

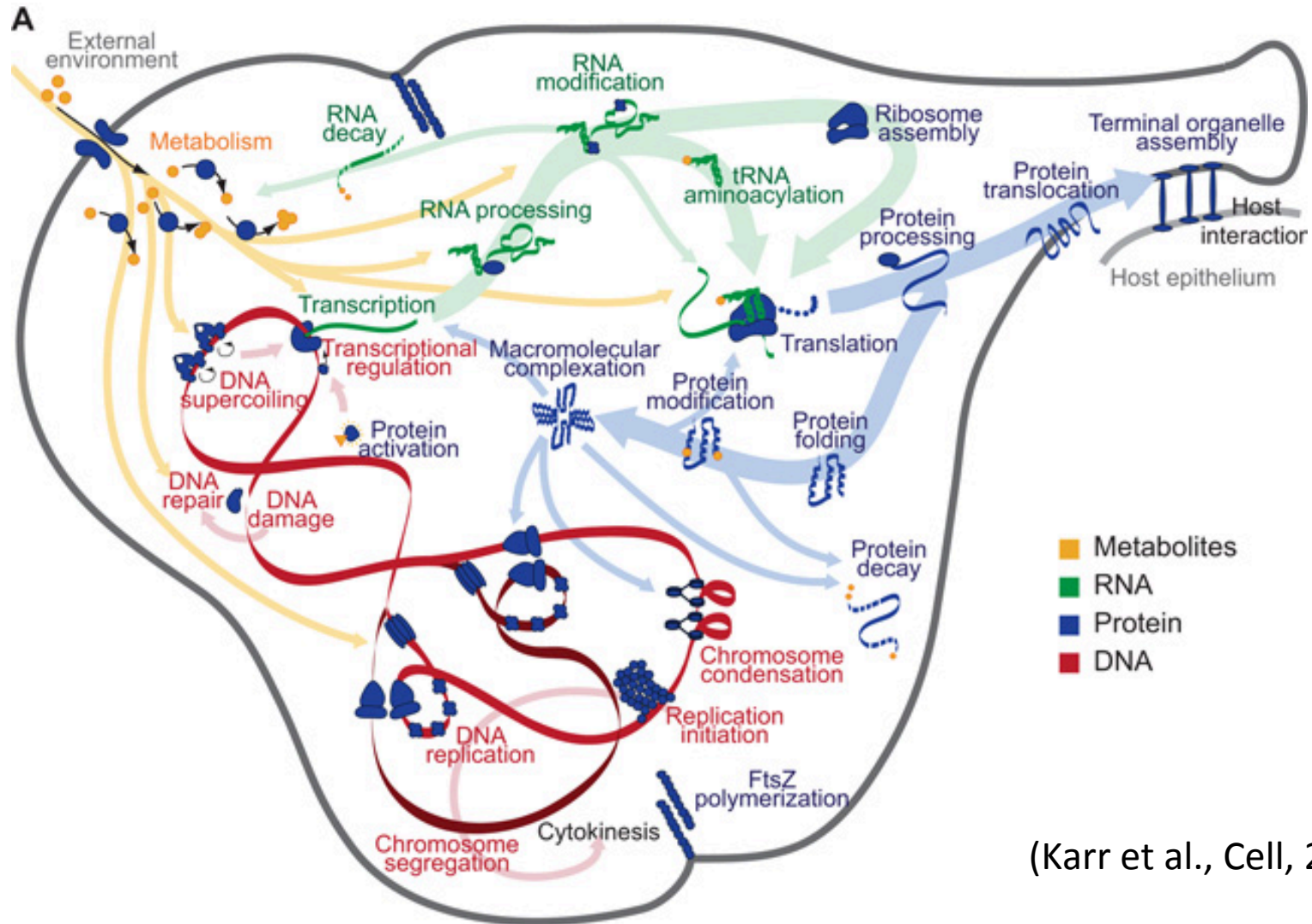
Network inference and modeling

BIOINF 525

Module 3, Lecture 3

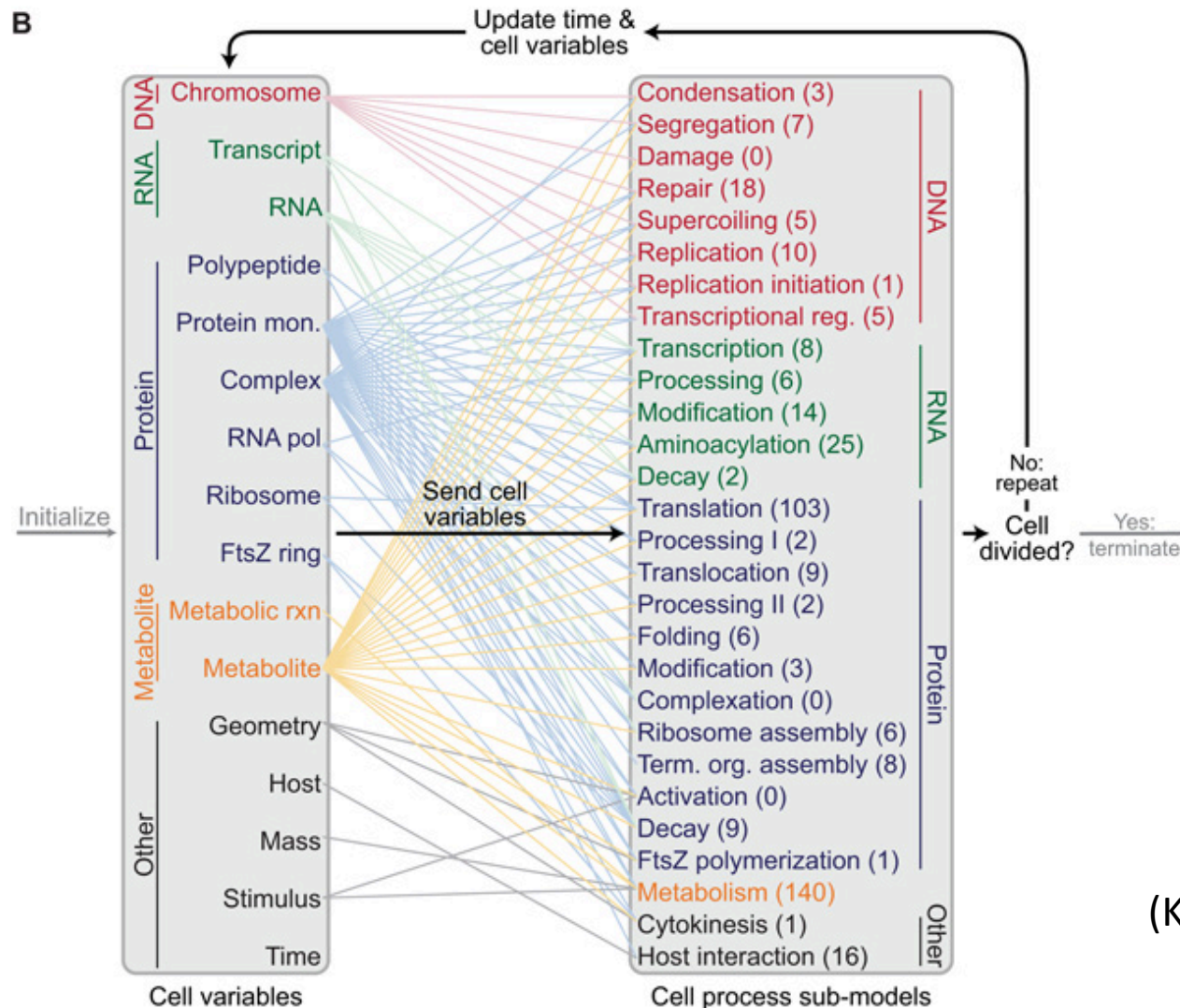
4/4/2017

Whole cell models allow prediction of phenotype from genotype



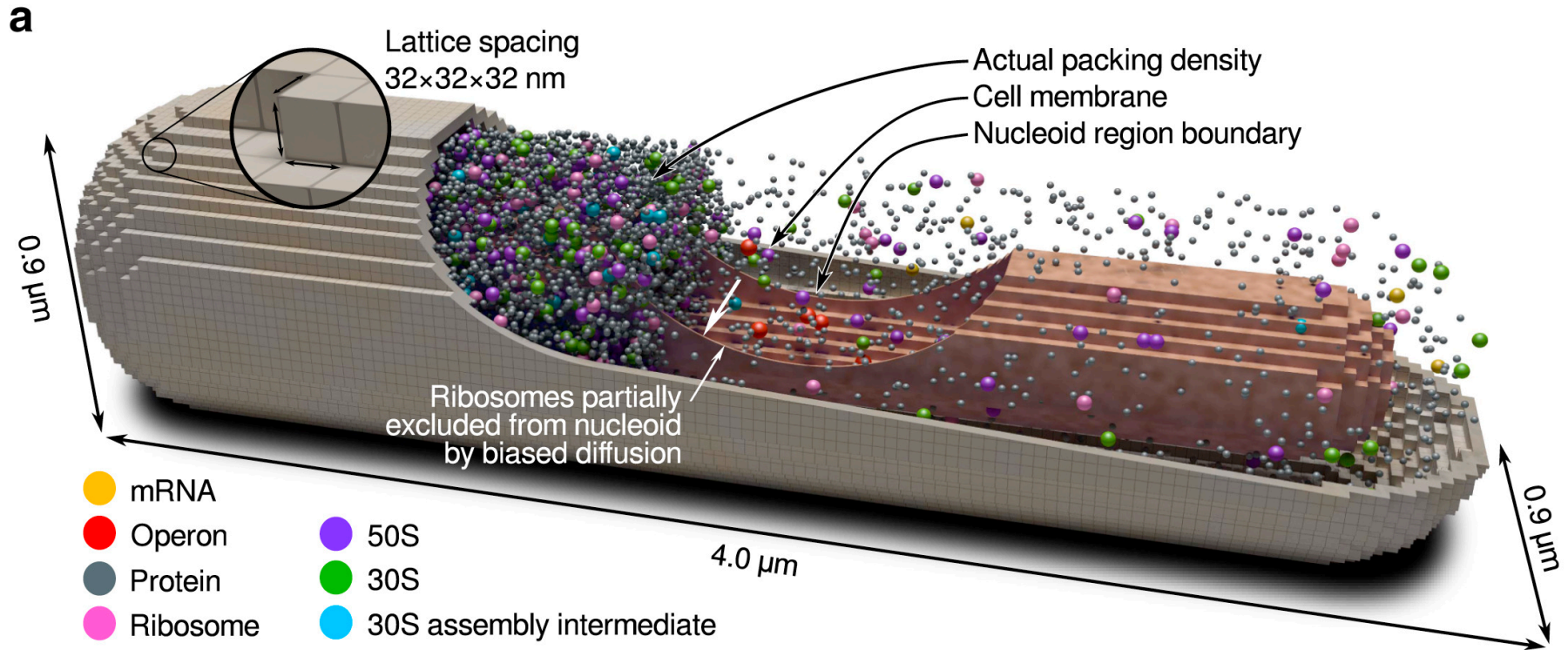
(Karr et al., Cell, 2012)

Whole cell models allow prediction of phenotype from genotype



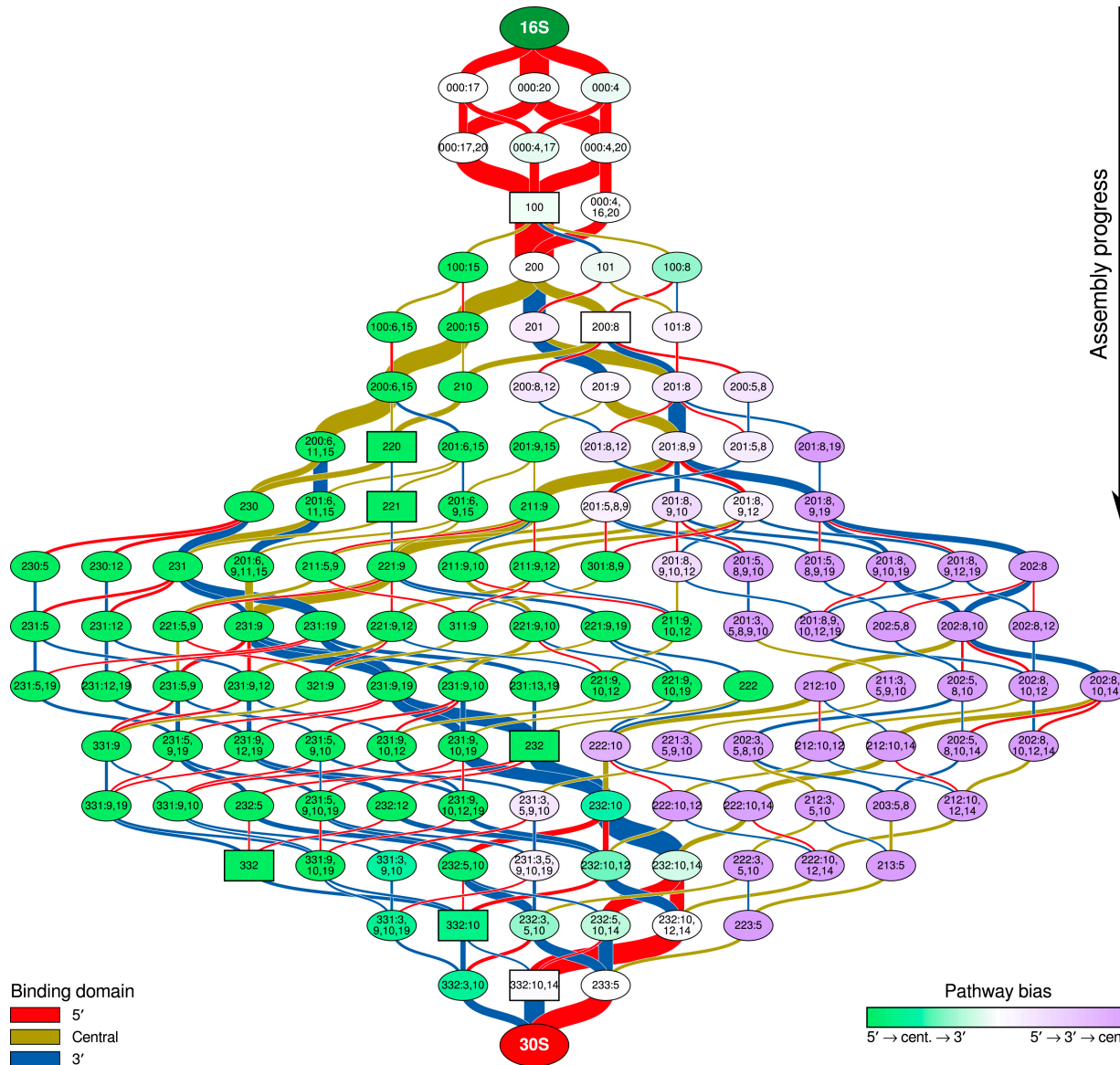
(Karr et al., Cell, 2012)

Spatially resolved cell-scale models



(Earnest et al., Biophys. J. 2015)

Spatially resolved cell-scale models



Assembly progress

(Earnest et al.,
Biophys. J. 2015)

Building systems-level models

- Motif and GO term analysis
- Kinetic modeling of simple networks
- Constraint-based modeling of cellular metabolism

Building systems-level models

- **Motif and GO term analysis**
- Kinetic modeling of simple networks
- Constraint-based modeling of cellular metabolism

So what do you do once you have peaks/expression calls/etc.?

- Direct inspection of known biological targets
- Literature-driven inference and hypothesis generation
- Gene set enrichment analysis
- Motif analysis
- Network inference

Gene set enrichment analysis

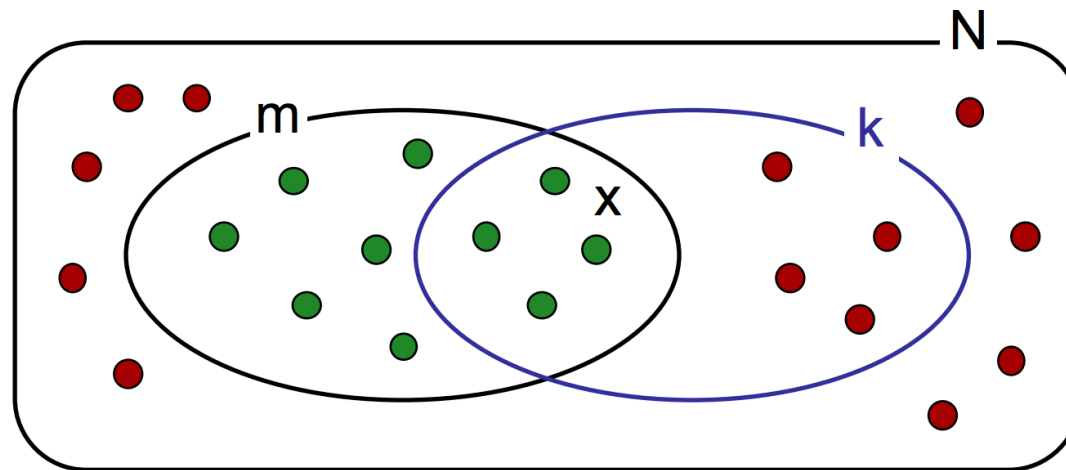
Identification of gene categories (e.g., GO terms) that are correlated with another data set

Common Tools: GSEA, DAVID, iPAGE

Gene set enrichment analysis

Identification of gene categories (e.g., GO terms) that are correlated with another data set

Common Tools: GSEA, DAVID, iPAGE



N = total number of elements

m = number of marked elements

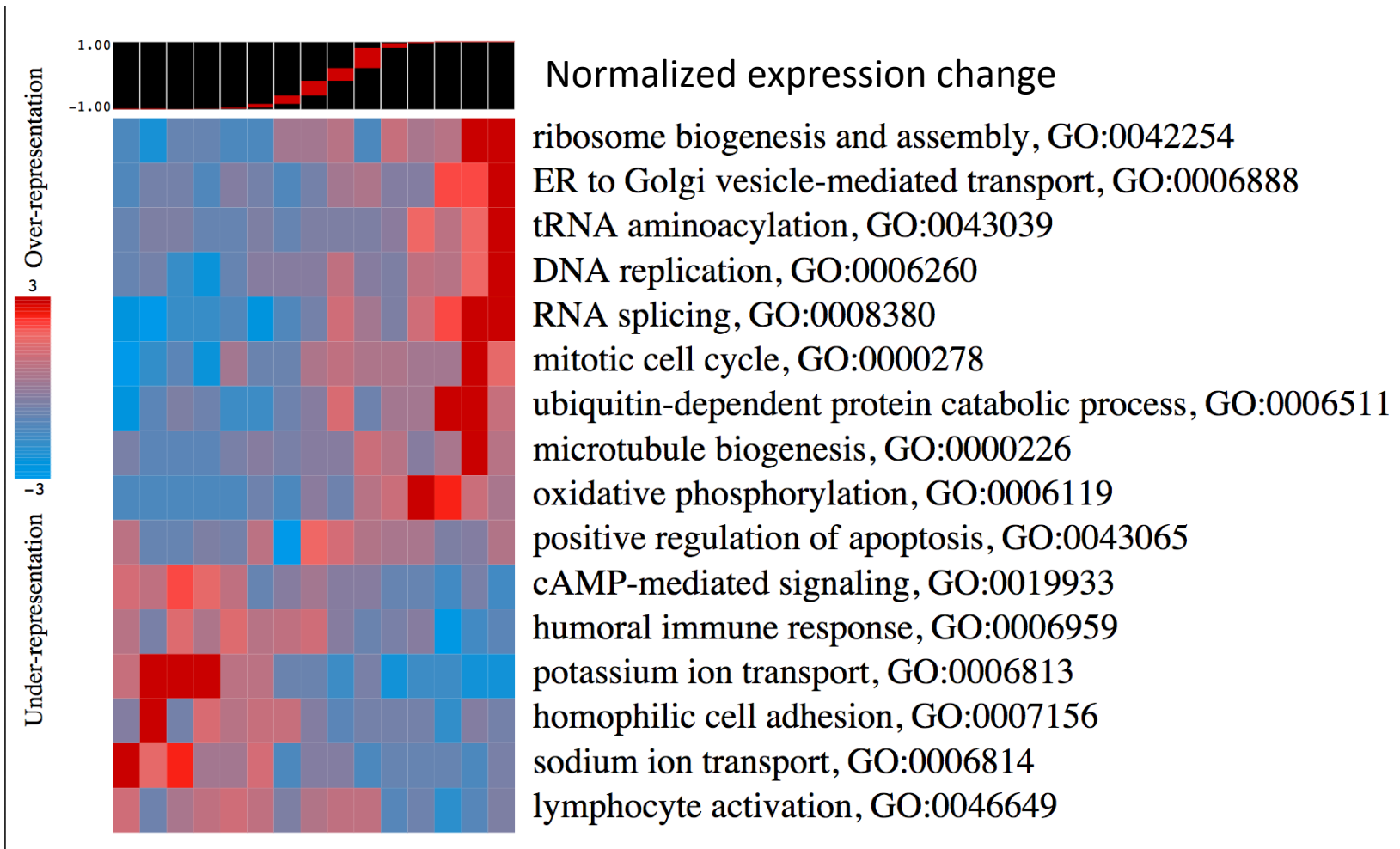
k = number of sampled elements

x = number of marked sampled elements

Gene set enrichment analysis

Identification of gene categories (e.g., GO terms) that are correlated with another data set

Example: Gene expression

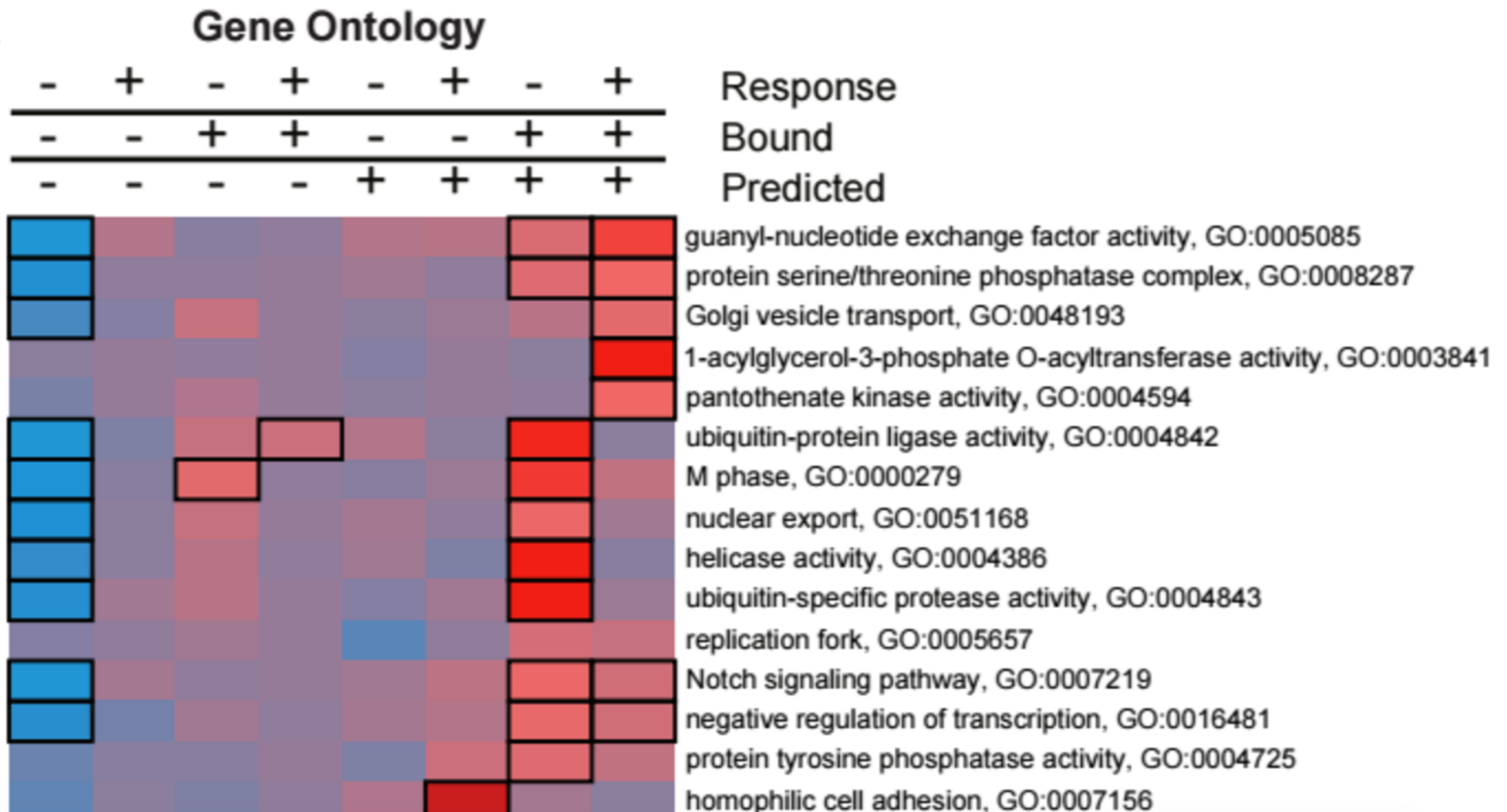


(From Goodarzi et al., Mol. Cell, 2009)

