“Building with a scaffold: emerging strategies for high- to low-level cellular modeling”

Trey Ideker$^1$ and Douglas Lauffenburger$^2$

$^1$Whitehead Institute for Biomedical Research, Cambridge, MA
$^2$Massachusetts Institute of Technology, Biological Engineering Division, Cambridge, MA

*TRENDS in Biotechnology Vol. 21 No. 6 June 2003*

BBSI Journal Club
June 17, 2004

**Presented By:**

Deanna Nachreiner, Nick Morsillo, Jordan Torok
Introduction to Modeling...

- Used to see what we cannot in wet labs
- Basis of models must come from concrete physiological data
- Modeling takes place at the molecular, sub-cellular, cellular and tissue levels
- In this paper: computational models are executed based on biological pathway scaffolds at different levels of abstraction
Why Study Biological Pathways?

- **Focus shift:** from studying individual protein function to concentrating on how proteins interact together in biological pathways
  - 30-40,000 genes = many more different proteins
  - Tedious and inconclusive!
- Many different pathway types: signaling, regulatory, metabolic, structural
  - ligand binding events, growth, regulation of cell cycle, glycolysis, and cell movement processes
- Through physiological pathway data, better computational simulations can be done, which will lead to new hypotheses to be studied in the lab. These experiments will lead us to new and exciting applications of models in the fields of biology and medicine.
- Understanding SCAFFOLDS: an overall network of the different molecular interactions seen throughout signaling pathways
Goals of this Paper...

- Review ways that pathway scaffolds at different levels of abstraction are built and used
- Examine the difference between “high-” and “low-level” computational models
- Present the need for a systematic approach for building and manipulating pathways
- Assess the phases of experiment design
- Look to the future of computational pathway models and their applications in biology
Identification and Characterization of Pathways...

**Method: Molecule-Directed**

- “Bait” protein identified, other proteins in pathway are identified next, a pathway is constructed molecule by molecule basis

- Problems: time consuming, no revealing of cross-talk data, lack of data organization
Identification and Characterization of Pathways...

**Method:** Systematic Development of Comprehensive Scaffolds of Molecular Interactions

- Details: requires a lot of work initially, but accurately allows for mapping and querying of pathways in an organized fashion

- Broadly covers aspects of cellular function and physiological responses
  - Different levels of abstraction...
“High-Level” or L1 Models

- High level of abstraction, low levels of detail

- Identify basic components and connections in the pathway of interest – build a framework

- Correlate dependent and independent variables to examine possible interrelationships between components
“Low-Level” or L2 Models

- Low level of abstraction, high levels of detail

- Shows how biological information flows from one component to another and the affects of kinetics and binding affinities on the system

- Markov Chains and Differential Equations are used at this level to predict component states and to model rates and diffusion of the system
The Spectrum of High-to Low-Level Modeling Approaches
The Problem with Modeling...

Determining how to use data acquired in high-level models to produce more detailed, computationally-useful low-level models, and further, how to bridge this information in order to maximize its use!

In short: getting from L1 to L2!
Constructing L1 Models

1. Create a Global Molecular Interaction Scaffold

2. Filter scaffold against pathway changes from perturbations

3. Assemble the parts of the pathway that are “activated”

This process creates an L1 model that is composed of only the relevant pieces of data which can then be used for lower-level modeling.
Characterizing Networks and Component States...

- **Step 1 in building high-level models**
- **Helps to identify interactions between molecules and their connectivities**

**Types of Interactions to follow:**
- Protein-Protein
- Protein-DNA
- Small Molecule

**Ways to follow them:**
- Measure Molecular Interactions
- Observe States Induced by Interactions
Protein-Protein Interactions...

- **Measured by:** yeast two-hybrid system, co-immunoprecipitation followed by mass spec
- Information stored in databases like:
  - BIND
  - DIP
  - BRITE
  - MIPS
- **Observed with:** mass spec, 2D PAGE
- Information stored in databases like:
  - SWISS-2D PAGE
  - TRIPLES
  - Scansite
Protein-DNA Interactions...

- **Measured by:** chromatin immunoprecipitation with microarray analysis
- Information stored in databases like:
  - TRANSFAC
  - BIND

- **Observed with:** DNA microarrays, SAGE
- Information stored in databases like:
  - GEO
  - ArrayExpress

IHF interacting with DNA
Small Molecule Interactions...

