Enabling Technologies from the Biology Perspective

H. Steven Wiley

January 22nd, 2002

U.S. Department of Energy
Pacific Northwest National Laboratory

What is a “Systems Approach” in the Context of Biological Organisms?

- Looking at cells as integrated systems and not as mere collections of parts
- Seeks to understand how changes in any given component will affect the behavior of the entire system
- Requires integration of information from molecular biology, biochemistry, physics and genetics to understand relationship between different sets of scientific data.
- Goal is a “Higher order” understanding of life processes.

Why should DOE care about Systems Biology?

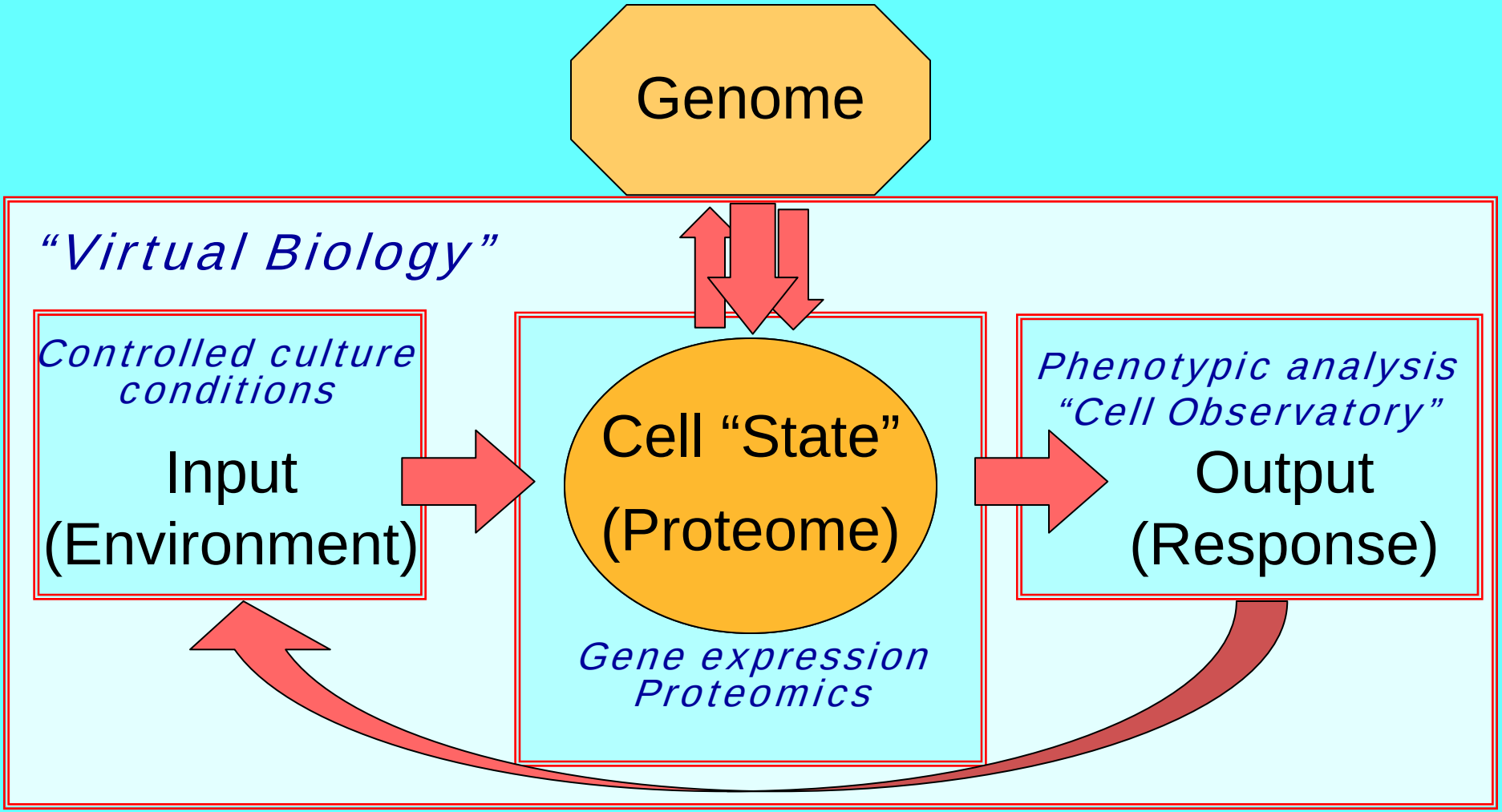
- It offers a chance to actually solve problems.
- Conversion of biology from a descriptive to a quantitative science.
- This will enable bioengineering and rational design of biological systems.
- DOE is one of the only organizations that is capable of applying a systems approach to biology

What are the Computational Requirements of Systems Biology?

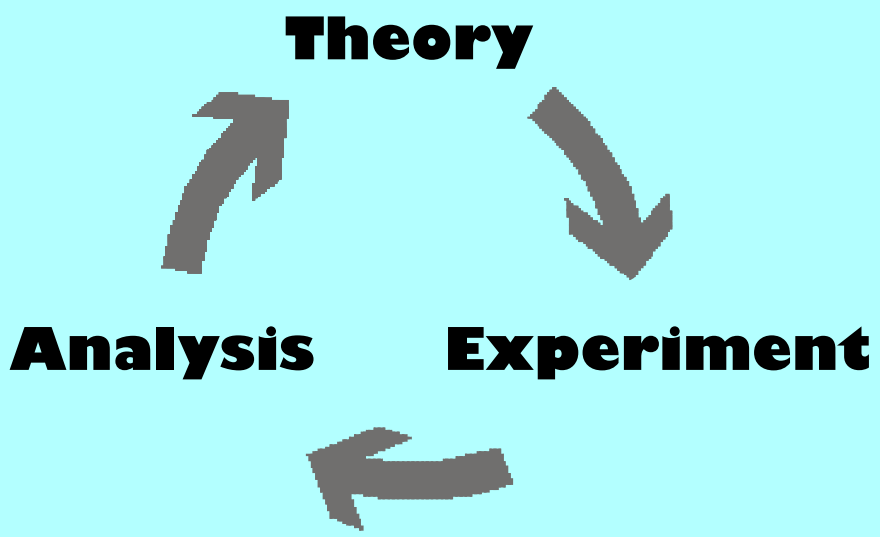
In order to integrate the data and move biology to being a system science, we need to:

- develop new theories and mathematics,
- develop new algorithms,
- implement on the largest available computer systems,
- make full use of large, distributed databases, and
- make the software and computer system usable by the biologist.

Biological Organisms are Dynamic and Recursive



Systems Biology Requires Scientific Integration

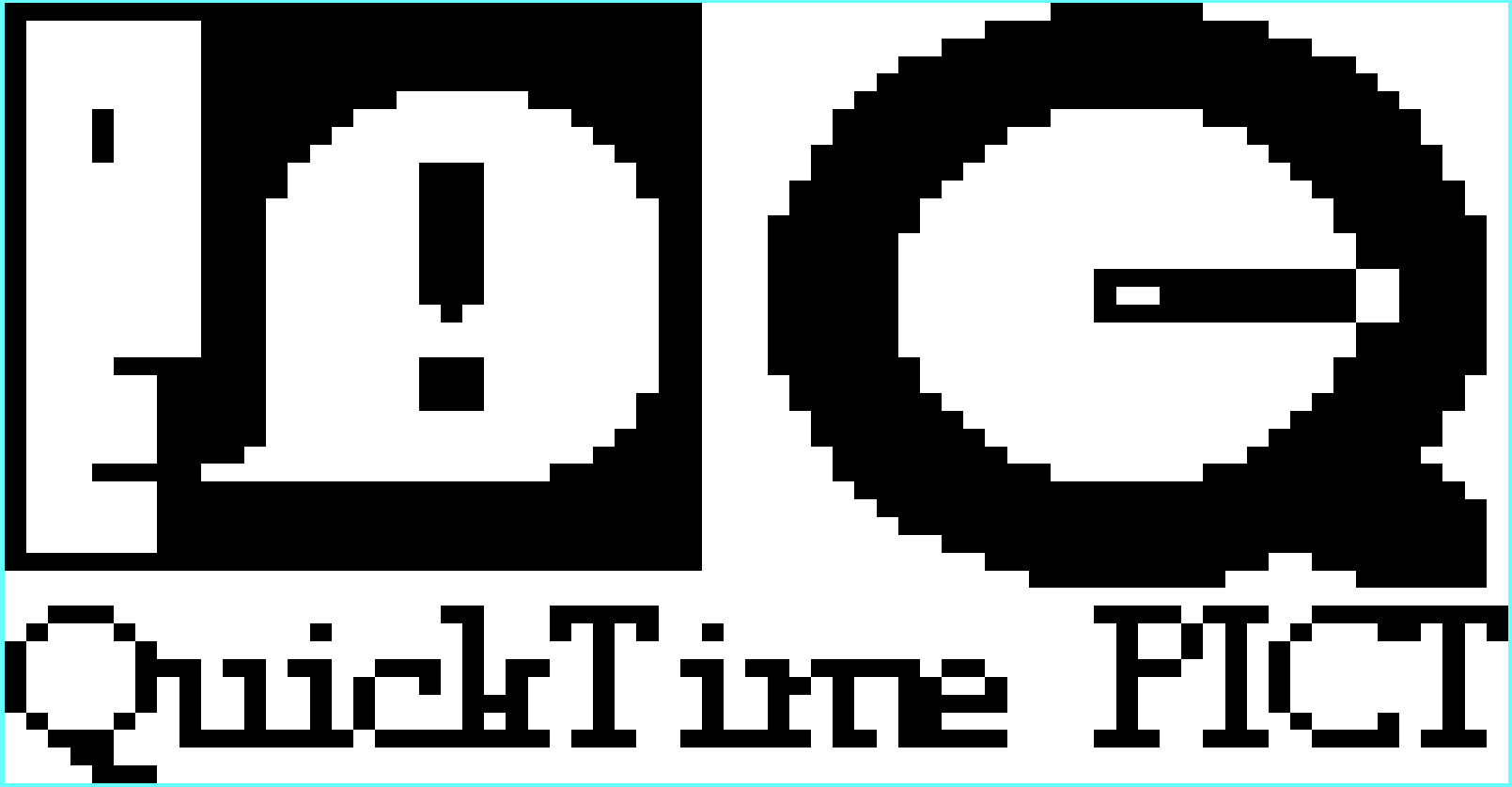


To solve specific problems using a "systems approach", you must couple theory with experiments and analysis in a recursive manner. This requires close coordination between program elements.

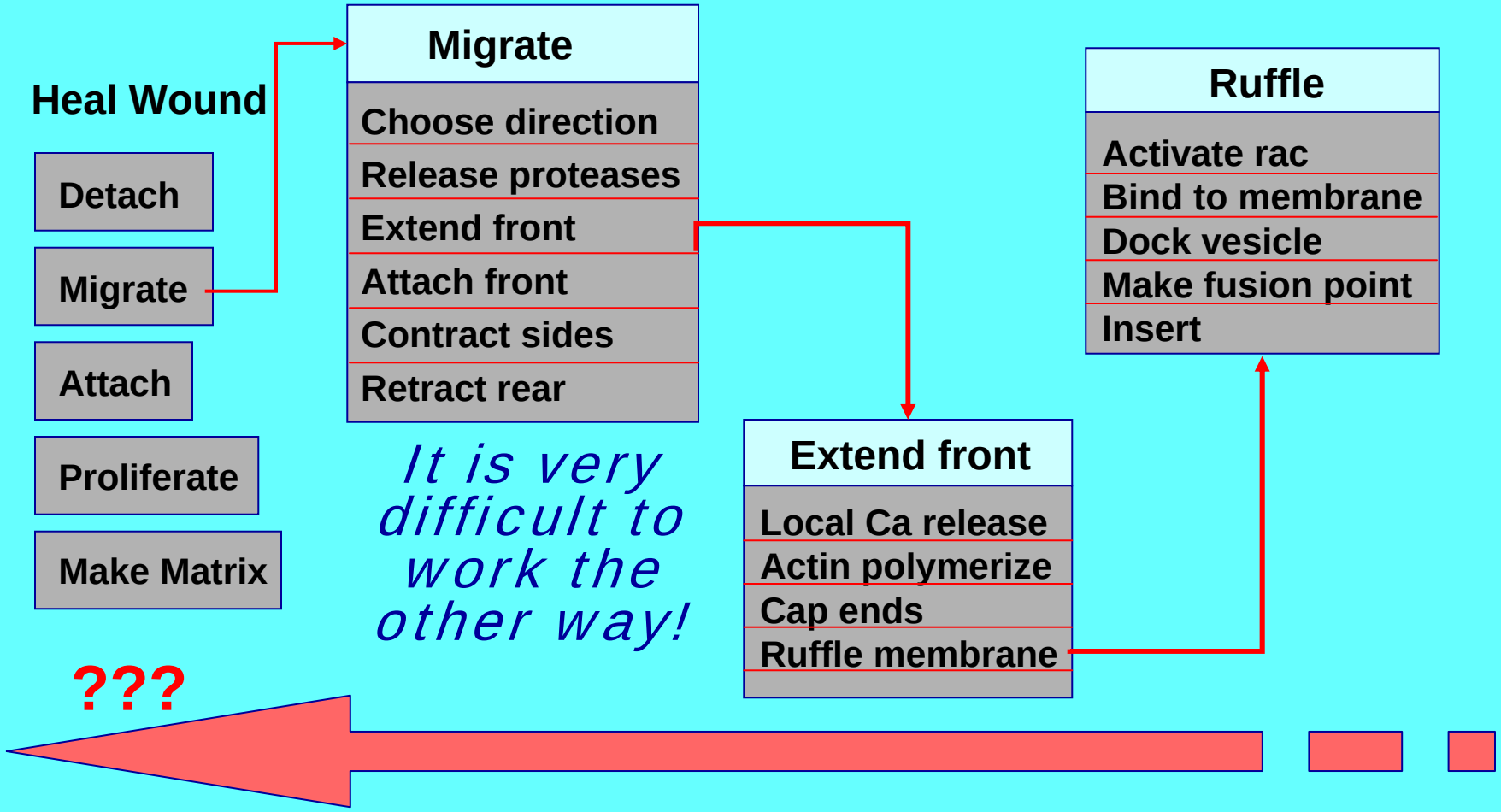
Understanding Complex Biological Systems Using a Top-Down:Bottom-Up Hybrid Approach