Gene set enrichment analysis

Identification of gene categories (e.g., GO terms) that are correlated with another data set

Example: Integration of data sets



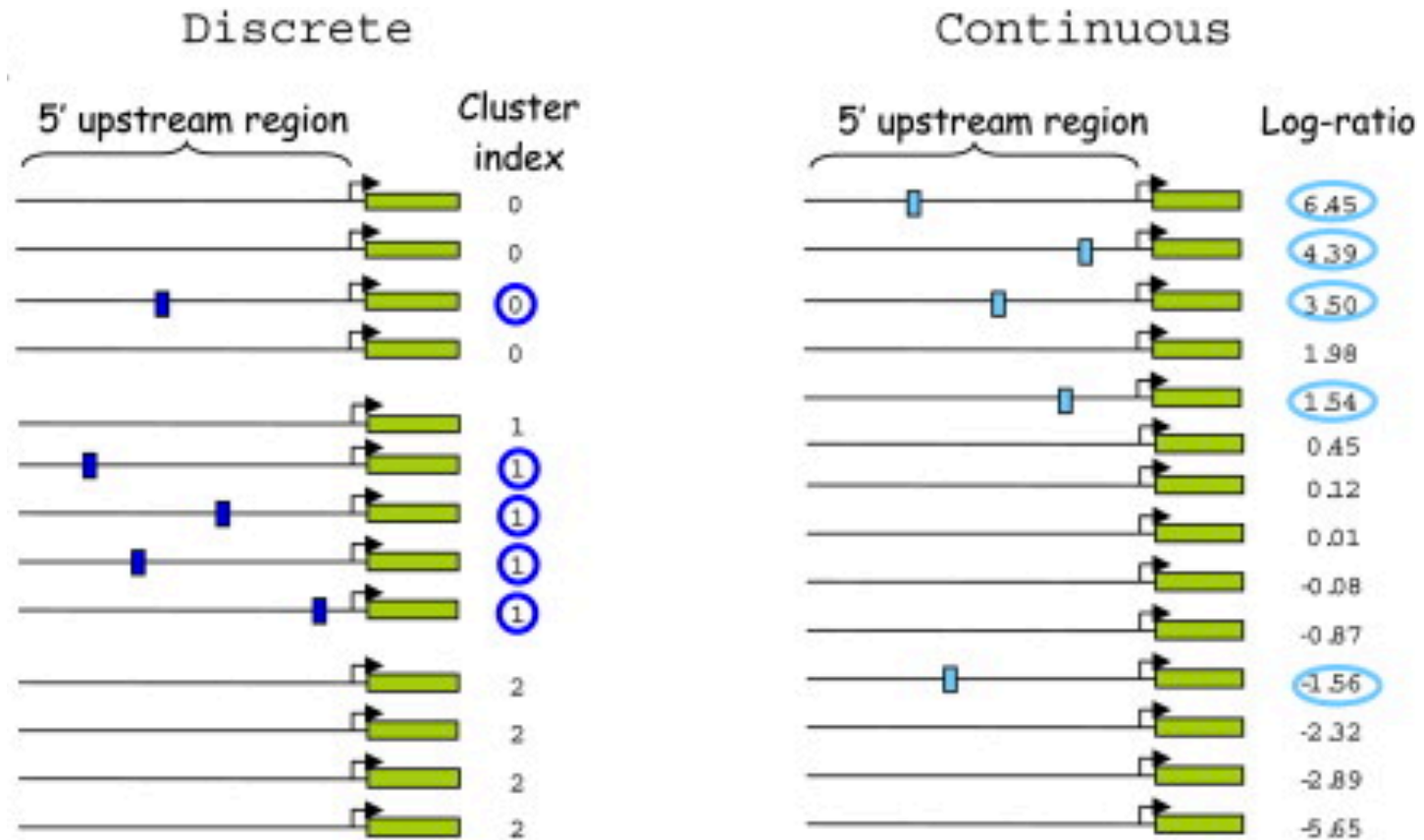
Motif analysis

Identify motifs (typically nucleic acid sequences) correlated with a data set of interest

Used in a variety of applications (RNA-seq, ChIP-seq, ribosome profiling, etc.)

Example tools: MEME suite, FIRE/TEISER, kmersvm

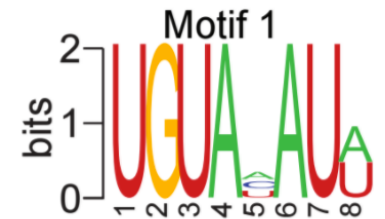
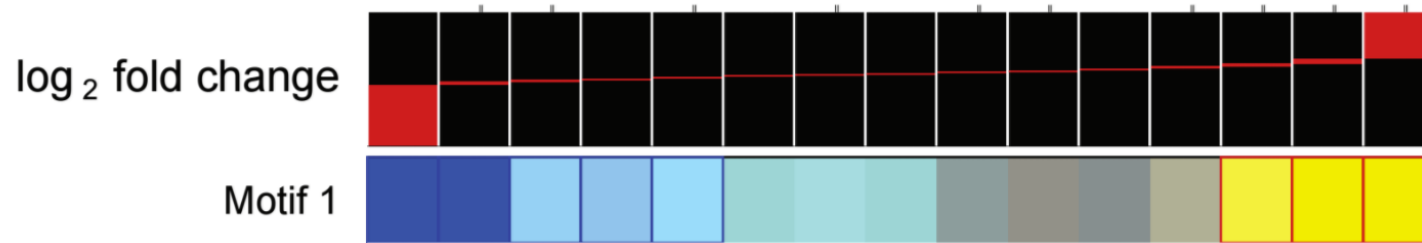
Motif analysis



(Image from Elemento *et al.*, Mol. Cell 2007)

Motif analysis

Inferred from knockdown:



From PAR-CLIP:



(Bottom adapted from Hafner, Cell 2010)

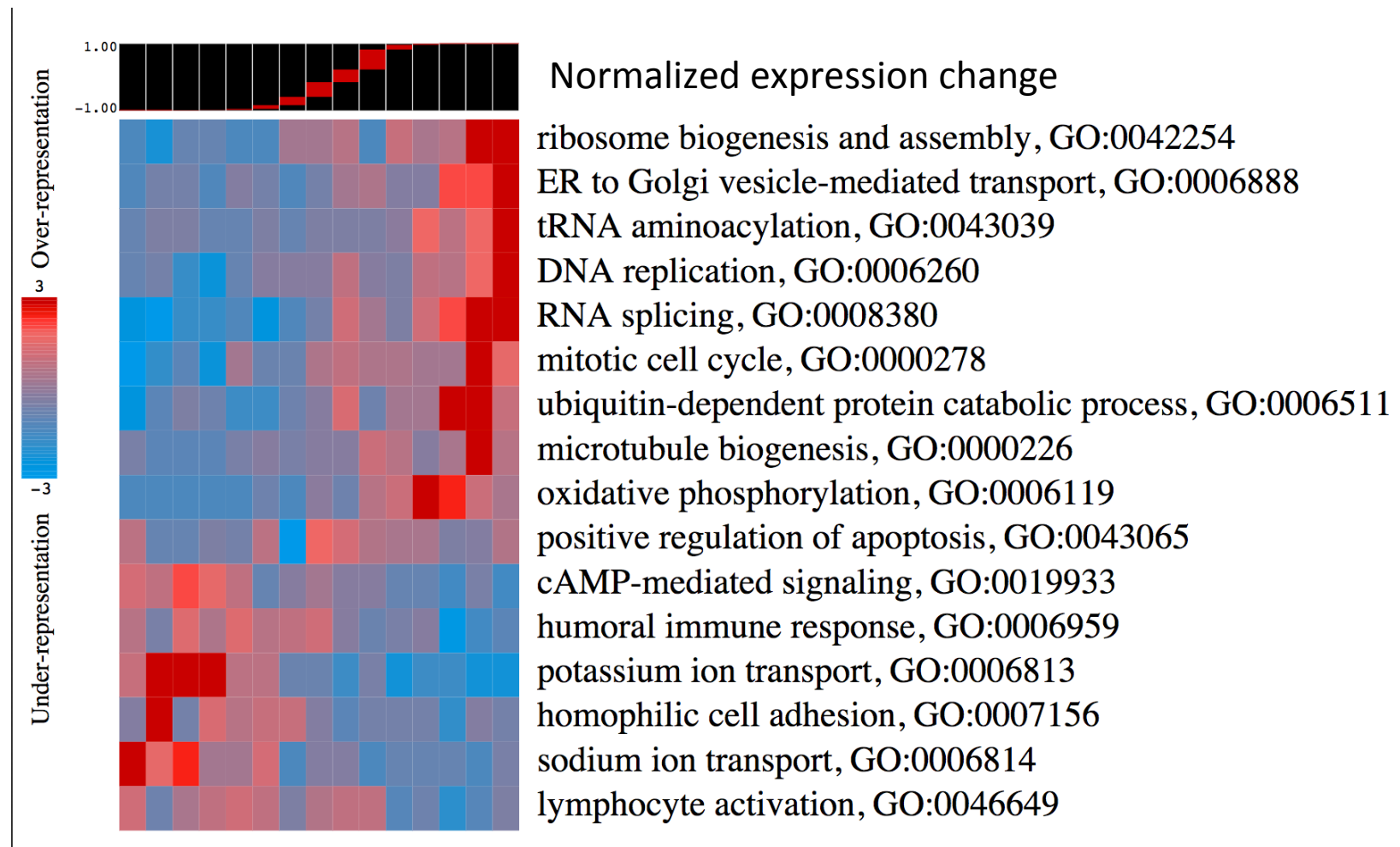
Interpretation of GO term/motif analysis

- Where possible find consensus from multiple programs
- Use as a starting point for more experiments (hypothesis **generating** tool)
- Keep in mind often high false discovery rates
- Look in detail at constituents giving rise to observations

Gene set enrichment analysis

Identification of gene categories (e.g., GO terms) that are correlated with another data set

Example: Gene expression



(From Goodarzi et al., Mol. Cell, 2009)

Building motifs into networks


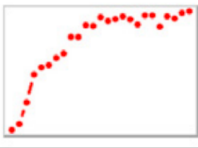
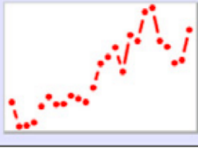
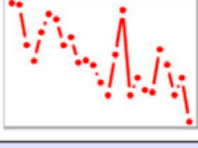
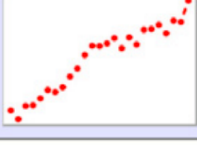
ARMADA: Inference from time courses

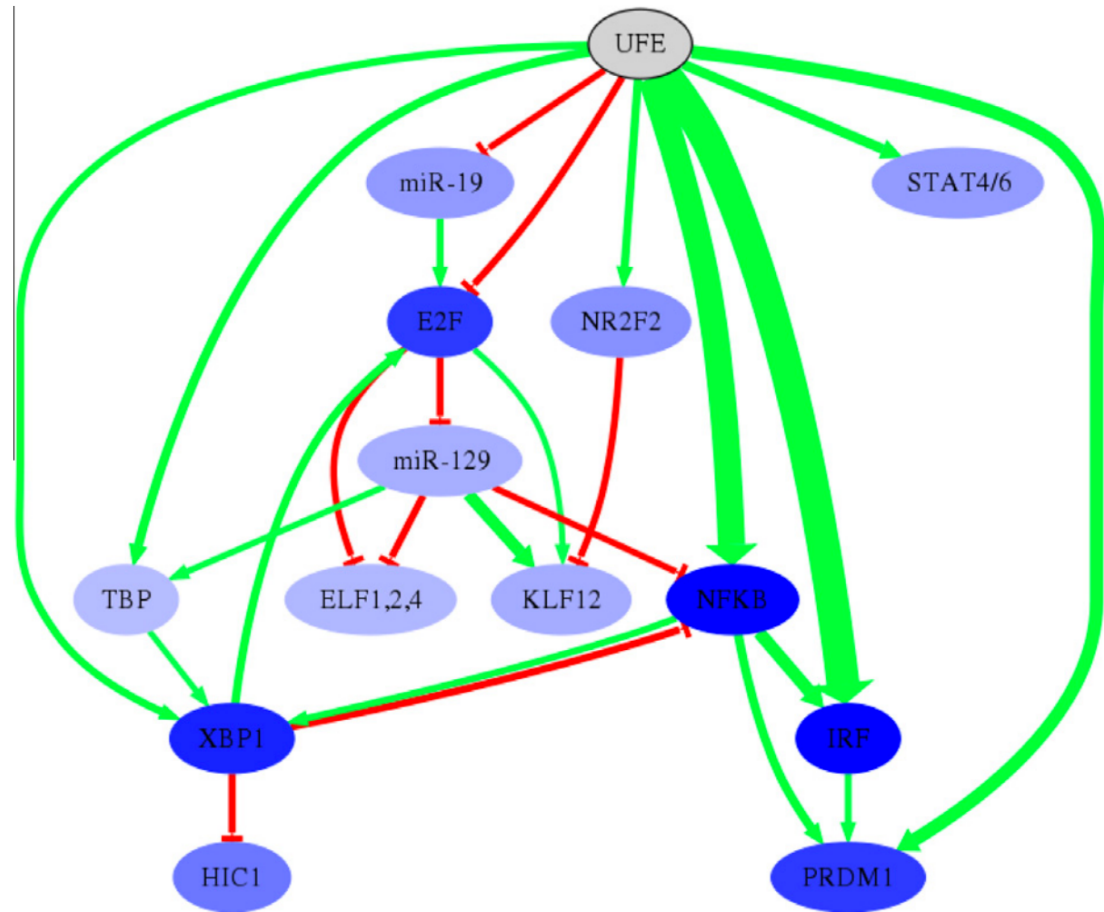
Motif name	Z-value	Associated genes	Profile	Logo
IRF1.2.7.p3	10.966	IRF1 (MAR, IRF-1) IRF2 (IRF-2, DKFZp686F0244) IRF7 (IRF7A, IRF-7H)		
NFKB1_REL_REL.p2	6.611	RELA (p65) REL (I-Rel) NFKB1 (KBF1, p105)		
XBP1.p3	3.173	XBP1		
E2F1..5.p2	2.979	E2F4 (E2F-4) E2F5 E2F2 (E2F-2) E2F1 (RBP3) E2F3		
PRDM1.p3	2.973	PRDM1 (PRDI-BF1)		

(Pemberton-Ross et al., Methods, 2015)

Building motifs into networks

ARMADA: Inference from time courses

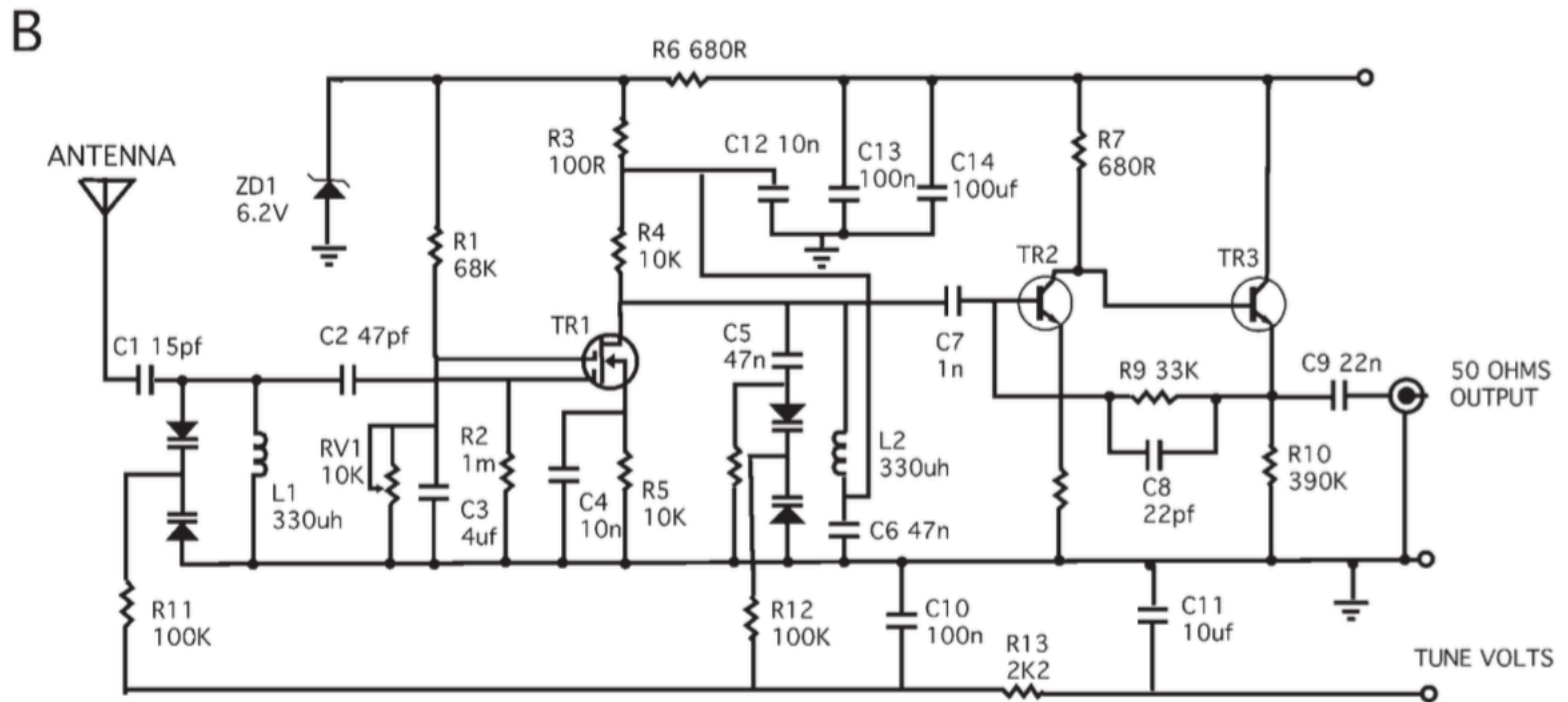
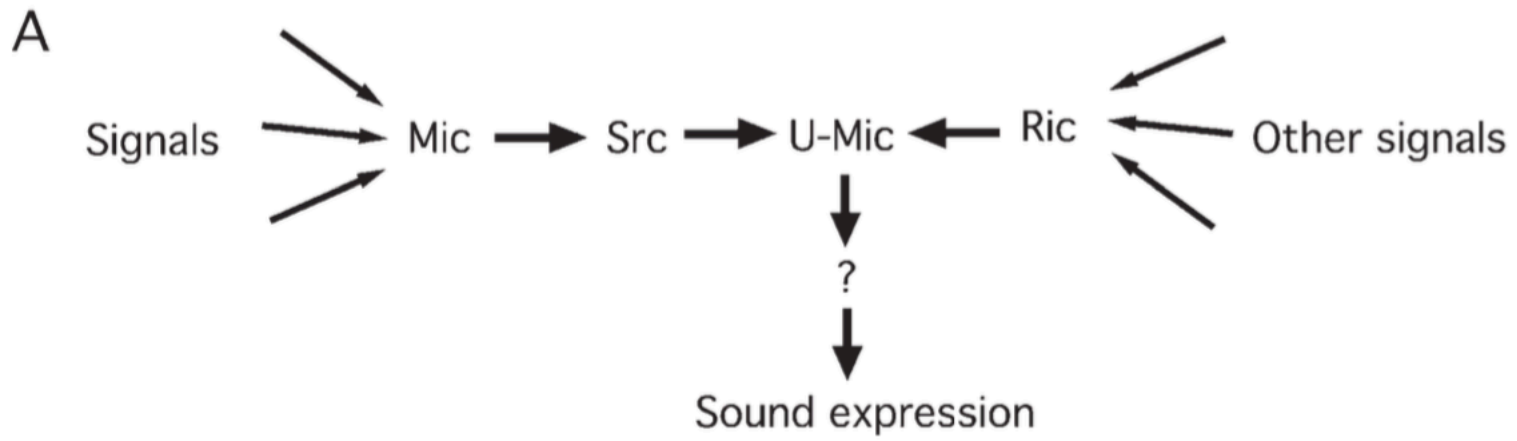
Motif name	Profile
IRF1.2.7.p3	
NFKB1_REL_REL.A.p2	
XBP1.p3	
E2F1..5.p2	
PRDM1.p3	



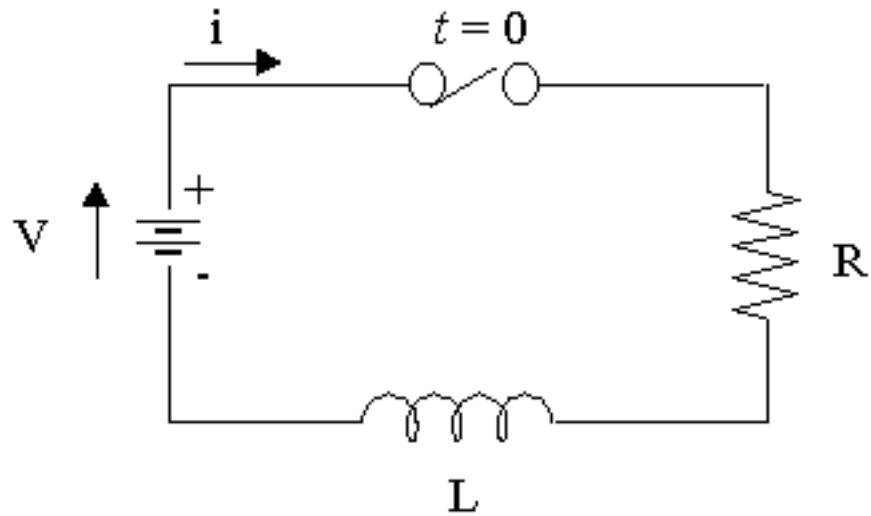
(Pemberton-Ross et al., Methods, 2015)

