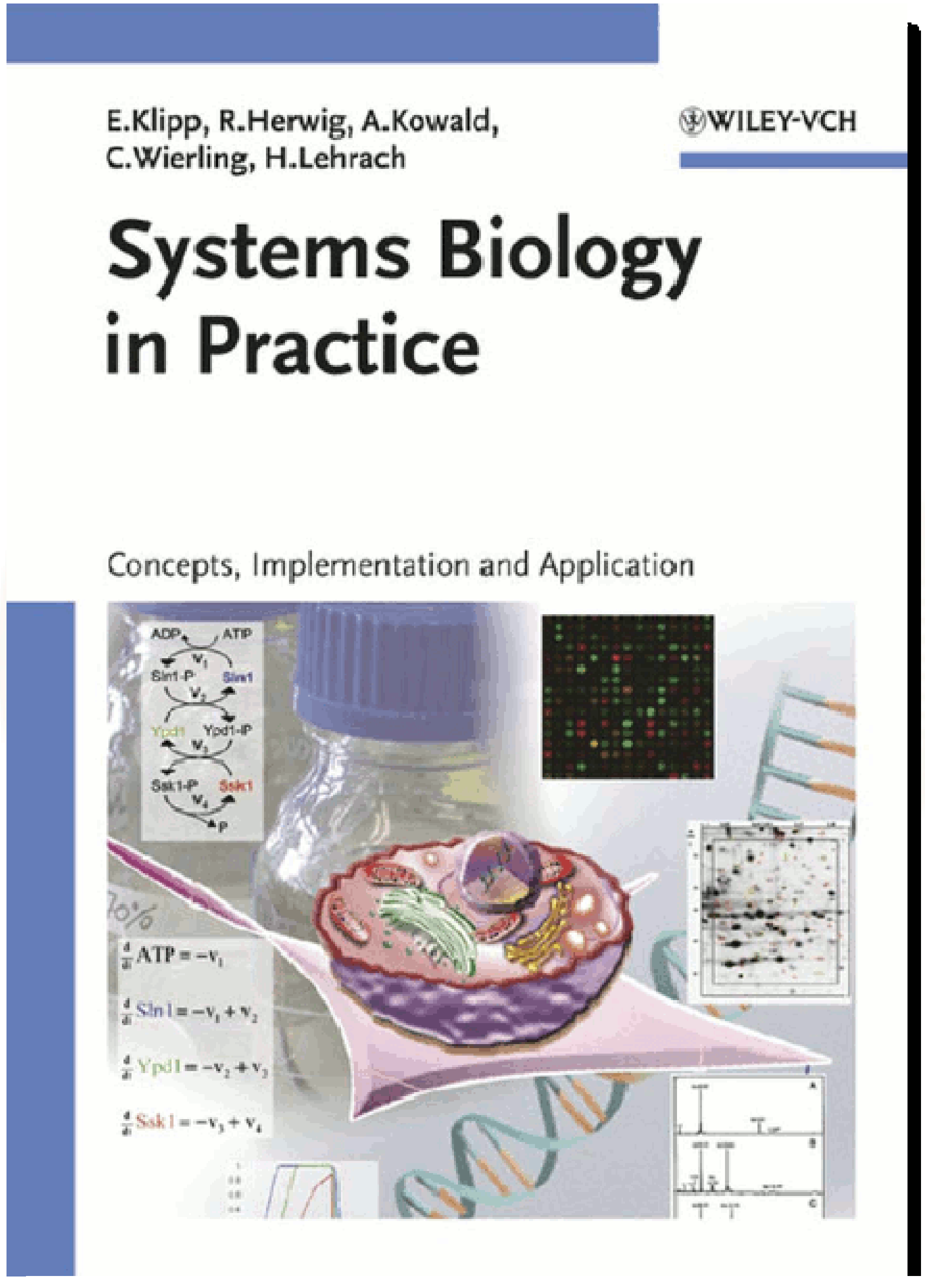
# Systems Biology-Models and Approaches

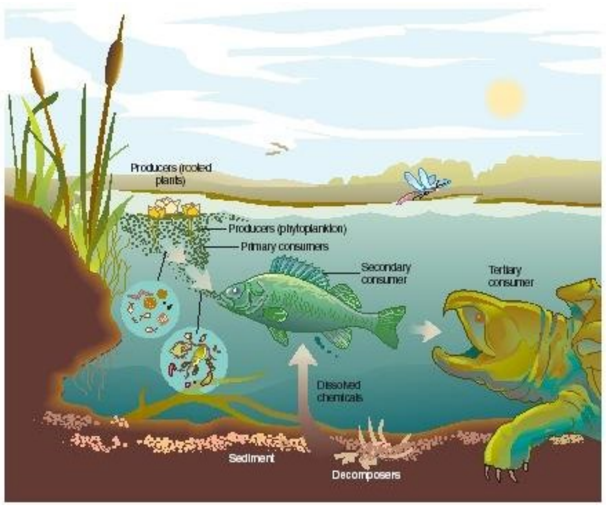

E.Klipp, R.Herwig, A.Kowald,
C.Wierling, H.Lehrach
WILEY-VCH
Systems Biology
in Practice
Concepts, Implementation and Application
ADP
ATP
Sln1-P
Sln1
Ypd1
Ypd1-P
Ssk1-P
Ssk1
P
$\frac{d}{dt}$ATP = $-v_1$
$\frac{d}{dt}$Sln1 = $-v_1 + v_2$
$\frac{d}{dt}$Ypd1 = $-v_2 + v_3$
$\frac{d}{dt}$Ssk1 = $-v_3 + v_4$

## Introduction

### Biology before Systems Biology:

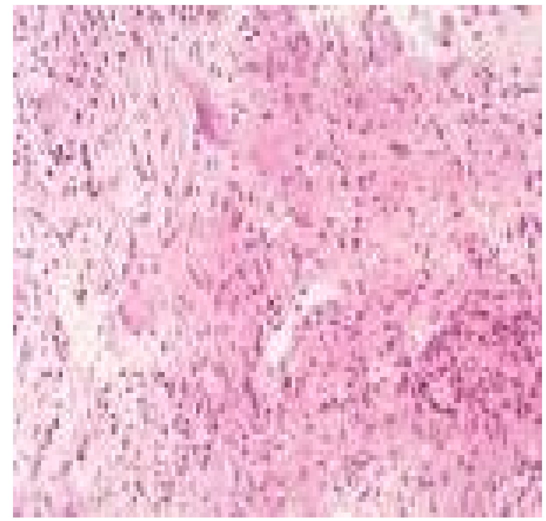
- Reductionism – Reduce the study from the whole organism to inner most details like protein or the DNA.
- Taxonomy → Study external morphology of and classify
- Physiology → Study the parts – how a root of a plant, or a leg of an animal grow.
- Cell biology, Biochemistry → Study cells – what are the reactions inside a cell.
- Molecular Biology → Study the molecules – DNA & Proteins.
- Result → Huge genomic data, millions of bases of sequence . . .

