277C Computational Systems Biology

**Course description:** Computational inference and modeling of gene regulation networks, signal transduction pathways, and the effects of regulatory networks in cellular processes, development, and disease. Readings, discussion and final project.

**Prerequisites:** Previous course in bioinformatics (such as ICS 277A or 277B), or molecular biology (such as Biological Sciences 99 or equivalent) together with programming experience; also multivariable calculus and linear algebra (such as Mathematics 2D and 2J or equivalent); or permission of instructor.

**Course Overview and Goals:** The expression “systems biology” is often used to describe attempts at unraveling living molecular systems above the traditional level of single genes and single proteins. Usually the focus is on the level of pathways and groups of pathways and their consequences for multicellular systems including development in tissues and higher levels of organization. A confluence of new instrumentation and data sources have made it possible and often essential to understand such biological systems computationally. A mix of lectures, readings, and a computational project will introduce students to the area.
Lecture 1 Notes

Define computational systems biology:

Achieving a predictive, scientific understanding of complex living systems whose complexity outruns intuitive understanding. Also developing mathematical and computational tools for that purpose.

A major component of that complexity is due to large networks. Fortunately data and computations can be obtained for large networks, all the way up to genomic scale.

Networks

There are a number of relevant computational analogies including the “noisy circuit” analogy for large molecular information-processing networks.

Some major stand-alone network types:

Transcriptional regulation
- Example: transcription factor network determining longitudinal coordinate systems in the fruit fly *Drosophila melanogaster*.

Signal transduction – protein modification
- Example: “MAP Kinase” cascade for detecting mating pheromone in bakers’ yeast *Saccharomyces cerevisiae*.

Metabolism
- Example: biosynthesis of branched chain amino acids in bacterium *Escheria coli* K-12.

Mechanical networks
- Mechanical interaction of cells and compartments. Space x circuitry x time.
- Major types of compartments: cell, nucleus, cytosol, membranes, spatial subdivisions eg. quadrants or octants.
- Example: development of the shoot apical meristem of the plant *Arabidopisis thaliana*.

Compound networks composed of the above
- Example: transcriptional feedback to modulate signal transduction cascades. Example: yeast cell cycle network with interacting protein modification and transcriptional regulation sectors.
Example: MAPK signal transduction cascade for yeast.


Note the existence of cross talk e.g. Dig1,2, Ste12 between parallel pathways.
**Basic computational strategy:** bio to computation only by way of mathematical models. This allows for separation of concerns in resolving incorrect predictions, and makes available physical science tools in applied mathematics.

To this end, for each type of network we will need to supply appropriate dynamics.

A more general framework than network-by-network analysis is objects, relations, and processes (transformations and/or dynamics). Objects and relations are typed nodes and links in a global super-network. Processes govern how the existence and state of these objects and relationships change over time.
**Incorporating biological knowledge**

Introduce fundamental biological knowledge: the Central Dogma of molecular biology, suitably augmented to show information flows via regulatory relationships among genes, RNA, and protein.

Central Dogma: DNA contains information coding for a protein or small group of related proteins and their regulation – together, a “gene”. This information is transcribed to message RNA and then translated (by 3 x 1/out of 4 bp \(\rightarrow\) 1 out of 20 amino acid map) to protein.

This basic information progression is refined by (1) feedback via transcription factors, proteins which bind to DNA or other TF’s and thereby affect transcription, and by (2) protein modifications such as phosphorylation (addition of a PO4) group of selected amino acid residues in a protein which (a) carries information and (b) affects protein charge and Xxx and thereby changes protein conformation and function. Also (3) protein assembly/disassembly into complexes which new form and function, and (4) actively regulated degradation of protein. Further information processing novelties not fully reflected below: RNAi, active transport within and between compartments, chromatin modification, prokaryotic gene clustering, action of small molecules including gasses, … biomineralization … and no doubt a host of others.

**Illustration:**
Pathway databases


S. cerevisiae:
S. cerevisiae:
Gene Ontology


Source: Saccharomyces Genome Database
Objects, relationships, and processes

What are the objects, relationships, and processes to be modeled?

central dogma diagram
  DB schema: reactants, processes/reactions, knowledge sources,
  models, behaviors (observable, selectable phenotypes)

objects
  Reactants
    central dogma: DNA, RNA, protein(with state info)
  compartments
    cells, cytosol, nucleus, organelles, membranes, spatial
    subcompartments …
  protein/DNA modifications:
    phosphorylation, methylation, acetylation, ubiquitination, ..
  localization to compartments, membranes and regions
  binding sites: DNA, protein, …

Processes and Reactions
  Metabolic network steps
  Allosteric enzymes - cooperativity
  Protein-protein regulatory interactions
  Transcriptional regulation, with feedback
  Diffusion, transport, and signaling
    e.g. auxin and PIN1 auxin transporter
  Processes (we have models and literature)
    Basic issues: resources, information, replication
    metabolism
    Cell cycle – e.g. budding yeast
    Signal transduction pathways
      e.g. yeast pheromone response, stress response, …
  Multicellular development
    Gene/Signal Regulatory Network
      e.g. plant growth, Drosophila blastoderm,
      many others.