Building systems-level models

- Motif and GO term analysis
- **Kinetic modeling of simple networks**
- Constraint-based modeling of cellular metabolism

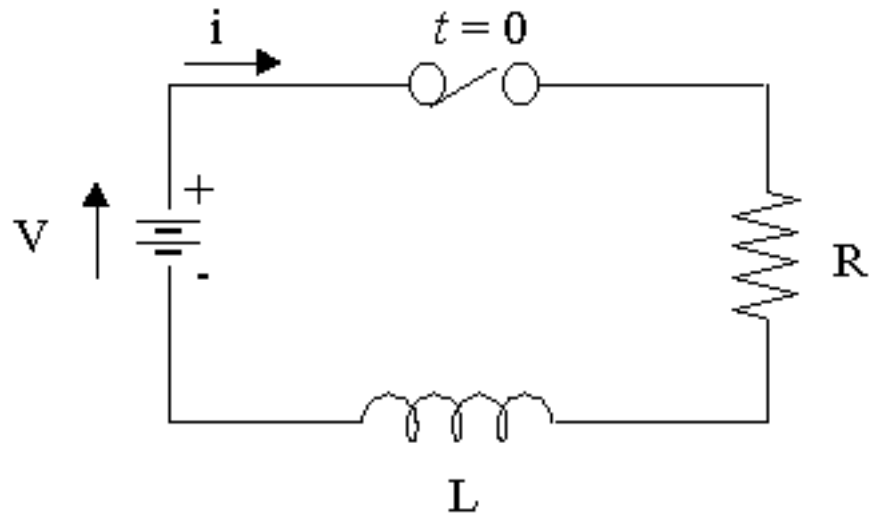


We know how to model electronic circuits



(Images from intmath.com)

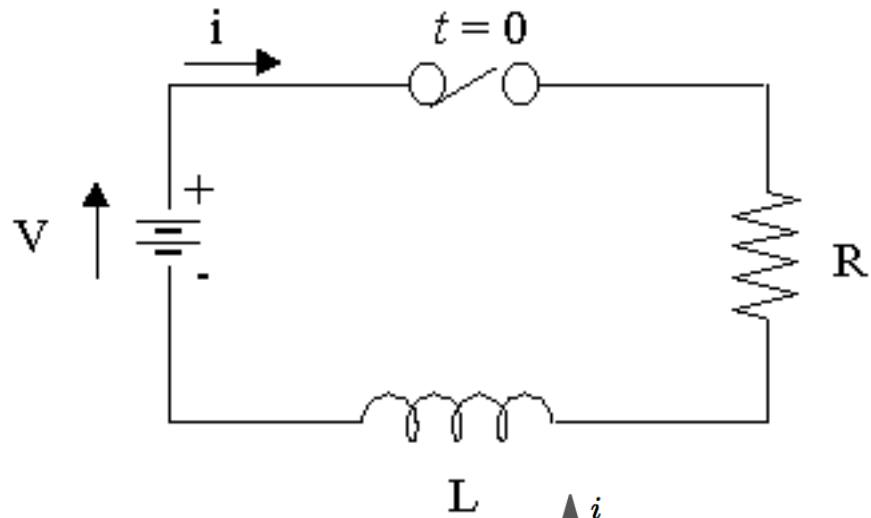
We know how to model electronic circuits



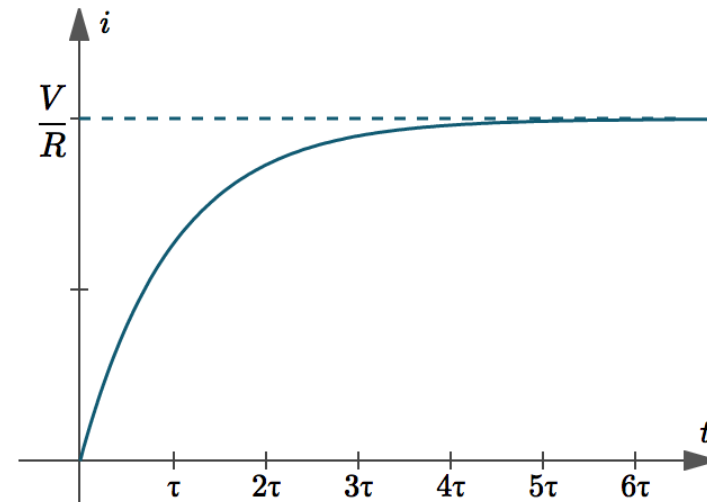
$$i = \frac{V}{R} \left(1 - e^{-(R/L)t} \right)$$

(Images from intmath.com)

We know how to model electronic circuits



$$i = \frac{V}{R} \left(1 - e^{-(R/L)t} \right)$$



(Images from intmath.com)

Differential equations are the language of change

$$dx/dt = \dots$$

$$x' = \dots$$

$$\dot{x} = \dots$$

Differential equations are the language of change

$$dx/dt = \dots$$

$$x' = \dots$$

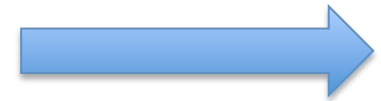
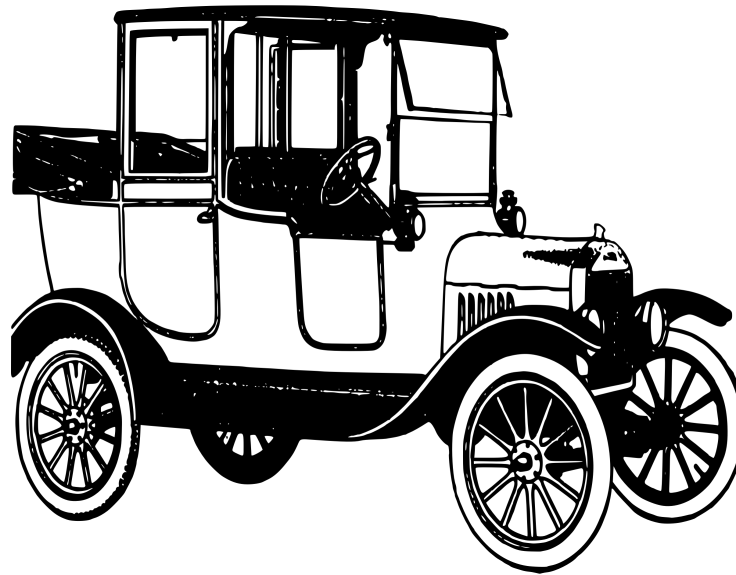
$$\dot{x} = \dots$$

All mean:

The rate of change in x per unit time is ...

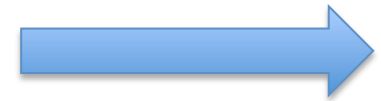
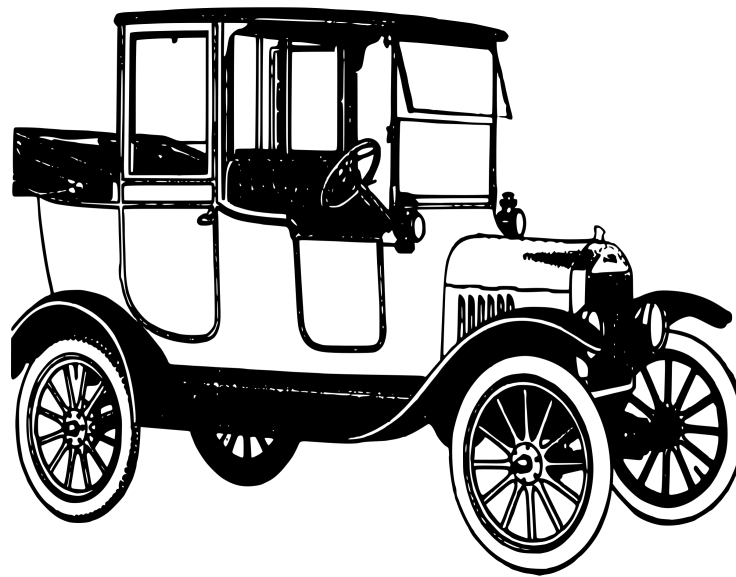
Example: A car moving at
constant speed

$$\frac{dx}{dt} = 10$$



Example: A car moving at
constant speed

$$\frac{dx}{dt} = 10$$



Our constant goal: What is $x(t)$?

Example: A car moving at
constant speed

$$\frac{dx}{dt} = 10$$

$$x(t) = 10t + C$$

Example: A car moving at constant speed

$$\frac{dx}{dt} = 10$$

- Start at some position – we'll call it C
- Every second, add to the current position the distance that the car travels in one second
- Keep doing that until we reach the time that we are interested in

Example: A car moving at constant speed

$$\frac{dx}{dt} = 10$$

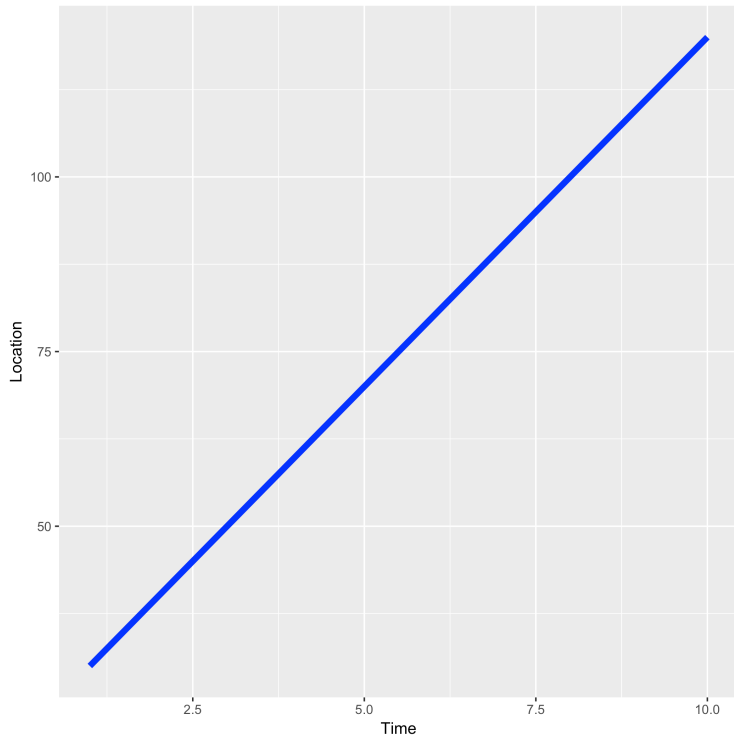
```
carloc <- function(t,v,C)
{
  x.curr <- C
  t.curr <- 0
  while (t.curr < t) {
    t.curr <- t.curr+1
    x.curr <- x.curr + v
  }

  return(x.curr)
}
```

Example: A car moving at constant speed

```
carloc <- function(t,v,C)
{
  x.curr <- C
  t.curr <- 0
  while (t.curr < t) {
    t.curr <- t.curr+1
    x.curr <- x.curr + v
  }

  return(x.curr)
}
```



Example: A car moving at constant speed

$$\frac{dx}{dt} = 10$$

$$x(t) = 10t + C$$

```
carloc <- function(t,v,C)
{
  x.curr <- C
  t.curr <- 0
  while (t.curr < t) {
    t.curr <- t.curr+1
    x.curr <- x.curr + v
  }

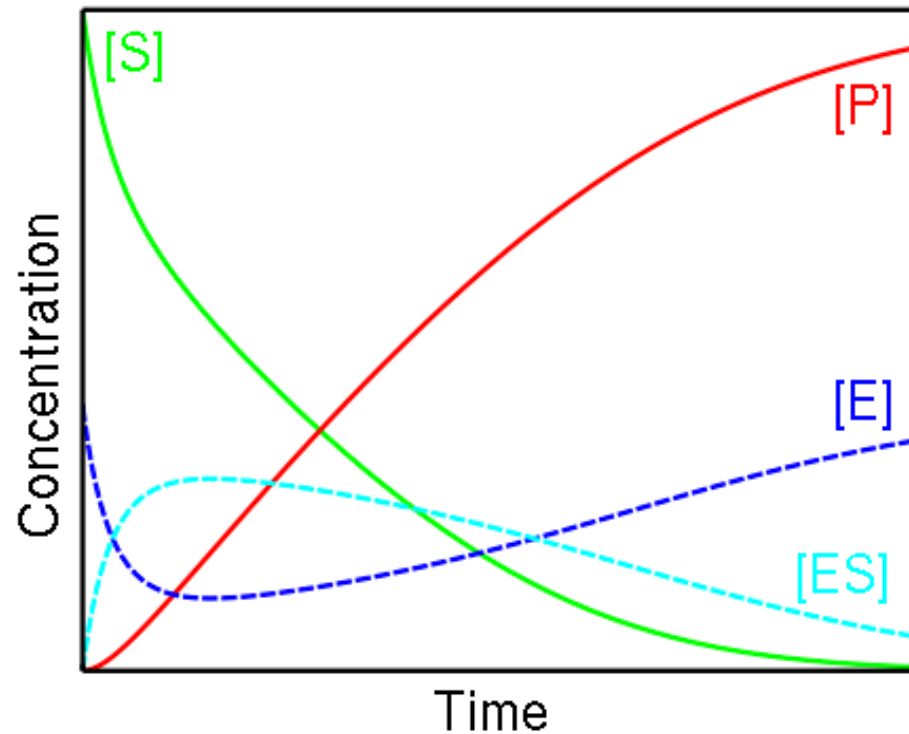
  return(x.curr)
}
```

Analytical

Numerical

Differential equations are the language of biology

$$v = \frac{d[P]}{dt} = \frac{V_{\max} [S]}{K_M + [S]}$$



(Images from wikipedia)

Two common motifs

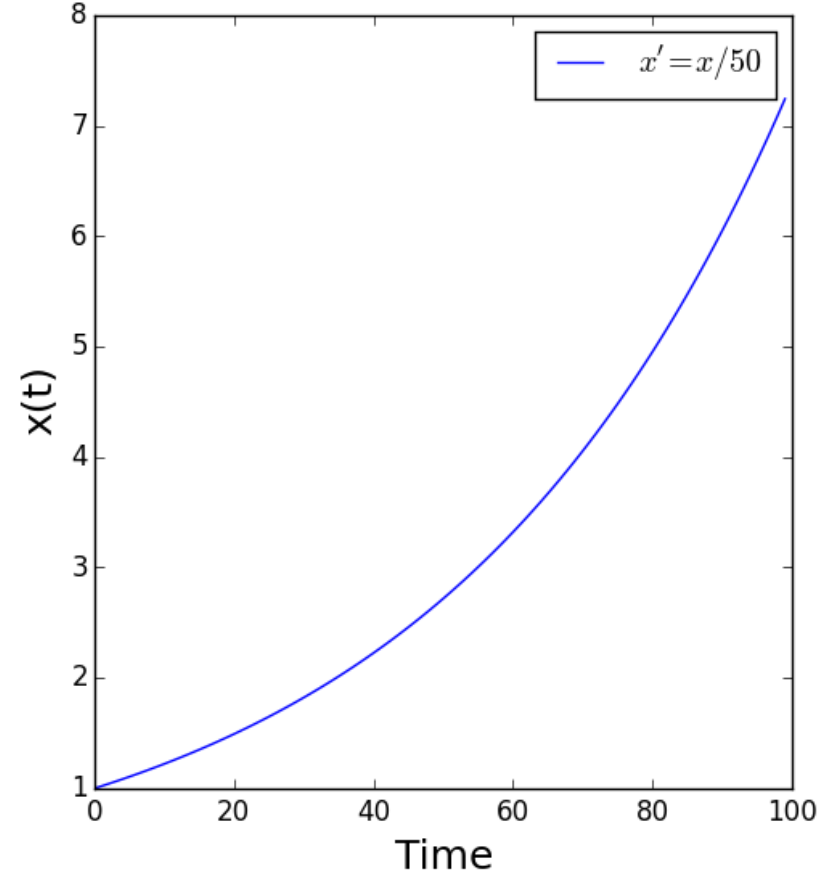
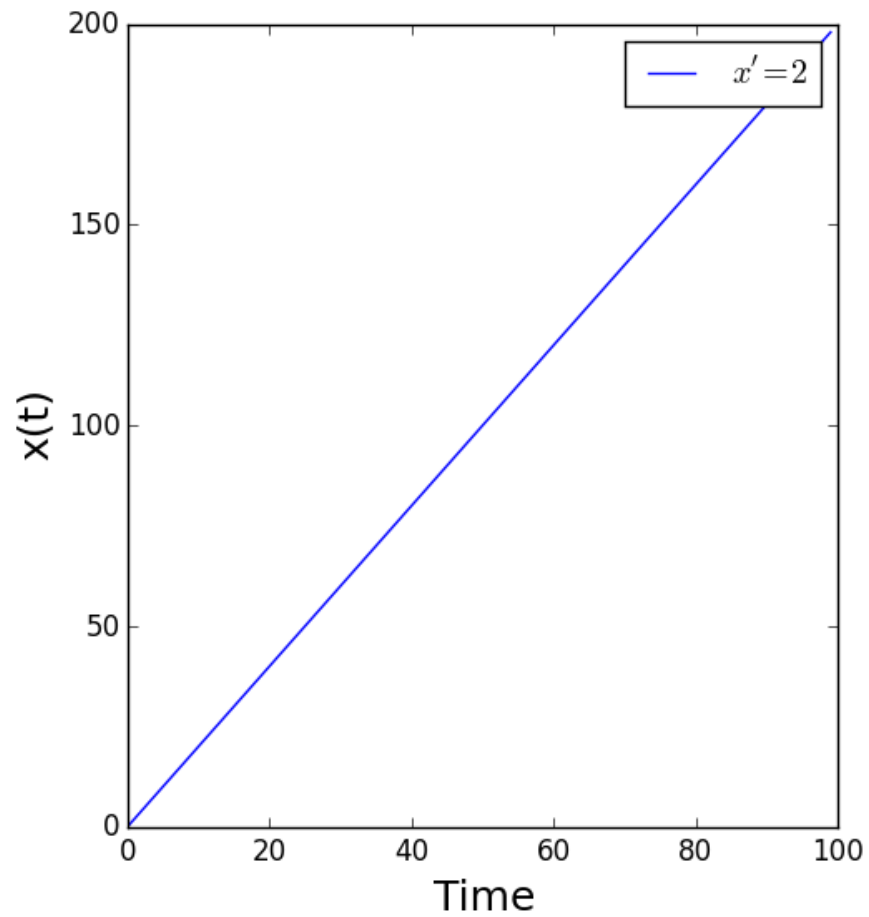
$$x' = k$$

$$x' = kx$$

Two common motifs

$$x' = k$$

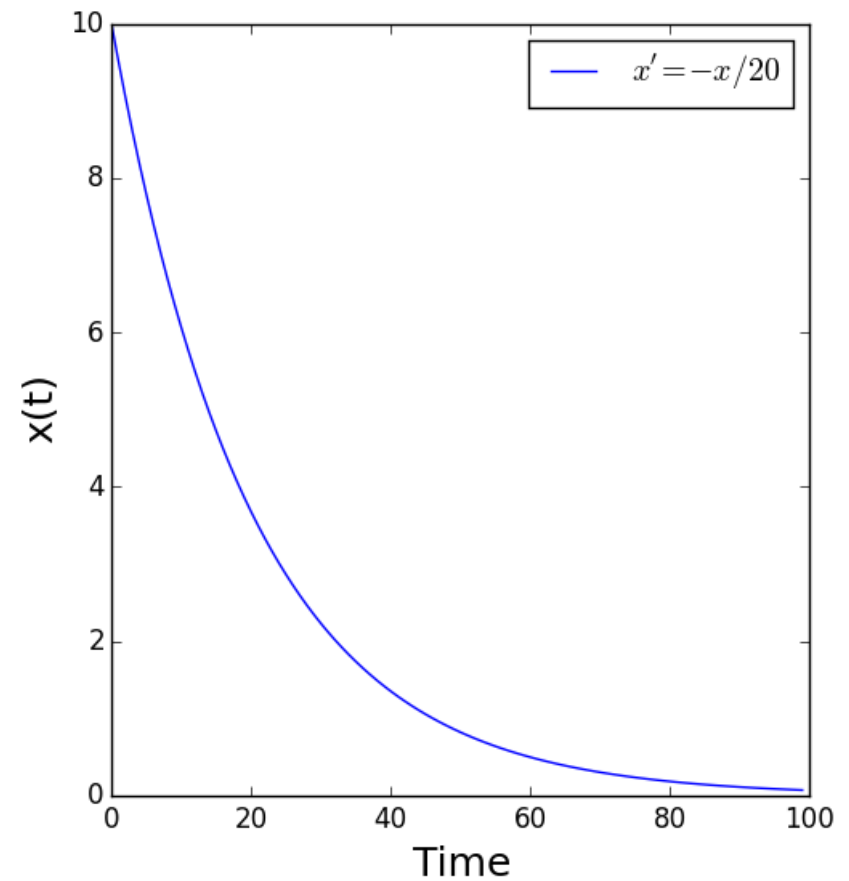
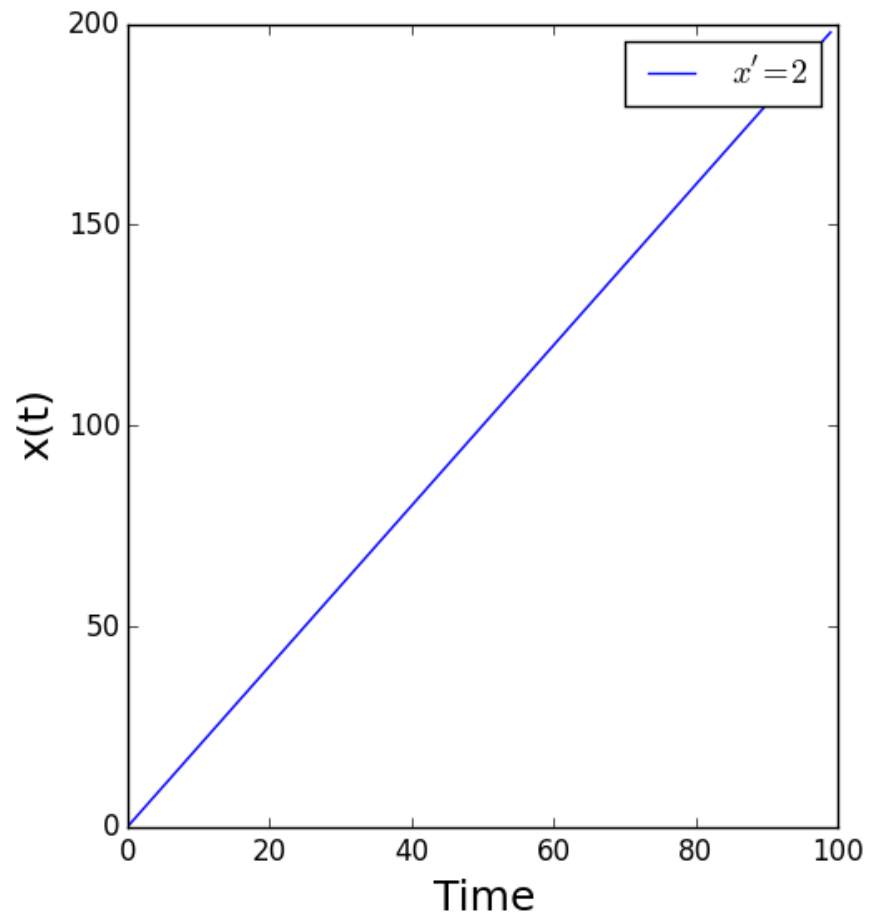
$$x' = kx$$