# Reductionist approach



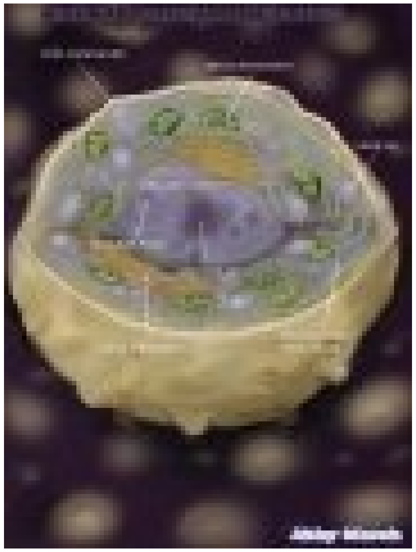

Ecosystem scale
1 km – 1000 km
Environmental impact
Nutrient flow
1 yr – 1000 yrs

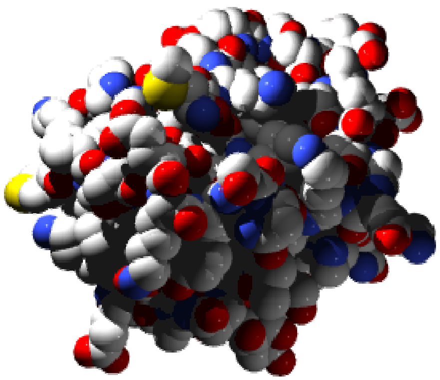
Organism scale
0.01m – 4.0 m
Behaviors
Habitats
1 hr – 100 yrs



Tissue Scale
0.01m - 1.0 m
Metabolic input
Metabolic output
1 s – 1 hr

# Reductionist approach





Cellular Scale
10 - 100 nm
Concentrations
Diffusion rates
10 ms - 1000 s

Molecular Scale
1.0 - 10 nm
Interaction data
10 ns - 10 ms

Atomic Scale
0.1 - 1.0 nm
Coordinate data
Dynamic data
0.1 - 10 ns

## Where has the reductionist approach reached?

### The Human Genome: numbers

- 23 pairs of chromosomes
- ~3,200,000,000 bases
- ~25,000 genes
- Gene length: 1000-3000 bases, spanning 30-40,000 bases
- ~1,000,000 protein variants

- Have we reached the goals of HG Project?

"*The HGP has given us a parts list. The next step is to find the relationships between the elements in a biological system: DNA, proteins, cells, tissues and organs*"
The Economist Technology Quaterly, september 17-23 2005

*"Science is built up of facts, as a house is with stones. But, a collection of facts is no more a science than a heap of stones is a house"*

- **(H. Poincaré)**

*"Life is an emergent, rather than an immanent and inherent, property of matter. Although it arises from the material world, it cannot be reduced to it"*

- **(E. Schrödinger)**

## Advancements at the molecular level

- **Complete DNA sequencing -** *mycoplasma, Escherichia coli (E. coli), Caenorhabditis elegans* **(***C. elegans***) and** *Drosophila melanogaster***.**
- **Methods to measure the mRNA level.**
- **Measurement of protein level and their interactions.**
- **Huge data on every level is available.**
- **At molecular level, the techniques are highly advanced and almost anything can be studied in detail.**

## Is it enough?

- **Nevertheless, such knowledge does not provide an understanding of biological systems as systems.**
- **Genes and proteins are components of the system.**
- **While an understanding of what constitutes the system is necessary for understanding the system, it is not sufficient.**
- **We must understand biological systems as a whole.**

**BIRTH OF SYSTEMS BIOLOGY**

## What is Systems Biology?


"Systems Biology is used in so many different contexts, nobody is really clear what you mean by it," - John Yates III.

Many definitions exist and each one is correct in its own way.

## Definitions of Systems Biology

- A branch of science that seeks to integrate different levels of information to understand how biological systems function.
- It is not the number and properties of system elements but their relations!!
- A biology-based inter-disciplinary field of study that focuses on complex interactions within biological systems, using a more holistic perspective (instead of the traditional reductionism) approach to biological and biomedical research.
- “Systems biology defines and analyses the interrelationships of all of the elements in a functioning system in order to understand how the system works.” - L. Hood.

**The study of the mechanisms underlying complex biological processes as integrated systems of many interacting components. Systems biology involves**

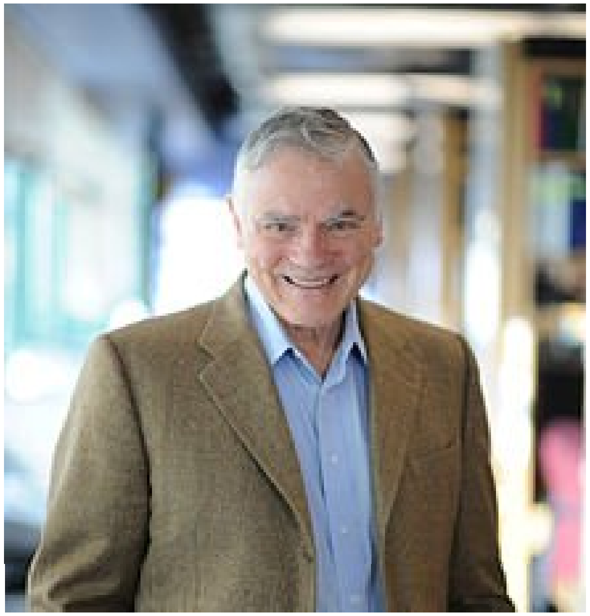
**(1) collection of large sets of experimental data**