Knowledge sources
  Gene expression images
  Microarray expression data
    e.g. yeast cell cycle (Spellman et al. 1998)
    ChIP-chip DNA:protein binding data (Lee et al. 2002)
  p:p interactions
    Y2H – pairwise, noisy
    Mass spectroscopy – group, better
  sequence data
    coding region motifs
    binding site motifs (Kellis et al 2003, others)
  textual information retrieval on scientific literature
lower the cost of human curation


Network types

Regulation
  Transcriptional regulation
  Signal transduction – protein modification
  Mechanical networks
    Mechanical interaction of cells and compartments
    space x circuitry x time

  major types of compartments: cell, nucleus, cytosol, membranes, spatial subdivisions eg. quadrants or octants.
Software architecture

Desired data flow in applications of systems biology:

We need to create software to support this flow.

For example:
Lecture 2. Signal Transduction in Yeast.

Observe again the overlapping signal transduction cascades, showing only the feedforward portions that are currently well known:

Madhani, HD. Fink, GR.
THE RIDDLE OF MAP KINASE SIGNALING SPECIFICITY [Review].
Modeling signal transduction

Bimolecular reactions in solution:

\[ A, B \leftrightarrow \{C \} \]

Example:

Yeast Fus3 phosphorylates Far1, arrests cell cycle. 
FUS3, KSS1 also phosphorylate Ste12 TF/Dig1/Dig2, leads to mating 
(discussed in Kusari et al. 2004)

Binding/unbinding at a site:

\[ A, S \leftrightarrow S-A \]

Example: Ste12 binds to DNA activating a battery of 200 mating response genes

In-solution mathematics

Law of mass action: small times, reaction probability is small, therefore proportional to 
the product of the concentrations of the inputs.

E.g. \[ A + B \rightarrow C, \ C \rightarrow A + B \]

Build out of subreactions \[ A+ B \leftrightarrow [AB] \rightarrow C \]

Use to build superreactions \[ S + E \leftrightarrow [SE] \rightarrow P + E \beta. \]

Binding site mathematics

There are two basic transformation: binding and unbinding. One can think of them as 
“reactions” for continuous probabilities rather than for continuous concentrations, or as 
stochastic grammar rules:

\[
\begin{align*}
\text{site_empty}(t) & \rightarrow \text{site_occupied}(t+\Delta t) \\ 
\text{site_occupied}(t) & \rightarrow \text{site_empty}(t+\Delta t)
\end{align*}
\]

with \( \Pr(e \rightarrow o | \Delta t) \) and \( \Pr(o \rightarrow e | \Delta t) \)

Transition probabilities for small time steps are proportional to delta-t and, for 
occupation, to the concentration of the ligand:

\[
Pr(\text{empty} \rightarrow \text{occupied}) = \alpha \Delta t [A]
\]

\[
Pr(\text{occupied} \rightarrow \text{empty}) = \beta \Delta t
\]

In matrix form:

\[
\begin{pmatrix}
Pr(\text{empty}) \\
Pr(\text{occupied})
\end{pmatrix}
(t+\Delta t) =
\begin{pmatrix}
1 - \alpha \Delta t [A] & \beta \Delta t \\
\alpha \Delta t [A] & 1 - \beta \Delta t
\end{pmatrix}
\begin{pmatrix}
Pr(\text{empty}) \\
Pr(\text{occupied})
\end{pmatrix}^t
\]
Fundamental stochastic “Master equation” for this two-state system [cite van Kampen here? Gillespie?]:

\[ \frac{dP}{dt} = M \cdot P \]

\[ M = \begin{pmatrix} -\alpha [A] & \beta \\ \alpha [A] & -\beta \end{pmatrix} \]