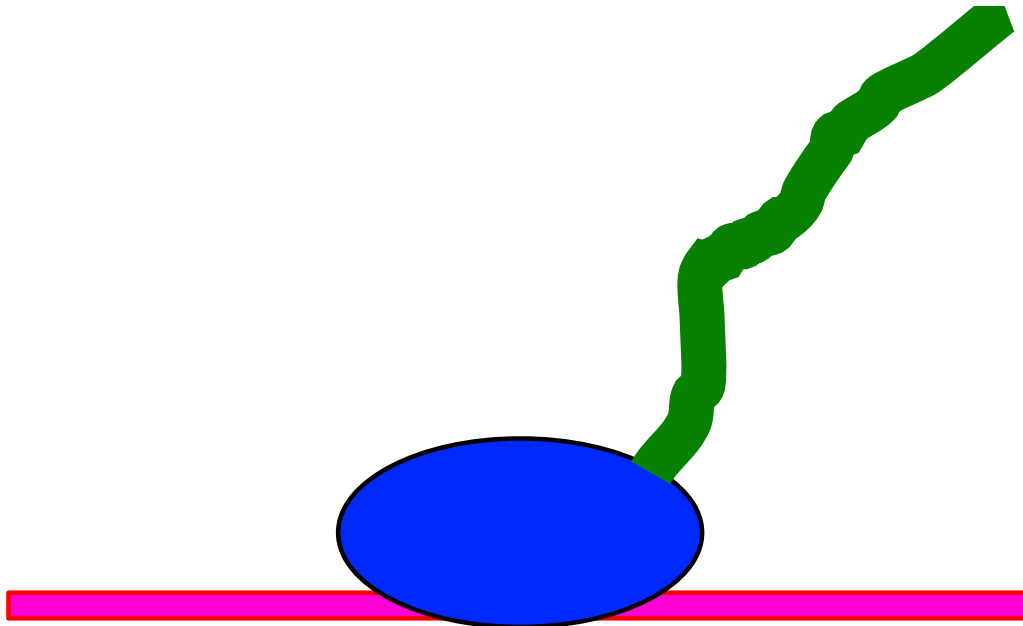
Two common motifs

$$x' = k$$

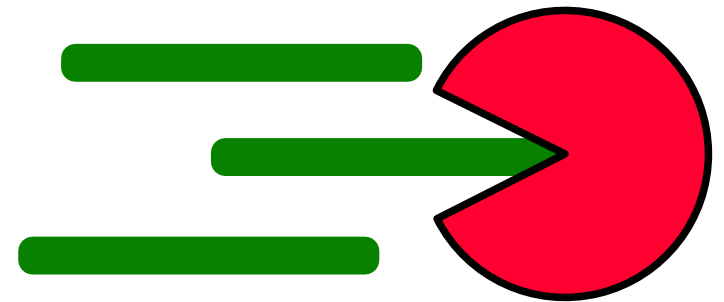
$$x' = -kx$$



A day in the life of a transcript

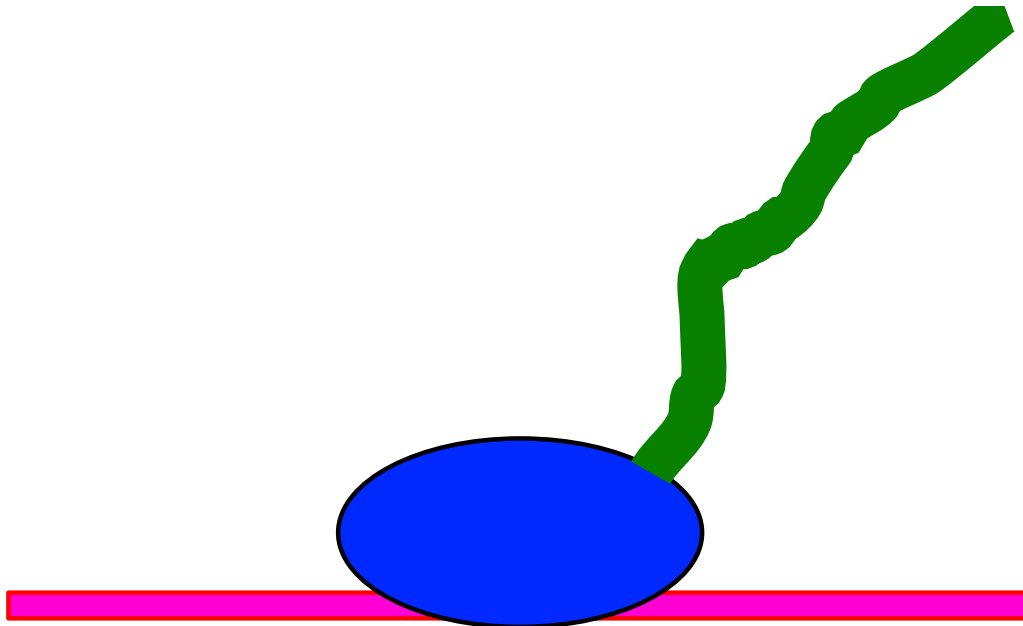


Synthesis

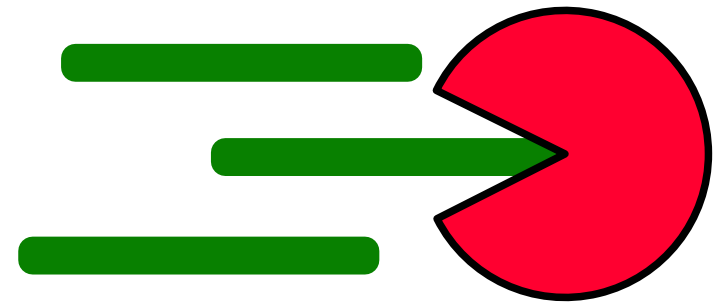


Degredation

A day in the life of a transcript



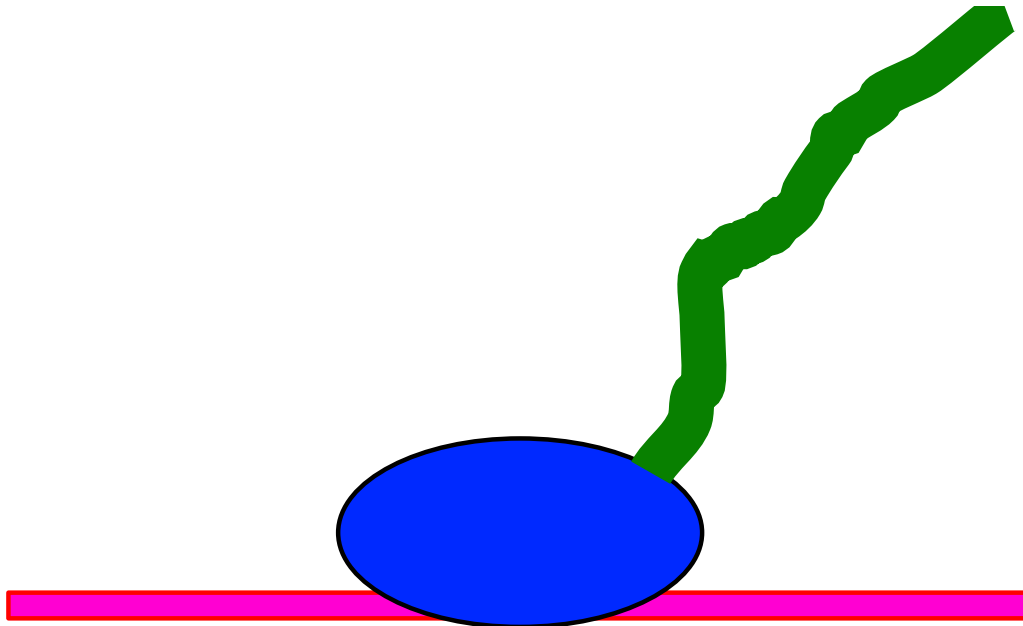
Synthesis



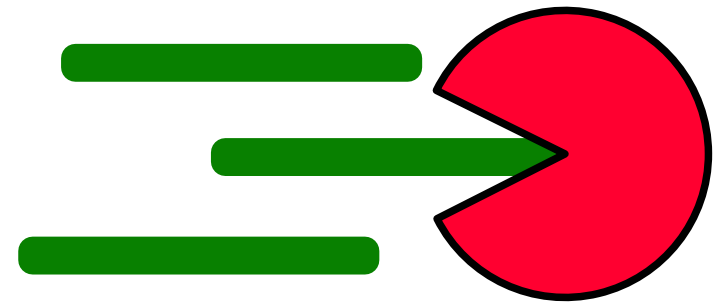
Degredation

How can we model the level of the transcript at any given time?

A day in the life of a transcript



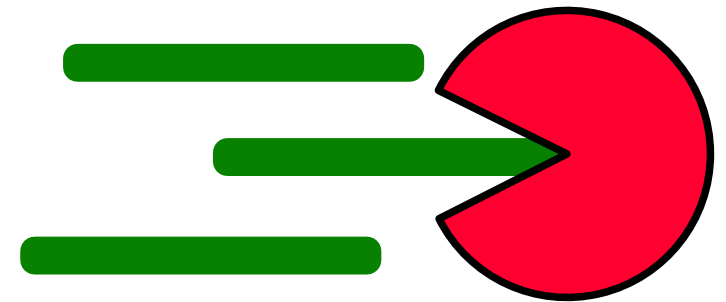
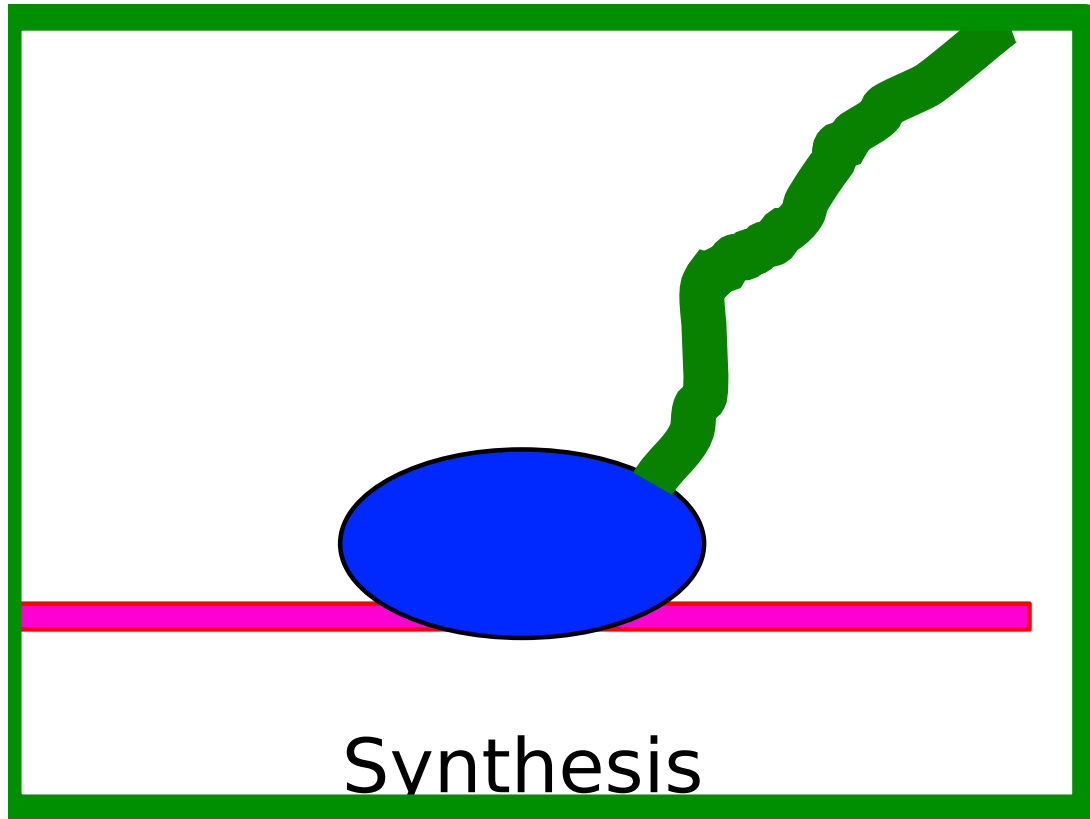
Synthesis



Degredation

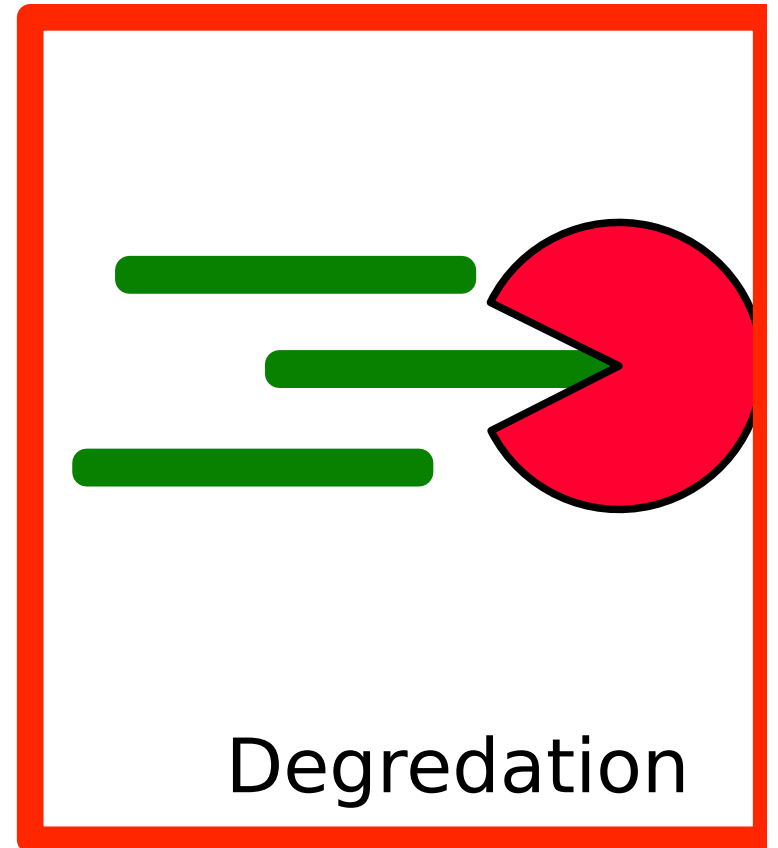
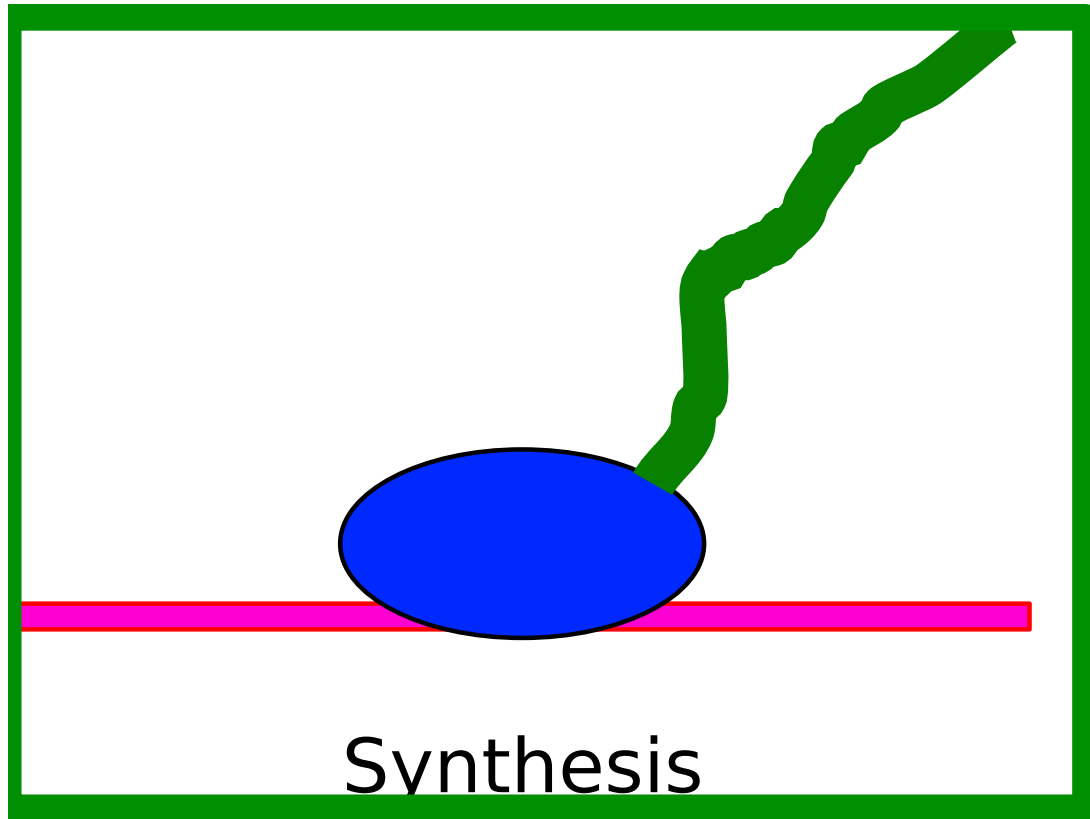
$$x' =$$