**(2) proposal of mathematical models that might account for at least some significant aspects of this data set,**

**(3) accurate computer solution of the mathematical equations to obtain numerical predictions, and**

**(4) assessment of the quality of the model by comparing numerical simulations with the experimental data.**

**- (Leroy Hood, 1999)**



**Leroy Hood**

- Invented automated DNA sequencer
- automated tool for synthesizing DNA.
- Co-founded the Institute for Systems Biology @ Seattle, Washington, USA

**Essence of living systems is flow of mass, energy, and information in space and time.**

**The flow occurs along specific networks**

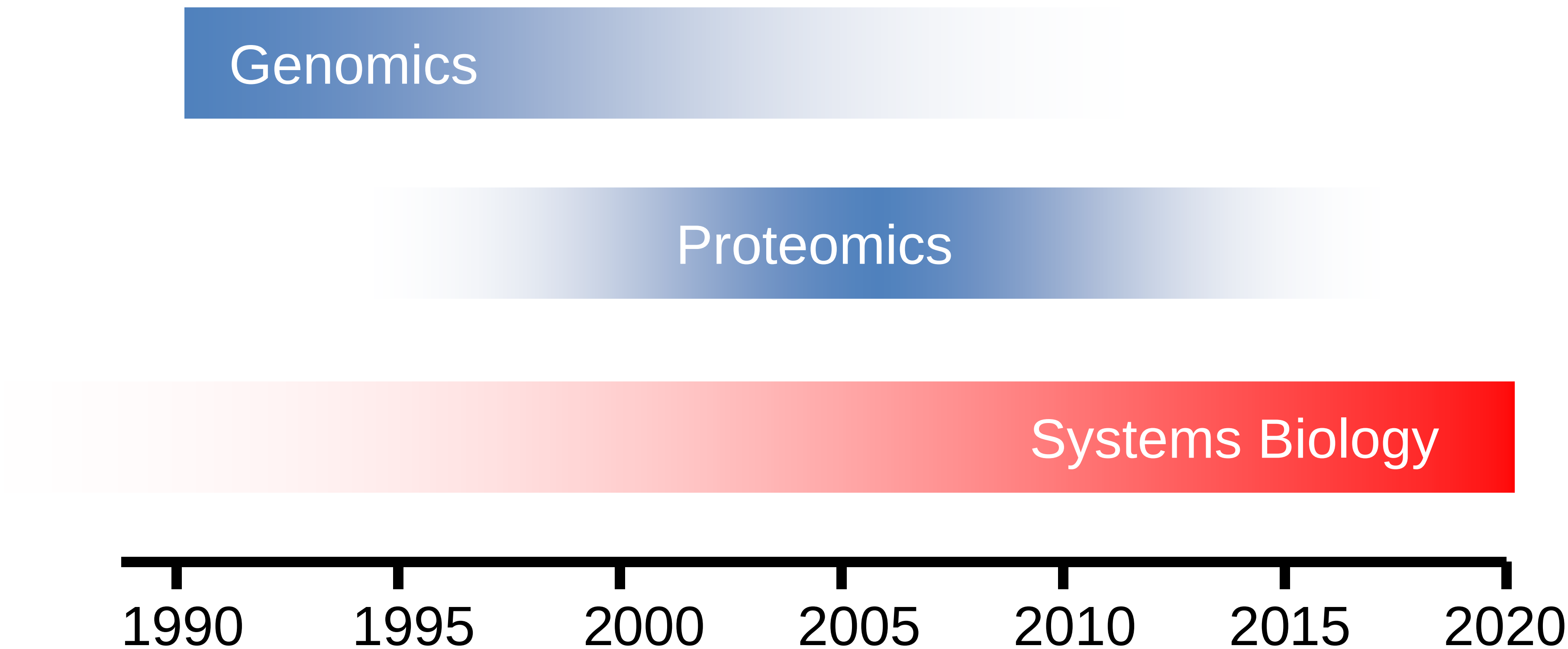
- **Flow of mass and energy (metabolic networks)**
- **Flow of information involving DNA (transcriptional regulation networks)**
- **Flow of information not involving DNA (signaling networks)**

**The Goal of Systems Biology:**

**To understand the flow of mass, energy,and information in living systems.**

**Predictive, Preventative and Personalized Medicine**
**-- Leroy Hood**

# Genomics, Proteomics & Systems Biology



# Why Systems Biology?

## On the technology side:

- Capabilities for high-throughput data gathering
- Biological networks have many more components than we previously surmised.

## On the biology side:

- The realization that we don't characterize biological systems quantitatively in their full complexity
- Hence, the scope and accuracy of our understanding of those systems will be compromised.
- Uncontrolled variables in the system will undermine our confidence in the conclusions from our experiments and observations.

## Systems Biology vs. traditional cell and molecular biology

- **Experimental techniques in systems biology are high throughput.**
- **Intensive computation is involved from the start in systems biology, in order to organize the data into usable computable databases.**
- **Exploration in traditional biology proceeds by successive cycles of hypothesis formation and testing; data accumulates during these cycles.**
- *Systems biology initially gathers data without prior hypothesis formation*; **hypothesis formation and testing comes during post-experiment data analysis and modeling.**