- Bottom-Up defines the molecular components of the system
  - This sets the physical constraints of system behavior
  - This provides basis for generating mechanistic models
- Top-Down computational approach is used to attribute behavioral characteristics to system
  - Encapsulate complexity in terms of modular structure
  - Attribute behavioral outcome to molecular composition

Cell Migration Can Be Broken Down Into a Series of Steps...



Extend Model From High-level to Low Level Example: Wound Healing Response



Modular Perspective in Cell Biology

- Consider cells as composed of groups of interacting modules
- Modules are functionally independent
- Modules interact at defined (and limited) regulatory nodes
- Function of cells defined in terms of collection of modules

Systems Biology = Quantitative Biology

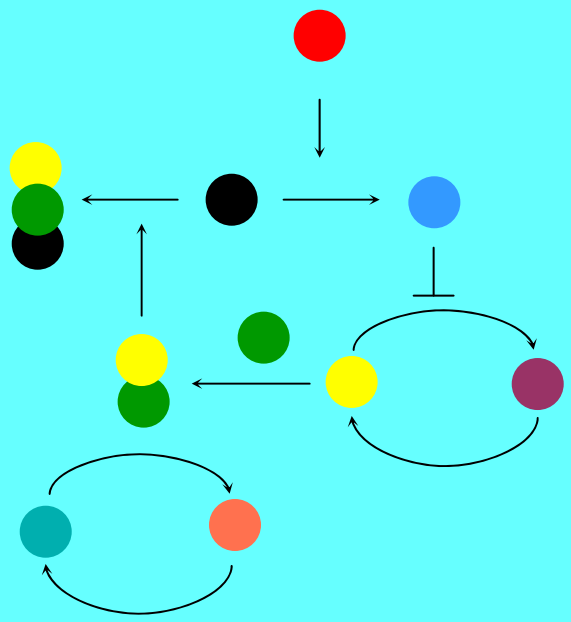
- System is defined as group of interacting parts
- Interactions are dependent on quantities
- Cellular processes are best described as rates
- Computer simulations require quantification

What, Where, Quantity, Quality?

To successfully model a complex biological system, one must minimally know the following information:

- What parts are being made? (identity)
- Where are the proteins located in cell? (location)
- What are their levels? (quantity)
- How do they interact with their partners? (activity)
  - As a function of covalent modification
  - Contribution of steric restrictions
  - Forward and reverse rate constants

What Is Necessary to Understand a Network?



Nodes
(Components)

Connectivity
(Interactions)

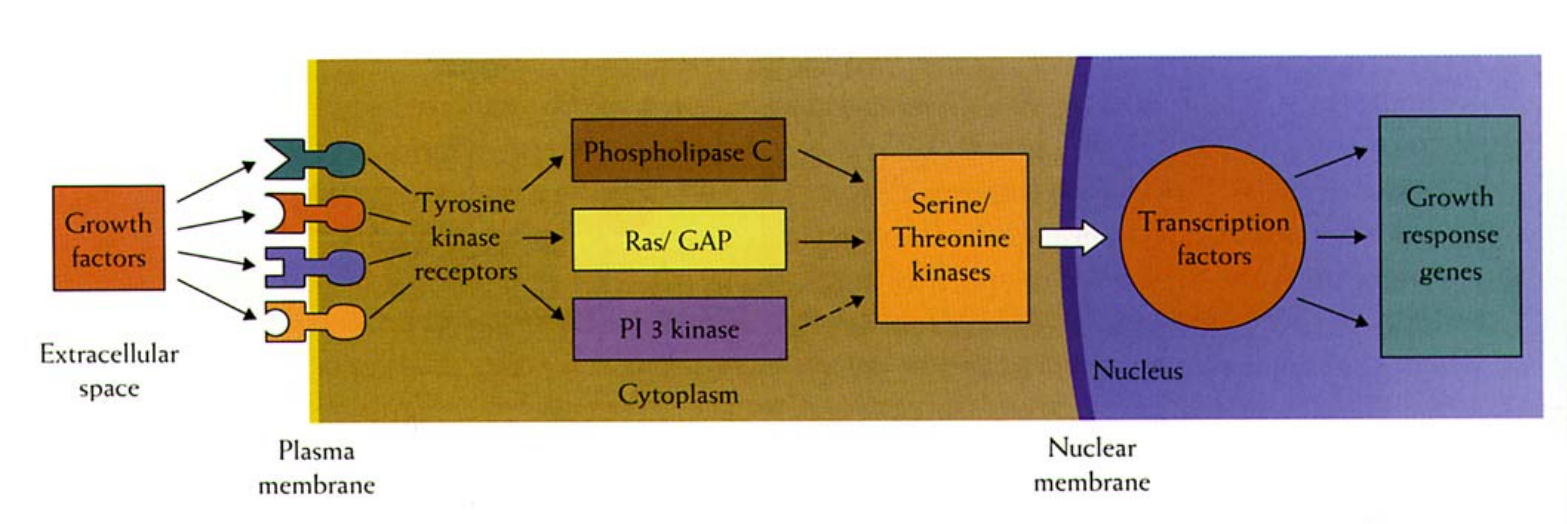
Experimental Biology

Spatial Organization
(Transport)

Dynamic Range
(Kinetics)

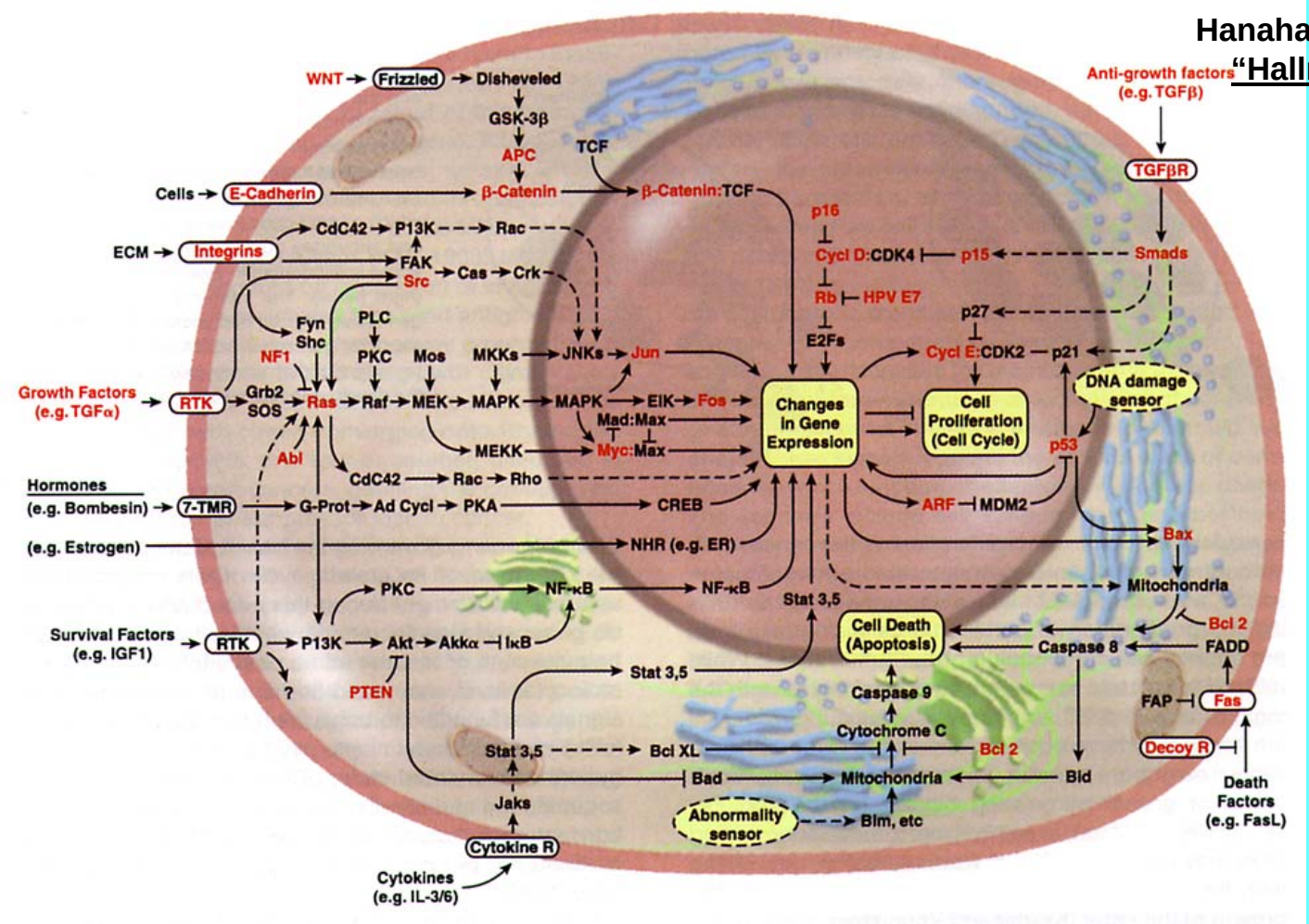
Modeling & Analysis

Growth Factor Pathways



How Can We Understand Signaling Networks?

Hanahan & Weinberg, 2000
"Hallmarks of Cancer"

