Multiscale Molecular Systems Biology: Reconstruction and Model Optimization

Dr. Ronan M.T. Fleming,
Systems Biochemistry Group,
Luxembourg Centre for Systems Biomedicine,
University of Luxembourg.

Friday, August 16, 2013
Interagency Modeling and Analysis Group Webinar

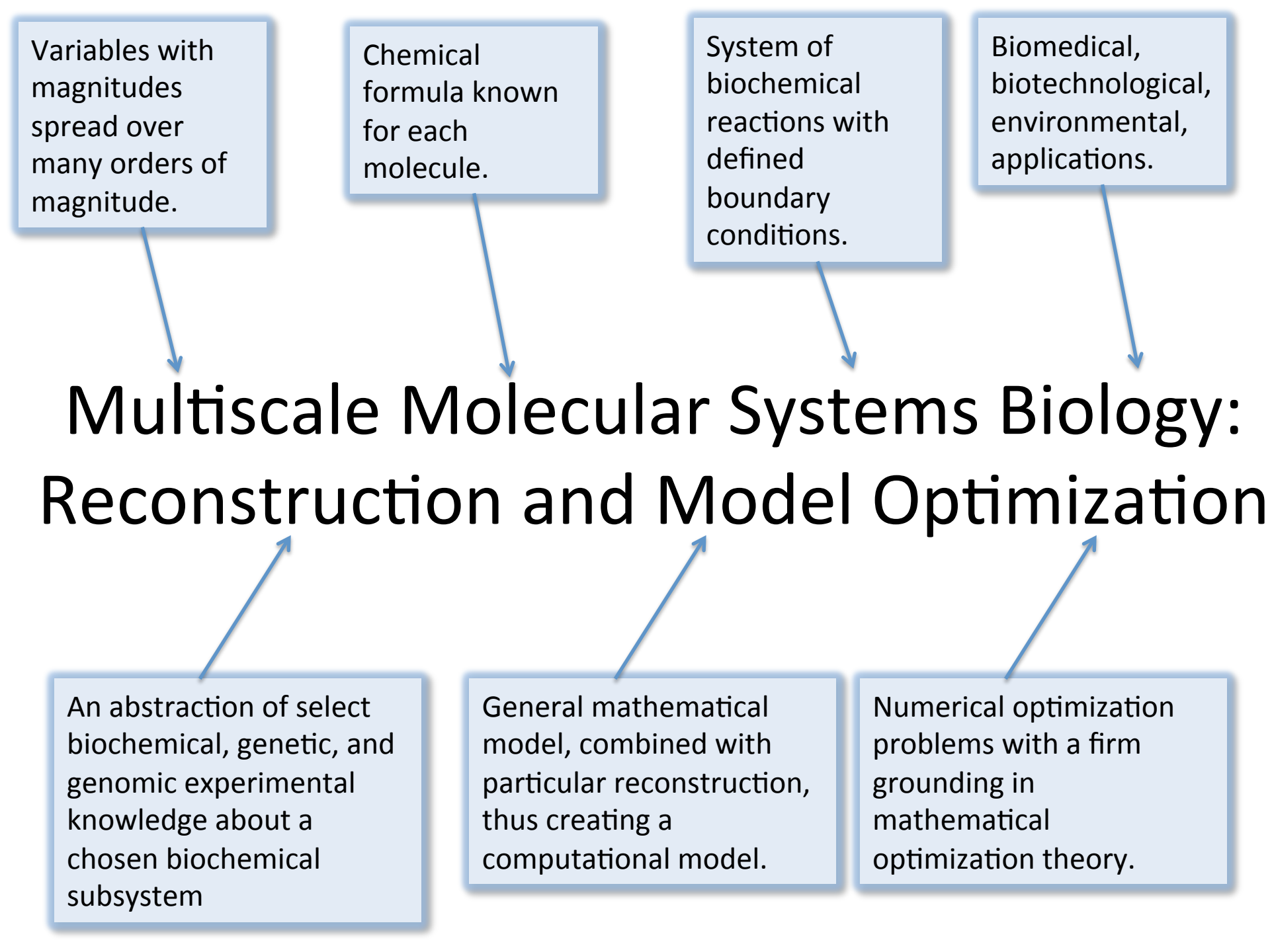
Multiscale Systems Biology Collaboration

- Molecular Systems Physiology Group
  - Ines Thiele, Luxembourg Centre for Systems Biomedicine.

- Systems Biology Research Group
  - Bernhard Palsson, University of California, San Diego.

- Systems Optimization Laboratory
  - Michael Saunders, Stanford University.

- Systems Biochemistry Group
  - Ronan Fleming, Luxembourg Centre for Systems Biomedicine.


Variables with magnitudes spread over many orders of magnitude.
Chemical formula known for each molecule.
System of biochemical reactions with defined boundary conditions.
Biomedical, biotechnological, environmental, applications.
Multiscale Molecular Systems Biology: Reconstruction and Model Optimization
An abstraction of select biochemical, genetic, and genomic experimental knowledge about a chosen biochemical subsystem
General mathematical model, combined with particular reconstruction, thus creating a computational model.
Numerical optimization problems with a firm grounding in mathematical optimization theory.

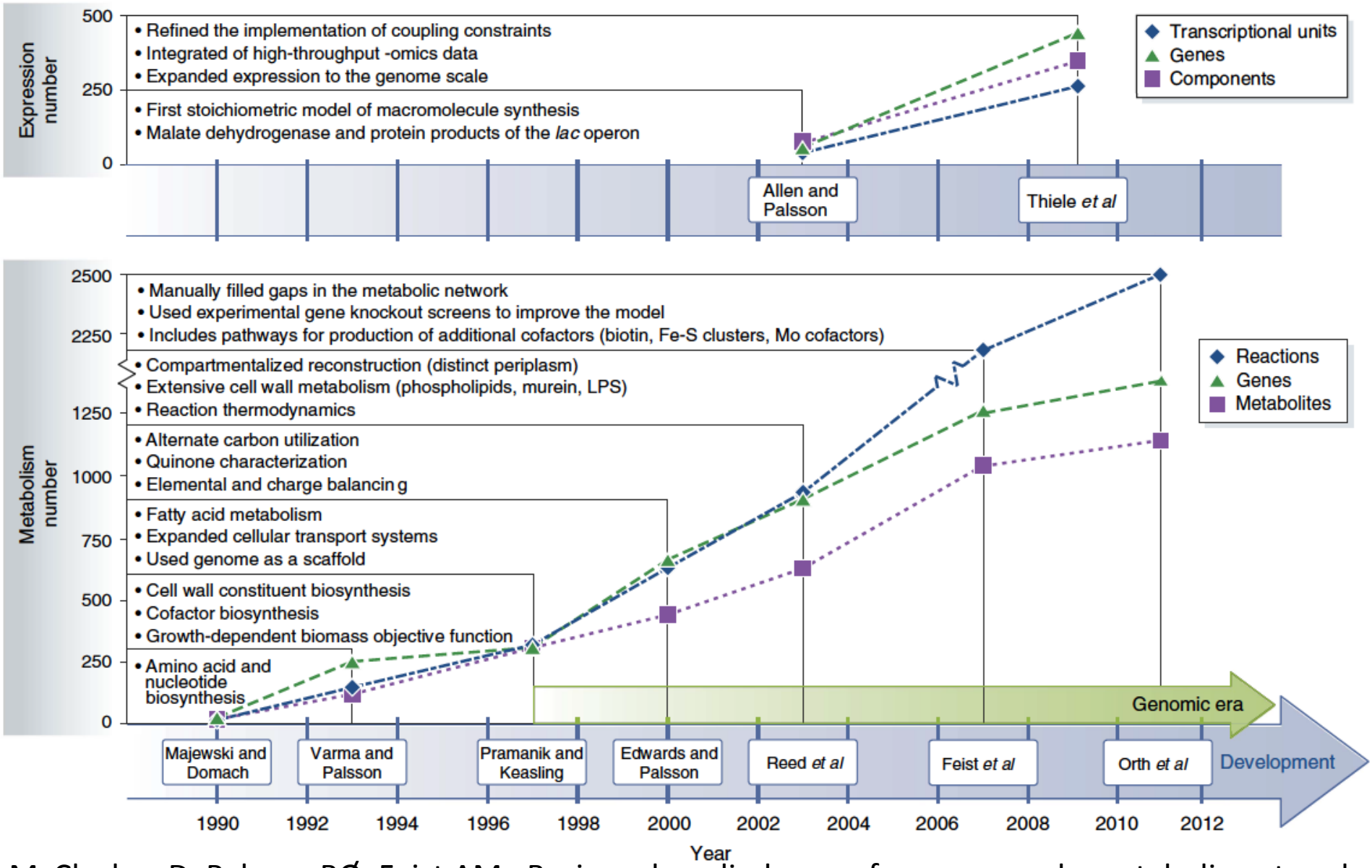
Increasing the comprehensiveness of genome scale computational models ….

- Increasing size
  - e.g. single microbe versus whole microbial community
    - 1 microbial species ~ 1e3 reactions
    - 1 community (1000 species) ~ 1e6 reactions
- Increasing ratio of fastest to slowest timescale
  - e.g. genome scale metabolic model versus integrated model of metabolism and macromolecular synthesis
    - Metabolic reactions
    - Macromolecular synthesis reactions
- Increased simulation fidelity
  - e.g. mass conservation alone, versus mass conservation, energy conservation, second law of thermodynamics, reaction kinetics, etc.

… leads to a mathematical and numerical optimization challenge:

- Large scale numerical optimization
  - Reduce computational complexity of algorithms to solve optimization problems
- Multiscale numerical optimization
  - Standard optimization software ideal for O(1) variables
- Mathematical formulation
  - Biochemical function is an inherently nonlinear process
  - How to formulae a mathematical modeling problem in a form amenable to a polynomial time algorithm

History of the Multiscale Systems Biology Collaboration

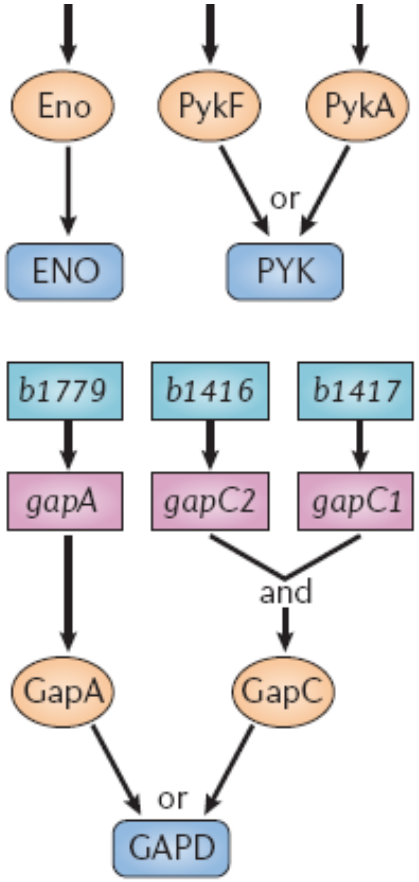


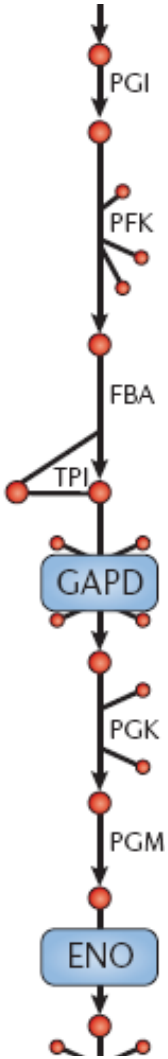
McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. Mol Syst Biol. 9:661, 2013.