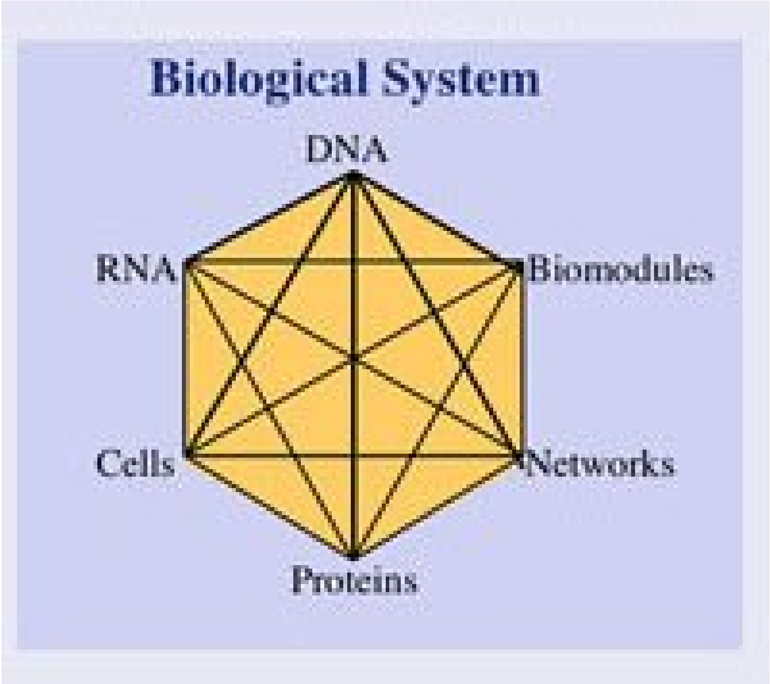
## The Core Concepts of Systems Biology

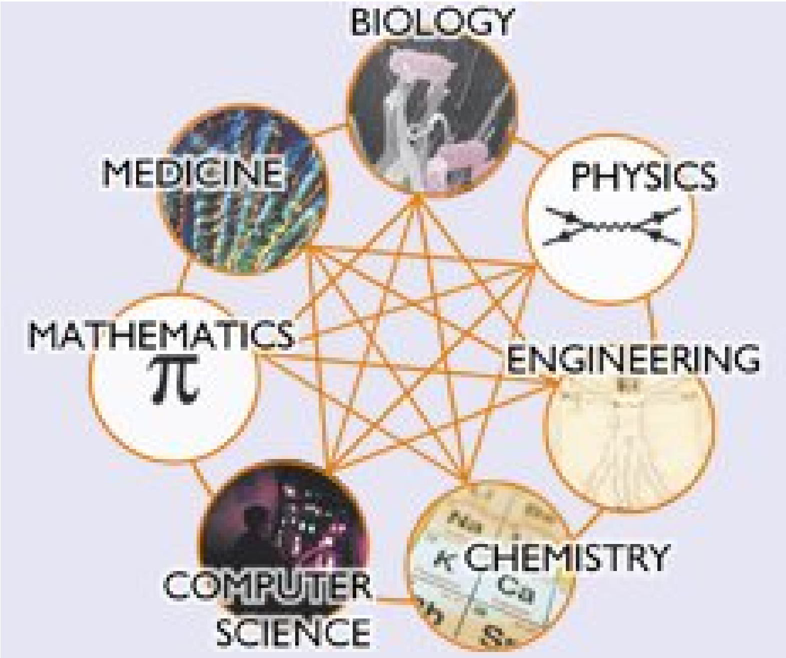
(i) Living systems are complex at all levels.

(ii) Properties of a living system emerges from the interactions of its components.

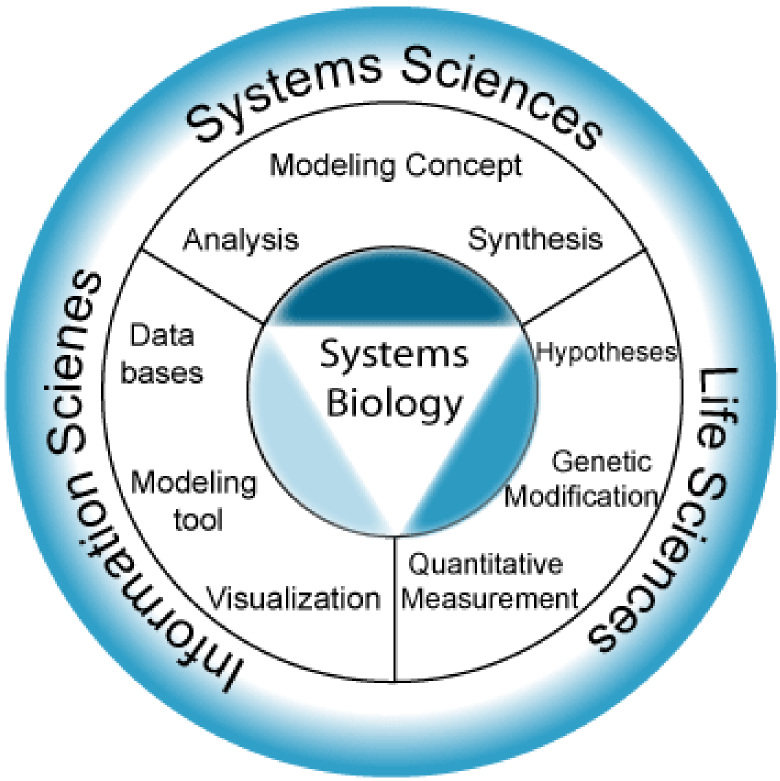
(iii) The whole is more than the sum of the parts.

(iv) Mathematics provides approaches to modeling biological systems.



