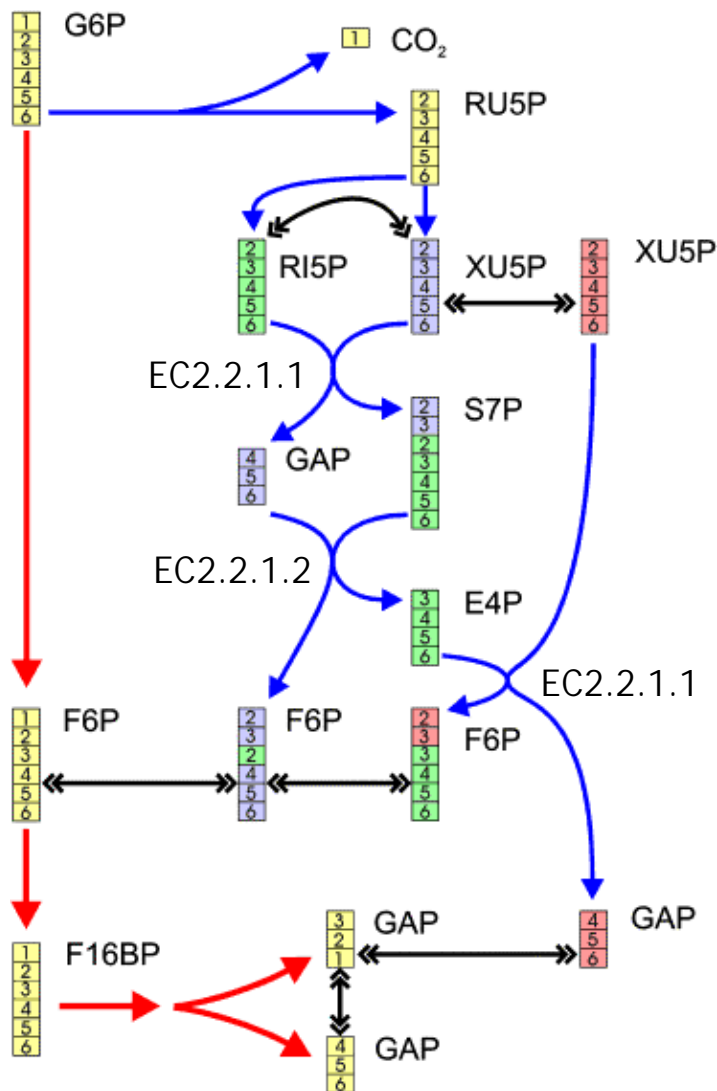
各反応を「部分構造の対応」に分解する必要



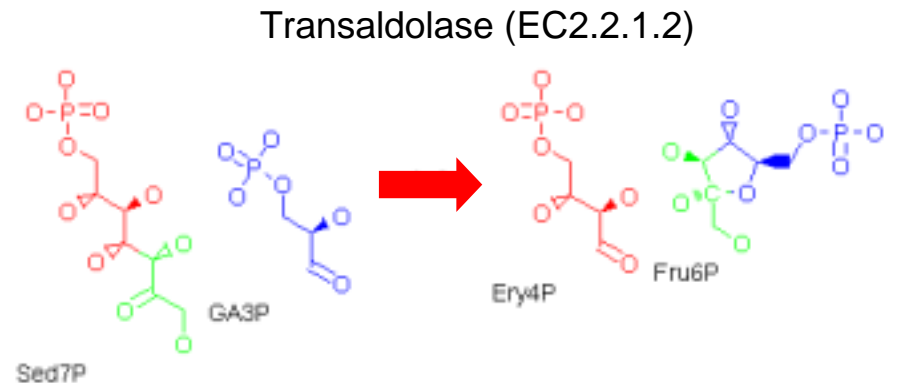
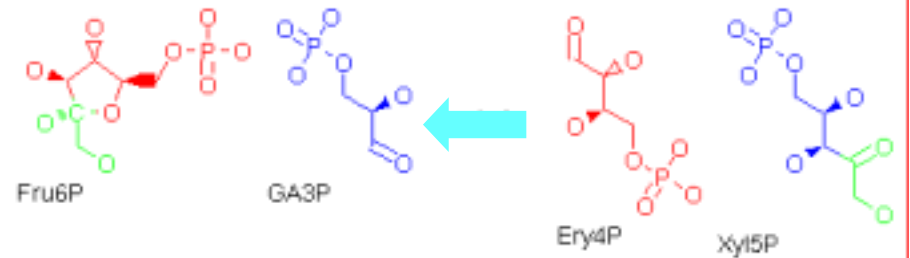
