Modelos Formales en Bioinformática

Javier Campos, Jorge Júlvez
University of Zaragoza
2011

Outline

Outline

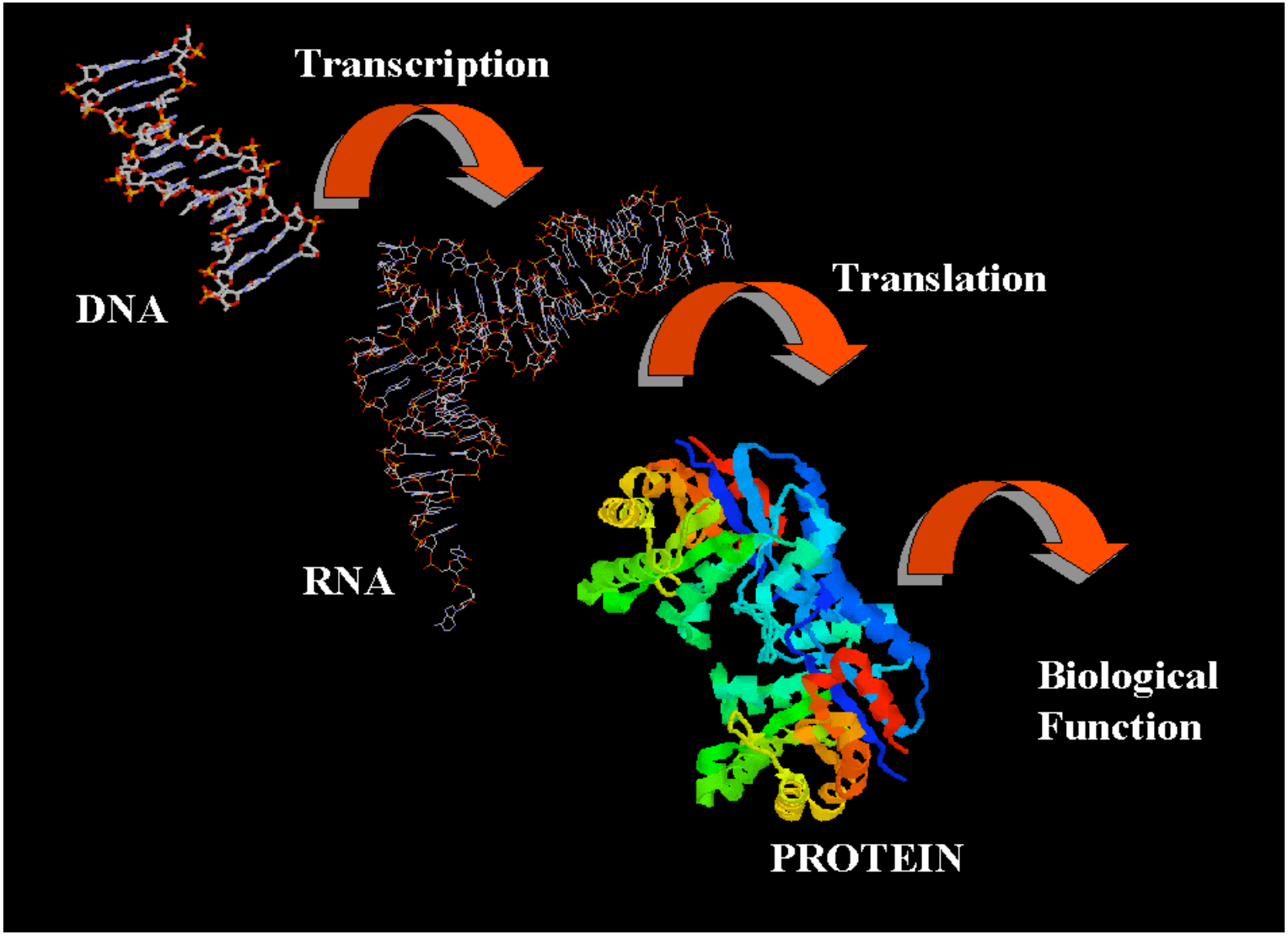
Formal Models in Bioinformatics

What is systems biology?

- Systems biology is the study of all the elements in a biological system (all genes, mRNAs, proteins, etc) and their relationships one to another in response to perturbations.
- Systems approaches attempt to study the behaviour of all the elements in a system and relate these behaviours to the systems or emergent properties.

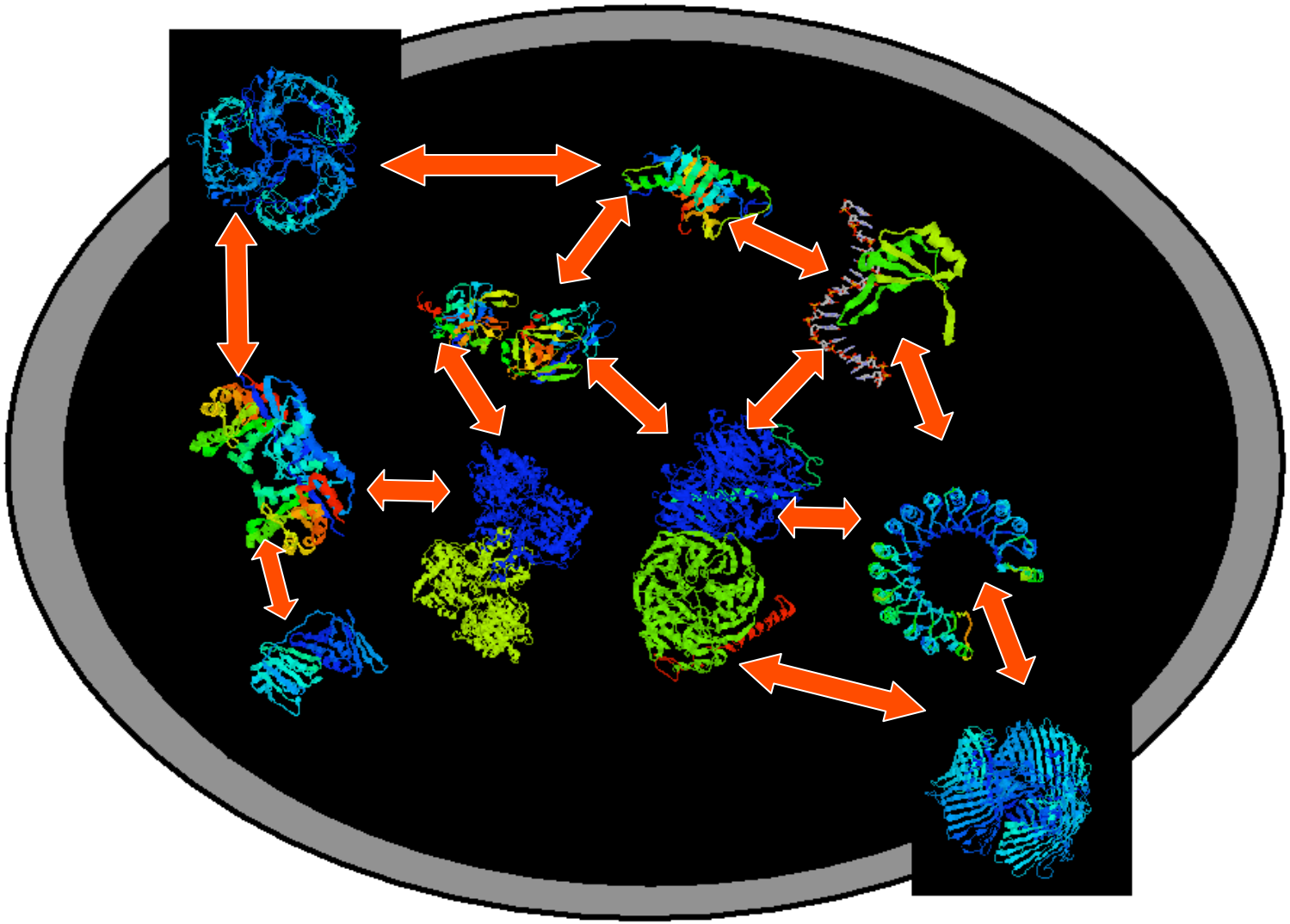
Formal Models in Bioinformatics

Bioinformatics Basics



Formal Models in Bioinformatics

Systems Biology: Interaction in Networks



Formal Models in Bioinformatics

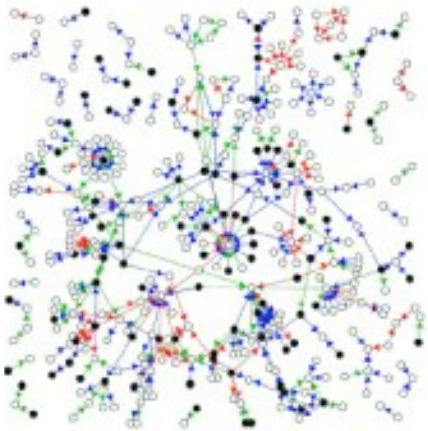
What is systems biology?

