

BioInformatic Tools and Genomes To Life

By Karan Sundlass

Goals

- Identify and characterize the molecular machines of life -- the multiprotein complexes that execute cellular functions and govern cell form.
- Characterize gene regulatory networks.
- Characterize the functional repertoire of complex microbial communities in their natural environments at the molecular level.
- Develop the computational methods and capabilities to advance understanding of complex biological systems and predict their behavior.

What computational methods are needed?

- Refining raw data from lab
- Analyzing data for useful information
- Incorporating data from many different sources to see overall patterns.

Refining Raw Data

DNA



**Genome
Sequence**

RNA



ESTs

Protein



**Sequence and
Structure**

DNA and RNA

- Assemble contigs and mRNAs into whole sequences
- Finding sequencing errors, alternative splicing
- Reading microarrays to determine expression patterns

Proteins

- Original data comes from sequencing proteins through digestion and chemical analysis

Proteins (Mass Spec)

- Proteins are run on 2D-PAGE
- Spot is digested, peptide fragments run through a mass spectrometer
- Data is run through a database of mass spec data, like Protein Prospector and compared
- Confirmation accomplished through further digestion and mass spec

Proteins (X-Ray Crystallography)

- Proteins must be grown and crystallized
- Useful to determine protein structure

Analyzing Data

**Genome
Sequence**

GRAIL, GenScan

**Genes,
Regulatory Regions**

ESTs

CAT, ArrayViewer

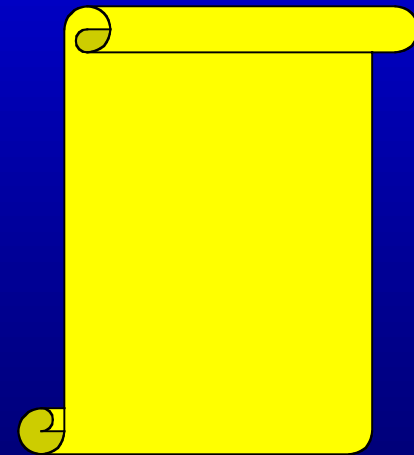
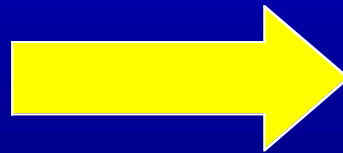
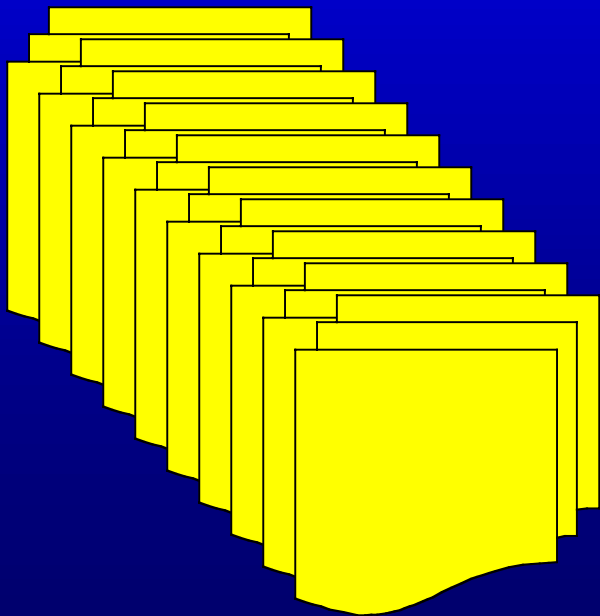
Expression Pattern

**Protein
Sequence**

Blast, Motif tools

Function and form

Analyzing Data



Incorporation of Related Data

Right now, this information is being generated at a very fast pace, and is being stored in many different databases:

- Annotated genome databases
- EST cluster databases
- Protein homology databases

Incorporation of Related Data

Using current tools, the following information can be obtained for a gene:

- Sequence
- When and where it is expressed
- What it codes for

Genomes to Life

Function

Expression Pattern

Gene Ontology

Gene Regulatory
Networks

Metabolic
Pathways

Complete Response Map

Gene Regulatory Networks (GRN)

- The determination of the relationships between genes requires the use of all the tools presented.
- A study that mapped part of the cell cycle regulatory network shows the the use of bioinformatic tools to elucidate the network.