Solution of linear master equation \( \frac{dP}{dt} = M \cdot P \):

\[
\begin{aligned}
\frac{-\Delta t}{e} & \begin{pmatrix} -\alpha [A] & \beta \\ \alpha [A] & -\beta \end{pmatrix} \\
& \approx \begin{pmatrix} 1 - \alpha \Delta t [A] & \beta \Delta t \\ \alpha \Delta t [A] & 1 - \beta \Delta t \end{pmatrix} \begin{pmatrix} 1 - P \\ P \end{pmatrix}
\end{aligned}
\]

Fixed point analysis:

\[
\begin{pmatrix} -\alpha [A] & \beta \\ \alpha [A] & -\beta \end{pmatrix} \begin{pmatrix} 1 - P \\ P \end{pmatrix} = 0
\]

\[
\begin{aligned}
(1 - P)\alpha [A] &= \beta P \\
\alpha [A] &= P(\beta + \alpha [A])
\end{aligned}
\]

\[ P = \frac{\alpha [A]}{\beta + \alpha [A]} \]

This is a Hill function with \( n=1 \). Homodimer: \([A] \rightarrow [B \, B] = k [B]^2 \) (mass action).

\[ P = \alpha k [B]^2 / (\beta + \alpha k [B]^2) \]

This is a Hill function with \( n=2 \).
Transcriptional regulatory models – elementary binding sites

Some of these are demonstrated in the Cellator notebook transcriptionalmodels_2.nb.

Model 1. monomer + empty site $\leftrightarrow$ occupied site (initiates transcription)

Math as above. Transcription rate = Hills function (n=1)

Cellator:

$$\text{interpret}[[\{\text{TF} + \text{SiteEmpty} = \text{SiteOcc}, k_{\text{on}}, k_{\text{off}}\}]]$$

$$\{\text{SiteEmpty}[t] = k_{\text{off}} \text{SiteOcc}[t] - k_{\text{on}} \text{SiteEmpty}[t] \text{TF}[t],$$
$$\text{SiteOcc}[t] = -k_{\text{off}} \text{SiteOcc}[t] + k_{\text{on}} \text{SiteEmpty}[t] \text{TF}[t],$$
$$\text{TF}[t] = k_{\text{off}} \text{SiteOcc}[t] - k_{\text{on}} \text{SiteEmpty}[t] \text{TF}[t],$$
$$\{\text{SiteEmpty}, \text{SiteOcc}, \text{TF}\}$$

$$\text{Solve}[[0 = k_{\text{off}} \text{SiteOcc}[t] - k_{\text{on}} \text{SiteEmpty}[t] \text{TF}[t],$$
$$0 = -k_{\text{off}} \text{SiteOcc}[t] + k_{\text{on}} \text{SiteEmpty}[t] \text{TF}[t],$$
$$\text{SiteOcc}[t] + \text{SiteEmpty}[t] = 1],$$
$$\{\text{SiteOcc}[t], \text{SiteEmpty}[t]\}]$$

$$\{\text{SiteOcc}[t] \rightarrow \frac{k_{\text{on}} \text{TF}[t]}{k_{\text{off}} + k_{\text{on}} \text{TF}[t]}, \text{SiteEmpty}[t] \rightarrow \frac{k_{\text{off}}}{k_{\text{off}} + k_{\text{on}} \text{TF}[t]}\}$$

Model 2. homodimer + empty site $\leftrightarrow$ occupied site, monomer + monomer $\leftrightarrow$ homodimer.

Math as above. Replace A = kB^2. Transcription rate = Hills function (n=2)
Model 3. TF activation required. TF+K ↔ TF* + K, TF* + empty site ↔ occupied site.

Math: exercise for the reader.

Model 4. As above with heterodimers. Exercise.

Model 5. Four-state model:
Still missing: direct transition from empty to dimer-bound state by dimers created in solution.

Homodimeric case: see Cellerator notebook for general formula and for reduction to n=2 Hill function in case two conditions are met. They imply that the off rate for P (and/or Q) is much higher than that of the dimer, i.e. the second slot occupancy stabilizes the first. Example of that is Pho2-Swi5 in yeast; 2 min, 15 sec, and 20 minute binding half-lives.

```
interpret[{{TF + Site[0, 0] = Site[1, 0], kon, koff1},
{TF + Site[0, 0] = Site[0, 1], kon, koff1},
{TF + Site[1, 0] = Site[1, 1], kon, koff2},
{TF + Site[0, 1] = Site[1, 1], kon, koff2}}]

{[site[0, 0][t] -> koff1 koff2
  koff1 koff2 + 2 koff2 kon TF[t] + kon^2 TF[t]^2 },
[site[0, 1][t] -> koff2 kon TF[t]
  koff1 koff2 + 2 koff2 kon TF[t] + kon^2 TF[t]^2 },
[site[1, 0][t] -> koff2 kon TF[t]
  koff1 koff2 + 2 koff2 kon TF[t] + kon^2 TF[t]^2 },
[site[1, 1][t] -> kon^2 TF[t]^2
  koff1 koff2 + 2 koff2 kon TF[t] + kon^2 TF[t]^2 }]
```

