

Modeling and Simulation in Computational Biology 1

BCB 5250 Introduction to Bioinformatics II

Spring 2020

Tae-Hyuk (Ted) Ahn

Department of Computer Science
Program of Bioinformatics and Computational Biology
Saint Louis University



SAINT LOUIS
UNIVERSITY™

— EST. 1818 —

Learning Outcome:

- Learn about basic concept of model and simulation in biology
- Understand the difference of the types of model
- Obtain the knowledge of mathematical modeling and simulation

Computer Simulations

Computer simulation is the process of making a computer behave the same as ...whatever it is we are interested in.

- Atoms
- Cooling metal alloy
- A society of voters
- Climate change
- Biological pathway
- A galaxy

Computer Simulations

Simulations have applications across a range of disciplines:

- Physics – solids, gases, fluids, solar systems
- Chemistry – molecular dynamics
- Biology – gene networks, predator-prey populations
- Sociology – socio networks, opinion propagation
- Technology – internet traffic, local networks
- Management – queuing, workflow models
- Finance & Economics – stock markets, supply-demand

Computer Simulations

To create a computer simulation to approximate a system, a **model** of that system must first be made. These are most often **mathematical models**.

"A model is a description of some system intended to predict what happens if certain actions are taken"

– Bratley, Bennet & Schrage (1987)

Models

- Modeling is a large discipline in itself and creating a system requires a lot of mathematical ability and understanding of the system.
- Models are usually composed of **variables** and **relationships** between them. Exactly what these variables represent and what the relationships between them are can vary.

Models

The variables of the model must represent the *state* of the system. The state is split into different components to represent the different parts of the system. These are sometimes called *model components*.

For example:

- A car in a traffic simulator may have a *position*, a *size* and a *velocity*.

Models

The **relationships** between these model components define the behavior of the system.
Going back to our previous example some rules may be:

- The position of the car changes based on the velocity.
- If the distance to the car in front is less than X, decelerate.

Models

The types of relationships depend on the type of model. Some categories of model include:

- Linear vs Nonlinear
- Static vs Dynamic
- Explicit vs Implicit
- Discrete vs Continuous
- Deterministic vs Probabilistic

Models

Models are also limited in the accuracy with which they describe the model. The usefulness of a model depends on a number of factors

- model validity
- level of simplification
- credibility
- tractability

Models

When using computer simulations, it is important to understand the limitations of the model you are using. A simulation (no matter how accurate) cannot provide useful results if the model is not suitable for the system you are studying.

A model is considered **valid** if the system it describes *sufficiently near* to the real system.

Simulations

Models are approximations of a real system (for the most part). Simulations approximate the behavior of the system described by the model.

- This type of model describes a system that continuously changes with time. Computers don't do things continuously so we resort to using a time-step to jumping forward in time repeatedly.
- This is an approximation of the model because we are simulating a continuous system with discrete time-steps.

Continuous Simulations

Continuous simulations are simulations that compute models that change continuously (usually over time). This type of simulation is extremely common in the physical sciences.

- Continuous simulations are most often based on models described by *ordinary differential equations (ODEs)* or *partial differential equations (PDEs)*.
- Because computers are limited to discrete calculations, these simulations update the system with discrete time-steps.

Continuous Simulations

To do this, the continuous model must be integrated over that time-step to calculate the total change in each of the variables representing our model.

- There are a number of different numerical methods that can be used for this purpose.

Discrete Event Simulations

The other main category of simulations are discrete event simulations. This type of simulation is no longer computing a continuously changing system but one that considers discrete events.

Biological Systems

- **What is a biological system?**

- It is a complex system consisting of very many simple and identical elements interacting to produce what appears to be complex behavior

- **Examples:**

- Macroscopic: circulatory system, digestive system, reproductive system
- Microscopic: cells, organelles, regulatory pathways

Importance

Applying mathematics to biology has a long history, but only recently has there been an explosion of interest in the field. Some reasons for this include:

- the explosion of data-rich information sets, due to the genomics revolution, which are difficult to understand without the use of analytical tools,
- recent development of mathematical tools such as chaos theory to help understand complex, nonlinear mechanisms in biology,
- an increase in computing power which enables calculations and simulations to be performed that were not previously possible, and
- an increasing interest in in silico experimentation due to the complications involved in human and animal research.