- ...systematic study of complex interactions in biological systems, thus using a new perspective (integration instead of reduction) to study them... one of the goals of systems biology is to discover new emergent properties (Wikipedia)
- Systems biology is the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions... systems biologists focus on all the components and the interactions among them, all as part of one system (Institute for Systems Biology, Washington)
- To understand complex biological systems requires the integration of experimental and computational research – in other words a systems biology approach (Kitano, 2002)

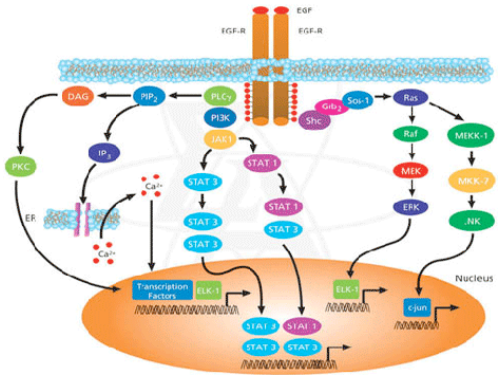
Formal Models in Bioinformatics

Networks

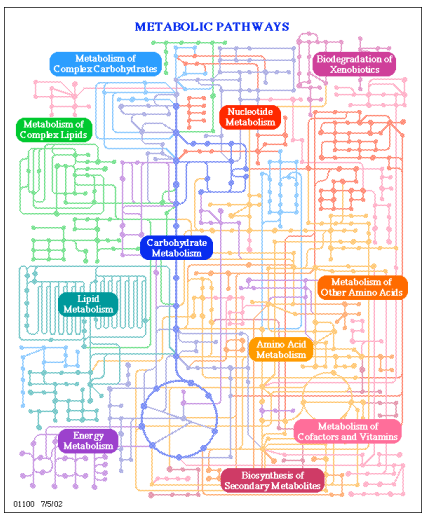
Gene regulation



Signalling

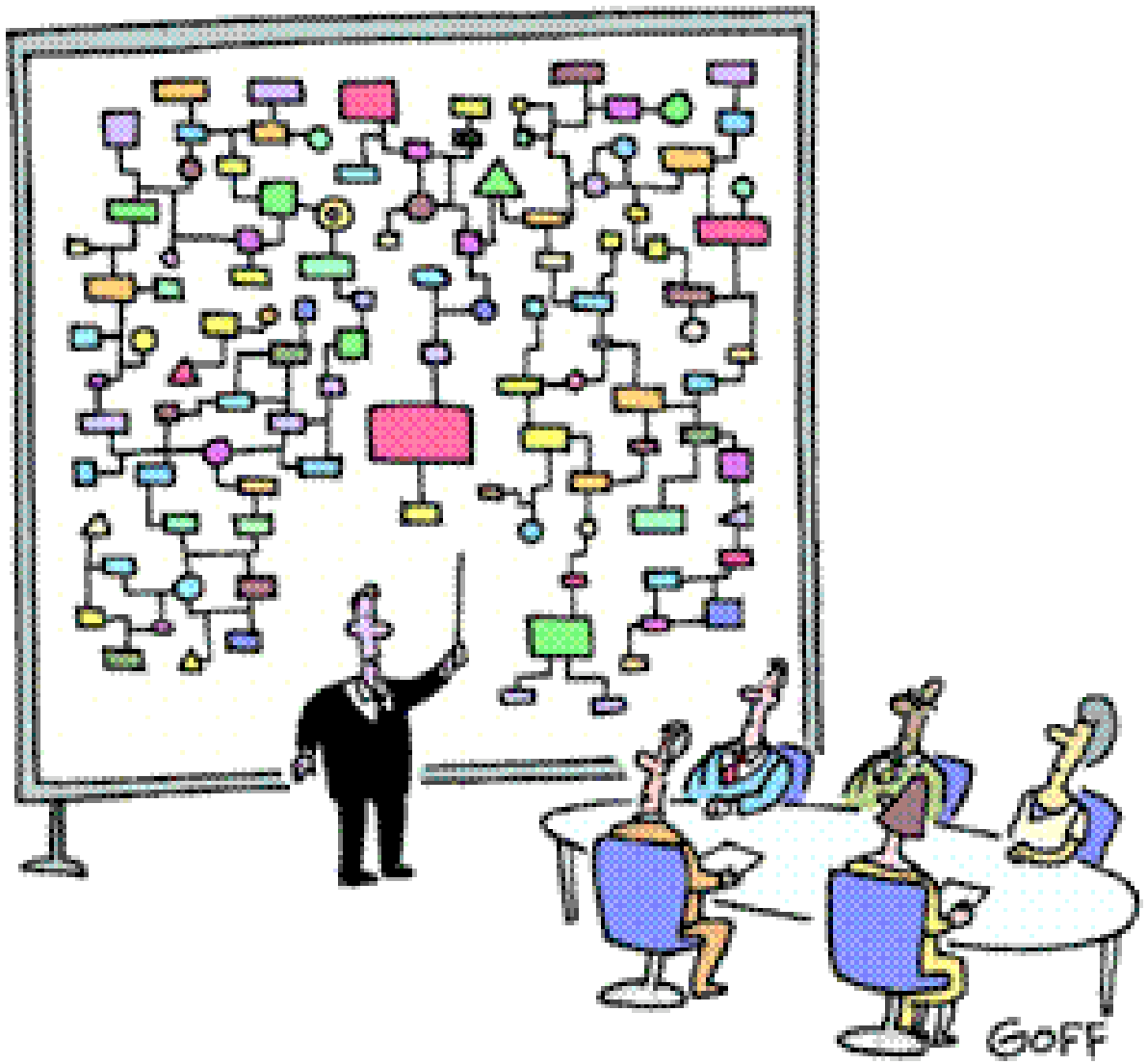


Metabolic Pathway



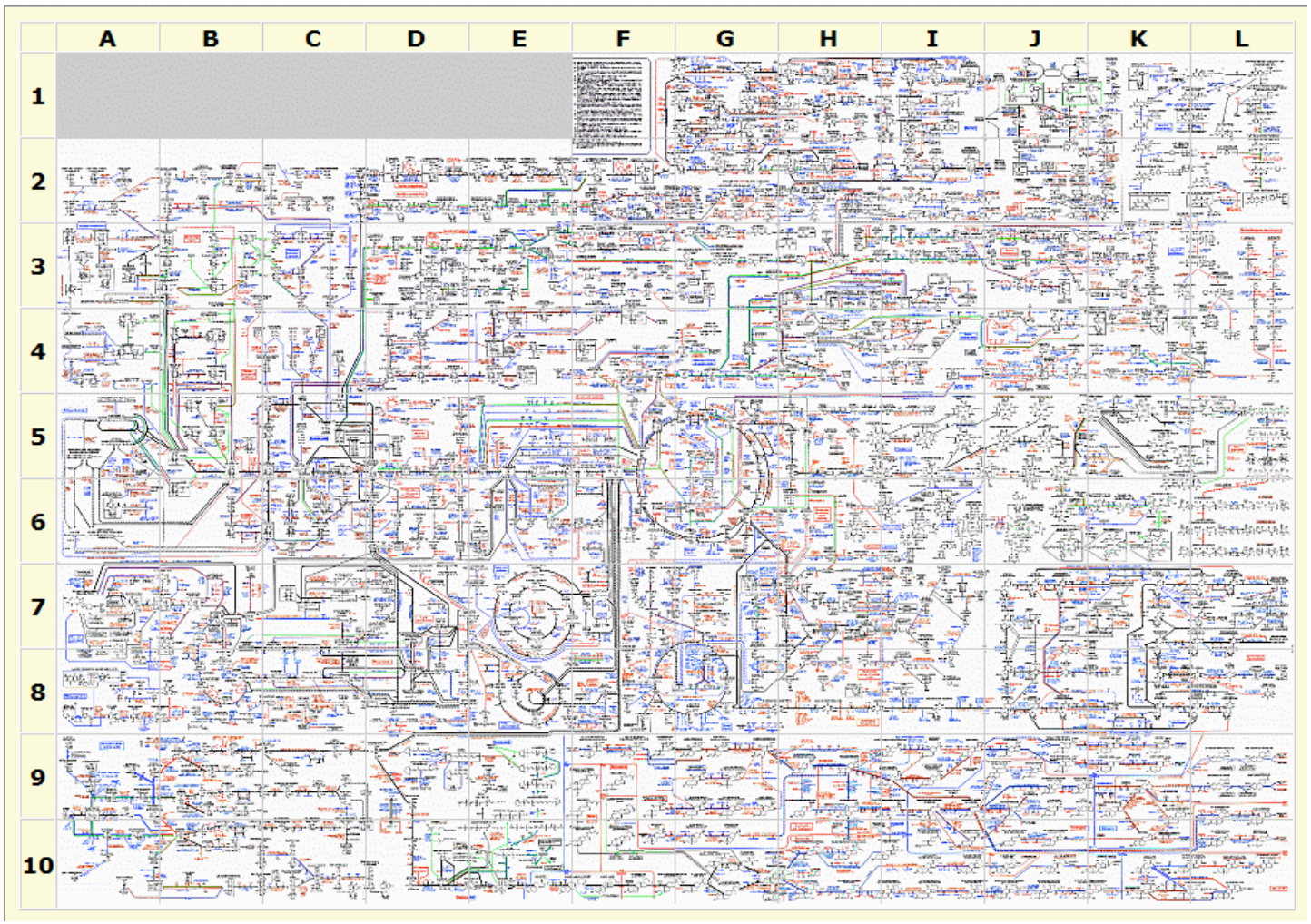
Protein-protein interaction

Formal Models in Bioinformatics



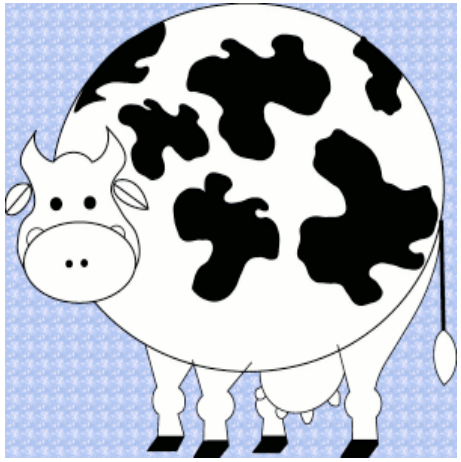
Formal Models in Bioinformatics

Metabolic Pathway



Systems Biology is a multidisciplinar field

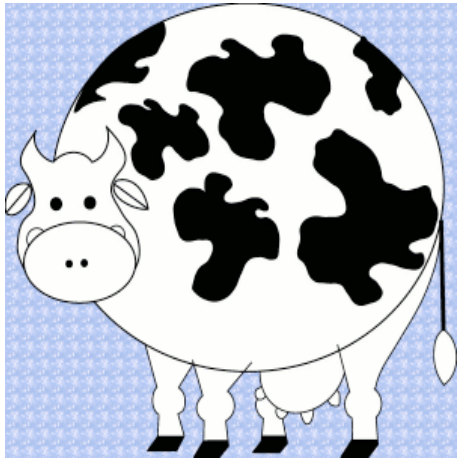
Which is the volume of a cow?



- A chemist perspective: Dissolve the cow in H_2SO_4, weight the result and measure the volume.
  - Sometimes it is important not to destroy the cow.

Systems Biology is a multidisciplinar field

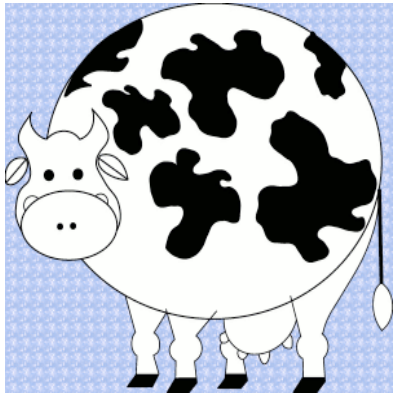
Which is the volume of a cow?



- A chemist perspective: Dissolve the cow in H_2SO_4, weight the result and measure the volume.
  - Sometimes it is important not to destroy the cow.
- An engineering perspective: Immerse the cow in a tank of water and measure the volume.
  - Some cows cannot swim. Water pressure might change volume.

Systems Biology is a multidisciplinar field