Condition to obtain a Hill function, n=2:

```
(* if koff2 << kon TF[t] << koff1: 
koff2 kon TF[t] << kon^2 TF[t]^2, koff1 koff2 *)
```
question: stochastic modeling better for transcription factor binding?
Answer: depends on numbers and separation of time scale. E.g. even-skipped. Show eve data.
Example TF binding sites

*Example: Yeast MCM1*

![Diagram of TF binding sites]

Figure 4– Transcriptional regulation at a-specific gene (asg) operator in budding yeast, for ? and a/? cells. Redrawn from (Johnson 1995). Note multiple protein-protein and protein-DNA interactions in a complex. DNA is at the bottom and includes the asg opera

Source: Gibson and Mjolsness, “Modeling Single Genes …”

Molecular Mechanisms of Cell-Type Determination in Budding Yeast

*Example: Yeast Ste12.*

Aspects of these pathways are well conserved:
Figure 1. Yeast MAPK protein interactions and the CD/sevenmaker region. (A and B) Subset of the protein interactions in which the yeast MAPKs Kss1 and Fus3 participate. (C) Sequences of *D. melanogaster* rolled (GenBank/EMBL/DDBJ accession no. P40417), *H. sapiens* ERK2 (NP_620407), and *S. cerevisiae* Kss1 and Fus3 proteins (P14681 and S28548), in the CD/7m region. The COOH-terminal–most residue shown is indicated on the right. The site of the sevenmaker mutation in rolled (D334N) is underlined. Residues identical in all four proteins are denoted at the bottom by an asterisk.


A conserved MAPK protein interaction network. Shown are selected components of the yeast mating (A), mammalian ERK1/2 (B), and mammalian JNK (C) MAPK cascades. Components shown include the MAPKs themselves (circles), MEKs (ovals), phosphatases (unshaded rectangles), scaffolds (eight-sided polygons), and transcription factors (shown bound to DNA). D sites are indicated by a boxed or circled ‘D’. FXFP docking sites are indicated by a boxed or circled ‘F’. The CD/7m region of the MAPKs is indicated by a circled ‘7’.

Master equation and its solution

One unimolecular reaction

Consider the simple unimolecular reaction

\{0 \rightarrow A, k\}

deterministic model:
\[ \frac{dA}{dt} = k \]
solution is
\[ A = A_0 + kt \]

stochastic model:
\[ n \geq 0: \frac{dPr(A=n)}{dt} = k Pr(A=n-1) - k Pr(A=n) \]
i.e. “Master equation”:
\[ \frac{dPr(n)}{dt} = M Pr(n) \]
general solution:
\[ Pr(t) = [exp Mt] Pr(t=0) \]

here \( M_{nm} = k [\delta(n,m+1) - \delta(m,n)] = k (S-I)_{nm} \)
for integers \( n, m \geq 0 \) (infinite dimensional matrices)

Then we calculate
\[ S^T S = I, S S^T = I [1-\delta(n,0) \delta(m,0)] \]
\[ S S = \delta(n, m+2) \]
\[ S^p = \delta(n, m+p) \]

Taylor series expansion:
\[ \exp k(S-I)t = \exp -kt \sum_{p=0}^{\infty} S^p (kt)^p / p! \]
e.g. take \( Pr(n)(t=0) = \delta(n,n_0) \), i.e. a definite starting vector “\( | n_0 \rangle \)”.
\[ Pr(n,t) = \exp -k t (k t)^{(n-n_0)} / (n-n_0)! \] if \( n \geq n_0 \); zero if \( n < n_0 \).

This is a Poisson distribution in \( n-n_0 \). The Poisson distribution has the generating function
\[ Z(\lambda) = \sum_n \lambda^n / n! \],
from which we can compute that it has mean value
\[ <n> = n_0 + k t \]

(just like the deterministic solution) and variance \(<(n-<n>)^2> = kt\), hence standard deviation \(\sqrt{kt} = O(t^{1/2})\), as one would expect for a random walk. Also the standard deviation is proportional to (here, equal to) the square root of the mean number of molecules \(<n>\).