Why to Model Biological Systems

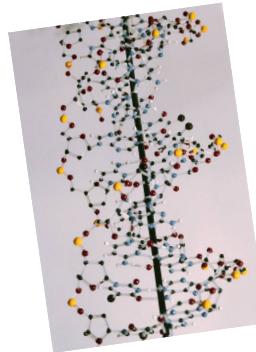
- To help reveal possible underlying mechanisms involved in a biological process
- To help interpret and reveal contradictions/incompleteness of data
- To help confirm/reject hypotheses
- To predict system performance under untested conditions
- To supply information about the values of experimentally inaccessible parameters
- To suggest new hypotheses and stimulate new experiments

Limitations of Mathematical Models

- Not necessarily a ‘correct’ model
- Unrealistic models may fit data very well leading to incorrect conclusions
- Simple models are easy to manage, but complexity is often required
- Realistic simulations require a large number of hard to obtain parameters
- Models are not explanations and can never alone provide a complete solution to a biological problem.

Molecular Biology vs. Systems Biology

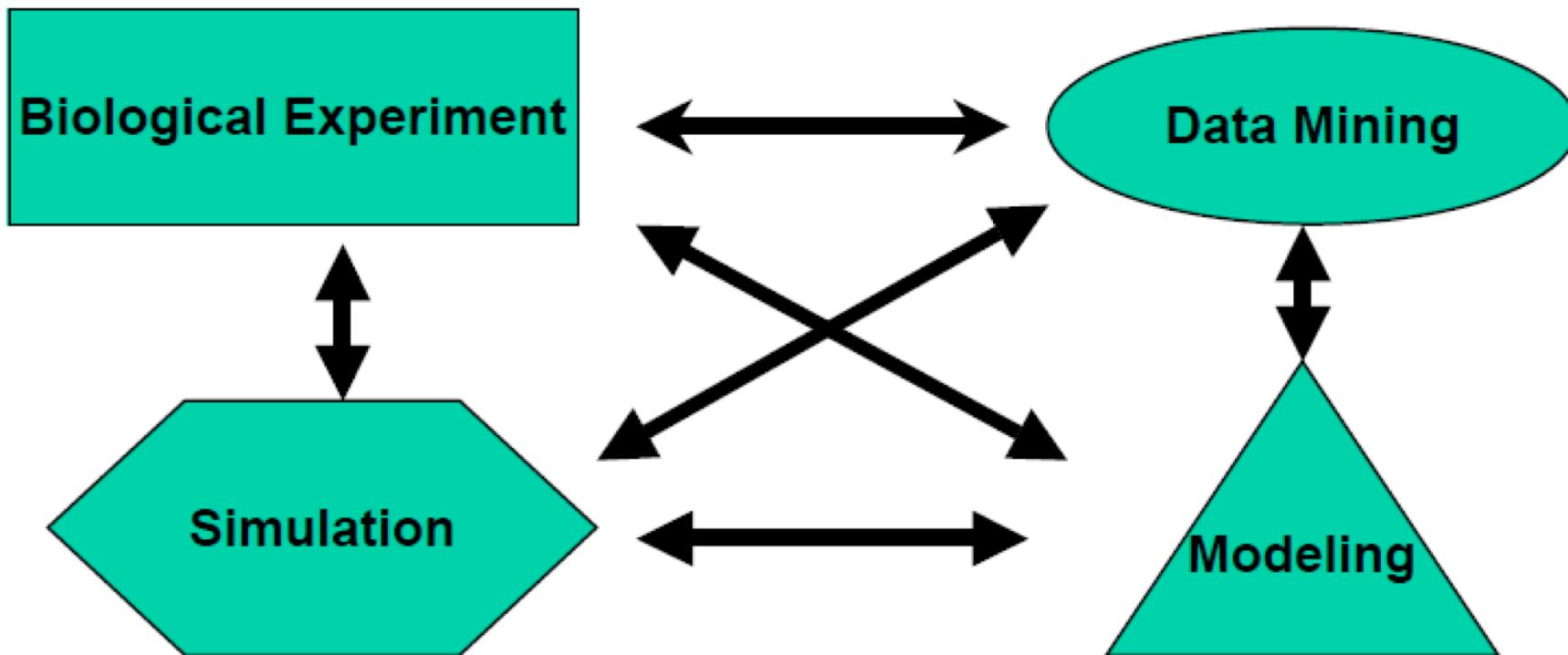
- In **molecular biology**, gene structure and function is studied at the molecular level.
- In **systems biology**, specific interactions of components in the biological system are studied – cells, tissues, organs, and ecological webs.