Reconstruction of reaction stoichiometry

FBA	[c]FDP ↔ DHAP + G3P	*fbaA, fbaB*
TPI	[c]DHAP ↔ G3P	*tpiA*
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	*gapA, gapC1, gapC2*
PGK	[c]13DPG + ADP ↔ 3PG + ATP	*pgk*
PGM	[c]3PG ↔ 2PG	*gpmA, gpmB*
ENO	[c]2PG ↔ H_2O + PEP	*eno*
PYK	[c]ADP + H + PEP → ATP + PYR	*pykA, pykF*

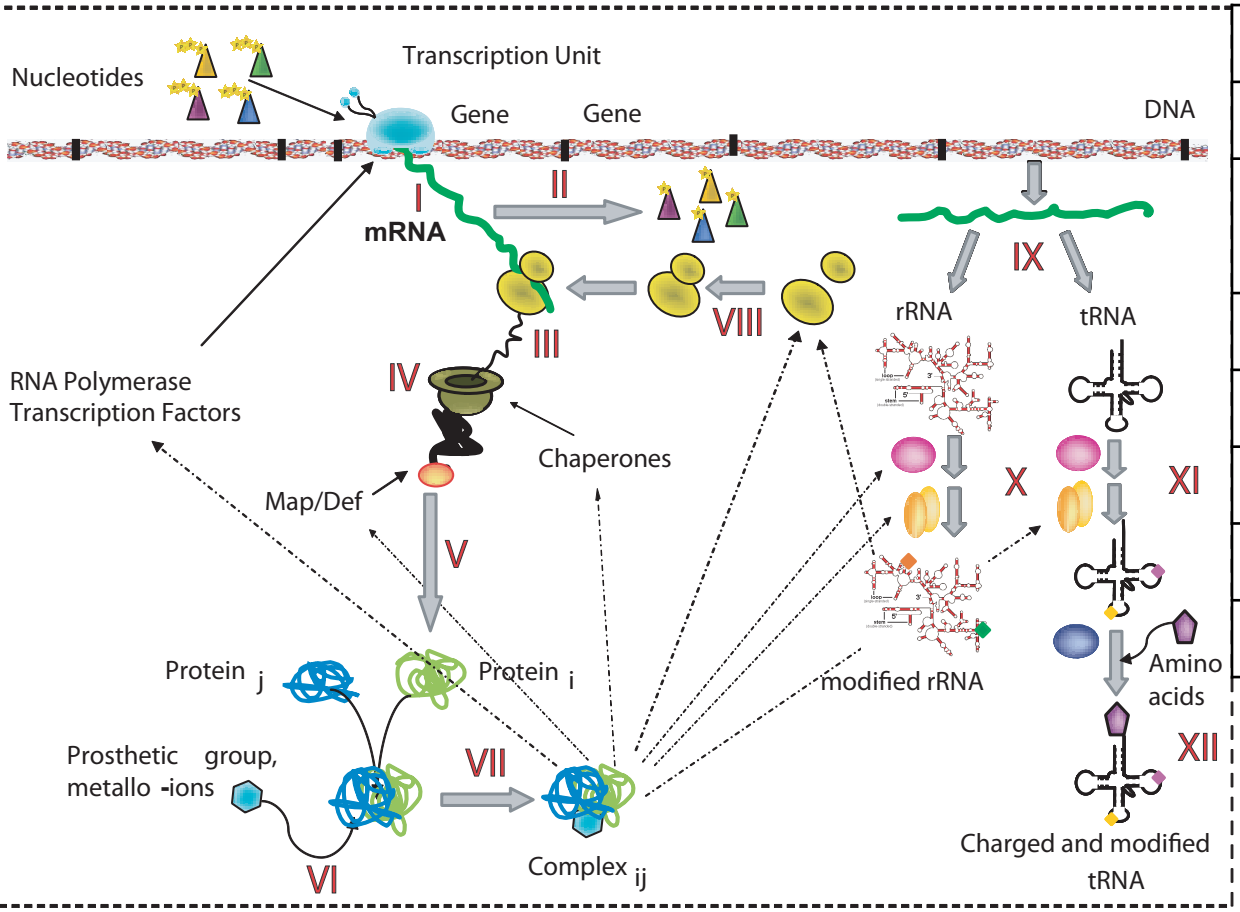
ATP	−1	0	−1	0	0	0	1	0	0	1
GLC	−1	0	0	0	0	0	0	0	0	0
ADP	1	0	1	0	0	0	−1	0	0	−1
G6P	1	−1	0	0	0	0	0	0	0	0
H	1	0	1	0	0	1	0	0	0	−1
F6P	0	1	−1	0	0	0	0	0	0	0
FDP	0	0	1	−1	0	0	0	0	0	0
DHAP	0	0	0	1	−1	0	0	0	0	0
G3P	0	0	0	1	1	−1	0	0	0	0
NAD	0	0	0	0	0	−1	0	0	0	0
PI	0	0	0	0	0	−1	0	0	0	0
13DPG	0	0	0	0	0	1	−1	0	0	0
NADH	0	0	0	0	0	1	0	0	0	0
3PG	0	0	0	0	0	0	1	−1	0	0





Stoichiometric Matrix (denoted S)

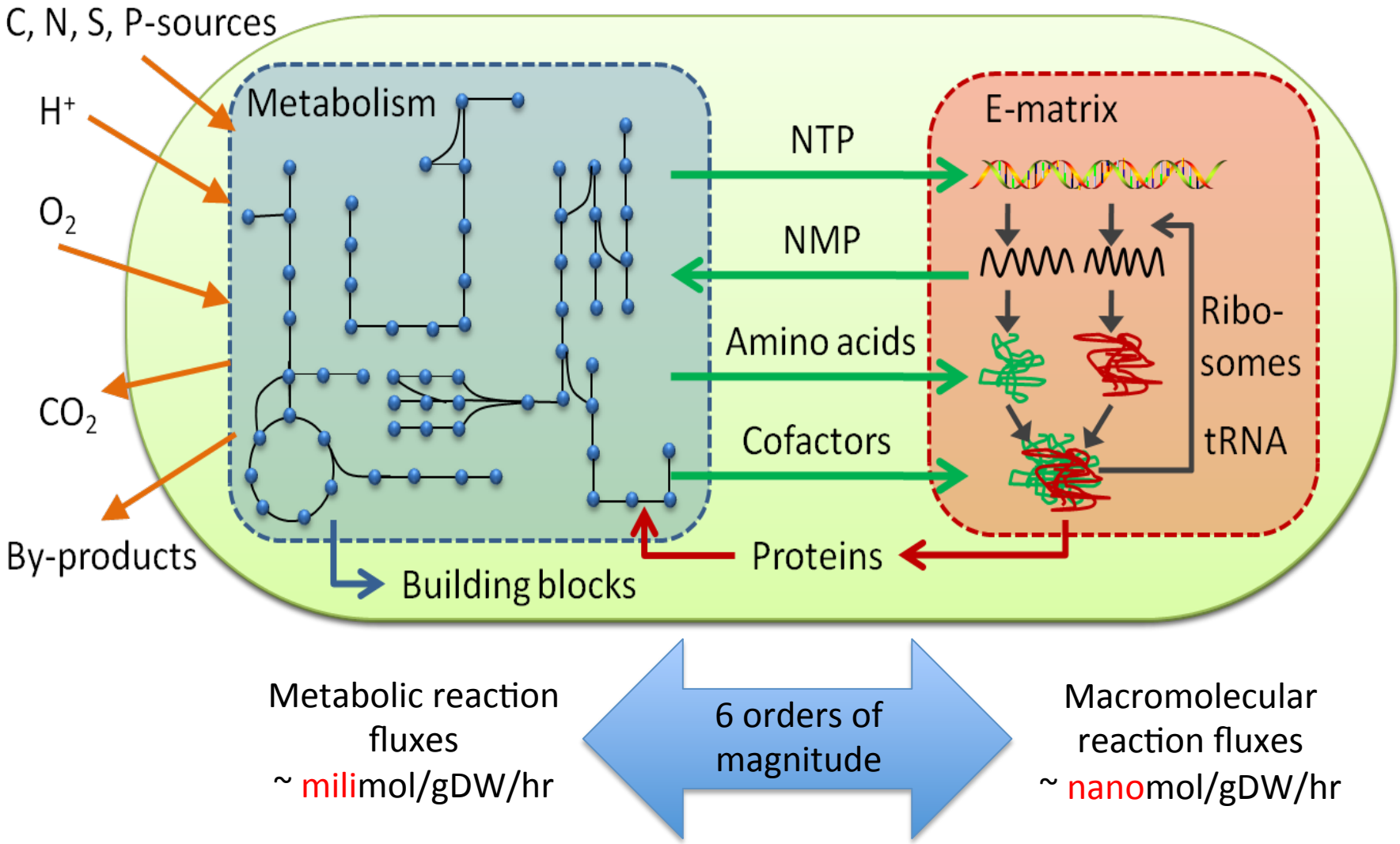
Reconstruction of macromolecular synthesis machinery



# transcription units	249
# genes	423
tRNA/rRNA/misc RNA	86/22/1
proteins (w/o genes)	228 (34)
# subsystems	27
# reactions	13,694
# components	11,991
# references	+500

Thiele I, Jamshidi N, Fleming RMT, Palsson BØ. Genome-scale reconstruction of Escherichia coli's transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS computational biology. 5(3):e1000312., 2009.

Integration of metabolism with macromolecular synthesis

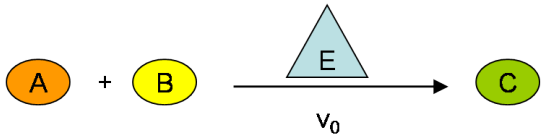


Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. PLoS One. 7(9):e45635, 2012.

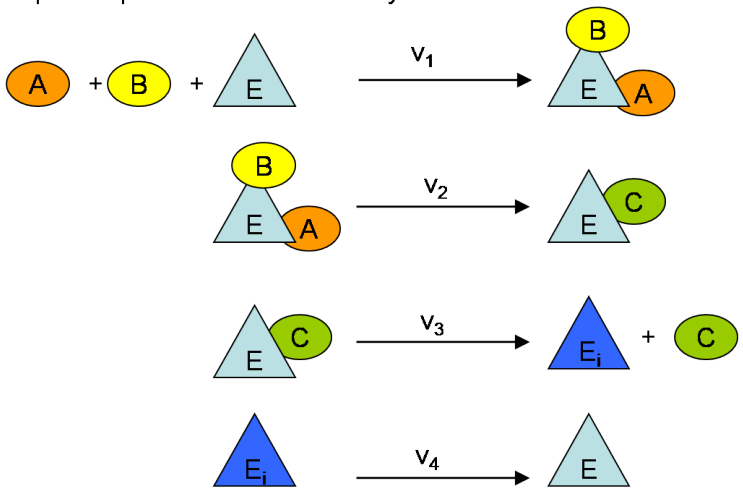
Conversion of integrated reconstruction of metabolism and macromolecular synthesis into a computational model

Canonical steady-state modeling

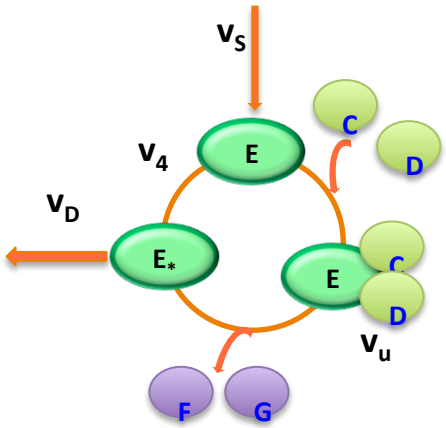
Implicit representation of an enzymatic reaction:

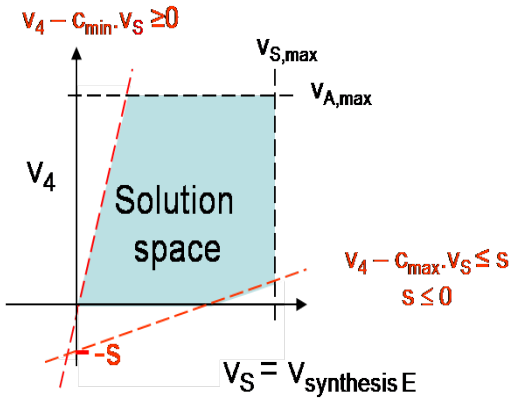


Explicit representation of an enzymatic reaction:



Coupling constraints





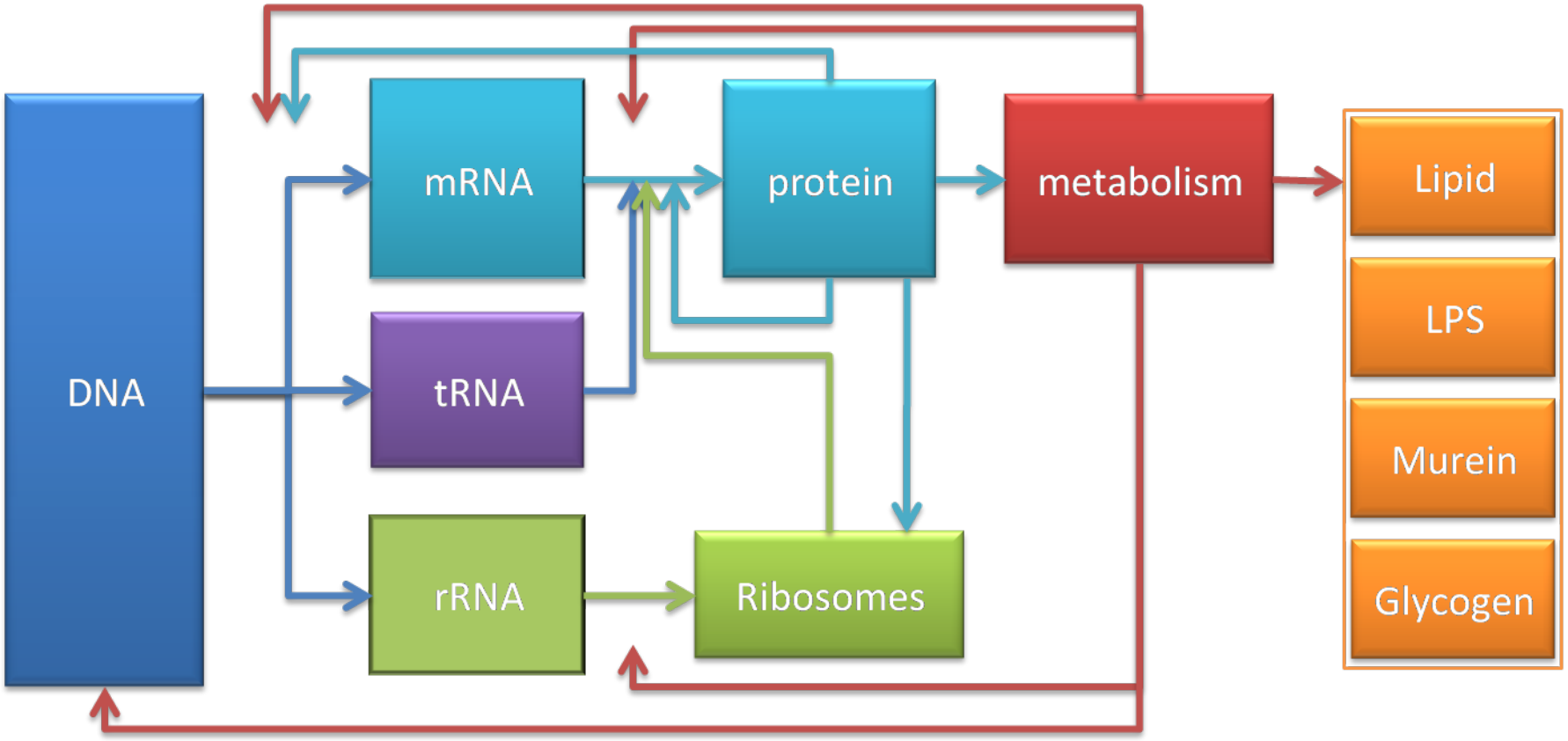
✓*If metabolic reaction is used, then protein & mRNA need to be produced*

✓*If flux through metabolic reaction increases, the synthesis rate of protein and mRNA needs to increase accordingly*

Thiele I, Fleming RMT, Bordbar A, Schellenberger J, Palsson BØ. Functional characterization of alternate optimal solutions of Escherichia coli's transcriptional and translational machinery. Biophysical journal. 98(10):2072-81, 2010.

Conversion of integrated reconstruction of metabolism and macromolecular synthesis into a computational model

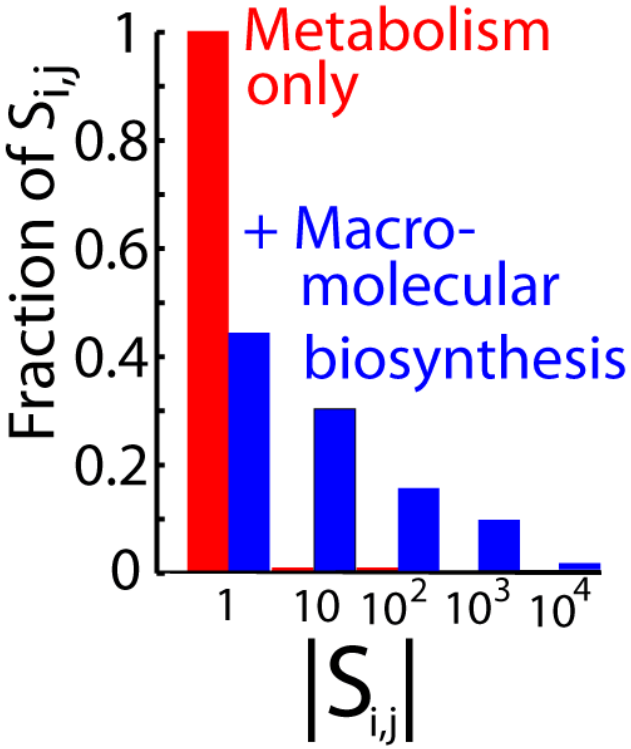
- Increasing scope of molecular processes represented



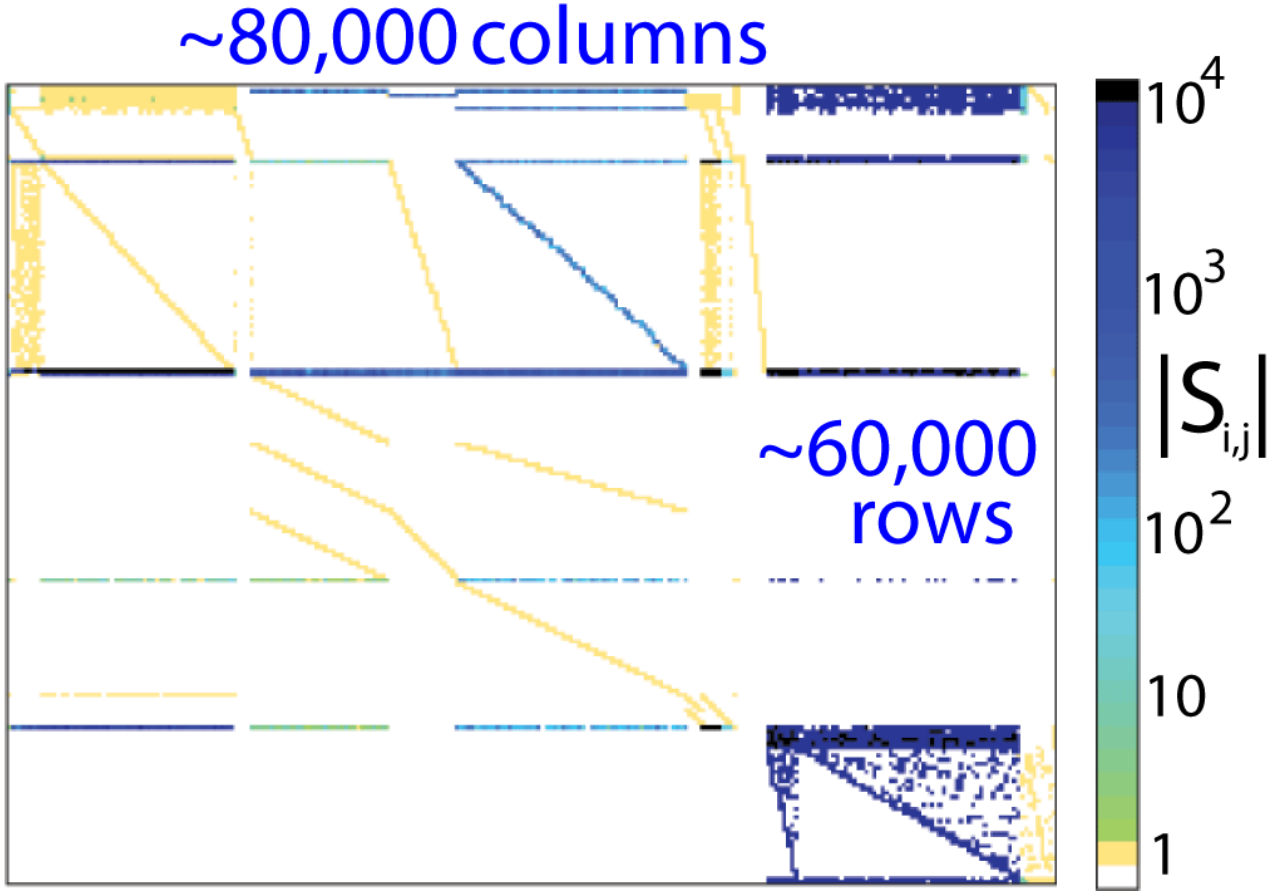
- However, molecular processes are intrinsically on different timescales scales…

Computational modeling requires numerical optimization involving large, sparse & stiff stoichiometric matrices: numerical analysis challenge

Many metabolic moieties in one macromolecule



Reaction rates over many orders of magnitude



SOFTWARE Open Access

Robust flux balance analysis of multiscale biochemical reaction networks

Yuekai Sun[1*], Ronan MT Fleming[2,3], Ines Thiele[2,3] and Michael A Saunders[4]

$$\begin{aligned} \underset{v}{\text{maximize}} \quad & c^T v \\ \text{subject to} \quad & Sv = 0, \\ & Cv \leq d, \\ & v_l \leq v \leq v_u, \end{aligned}$$

Reformulation involves a trade off between computational efficiency and reliability.

$$A + 10^9 B \rightarrow C + D$$

reformulate

$$\begin{aligned} A + 1000B_1 &\rightarrow C + D, \\ 1000B_2 &\rightarrow B_1, \\ 1000B &\rightarrow B_2 \end{aligned}$$

$$0.0001 \leq \frac{v_1}{v_2} \leq 10000.$$

reformulate

$$\begin{aligned} v_1 &\leq 100 s_1, & s_1 &\leq 100 v_2 \\ v_2 &\leq 100 s_2, & s_2 &\leq 100 v_1 \end{aligned}$$

BMC Bioinformatics 2013, 14:240 doi:10.1186/1471-2105-14-240

Robust flux balance analysis of multiscale biochemical reaction networks

Yuekai Sun[1*], Ronan MT Fleming[2,3], Ines Thiele[2,3] and Michael A Saunders[4]

Table 1 FBA results for ME76664 before and after lifting

68299 rows	Simplex		Barrier	
76664 columns	Before	After	Before	After
Iterations	48603	58288	56490	9985
CPU time	242	292	384	93
Infeasibilities	1.3×10^{-4}	2.9×10^{-6}	1.4×10^{-1}	3.4×10^{-6}

FBA results for the *E. coli* Metabolic-Expression model ME76664 using CPLEX primal simplex and barrier solvers. Iterations, time, and sum of infeasibilities before and after lifting. The iterations in columns 4 and 5 include about 100 for the barrier solver and the remainder for the simplex crossover.