Exercise:

Perform the same calculation for

(a) \{A \rightarrow 0, k\}

hint:
\[ S_{nm} = m \delta(n+1,m) \]
\[ T_{nm} = m \delta(n,m) \]
\[ M_{nm} = k(S_{nm} - T_{nm}) \]
To calculate \(M^p\), use the “commutation” relationship
\[ ST = TS + S \]
to put all strings like STTSSSTSTT … into a sum of terms of the form \(T^{(p-k')}S^{k}\). Then multiply by a definite starting state \(|n_0>\).

(b) \{0 \leftrightarrow A, kf, kr\}

Note that this is like a two-rule stochastic grammar. Context free stochastic grammars have generating functions \cite{Flajoulet}.

(c) \{0 \rightarrow A, 0 \rightarrow B, A+B \rightarrow C, C \rightarrow 0\}
Solution for exercise (a) above.

In group theory, when multiplication isn’t necessarily commutative (as for our infinite dimensional matrices), commutators are defined as \([A, B] = AB - BA\).


\[ e^A B e^{-A} = B + \{A, B\} + (1/2!) \{A, [A,B]\} + (1/3!) \{A, [A, [A,B]]\} + \ldots . \]

Taking \(A = aS\) and \(B=bT\), where \(a\) and \(b\) are any scalars, we have

\[\begin{align*}
[A, B] &= ab [S,T] = ab S \\
[A, [A,B]] &= a^2 b [S,S] = 0 \\
&\quad \text{all higher terms} = 0
\end{align*}\]

and therefore

\[\begin{align*}
[\exp(aS), bT] &= ab S \exp(aS) \\
&\quad \text{(note that T is acting like a derivative operator)}
\end{align*}\]

which implies

\[\begin{align*}
\exp(aS) (bT) &= (bT + abS) \exp(aS) , \quad \text{and therefore} \\
\exp(aS) (bT)^n &= (bT + abS)^n \exp(aS) \quad \text{for all } n \geq 0.
\end{align*}\]

Now we can use the Taylor series expansion to calculate

\[\begin{align*}
\exp(aS) \exp(bT) &= \sum_n \exp(aS) (bT)^n/n! \\
&= \sum_n (bT + abS)^n \exp(aS)/n! \\
&= \exp(bT + abS) \exp(aS)
\end{align*}\]

whence

\[\exp b(T+aS) = \exp(aS) \exp(bT) \exp(-aS).\]

Taking \(a = -1\) and \(b = -kt\) we get

\[\exp Mt = \exp kt(S-T) = \exp(-S) \exp(-kt T) \exp(S).\]

We now need to calculate the matrix element

\[\Pr(nl m, t) = \langle nl \exp Mt | lm \rangle = \sum_{\{ij=0, \text{infty}\}} ((-1)^i/i!j!) \langle nlS^i \exp (-kt T S^j) lm \rangle.\]
From the definition of $S$,
\[
S^j |m\rangle = (m! / (m-j)! ) |m-j\rangle , = 0 \text{ if } j> m, \text{ and}
\]
\[
<n|S^i = <n+i| (n+i)! / n! , \text{ and}
\]
\[
\exp -kt T = \sum_p |p\rangle \exp -pkt <p| .
\]

Thus
\[
\Pr(n| m, t) = <n| \exp Mt |m> = \sum_{i,j,p=0, \infty} ((-1)^i m! (n+i)!/( i!j! (m-j)!n! ) \exp -kt p
\]
\[
<n+i | p><p | m-j>
\]
\[
(\text{last line } = \delta(n+i,p) \delta(p, m-j)
\]
\[
= \sum_i (-1)^i m!/(i!j!n!) \exp -kt (n+i)
\]

using the generating function
\[
g_m(x,y,z) = \sum_{i+j+n=m} x^i y^j z^n = (x + y + z)^m ,
\]
we can finally evaluate that
\[
\Pr(n| m, t) = <n| \exp Mt |m> = C(m,n) p^n (1-p)^m-n,
\]
where $C(n,m)$ is the binomial coefficient $n!/(m! (n-m)!)$ and $p = \exp -kt$. This is a Binomial distribution with parameters $m$ and $p$. Using its generating function $(1-p + xp)^m$, we can compute its mean and variance,
\[
<n> = mp
\]
\[
<(n - <n>)^2) = m(m-1) p .
\]
Note that at its maximal variance (for $p = 1/2$), the standard deviation of $n$ is again proportional to the square root of the mean $<n> .

This methodology can be tried out whenever the reactions can be decomposed into a set of operators whose commutators form a closed, finite algebra, even though they represent potentially infinite numbers of particles and therefore dimensions.
Generalization: Master equation to Langevin Equation.