Global Analysis of the Genetic Network Controlling a Bacterial Cell Cycle. Laub, M. et al

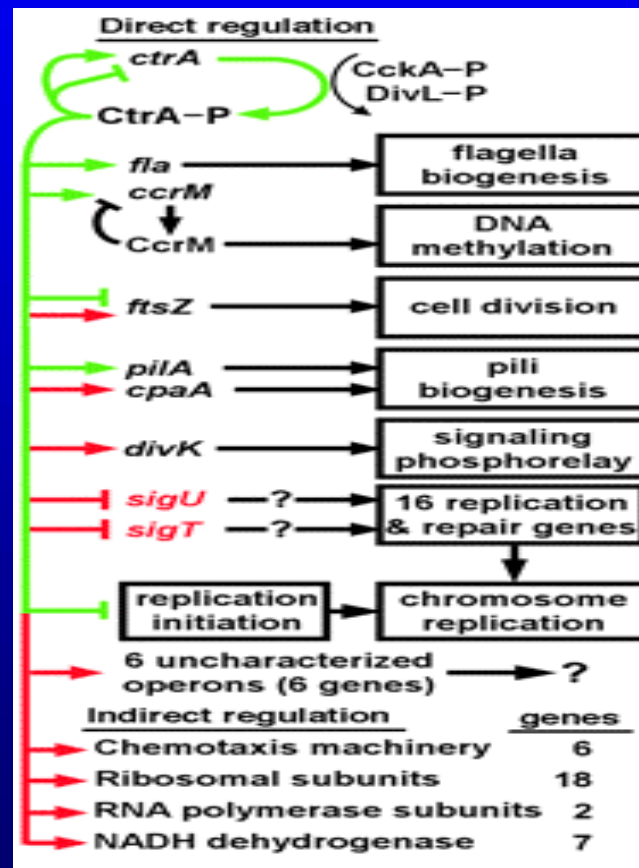
Science, December 15 2000. V. 290 2144-2148.

Gene Regulatory Networks (GRN)

- Using DNA and RNA microarrays, made expression profiles at different times.
- Identified genes that were expressed in cycles with a transform algorithm.
- Clustered genes that expressed together
- Noted changes in phenotypes at same times
- Mutated CtrA, a master regulator that inhibits the expression of G1- S phase factors.

Gene Regulatory Networks (GRN)

- Results:
 - CtrA regulates 26% of cell cycle regulated genes.
 - Constructed a GRN for CtrA