BMC Bioinformatics 2013, 14:240 doi:10.1186/1471-2105-14-240

SOFTWARE Open Access

Robust flux balance analysis of multiscale biochemical reaction networks

Yuekai Sun[1*], Ronan MT Fleming[2,3], Ines Thiele[2,3] and Michael A Saunders[4]

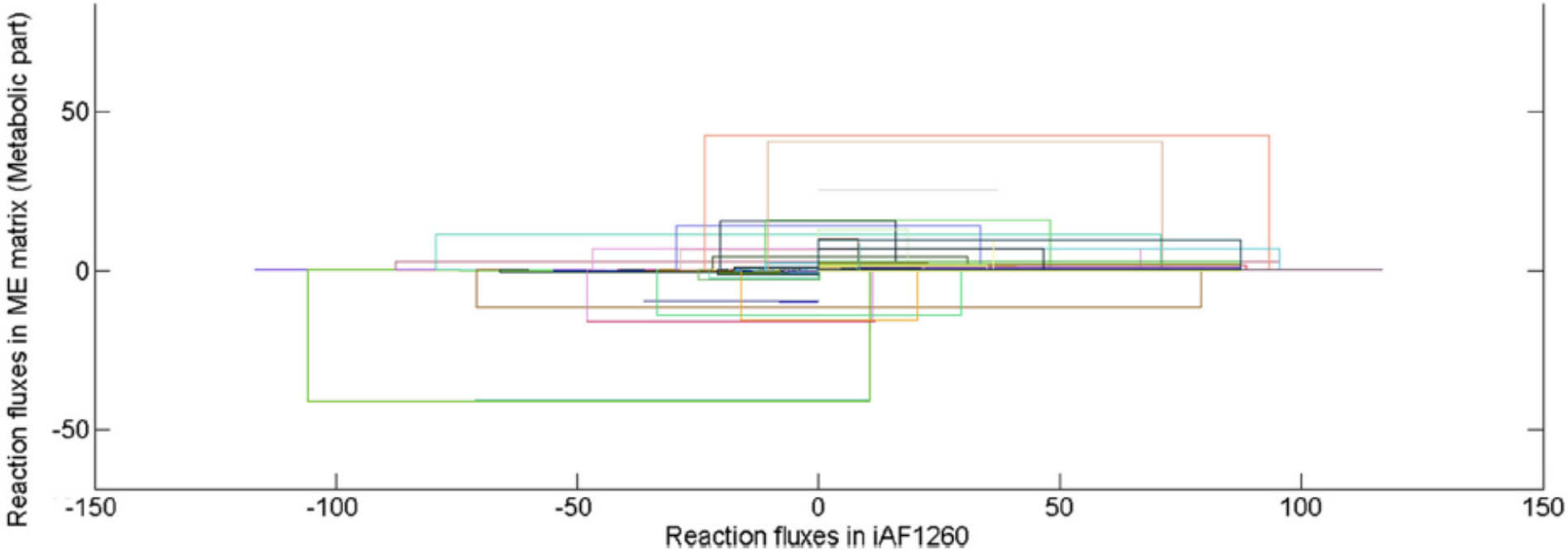
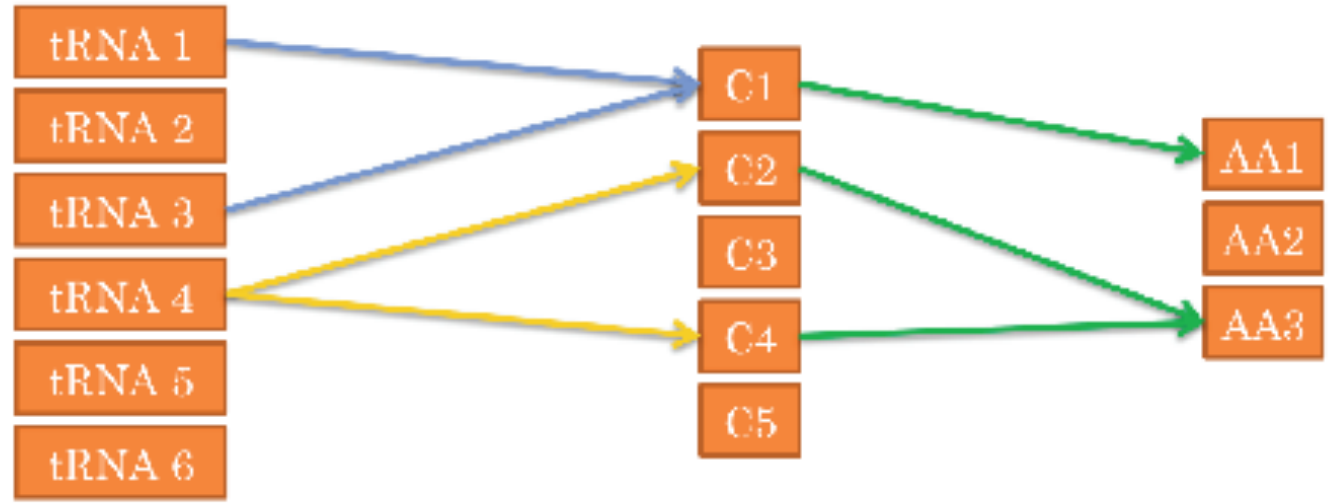


Figure 2 Flux variability analysis of the *E. coli* Metabolic-Expression model. Minimum and maximum flux for iAF1260 (which only accounts for metabolic reactions) versus the minimum and maximum flux for the Metabolic-Expression model. Each colored box corresponds to a different reaction in metabolism. The boxes are always longer on the axis for the metabolic model (iAF1260) than on the axis for the Metabolic-Expression model. This demonstrates that increasing the comprehensiveness of the model toward whole cell modeling leads to a substantial shrinkage of the steady state solution space. (Fluxes are plotted in mmol $\cdot$ g_{dw}^{-1} $\cdot$ hr^{-1}).

BMC Bioinformatics 2013, 14:240 doi:10.1186/1471-2105-14-240

New insights from multiscale systems biology models: *mechanics of the genotype-phenotype relationship*

- The genetic code has redundancy but no ambiguity
  1. can be multiple codons per amino acid
  2. multiple tRNA can read the same codon
  3. tRNA can read multiple synonymous codons
- Codon usage bias i.e. frequency of synonymous codons differs between organisms, within genomes, and along genes.



Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. PLoS One. 7(9):e45635, 2012.

New insights from multiscale systems biology models: *mechanics of the genotype-phenotype relationship*

Genetic code of E. coli (RNA perspective)

	U		C		A		G	
U	UUU	Phe (2)	UCU	Ser (5)	UAU	Tyr (3)	UGU	Cys (1)	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu (8)	UCA		UAA	Ochre	UGA	Opal	A
	UUG		UCG		UAG	Amber	UGG	Trp (1)	G
C	CUU	Leu (8)	CCU	Pro (3)	CAU	His (1)	CGU	Arg (7)	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln (4)	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile (5)	ACU	Thr (4)	AAU	Asn (4)	AGU	Ser (5)	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys (6)	AGA	Arg (7)	A
	AUG	Met (6)	ACG		AAG		AGG		G
G	GUU	Val (7)	GCU	Ala (5)	GAU	Asp (3)	GGU	Gly (6)	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu (4)	GGA		A
	GUG		GCG		GAG		GGG		G

E. coli has 86 tRNA molecules.

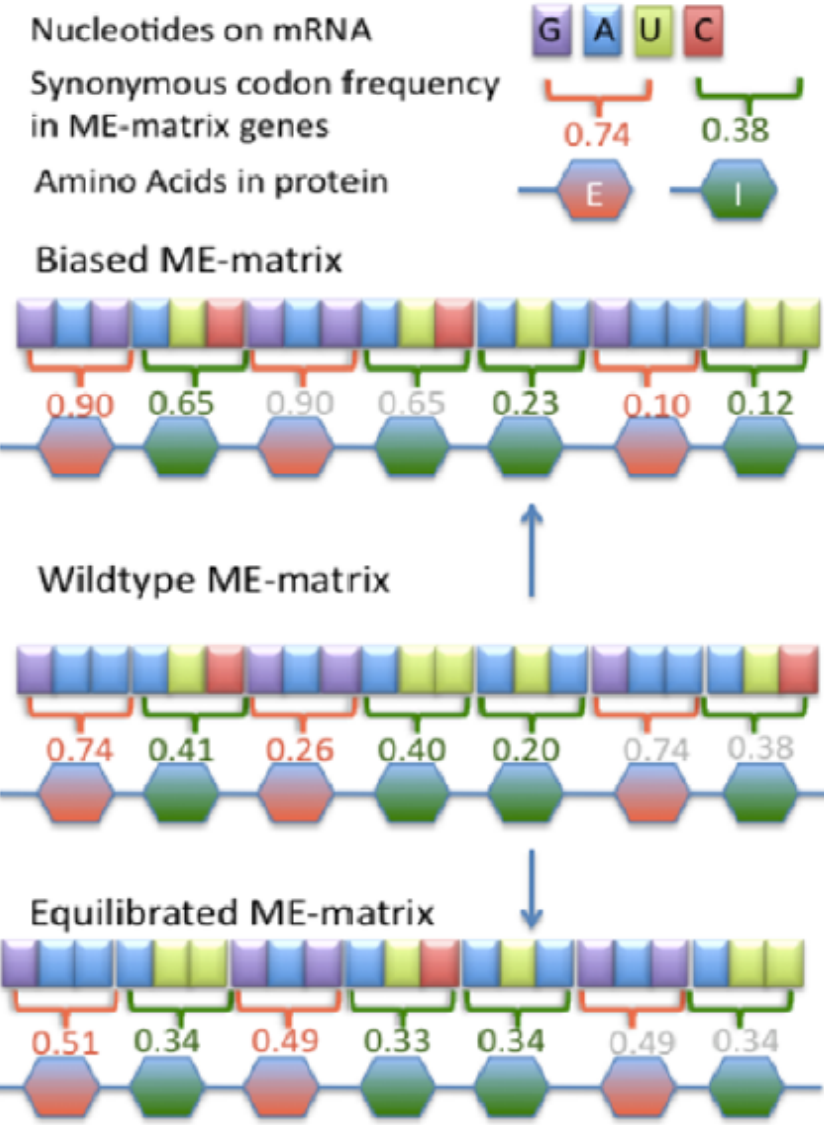
Number of distinct tRNA molecules per amino acid are given in parenthesis.

Leu = Leucine

- 6 different synonymous
- 8 different tRNA
- In wild type E. coli, CUG is the dominant synonymous codon

Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. PLoS One. 7(9):e45635, 2012.

New insights from multiscale systems biology models: *mechanics of the genotype-phenotype relationship*



The biased strains were generated using the following algorithm:

Input: model, sequence for each gene in model, number of iterations m

Output: model_biased

Algorithm:

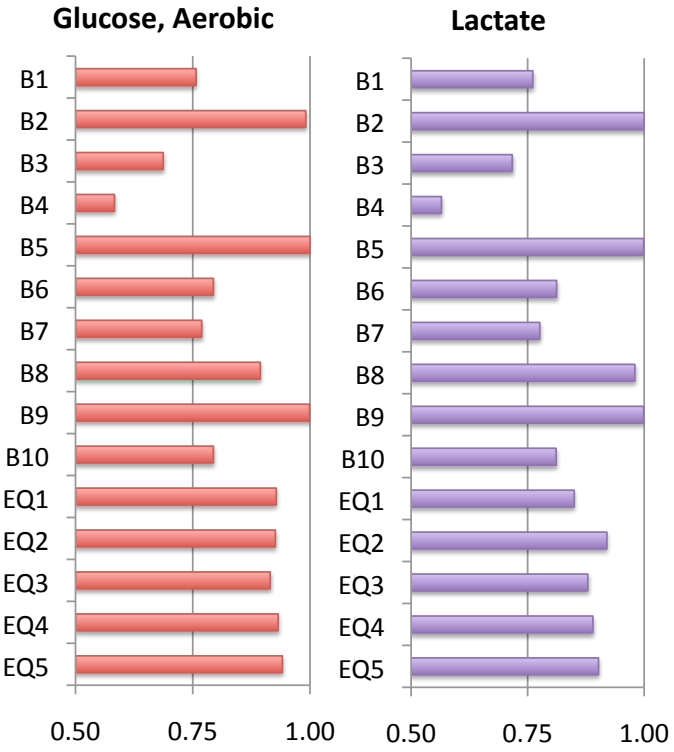
1. Choose randomly a codon, c_1
2. Identify possible synonymous codons: $c_s = \{c_1 = c_{s1}, c_{s2}, \ldots, c_{sk}\}$
3. Choose randomly one codon from c_s: c_{si}
4. Replace all instances of c_1 with c_{si}
5. Update ME-matrix for all genes based on new gene sequence:
   (a) Transcription reactions.
   (b) mRNA degradation reactions.
   (c) Translation reactions (tRNA molecule will be updated based on codon recognition).
6. Repeat 1 through 5 m times, m = 100.