## Some applications




**Basic Science/"Understanding Life"**

**Predicting Phenotype from Genotype**

**Understanding/Predicting Metabolism**

**Understanding Cellular Networks**

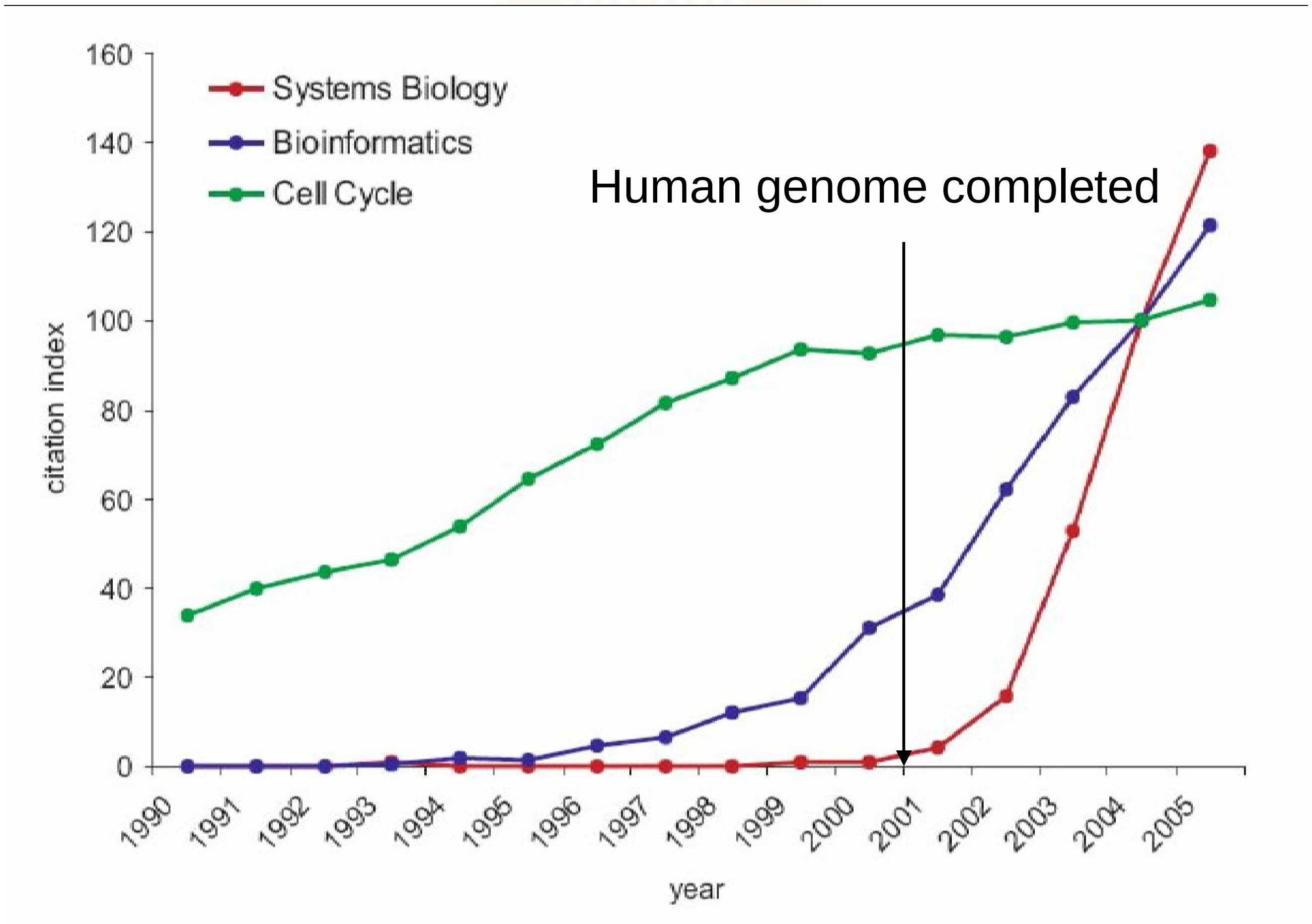
**Predicting Disease Outcome/Prognosis**

**Understanding Pathogenicity/Toxicity**

**Predicting Adverse Drug Reactions**

**Improving Medical Efficiency**

# Pubmed Abstracts – A comparison



## Basic Principles

- Life is one of the most complex phenomena in the universe.
- Biologists have thoroughly studied how parts of the cell work:

    biochemistry of small and large molecules,

    structure of proteins, DNA and RNA,

    principles of DNA replication, transcription and translation

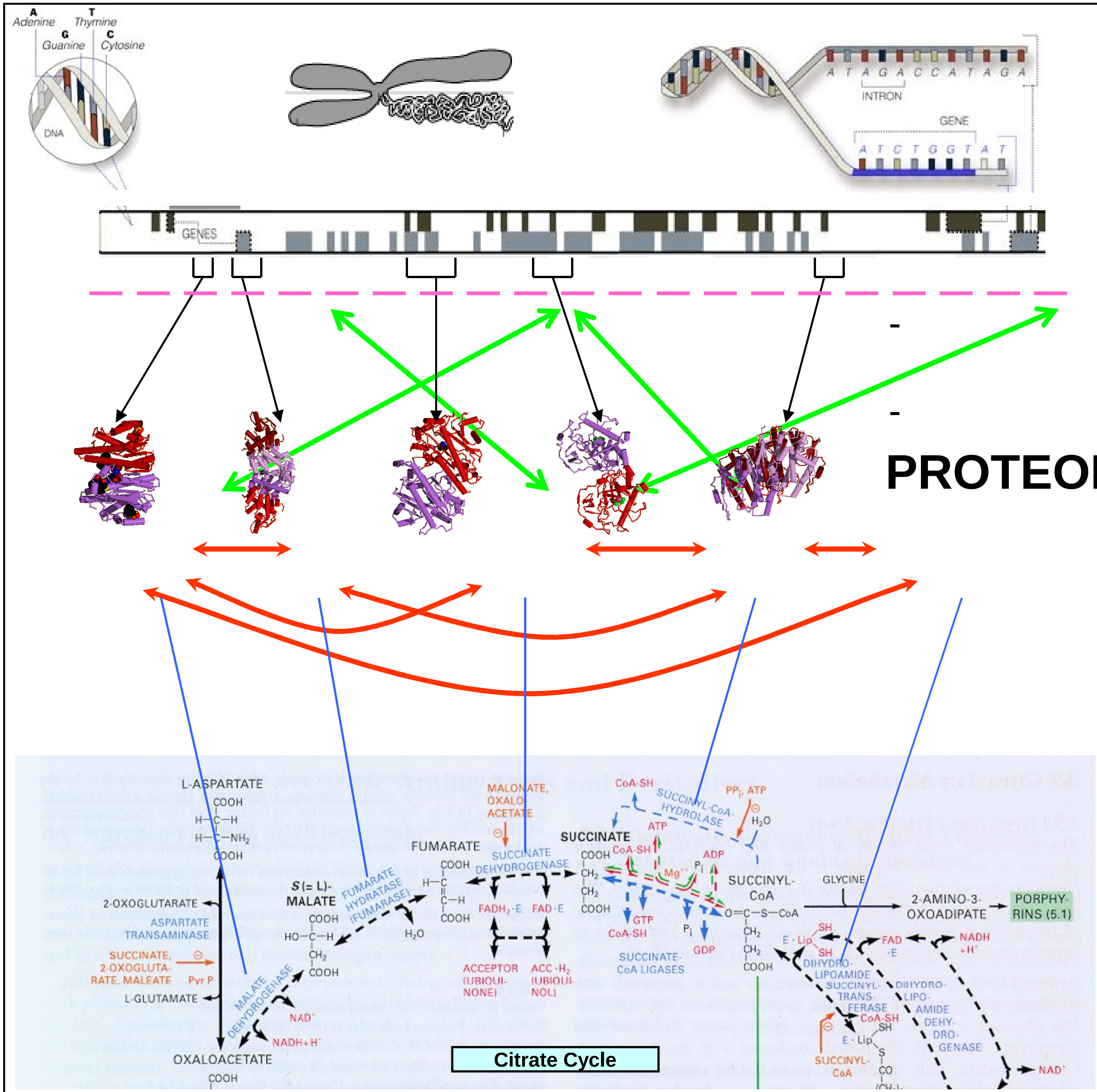
    theoretical concepts about the interaction of the molecules

- The next step is efforts towards a systematic investigation of

    cells, organs, and organisms (as a whole system)

    cellular communication, division & adaptation.