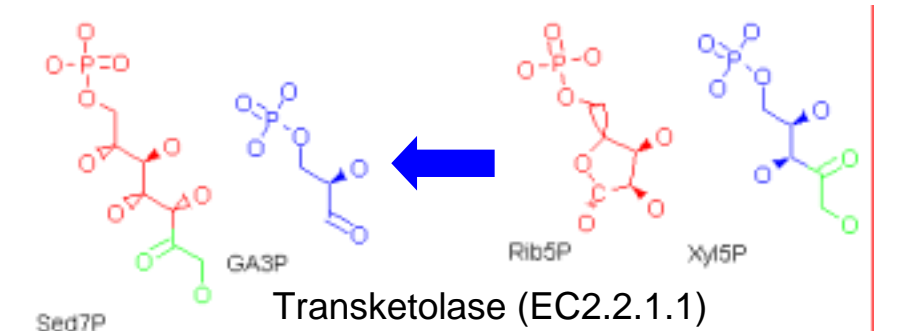
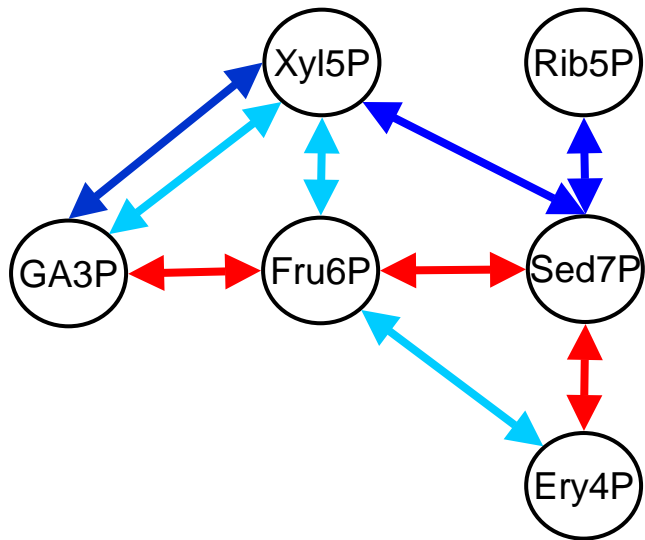
原子と原子の対応にまで分解してしまえば、質量保存則から、1対1の関係になっている。
(計算機で扱いやすい。)

