UCL

Computational Challenges of Systems Biology

Anthony Finkelstein
University College London
CoMPLEX & Computer Science

Work with Jonathan Ashmore, David Bogle, James Hetherington, Linzhong Li, Ofer Margoninski, Peter Saffrey, Rob Seymour & Anne Warner

Modeling and identification of metabolic systems

CARSON, E. R., C. COBELLI, AND L. FINKELSTEIN. *Modeling and identification of metabolic systems.* Am. J. Physiol. 240 (Regulatory Integrative Comp. Physiol. 9): R120–R129, 1981.—Introductory principles of physiological systems analysis by computer simulation or computation are introduced. The problems of model formulation, identification, and validation are examined. Selected guidelines for successful modeling are suggested, and examples of such applications in the field of endocrinology and metabolism are given.

mathematical model; compartment; physiological systems analysis; control system; system identification

Challenging Applications

- New applications of computing rarely attract much attention from computer scientists unless they pose novel computational challenges, stretch the state-of-the-art or open an unanticipated use of computing concepts.
- Bioinformatics is an example of an application that has attracted such attention …

The Molecular Revolution

A revolution that has reshaped the life sciences

- We now understand:
  - the DNA sequence of many genes, up to whole genomes
  - the mechanics of much of RNA synthesis in exquisite detail
  - the genetic code for specifying amino acids so that the backbone of a protein can be directly predicted from DNA sequence information
  - some of the complexities of RNA splicing, the means by which one gene can generate many RNAs and therefore proteins

The Molecular Revolution

  - how DNA sequences, called promoters, determine which genes are expressed
  - how DNA binding proteins, called transcription factors, modify gene expression
- Knocking-out and over-expressing genes and RNAs have revealed how particular genes contribute to certain biological processes; it has also revealed substantial functional redundancy.

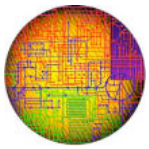
Bioinformatics

- In the process of achieving this revolution in understanding we have accumulated very large amounts of data
- The scale of the data, its structure and the nature of the analytic task have merited serious attention from computer scientists and have prompted work in intelligent systems, data-mining, visualisation and more
- It has also demanded serious efforts in large-scale data curation and worldwide infrastructure
  - Bioinformatics the handmaiden of molecular biology

Limits of Bioinformatics

- Bioinformatics is only the first step in reshaping the life sciences
- For further progress, we must return to the study of whole biological systems: the heart, the cardiovascular system, the brain, the liver
  - systems biology
- To succeed we must combine information from the many rich areas of biological information. Alongside the *genome*, our knowledge about genes, we place the *proteome*, *metabolome*, and *physiome,* our information about proteins, metabolic processes, and physiology

Systems Biology

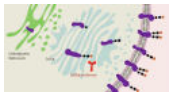


- We must build an *integrated* physiology of whole systems
- Systems biology is at least as demanding as, and perhaps more demanding than, the genomic challenge that has fired international science and excited the attention of the public
- To achieve it will involve computer scientists working in close partnership with life scientists and mathematicians. By contrast with the molecular biology revolution, computer science will be proactively engaged in shaping the endeavour rather than clearing up afterwards!

The Prize

- The prize to be attained is immense!
  - From 'in-silico' drug design and drug testing
  - To individualised medicine that will take into account physiology and genetic profile
- Systems biology has the potential to have a profound impact on healthcare and beyond.

Cataloguing is not Understanding

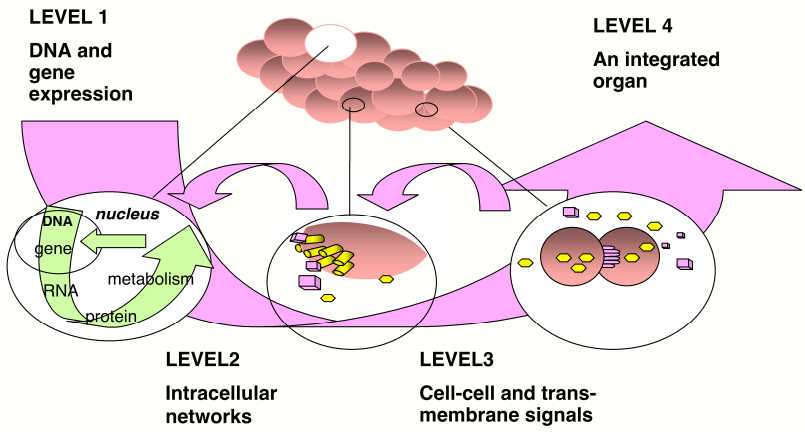


- Even if we had a catalogue of all the gene sequences, how they are translated to make proteins, which protein can interact with which, and the way in which the protein back bones fold, we would not be able to put them into a functionally meaningful framework:
  - All proteins are post-translationally modified. These additions influence the actual shape of proteins
  - Just because two proteins can interact, it does not mean that they do so in real cells
  - Many functionally important, small molecules are synthesized by metabolism

