277C Computational Systems Biology


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 → 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]. 

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’.
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

Anasua B. Kusari, Douglas M. Molina, Walid Sabbagh, Jr., Chang S. Lau, and Lee
Bardwell. A conserved protein interaction network involving the yeast MAP kinases Fus3

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:

\[
\begin{align*}
\text{site}_\text{empty}(t) & \rightarrow \text{site}_\text{occupied}(t+\Delta t) \quad \text{with Pr}(e \rightarrow o \mid \Delta t) \\
\text{site}_\text{occupied}(t) & \rightarrow \text{site}_\text{empty}(t+\Delta t) \quad \text{with Pr}(o \rightarrow e \mid \Delta t)
\end{align*}
\]

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?]:
Solution of linear master equation:

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

Fixed point analysis:

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

\[
(1 - P)\alpha[A] = \beta P
\]

\[
\alpha[A] = P(\beta + \alpha[A])
\]

\[
P = \frac{\alpha[A]}{\beta + \alpha[A]}
\]
MAPK and other signal transduction pathways in cancer research.

Apoptosis (homo sapiens):


RTK

Survival

Transformation

Cytoskeletal rearrangement

Transformation

p53

Growth

Translation

Apoptosis

Cell cycle arrest

Cell cycle

Glucose metabolism