An Assessment of Computational Systems Biology

From a computing perspective

Steve Burbeck
IBM Life Sciences
April, 2003

Famous Last Words

"Every attempt to employ mathematical methods to study of biological questions must be considered profoundly irrational and contrary to the spirit of biology. If mathematical analysis should ever hold a prominent place in biology - an aberration which is happily almost impossible - it would occasion a rapid and widespread degeneration of that science."

Auguste Comte (1830)

Founder of Positivism

The Growth of Computational Biology

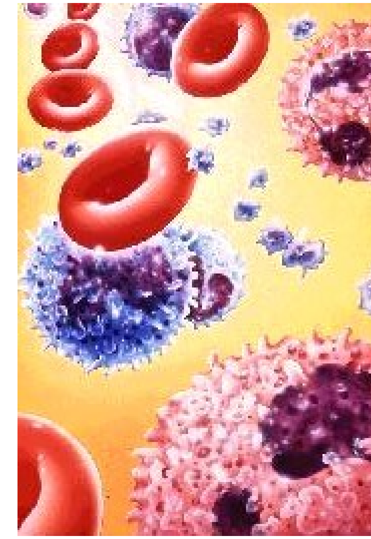
- By the end of this decade, biology may be the single largest consumer of scientific computing resources.
- We are already starting to see that phenomenon at academic institutions

Systems Biology: Opportunity and Caution

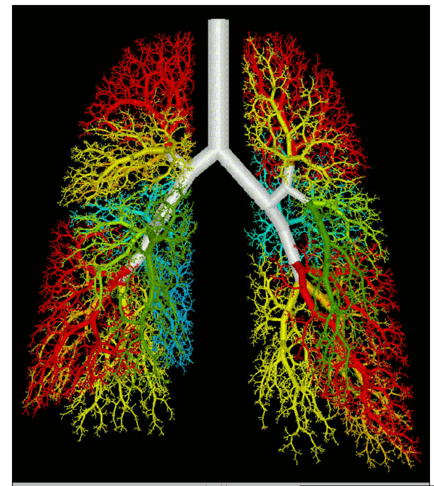
- Opportunity: the high-throughput biology gold rush
  - analyzing the flood of new parts
  - understanding how all these parts fit and function together in living systems promises great rewards
- Caution: all that glitters is not gold
  - Interactions between parts are complex and produce surprising emergent behavior
  - Simulations can help us understand these emergent behaviors
  - Or generate glittering simulation equivalents of fools gold

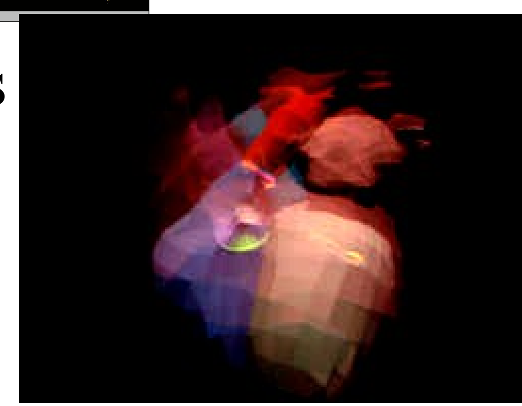
Modeling/Simulation applicable at:

Organism Level
Organ Level
Cellular Level
Molecular Level



Complex Fluid





Heart/ Lungs

Circulatory System and Angiogenesis

The Major Themes in Systems Biology

- Automated analysis and annotation
  - Model-based prediction of ORFs, protein function, protein structure prediction, sequence homology searching, etc.
  - Results in “derived” data, perhaps more voluminous than the raw data
- Modeling
  - Build descriptive models that summarize and organize data, e.g., pathway models
- Simulation
  - Build dynamic computational models to investigate behavior of systems
  - Results in executable simulations that predict and explain behavior
- Iterative collaboration between computational and experimental biology
  - Iterate, iterate, iterate
  - Without this step, simulation is little more than a game not unlike SimCity
  - The problem usually is lack of key biological data

Hierarchy of Systems

Level of Abstraction	Entities	Computational Tools and Techniques	Unsolved Issues
Parts	Genes, Proteins, DNA binding sites, splice sites, membrane targeting signals, etc.	Sequence matching Gene prediction Protein structure prediction Data mining	Consistent naming Referential integrity Curation and accuracy Ontologies
Networks and Pathways	Networks based on pair-wise relationships between parts	Path tracing in networks, shortest path finding Cluster analysis Descriptive models and simulations Entity-Relationship diagram tools (e.g., Gene Ontology tools) Visualization	Separating meaningful relationships from artifacts Temporal relationships Understanding effects of perturbations
Assemblies and Complexes	3-D structures of parts, e.g., proteins and RNAs	Modeling Simulation Visualization	Description and modeling of location in the cell How structure determines function Relationship to pathways Interactions with other assemblies
Systems of Systems	Structures of substructures, e.g., organelles, cytoskeleton, cells, biofilms, tissues, organs	Modeling Simulation Visualization	Prediction and explanation across meta-levels Building multi-scale models, both physical and temporal 3-D structural connections Scientific validity