2. How does it happen that the rigorous but computationally intractable CME,

$$\frac{\partial P(x, t | x_0, t_0)}{\partial t} = \sum_{j=1}^{M} \left[ a_j(x - v_j)P(x - v_j, t | x_0, t_0) - a_j(x)P(x, t | x_0, t_0) \right],$$

segregates for “large” systems to the heuristic but computationally efficient RRE,

$$\frac{dX(t)}{dt} = \sum_{j=1}^{M} v_j a_j(X(t)) ?$$

**Theorem:** If $\Delta t$ is a macroscopic infinitesimal, in that during $\Delta t$,
- no propensity function changes its value significantly, yet
- every reaction channel fires many more times that once,
then we can approximate the $t$ to $t + \Delta t$ system update by

$$X(t + \Delta t) = X(t) + \sum_{j=1}^{M} v_j a_j(X(t)) \Delta t + \sum_{j=1}^{M} v_j \sqrt{a_j(X(t))} N_j(t) \sqrt{\Delta t}.$$

Here, the $N_j(t)$ are statistically independent, temporally uncorrelated, normal random variables with means 0 and variances 1.

- This is the Chemical Langevin Equation (CLE).
- It approximates $X(t)$ as a continuous (versus a jump) Markov process.
- It is mathematically the same as the SDE

$$\frac{dX(t)}{dt} = \sum_{j=1}^{M} v_j a_j(X(t)) + \sum_{j=1}^{M} v_j \sqrt{a_j(X(t))} \Gamma_j(t),$$

where $\langle \Gamma_j(t) \Gamma_{j'}(t') \rangle = \delta_{jj'} \delta(t - t') \quad (j, j' = 1, \ldots, M).$

The corresponding chemical Fokker-Planck equation (CFPE) can be shown to be

$$\frac{\partial P(x, t | x_0, t_0)}{\partial t} = \sum_{j=1}^{M} \frac{\partial}{\partial x_j} \left[ \sum_{j=1}^{M} v_j a_j(x) \right] P(x, t | x_0, t_0)$$

$$+ \frac{1}{2} \sum_{j=1}^{M} \frac{\partial^2}{\partial x_j^2} \left[ \sum_{j=1}^{M} v_j^2 a_j(x) \right] P(x, t | x_0, t_0)$$

$$+ \sum_{i,j=1}^{M} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \left[ \sum_{j=1}^{M} v_j v_j a_j(x) \right] P(x, t | x_0, t_0)$$
Small Systems

Basic behaviors: fixed points, oscillations, and amplification

**Bistable fixed points**

E.g. a network with 2 Hills functions, as derived above for transcriptional regulation or other binding site dependent regulation:

```plaintext

gnet = {
     {A \rightarrow B, hill[vmax \rightarrow 1, nhill \rightarrow 2, khalf \rightarrow 1, basalRate \rightarrow 0]},
     {B \rightarrow A, hill[vmax \rightarrow 1, nhill \rightarrow 2, khalf \rightarrow 1, basalRate \rightarrow 0]},
     {A \rightarrow \emptyset, la1}, {B \rightarrow \emptyset, la2}
}
```

alternative regulatory mechanisms: transcriptional, protein modification (e.g. phosphorylation, complex assembly/disassembly, regulated protein degradation. Different cellerator notations and reaction translations for each of these.

Fixed points:

```plaintext
ParametricPlot[{{A, A^2 / (la (1 + A^2))},
               {A, Sqrt[(la A) / (1 - (la A))]}}, {A, 0, .90 * 1 / la}]
```

note that number of fixed points depends on n and, for n=1, on the slopes at the origin.
Oscillations

Negative feedback can build them. For example:

\[
\begin{align*}
\frac{dx}{dt} &= y \\
\frac{dy}{dt} &= -x
\end{align*}
\]

A biological example is provided by the mathematical model for NFκB signal transduction in Hoffmann, Levchenko, Scott, Baltimore. Science 298:1241, 2003.

Note mechanism depends on active degradation, transcriptional regulation, and protein complex formation to make a negative feedback loop.

Also, a three-node ring oscillator is defined by A --\| B --\| C --\| A, where the inhibitory connection “A --\| B” may be implemented transcriptionally (Hill function or GRN model), or by enzyme-catalytic regulation in which activated A deactivates B and/or deactivated A activates B (under mass action kinetics or Michaelis-Menton approximation leading to a Hill function regulation), or by protein complex assembly/disassembly, etc..