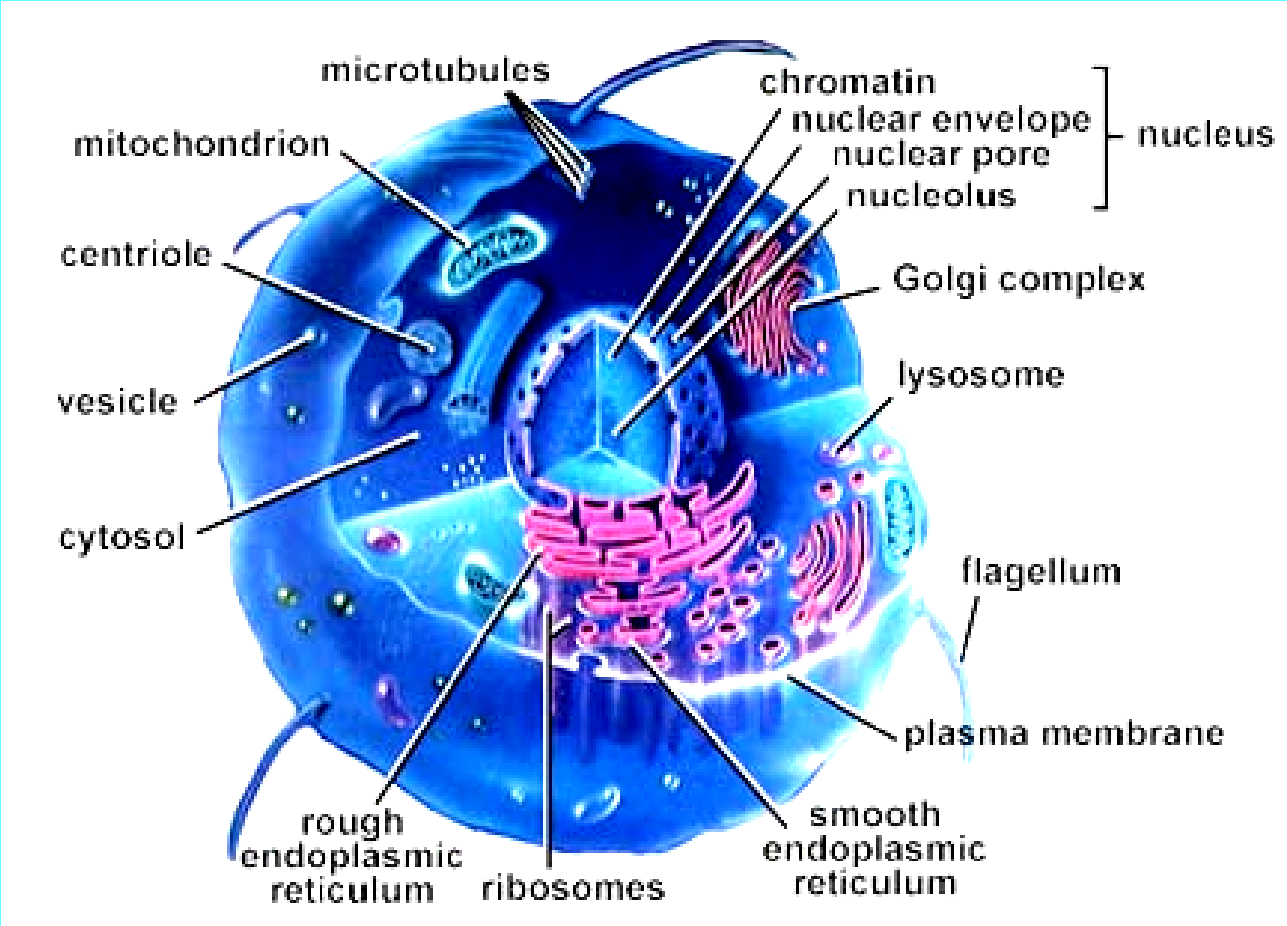


Battelle

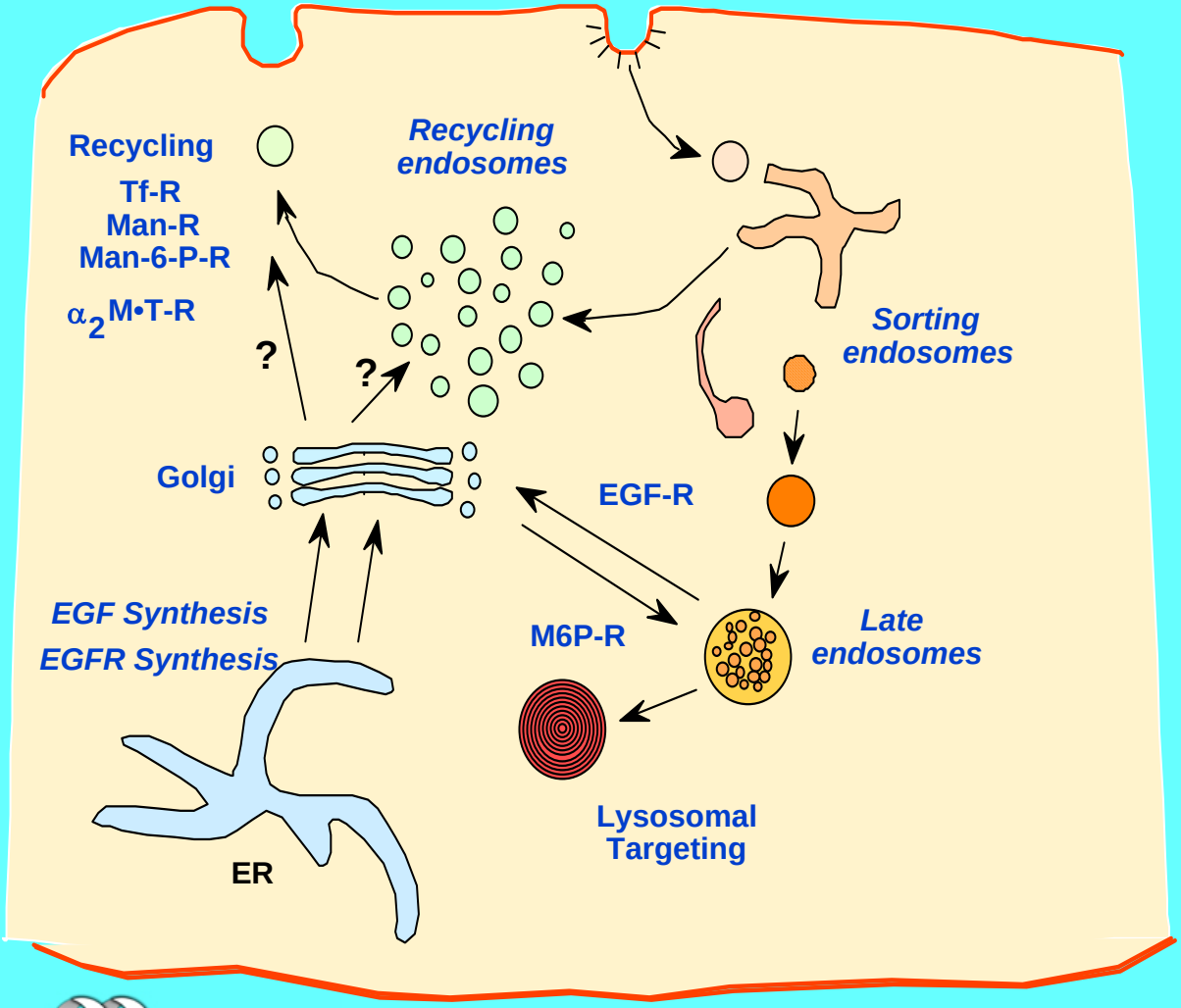
Bio Molecular Networks

U.S. Department of Energy
Pacific Northwest National Laboratory

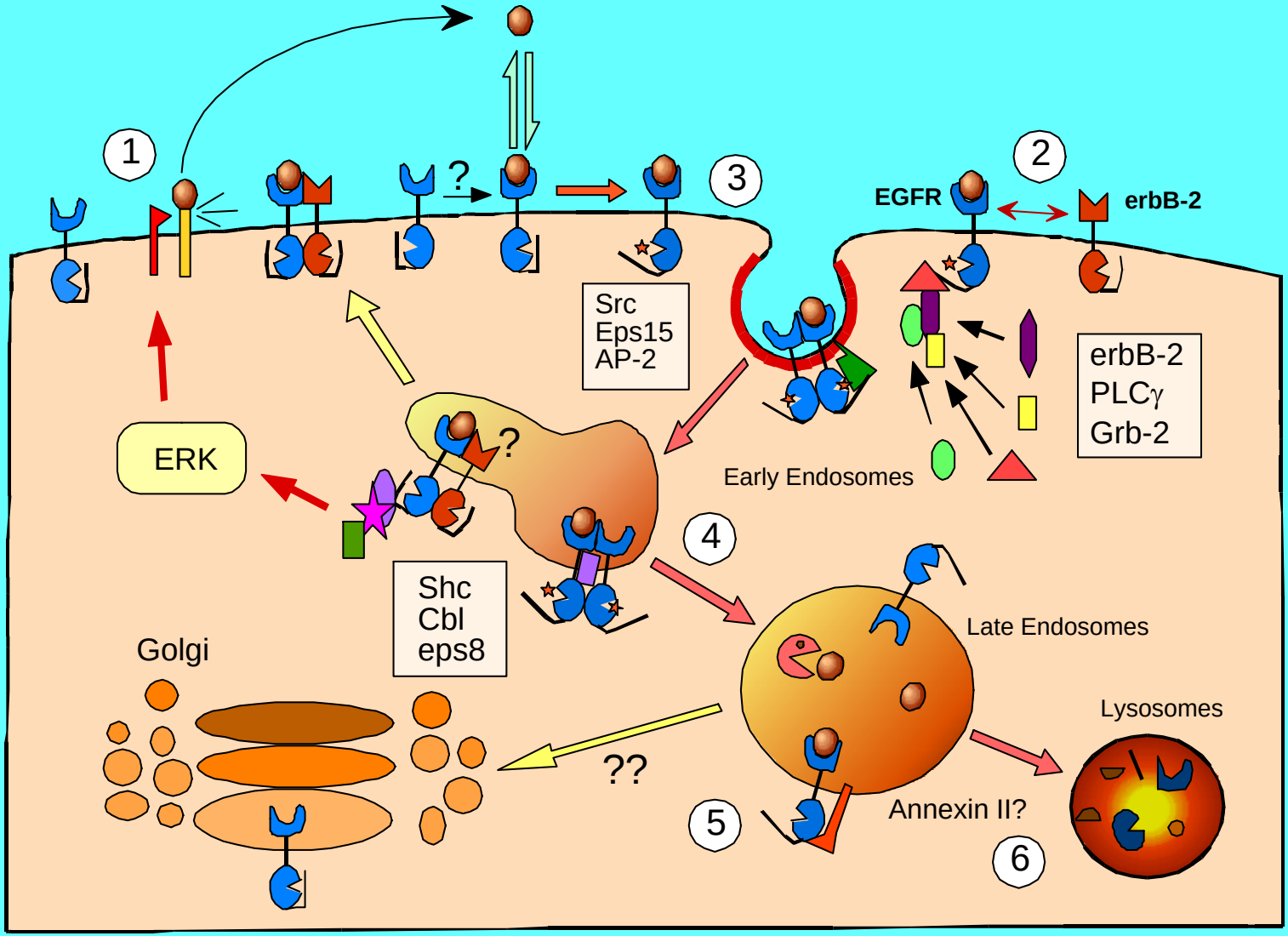
Structure of Cells



Intracellular Trafficking Pathways



The EGFR System



Battelle

Bio Molecular Networks

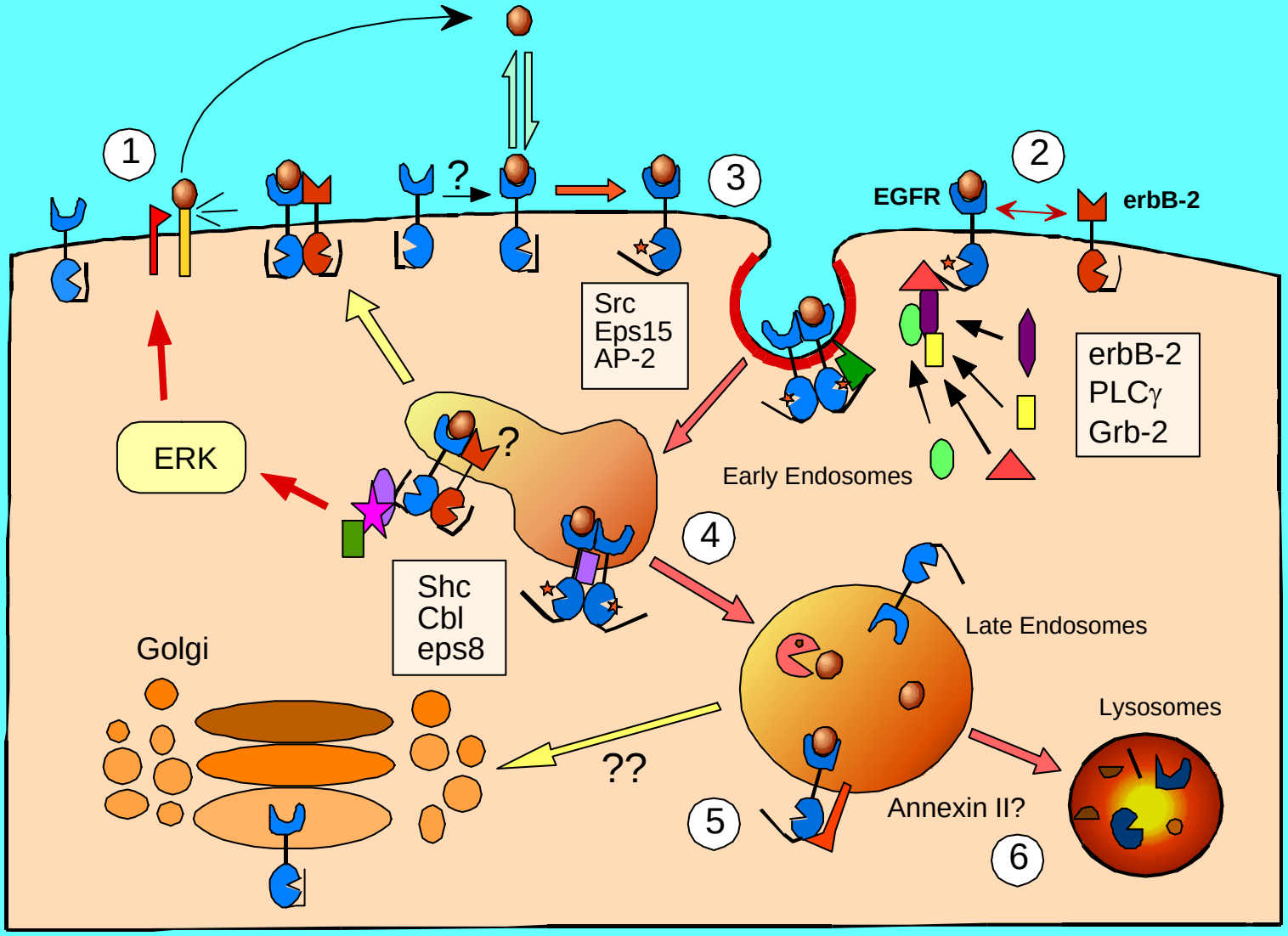
U.S. Department of Energy
Pacific Northwest National Laboratory

How Do We go From Cartoon Model to Realistic Model?