代謝マップに出てくる反応を、原子と原子の対応にまで分解するとどうなるか？

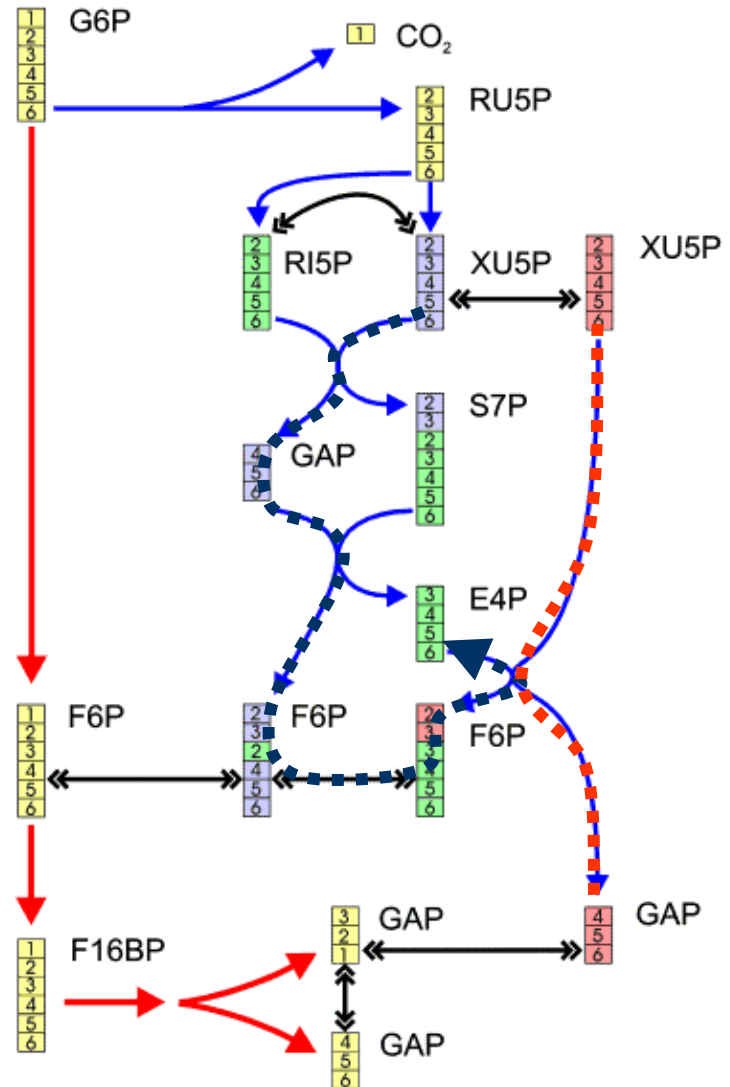
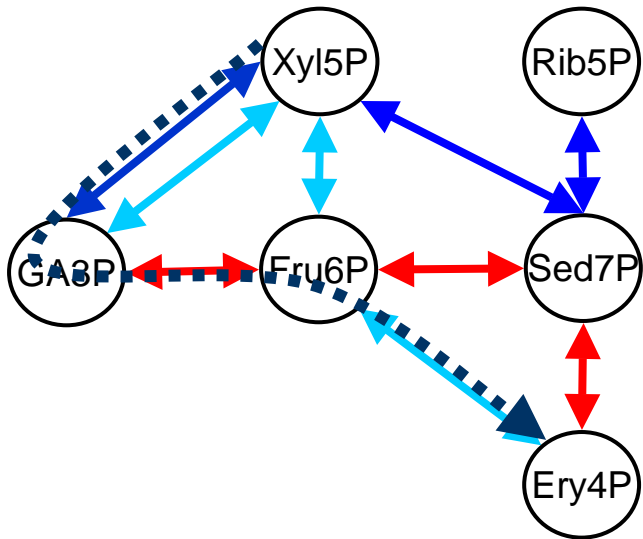
ペントース・リン酸経路 原子レベルのマッピング