Metabolic Pathways

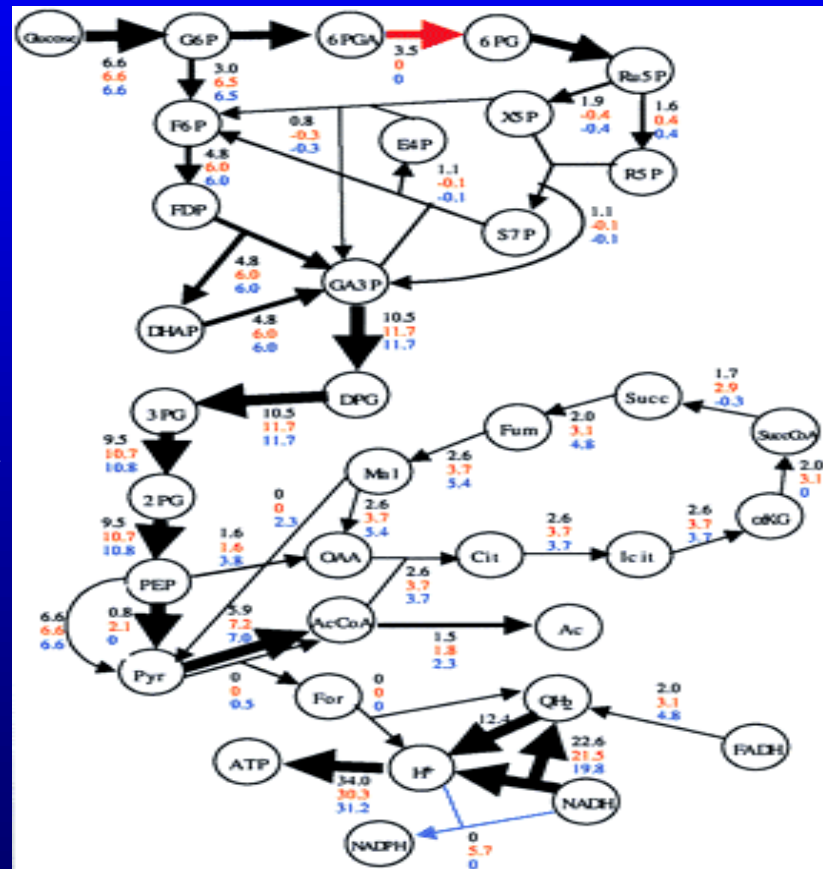
- This study constructed a computer model of the metabolic pathway of glucose, and tested the accuracy of the model against experimental observations.
- ***The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities.***
J. S. Edwards and B. O. Palsson
- ***Proc. Natl. Acad. Sci. USA, Vol. 97, Issue 10, 5528-5533, May 9, 2000v***

Metabolic Pathways

- Used biochemical properties of reactions between metabolic enzymes and metabolites to create a matrix of interactions.
- Matrix was 436 metabolites by 720 enzyme catalyzed reactions.
- Found best model with flux balance analysis of matrix and an optimizing algorithm, very computationally intensive.
- Compared predictions of deleted intermediaries to experimental observations.

Metabolic Pathways

- Results:
 - 86% of predictions were experimentally observed to be true
 - Essential central metabolic enzymes of glucose pathway determined



So What?

- With the continuing discovery of new GRNs and metabolic pathways, the rules that shape them are being more clearly understood.

So What?

- Predictive methods are improving, and it should be possible to recognize systems that have not been studied from sequence, expression, and functional data.

Applications

- As the rules regulating organisms are understood, the ability to manipulate them will grow.
- Uses:
 - Bioremediation
 - Carbon Sequestration
 - Defense from Biological and Chemical Attacks
 - Tests for worker susceptibility to hazardous materials

Not So Fast

- Just as the highly automated technology used to sequence DNA allowed genomic projects to proceed, automated techniques need to be developed to find GRNs and metabolic pathways.
- The algorithms to interpret and store this data efficiently must also be implemented.

Genomes to Life

- The fulfillment of the Genomes to Life project's goals will be greatly facilitated by the development of these predictive models and tools to quickly find homologies in these networks, just as the determination of protein function was accelerated by tools that found protein homology.

The Future

- While the DOE plans to mainly focus on microbial organisms, the extension of these models will help the discovery of human analogs.

The Future

- The GRN of a human would allow immediate diagnosis and identification of drug targets for almost any genetic disease

The Future

“Deciphering the complex relation between the genotype and the phenotype will involve the biological sciences, computer science, and quantitative analysis, all of which must be included in the bioengineering of the 21st century.”

References

- Dr. Brutlag
- www.doegenomestolive.org
- “Computing the Genome”, Ed Uberbacher,
www.ornl.gov/ornlreview/v30n3-4/genome.htm
- *Global Analysis of the Genetic Network
Controlling a Bacterial Cell Cycle.* Laub, M. et al
Science, December 15 2000. V. 290 2144-2148.
- *The Escherichia coli MG1655 in silico metabolic
genotype: Its definition, characteristics, and
capabilities.* J. S. Edwards and B. O. Palsson.
Proc. Natl. Acad. Sci. USA, Vol. 97, Issue 10,
5528-5533, May 9, 2000v