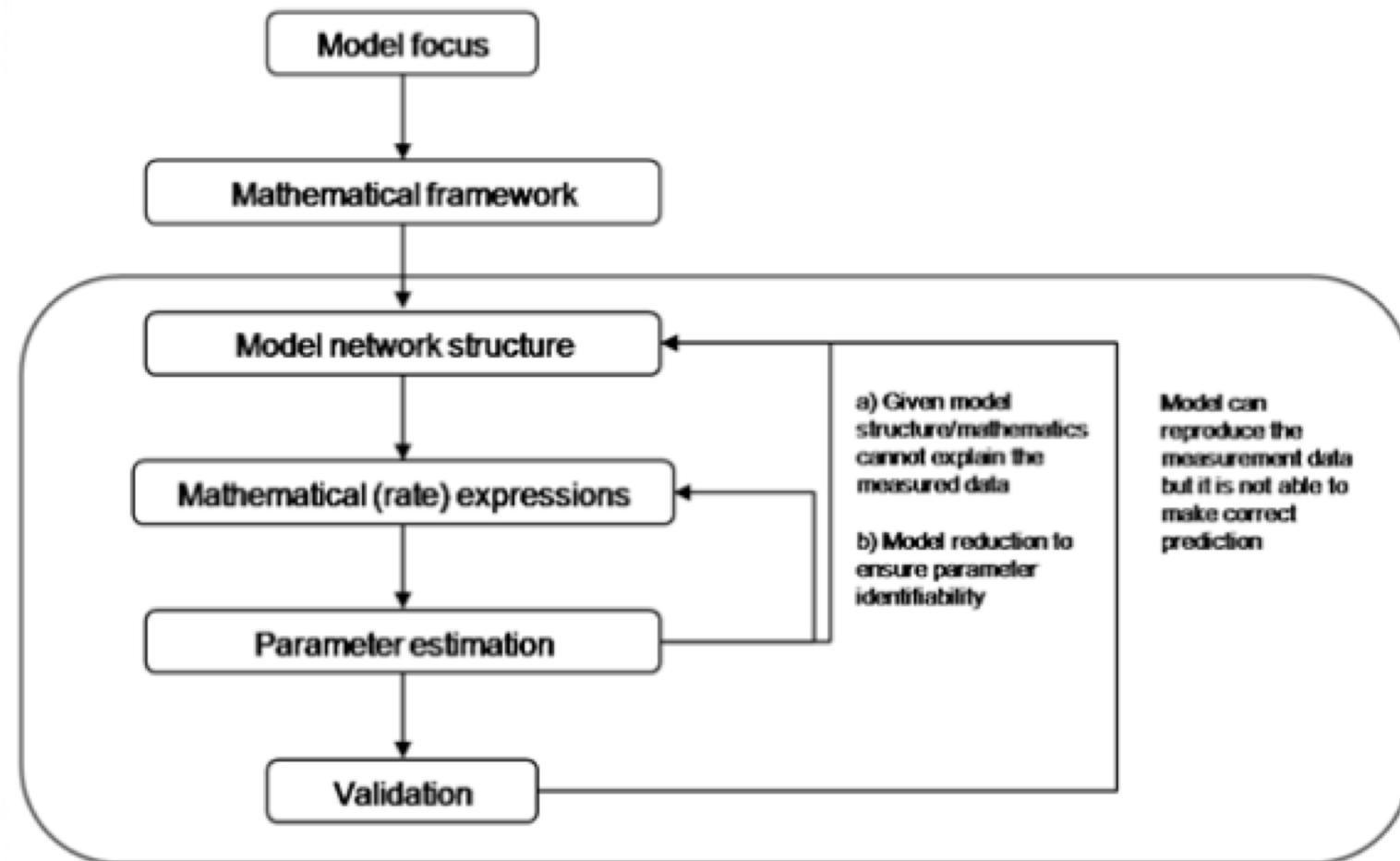
Modeling Biological Systems

- Modeling is from Simple Structures to Complex Systems.
- Modeling of Biological Systems is evolving into an important partner of experimental work.
- Biological Systems are complex, thus, a combination of experimental and computational approaches are needed.
- All facets of biology, environmental, organismic, cellular and molecular biology are becoming more accessible to chemical, physical and mathematical approaches.