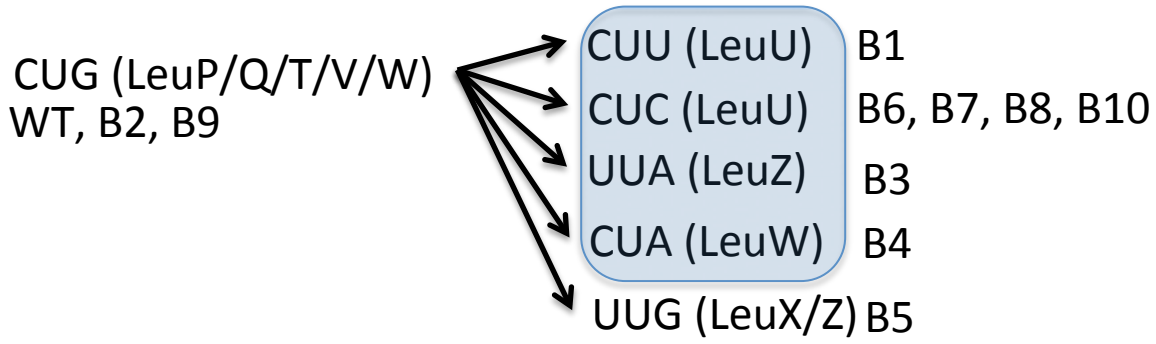
Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. PLoS One. 7(9):e45635, 2012.

New insights from multiscale systems biology models: *mechanics of the genotype-phenotype relationship*

- Changes in codon usage affect
  - ability to grow
  - maximal possible growth rate *in silico.*

- Analysis of numerical properties of flux balance analysis solutions used to derive causal molecular mechanistic hypothesis connecting codon usage with growth rate
- Limit to growth was ribosomal RNA operon transcription rate in wild type, but leucyl-tRNA transcription rate in biased strains



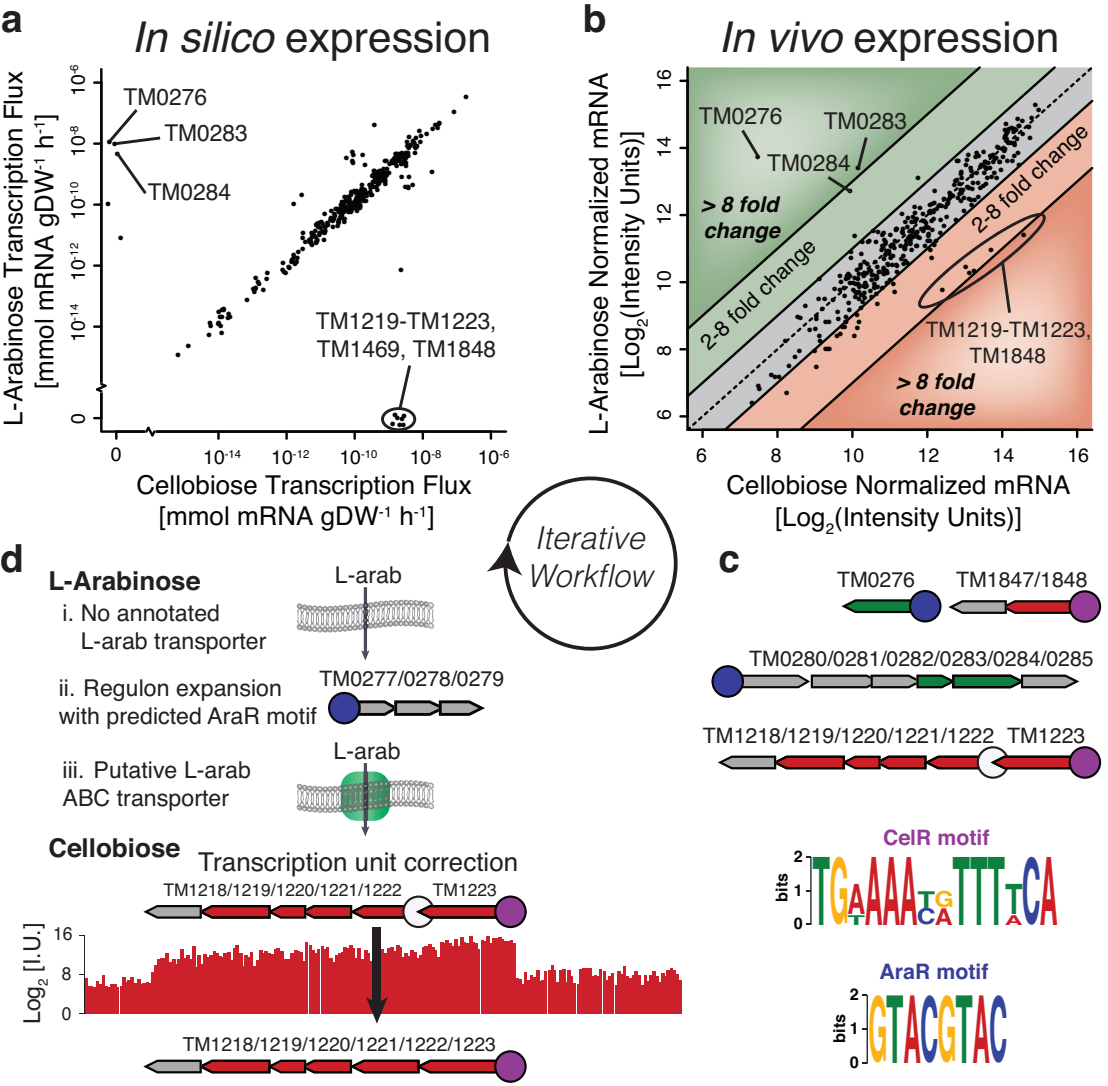


Increased tRNA demand may be met by augmenting supply

- e.g. modification of a tRNA to expand its set of read codons
  - In *E. coli* MAS39, a second leucyl-tRNA ($tRNA^{leuW}$) is able to read CUU due to a uridine-5-oxyacetic acid modification.
  - It remains to be experimentally established if E. coli MG1655 $tRNA^{leuW}$ can also read CUU.

Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. PLoS One. 7(9):e45635, 2012.

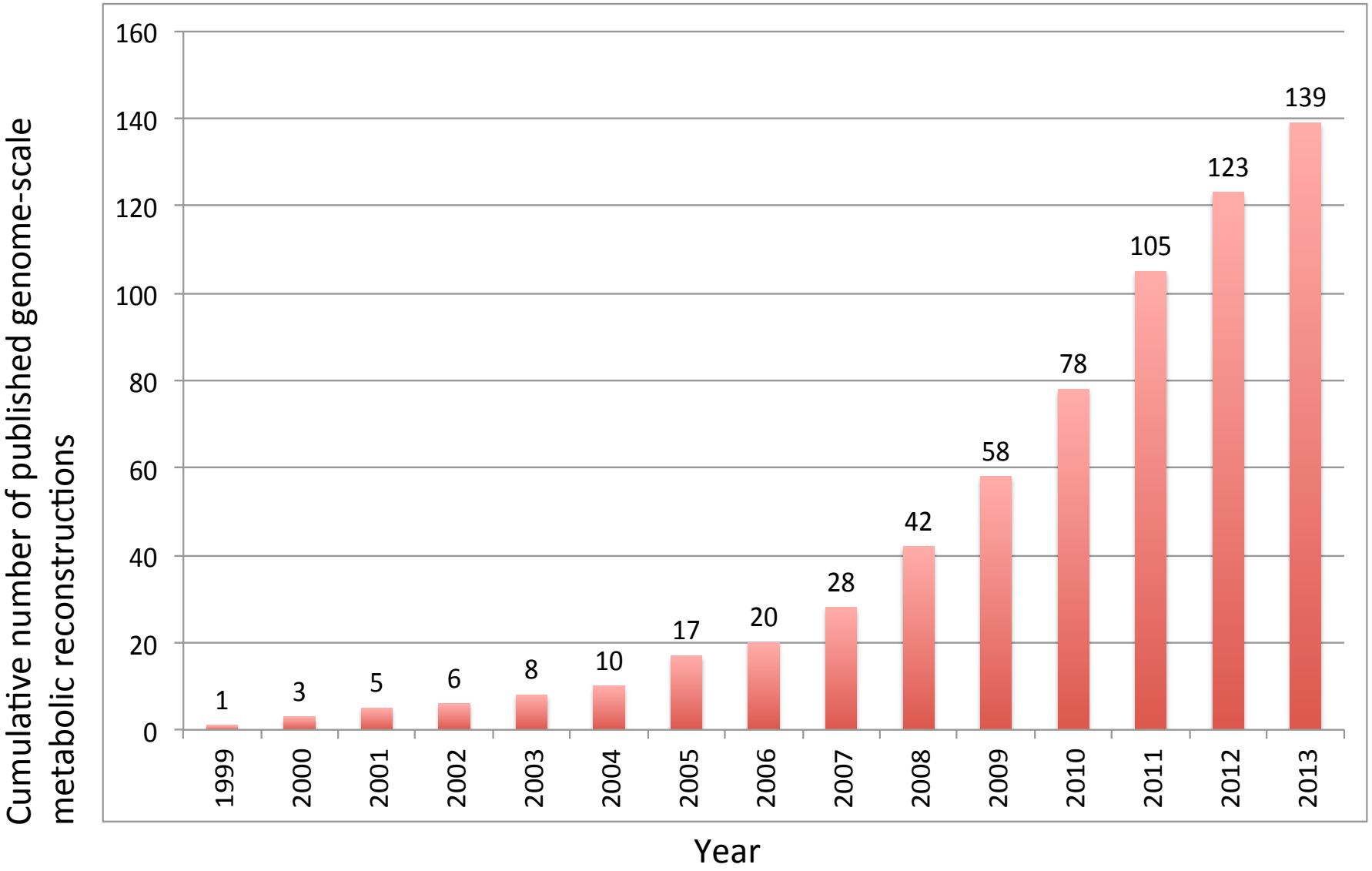
New insights from multiscale systems biology models: iterative annotation of gene function



- *(a & b)* ***In vivo*** **transcriptome measurements confirm the *in silico* transcriptomics** predictions for differential expression of genes when growing on L-Arabinose or cellobiose minimal medium.
- (c) scanning of promoter and upstream regions of essential genes **identified high-scoring motifs**
- (d) scan of remaining genome for AraR motif identified different genes, within a single transcriptional unit, with sequences similar to
  - a sugar-binding protein for an arabinose ABC transporter
  - a permeases of an ABC transporter
- **iterative workflow**: reconstruction-> model-> prediction -> new annotation-> better model -> …

Lerman JA, et al. In silico method for modelling metabolism and gene product expression at genome scale. Nat Commun. 3:929, 2012.