- Model must be quantifiable
- Model must make very specific predictions
- Model must account for changes in components as well as input parameters
- Must be testable with experimental data

Models and experiments need to be developed together

The EGFR System



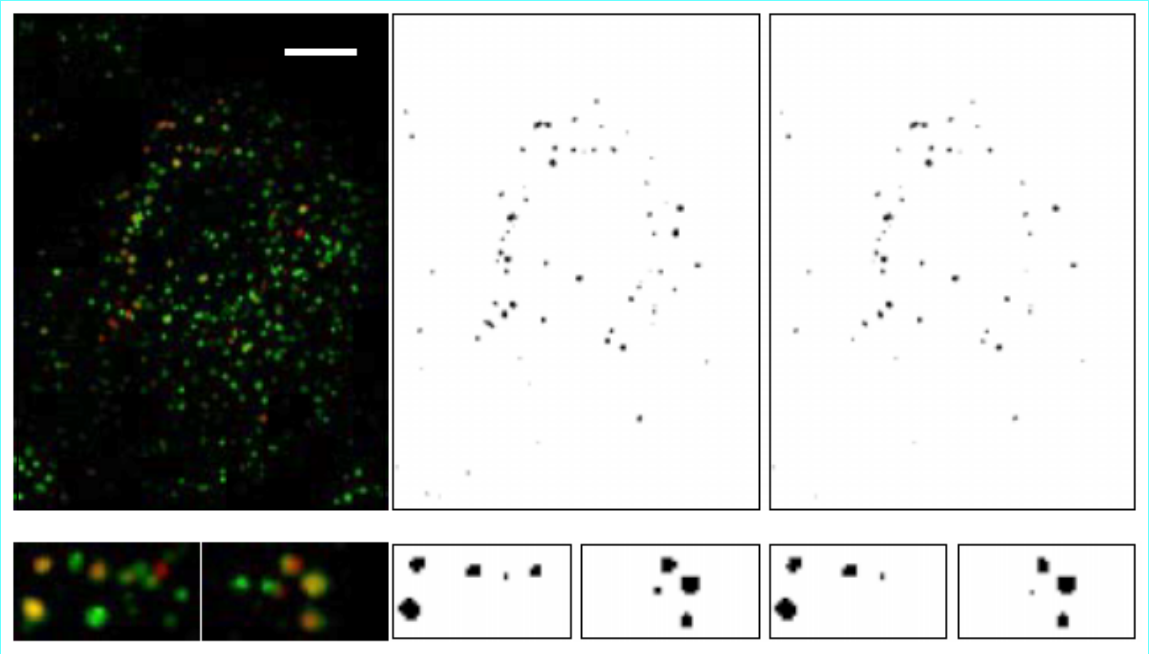
Battelle

Bio Molecular Networks

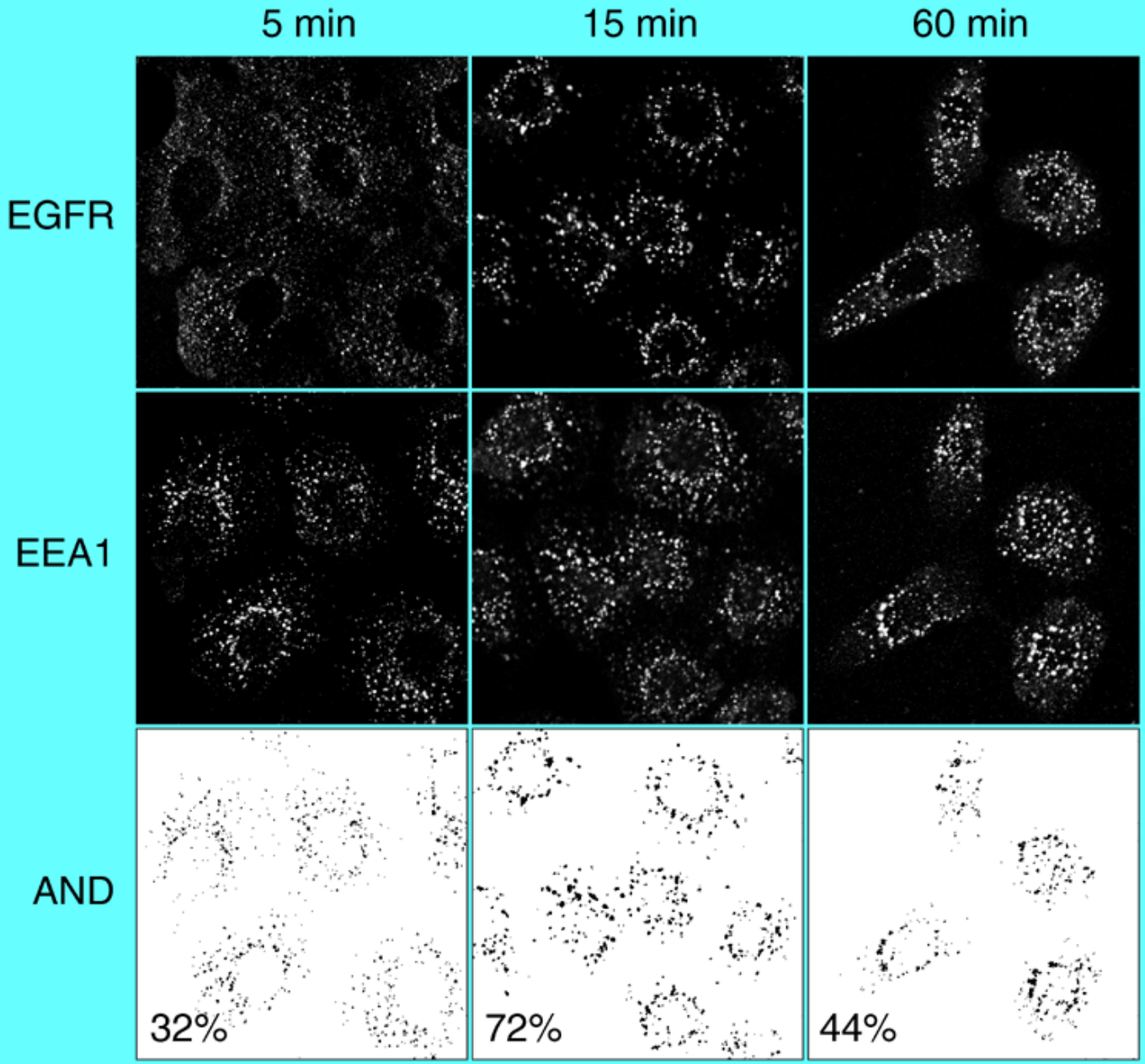
U.S. Department of Energy
Pacific Northwest National Laboratory

Spatial Colocalization by Image Analysis

For two molecules to interact, they must be located in the same space at the same time. This can be determined by image analysis



Movement of EGFR through early endosomes as determined by image analysis



Battelle

Bio Molecular Networks

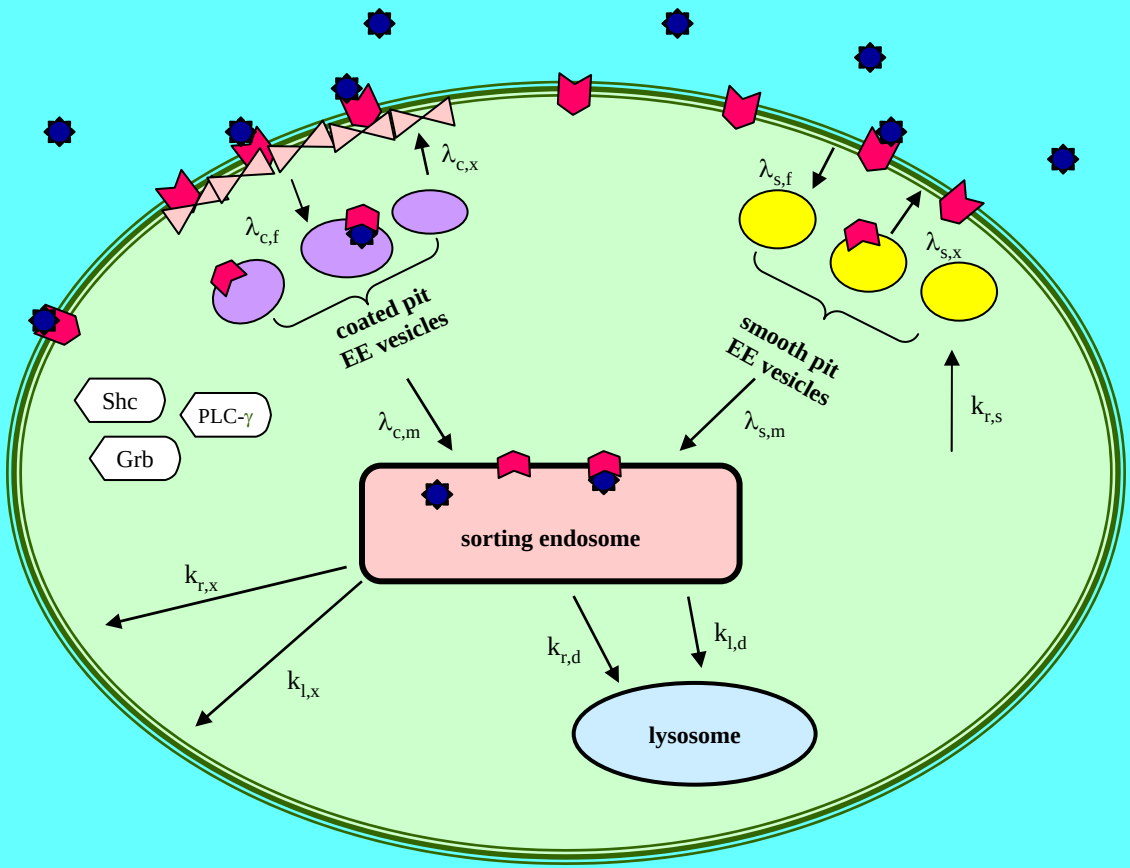
U.S. Department of Energy
Pacific Northwest National Laboratory

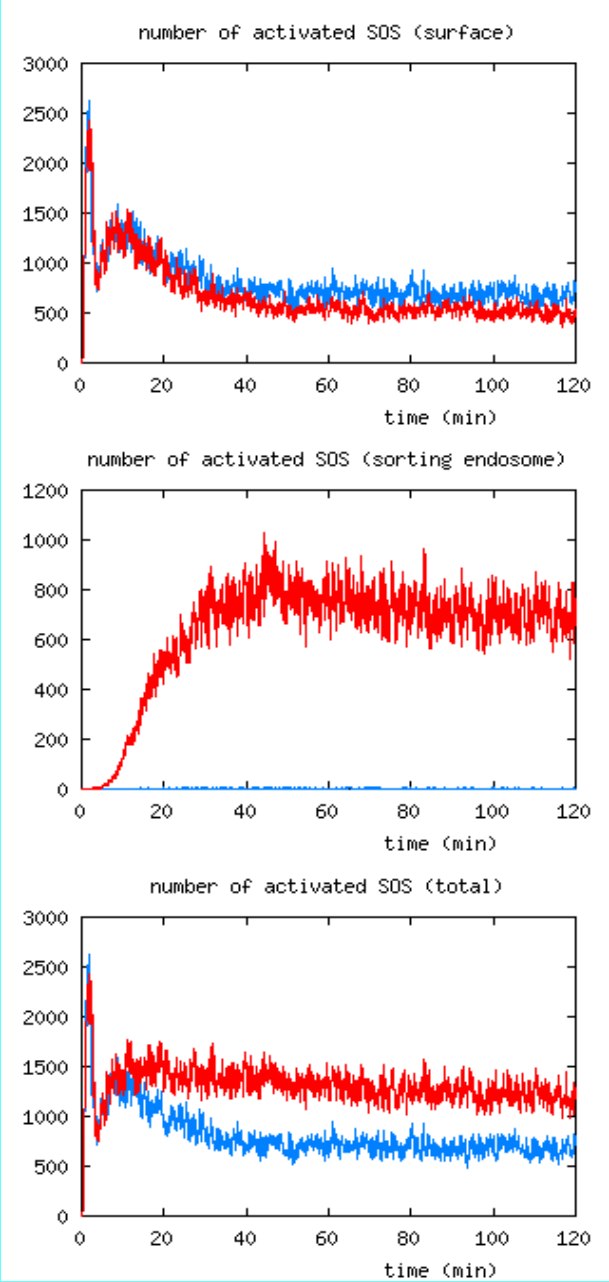
Colocalization of EGFR and HER2

QuickTime™ and a
DV - NTSC decompressor
are needed to see this picture.