Computational Biology



Modeling Process



Systematic modeling approach – an overview

Avi Ericsson, 2007

Modeling Decision

- **Resolution**
 - What should be the resolution of the model?
 - **What are our model variables representing?** Given the recent development of experimental techniques resulting in quantitative molecular data it is now possible to create models describing molecular contents (numbers or concentrations) and compare directly with experiments.
 - Let us look at dynamical models of molecular contents describing biochemical networks within single cells and briefly extend the approach to multicellular systems.

Modeling Decision

- **Continuous vs. Discrete?**

- Molecules are individual objects, and the quantitative measure of molecular content is in principle number of molecules.
- On the other hand, the number of molecules (of the same type) within for example a cell is often large and a continuous variable for the concentration (number of variables per unit volume) is then applicable to describe the system behavior.
- Continuous modeling methods was used in most biochemical pathway modeling.
- The limit for when the concentration is sufficient to describe a system depends on the details of the system but typically when the number of molecules are more than 10–100 it is safe to use concentration as a measure of molecular content.

Modeling Decision

- **Deterministic vs. Stochastic?**

- This point is somewhat related to the previous.
- In principle there is a probability connected to an individual reaction to occur.
- This can be taken into account using a stochastic update of the system variables (reactions happen with a specific probability).
- In deterministic modeling, we assume systems with a large number of molecules where it is applicable to use a deterministic description of the system update.

Deterministic processes (dynamical systems)

- A fixed mapping between an initial state and a final state. Starting from an initial condition and moving forward in time, a deterministic process will always generate the same trajectory and no two trajectories cross in state space.
- Ordinary differential equations (Continuous time. Continuous state space. No spatial derivatives.)
- Partial differential equations (Continuous time. Continuous state space. Spatial derivatives.)
- Maps (Discrete time. Continuous state space)

Stochastic processes (random dynamical systems)

- A random mapping between an initial state and a final state, making the state of the system a random variable with a corresponding probability distribution.
- Non-Markovian processes -- Generalized master equation (Continuous time with memory of past events. Discrete state space. Waiting times of events (or transitions between states) discretely occur and have a generalized probability distribution.)
- Jump Markov process -- Master equation (Continuous time with no memory of past events. Discrete state space. Waiting times between events discretely occur and are exponentially distributed.) See also Monte Carlo method for numerical simulation methods, specifically Continuous-time Monte Carlo which is also called kinetic Monte Carlo or the stochastic simulation algorithm.
- Continuous Markov process -- stochastic differential equations or a Fokker-Planck equation (Continuous time. Continuous state space. Events occur continuously according to a random Wiener process.)

Constructing the network

Using what we know about our system to construct a biochemical network.

- E.g what are the entities involved? in what processes? Where does it all take place?
Etc.

Components in Biochemical Networks

Biochemical *entities (species)*:

- Metabolites
- Proteins (enzymes, transporters, building-blocks ...)
- Lipids
- Nucleic Acids (DNA, RNA, ...)

Components in Biochemical Networks

Quantification:

- Amount (1 mol, 1 mmol, 1 nmol, ...)
- Concentration ($1 \text{ mol/m}^3 = 1 \text{ mmol/dm}^3 = 1 \text{ M}$, ...)
- Activity (% , ...)

Components in Biochemical Networks

The biochemical entities are involved in different processes :

- Reactions (catalysis, transfer, binding, assembly/disassembly)
- Translocations (e.g., transport – active/passive)
- Expression

Components in Biochemical Networks

The processes themselves might be influenced by other players:

- Enzymes catalyze reactions (activator)
- Transcription factors facilitates expression (activator)
- Metabolites may inhibit reaction rates (inhibitor)

Components in Biochemical Networks

The processes take place somewhere:

- Reactions of entities within a compartment
- Transport of an entity over a membrane (between compartments)
- Recruitment of proteins to a membrane
- Dimerization of receptors on a membrane

Components in Biochemical Networks

And more....