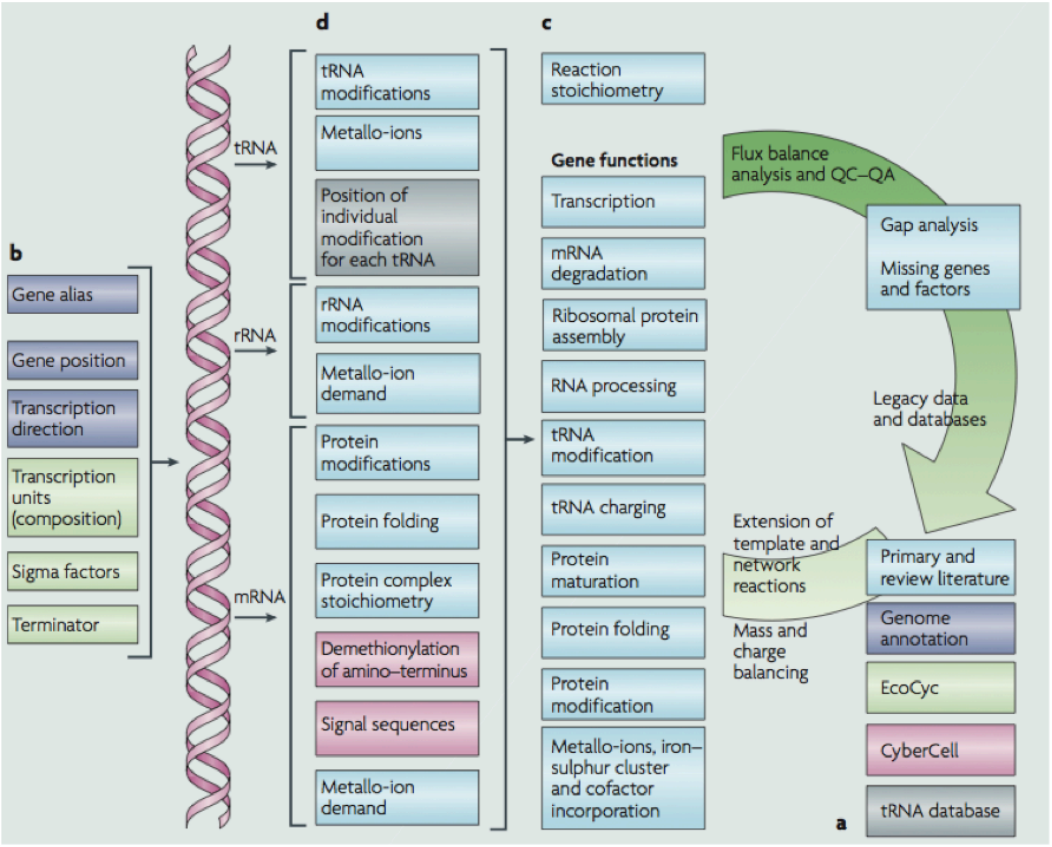
- Metabolic models for increasing number of species
- We envisage a demand for integrated models of metabolism and macromolecular synthesis for all of these species



U01 Aims 2012-2017:
Systems Biology Research Group, UCSD

Prototyping reconstruction & validation procedure on individual organisms, e.g. *E. coli* & *T. maritima*

Development of software workflows for multiscale model reconstruction and systematic validation using transcriptomic data



Feist, A. et al. Nat Rev Microbiol, 7(2):129–143, 2009.

- Develop software for reconstruction of biochemical networks spanning multiple cellular subsystems.
- Develop techniques for quantitative prediction and validation of transcript abundance in an integrated model of metabolism, macromolecular synthesis and regulation
  - Comparison with new experimental data
  - e.g. for Geobacter spp.
    - biogeochemical cycling of carbon and metals
    - bioenergy applications.

U01 Aims 2012-2017: Systems Optimization Laboratory, Stanford.

- Development of quad precision versions of large-scale LP/QP/NLP solvers (SQOPT , SNOPT), and their linear sparse-matrix algebra 'engine' (LUSOL)
  - re-implement in Fortran 2003 with quad precision variables
  - redesign of key components, e.g., storage allocation
  - Software: www.stanford.edu/group/SOL/multiscale/software.html

	Decade	Variables	Computation
1a	1970 Fortran	single	single
1b	1970 C	single	double
2a	1980	single	single
2b		double	double
3a	2010	single	single
3b		double	double
3c		quad	quad

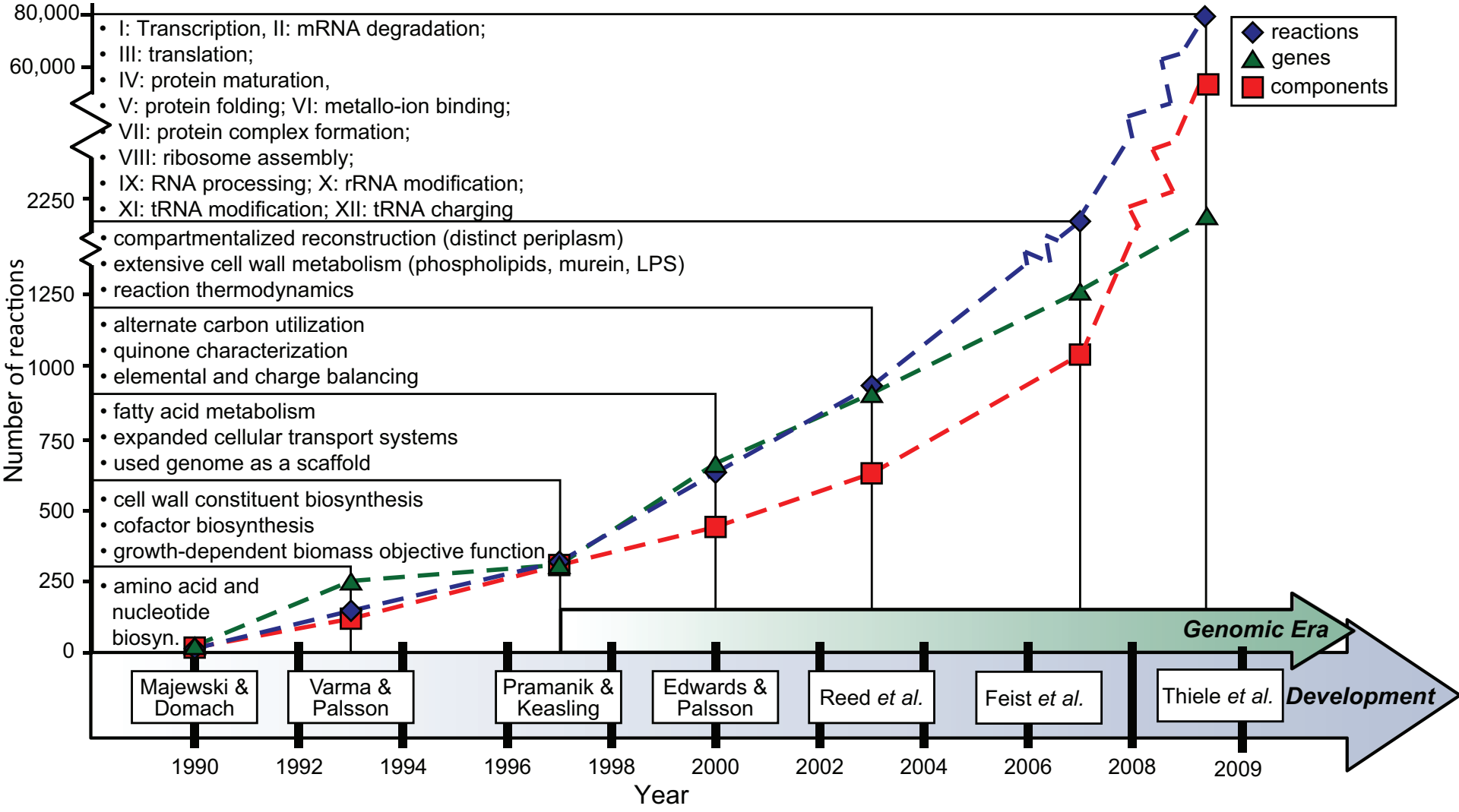
Table 2: History of Scientific Computing

Increasing the reliability, while maintaining efficiency of numerical optimization solvers

Current large-scale LP & QP solvers

New large-scale LP & QP solvers

Increasing size of *E. coli* reconstructions



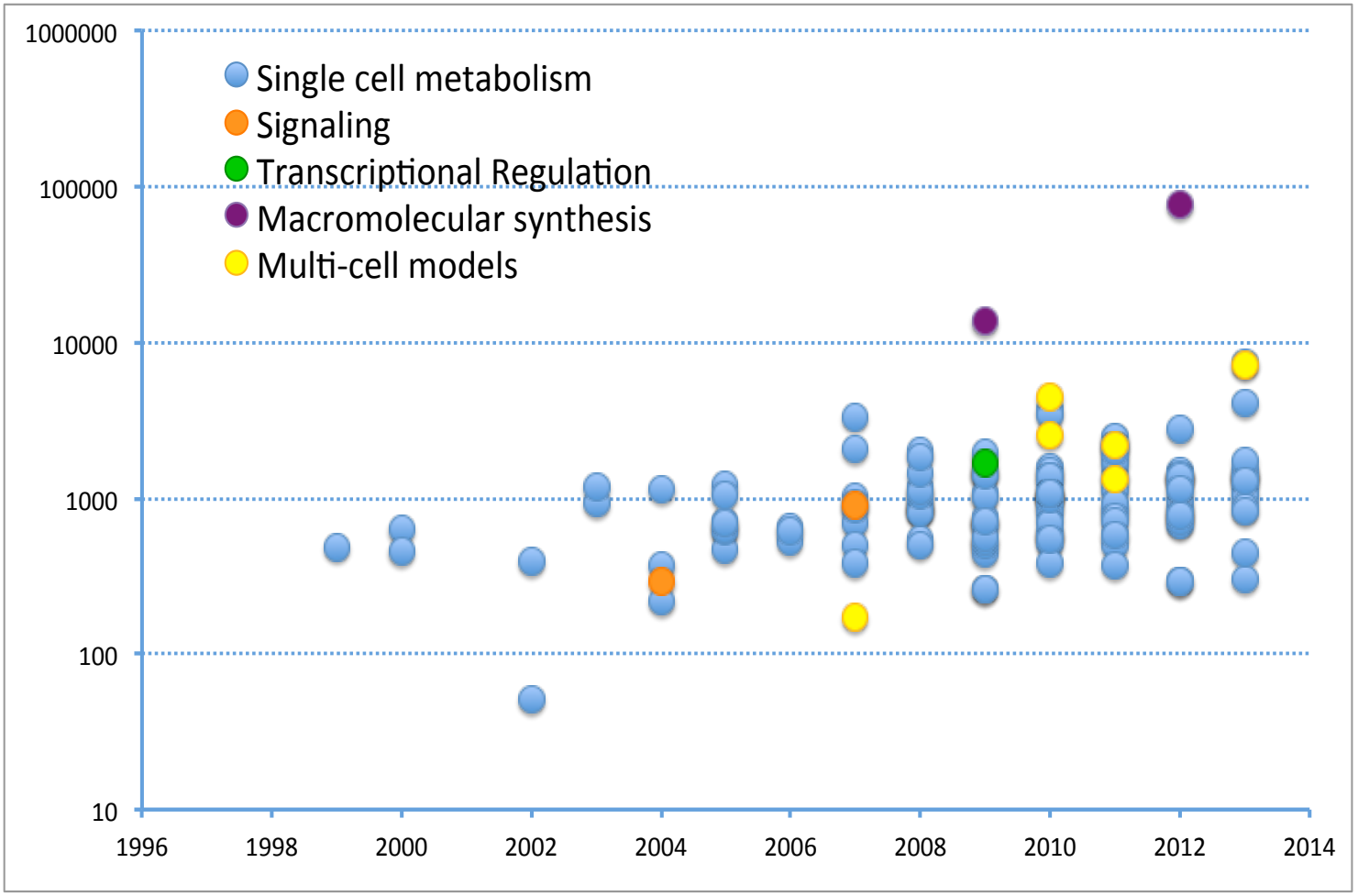
Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. PLoS One. 7(9):e45635, 2012.

U01 Aims 2012-2017: Systems Optimization Laboratory, Stanford.

- Development of hypergraph network flow algorithms for optimization with biological networks
  - Standard network flow algorithms
    - Complexity O(n)
    - Designed specifically for graphs, whereas biological networks are hypergraphs
  - Standard linear optimization algorithms
    - Complexity O(n^3)
    - Used for, amongst many other things, optimization over biochemical networks
      - e.g. Flux Balance Analysis
  - Open research question: Does there exist an algorithm of lower computational complexity than a standard linear optimization solver, specifically for biochemical hypergraph flow problems?

$$\begin{aligned} \underset{v}{\text{maximize}} \quad & c^T v \\ \text{subject to} \quad & Sv = 0, \\ & Cv \leq d, \\ & v_l \leq v \leq v_u, \end{aligned}$$

Biochemical reaction network models are increasing in scope



However, the currently available approaches for large scale modeling of biochemical reaction networks only explicitly represent reaction flux, not molecular abundance.