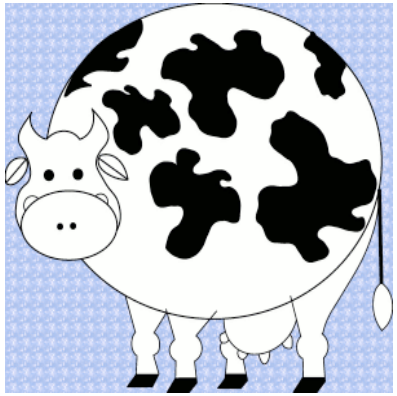
Which is the volume of a cow?



- A mathematician perspective: Cut the cow into pieces and sum up the pieces.
  - The total might not be equal to the sum of its parts → emergent properties.

Systems Biology is a multidisciplinar field

Which is the volume of a cow?



- A mathematician perspective: Cut the cow into pieces and sum up the pieces.
  - The total might not be equal to the sum of its parts → emergent properties.
- A physicist perspective: Consider a spherical cow with negligible mass...
  - Even E. coli is not that spherical

Outline

Formal Models in Bioinformatics

Mathematics to model the time evolution of a population

Formal Models in Bioinformatics

Mathematics to model the time evolution of a population

From penguins to rabbits..

Modelling the time evolution of a population

Early attempts to make use of mathematics in biology

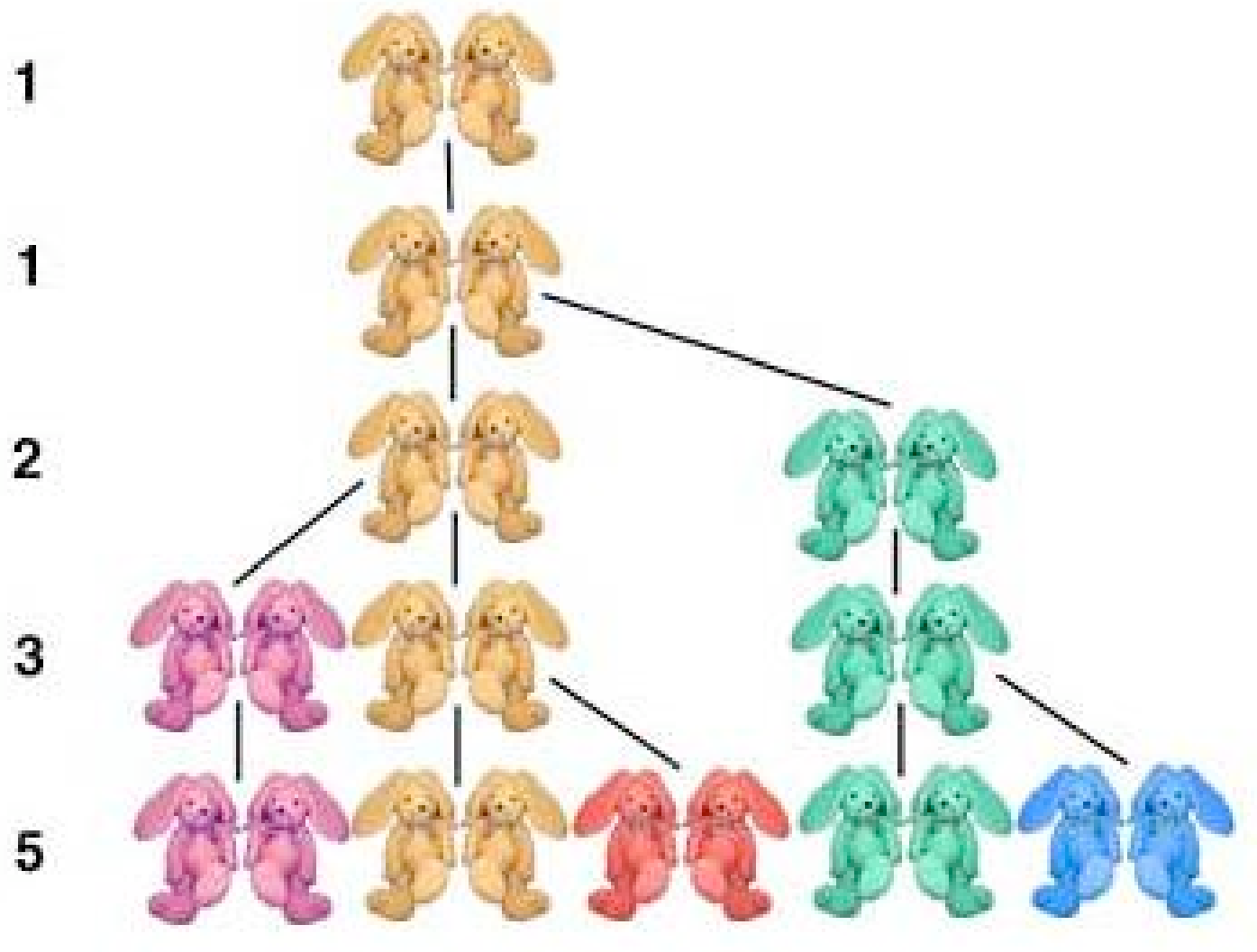
Leonardo de Pisa or Fibonacci (1202)

- At month 0 there is a pair of rabbits (one female and one male).
- Every pair of rabbits (one female and one male) can mate at the age of one month.
- The female rabbit always produces a new pair of rabbits (one female and one male) every month from the second month on.
- As there is no death, all rabbits survive.

What is the number of pairs of rabbits in month n?

Modelling the time evolution of a population

Rabbits population



Modelling the time evolution of a population

Leonardo de Pisa or Fibonacci (1202)

- At month 0 there is a pair of rabbits (one female and one male).
- Every pair of rabbits (one female and one male) can mate at the age of one month.
- The female rabbit always produces a new pair of rabbits (one female and one male) every month from the second month on.
- As there is no death, all rabbits survive.