- This approach has been termed systems biology.




# GENOME

miRNA regulation?

protein-gene interactions

# PROTEOME

protein-protein interactions

# METABOLISM

Bio-chemical reactions

- Time to integrate different fields of biology and natural science
  - to understand how cells work, and the processes regulated,
  - how cells react to environmental changes and anticipate those reactions.
- Systematic view of biological processes is achieved by sophisticated experimental techniques and methodologies.
- High-throughput methods – measure the expression levels of all genes of a cell at the same time – with temporal resolution.
- A complete study of cell components and processes in time and in space is important for the further elucidation of cellular regulation.

**Systems Biology has many potential applications;**

- **Biotechnological production: requires tools with high predictive power to design cells with desired properties cheaply and reliably.**
- **Health care: models are necessary to understand the diseases and to develop methods to cure the disease.**
- **Individualized and predictive medicine**
  - **- increasing need for the exact models of cellular networks**
  - **- need for the prediction of systems behavior with reference to drug development.**
- **Once a detailed model has been constructed, all effects of possible perturbations (disturbances to the network) can be predicted fairly cheaply *in silico*.**

**Predict the outcome of complex processes**

**e. g. cancer treatment on the tumor**

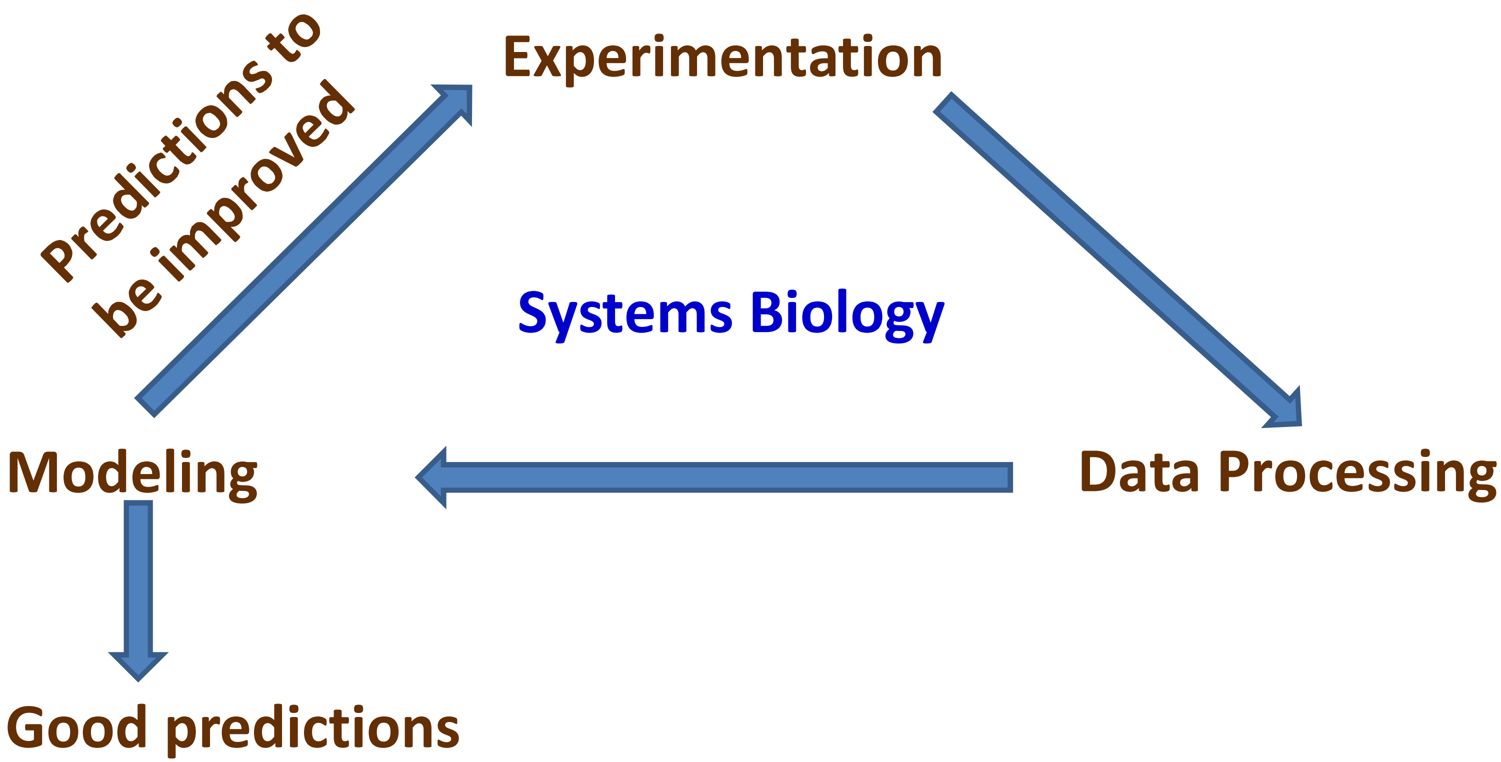
**– how effectively the treatment eliminates the tumor as well as possible metastatic cells on the patient;**

**– what the cancer treatment does to other rapidly growing tissues;**

**– how bad the predicted side effects of a specific treatment in a specific patient are;**

**Cancer and other such problems are very complex and hence, cannot be approached without systems biology.**

**The present reductionist approach like studying a single gene or process will not give enough insight to solve complex problems.**