QuickTime™ and a
Cinepak decompressor
are needed to see this picture.

Trafficking model for EGFR signaling





20 nM of ligand is injected in 15 seconds
(red: EGF, blue: TGF-α)

Stochastic/Kinetic Simulations → Dynamic Monte Carlo method
(Probability weighted)
A probabilistic approach as opposed to deterministic
Extension of the Monte Carlo method to the time domain

H Resat, HS Wiley, DA Dixon

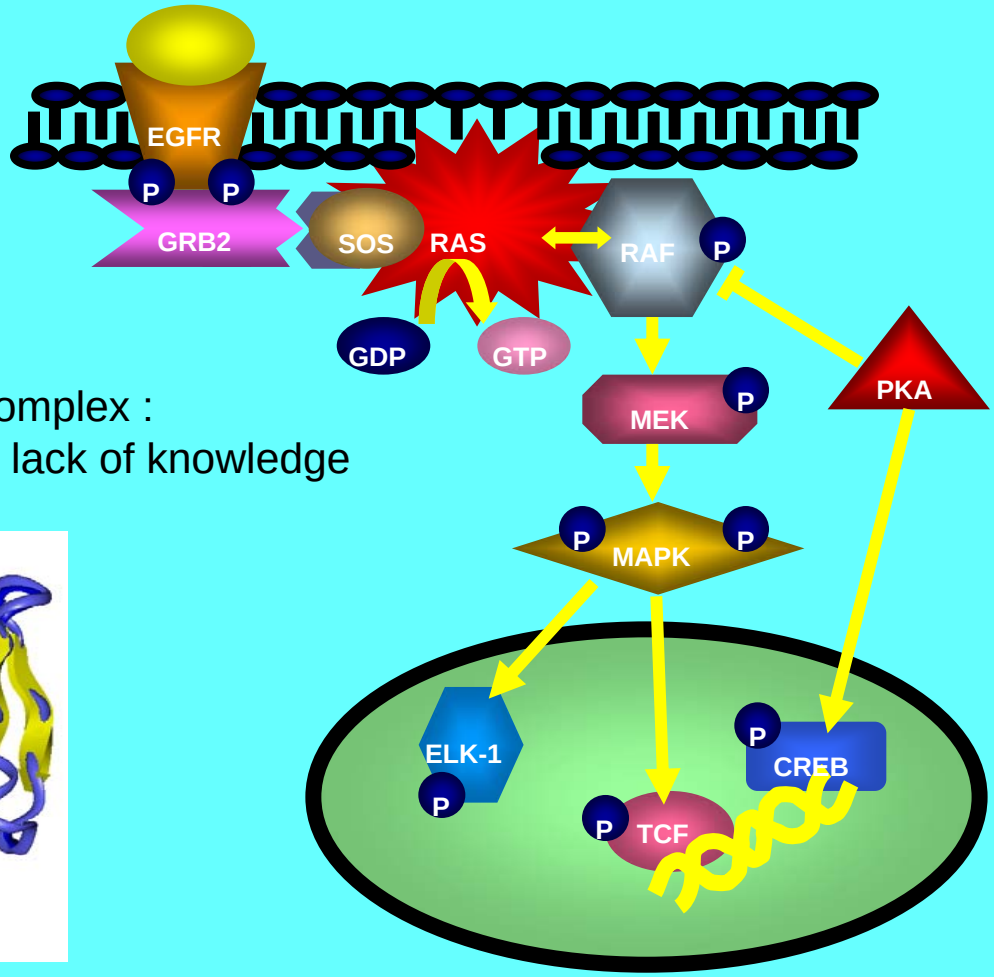
Battelle

Bio Molecular Networks

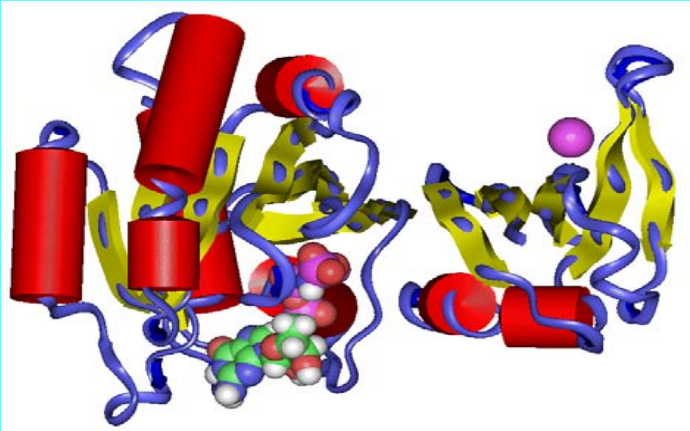
U.S. Department of Energy
Pacific Northwest National Laboratory

Signal Pathways: Computer Simulation of Protein-Protein Complexes

- Ras acts as the GTP-activated molecular switch in the cell signaling pathway for gene expression, controlling cell proliferation and differentiation. Mutations found at Gln-61 in 30% of human tumors.



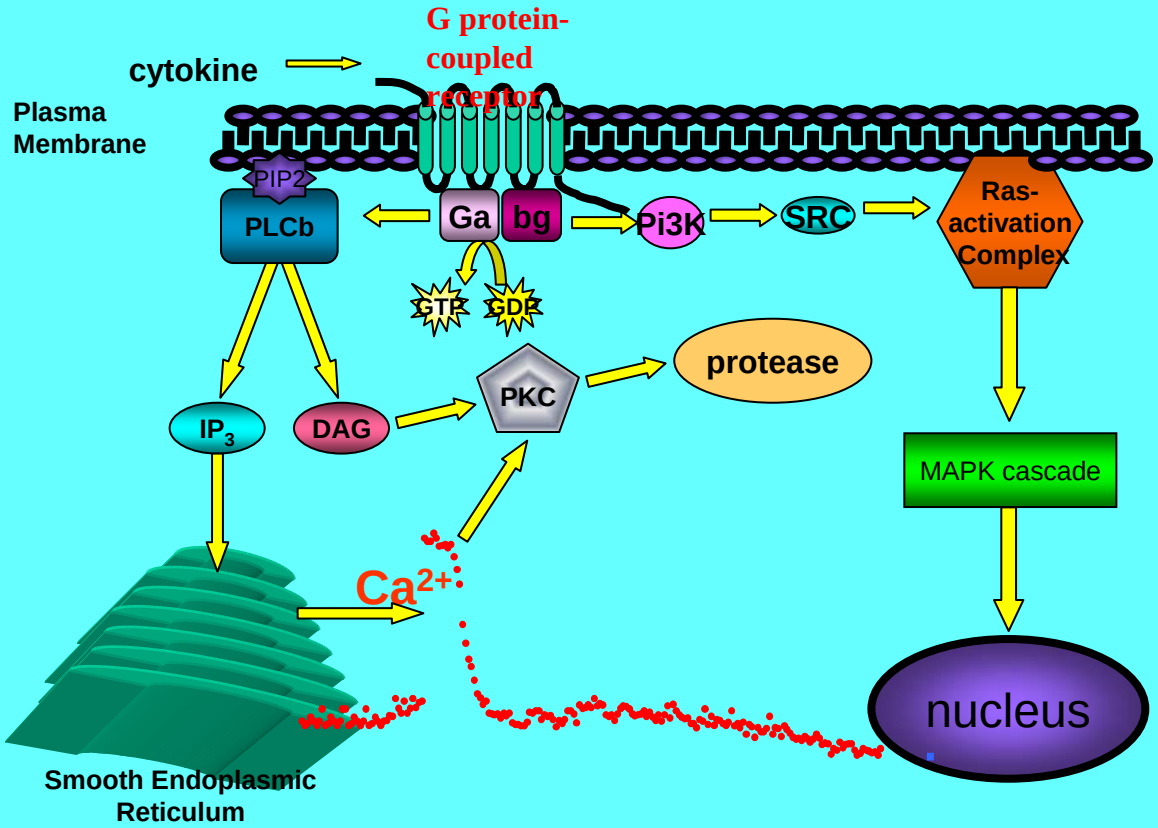
- Issues for modeling the Ras:Raf signaling complex : 50,000 atoms, long-time simulations, solvent, lack of knowledge of the entire structure

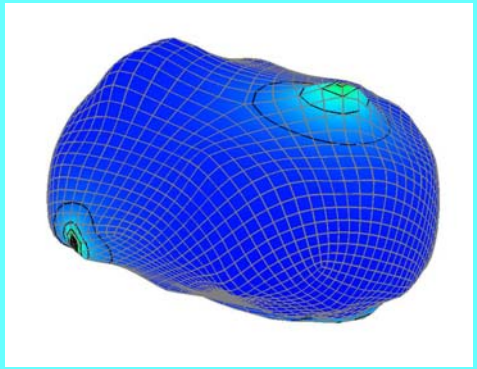


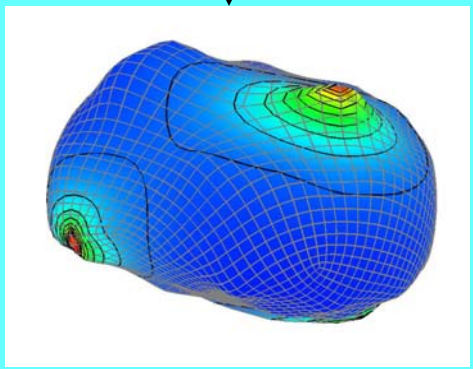
TP Straatsma, H Resat, JH Miller, T Soares, DA Dixon, PNAS, 2001

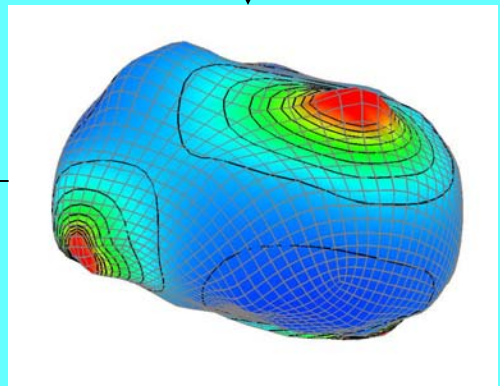
Model for the Role of Ca^{2+} in Stress-Response Pathways: NWGrid/NWPhys