The number of pairs in month n, R_n, satisfies:

$$R_{n+1} = R_n + R_{n-1}$$

$R_0 = 1$
$R_1 = 1$
$R_2 = 1 + 1 = 2$
$R_3 = 2 + 1 = 3$
$R_4 = 3 + 2 = 5$
$R_5 = 5 + 3 = 8$

Modelling the time evolution of a population

Deterministic modelling

- Let $N(t)$ be the population of those penguins at time t.
- $N(t)$ is the number of individuals in the population at time t.
- The change in the number of penguins in a small time interval, from t to $t + \Delta t$, is given by:

$$N(t + \Delta) = N(t) + \textit{births} - \textit{deaths} + \textit{migration}$$

- This equation is a conservation equation for the number of individuals of the population.
- The form of the various terms on the right-hand-side requires essential feedback from biologists.

Modelling the time evolution of a population

Deterministic birth process

Let us assume that:

- There are no death events in the population.
- There are only birth events in the population.
- The birth rate (number of births per unit of time), b, is the same for all individuals of the population.
- We have:

$$N(t + \Delta t) = N(t) + \ births$$

Modelling the time evolution of a population

Deterministic birth process

- Change in the population is due to birth events.
- The births in the time interval $[t, t + \Delta t]$ due to a single individual is $b\Delta t$.

Modelling the time evolution of a population

Deterministic birth process

- Change in the population is due to birth events.
- The births in the time interval $[t, t + \Delta t]$ due to a single individual is $b\Delta t$.
- The births in the time interval $[t, t + \Delta t]$ due to all individuals is $N(t)b\Delta t$.

Modelling the time evolution of a population

Deterministic birth process

- Change in the population is due to birth events.
- The births in the time interval $[t, t + \Delta t]$ due to a single individual is $b\Delta t$.
- The births in the time interval $[t, t + \Delta t]$ due to all individuals is $N(t)b\Delta t$.
- $N(t+\Delta t) = N(t)+N(t)b\Delta t \implies \dfrac{N(t+\Delta t) - N(t)}{\Delta t} = bN(t)$

Modelling the time evolution of a population

Deterministic birth process

- Change in the population is due to birth events.
- The births in the time interval $[t, t + \Delta t]$ due to a single individual is $b\Delta t$.
- The births in the time interval $[t, t + \Delta t]$ due to all individuals is $N(t)b\Delta t$.
- $N(t+\Delta t) = N(t)+N(t)b\Delta t \implies \dfrac{N(t + \Delta t) - N(t)}{\Delta t} = bN(t)$
- For a very small time interval, $\Delta t \to 0$,

$$\lim_{\Delta t \to 0} \frac{N(t + \Delta t) - N(t)}{\Delta t} = \frac{dN(t)}{dt} = bN(t)$$

- This equation can be easily solved by integration.

Modelling the time evolution of a population

Deterministic birth process

$N(t)$: Number of individuals at time t

$$\frac{dN(t)}{dt} = bN(t)$$

- If the population at time $t = t_0$ is given by N_0, we have:

$$N(t) = N_0 e^{b(t-t_0)}$$

- In a deterministic birth process the population size is predicted at time t with **absolute certainty**, once the initial size N_0 and birth rate b are given.
- The population size $N(t)$ and time t are both continuous variables (both take real values) and not discrete (take integer values).

Modelling the time evolution of a population

Deterministic birth process

$N(t)$: Number of individuals at time t

$$\frac{dN(t)}{dt} = bN(t)$$

- If the population at time $t = t_0$ is given by N_0, we have:

$$N(t) = N_0 e^{b(t-t_0)}$$

- In a deterministic birth process the population size is predicted at time t with **absolute certainty**, once the initial size N_0 and birth rate b are given.
- The population size $N(t)$ and time t are both continuous variables (both take real values) and not discrete (take integer values).
- Is this a good mathematical population growth model?

Modelling the time evolution of a population

Stochastic birth process

- Let $\mathbf{X}_t$ be the dicrete random variable that describes the number of individuals of the population at time t.
- The stochastic process that describes the population satisfies:

$$\mathbf{X}_t \in \{1, 2, \ldots\} \text{ and } t \in [0, +\infty)$$

- Denote by $p_n(t)$ the probability that at time t the size of the population is n, i.e., the probability that at time t there are n individuals in the population:

$$p_n(t) = Prob(\mathbf{X}_t = n)$$

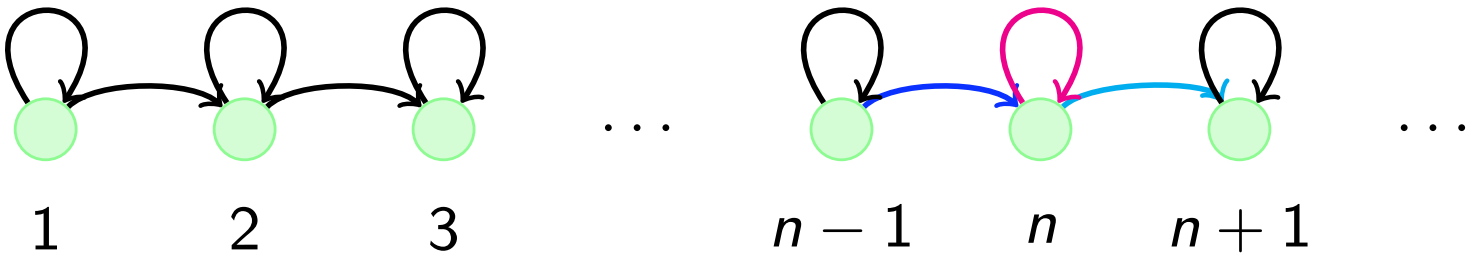
Modelling the time evolution of a population

Stochastic birth process

- Consider a small time interval $[t, t + \Delta t]$
- How is $\mathbf{X}_{t+\Delta t}$ related to $\mathbf{X}$
- We have the following rules:
  - There are no death events in the population.
  - There are birth events in the population: the probability that a birth takes place in Δt is $b\Delta t$.
  - The probability of more than one birth in a time interval Δt is negligible (no twin births allowed).

Modelling the time evolution of a population

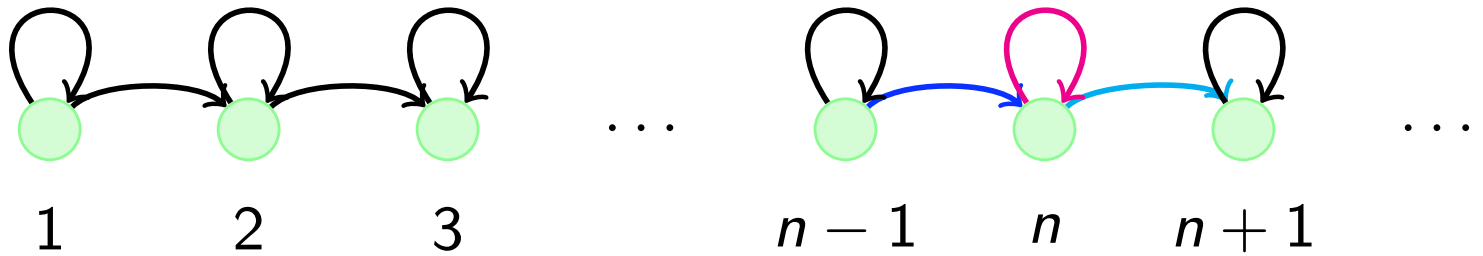
Stochastic birth process



- The probability that a population of size $n-1$ increases to n in the time interval $(t, t+\Delta t)$ is $(n-1)b\Delta t$.
- The probability that a population of size n increases to $n+1$ in the time interval $(t, t+\Delta t)$ is $b\Delta tn$.
- If at time t the population has n individuals, the probability that no birth event takes place in the time interval $(t, t+\Delta t)$ is $1 - b\Delta tn$.

Modelling the time evolution of a population

Stochastic birth process

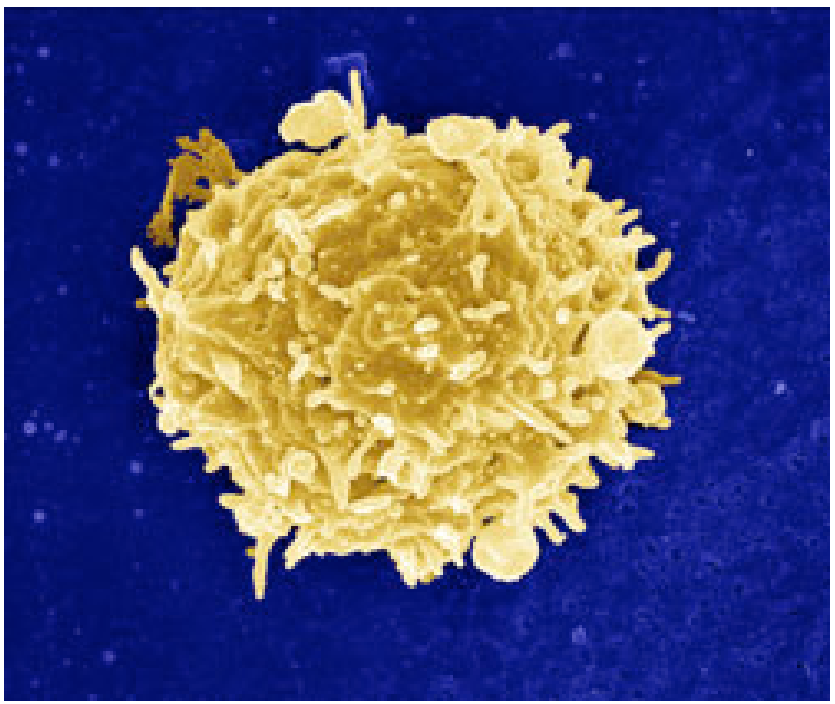


- The probability that a population of size $n-1$ increases to n in the time interval $(t, t+\Delta t)$ is $(n-1)b\Delta t$.
- The probability that a population of size n increases to $n+1$ in the time interval $(t, t+\Delta t)$ is $b\Delta tn$.
- If at time t the population has n individuals, the probability that no birth event takes place in the time interval $(t, t+\Delta t)$ is $1 - b\Delta tn$.
- Evolution equation for $p_n(t)$:

$$p_n(t+\Delta t) = (n-1)b\Delta t p_{n-1}(t) + (1 - nb\Delta t)p_n(t)$$

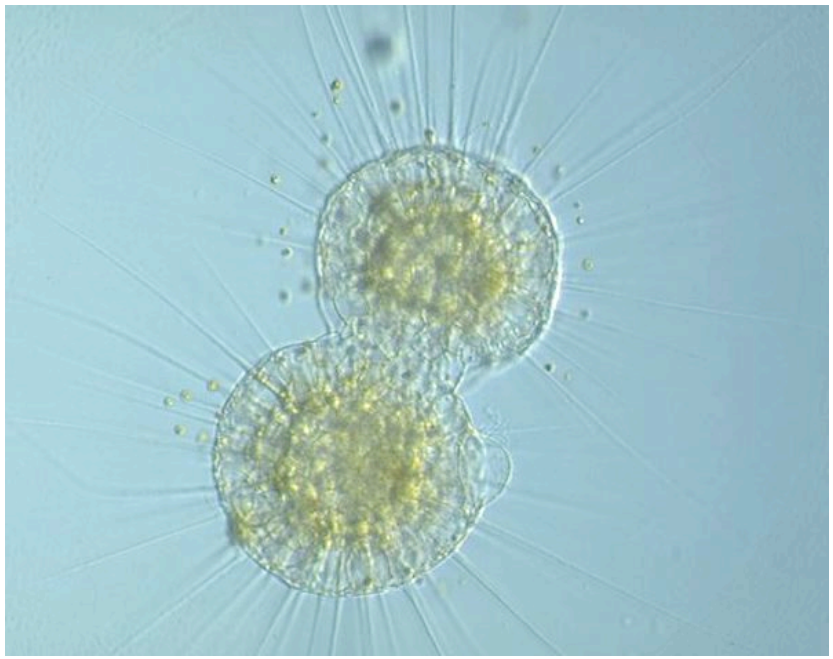
Modelling the time evolution of a population

T cell



Modelling the time evolution of a population

Cell division

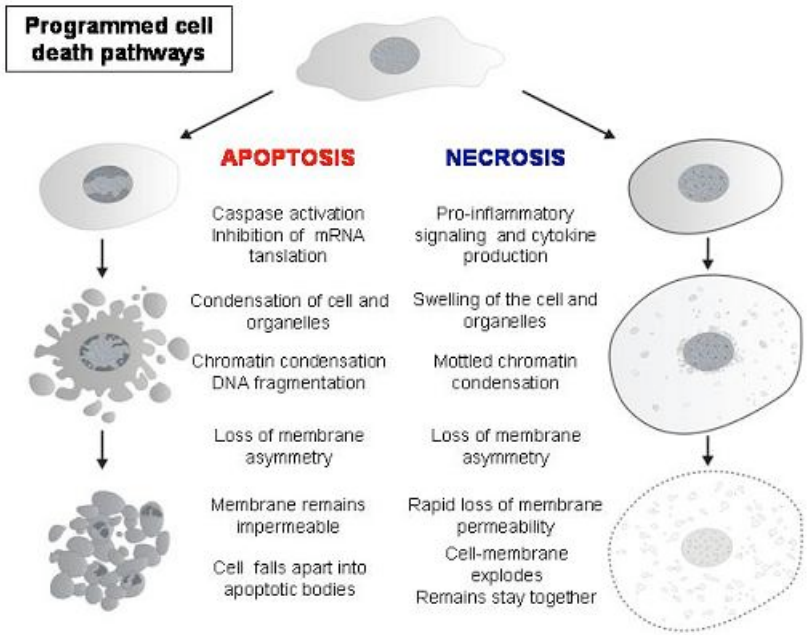


Birth event:

- At time t there are n cells.
- During the time interval Δt there is a single birth event.
- At time $t + \Delta t$ there are $n + 1$ cells.

Modelling the time evolution of a population

Cell death

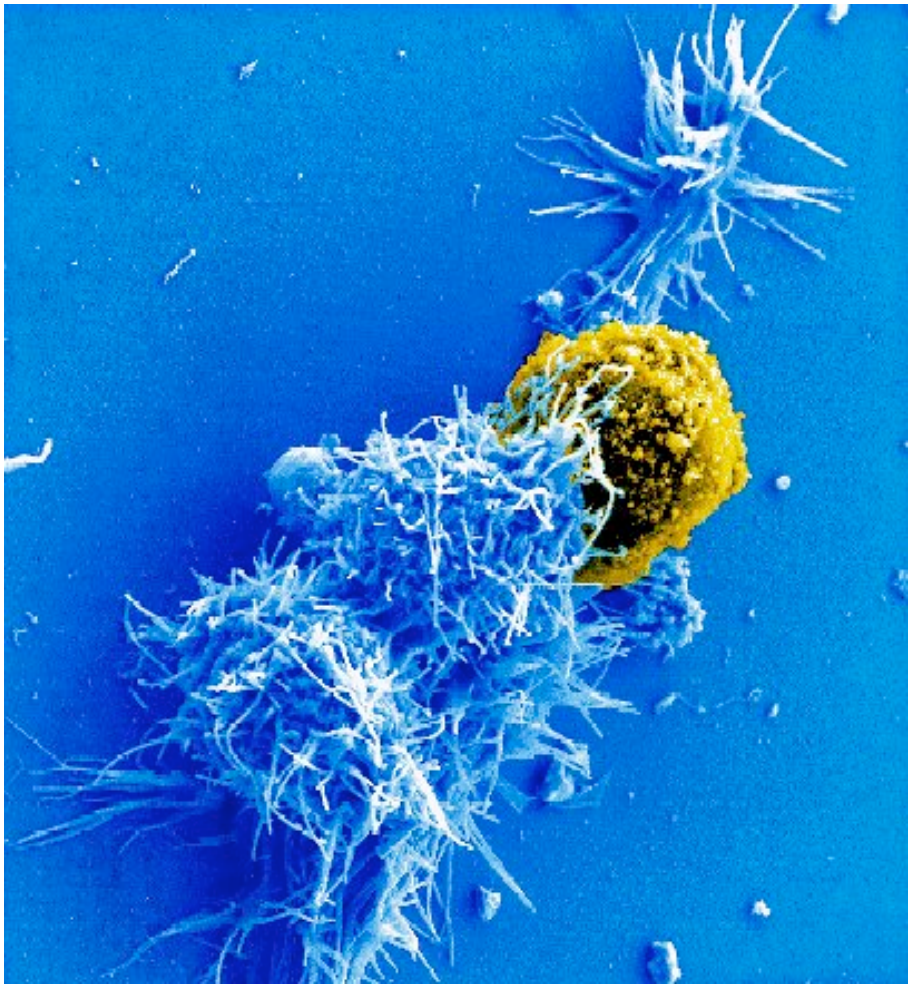


Death event:

- At time t there are n cells.
- During the time interval Δt there is a single death event.
- At time $t + \Delta t$ there are $n - 1$ cells.