ペントース・リン酸経路 原子レベルのマッピング

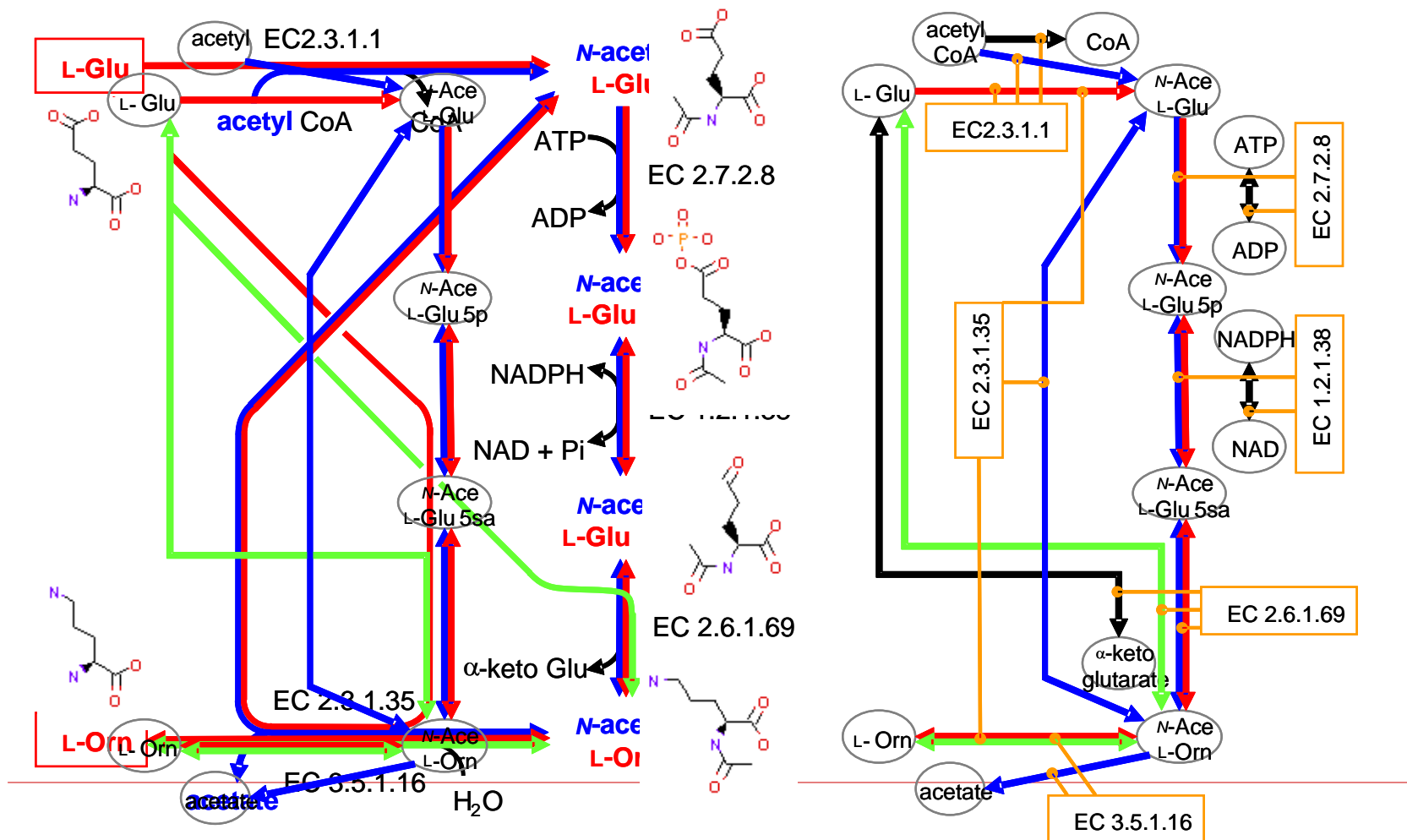


ペントース・リン酸経路 原子レベルのマッピング



Arita, M. *Genome Research* **13**(11), 2455-2466, 2003.