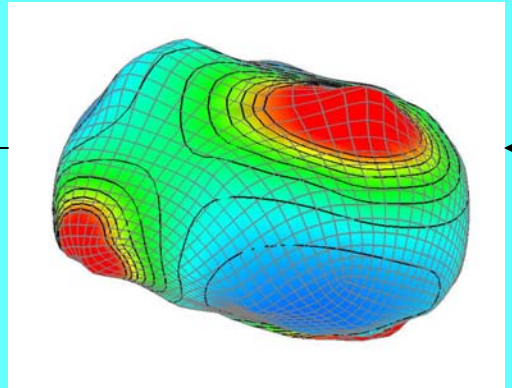
H Trease, JH Miller

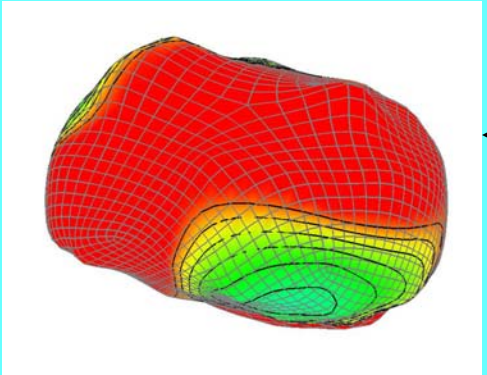




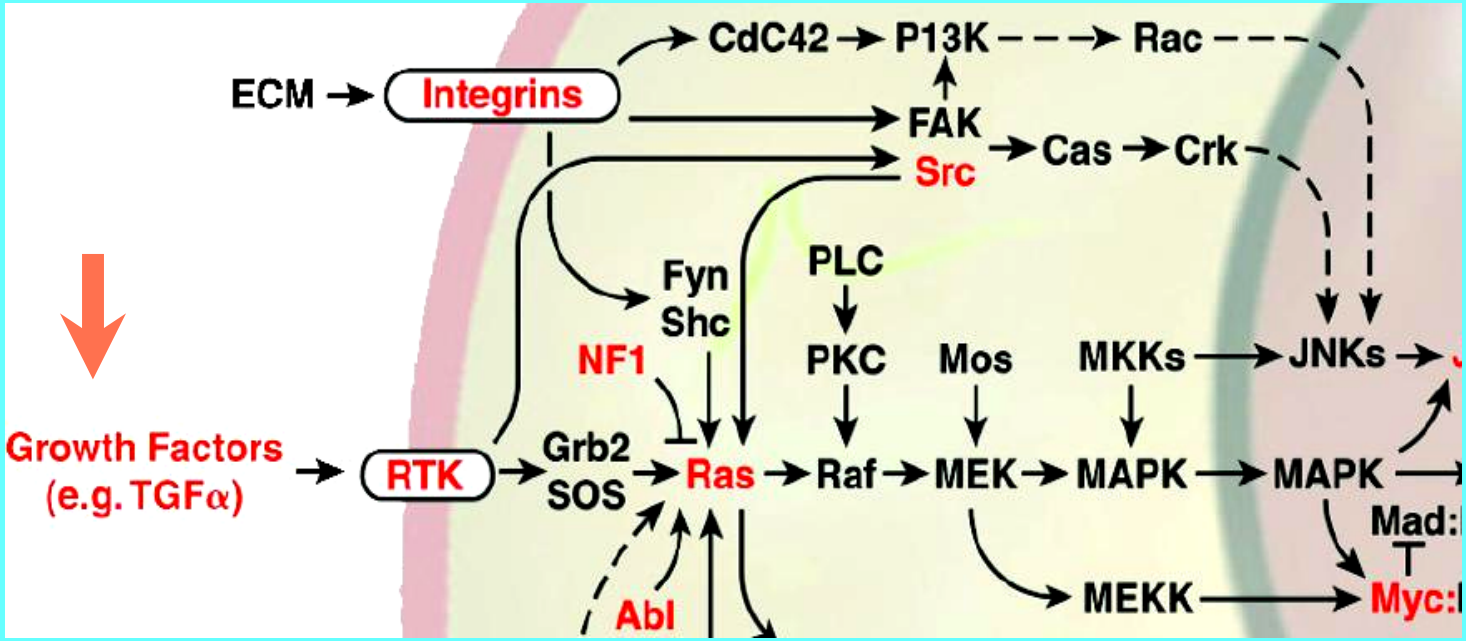






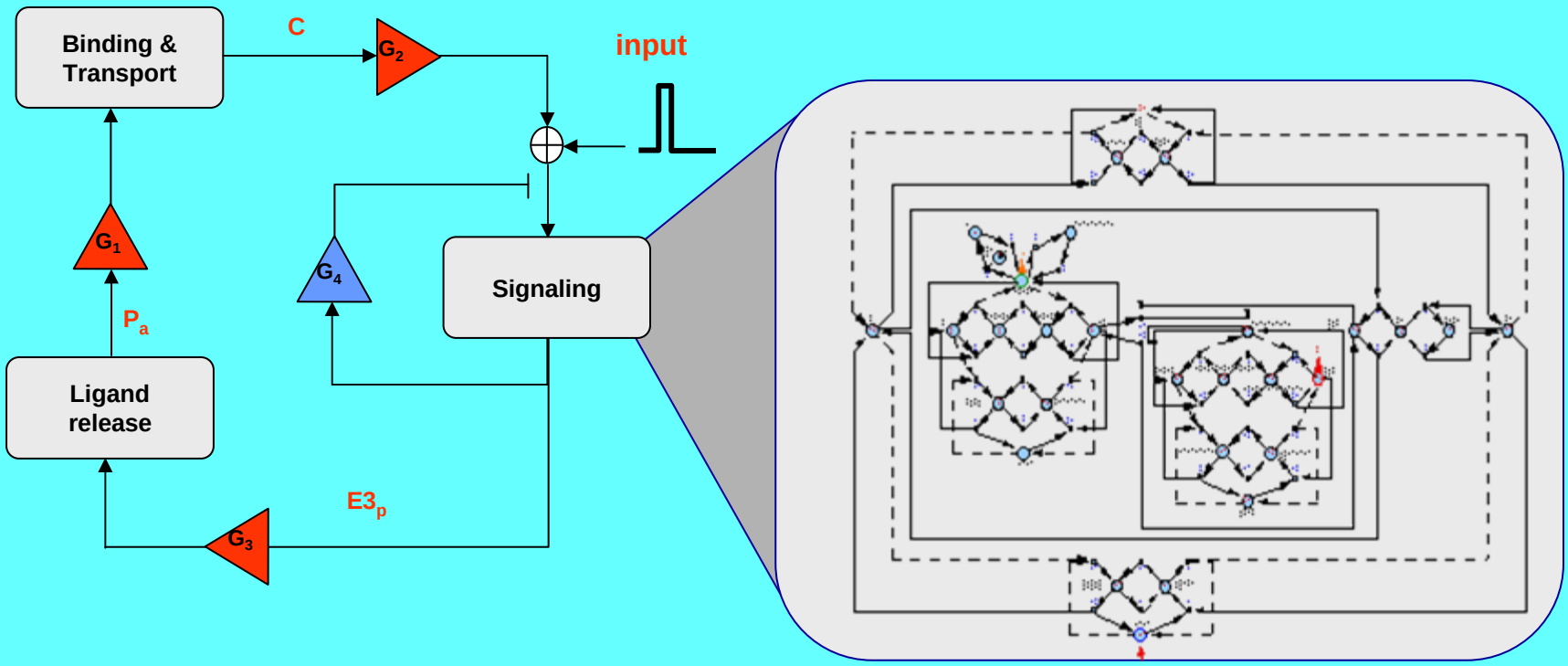


Signaling Through Growth Factor Receptors



Ligand (Extracellular molecule)	→	Membrane **receptor**	→	Intracellular signaling cascade	→	Physiological response
INPUT	→	SENSOR	→	ACTUATOR	→	OUTPUT

Central Role of Computation and Simulation



Computer simulations provide a conceptual framework for organizing system behavior

Cell State Dictates Cell Response

- DNA is information storage of cell
- Proteins are the “program” running the cell
- The running program dictates cellular response

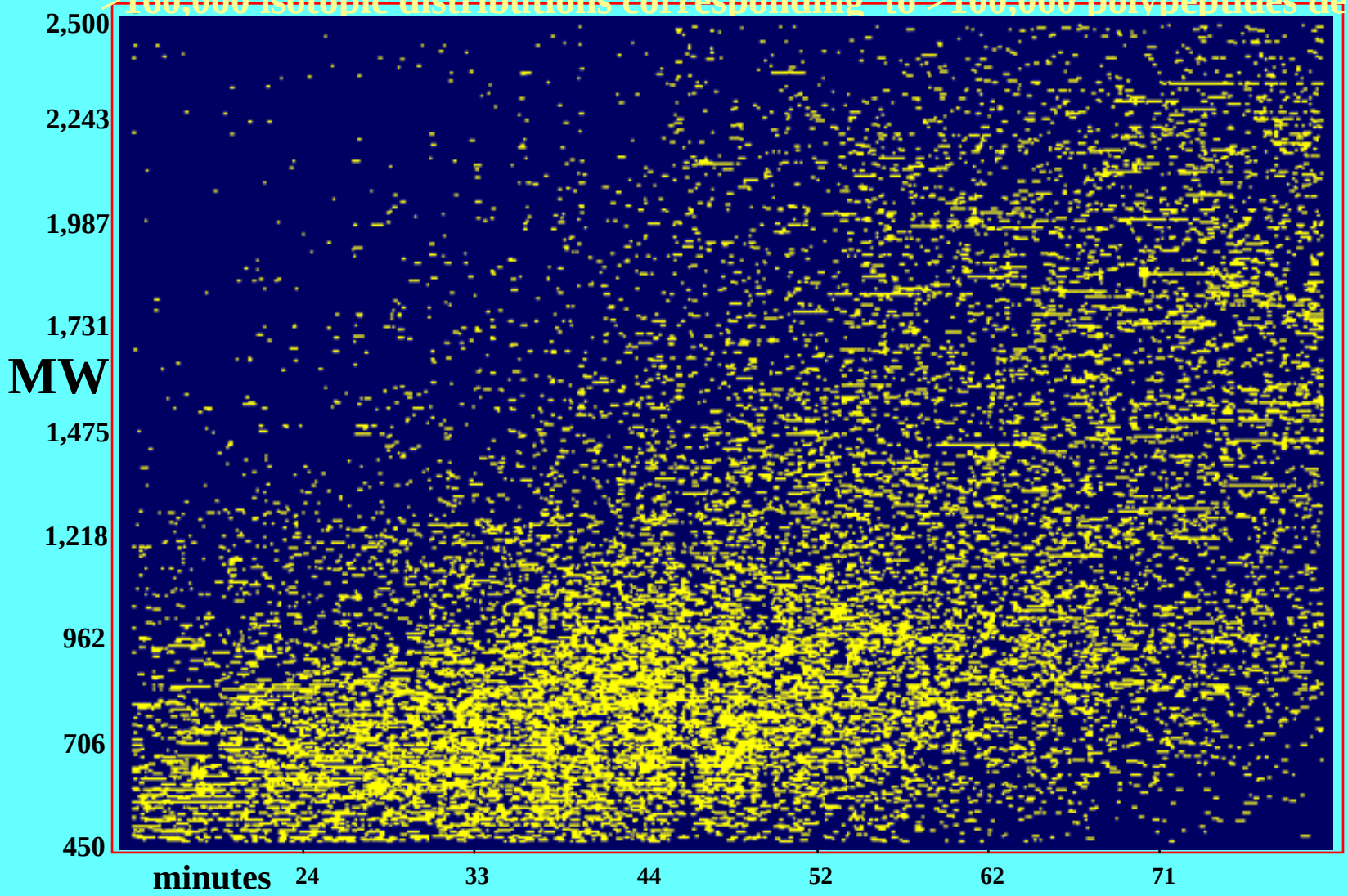
Thus, one cannot adequately model or predict cellular responses UNLESS the proteome of a given cell is first specified!

Proteomics

- What parts are being made? (identity)
  - Develop technologies for global proteomics
  - Develop technologies for selective proteomics
- What is their state of modification? (activity)
  - Phosphorylation state
  - Other covalent modifications
- What are their levels? (quantity)
  - High-throughput quantification
  - Rapid quantification