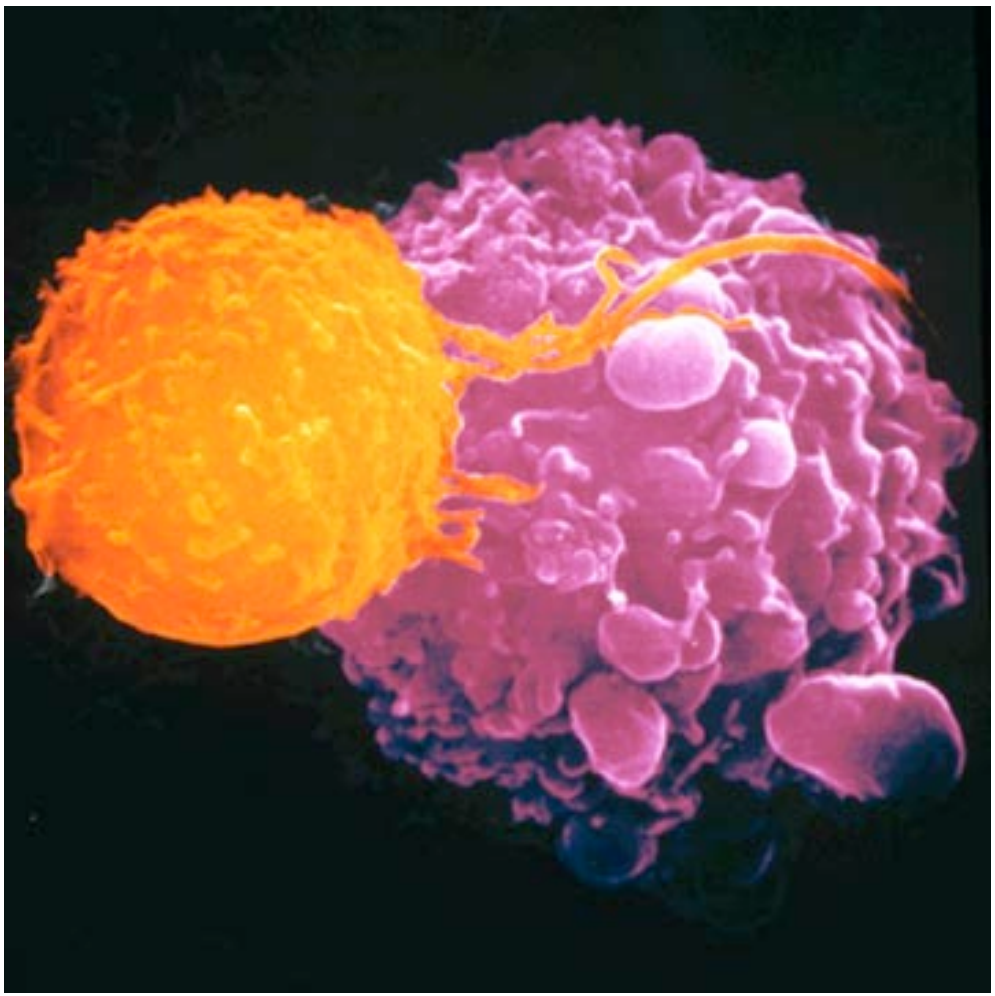
Modelling the time evolution of a population

Cell-cell interactions lead to events



Modelling the time evolution of a population

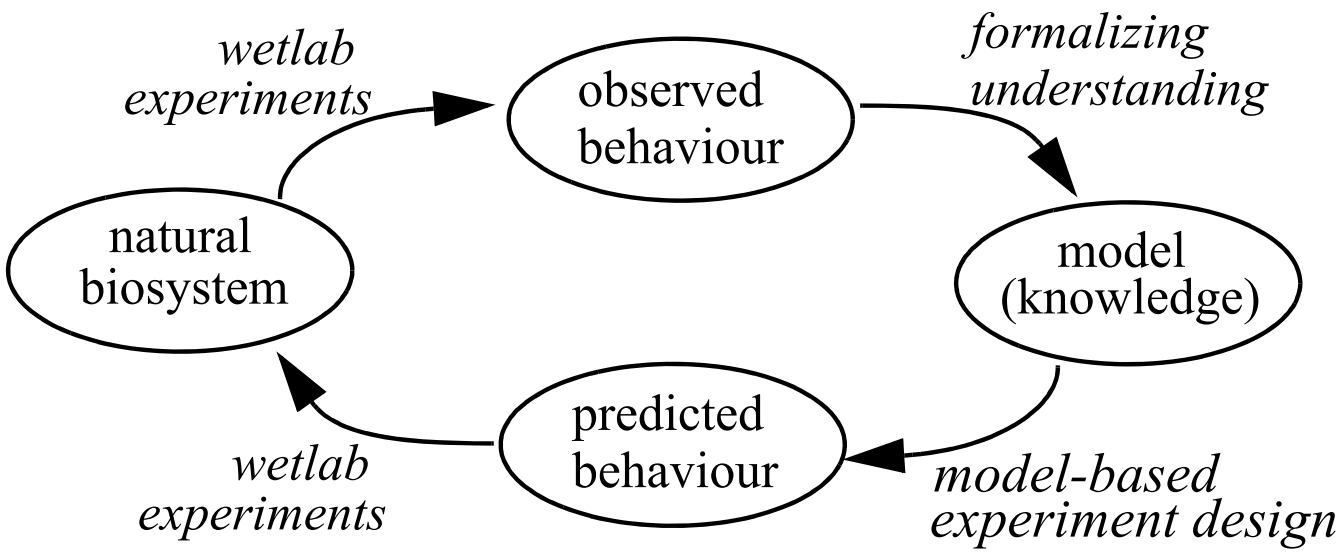
T cell and Tumour cell



Outline

Modelling in systems biology

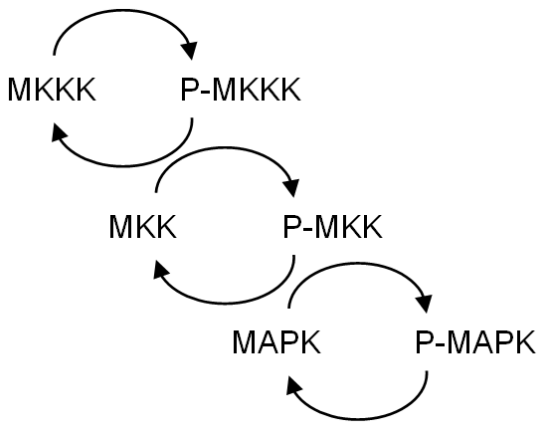
Modelling: Design and construction of models of existing biological systems, which explain observed properties and predict the response to experimental interventions.



Modelling in systems biology

Model:

- Formal representation of the real world.
- Simplified abstract view of the complex reality



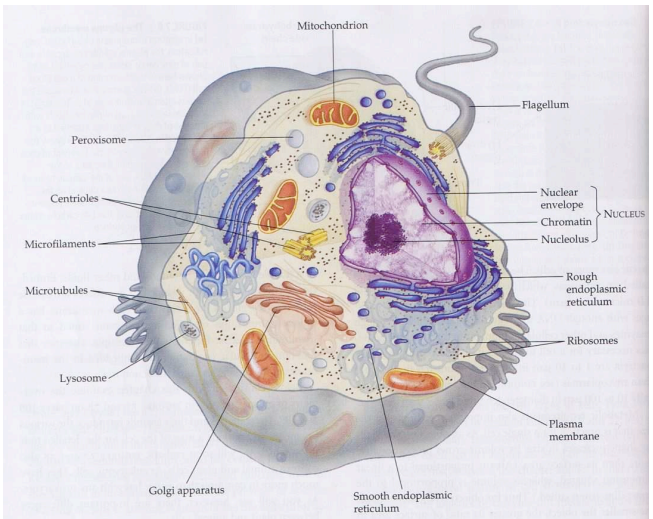
Modelling in systems biology

Why model?

- A model can generate new insights
- A model can make testable predictions
  - E.g., predict the effect of drugs on an organism
  - E.g., predict the effect on an inhibitor on a pathway
- A model can test conditions that may be difficult to study in the laboratory
- A model can rule out particular explanations for an experimental observation
- A model can help you identify what's right and wrong with your hypotheses

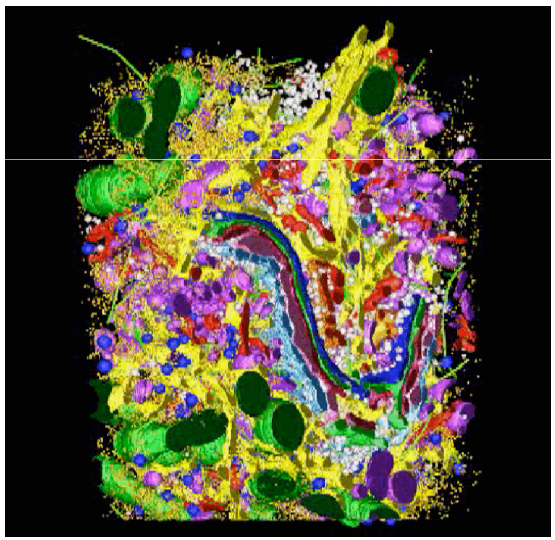
Modelling in systems biology

Textbook view of the cell and reality



Campbell, Reece & Mitchell (1998) *Biology*, 5th Edition

Three D EM image of a pancreatic Beta cell



Modelling in systems biology

In silico humans - spatial & temporal scales

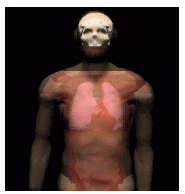
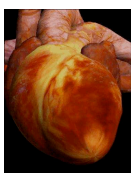
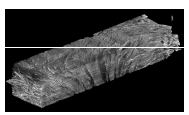
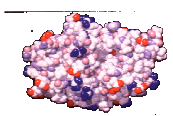
- **1 m** person
- **1 mm** electrical length scale of cardiac tissue
- **1 mm** cardiac sarcomere spacing
- **1 nm** pore diameter in a membrane protein

Range = 10^9

- **10^9 s (70 yrs)** human lifetime
- **10^6 s (10 days)** protein turnover
- **10^3 s (1 hour)** digest food
- **1 s** heart beat
- **1 ms** ion channel HH gating
- **1 ms** Brownian motion