Electrical circuit oscillators can be built by positive feedback as well.
Minimal mitotic oscillator:


Note importance of protease. There is a whole network surrounding that, including Sic1 and its regulatory network neighbors.

```
GoldbeterMinimalSystem = {
  {σ = C, vi, kd},
  \{C \rightarrow σ, h\ill[\vmax \rightarrow \vd, \khalf \rightarrow \kd]\},
  \{\Comp[M] \rightarrow M, h\ill[\vmax \rightarrow \vl, \khalf \rightarrow \k1]\},
  \{M \rightarrow 0, h\ill[\vmax \rightarrow \v2, \khalf \rightarrow \k2]\},
  \{\Comp[X] \rightarrow X, h\ill[\vmax \rightarrow \v3+M[t], \khalf \rightarrow \k3]\},
  \{X \rightarrow 0, h\ill[\vmax \rightarrow \v4, \khalf \rightarrow \k4]\}
};
gms = interpret[GoldbeterMinimalSystem] /. 
{v1 \rightarrow \vM1+C[t] / (Kc+C[t])}
```

\[
\begin{align*}
  &\{c'[t] = \v1 - kd \cdot c[t] - \frac{\vd \cdot c[t] \cdot x[t]}{kd + c[t]}, \\
  &m'[t] = \frac{\vM1 \cdot c[t] \cdot (1 - m[t])}{(Kc + c[t]) \cdot (1 + K1 \cdot m[t])} - \frac{\v2 \cdot m[t]}{K2 + m[t]}, \\
  &x'[t] = \frac{\vM3 \cdot m[t] \cdot (1 - x[t])}{1 + \k3 \cdot x[t]} - \frac{\v4 \cdot x[t]}{K4 + x[t]}, \{c, m, x\}\}
\end{align*}
\]
Amplification (??): MAPK cascade

$K_3 \leftrightarrow K_2$ etc written out here.

Negative feedback loops may go from MAPK*, transcriptionally, to phosphatases for the last or the first steps in the above pathway.
Network level yeast signal transduction models:

Bayes net inferred from expression and p:dnase binding data for yeast cell cycle. Compare to KEGG, other curated pathways.

Network level yeast cell cycle models:


Note the presence of many nontranscriptional mechanisms for regulation: protein phosphorylation, complex formation, protein degradation, etc.

---

![Diagram](image1)

**Fig. 1.** The cell cycle. Outer ring illustrates the chromosome cycle. The nucleus of a newborn cell contains unreplicated chromosomes (represented by a single bar). At Start, the cell enters S phase and replicates its DNA (initiated by replication bubbles on the "chromosome"). At the end of S phase, each chromosome consists of two sister chromatids (X held together by tethering proteins). After a gap (G2 phase), the cell enters M phase, when the replicated chromatids are aligned on the metaphase plate, with sister chromatids attached to opposite poles of the spindle. At Finish, the sister proteins are removed so that the sister chromatids can be segregated to opposite sides of the cell (anaphase). Shortly thereafter the cell divides to produce two daughter cells in G1 phase. The inner icons represent the fundamental molecular machinery governing these transitions. Start is triggered by a protein kinase, Cdc2, whose activity depends on association with a cyclin subunit. Cdc2 activity drives the cell through S phase, G2 phase, and up to metaphase. Finish is accomplished by proteolytic machinery, APC, which destroys the tethers and cyclin molecules. In G1 phase, APC is inactive and Cdc2 inactive, because it lacks a cyclin partner. At Start, the APC must be turned off so that cyclins may accumulate. Cdc2 and APC are antagonistic proteins: APC destroys Cdc2 activity by degrading cyclins, and cyclin/Cdc2 dough inhibits APC by phosphorylating one of its subunits.

---

![Table](image2)

**Table 2.** Cell cycle regulatory proteins in yeasts and culminates.

<table>
<thead>
<tr>
<th>Component</th>
<th>Binding partner</th>
<th>Phosphorylation</th>
<th>Metaphase control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cdc2</td>
<td>Cdc2</td>
<td>Cyclin B</td>
<td>Cdc2</td>
</tr>
<tr>
<td>CycB</td>
<td>Cdc2</td>
<td>Cyclin B</td>
<td>Cdc2</td>
</tr>
<tr>
<td>Cdc1</td>
<td>Cdc2</td>
<td>Cyclin B</td>
<td>Cdc2</td>
</tr>
<tr>
<td>Cdc20</td>
<td>Cdc2</td>
<td>Cyclin B</td>
<td>Cdc2</td>
</tr>
<tr>
<td>IE</td>
<td>Cdc2</td>
<td>Cyclin B</td>
<td>Cdc2</td>
</tr>
<tr>
<td>CKI</td>
<td>Cyclin B</td>
<td>Cyclin B</td>
<td>Cdc2</td>
</tr>
<tr>
<td>SK</td>
<td>Cyclin B</td>
<td>Cyclin B</td>
<td>Cdc2</td>
</tr>
</tbody>
</table>