Good predictions tell us about

how a system will behave under certain conditions,

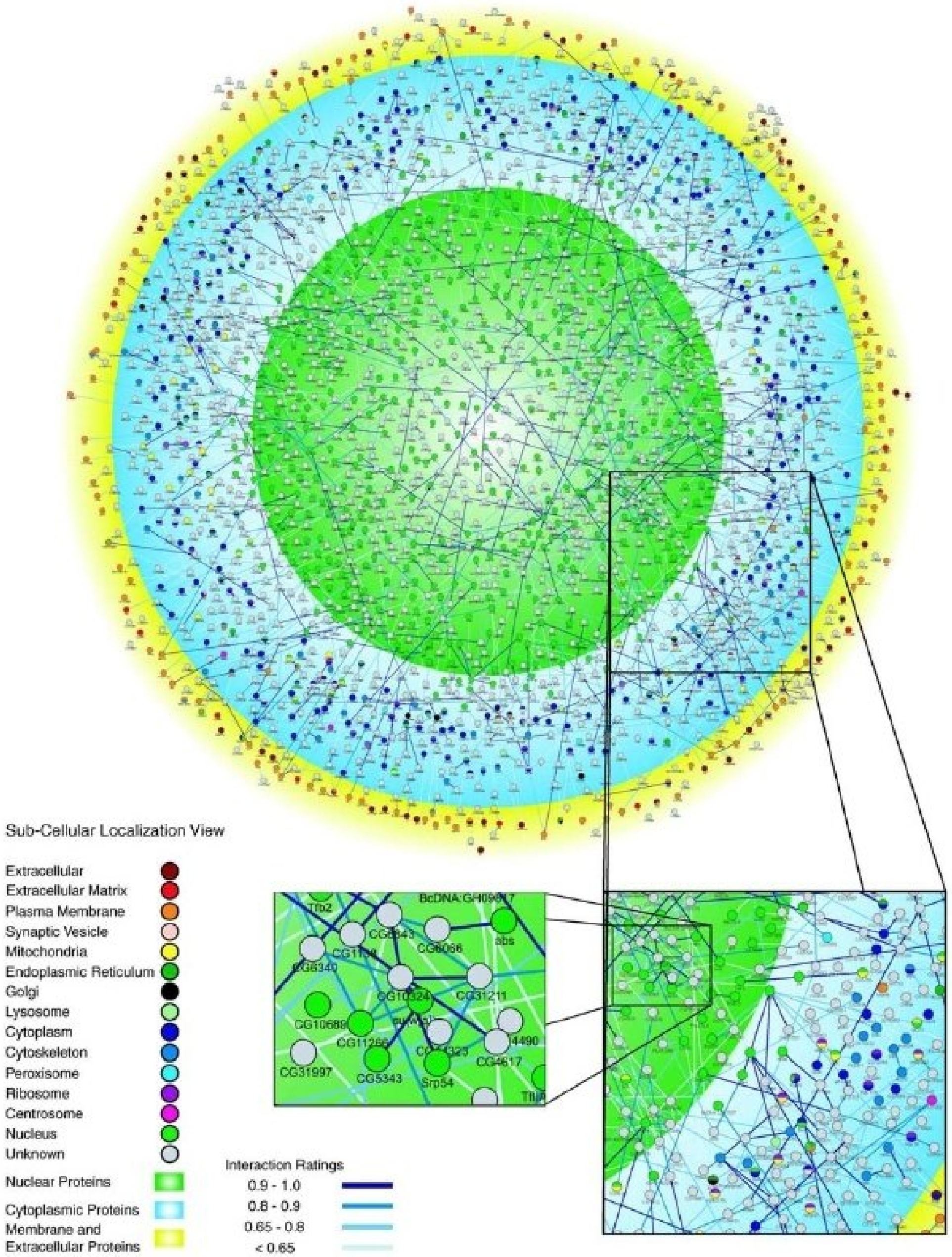
how a drug will act on a particular patient

etc. . . . . .

- A major topic of current systems biology is the analysis of **networks**:
  - gene networks,
  - protein interaction networks,
  - metabolic networks,
  - signaling networks, etc.
- The biological networks are of different types - Different theoretical methods are there for different types of networks.
- The individual networks (gene, protein interaction, etc.) must be unified to get an understanding of the system as a whole.
- Simplified mathematical models can be used to study the biological networks.
- Computational models can be used to study the biological networks.

# The Protein Network of Drosophila

CuraGen Corporation
Science, 2003



- **Biological systems are subject to evolution;**
- **This is helpful because;**

  **- organisms from different species may be similar, and hence one model could be used for several species.**

  **- prediction of protein function, prediction of network properties etc. can be easily done using similar models.**
- **This is a problem because;**

  **- variations due to evolution within the species makes it difficult to predict certain networks and their behavior.**

  **- for example for some genetic diseases, the medicine has to be personalized – a major challenge in systems biology.**

## Systems biology is Modeling

- **Mathematical modeling and computer simulations can help us to;**

  **- understand the nature and dynamics of biological processes**

  **- to predict the future of the biological process**

  **- to predict their interactions with the environment.**

- **What is a model?**

  **- a model is an abstract representation of objects or processes that explains features of objects or processes.**

  **- e.g. the letters A, C, G, and T are used as a model for DNA sequences.**

## Properties of models

### Model Assignment is not Unique

**Biological phenomena can be described in mathematical models.**

**These models are not specific or unique to an organism or process.**

**→An organism can be studied with different experimental methods.**

**→A biological process can be described with different models.**

**→A model may be applied to different biological objects.**

**→There is choice in selecting a model or an algorithm to describe a biological object.**

**→We may use different models but the modeling has to reflect essential properties of the system.**

**→Different models may highlight different aspects of the same process or an organism.**

## System State

- The state of a system is a snapshot of the system at a given time that contains enough information to predict the behavior of the system for all future times.
- Different models have different representations of the state.
- Each model defines what it means by the state of the system.
- Given the current state, the model predicts which state or states can occur next, thereby describing the change of state.

## Steady State

- Steady states, (stationary states or fixed points) – the values of all state variables remain constant in time.