- Rate expression, parameter estimation and validations

Modeling Tools

- Graphical modeling tools
- Mathematical modeling tools

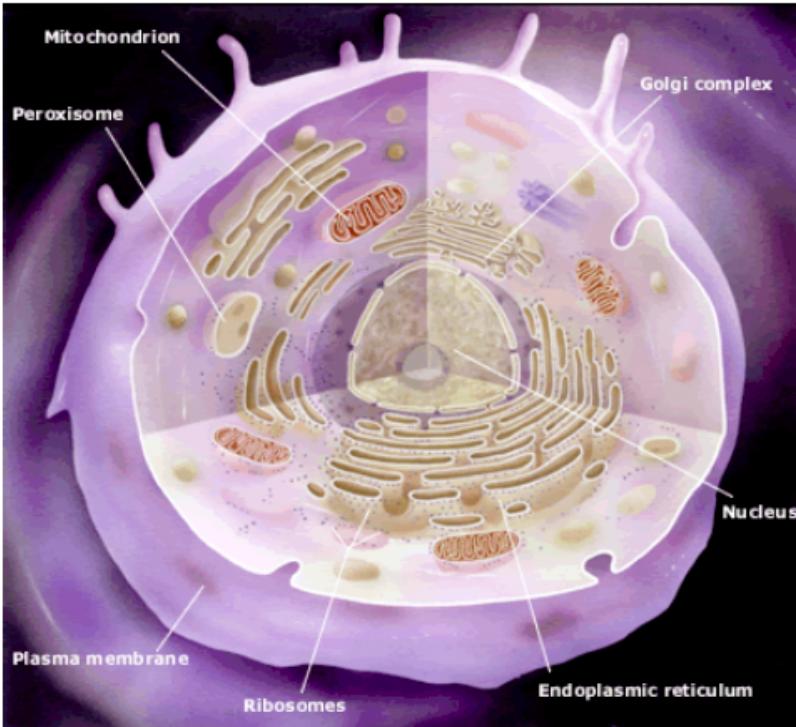
Standardization

- Computational, dynamical modeling is essential to achieving a deeper understanding of mechanisms underlying complex biological systems
- There is currently a proliferation of software tools
- Different packages have complementary strengths
- Model editing, simulation, analysis, display
- No single tool is able to do everything
- New techniques (or new tools) evolve all the time
- Researchers need to use more than one tool
- The field needs agreed-upon standards enabling tools to be used cooperatively

SBML

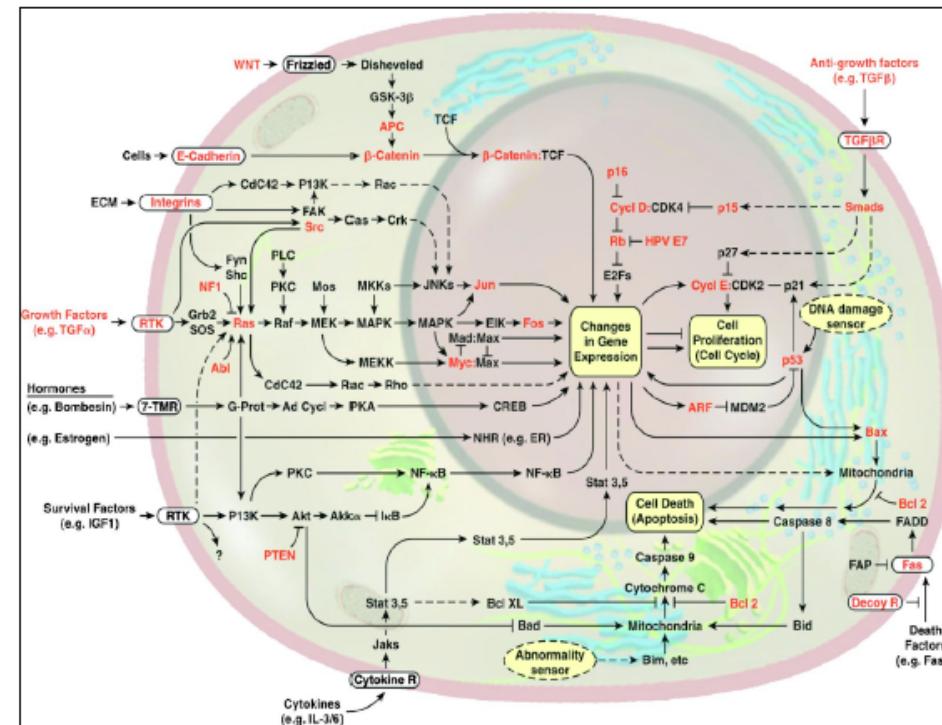
- A machine-readable format for representing computational models in systems biology
- Expressed in XML (Extensible Markup Language)
- Intended for software tools— not for humans
 - (Although it is text-based and therefore readable)
- Intended to be a tool-neutral exchange language for software applications in systems biology
- Simply an enabling technology