Modelling

- A bottom-up, 'data-driven' strategy, will not work — we cannot build an understanding of biological systems from an understanding of the components alone
- What other approaches might be tried?
  - We can use experimental information to build models at different biological scales, integrating these models to create an 'orchestrated' assemblage of models ranging from gross models of physiological function through to detailed models that build directly on molecular data

From Gene to Organ

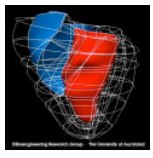


Key Concepts

- Key concepts for systems biology forced upon us by the peculiar complexity of biological systems
  - the importance of simplification
  - the importance both of modularity and of the integration of the modules
  - iteration between model and experiment as the key to ensuring that models are realistic

Models: State-of-the-Art

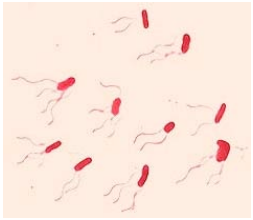
- The paradigmatic example of systems biology is the model of the heart developed by Denis Noble
  - a computational model of the electrical and mechanical activity of the heart in health and disease, linked to sophisticated visualisations
  - invaluable in developing an understanding of cardiac arrhythmia with consequences both for drug design and testing
  - grown from relatively simple beginnings in 1962 as an adaptation of the classic Hodgkin-Huxley squid axon model (one of the landmark achievements of modern biology), to its current state involving hundreds of equations and adjunct models, such as a finite element model



Models: State-of-the-Art

- Only covers a small part of the mechanical, electro-physiological, and chemical phenomena of the heart, hence reveals not just what can be achieved but suggests the scale of the challenge that Systems Biology presents
- There exist a plethora of 'stand-alone' models of various biological phenomena produced by different researchers. They are mostly relatively simple, although a few are more sophisticated

the bacterial model of Dennis Bray that models flagellar motion (flagella are thin projections from cells) and chemo-sensitivity



Exemplar

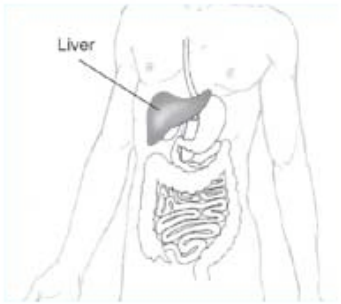
- We need additional convincing exemplars of systems biology of the general type of the heart model:
  - explicitly 'engineered' with some systematic modularity and separation of concerns among component models
  - the liver is the subject of a major UK research project funded by a Department of Trade & Industry 'Beacon' scheme that supports 'high adventure science' with the possibility of advances that have significant industrial potential. The aim of the project is to produce a physiological model of the liver that is integrated across scales …

Exemplar

- The liver has been selected as a good exemplar of systems biology:
  - it is medically important and structurally relatively homogeneous
  - it is also challenging – the liver is primarily a chemical system, where the heart is electromechanical
  - there are also a number of ongoing efforts to build 'in vitro' livers, that is artificial livers that can be used while patients who have suffered liver damage are recovering

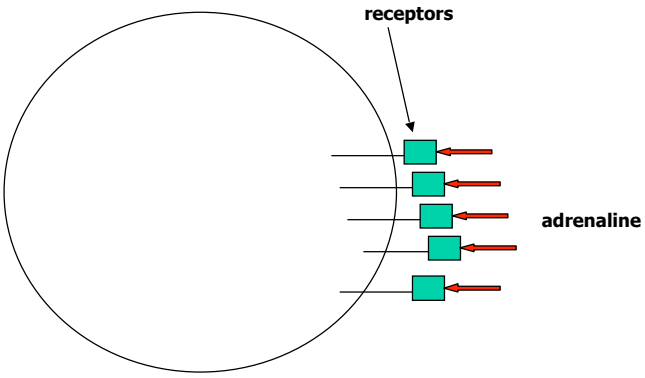
The Liver

- The liver has three principal functions:
  - it stores materials to be released into the blood stream when needed
  - it synthesizes proteins and peptides from amino acids
  - it detoxifies the system by breaking down harmful materials such as alcohol, which are then excreted



Taster: Glucose Release from Hepatocytes

Adrenaline circulating in the blood stream binds to ß - adrenergic receptors on the membrane of the hepatocyte



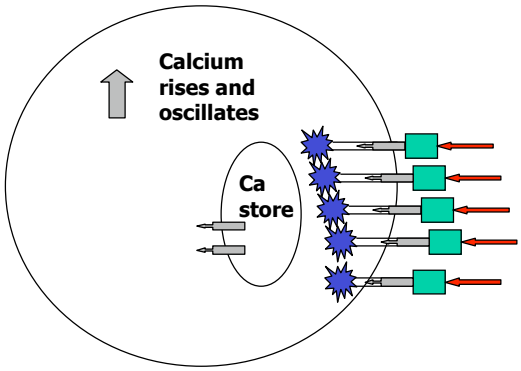
Taster: Glucose Release from Hepatocytes