A day in the life of a transcript



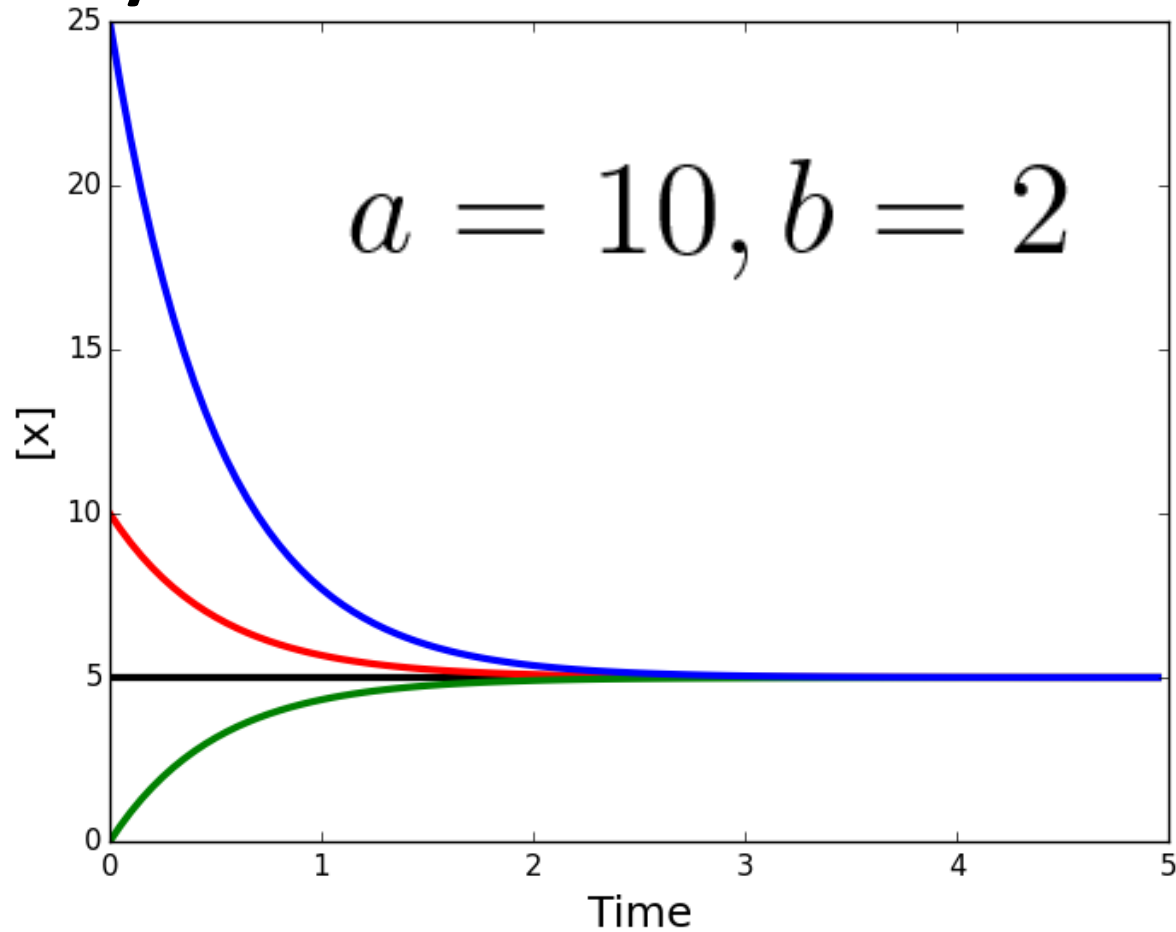
$$x' = \boxed{a}$$

A day in the life of a transcript



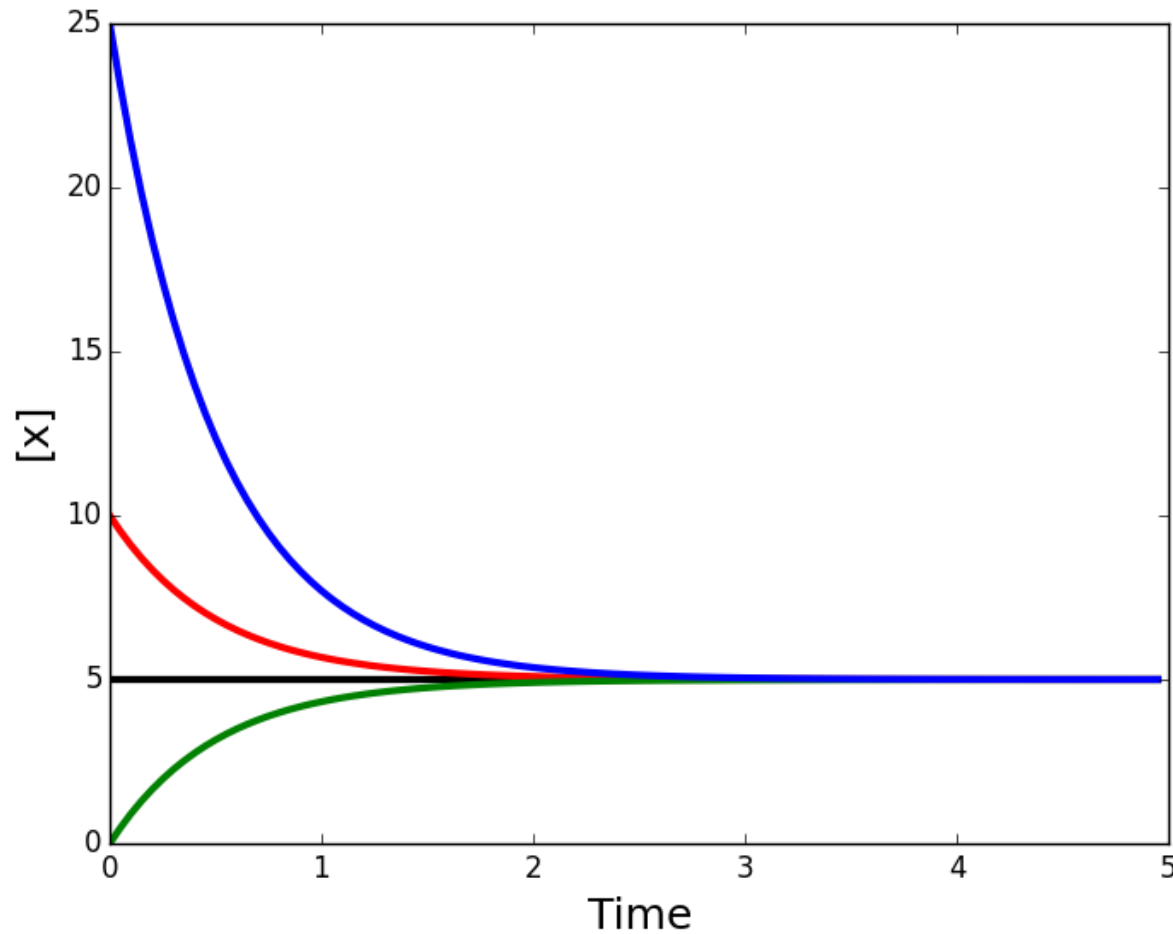
$$x' = a - bx$$

A day in the life of a transcript



$$x' = a - bx$$

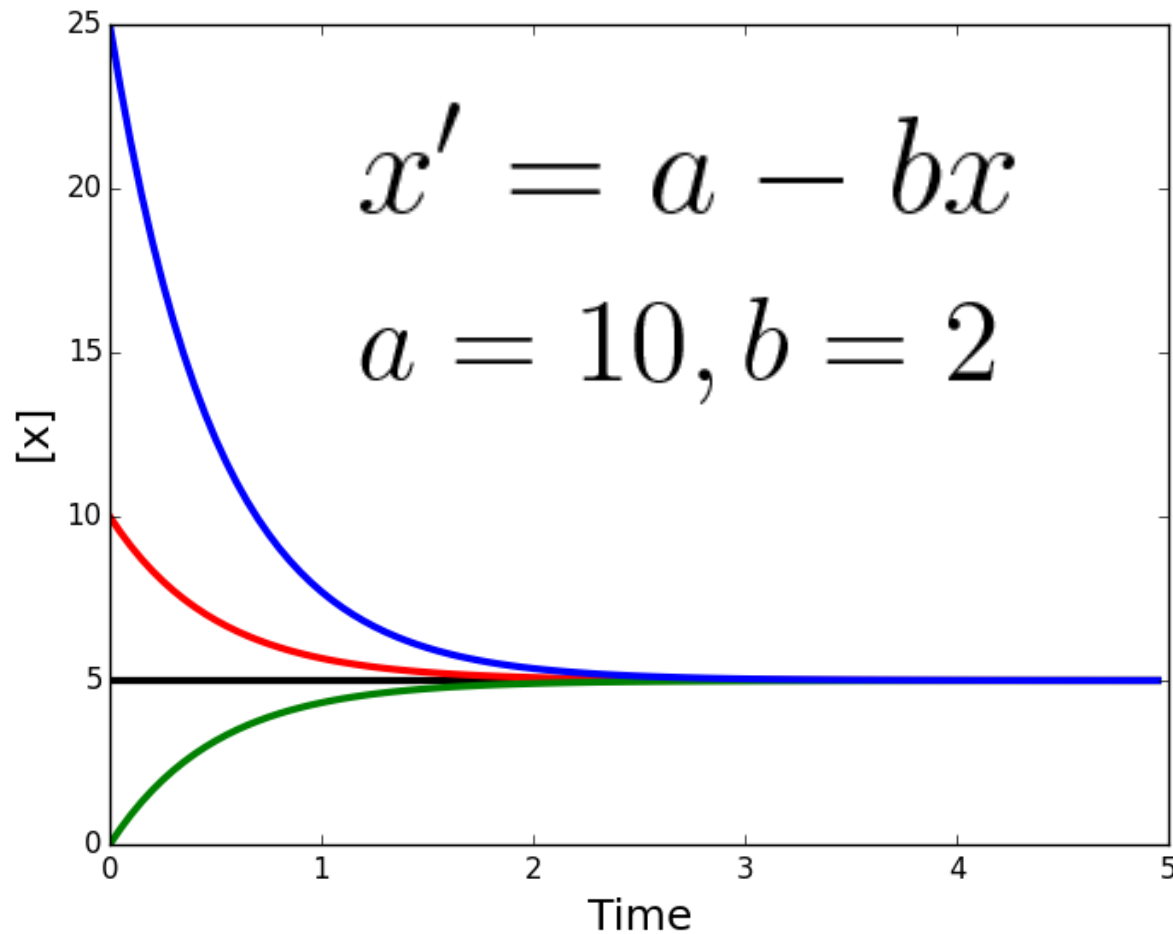
Stationary point analysis



Stationary point:
 $dx/dt = 0$ (for all variables)

Finding these gives
steady state values

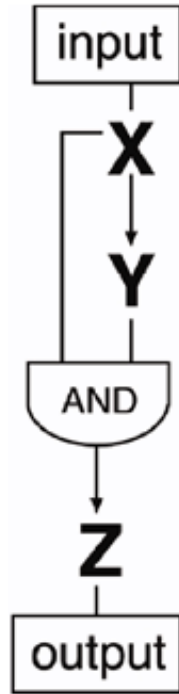
Stationary point analysis



Stationary point:
 $dx/dt = 0$ (for all variables)

Finding these gives
steady state values

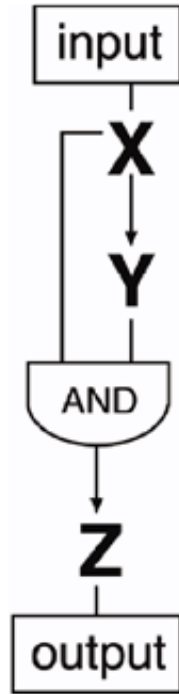
Application to feed forward loops



$$\frac{dY}{dt} = F(X, T_y) - \alpha Y$$

(Shen-Orr *et al.*, Nat. Gen. 2002)

Application to feed forward loops

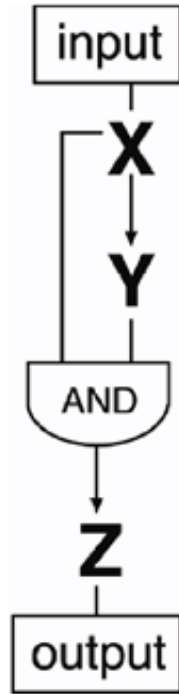


Rate of change in [Y] $\rightarrow \frac{dY}{dt} = F(X, T_y) - \alpha Y \leftarrow$ Degradation rate of Y

Threshold on X value

(Shen-Orr *et al.*, Nat. Gen. 2002)

Application to feed forward loops



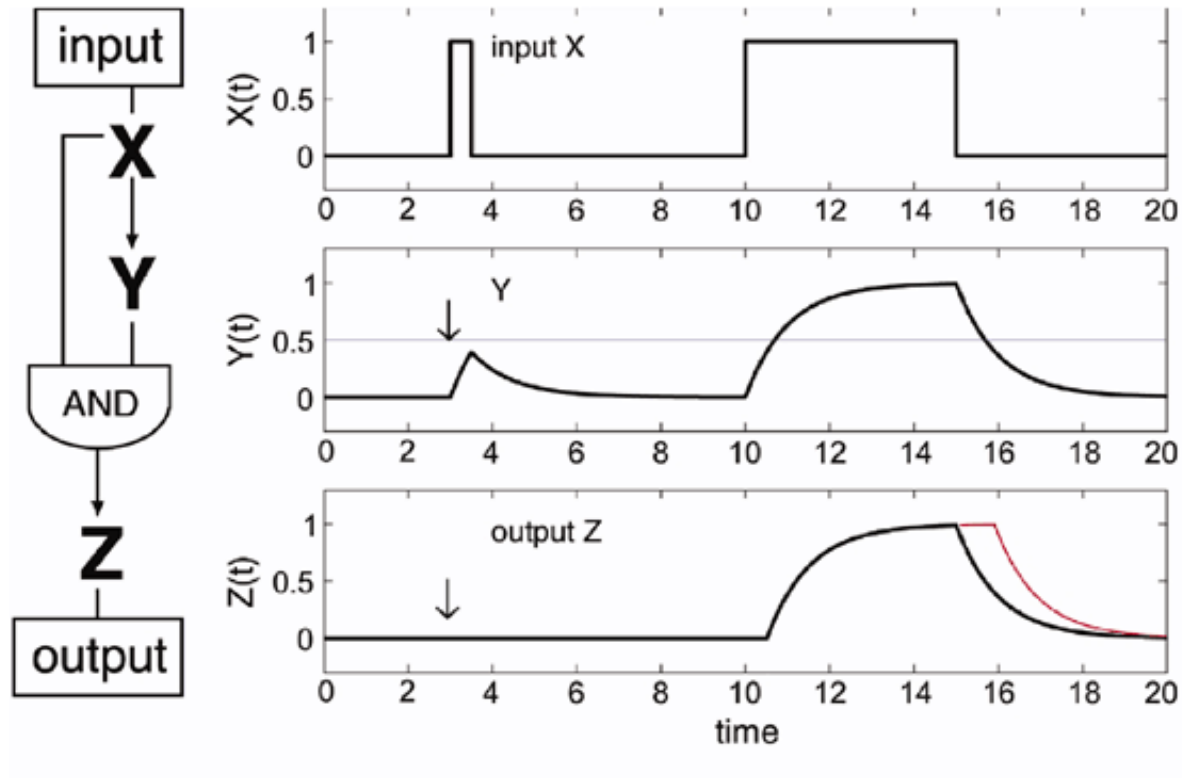
$$F(X, T_y) \equiv \begin{cases} 1 & \text{if } X > T_y; \\ 0 & \text{otherwise} \end{cases}$$

Rate of change in [Y] $\rightarrow \frac{dY}{dt} = F(X, T_y) - \alpha Y \leftarrow$ Degradation rate of Y

Threshold on X value

(Shen-Orr *et al.*, Nat. Gen. 2002)

Application to feed forward loops



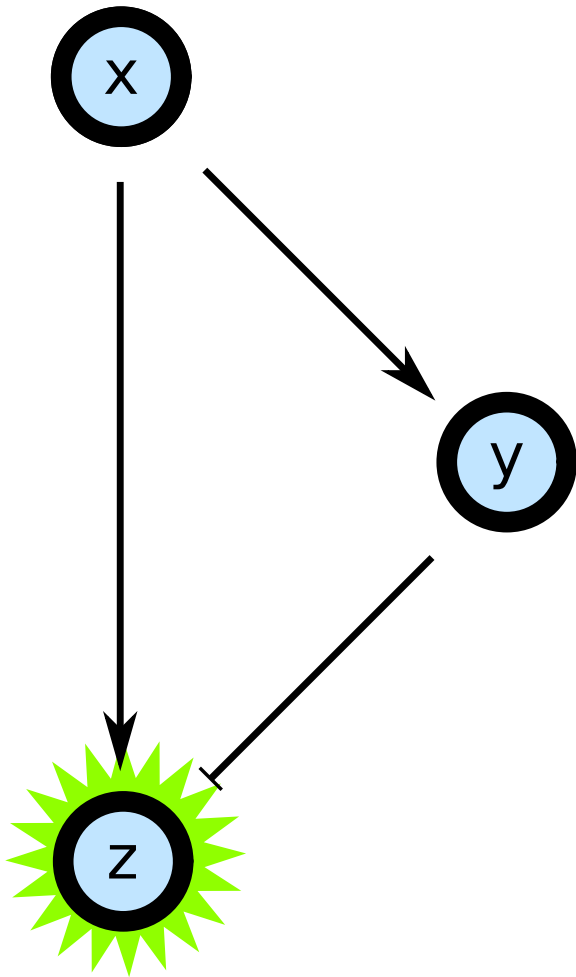
$$\frac{dY}{dt} = F(X, T_y) - \alpha Y$$
$$\frac{dZ}{dt} = F(X, T_y)F(Y, T_z) - \alpha Z$$

(Shen-Orr *et al.*, Nat. Gen. 2002)

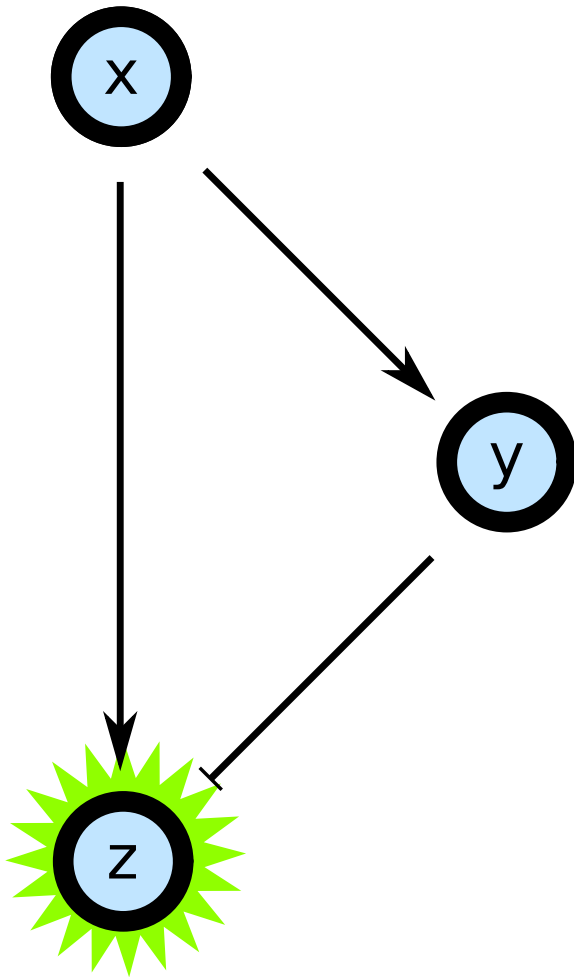
Can we test our answers from Lab 1?

Working with an assigned group of peers, design a BioBrick-based construct that would yield a transient burst of GFP expression when *E. coli* cells bearing the plasmid undergo cold shock (20 C) while growing in glucose minimal media. **Indicate the part numbers to be assembled (in order), draw a schematic of the resulting mini-network, and explain why your construct will implement the desired function.**

Common solutions

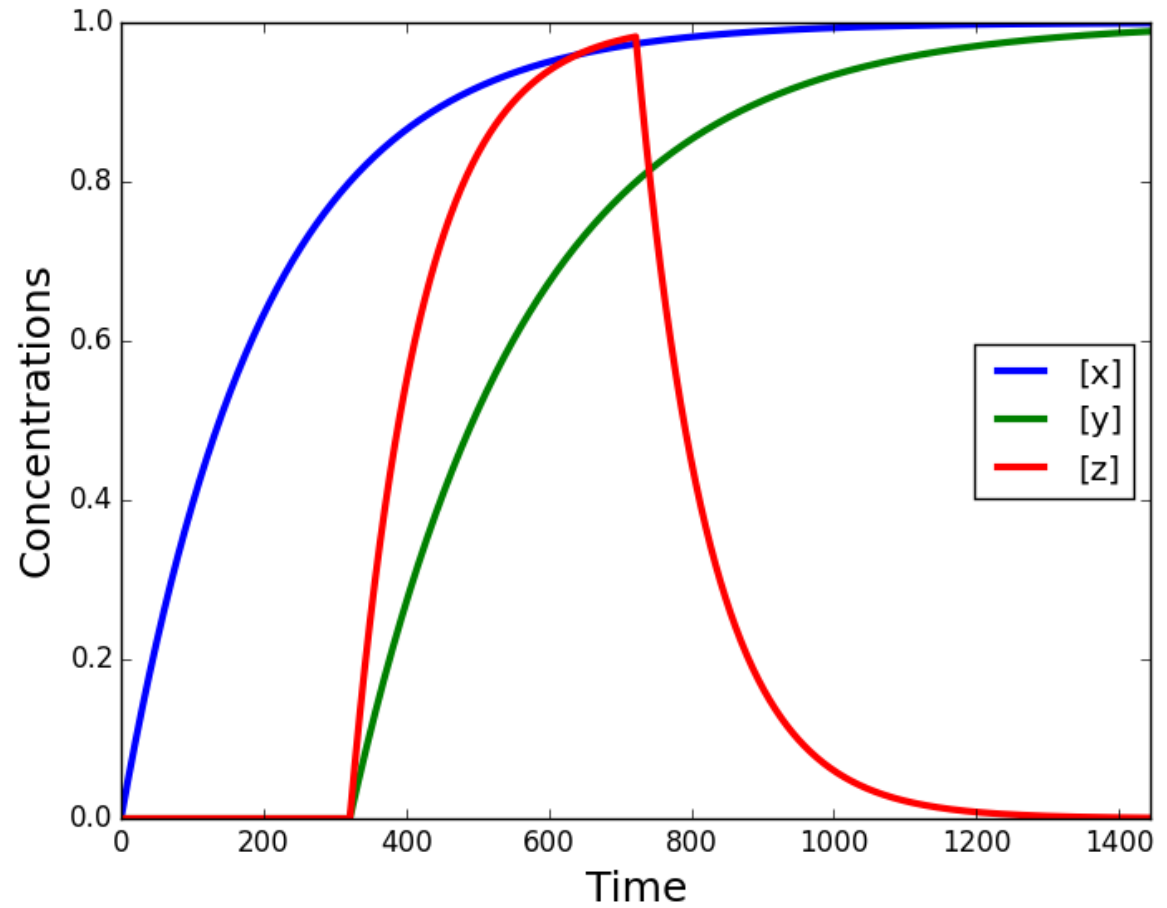
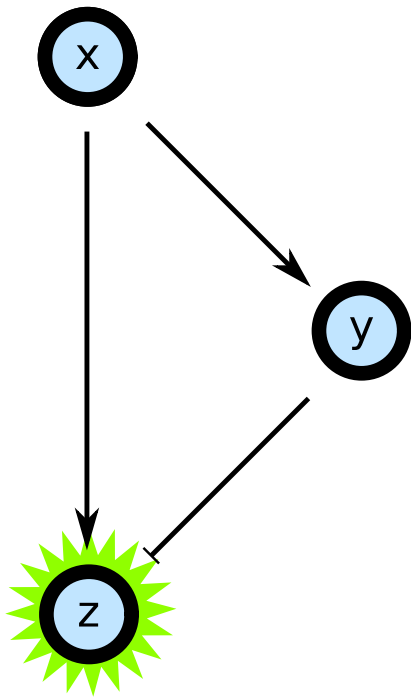


Common solutions

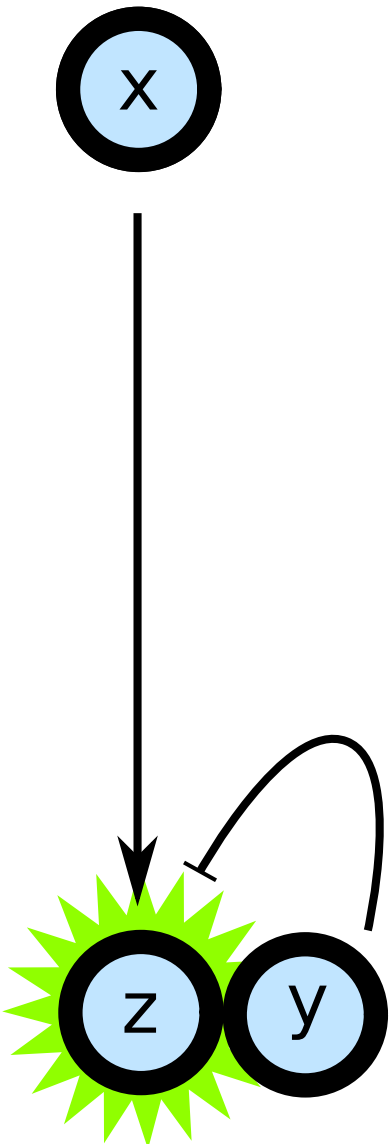


$$x' \equiv k_x - \alpha x$$
$$y' \equiv \begin{cases} -\alpha y & x < t_y \\ k_y - \alpha y & \text{otherwise} \end{cases}$$
$$z' \equiv \begin{cases} k_z - \alpha z & x > t_{z,x}; y < t_{z,y} \\ -\alpha z & \text{otherwise} \end{cases}$$