Open question: does there exist a numerically scalable approach to model fluxes and concentrations explicitly?

$$\begin{aligned} \underset{v}{\text{maximize}} \quad & c^T v \\ \text{subject to} \quad & Sv = 0, \\ & Cv \leq d, \\ & v_l \leq v \leq v_u, \end{aligned}$$

- Consider a biochemical network with m *molecular species* and n reversible chemical reactions

- Define forward and reverse stoichiometric matrices, $F, R \in \mathbb{Z}_{\geq 0}^{m,n}$, respectively, where F_{ij} denotes the stoichiometry[1] of the i^{th} molecular species in the j^{th} forward elementary reaction and R_{ij} denotes the stoichiometry of the i^{th} molecular species in j^{th} reverse elementary reaction.

- We assume that every elementary reaction conserves mass, that is, there exists at least one positive vector $l \in \mathbb{R}_{>0}^{m}$ satisfying $(R-F)^T l = 0$
  where $R - F$ represents net reaction stoichiometry.

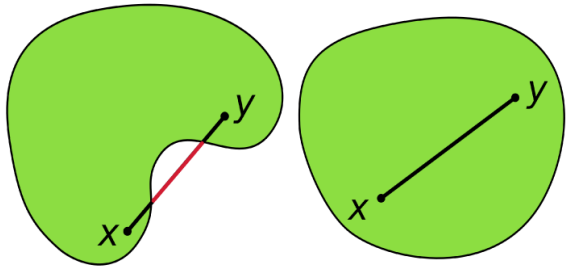
- Let $c \in \mathbb{R}_{>0}^{m}$ denote a variable vector of molecular species concentrations.

- Assuming constant non-negative *elementary kinetic parameters* $k_f, k_r \in \mathbb{R}_{\geq 0}^{n}$, we assume *elementary reaction kinetics* for forward and reverse elementary reaction rates as $v_f(k_f, c) \equiv \exp(\ln(k_f) + F^T \ln(c))$ and $v_r(k_r, c) \equiv \exp(\ln(k_r) + F^T \ln(c))$, respectively.

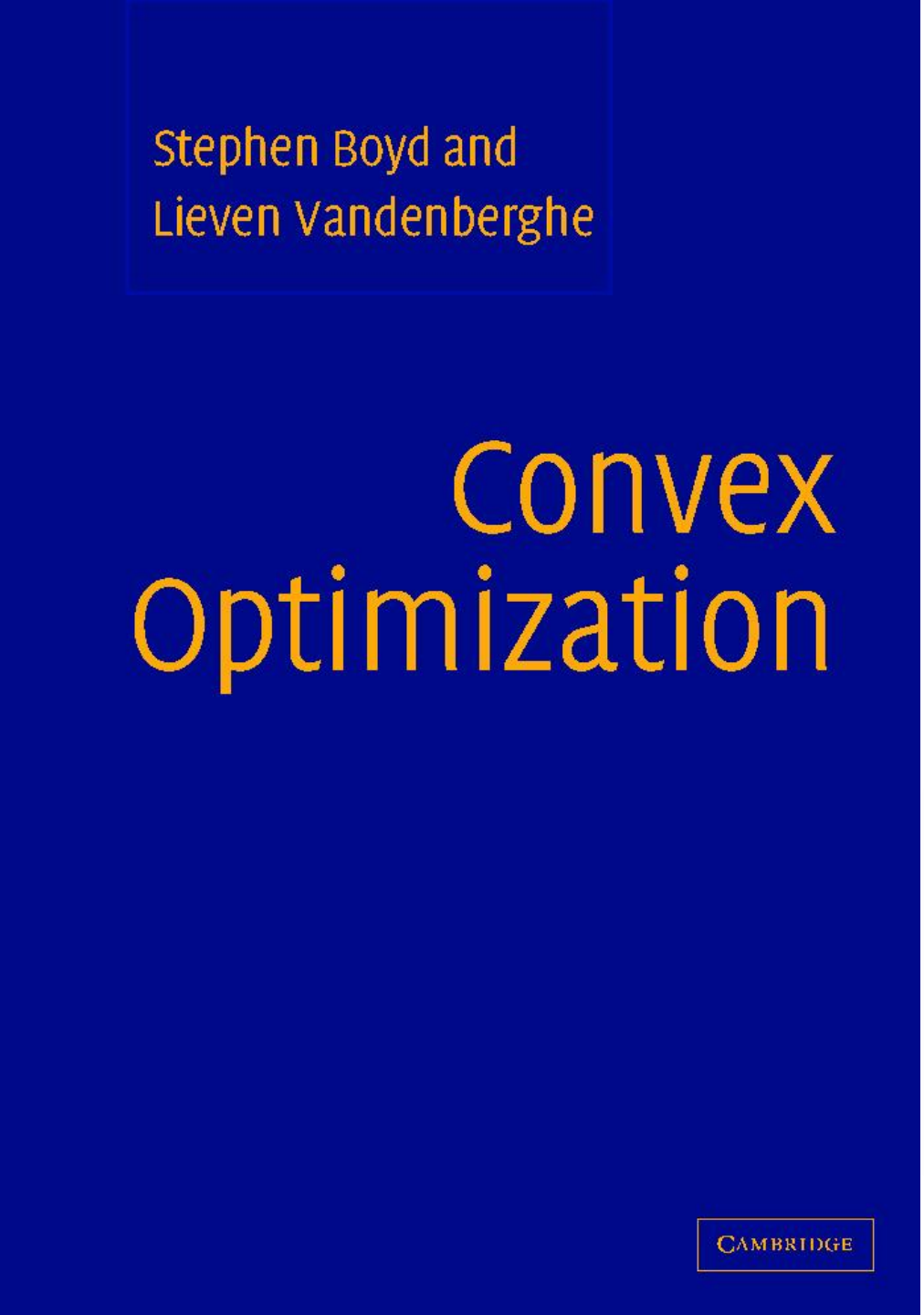
- The deterministic dynamical equation for time evolution of molecular species concentration may then be expressed as

$$\begin{aligned}\frac{dc}{dt} &\equiv (R-F)(v_f(k_f,c)-v_r(k_r,c)),\\ &= (R-F)(\exp(\ln(k_f)+F^T\ln(c))-\exp(\ln(k_r)+F^T\ln(c)))\equiv v(c)\end{aligned}$$

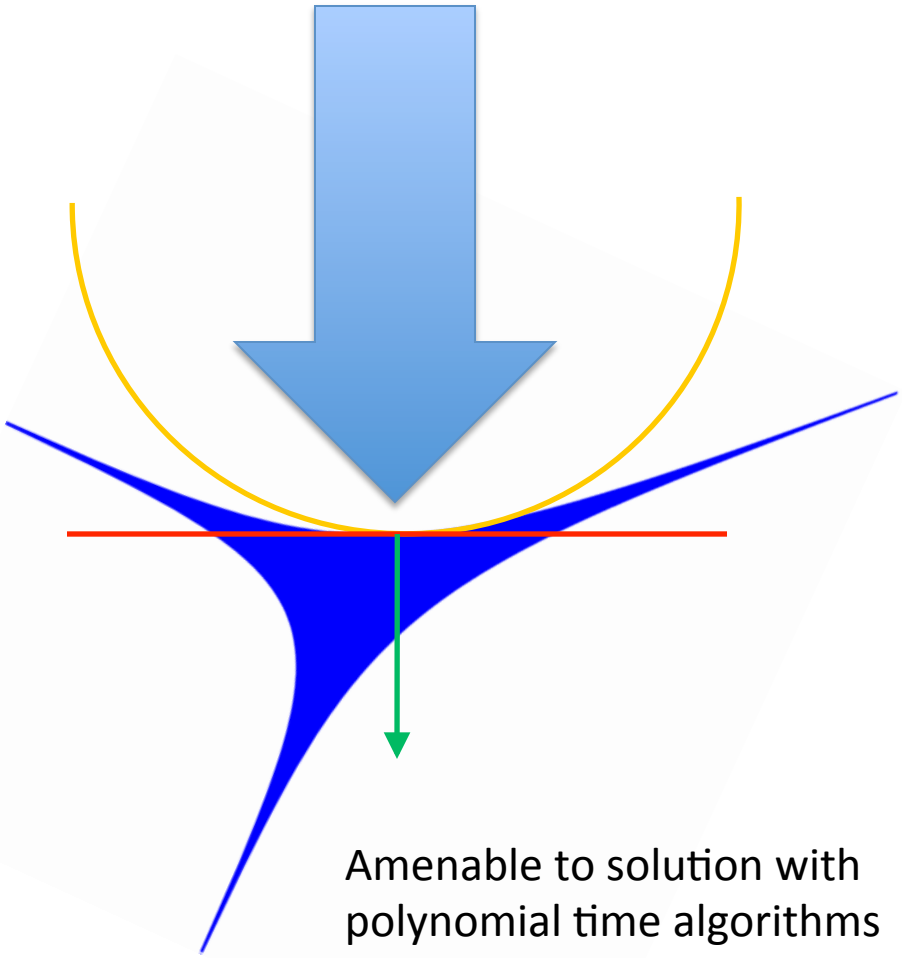
Assuming a non-equilibrium steady state, $\frac{dc}{dt}=0$ and $v_f(k_f,c)\neq v_r(k_r,c)$, one is then interested in the nonlinear, nonconvex set of steady state molecular species concentrations $\{c|v(c)=0\}$.



- Modeling challenge
  - High dimensional
    - many molecular species and reactions
  - Nonlinear
    - nonlinear relationship between molecular species abundance and reaction rate
  - Need algorithms with low polynomial time complexity, guaranteed convergence, certificate of infeasibility
    - mathematical model & algorithm formulation problem
  - Multiscale
    - molecular species concentrations vary over many orders of magnitude
      - e.g. transcript abundance versus metabolite abundance
  - Paucity of kinetic parameters



Nonlinear relations between reaction rates and metabolite concentrations satisfied at optimum of a convex optimization problem



Amenable to solution with polynomial time algorithms

A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks

R.M.T. Fleming [a,*], C.M. Maes [b], M.A. Saunders [c], Y. Ye [c], B.Ø. Palsson [d]

Journal of Theoretical Biology 292 (2012) 71–77

A B S T R A C T

We derive a convex optimization problem on a steady-state nonequilibrium network of biochemical reactions, with the property that energy conservation and the second law of thermodynamics both hold at the problem solution. This suggests a new variational principle for biochemical networks that can be implemented in a computationally tractable manner. We derive the Lagrange dual of the optimization problem and use strong duality to demonstrate that a biochemical analogue of Tellegen's theorem holds at optimality. Each optimal flux is dependent on a free parameter that we relate to an elementary kinetic parameter when mass action kinetics is assumed.

Theorem 1. *Let $v_e^\star$ be any set of optimal exchange fluxes from problem (FBA). Define $b=-S_e v_e^\star$, and let c be any vector in $\mathbb{R}^n$. The convex equality-constrained problem*

$$\underset{v_f, v_r > 0}{\text{minimize}} \quad \phi \equiv v_f^T(\log(v_f)+c-e)+v_r^T(\log(v_r)+c-e)$$

$$\text{subject to} \quad Sv_f - Sv_r = b : y \qquad \text{(EP)}$$

is then feasible, and its solution $(v_f^\star, v_r^\star)$ is a set of thermodynamically feasible internal fluxes. The combined vector $(v_f^\star, v_r^\star, v_e^\star)$ is thermodynamically feasible and optimal for problem (FBA). The associated chemical potentials u may be obtained from the optimal Lagrange multiplier $y^\star \in \mathbb{R}^m$ for the equality constraints according to $u=-2\rho y^\star$.

- Constraint on ratio of concentrations, not absolute concentration.
- Biochemical reaction directions are an evolved subset of thermodynamically feasible directions.