- **The steady state is actually an abstraction that is based on a separation of time scales.**
- **Fast and slow processes are coupled in the biological world – -**

  **Fast - formation and release of chemical bonds – nano secs.**

  **Slow - growth of organisms – years.**
- **Fast processes reach a quasi-steady state after a short period.**
- **Slow processes are mostly in the steady state or their change is negligible.**
- **Each steady state can be regarded as a quasi-steady state of a system that is in a larger non-stationary (steady) environment.**
- **Steady states point to typical behavioral modes of the system and hence the respective mathematical problems are easier to solve.**

## Variables, Parameters, and Constants

- **All models have variables, parameters, and constants.**
- **A <u>constant</u> is a quantity with a fixed value, such as the natural number $e$ or Avogadro's number.**
- **<u>Parameters</u> are quantities that are assigned a value,**

  **Eg. enzyme concentration – the value may change.**
- **<u>Variables</u> are quantities with a changeable value.**
- **<u>State variables</u> are a set of variables that describe the system behavior completely. They are independent of each other and each of them is necessary to define the system state.**
- **For example, diameter $d$ and volume $V$ of a sphere obey the relation $V = \pi d^3/6$.**

- *Volume of a sphere* $V = \pi d^3/6$
- **π and 6 are constants and** $V$ **and** $d$ **are variables, but only one of them is a state variable, since the mentioned relation uniquely determines the other one.**
- **Whether a quantity is a variable or a parameter depends on the model**

  **- The enzyme concentration is frequently considered a parameter in biochemical reaction kinetics.**

  **- no longer valid if, in a larger model, the enzyme concentration may change due to gene expression or protein degradation.**

## Model behavior

- Model behavior is determined by

  (1) influences from the environment (input) and

  (2) processes within the system.

- Further, the system structure, (i. e., the relation among variables, parameters, and constants) determines how the external inputs are processed and how the internal processes are controlled.

## Process classification

- Reversible – process can proceed in forward or backward direction.
- Irreversible – process can proceed in only one direction
- Periodicity – at specific time intervals the state of the process may change.

## Process classification – continued

- <u>Deterministic</u> – the future state of the process may be determined by the current state.
- <u>Stochastic</u> – future state cannot be determined – only a probability can be predicted.
- <u>Continuous</u> – the values of the state are continuous.
- <u>Discrete</u> – the values of the state are discrete (not continuous).

## Advantages of Computational Modeling

- Modeling gives conceptual clarification – makes things understandable – Eg. AGTC for nucleotides.
- Modeling also highlights gaps in knowledge or understanding – if a model fails to mimic the real experiment then there is a gap.

- **Modeling provides independence from the modeled object. Time and space may be stretched or compressed *ad libitum*.**
- **Modeling is cheap compared to experiments. Models do no harm on animals or plants and help to reduce them in experiments.**
- **They do not pollute the environment. Models do not interact with the environment or with the modeled system.**
- **Modeling can assist experimentation. With an adequate model one may test different possibilities that are not available in experiment.**
- **One may follow time courses of compounds that cannot be measured in an experiment. Eg. Aging process**
- **One may impose perturbations that are not feasible in the real system. Eg. Effect of a toxin.**

- **One may cause precise perturbations without directly changing other system components, which is usually impossible in real systems.**
- **Model simulations can be repeated for many different conditions.**
- **Model results can often be presented in precise mathematical terms that allow for generalization.**
- **Graphical representation and visualization make it easier to understand the system.**
- **Finally, modeling allows for making well-founded and testable predictions.**

## Model Development – Modeling Workflow

**1. Formulation of the problem:**

- which questions shall be answered with the model
- clear statement about the background, problem, and hypotheses

**2. Verification of available information:**

- the existing data about the structure of the system has to be collected and checked.

**3. Selection of model structure:**

- determine the general type of the model like 1) microscopic or macroscopic, 2) deterministic or stochastic approach, 3) discrete or continuous variables, 4) steady-state, temporal, or spatio-temporal description. Find variables for external influence, internal structure.

**4. Establishing a simple model:**

- **model can be in words, scheme in mathematical formulation.**

**5. Sensitivity analysis:**

- **mathematical simulation may be highly sensitive to parameter changes. Verify the parameter choice.**

**6. Experimental tests of the model predictions:**

**7. Stating the agreements and divergences between experimental and modeling results:**

- **whether the model results agree with the experimental behavior**
- **if results do not agree, check for false assumptions, over simplification, wrong model structure, inadequate experimental design, or other inadequately represented factors.**

**8. Iterative refinement of model:**

**- usually initial models may not be correct. Refine continuously till the model adequatly represents the biological system.**

## Typical Aspects of Biological Systems and Corresponding Models

**Network Versus Elements :**

**A system has elements that interact and form a network.**

**The elements have certain properties. In the network, the elements have certain relations to each other (and, to the environment).**

**The system has properties that rely on the individual properties and relations between the elements.**

**System may show properties and characteristics that often cannot be deduced from the individual properties of the elements.**

**Modularity:**

**Modules are subsystems of molecular networks that can be treated as functional units, which perform identifiable tasks.**

**Parts of a networks are separated as modules by the criterion that mass transfer occurs internally but not between the modules - modules are linked by regulatory effects - one chemical from a module regulates other module.**

**Advantage of modularity – models can be first studies separately and then integrated in a network. Disadvantage – ignore the connectivity between different modules – sort of reductionist approach which is against the concept of systems biology..**

## Systems Biology is data integration

Important part of systems biology is integration of data from high-throughput techniques like genomics, transcriptomics and proteomics .

Lowest complexity - defines common schemas for storage, representation and transfer of data – Eg. Microarray data.

More complex level, schemas have been defined for biological models and pathways – Eg. SBML, and CellML which use an XML-like language style.

Second level of complexity, query-based information information retrieval, connection of different data types in different databases – and the visualization and presentation of the data. Eg. SRS

**Next level of complexity – data correlation – combine information from diverse datasets to learn / explain natural processes. Eg. Integrate information from transcriptome or proteome with genome sequence annotations.**

**At this level errors /misinterpretations are common. Normalization or minimizing critical errors is important. Requires sophisticated data analysis tools, data mining tools and algorithms.**

**Highest level of data integration – mapping of the integrated experimental data into networks in order to model interactions of the biological system.**

**Currently this level is not well achieved – due to the lack of standardization – need of the hour in systems biology.**

I

- .