Common solutions



Common solutions

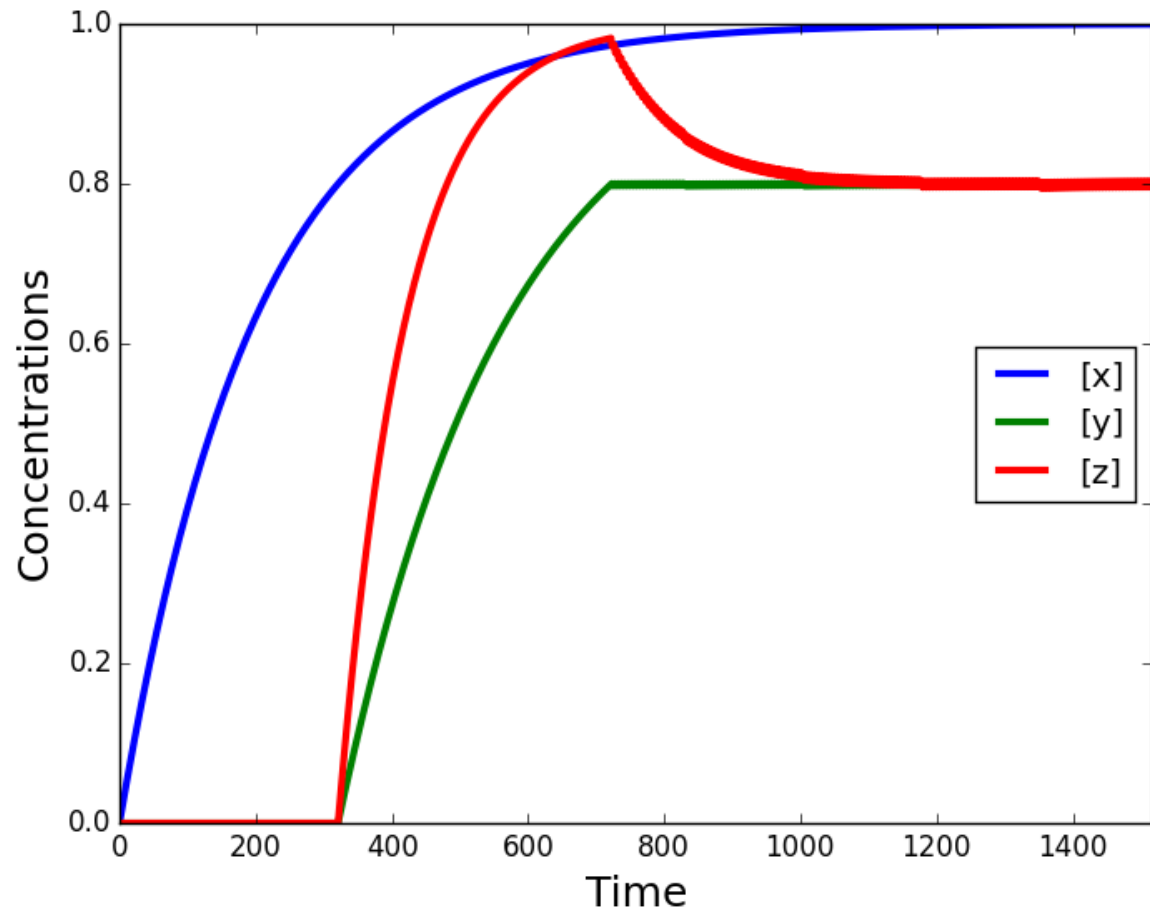
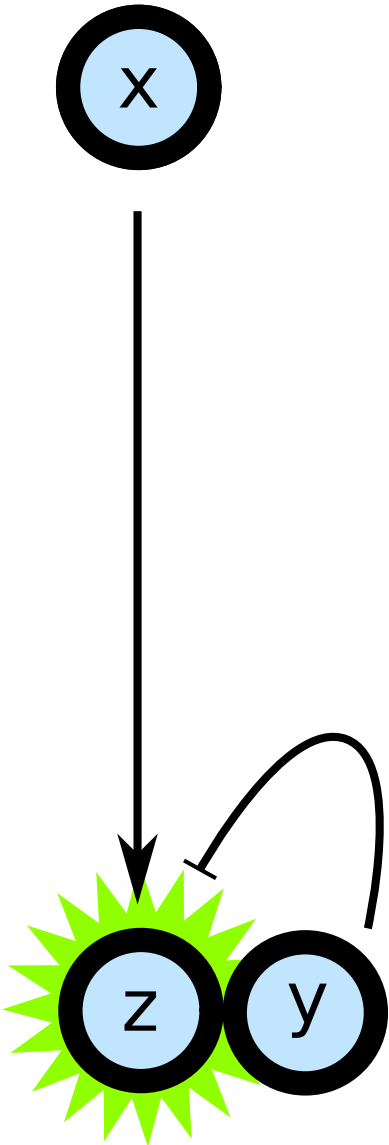


$$x' \equiv k_x - \alpha x$$

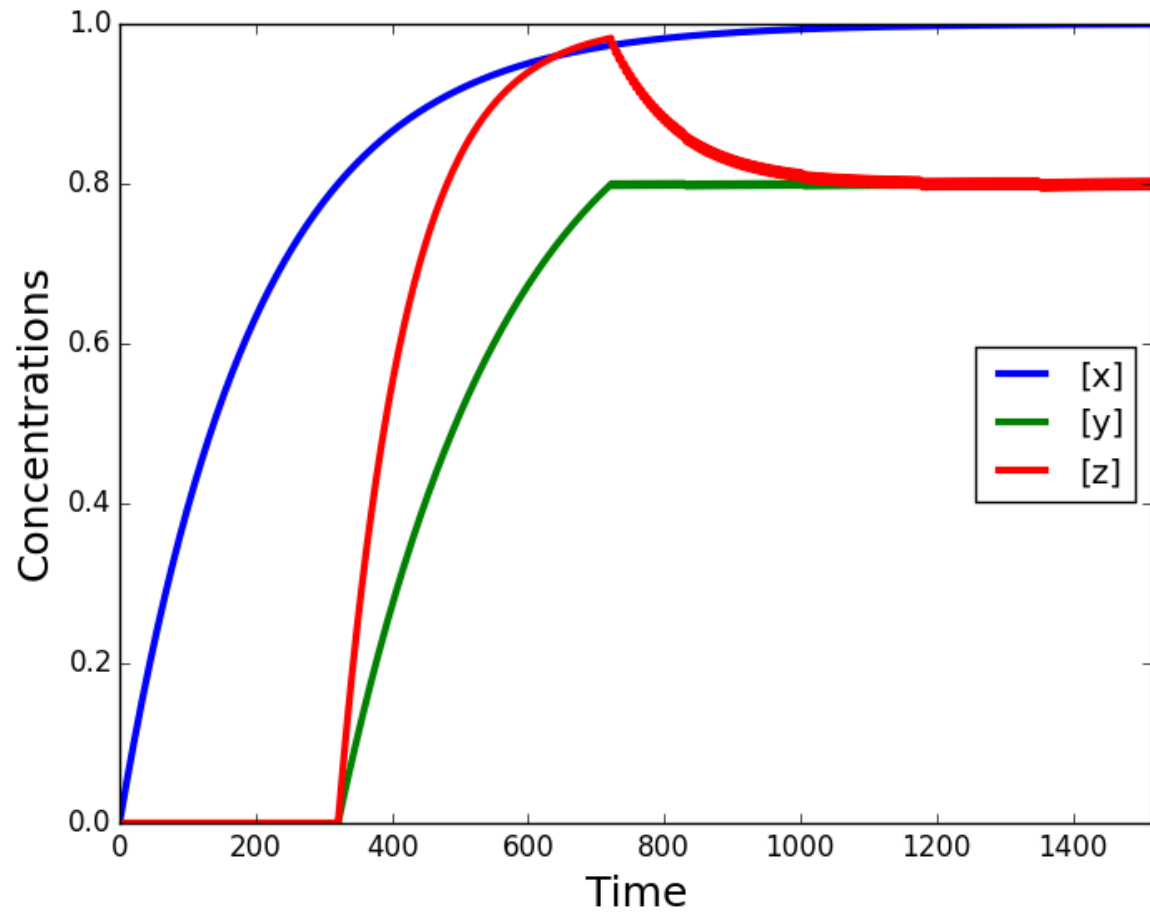
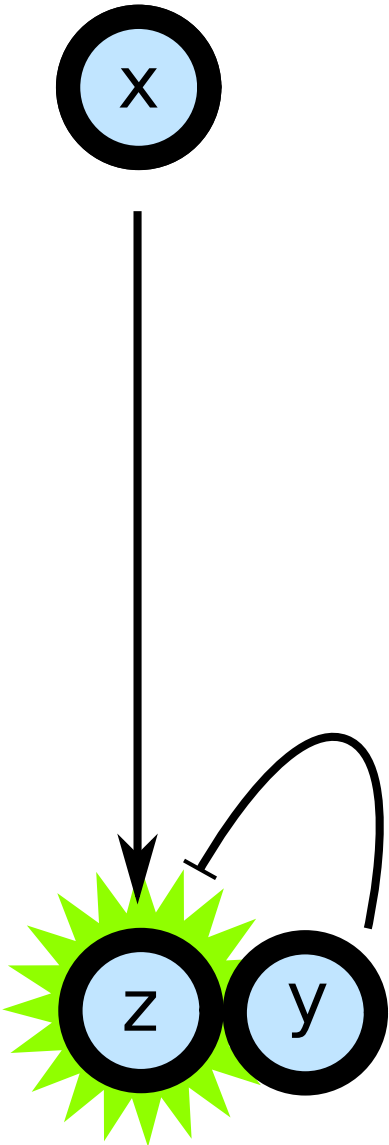
$$y' \equiv \begin{cases} k_y - \alpha y & x > t_{y,x}; y < t_{y,y} \\ -\alpha y & \text{otherwise} \end{cases}$$

$$z' \equiv \begin{cases} k_z - \alpha z & x > t_{z,x}; y < t_{z,y} \\ -\alpha z & \text{otherwise} \end{cases}$$

Common solutions

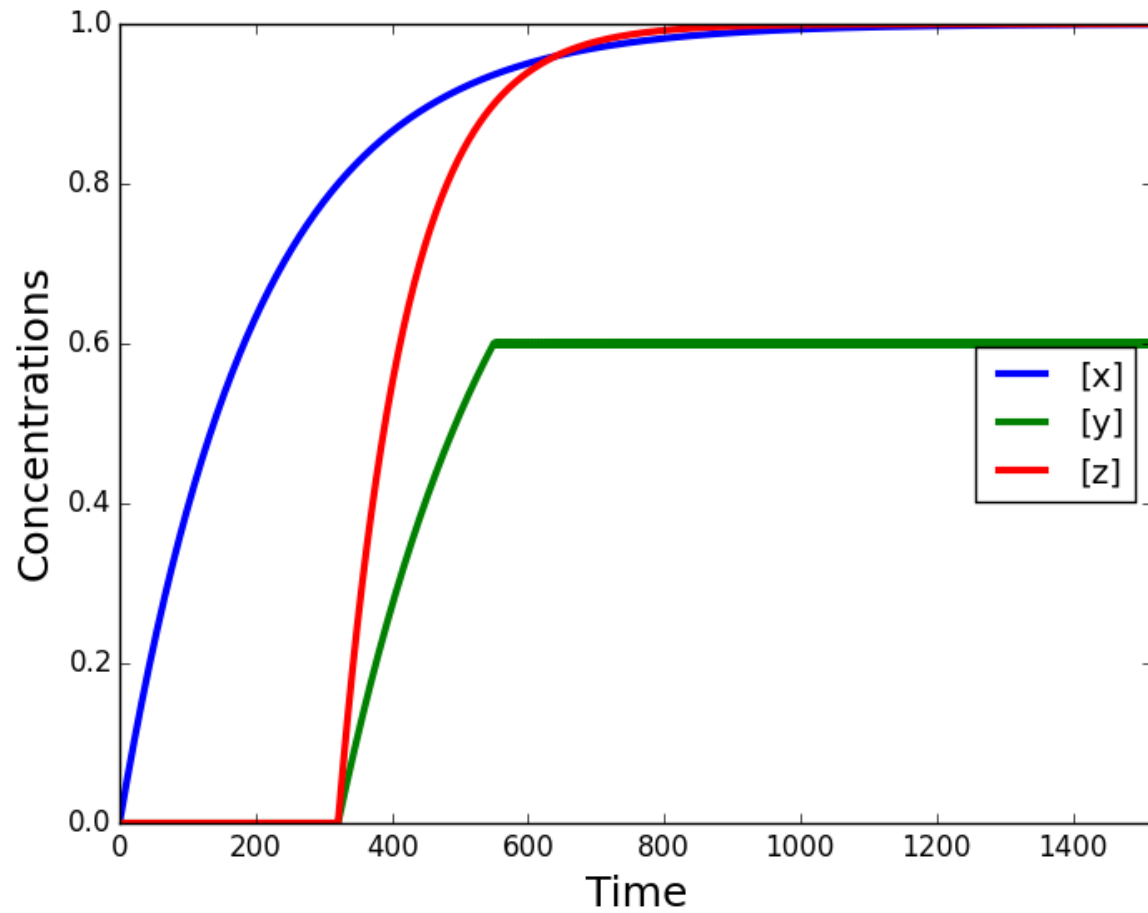
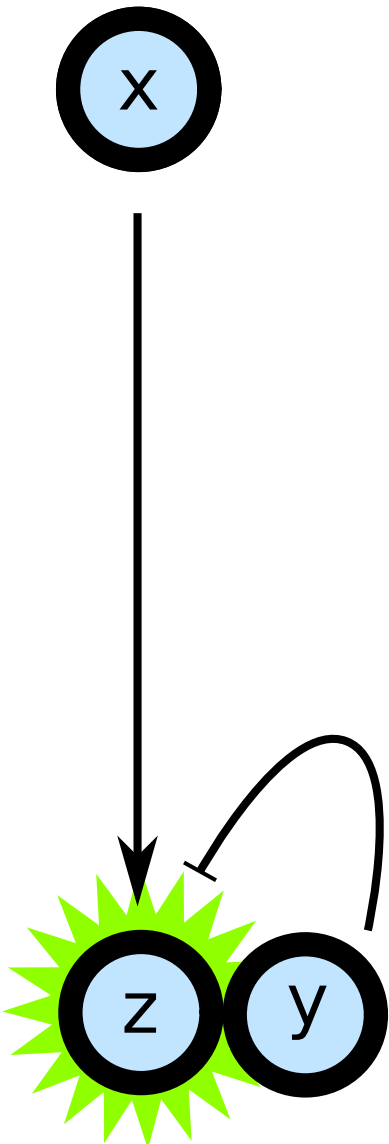


Common solutions



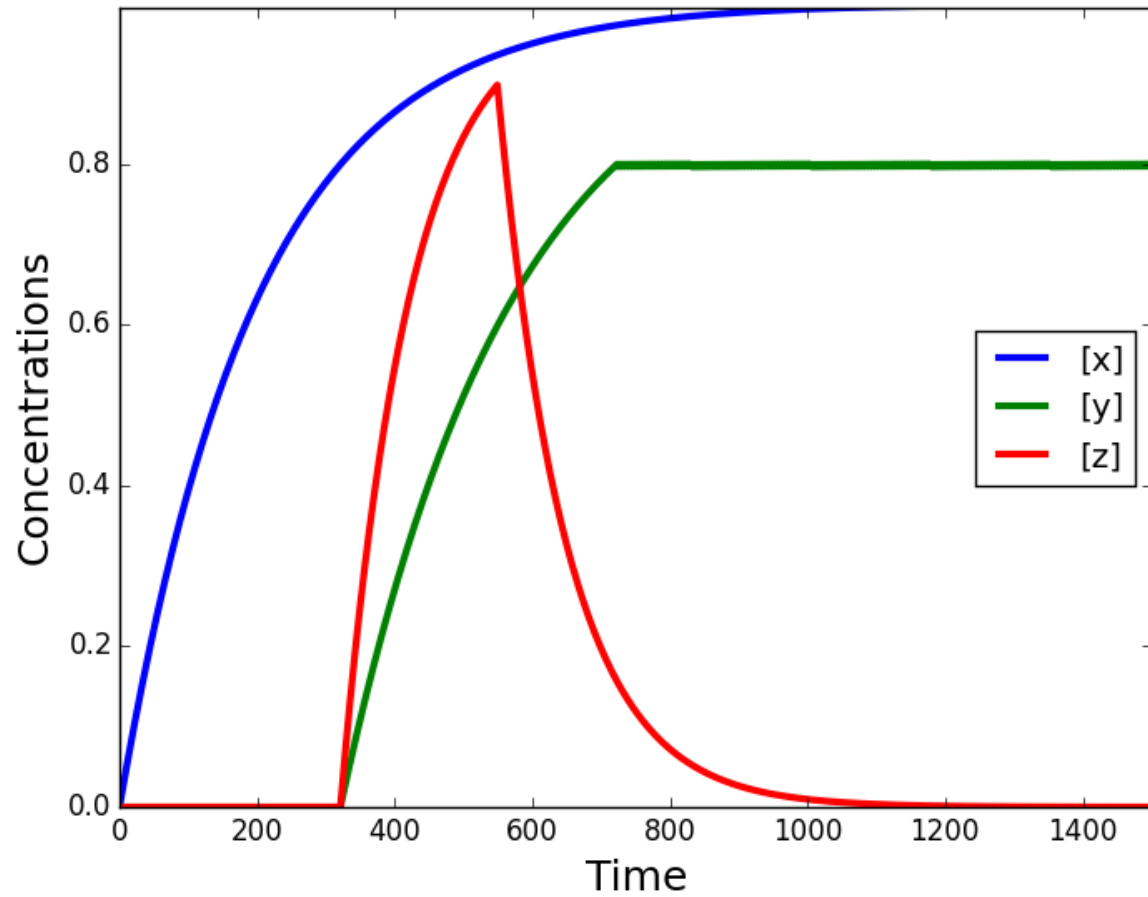
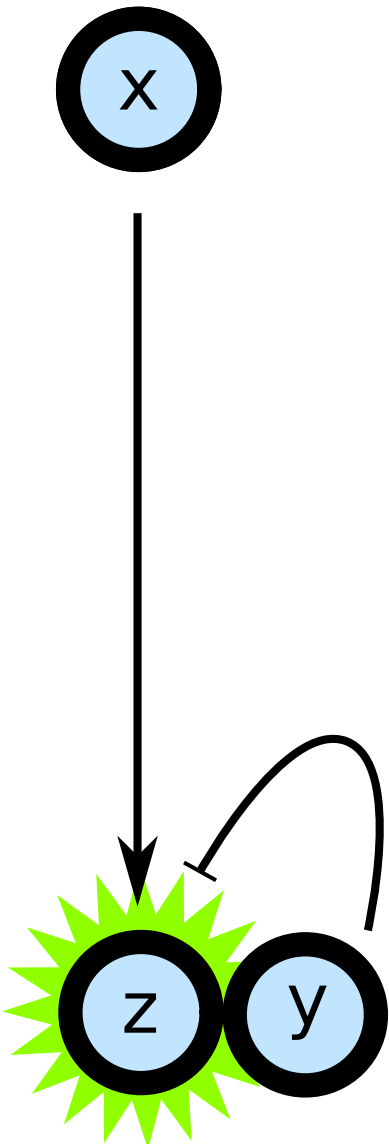
Equal thresholds

Common solutions



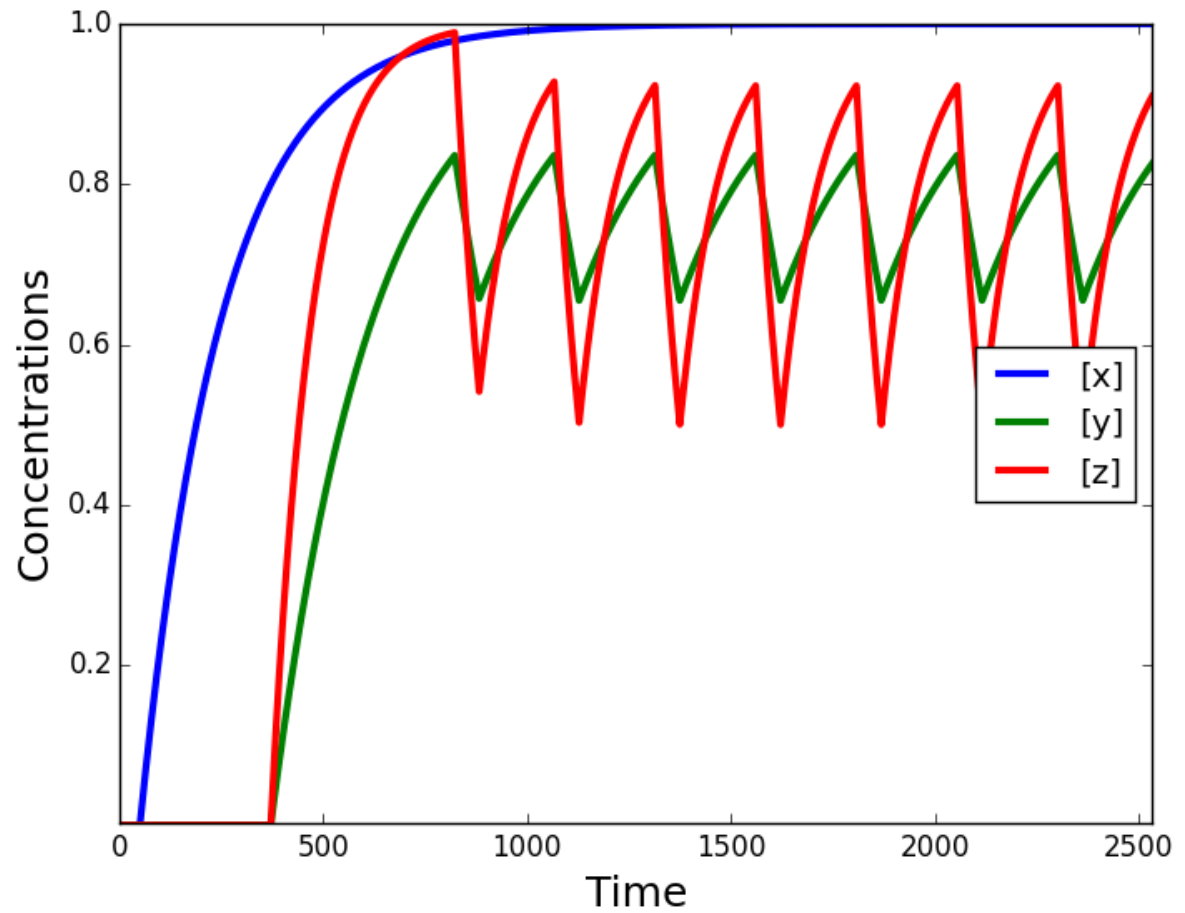
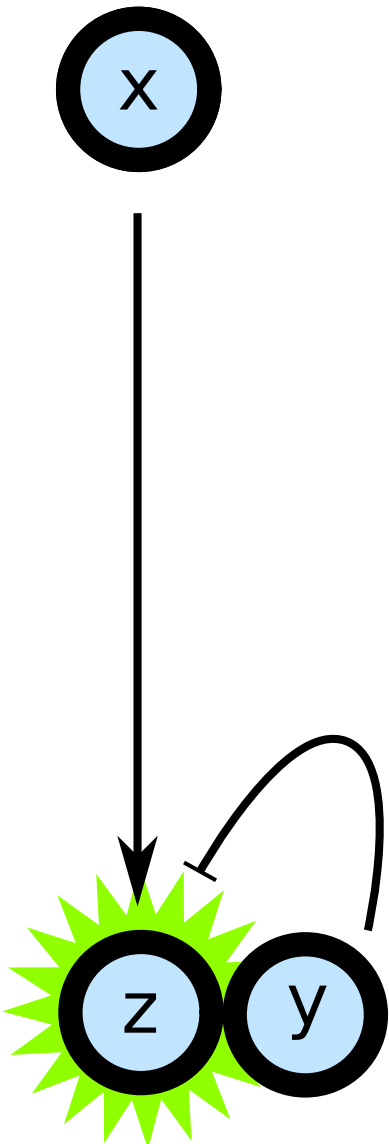
Lower threshold on y

Common solutions



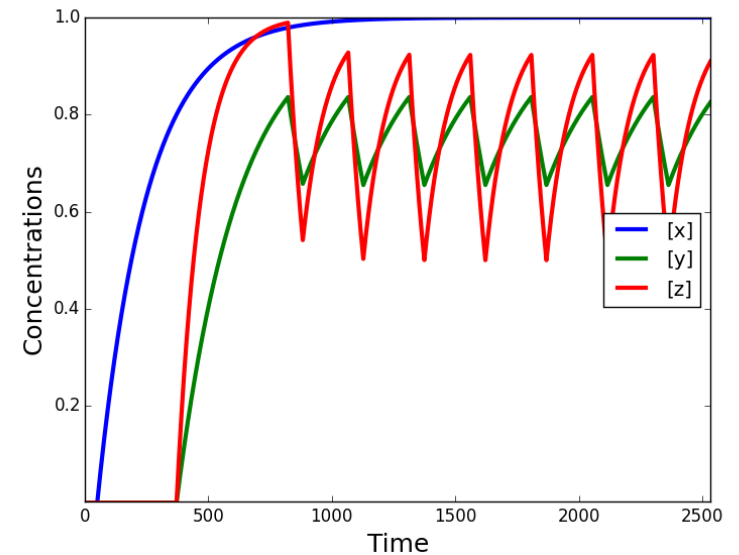
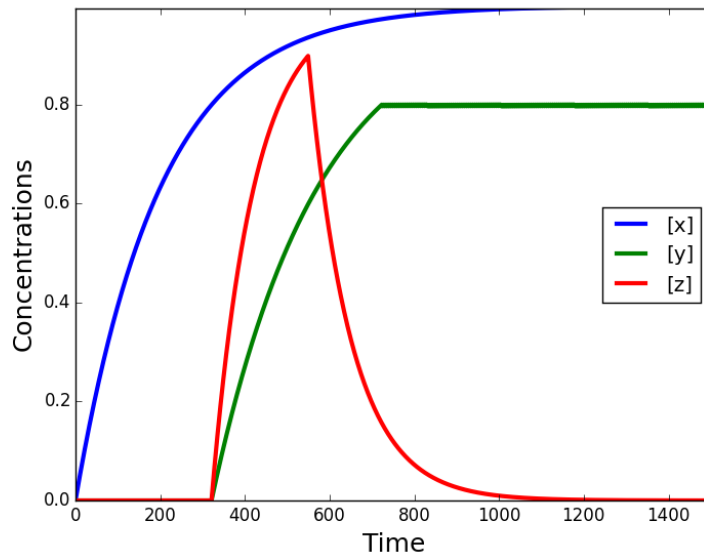
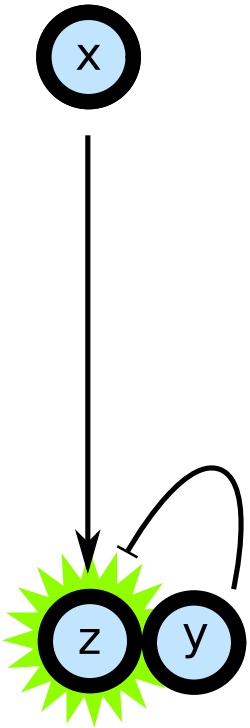
Lower threshold on y