Capillary LC-FTICR 2-D display of peptides from a yeast soluble protein digest

>160,000 isotopic distributions corresponding to >100,000 polypeptides detected

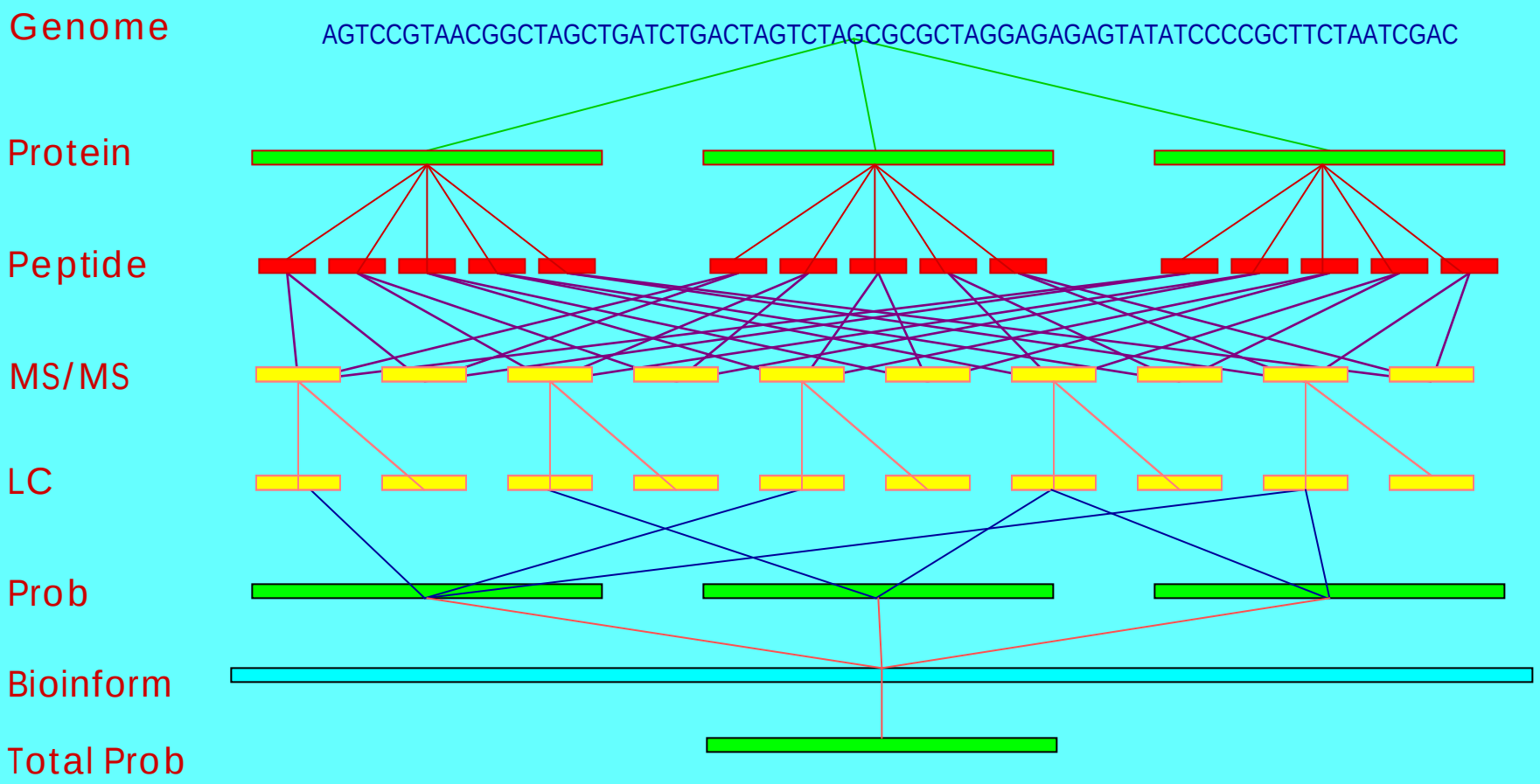


Battelle

Bio Molecular Networks

U.S. Department of Energy
Pacific Northwest National Laboratory

Probability Network for Proteomics

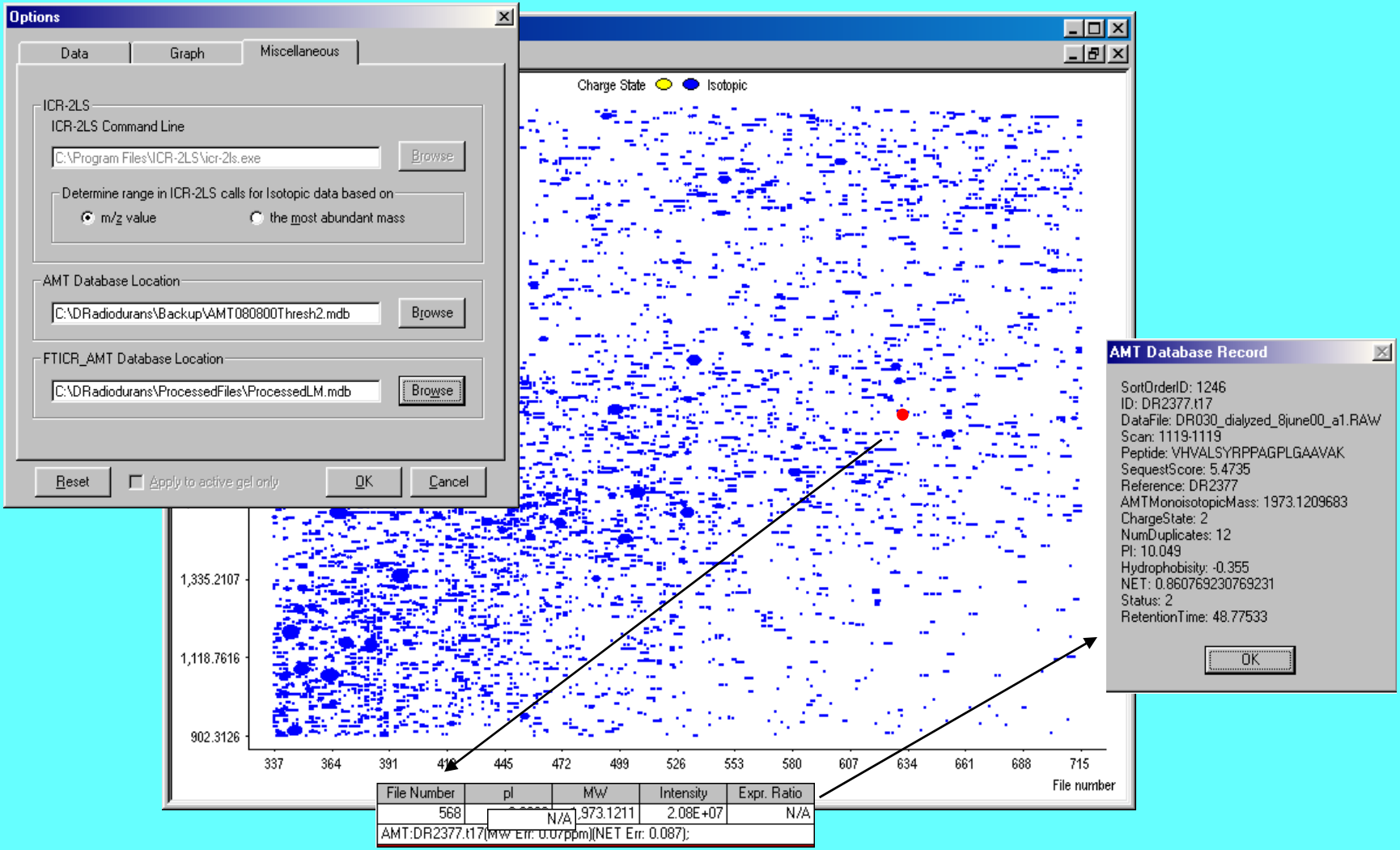


Battelle

BioMolecular Networks

U.S. Department of Energy
Pacific Northwest National Laboratory

Software provides connection between visual displays and data

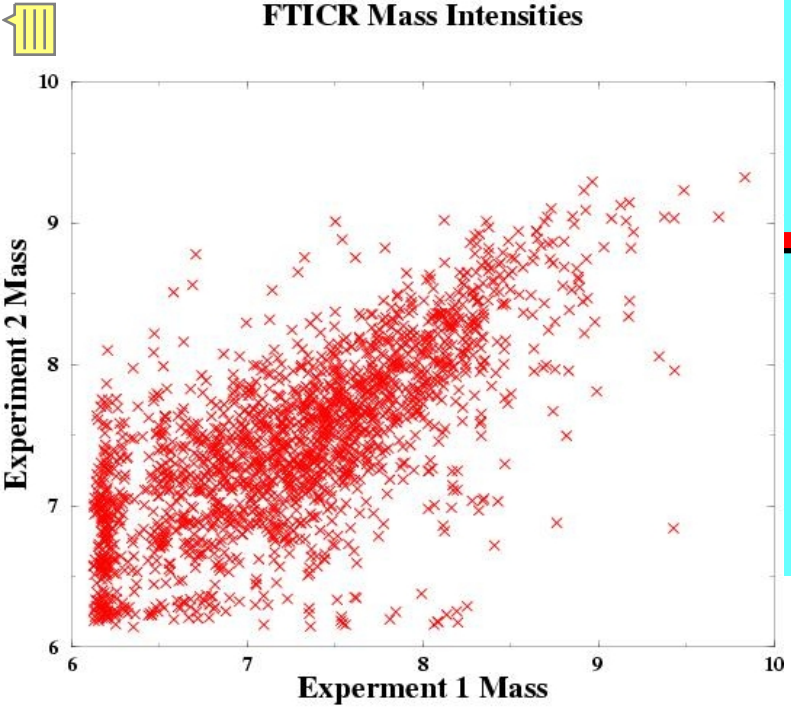


Battelle

Bio Molecular Networks

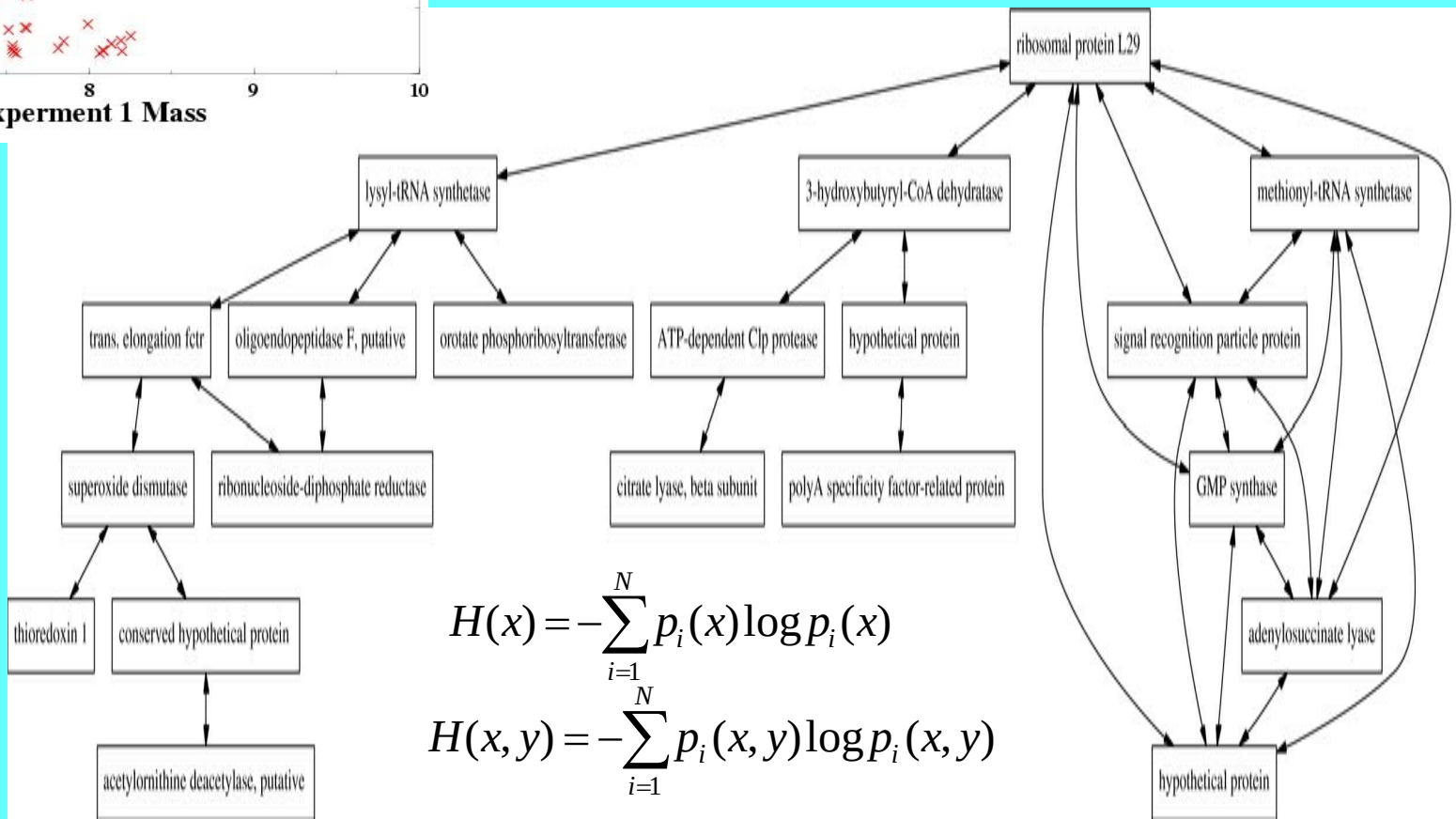
U.S. Department of Energy
Pacific Northwest National Laboratory

Expression analysis from MS Proteomic Data



Expression levels between cell populations as measured by FTICR MS are correlated and quantitative…

…allowing for expression and network analysis.



$$H(x) = -\sum_{i=1}^{N} p_i(x) \log p_i(x)$$

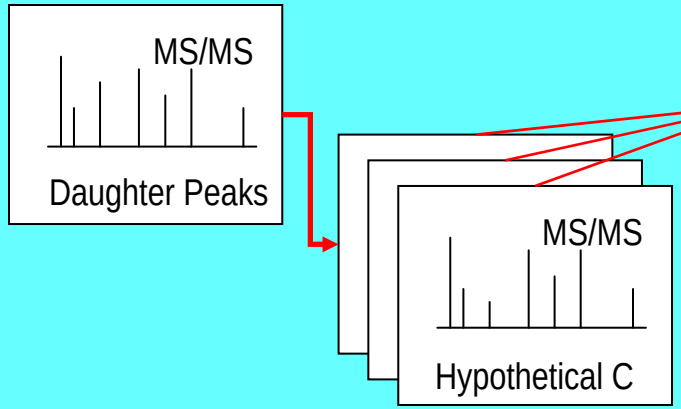
$$H(x,y) = -\sum_{i=1}^{N} p_i(x,y) \log p_i(x,y)$$

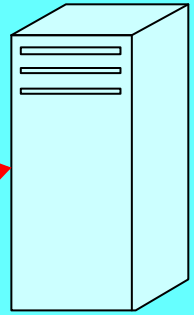
WR Cannon

Database Requirements for Proteomics

1. Build Knowledge Base

Pooled Proteins → Pooled Peptides → $>10^3$



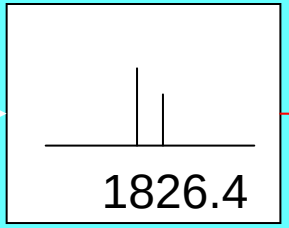


Archive data

2. High-throughput Experiments

Pooled Proteins → Pooled Peptides →

FTICR-MS



Proteomic State

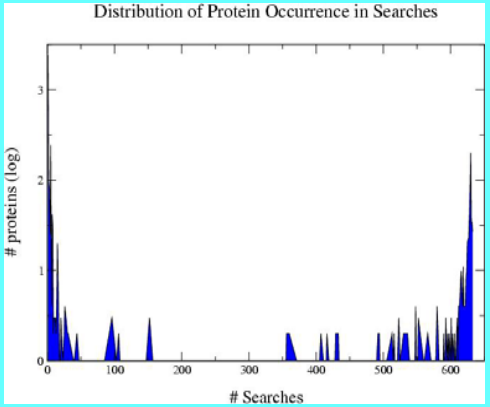
Automated data mining of genome superfamily sequences – Data integration and method validation

