Computational Biology

Christine Beveridge
Level 2, Building 62
c.beveridge@uq.edu.au

1

Reading

- Haefner J.W. Modeling Biological Systems - Principles and Applications, QH308.2 H325
- G. Polya - How to solve it: A new aspect of mathematical method. QA 11 P6

2

The Next Three Weeks

- Today – brief overview of bioinformatics and DNA sequence analysis and proteins analysis;
- The Rest:
  - *Not* about new theorems;
  - Learn how to think about systems, analyse them and build useful models;
  - Learn how to apply your knowledge of maths to solve problems using mathematical models;
  - You are expected to be able to create simple algebraic and differential equations and to write Matlab code to create corresponding models.

3

What is Computational Biology?

- Computational biology includes: bioinformatics, biomathematics and systems biology
- It includes solving problems involving: DNA, RNA, proteins, metabolic networks and/or systems fromsingle organisms to populations of many organisms.
- It is <u>not</u> simply maths + computing + biology
- It is a new discipline which spans all of biology and addresses questions that have a substantial size, complexity or predictive component.

4

Why Study or Use Computational Biology?

- Like much of biology, the genetic and protein codes are made of simple units.
- Simple interactions lead to complex emergent properties
- The complexity and size of the data sets requires computational approaches.
- Understanding and interpreting some parts of the genetic code and the regulation and relationships within and between various levels (DNA, RNA, Protein) has not been fully understood – computational approaches provide a rational platform that accommodates complexity that is inherent due to size and/or interactions.

5

Some Questions that Require Bioinformatics

- Comparing human genome with mouse or plant. All have about 40,000 genes each with about 5000 different bases.
- Searching for motifs which are related copies of sequence in the part of the gene that determines how many copies of RNA are produced.
- Searching in lists of all known genes of any species for a close match with a gene of interest (BLAST).
- Determining the 3-D structure of a protein

6

DNA to Ecosystems

- DNA is composed of strings of information
  - Text searching is important including similar and exact matches across huge data sets from multiple origins (species)
- DNA is the template from which proteins are made; proteins are what does the work, not the DNA!
  - Proteins are 3D structures and interact with other proteins based on that structure and biochemical properties – modelling physical properties of proteins is important

7

DNA Structure

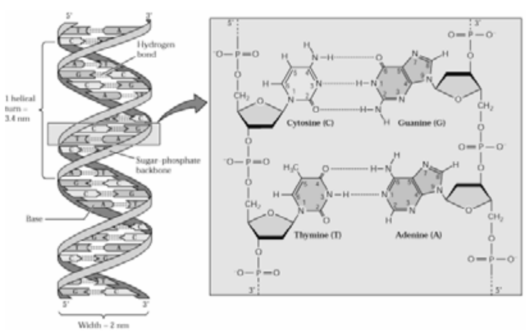
DNA is always made of combinations of bases: A, T, C, G

Adenine
Cytosine
Guanine
Thymine

These are held in a double helix by hydrogen bonding and a sugar-phosphate backbone

8

DNA Structure

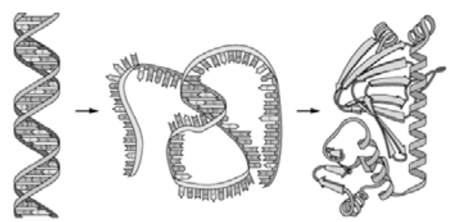


Double stranded DNA is always paired A-T and C-G

9

DNA Sequence to Protein Sequence

DNA **RNA** **Protein**



A, T, C, G **A, U, C, G** **20 amino acids**

Uracil

10

Transcription and Translation

DNA	**5'---CTCAGCGTTACCAT---3'** **3'---GAGTCGCAATGGTA---5'**
	↓ *transcription*
RNA	**5'---CUCAGCGUUACCAUG---3'**
	↓ *translation*
PROTEIN	**N---Leu-Ser-Val-Thr---C**

- The sequence of amino acids in the protein is determined by the sequence of bases (triplet codons) in the gene.

11

Triplet codons (3 bases) are programmed to make aminoacids (shown in yellow) that string together to make proteins

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC		UCC		UAC		UGC	
	UUA	Leu	UCA		UAA	STOP	UGA	STOP
	UUG		UCG		UAG		UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC		CCC		CAC		CGC	
	CUA		CCA		CAA	Gln	CGA	
	CUG		CCG		CAG		CGG	
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC		ACC		AAC		AGC	
	AUA		ACA		AAA	Lys	AGA	Arg
	AUG	Met	ACG		AAG		AGG	
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC		GCC		GAC		GGC	
	GUA		GCA		GAA	Glu	GGA	
	GUG		GCG		GAG		GGG	

12

Some Questions that Require Bioinformatics

- Comparing human genome with mouse or plant. All have about 40,000 genes each with about 5000 different bases.
- Searching for motifs which are related copies of sequence in the part of the gene that determines how many copies of RNA are produced.
- Searching in lists of all known genes of any species for a close match with a gene of interest (BLAST).
- Determining the 3-D structure of a protein

Systems Biology

- "Unlike traditional science which examines single genes or proteins, **systems biology** studies the complex interaction of all levels of biological information: genomic DNA, mRNA, proteins, functional proteins, informational pathways and informational networks to understand how they work together."
- May be considered holistic.
- Understanding all or most components in a system or underlying a problem allows for better solutions and predictions.