代謝の表現



ネットワークのスケールフリー性

代謝ネットワークはスケールフリーだといわれる。

1. ほとんどの頂点は接続数 (degree) が低く、べき分布をなす。(scale-freeness)
 2. すべての頂点を結ぶ最短距離の平均値 (AL ... average path length) は非常に小さく、ランダムグラフのそれに近い。
 3. ハブ (hub) と呼ばれる接続数の大きな頂点がクラスタを形成し、ネットワークを覆う。
-

べき分布とは？

□ べき分布 $p = Cx^{-\lambda}$

□ 両対数プロットをとると、標本点が直線上に並ぶ。

$$\log p = \log C - \lambda \log x$$

□ この傾きは x 軸のスケールによらない。

($x = Kx$ としても傾きは不変。)

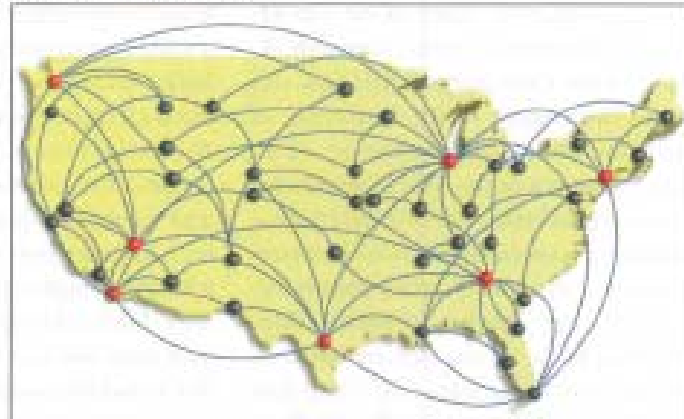
$$\log p = \log C - \lambda \log K - \lambda \log x$$

自然界のネットワーク

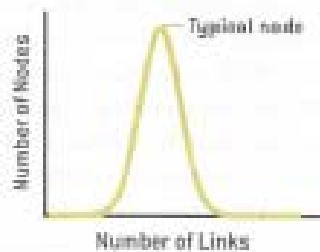
Random Network



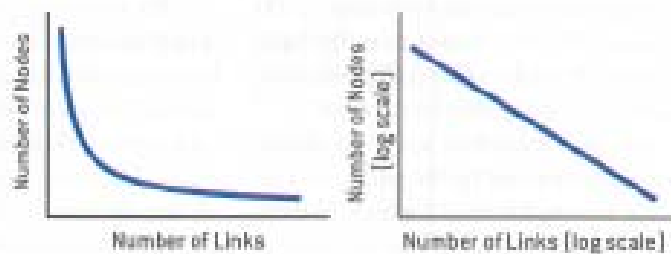
Scale-Free Network



Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages



自然界のネットワークは少ないステップ数で渡り歩ける

TABLE I. The general characteristics of several real networks. For each network we have indicated the number of nodes, the average degree $\langle k \rangle$, the average path length ℓ , and the clustering coefficient C . For a comparison we have included the average path length ℓ_{rand} and clustering coefficient C_{rand} of a random graph of the same size and average degree. The numbers in the last column are keyed to the symbols in Figs. 8 and 9.

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook <i>et al.</i> , 2001a, Pastor-Satorras <i>et al.</i> , 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási <i>et al.</i> , 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási <i>et al.</i> , 2001	9
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460 902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook <i>et al.</i> , 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

サイズ ↑

↑ ステップ数

スモールワールド

- 自然のネットワークは殆どsmall worldらしい。
 - 送電線のネットワーク
 - 線虫の神経網
 - ハリウッド映画俳優の共演関係
 - たんぱく質の相互作用ネットワーク
 - 代謝のネットワーク