---

![Diagram](image3)

**Fig. 6.** The basic cell cycle engine in eukaryotic cells. The generic components in this mechanism correspond to specific gene products in well-studied organisms (see Table 2). For details and parameters, see the referenced literature.
Hysteresis (flip/flop) diagram for Novak/Tyson budding yeast model (compare to our fixed point analysis above)

\[
\frac{d[C_{yB}]}{dt} = k_1 - (k_2' + k_2''[Cdh1])[C_{yB}], \quad (1)
\]

\[
\frac{d[C_{dh1}]}{dt} = \frac{(k_3' + k_3''A)(1 - [Cdh1])}{J_3 + 1 - [Cdh1]}
\]

\[
-\frac{k_4m[C_{yB}][C_{dh1}]}{J_4 + [C_{dh1}]}. \quad (2)
\]

**Fig. 3.** Bifurcation diagram for eqns (1) and (2). The steady-state concentration of CycB is plotted as a function of the bifurcation parameter, \( p = (k_3' + k_3''A)/k_4m \). Other parameters: \( \beta = \delta = 0.04 \). Saddle-node bifurcations occur at \( p_1 \approx 0.0545 \) and \( p_2 \approx 0.261 \).
Use of protein:DNA binding data to define transcriptional regulation in yeast cell cycle

**Fig. 4.** Model for the yeast cell cycle transcriptional regulatory network. A transcriptional regulatory network for the yeast cell cycle was derived from a combination of binding and expression data (see text). Yeast cell morphologies are depicted during the various stages of the cell cycle. Each blue box represents a set of genes bound by a common set of regulators and coexpressed throughout the cell cycle. Text inside each blue box identifies the common set of regulators that bind to the set of genes represented by the box. Each box is positioned in the cell cycle according to the time of peak expression levels for the genes represented by the box. Regulators, represented by ovals, are connected to the sets of genes they regulate by solid lines. The arc associated with each regulator effectively defines the period of activity for the regulator. Dashed lines indicate that a gene in the box encodes a regulator found in the outer rings.


Compare to protein-level model – Novak and Tyson
Figure 5 Global view and higher order organization of modules. The graph depicts inferred modules (middle; numbered squares), their significantly enriched cis-regulatory motifs (right; significant motifs from Fig. 4a) and their associated regulators (left; ovals with black border for transcription factors or with green border for signal transduction molecules). Modules are connected to their significantly enriched motifs by solid blue lines. Module groups, consisting of sets of modules that share a common motif, and their associated motifs are enclosed in bold boxes. Only connected components that include two or more modules are shown. Motifs connected to all modules of their component are marked in bold. Modules are also connected to their predicted regulators. Red edges between a regulator and a module are supported in the literature: either the module contains genes that are known targets of the regulator (Table I, G column) or upstream regions of genes in the module are enriched for the cis-regulatory motif known to be bound by the regulator (Table I, M column). Regulators that we tested experimentally are marked in yellow. Module groups are defined as sets of modules that share a single significant cis-regulatory motif. Module groups whose modules are functionally related are labeled (right). Modules belonging to the same module group seem to share regulators and motifs, with individual modules having different combinations of these regulatory elements.

References on yeast network models:

Continuous-valued Bayes Net inference:

Bayes network for yeast heat shock microarray dataset.


Cf. their PSB 2002 paper showing regression for continuous-valued Bayes Networks.

Imoto et al. 2002. Mostly, this network doesn’t agree with the Kegg one or the others.
Conservation of gene networks:


Integrating sequence information:


Yoseph Barash, Gal Elidan, Nir Friedman, Tommy Kaplan, Modeling Dependencies in Protein-DNA Binding Sites RECOMB 03, http://www.cs.huji.ac.il/labs/combio/TFBN
MAPK and other signal transduction pathways in cancer research.

Apoptosis (homo sapiens):
References


Yoseph Barash, Gal Elidan, Nir Friedman, Tommy Kaplan, Modeling Dependencies in Protein-DNA Binding Sites RECOMB 03, http://www.cs.huji.ac.il/labs/compbio/TFBN/