When adrenaline binds to the receptors, ion channels open in the membrane

Calcium enters the cell through ion channels

G-proteins are activated

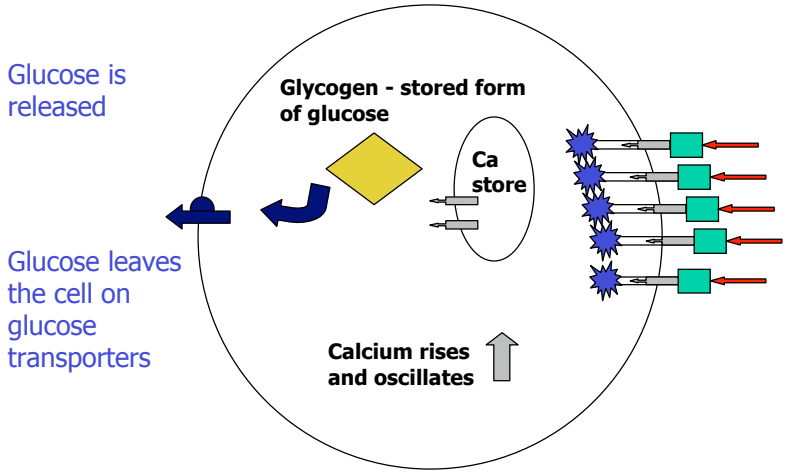
Calcium leaves the store



Cytoplasmic calcium rises; the rise in calcium suppresses further Ca release and cytoplasmic Ca oscillates

Taster: Glucose Release from Hepatocytes

The increase in calcium mobilizes glucose release from glycogen



Glucose Release from Hepatocytes

- Models of each of these sub-processes, such as G-protein activation or cytoplasmic calcium oscillation, may be constructed in isolation
  - typically these may be modelled as ordinary differential equations (ODEs) though certain processes appear to lend themselves to discrete event modelling
  - the processes have, in this case, been well studied experimentally and the parameters, that constitute the context, can be related systematically to values in the literature
  - by way of a mediating ontology such as the Gene Ontology

Glucose Release from Hepatocytes

- Assuming homogeneous models of the sub-processes we can connect these together to build a detailed model of the entire network
  - representational heterogeneity naturally makes this more difficult
- Alongside this model we can build a 'simplified' or gross model. Rather than the more complex behaviours built into the models of ion channel opening, protein activation etc. we assume these behave as perfect switches to make the system piecewise linear

Glucose Release from Hepatocytes

- The simplified system is biologically unrealistic, and many features, such as the shape or period of oscillations, are not preserved
  - some, however, are, and the simplified model allows us to use algebraic analysis, facilitating the development of understanding about the system

Systems Engineering 101

Only when you understand the task that lies in front of you can you devise an appropriate strategy. A mistake now - in the planning stage - of systems biology will be immensely costly subsequently

The models that we have built so far have been small and concerned with only a limited range of phenomena

They are tightly tied to the parameter data. This data is deployed for this purpose only and is shorn of context

The models are hand-crafted and stand-alone. They are presented with a limited range of metadata

They are usually presented statically with pre-digested results. Modelling assumptions are submerged and the models are stripped of their rationale

The models are flat and there is no indication or understanding of the level at which these models should be built. There is no 'information hiding'

Models are presumed to live till publication. They are not designed for change and they lack traceability. The models are 'not engineered'

Now … imagine what systems biology will be like if we succeed in our ambitions …

Many 10s of thousands of different highly complex models expressing different phenomena and at different levels of detail

Each part of heterogeneous assemblages of models with diverse ownership

Evolving and changing independently and as 'families' … sometimes contested. With 'interpretations' that themselves change and must be managed

Directly tied to the wealth of (changing) bioinformatics data and to the scientific literature

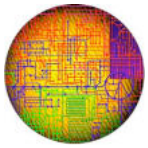
Directly tied to experimental data and protocols

Related to a complex set of (changing) computational resources required to interpret models

Responding to challenging usability and visualisation demands from a wide variety of stakeholders

Oh yes … add security for medical data, curation and reproducibility for science data, and verification for pharmaceutical regulation

The biggest (software) engineering problem in the world … ever! Means planning for the management and engineering challenges now. Not modelling, not biology but model management



Immediate Implications

Deterministic	⟷	Stochastic
Compartmental variables	⟷	Individual or functional
Spatially homogeneous	⟷	Spatially explicit
Uniform time scale	⟷	Separated time scales
Single scale entities	⟷	Cross-scale entities

Immediate Implications

- Intermediating infrastructure
- Configuration and version management
- Fine grain workflow management
- Methodological alignment

What we have done:

- A framework for selecting modelling schemes
- A metamodel for systems biology
- A model parameter repository
- A prototype middleware for integrating heterogeneous modelling platforms
- A unified ontology framework for systems biology models
- A new versioning and impact analysis tool for systems biologists

(the first) modular, integrative, scale-crossing, hybrid model of liver glucose homeostasis

Where Now?

- Unlike projects to map genomes there is no clear end-point for systems biology. Important staging posts:
  - models that provide 'thin' vertical slices across scales are one such
  - the development of models that are approved for drug testing, perhaps in place of animal models, and that satisfy the strict requirements of validity, reliability, transparency and traceability
  - the establishment of global 'collaboratories' in which models can be exchanged, reviewed and analysed
- Finally, when we can dependably diagnose health issues and identify novel treatments using our models, systems biology will have come of age