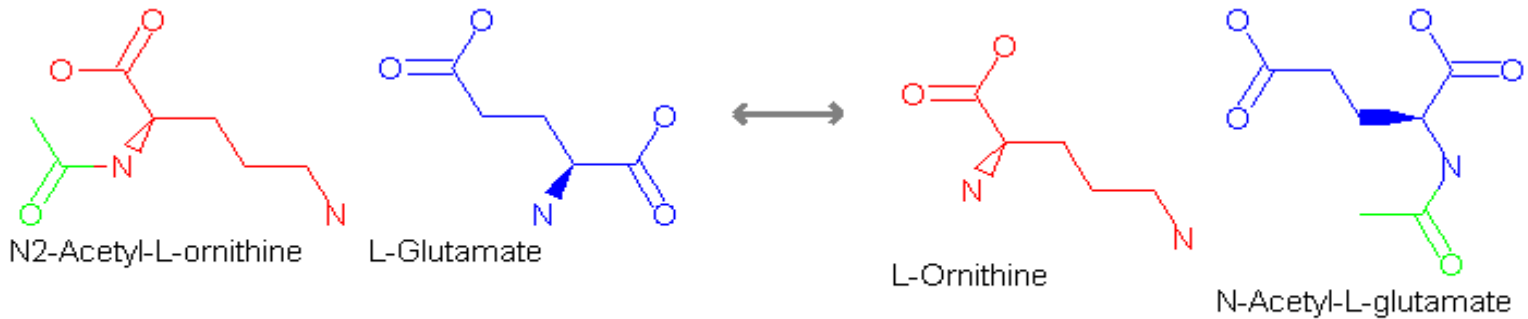
Watts DJ, Strogatz SH “Collective dynamics of 'small-world' networks”, *Nature* 393(6684):440-442, 1998

代謝ネットワークに関する従来 of 成果

Small world性に関して引用される代表論文

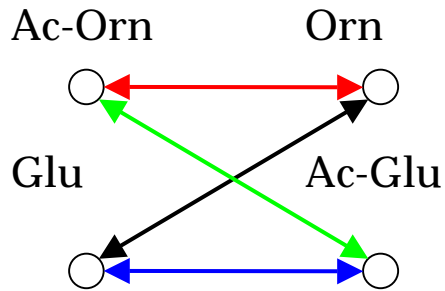
1. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabási, A.-L. “The large-scale organization of metabolic networks” *Nature* **407**: 651-654, 2000.
 2. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. and Barabási, A.-L. “Hierarchical Organization of Modularity in Metabolic Networks” *Science* **297**: 1551-1555, 2002.
 3. Fell, A.D. and Wagner, A. “The small world of metabolism” *Nat Biotechnol* **18**: 1121-1122, 2000.
 4. Ma, H. and Zeng, A-P. “Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms” *Bioinformatics* **19(2)**:270-277, 2003.
-