What are the properties of biological systems?

- Several interacting components
- Some components have different properties
- Properties of components may change over time
  - independently, or more commonly…
  - due to interactions with other components
- Emergent properties of systems might not be apparent from observation of the individual components and their interactions

Range of Biological Information and Systems

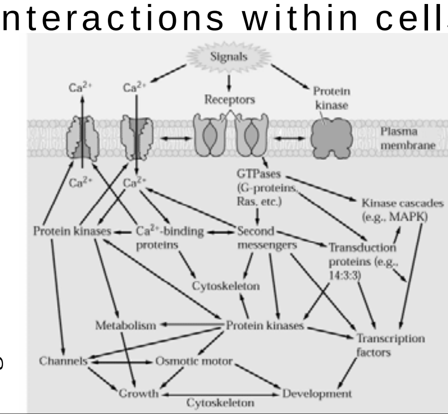
Genetic	(DNA and RNA)
Epigenetic	(DNA modification such as methylation)
Physiological-Cellular Level	(DNA-RNA interactions, Proteins, Structural, Metabolism, Signal transduction)
Physiological- Organism Level	(Structural, Metabolism, Long-distance signalling, Development, Immune system)
Populations	(Population dynamics, Evolution)
Ecosystem	(Interacting Populations, environment ←→ populations)

Some Questions in Systems Biology

- How will a 10% increase activity of a given protein (e.g., for sucrose synthase) affect a given property (e.g., sugarcane yield)?
- What will the rate of spread of chicken flue be if introduced into Sydney on Boeing 747?
- What effect will a 15% increase in sewage waste into Morton Bay at Scarborough have on the estuary health?
- What park-size is required to protect the hairy-nose wombat?

Interactions within cells

Signal transduction cascade

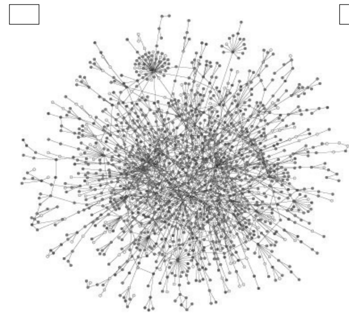


Protein Networks

- Protein interactions form an important part of cell function
- Protein interaction networks involve
  - a small number of proteins that interact with a large number of other proteins and
  - a larger number of proteins that interact with a small number of proteins.
- This is very similar to the structure of the internet.
- What are the properties of such networks and what can this tell us about biology?

19

Protein Network in Yeast

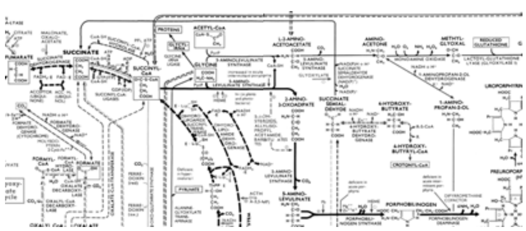


red, lethal
green, non-lethal

From Jeong et al Nature. 411:41, 2001

20

Metabolic Pathways in Cells



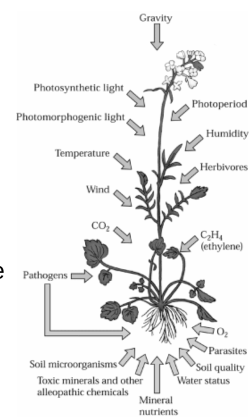
From: http://www.expasy.ch/ follow links to: Boehringer Mannheim's Biochemical Pathways

- Predictions are difficult due to complexity
- This complexity can be handled by computational approaches

21

Whole Organisms

- Shown here are the environmental influences on plant growth and development.
- Similar concepts as at the cellular and molecular levels can be used to model the interplay of these factors.



22

Why Model Systems?

- When the answer is not obvious without a model
- To better predict, understand or capture the emergent properties of the system
- To understand the mechanisms that lead to the emergent properties
- To test what hypotheses are consistent with the biological data
- To predict effects of change in the system
- To calculate relationships between parts of the system

23

Models are used to…

- Understand
  - gain insight on a biological system
  - test an idea, hypothesis or other system of logic
- Predict
  - future events or
  - properties of the system or object that are currently unknown
- Control
  - events or objects in a real system

24

Forms of models

- Conceptual or Verbal
  - Descriptions in words
- Diagrammatic
  - Drawings of the representations of objects and inter-relations
- Physical
  - E.g. an engine, molecular "lego" models
- Formal
  - Algebraic and differential equations

25

Types of Models

- Analytical
  - E.g., fluid dynamics to describe water flow, enzymatic rates and equilibrium
- Continuous
  - Dynamic model incorporating any fraction of time, space and/or other model parameter
- Discrete
  - Values in a dynamic model such as time and/or space occur only in intervals, each model containing a finite number of these values

26

Types of Models cont.