Automatic Superfamily Finder

Random hits

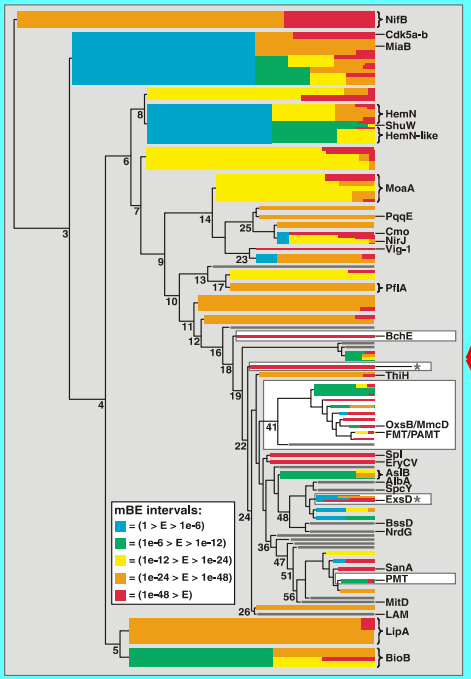
Fragments

Members



Extracts and validates superfamily sequences based on PSI-BLAST (automation tool under development)

Superfamily organizer



Organizes and visualizes superfamily sequences (Prototype built)

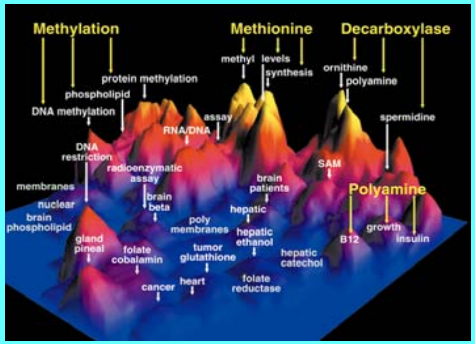
Pattern Integrator

Motif CxxxCx[F,Y,W]C

Domain N C

Operon

Text



Supports data mining

Battelle

Bio Molecular Networks

U.S. Department of Energy
Pacific Northwest National Laboratory

File Edit View Favorites Tools Help

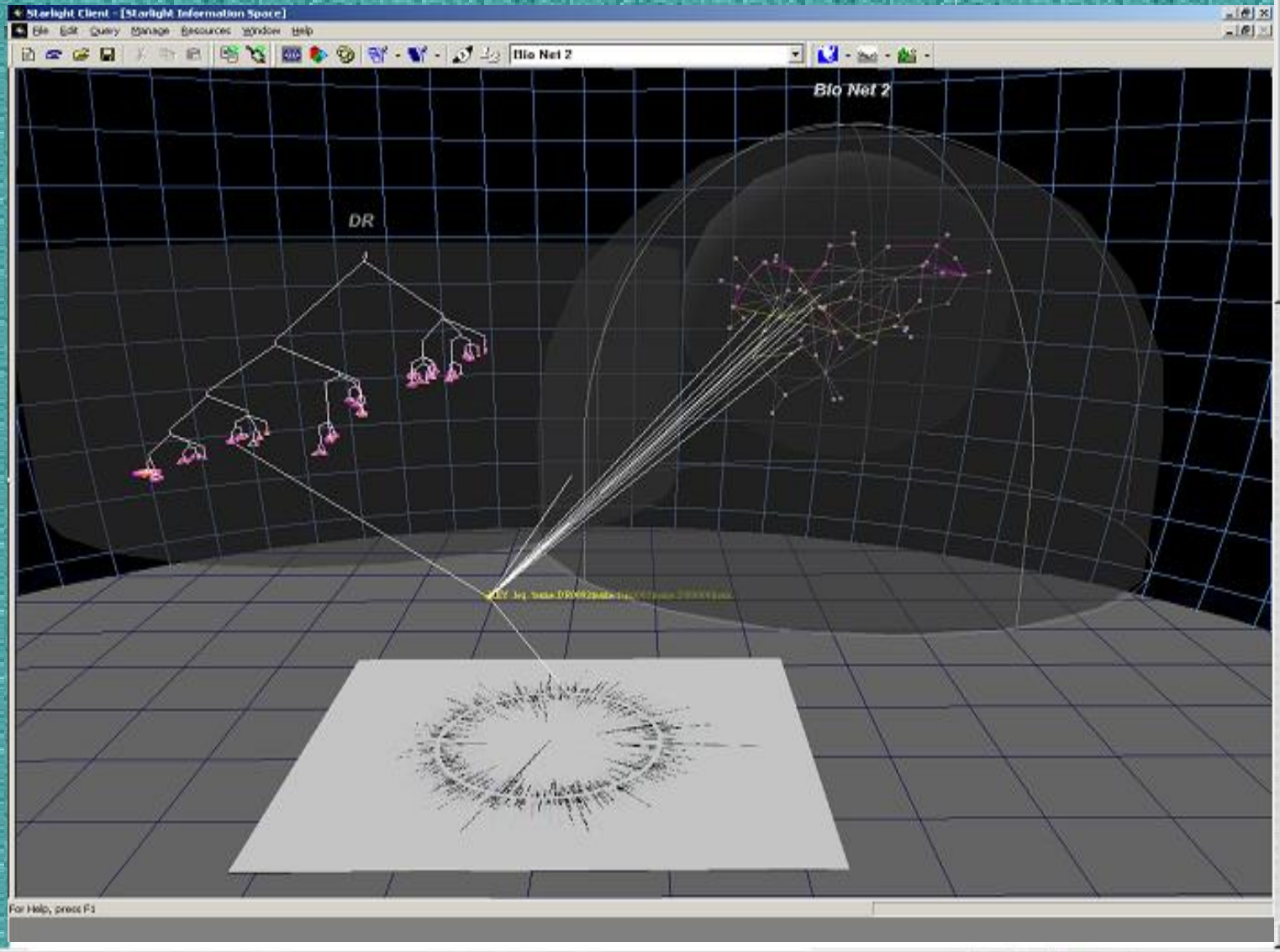
Back Search Favorites History

Address http://www.biomolecular.org Go Links

Bio Molecular Networks Home

Computational Cell Environment

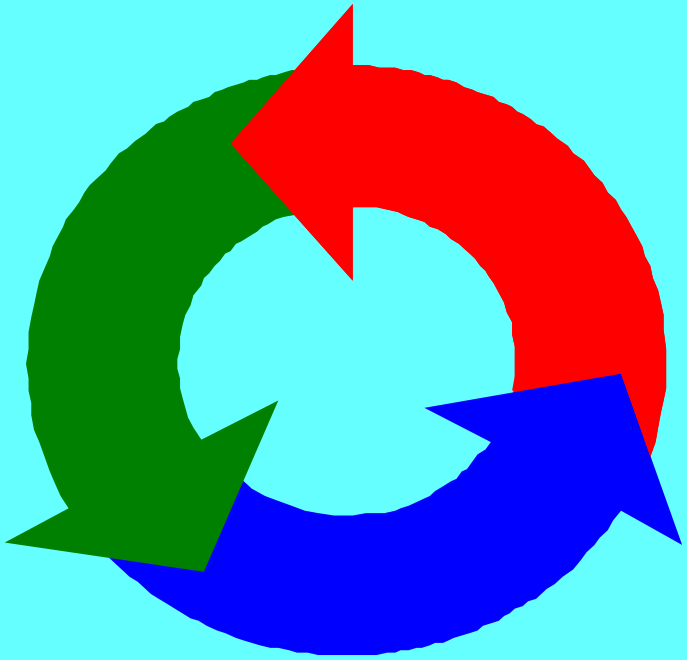
- Analysis Tools
  - Starlight
  - Similarity Box
  - Psort
  - Prosite
- Announcements
- Collaborators
- Data Sets
  - Experimental
    - FTICR
    - CARS
  - Computational
- Reference Links



Internet

New Model for Biological Analysis

Information Synthesis
Mathematical,
Computational,
Visualization



Analytical
Experimental,
Robotic
Computation

Hypothesis
Computation

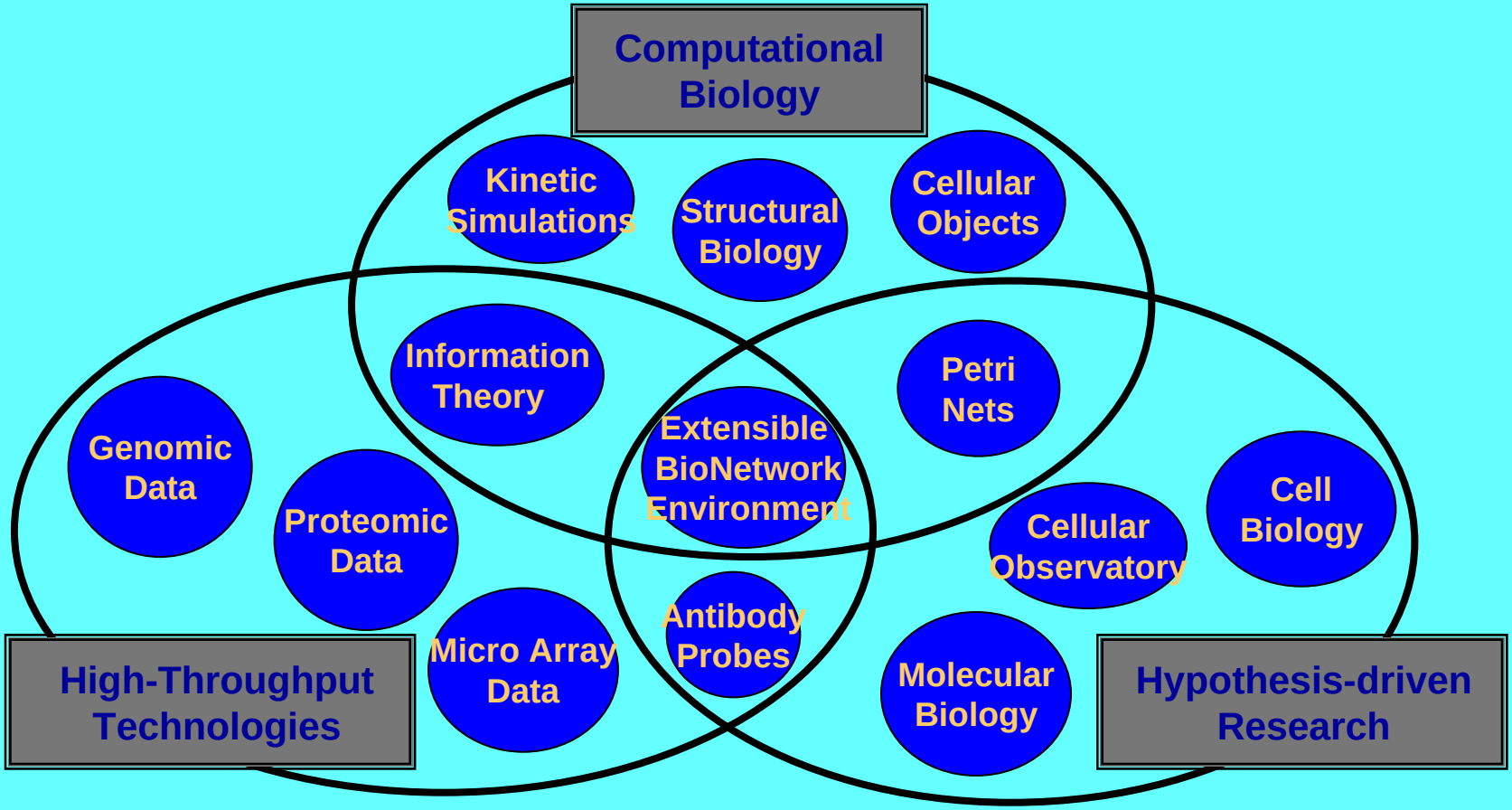
1. Enhance understanding of known systems

2. Provide qualitative/ quantitative insights → direct new experiments or design new systems

3. Provide quantitative results to:
- Replace experiments that are too difficult, dangerous or expensive
- Go into new temporal and spatial domains
- Extend limited experimental data into new domains of parameter space

Battelle

BioMolecular Networks

U.S. Department of Energy
Pacific Northwest National Laboratory

Computational Sciences Ties Systems Biology Together