Modeling the Cell

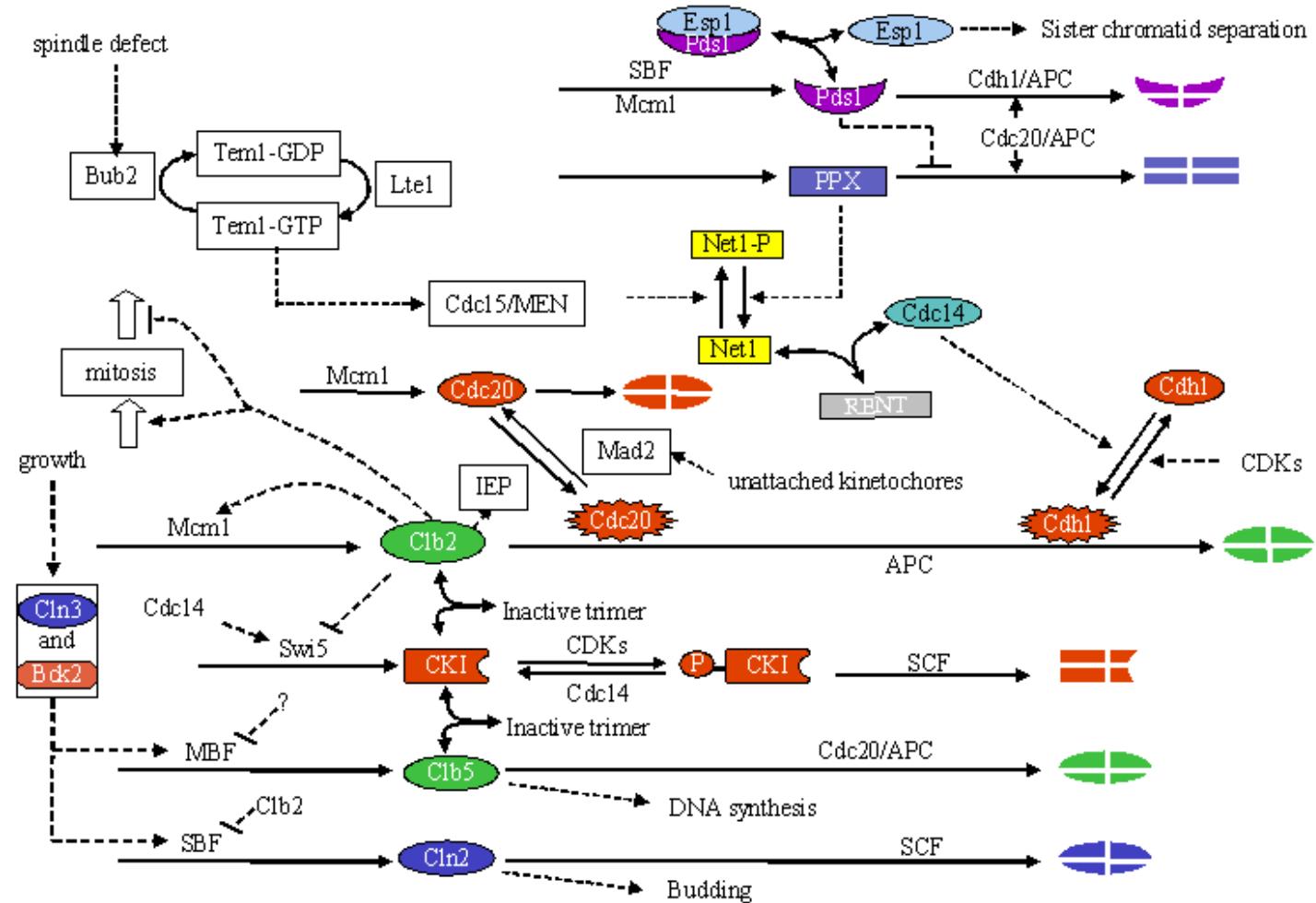


A nice picture of
Cell

From a modeler's
point of view



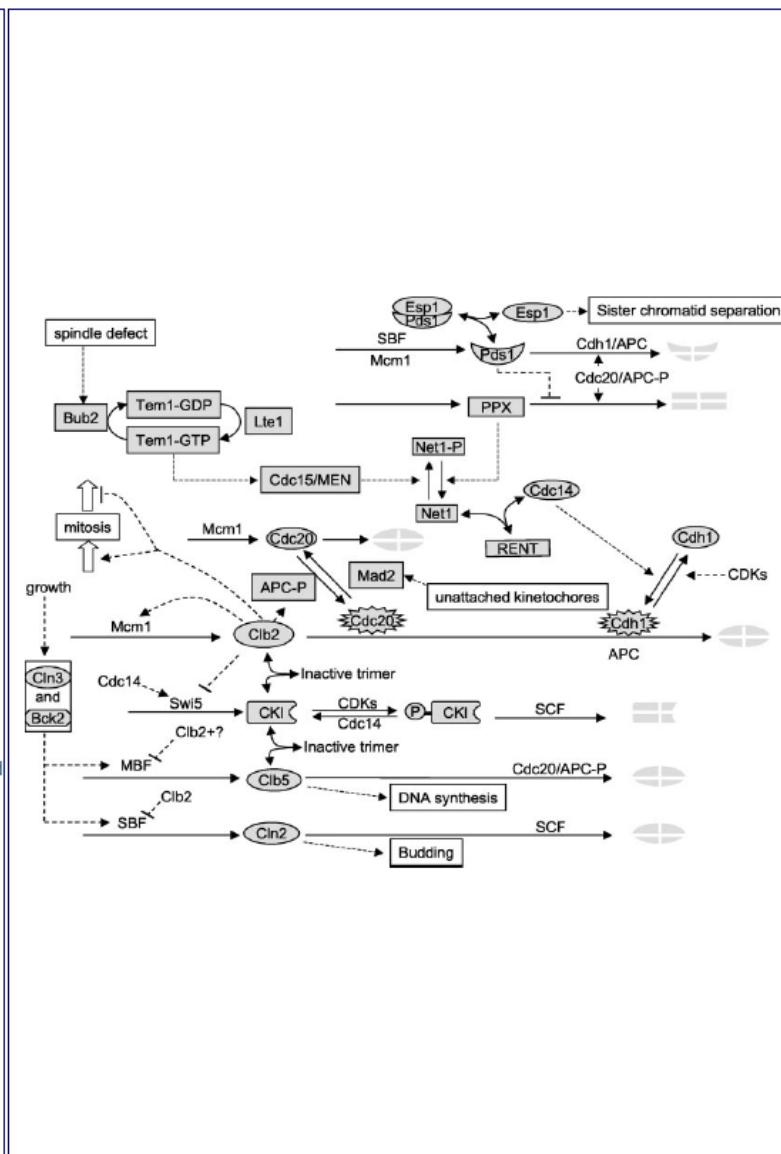
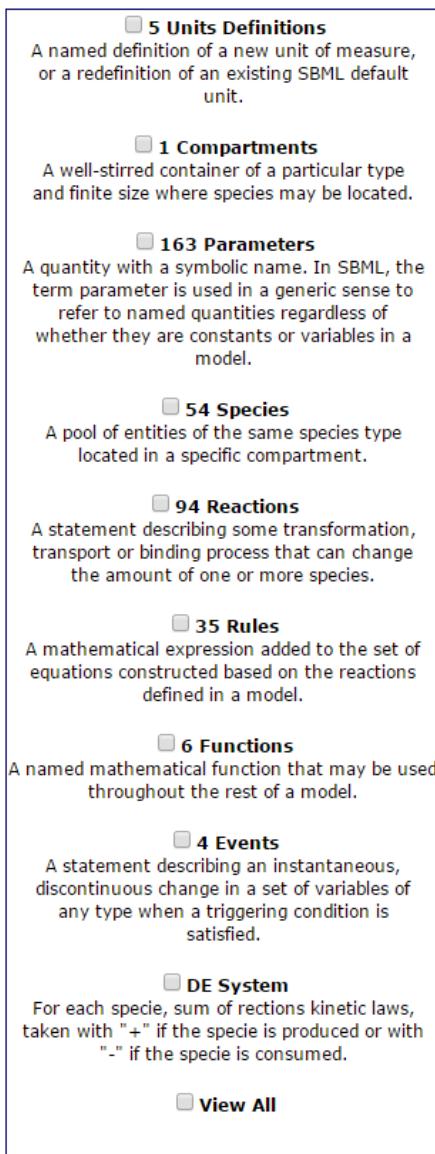
Budding Yeast Cell Cycle Model



The three major cell cycle events - DNA replication, mitosis and cell division, are controlled by **cyclin-dependent kinases (Cdks)**, which are active only when bound to a cyclin partner.