ネットワークのグラフ変換

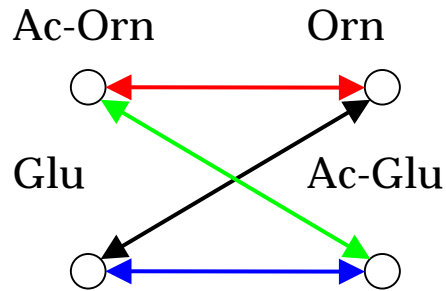


EC 2.3.1.35: glutamate N-acetyltransferase

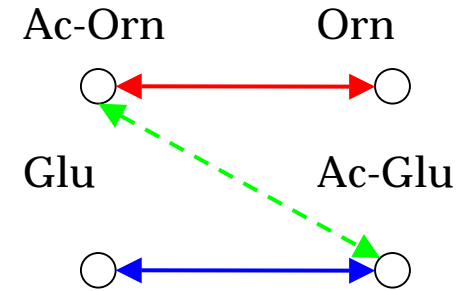
Jeong *et al.*



Wagner and Fell



Ma and Zeng



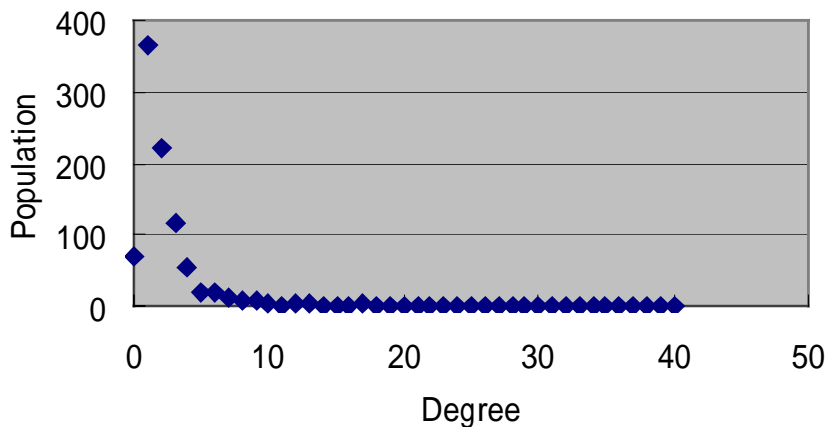
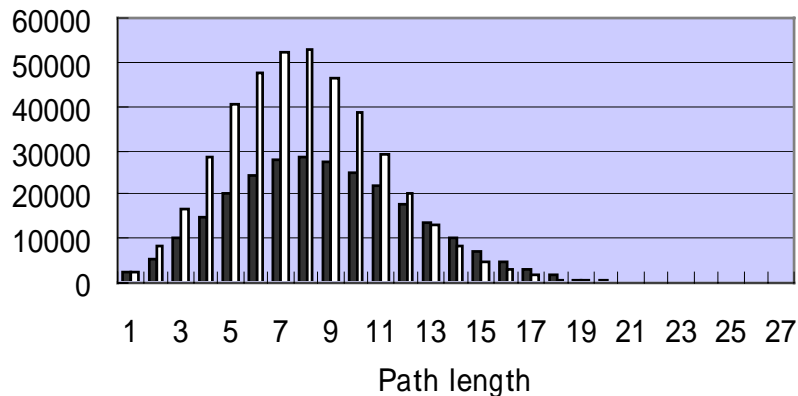
グラフ変換のまとめ

- Jeong *et al.*
同じ反応に「関与する」という関係をグラフ化
代謝物質間における、シグナル伝達をモデル化
 - Wagner & Fell
補酵素を除いて反応に「関与する」関係をグラフ化
代謝経路における、摂動(perturbation)の伝わりをモデル化
 - Ma & Zeng
反応における物質の対応を手作業でグラフ化
いわゆる代謝経路のモデル化
-

各研究報告の結果

	Jeong et al. (<i>Nature</i> 407: 651)	Wagner and Fell (<i>Nat Biotech</i> 18: 1121)	Ma and Zeng (<i>Bioinform</i> 19 :270)
Data	43 bacteria	<i>E.coli</i>	80 species
Graph	directed	undirected	directed
Top 10 Hubs	H ₂ O ADP orthophosphate ATP L-glutamate NADP+ PP NAD ⁺ NADPH ⁺ NADH	L-glutamate pyruvate coenzyme A 2-oxo-glutarate L-glutamine L-aspartate acetyl CoA phosphoribosyl PP tetrahydrofolate succinate	glycerate 3P D-ribose 5P acetyl CoA pyruvate D-Xylulose 5P D-Fructose 6P 5P-D-ribose 1PP L-glutamate propanoyl CoA
AL <i>E.coli</i>	3.2	3.8	8.2

代謝はスモールワールドではない



- 代謝経路の長さの統計を取ると、平均は8になる。
- これはNatureの論文等が3.3と計算した値と、同等のネットワーク。
- ネットワークのハブは、
CO₂, Pyruvate, Acetyl CoA, ATP, D-Glucose, L-Glutamate, D-Galactose ...

まとめ

- 代謝は原子レベルで電子化しないと意味がない。
 - 代謝ネットワーク = スモールワールド ではなく、その特徴はネットワークの見方に依存。
 - パスウェイのネットワーク ... 制約の多い経路
 - 情報伝達のネットワーク ... スケールフリー？
 - スケールフリー性は特別な性質ではない。
 - 他の網羅的解析におけるスケールフリー性の解釈は次回。
-