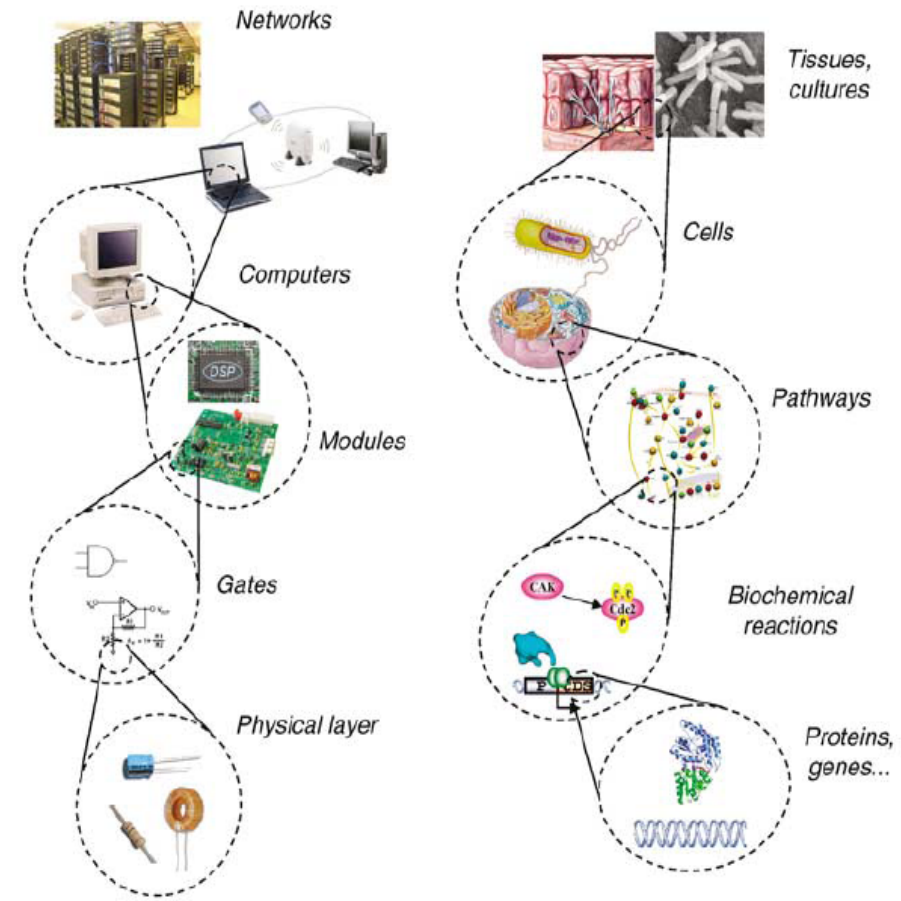
Range = 10^{15}

Requires a hierarchy of inter-related models


Modelling in systems biology

Levels of abstraction



Adrianantoandro et al. Mol Sys Bio 2006

Systems

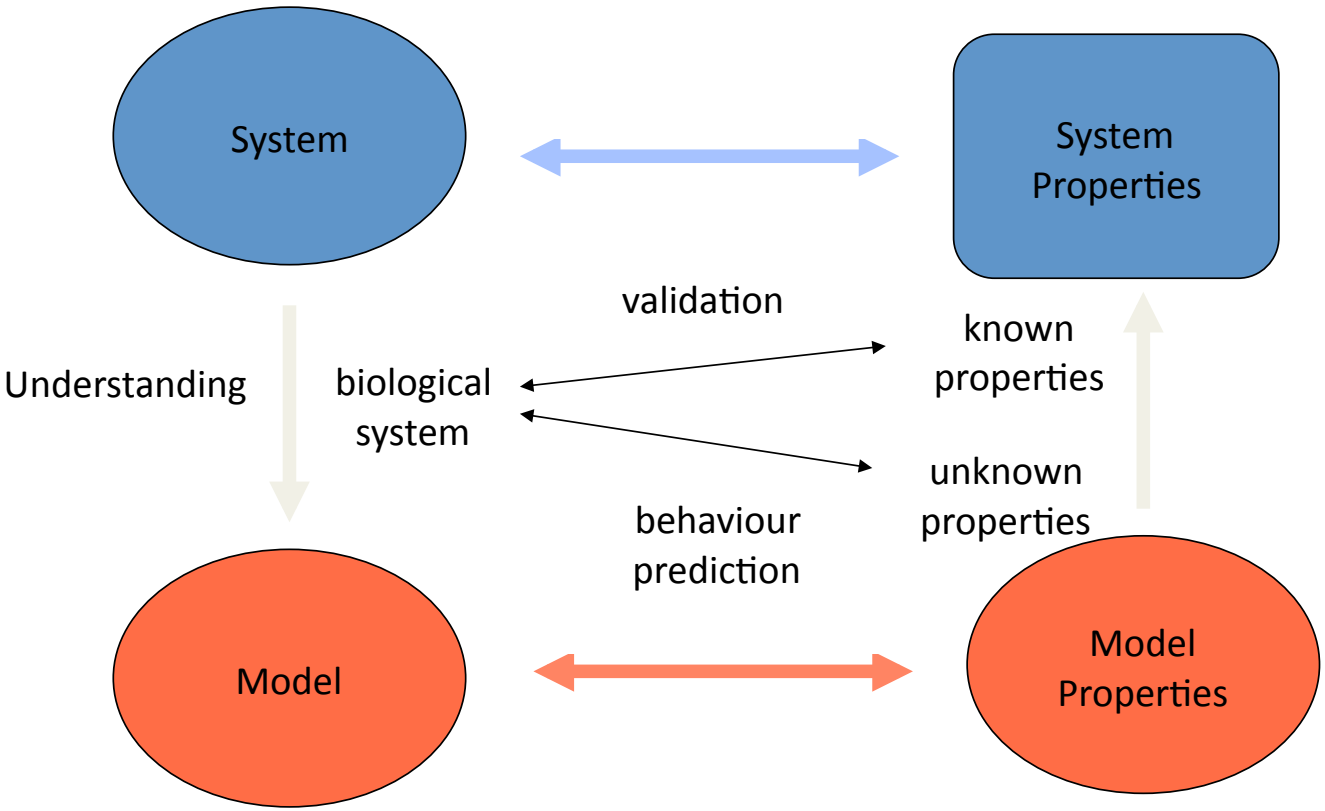
Devices

Parts

DNA

Modelling in systems biology

Models must be validated by experimental data: Simulations must be accurate representations of the real world.



Formal Models in Bioinformatics

A Framework for Modelling

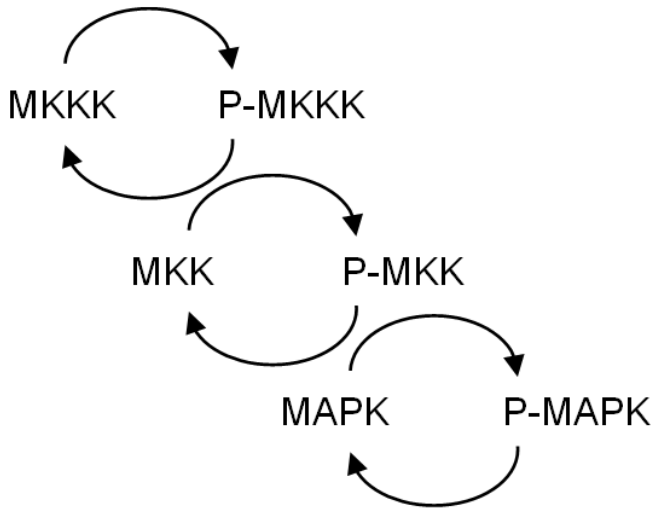
- Define all the components of the system
- Systematically perturb and monitor components of the system
- Reconcile the experimentally observed responses with those predicted by the model
- Design and perform new perturbation experiments to distinguish between multiple or competing model hypotheses.

(Ideker, Galitski & Hood, 2001)

Modelling in systems biology

Models should be:

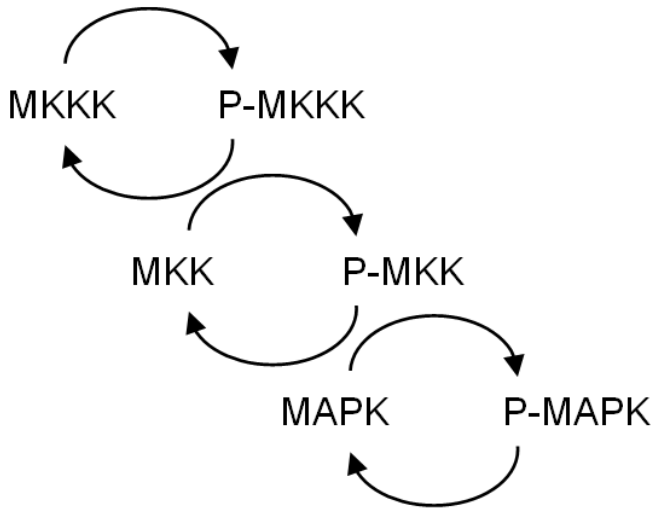
- Readable.
- Unambiguous.
- Analysable.
- Executable.