Common Issues

- Complex systems with emergent properties
  - e.g., a cell may have 10^{14} molecules and 10^{6} circuit nodes
  - parts interact in complex nonlinear ways
  - the whole is far more than the sum of its parts
- Multi-scale in space and/or time
- Modeling and simulation necessary to understand the systems

The Challenge of Emergence

- What are the “right” levels of abstraction?
- How does “function” or cause-and-effect cross levels, and on what time scales?
- Are physical perturbations (or simulation inaccuracies) suppressed or magnified as they cross levels?
- If a level is omitted in a model, what are the consequences to accurate simulation and prediction?
- These issues are general Complexity Theory issues common to the study of many kinds of emergent systems

The Limits of Occam's Razor

- Biology is a historical science that reflects the frozen consequences of 3.5 billion years of evolutionary accidents
  - A programmer might think of them as layer after layer of "clever hacks"
  - Each layer exploits the hacks that have come before
- In biology, all bugs are features
  - -1 programmed frameshift (the "off by one" bug)
  - Alternate splicing - many proteins from one gene
    - The GnRH receptor gene, for example, encodes a splice variant that acts as a repressor of the receptor itself
  - "If you can think of it (i.e., some surprising quirky cellular mechanism), you will find that somewhere the biological machinery does it." - Terry Gaasterland, Rockefeller Univ.
- Simple models are almost certainly wrong, perhaps expensively so

The Importance of 3-D Models

- *Modular* biology replaces molecular biology
  - "Nearly every important function in the cell is carried out by an assembly of 10 or more proteins" - Lee Hartwell
- Modeling 3-D structures at multiple scales
  - cytoskeleton, organelles, membrane compartments are at least two orders of magnitude bigger than smaller constituent parts
  - interactions between levels forces models to address multiple levels
- Self Assembly
  - Life-cycle issues - when does an assembly form and when and how is it destroyed?
  - E. coli flagella disappear in nutrient rich environments and reappear when needed
  - Models must eventually account for these issues, perhaps sooner than we think

The Importance of Stochastic Processes

- Stochastic processes are are often dismissed as just a nuisance…beware!
- Many processes require stochastic treatment
  - Small numbers of molecules can have large effects
    - e.g., single gene transcription
  - Randomness is inherent, and *necessary*, in some processes
    - stochastic switches, e.g., sporulation switch models (Denise Wolf at Berkeley)
    - probabilistic alternate splicing which produces necessary splice variants

A Marriage of Two Cultures

- Biological modeling and simulation requires expertise in both biology and computing
  - Pioneers in the field tend to be savvy about both
  - But such people are rare
  - So, biologists often must team with computational experts
- All too often, computational folks are from Mars and biologists are from Venus

The Computing Culture

- Assumes that complex systems reflect a functional design that can be reverse engineered
- Assumes clever algorithms and or data mining provide the crucial value
- Tends to underestimate testability issues and the difficulty of wet-lab verification
- tends to prefer to use modeling techniques based on familiarity, efficiency, or elegance rather than suitability for the domain
- tends to overconfidence about how to build, fit, and test the models and simulations
- Assumes that one large model is better than many smaller more focused models

The Biology Culture

- Assumes that complex systems result from messy evolution rather than design
- Assumes that biological insights provide the crucial value
- Tends not to understand the far-reaching implications of choosing a particular simulation approach
- Tends to discount the difficulty of making accurate computing models and simulations
  - All software development takes far longer than you'd expect
  - keeping track of the various versions, ever present bugs, and design revisions is harder than it looks
  - bugs that happen to give you the results you want aren't hunted down with the same fervor as ones that don't

Simulation Migrant Workers

- Simulation skills have been honed in many fields
  - Weather - modeling weather and climate
  - The Boeing 777 - based on a complete model
  - Financial models on Wall Street - simulating stochastic market movements for fun and profit
  - Astrophysics - modeling galactic formation
  - Chemical engineering - simulating bulk chemistry
  - Oil reservoirs - simulating oil flow rates
- Immigrants from all these fields are moving to systems biology
  - ***When all you have is a hammer, everything looks like a nail***
  - Living systems are so complex that every immigrant can find their favorite equivalent of a nail
  - The challenge is to pick the right model for the job and to weave many kinds of models together rather than focus on just one or two

Implications for Venture Capital

- Some startups are primarily about the biology
- Some are primarily about the computing
- When a startup requires a close marriage of the two cultures, be sure to understand
  - who's on top?
  - how well do they communicate with each other?
  - is the simulation approach chosen to fit the biological problem, or vice versa?

An Afterthought

Multi-cellular Computing

In the 20th century we sought to create ever more capable computers.

Now we seek to create ever more elegant groups of collaborating computers.

We recapitulate the biological evolution from single-cell to multi-cell organisms.