Common solutions



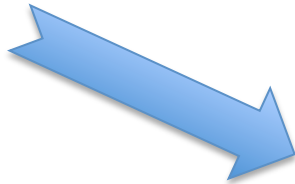
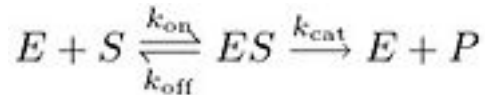
With lag on regulation

Bifurcation analysis



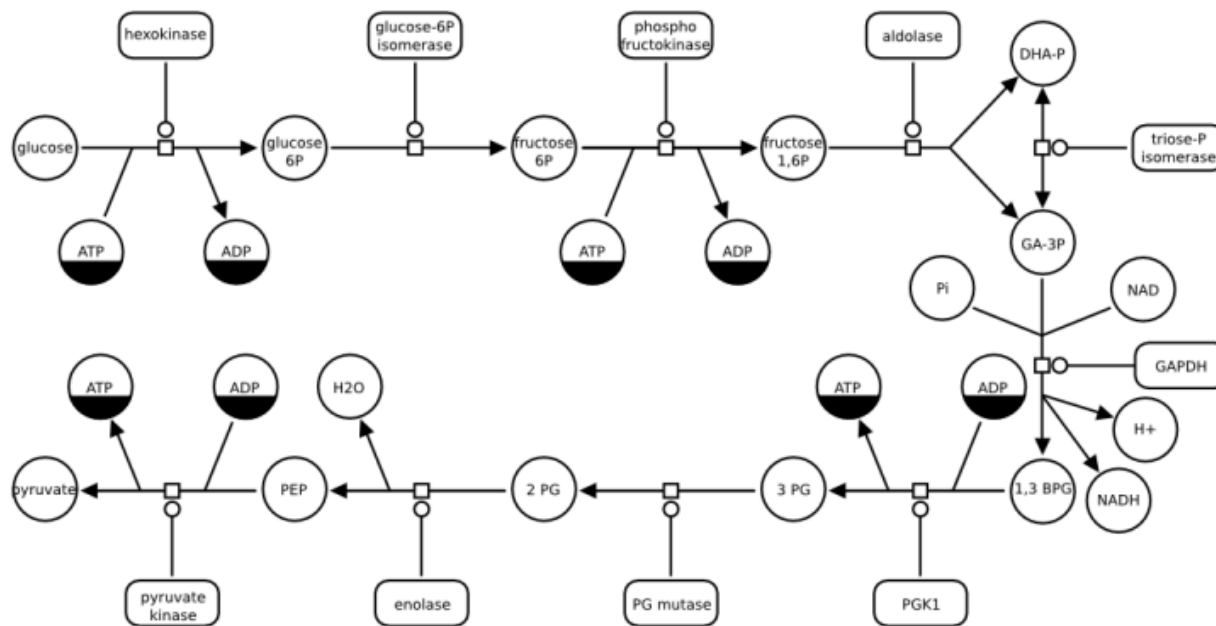
Bifurcation: Change in qualitative behavior of system as parameters change

Building models: SBML



```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="2" version="3" xmlns="http://www.sbml.org/sbml/level2/version3">
  <model name="EnzymaticReaction">
    <listOfUnitDefinitions>
      <unitDefinition id="per_second">
        <listOfUnits>
          <unit kind="second" exponent="-1"/>
        </listOfUnits>
      </unitDefinition>
      <unitDefinition id="litre_per_mole_per_second">
        <listOfUnits>
          <unit kind="mole" exponent="-1"/>
          <unit kind="litre" exponent="1"/>
          <unit kind="second" exponent="-1"/>
        </listOfUnits>
      </unitDefinition>
    </listOfUnitDefinitions>
    <listOfCompartments>
      <compartment id="cytosol" size="1e-14"/>
    </listOfCompartments>
    <listOfSpecies>
      <species compartment="cytosol" id="ES" initialAmount="0" name="ES"/>
      <species compartment="cytosol" id="P" initialAmount="0" name="P"/>
      <species compartment="cytosol" id="S" initialAmount="1e-20" name="S"/>
      <species compartment="cytosol" id="E" initialAmount="5e-21" name="E"/>
    </listOfSpecies>
    <listOfReactions>
      <reaction id="veq">
        <listOfReactants>
          <speciesReference species="E"/>
          <speciesReference species="S"/>
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="ES"/>
        </listOfProducts>
        <kineticLaw>
          <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
              <times/>
              <ci>cytosol</ci>
            </apply>
          </math>
        </kineticLaw>
      </reaction>
    </listOfReactions>
  </model>
</sbml>
```

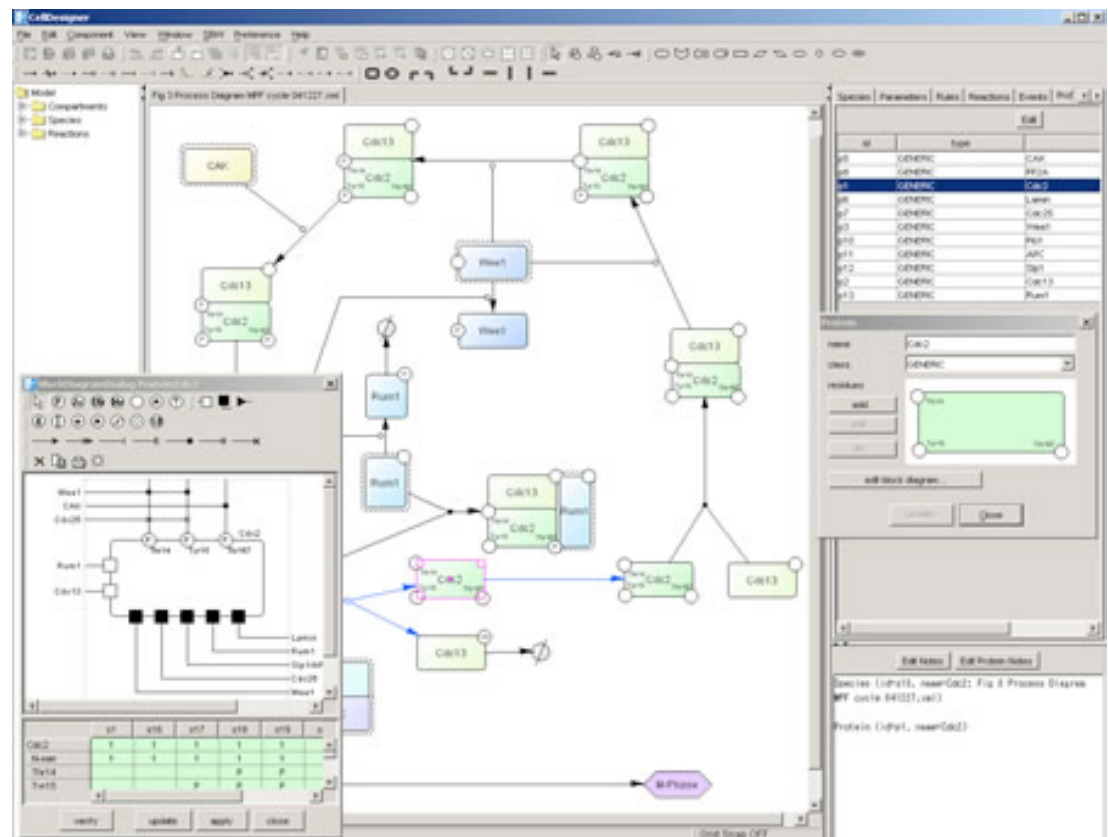
Building models for humans: SBGN



sbgn.github.io

Tools for building and simulating DE-based models

- COPASI
- **CellDesigner**
- BioSpice
- (and a bunch more)



Finding pre-built models

The screenshot shows the BioModels Database website. At the top left is the EMBL-EBI logo. The main header features the BioModels Database logo and title. A navigation bar includes links for BioModels Home, Models, Submit, Support, About BioModels, and Contact us. The main content area has a descriptive paragraph about the database. Below this, there are two main sections: 'Models published in the literature' and 'Models automatically generated from pathway resources (Path2Models)'. Each section has a 'Browse' area with colored boxes for 'Manually curated (627 models)' and 'Non curated (980 models)', and an 'Alternative access' area with icons for 'Gene Ontology classification', 'Gene Ontology tree', and 'Advanced search'.

EMBL-EBI

BioModels Database

BioModels Home | Models | Submit | Support | About BioModels | Contact us

BioModels Database is a repository of computational models of biological processes. Models described from literature are manually curated and enriched with cross-references. All models are provided in the Public Domain. More information about BioModels Database can be found in the [FAQ](#).

Models published in the literature

Browse

Manually curated (627 models)

Non curated (980 models)

Alternative access

Gene Ontology classification

Gene Ontology tree

Advanced search

Models automatically generated from pathway resources (Path2Models)

Browse

Metabolic (112,898 models)

Non-metabolic (27,531 models)

Alternative access

<http://www.ebi.ac.uk/biomodels-main>

How do we build models?

Networks/wiring from:

- Next gen sequencing
- Protein-protein interactions
- Enzyme characterization

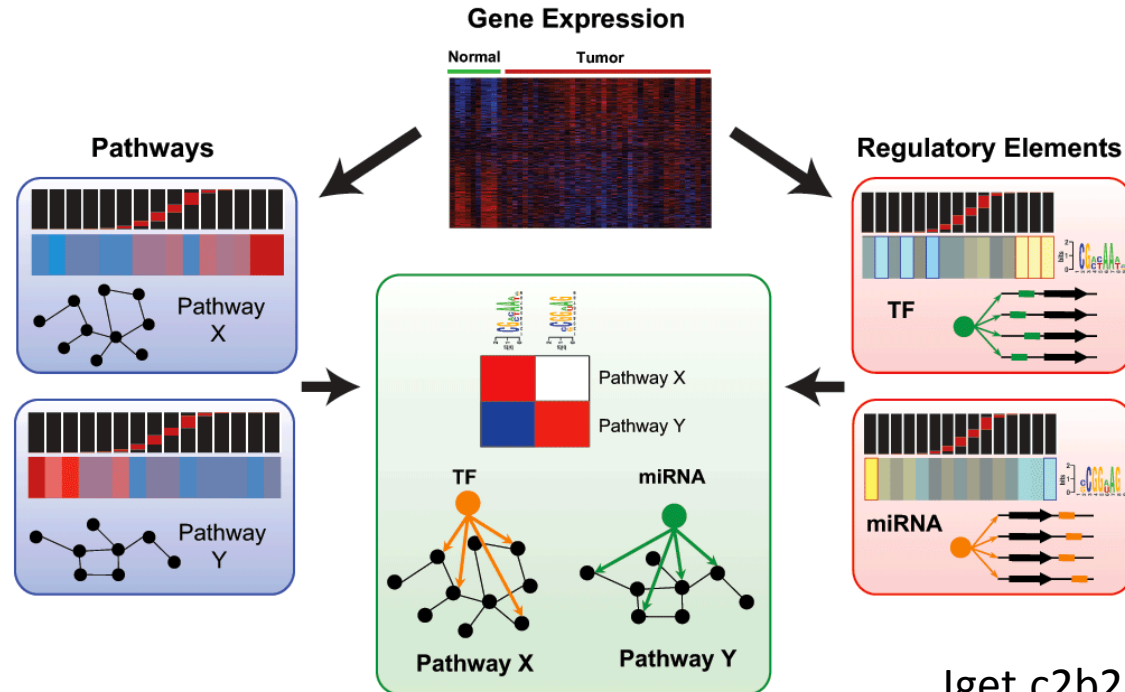
Parameters from:

- Direct measurement of key constants
- Fitting parameters to experimental results

How do we build models?

Networks/wiring from:

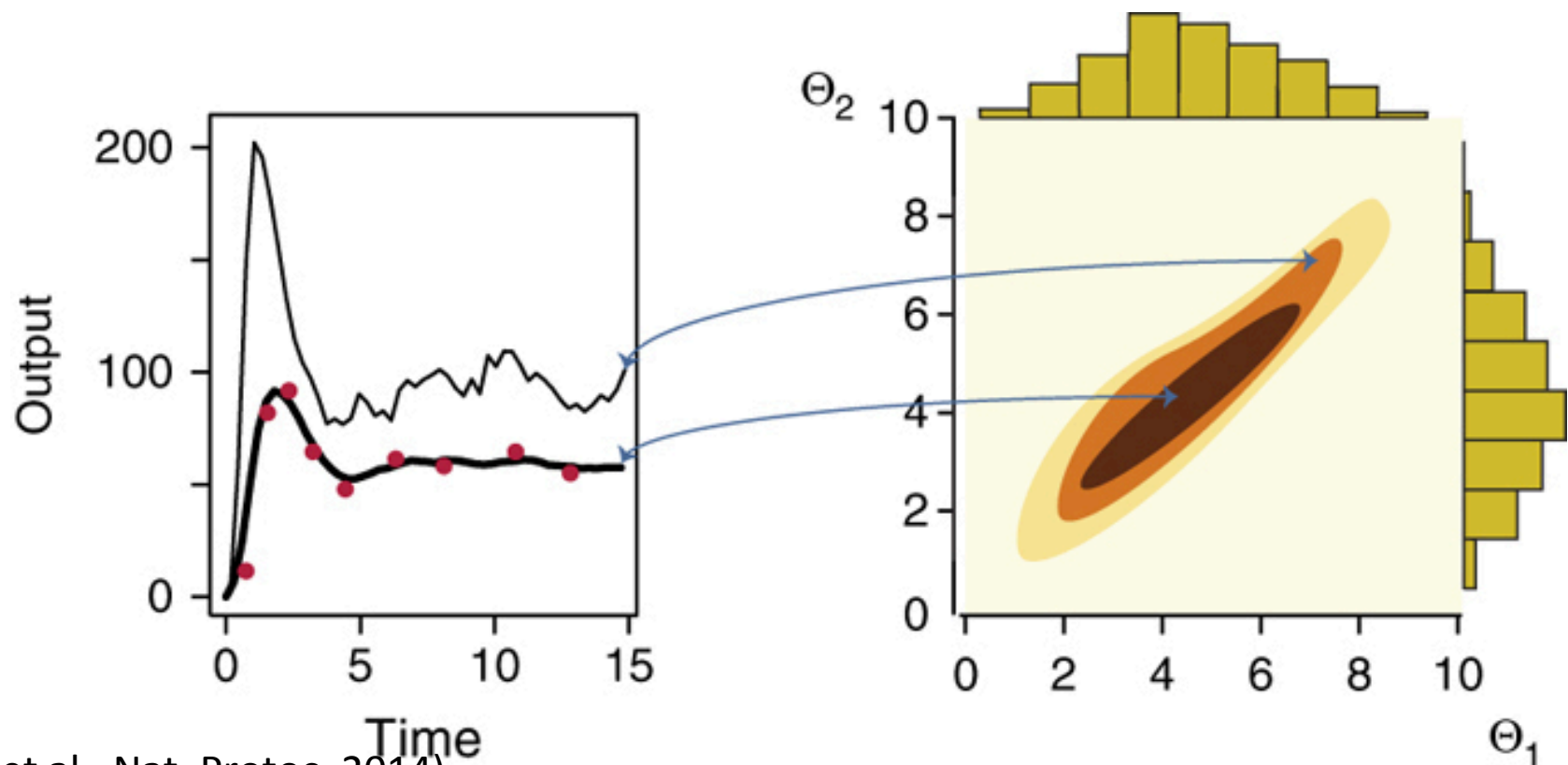
- Next gen sequencing
- Protein-protein interactions
- Enzyme characterization



How do we build models?

Parameters from:

- Direct measurement of key constants
- Fitting parameters to experimental results

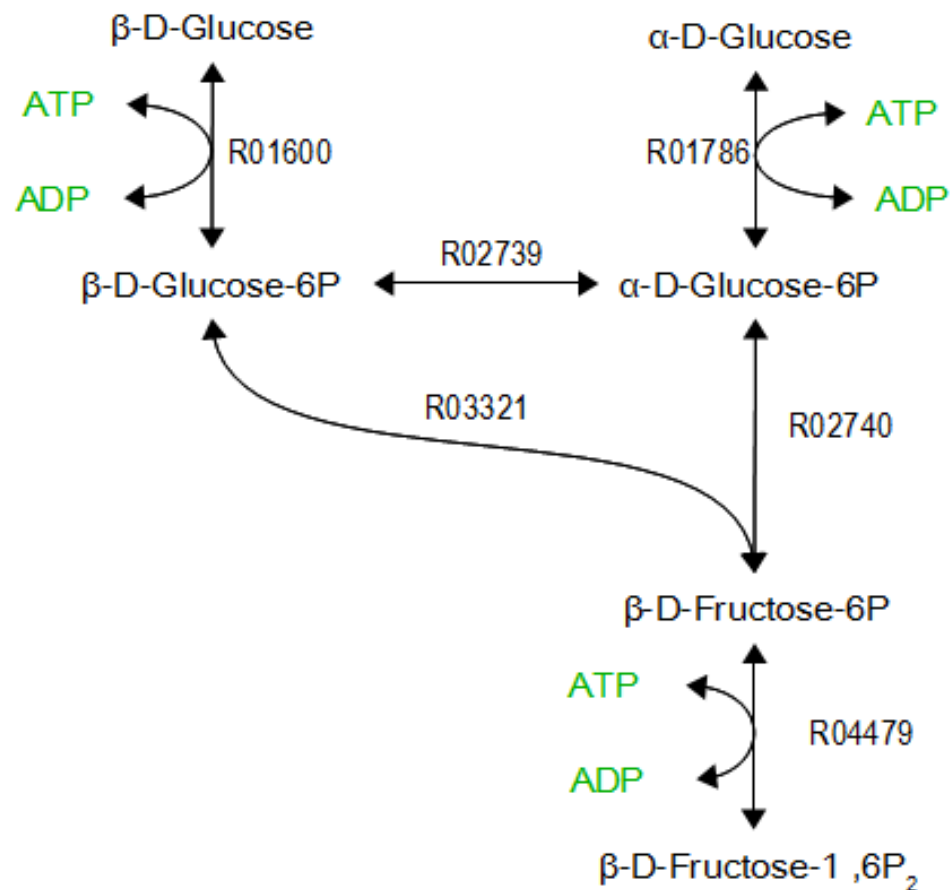


(Liepe et al., Nat. Protoc. 2014)

Building systems-level models

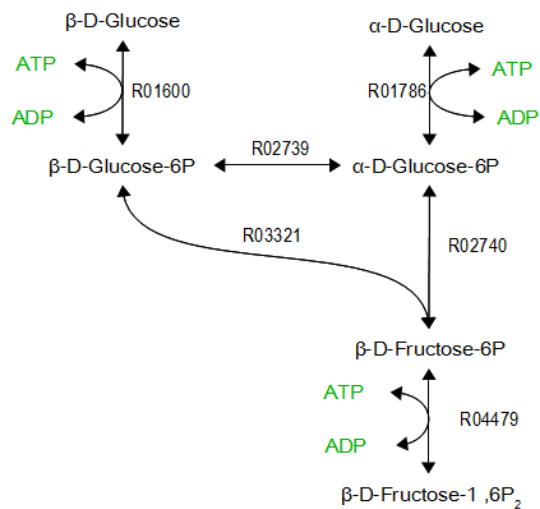
- Motif and GO term analysis
- Kinetic modeling of simple networks
- **Constraint-based modeling of cellular metabolism**

Constraint-based modeling allows simplified metabolic simulation

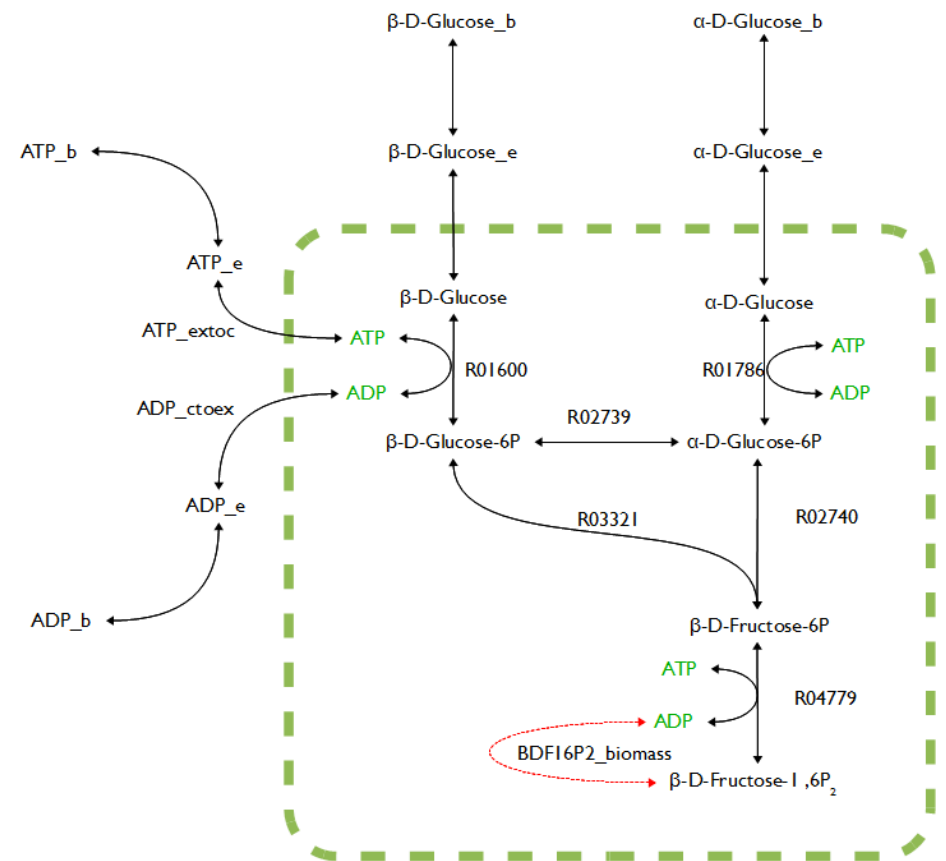


(Image from Thomas Forth)