SBML (Systems Biology Markup Language)

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1">
<model id="Ec_iAF1260" name="Ec_iAF1260">
<listOfCompartments>
    <compartment id="in" outside="out"/>
    <compartment id="out" />
</listOfCompartments>
<listOfSpecies>
    <species id="A_out" name="A_out" boundaryCondition="false"/>
    <species id="A_in" name="A_in" boundaryCondition="false"/>
    <species id="B_in" name="B_in" boundaryCondition="false"/>
    <species id="B_out" name="B_out" boundaryCondition="false"/>
</listOfSpecies>
<listOfReactions>
    <reaction id="R1" name="R1" reversible="false">
        <listOfReactants>
            <speciesReference species="A_out" stoichiometry="1" 
                compartment="out"/>
        </listOfReactants>
        <listOfProducts>
            <speciesReference species="A_in" stoichiometry="1" 
                compartment="in"/>
        </listOfProducts>
    </reaction>
    <reaction id="R2" name="R2" reversible="false">
        <listOfReactants>
            <speciesReference species="A_in" stoichiometry="1" 
                compartment="in"/>
        </listOfReactants>
        <listOfProducts>
            <speciesReference species="B_in" stoichiometry="1" 
                compartment="in"/>
        </listOfProducts>
    </reaction>
    <reaction id="R3" name="R3" reversible="false">
        <listOfReactants>
            <speciesReference species="B_in" stoichiometry="1" 
                compartment="in"/>
        </listOfReactants>
        <listOfProducts>
            <speciesReference species="B_out" stoichiometry="1" 
                compartment="out"/>
        </listOfProducts>
    </reaction>
</listOfReactions>
</model>
</sbml>
```

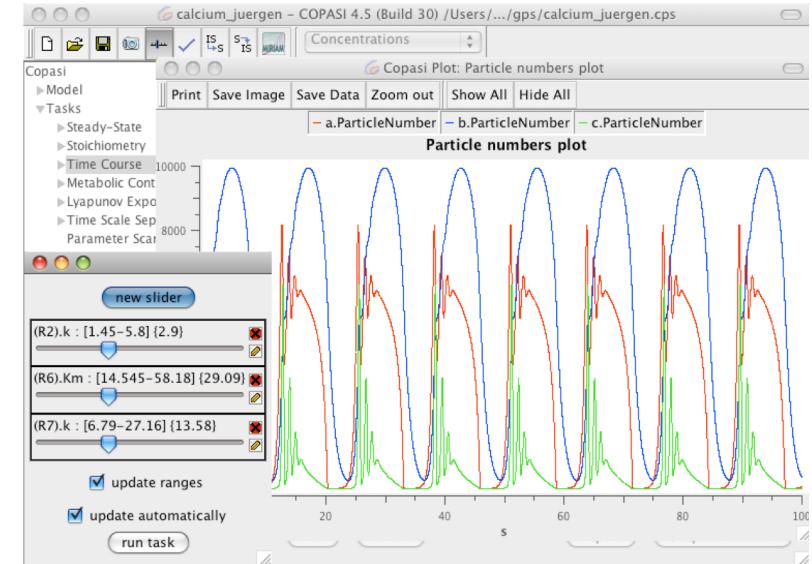


SBML (Systems Biology Markup Language)

- <http://sbml.org/>
- A free and open interchange format for computer models of biological processes.
- SBML is useful for models of metabolism, cell signaling, and more.
- It continues to be evolved and expanded by an international community.
- SBML Software Guide (http://sbml.org/SBML_Software_Guide)

COPASI

- <http://copasi.org/>
- Models in **COPASI** are **based** on **reactions** that convert a set of *species* into another set of *species*.
- Each *species* is located in a **compartment**, which is a physical location with a size (volume, area, etc).
- COPASI **automatically converts** the reaction network to a set of differential equations or to a system of stochastic reaction events.
- COPASI can import and export models in the **SBML** format.



Prepare Lab and Homework

- Install COPASI at your Laptop.
- Watch the COPASI Video Tutorials (http://copasi.org/Support/Video_Tutorials/). Especially, watch the first four videos (links are also provided in below).
 - <https://youtu.be/4pH16ema-Lg>
 - <https://youtu.be/Q3lumEn9204>
 - <https://youtu.be/BWUoRuYLJPQ>
 - https://youtu.be/qiYLMTsa-_Y