- Deterministic
  - A model with no random events and hence produce the same output in recursions with the same inputs
- Stochastic
  - Includes random events and can generate different outputs from the same inputs
- Statistical model
  - Examines distributional properties of the data - E.g., line of best fit

27

Types of Models cont.

- Process
  - Explicitly incorporates aspects of biological processes
- Rule-based
  - Behaviour of elements/agents/objects are governed by a set of rules and conditional statements
- Individual-based
  - Individual objects and interactions among them are modeled to produce the whole model.

28

Types of Models and Movement

- Compartment
  - Describe flow of physical material (e.g. water) between compartments (usually use differential or finite difference equations)
- Transport
  - Describe movement of material, energy or momentum in physical space (using partial differential equations and conservation principles
- Particle/Individual
  - Describe the flow of individual objects (e.g., blood cells, ants, disease)

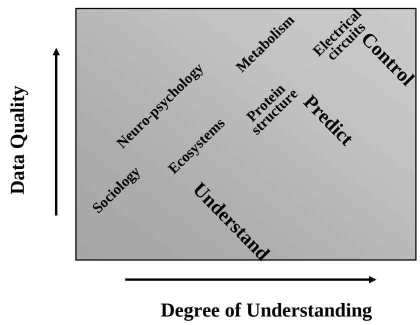
29

Choice of model & Level of Abstraction

- Carefully consider the biological problem
- How should objects, movement, time, space, processes and relationships be represented?
- What balance of:
  - Realism: the degree to which the model represents the physical properties
  - Precision: the accuracy of the model predictions
  - Generality: the number of systems and situations to which the model correctly applies
- Consider the constraints of:
  - Data quality
  - Degree of understanding

30

Relationship between data quality and understanding



31

Model Objectives

- Make a statement of the objectives of the model.
  - What is the system to be modelled?
  - What are the major questions to be addressed by the model?
    - How will the model be applied?
  - What is the "stopping rule" for the modelling activity?
    - How good must the model be - compared to what?
  - How will the output be analyzed, summarised and used?

32

Formulation:
Understand the problem

- What are the unknown(s)?
- What are the data?
- What is the condition/relationships among the data?
- Identify all the objects and relationships among them.
- Develop a list of specific hypotheses (verbal statements) that may meet the objectives.
- Create a conceptual model: Often a diagram or flowchart is used.

33

Formulation:
Devise a plan

- Can you make use of a related problem that has been solved?
- Could you simplify the problem by restating it or by first solving a related problem or part?

34

Mathematical Formulation

- Convert qualitative hypotheses to specific quantitative relations.
- Formulate these relations into mathematical equations.
  - This step is challenging as it requires that vague concepts and loose relations be translated into mathematics.

35

Algorithm, Implementation and Verification

- Translate the equations into an appropriate algorithm for the selected computer code/program.
- Verify that this translation/algorithm is correct for the defined mathematical relationships.

36

Calibration: Parameter Estimation and Trials

- Before meaningful output can be generated (simulations) the model may require initial values and constraints.
- This may require repeated running of the model to modify parameters until the output meets observed data values or trends.
- This process in itself may already meet some objectives of the modelling.

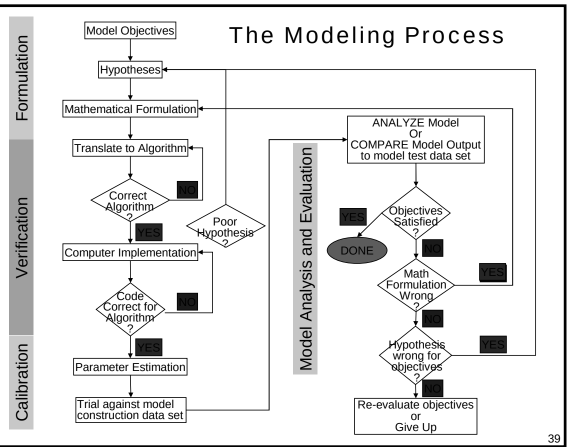
37

Outcome Evaluation

- Evaluate outcomes in light of the model objectives.
- Is hypothesis false or supported?
- Re-confirm algorithms (maths and computing) are valid for the system and hypotheses?
- Define another way to check your results?
- Can the results be derived differently?
- Will you be able to apply this approach to other problems?

38

The Modeling Process



39

Analysis and Evaluation

- The answer or output from the model should be evaluated against the objectives.
- At best, this requires an independent data set for comparison.
- One can debate here as to whether we are trying to use our model to prove (or state our understanding at a point in time) or disprove something. [Normally in science, we are trying to disprove something.]

40