- **Measured by:** Protein Arrays
- Information stored in databases like:
  - MetaCyc
  - KEGG
  - Klotho
- **Observed with:** mass spec, 2D NMR
- No databases of metabolic profiles currently exist

Following molecules like: carbohydrates, lipids, drugs, hormones, etc.
Continued Construction of L1 Models

- The next steps include:
  - Comparing the scaffold with pathway changes from cellular experiments under varying conditions
  - Assembling the parts of the pathway that are of interest – the parts that we may wish to research further in the future
L1 Construction Process...
Use of Probabilistic Approaches to Create L1 Models

- Begin with a group of genes that are expressed simultaneously
- Choose small groups of known transcription factors the evaluate and record their effects on the system
- Include only those transcription factors at specific levels which produce the maximal levels of expression observed initially

This method builds up, adding only known, useful information while the previous method works backwards from the final product (the entire scaffold), carving away at useless information
Pathway and Scaffold Searches

- Identification of pathways and scaffold types is not limited to the types of interactions mentioned earlier, many different kinds exist
  - Jensen *et al.* based their scaffold on known gene-gene associations by directly linking any 2 genes found to have been previously co-expressed in experiments
  - Matthews *et al.* based their scaffold not on the protein-protein interactions in yeast, but against a similar scaffold from *C. elegans*, including only those protein interactions found in both species
The Next Step: Visualization

- Many programs are currently available to view interaction scaffolds
  - Osprey, PIMRider, GenoMax, Cytoscape, Pathway Tools
  - Provide network visualization, layout, shows activated network against expression data, etc.

- DNA Microarrays are used to produce gene expression profiles and measure cellular states because they are widely available
  - though this method leaves out any contributing information at the small molecule level – leaving a fragmented pathway
Constructing L2 Models

- Used to predict cell behaviors at a very detailed physiological level
  - Example: altered DNA sequence → altered protein structure → altered rate constant → signaling pathways are changed → changes in cell growth or movement, signal processing

- Problem: lack of developed molecular networks and connectivity data for use in low-level modeling

- Goal: to increase the amount of experimental data at the molecular level in order to increase the amount of new data that can be retrieved from computational modeling
From L1 to L2 –
A Systematic Approach

- **Beginning at the L1 level:**
  - Representing key components
  - Model influences on each other
  - Model influences on surroundings
  - Bayesian Network Model

- **Look to Subsets of Networks for component states:**
  - Boolean Logic
  - Interpret and predict cell behaviors, test

- **Proceed to L2 level:**
  - Detailed reaction information
  - Kinetics
  - Binding
  - Activation
The Systematic Approach

- While no current model exits...
- Circuit Design Example
  - Large digital circuits:
    - Components: logic gates, memory units, timers, counters, clocks
  - Lower Level:
    - Circuit Layouts: logic-gate geometry (2D and 3D)
  - Still Lower:
    - Simulate analog behavior: use resistors, capacitors, transistors, batteries, etc.
- Bridge levels together
From L1 to L2 – An Example

- Ronen et al.

E. Coli with GFP reporter plasmid → High-resolution kinetics experiments on parallel reporter strains → Promoter activities vs. time → Comparison to parametrized trial function

Network of known structure → Active regulator concentration

Effective parameters: $k_1, k_2, k_3, k_n$ → $A(t)$

$X_i(t) = \frac{b}{1 + A(t)/k_i}$

TRENDS in Biotechnology
Experimental Design Phases

- **Phase A** — examine entire systems, consider all possibilities, perturb system for certain hypotheses and observe results used to build L1 models (COVERAGE)

- **Phase B** — target specific pathway components, focus only on pathways of interest, observe results used to build L2 models (LEVERAGE)
The Future of Pathway Modeling

- Currently computer modeling is used in almost every industry
  - Engineering Disciplines

- Systematic Approach + Bridge Gap between L1 and L2
  - Software development

- Applications in the fields of molecular biology and pharmacology are endless
  - Hypotheses made at the L2 level can be tested at multiple pathways as well as the L1 level, seeing the entire result
  - Example: pharmacological drug testing and side effects
Critique…

- While pathway scaffolds do show a promising future, aspects in real life physiology are left out of this model:
  - Cellular ultrastructure does not allow for these processes to occur as naturally as the paper presents
  - Fails to recognize how much of a scaffold component is present and how sensitive the pathway will be to changes in one component’s concentration
References

- http://www.chem.wisc.edu/~weisshaar/Frames/dna.htm
- http://www.chem.duke.edu/~toone/labgroup/PCInt.html
- http://www.ai.mit.edu/~murphyk/Bayes/la.times.html
- http://www.calresco.org/boolean.htm
- http://www.biochem.arizona.edu/classes/bioc568/two-hybrid_system.htm
Acknowledgments

- Joel Stiles, MD, PhD
- John Pattillo, PhD
- Jordan Torok, Nick Morsillo
- Everyone at BBSI
- University of Pittsburgh
- Carnegie Mellon University
- Pittsburgh Supercomputing Center
- NIH/NSF
- Mark Connell - caterer
Yeast 2 Hybrid System

Two-hybrid system

Structure-function properties of a typical transcription factor:

- DNA-binding domain (DBD)
- Activation domain (AD)
- Reporter gene

Two-hybrid system: two types of hybrids:

- DBD (or domain) of interest ("bait")
- Interacting protein (or domain) ("prey")

By itself, the DBD:bait fusion does not stimulate expression.
When bait and prey interact, the reporter gene is expressed.