Constraint-based modeling allows simplified metabolic simulation

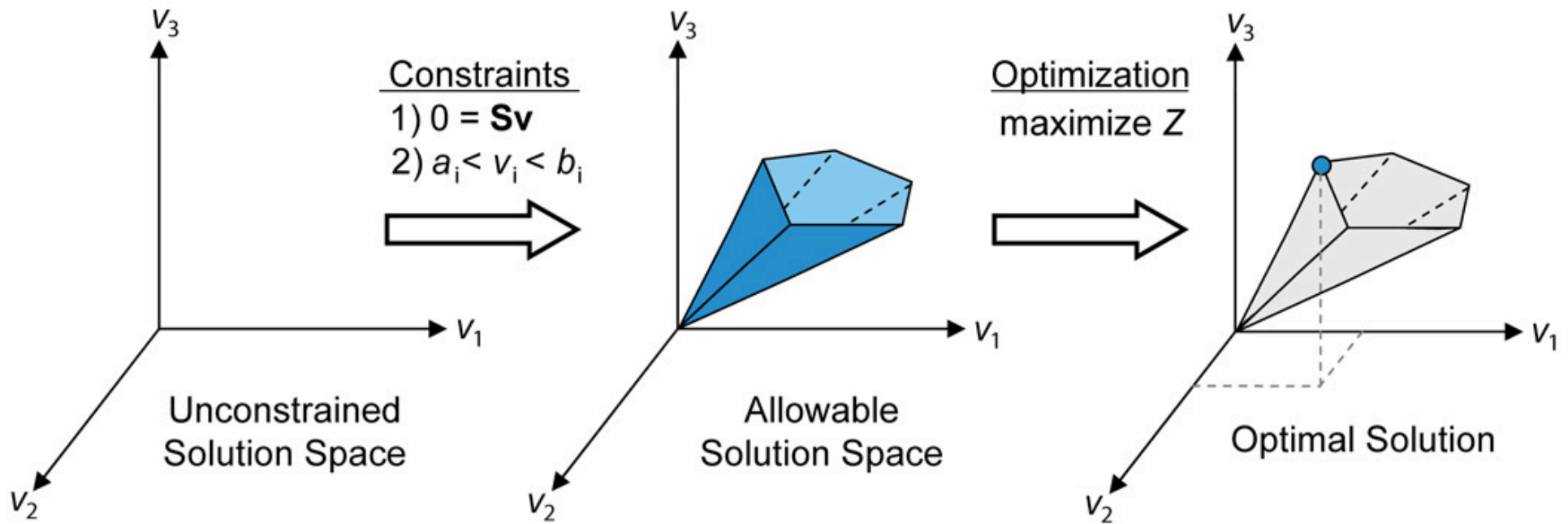


Prepare for FBA



(Image from Thomas Forth)

Constraint-based modeling allows simplified metabolic simulation



(Orth et al., Nat. Biotechnol. 2010)

Usage cases for constraint-based modeling

Strengths:

- Fast model evaluation
- Simple interpretation
- Can work from somewhat incomplete data

Weaknesses:

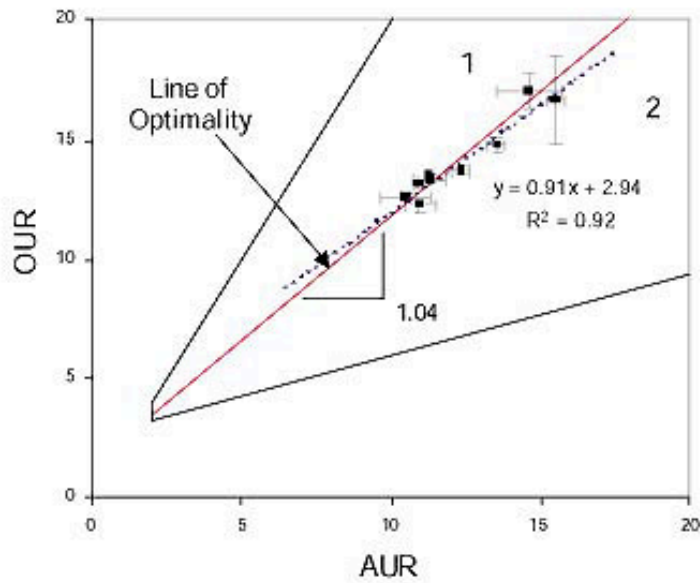
- No concentrations, only fluxes
- No dynamics
- Optimality assumption

Usage cases for constraint-based modeling

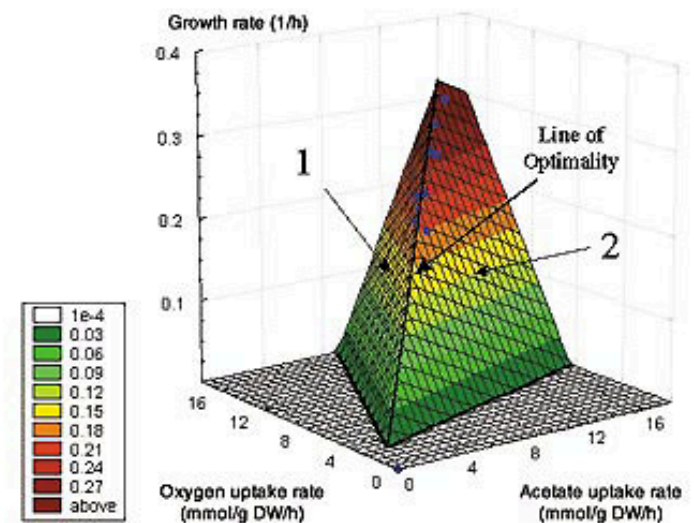
- Designing metabolic networks to make specific products
- Identifying/understanding effects of mutations
- Finding holes in current state of knowledge on metabolic networks
- Optimizing media for growth or production

Example: Identifying key metabolic parameters in E. coli

A

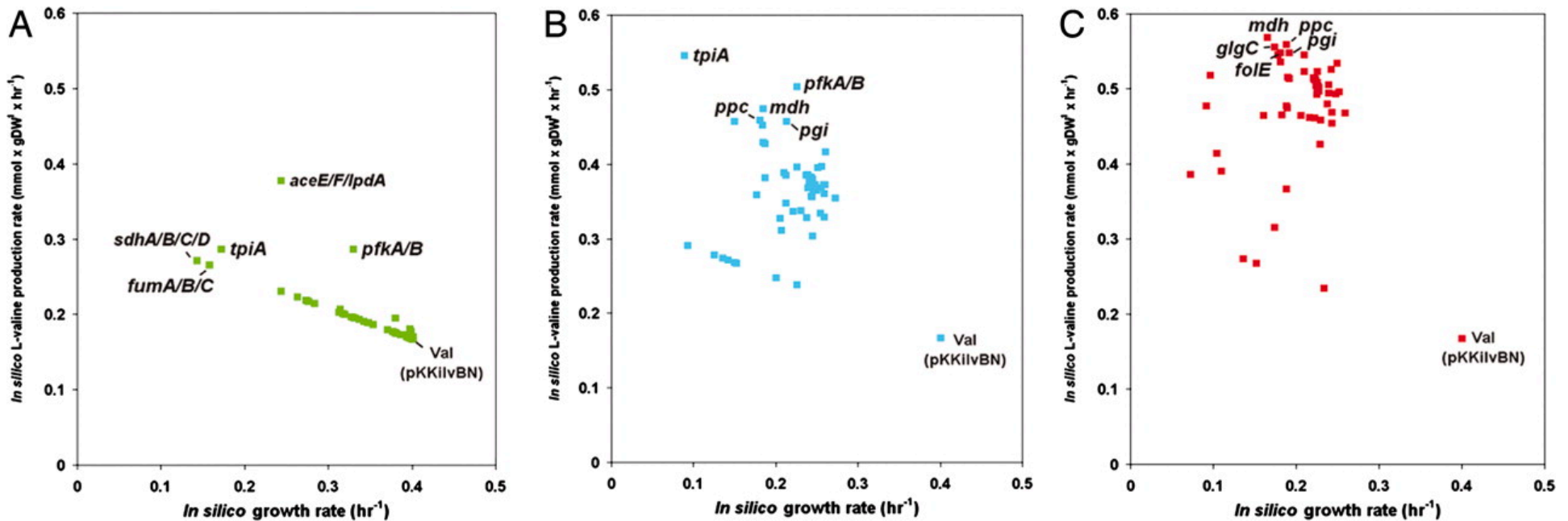


B



(Edwards et al., Nat. Biotech., 2001)

Example: Engineering strains to produce L-valine



Gave 45% improvement over rationally designed strain

(Park et al., PNAS, 2007)

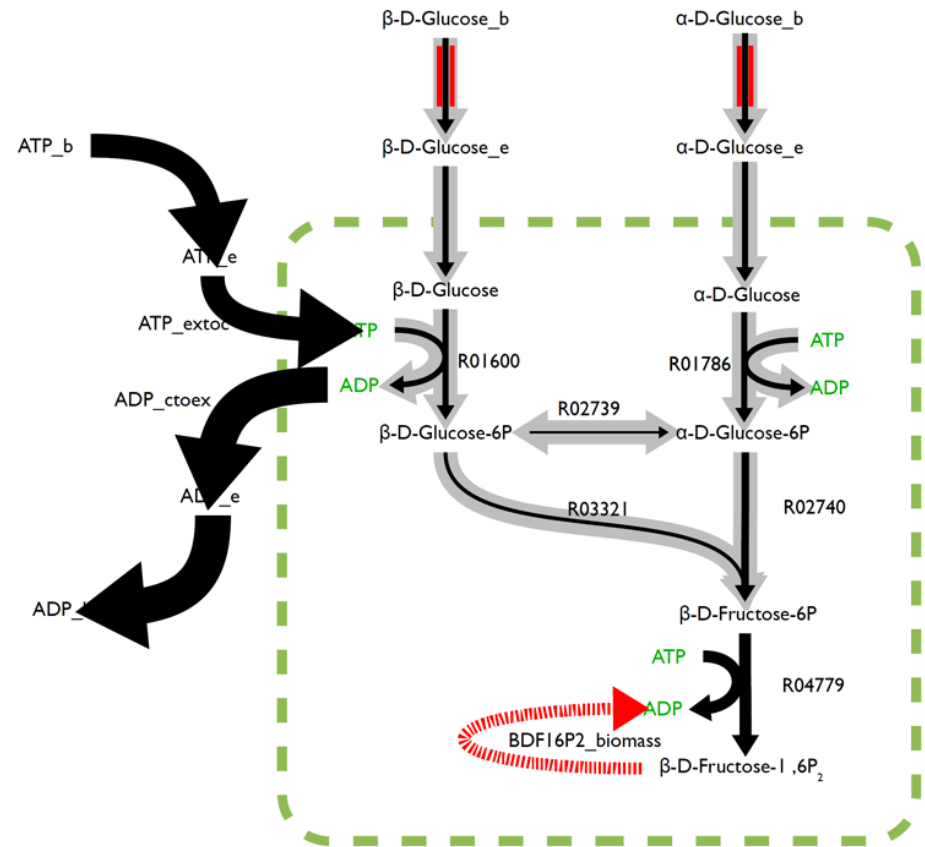
Tools for constraint-based modeling

- COBRA toolbox (matlab)
- MASS toolbox (mathematica)
- Sybil (R)
- cobrapy, PyFBA (python)

Most allow SBML import

Common FBA variations

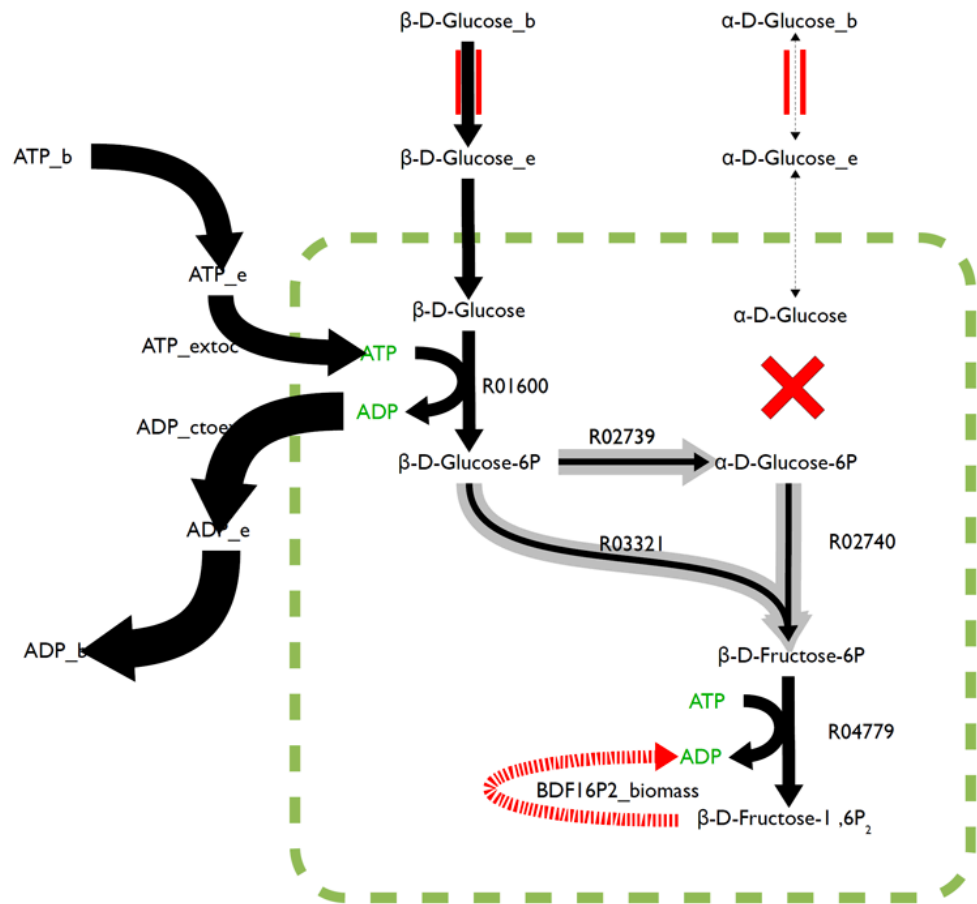
- **Flux Variability Analysis (FVA): Give boundaries on solutions**
- **Minimization of Metabolic Adjustment (MOMA): Find smallest possible perturbation**
- **Regulatory on-off minimization (ROOM): Minimize number of regulatory changes**



(Image by Thomas Forth)

Common FBA variations

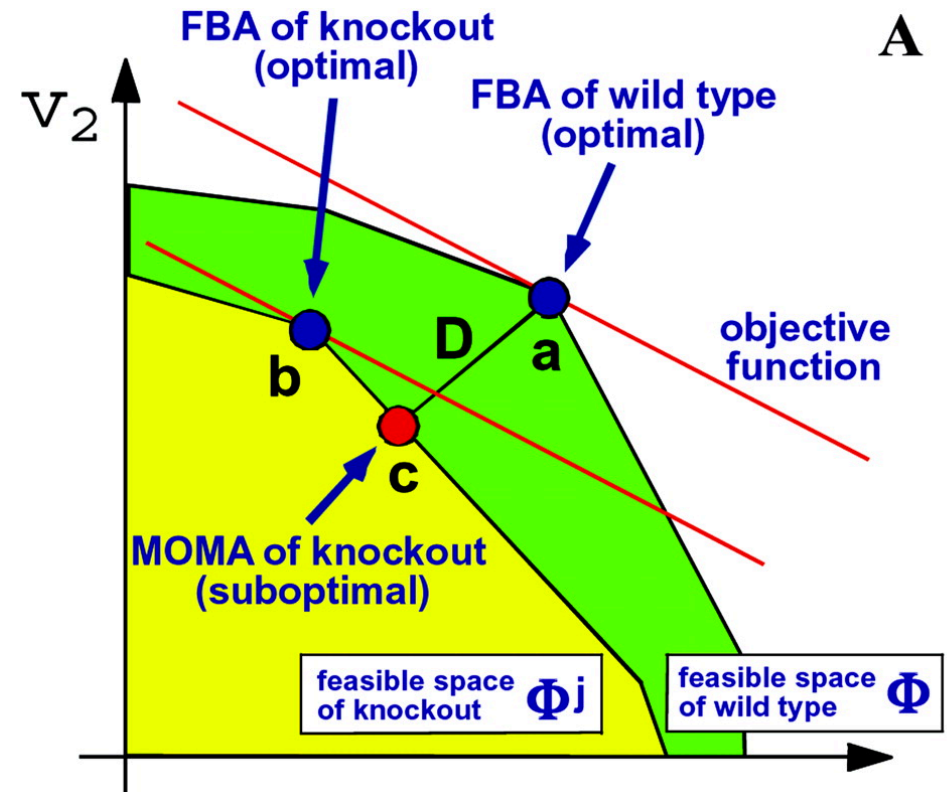
- **Flux Variability Analysis (FVA): Give boundaries on solutions**
- **Minimization of Metabolic Adjustment (MOMA): Find smallest possible perturbation**
- **Regulatory on-off minimization (ROOM): Minimize number of regulatory changes**



(Image by Thomas Forth)

Common FBA variations

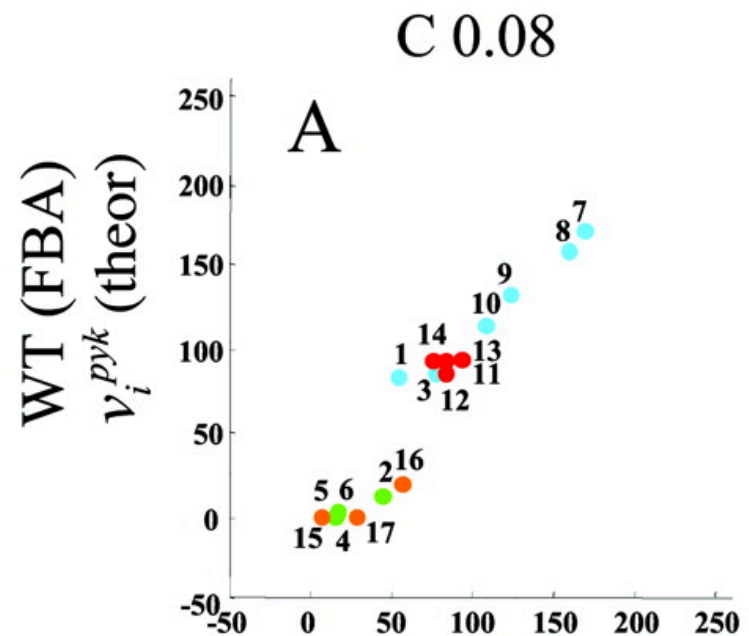
- Flux Variability Analysis (FVA): Give boundaries on solutions
- **Minimization of Metabolic Adjustment (MOMA): Find smallest possible perturbation**
- Regulatory on-off minimization (ROOM): Minimize number of regulatory changes



(Segre et al., PNAS, 2002)

Common FBA variations

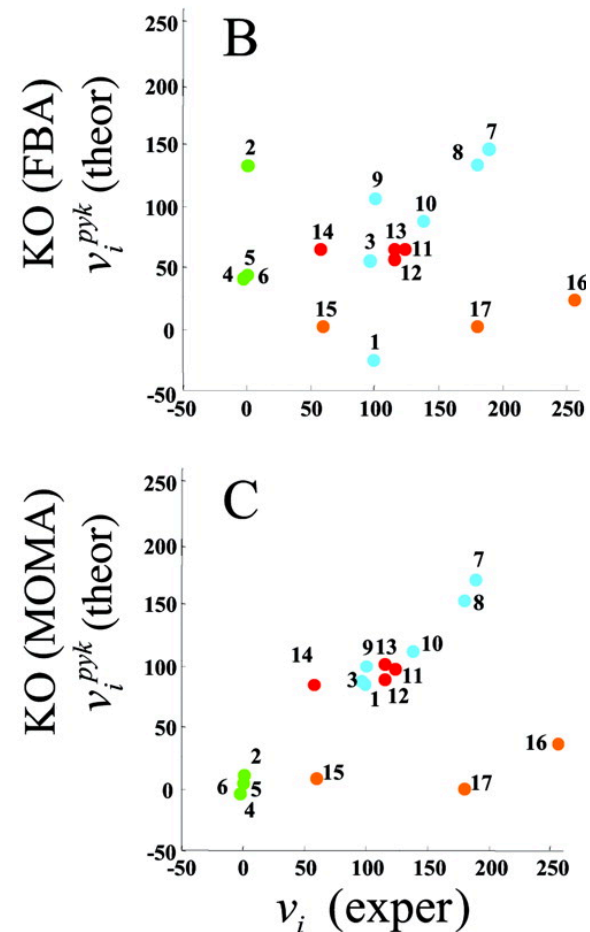
- Flux Variability Analysis (FVA): Give boundaries on solutions
- **Minimization of Metabolic Adjustment (MOMA): Find smallest possible perturbation**
- Regulatory on-off minimization (ROOM): Minimize number of regulatory changes



(Segre et al., PNAS, 2002)

Common FBA variations

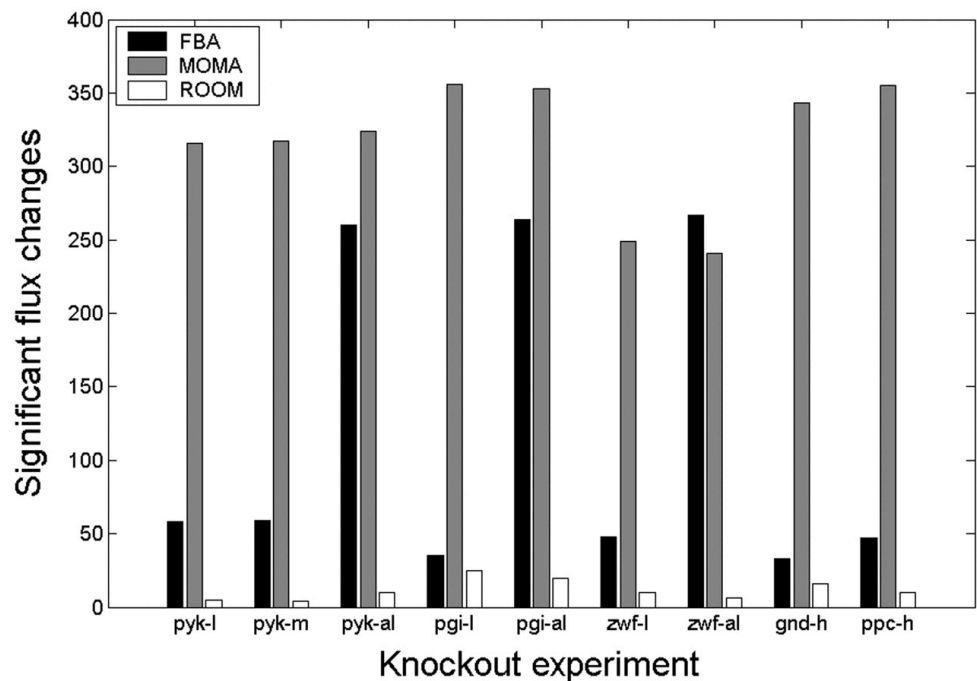
- Flux Variability Analysis (FVA): Give boundaries on solutions
- **Minimization of Metabolic Adjustment (MOMA): Find smallest possible perturbation**
- Regulatory on-off minimization (ROOM): Minimize number of regulatory changes



(Segre et al., PNAS, 2002)

Common FBA variations

- Flux Variability Analysis (FVA): Give boundaries on solutions
- Minimization of Metabolic Adjustment (MOMA): Find smallest possible perturbation
- **Regulatory on-off minimization (ROOM): Minimize number of regulatory changes**



(Shlomiet al., PNAS, 2005)

Common FBA variations

- Flux Variability Analysis (FVA): Give boundaries on solutions
- Minimization of Metabolic Adjustment (MOMA): Find smallest possible perturbation
- **Regulatory on-off minimization (ROOM): Minimize number of regulatory changes**

Prediction of knockout lethality:

	FBA	MOMA	ROOM
True-positive	449	399	449
False-positive	64	60	62
True-negative	23	27	25
False-negative	19	69	19
Positively predicted genes	96%	85%	96%
Negatively predicted genes	26%	31%	29%
Overall prediction	85.0%	76.7%	85.4%

(Shlomi et al., PNAS, 2005)