Modelling in systems biology

Models should be:

- Readable.
- Unambiguous.
- Analysable.
- Executable.

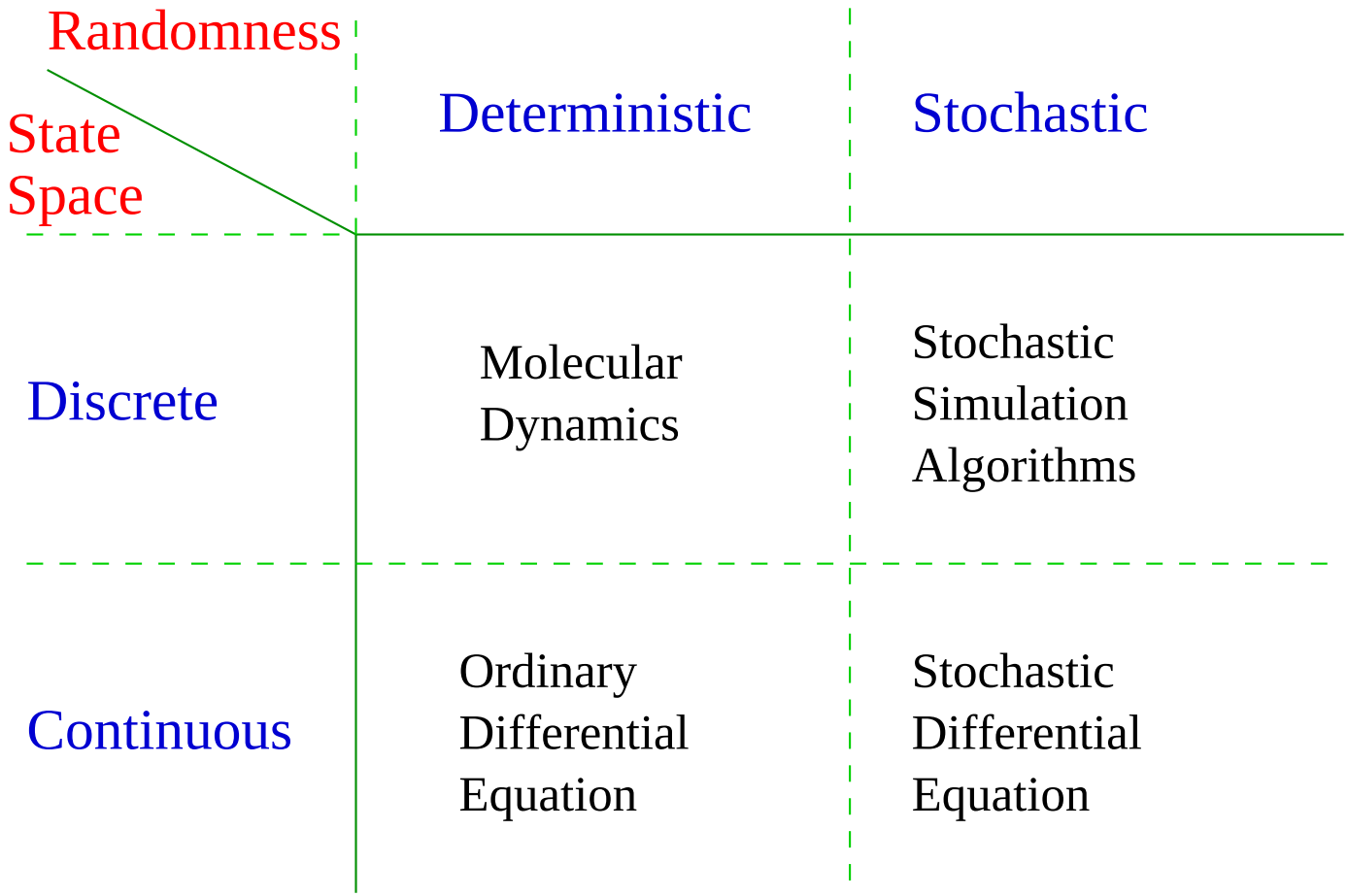


Suggestions:

- Occam's razor: Don't overcomplicate things.
- Einstein: Everything should be made as simple as possible, but not simpler.

Modelling in systems biology

Modelling Regimes



Modelling in systems biology

Modelling Regimes

- Discrete and stochastic: Small numbers of molecules. Exact description via Stochastic Simulation Algorithm (SSA) - Gillespie. Large computational time.

Modelling in systems biology

Modelling Regimes

- Discrete and stochastic: Small numbers of molecules. Exact description via Stochastic Simulation Algorithm (SSA) - Gillespie. Large computational time.
- Continuous and stochastic: A bridge connecting discrete and continuous models. Described by SDEs Chemical Langevin Equation.

Modelling in systems biology

Modelling Regimes

- Discrete and stochastic: Small numbers of molecules. Exact description via Stochastic Simulation Algorithm (SSA) - Gillespie. Large computational time.
- Continuous and stochastic: A bridge connecting discrete and continuous models. Described by SDEs Chemical Langevin Equation.
- Continuous and deterministic: Law of Mass Action. The Reaction Rate equations. Described by ordinary differential equations. Not valid if molecular populations of some critical reactant species are small.

Outline

Modelling in systems biology

Biological evidence of noise

- “Stochasticity is evident in all biological processes the proliferation of both noise and noise reduction systems is a hallmark of organismal evolution” Federoff et al.(2002).
- “Transcription in higher eukaryotes occurs with a relatively low frequency in biologic time and is regulated in a probabilistic manner” Hume (2000).
- “Gene regulation is a noisy business” Mcadams et al. (1999).
- “Initiation of gene transcription is a discrete process in which individual protein-coding genes in an off state can be stochastically switched on, resulting in sporadic pulses of mRNA production” Sano 2001.
- “It is essential to study individual cells and to measure the cell to cell variations in biological response, rather than averaging over cell populations” Zatorsky, Rosenfeld et al. 2006.

Modelling in systems biology

Origin of Stochasticity

- Intrinsic noise due to small numbers of molecules (e.g. mRNA, DNA loci, TFs).
- Uncertainty of knowing when a reaction occurs and which reaction it is.
- Relative statistical uncertainty is inversely proportional to the square root of the number of molecules.
- Applies equally well to studying channel behaviour via the concept of channel molecules.
- Extrinsic noise due to (external) environmental effects (extrinsic factors are: stage in cell cycle, number of RNAP or ribosomes, cellular environment).

Modelling formalisms

Markov chains

- Based on the concept of state of the system
- Solution techniques:
  - Enumerative
  - Transient and steady-state analysis
  - Exact and approximate analysis
- Drawbacks:
  - Low abstraction level
  - Model size equals number of states of the system
  - Only in very particular cases aggregation techniques exist

Modelling formalisms

Queueing networks

- High abstraction level
  - The number of states characterizing the system grows exponentially on the model size.
- Solution techniques:
  - Enumerative (based on Markov chains)
  - Reduction/transformation-based
  - Structurally based (product-form solution, exact)
  - Transient and steady-state analysis
  - Exact, approximate and bounds
- Drawbacks:
  - Lack of synchronization primitive
  - Extensions exist but destroying analysis possibilities

Modelling formalisms

Stochastic Petri nets

- Abstraction level similar to queueing networks
- With synchronization primitive
  - SPN =Petri nets+ timing interpretation=queueing networks+ synchronizations
- Wide range of qualitative (logical properties) analysis techniques:
  - Enumerative (based on Markov chains)
  - Reduction/transformation-based
  - Structurally based
- Petri nets as a formal modelling paradigm
  - a conceptual framework to obtain specific formalisms based on common concepts and principles at different life-cycle phases

Modelling formalisms

Stochastic Petri nets (cont.)

- Analysis techniques:
  - Exact: mainly enumerative (based on Markov chains)
  - Bounding techniques (structurally based)
  - Approximation techniques (reduction/transformation)
- Drawbacks:
  - Lack of a product-form solution for efficient exact analysis in most cases

Contents of the Course

1. Discrete and Continuous Markov chains
2. Birth and Death Processes
3. Stochastic Simulation
4. Hidden Markov Chains
5. Stochastic Petri nets

Course info:

`http://webdiis.unizar.es/asignaturas/SPN/`

Acknowledgments

Much of the material in the course is based on the following courses:

- Deterministic models in mathematical biology, Magic 042 - Lecture 1
  Carmen Molina-París, Department of Applied Mathematics, School of Mathematics, University of Leeds
- Una Introducción a la Biología de Sistemas
  Raúl Guantes (UAM), Juan F. Poyatos (CNB)
- A conceptual framework for BioModel Engineering (Systems Biology, Synthetic Biology)
  Rainer Breitling, Groningen, NL; David Gilbert, Brunel, UK; Monika Heiner, Cottbus, DE
- A Petri Net Perspective on Systems and Synthetic Biology
  Monika Heiner, Brandenburg University of Technology Cottbus, DE Dept. of CS
- Systems Biology: Stochastic models and Simulation
  Kevin Burrage, Institute for Molecular Bioscience, The University of Queensland, Australia.