PLoS Comput Biol 9(7): e1003098. doi:10.1371/journal.pcbi.1003098
OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Consistent Estimation of Gibbs Energy Using Component Contributions

Elad Noor[1☯], Hulda S. Haraldsdóttir[2☯], Ron Milo[1]*, Ronan M. T. Fleming[2,3]*

$$\exp\left(-S_j^T \cdot \frac{u^\circ}{RT}\right) = \frac{k_{fj}}{k_{rj}}$$

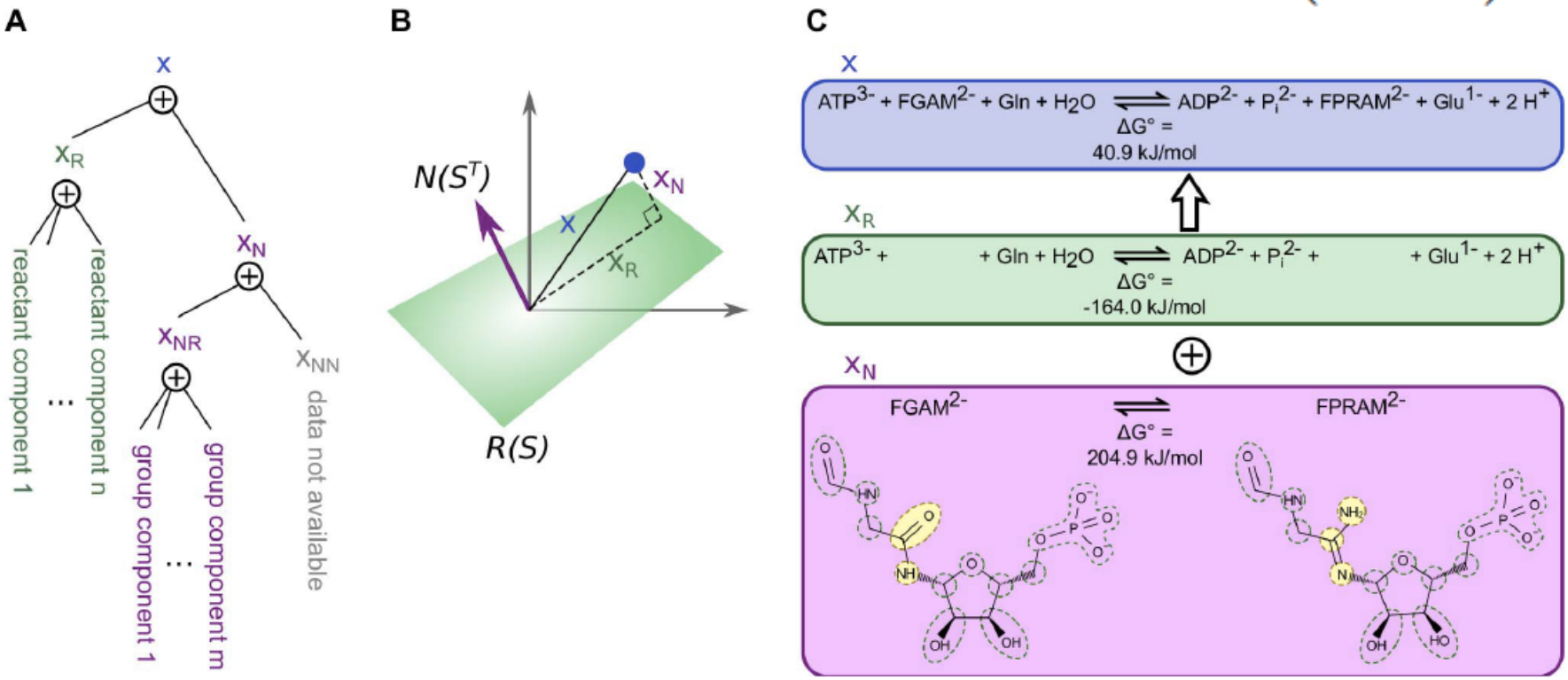


Figure 3. A diagram illustrating how the component contribution method projects the stoichiometric vector onto the different spaces. (A) The reaction vector x is decomposed into the two components x_R and x_N, where the reactant contribution and group contribution methods are used for the relevant components. Later, x_N is decomposed into x_{NR} and x_{NN}. The same projection is shown graphically in (B) where the green plane represents the range of S and the normal to that plane represents the null space of S^T. (C) An example for a reaction which decomposes into two non-zero components. In this case, the component x_{NN} is equal to 0, which means that the reaction is covered by the component contribution method.

Mass conserved elementary kinetics is sufficient for the existence of a non-equilibrium steady state concentration

R.M.T. Fleming [a,b,*], I. Thiele [a,c]

Journal of Theoretical Biology 314 (2012) 173–181

When does there exist a non-equilibrium steady state concentration?

Theorem 1. *Let the dynamical equation for mass conserved elementary kinetics be*

$$\dot{c} \equiv \frac{dx}{dt} = S \cdot (K_f \cdot \exp(F^T \cdot \ln(c)) - K_r \cdot \exp(R^T \cdot \ln(c))), \quad (4)$$

where $c \equiv c(t) \in R^m$ *is the molecule concentration at time* $t > 0$, $\dot{c} \in R^m$ *is the time derivative of concentration,* $K_f = \mathrm{diag}(k_f)$, $K_r = \mathrm{diag}(k_r)$ *and* k_f, $k_r \in R^n_{\geq 0}$ *are non-negative forward and reverse kinetic parameters.* $F, R \in R^{m,n}_{\geq 0}$ *are forward and reverse stoichiometric matrices.* $S \equiv -F + R$ *is a consistent stoichiometric matrix defined by the existence of at least one strictly positive vector* $l \in R^m_{>0}$, *such that* $S^T \cdot l = 0$. *Assuming a finite and strictly positive initial concentration* $c_0 \equiv c(0) \in R^m_{>0}$, *then there exists at least one finite and non-negative steady state concentration* $x^\star_{\geq 0}$, *such that* $\dot{c} = 0$.

U01 Aims 2012-2017: Systems Biochemistry & Molecular Systems Physiology Groups, Luxembourg.

- **Scalable algorithms for multiscale reconstruction and modeling**
  - **Software to enable high fidelity reconstruction of biochemical networks**
    - e.g. checking for consistency with known biochemistry
    - e.g. suggesting extension to existing reconstruction to account for known biochemical function
  - **Multiscale mass conserved elementary kinetic modeling**
    - **Forward problem**
      - given kinetic parameters, compute a non-equilibrium steady state
    - **Inverse problem**
      - given reaction stoichiometry, experimental boundary conditions thermodynamic constraints and net reaction directions consistent with biochemistry, search among consistent kinetic parameters
      - Gradient-based search of kinetic parameters in multiscale models.
        - existing algorithms do not search for kinetic parameters using gradient based methods due to a perception that the merit function for such a problem contains local minima
        - High risk, high gain project: does a formulation of the problem exist which is amenable to solution with a polynomial time algorithm guaranteed to reach a global optima?

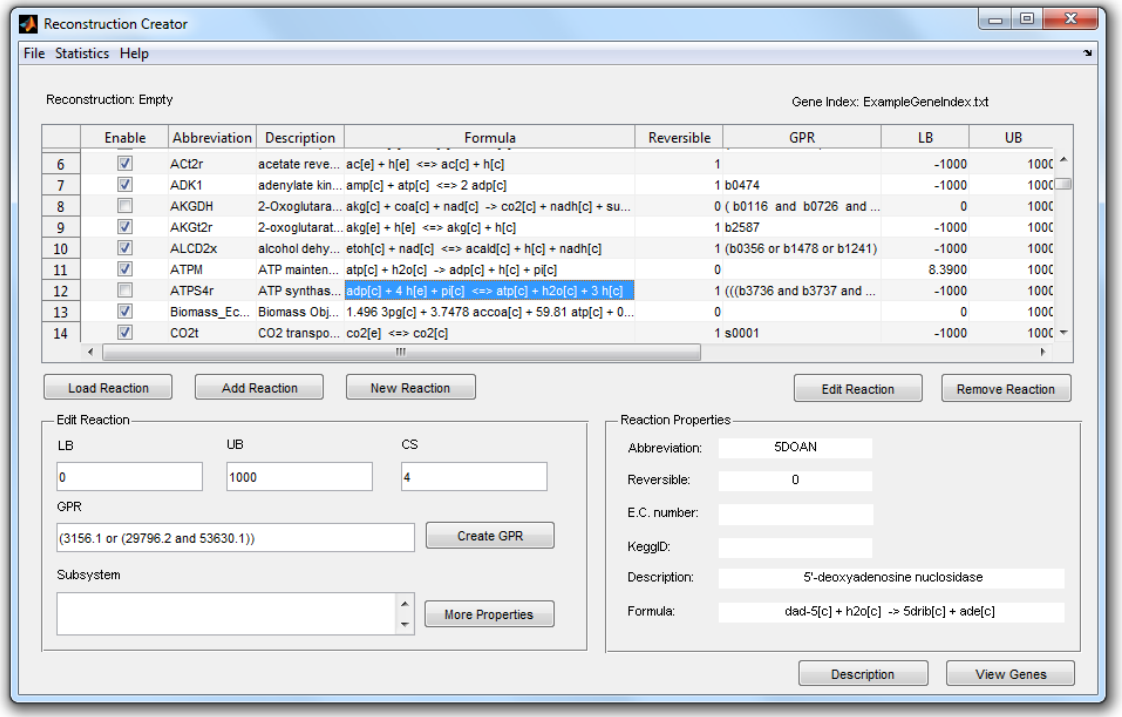
Sharing models

- Network reconstructions used to generate computational models are always made available with the accompanying paper
  - Systems Biology Markup Language (SBML)
    - all monoscale models
    - for multiscale models, standardised representation that is scalable needs to be developed.
    - multiscale models still distributed, but representation not yet standardised.

http://systemsbiology.ucsd.edu/Downloads
http://thielelab.eu/
http://www.stanford.edu/group/SOL/multiscale/models.html

Open source software

- **rBioNet** - a COBRA toolbox extension for reconstructing high-quality biochemical networks *(Thorleifsson & Thiele, Bioinf, 2011)*
- **von Bertalanffy 1.0** - a COBRA toolbox extension to thermodynamically constrain metabolic models *(Fleming & Thiele, Bioinf, 2011)*
- **COBRA Toolbox v2.0** - quantitative prediction of cellular metabolism with constraint-based models *(Schellenberger et al, Nat Protoc, 2011)*



- **COBRApy -** COnstraints-Based Reconstruction and Analysis for Python *(Ebrahim et al, BMC Syst Biol, 2013)*
- **fastFVA** – a tool for computationally efficient flux variability analysis *(Gudmunsson & Thiele, BMC Bioinf, 2010)*
- **robustFBA** - Robust flux balance analysis of multi- scale biochemical reaction networks *(Sun et al, BMC Bioinf, 2013)*

See http://www.stanford.edu/group/SOL/multiscale/software.html for links to software

Open source software continued:

- **LUSOL, LUMOD** Routines for dense and sparse LU factorization.
- **PDCO** A primal-dual interior point method for large-scale optimization with convex objective and linear constraints.
- **PNOPT** Proximal Newton-type methods for minimizing composite functions (unconstrained optimization of the sum of smooth and nonsmooth functions).
  - **Need help with cobra methodology?**
    - openCOBRA Google group
      - https://groups.google.com/forum/#!forum/cobra-toolbox
        - >2009
        - 430 posts
        - 300 members
        - Anyone can view content.
        - Anyone can apply to join.

See http://www.stanford.edu/group/SOL/multiscale/software.html for links to software

Thanks

- Colleagues in the Multiscale Systems Biology Collaboration
- Funding 2012-2017
  - National Institutes of Health (NIH) through the Multiscale Modeling Physiome Initiative, grant 1U01GM102098-01.
  - Department of Energy (DOE) through the Scientific Discovery Through Advanced Computing Program, grant DE-SC0010429.
- Program coordination
  - Interagency Modeling and Analysis group