From Bench to Bedside: Role of Informatics

Nagasuma Chandra
Indian Institute of Science
Bangalore

Apparent ‘disconnect’ among DATA pieces STUDYING THE SAME SYSTEM

Chest sounds

Blood tests

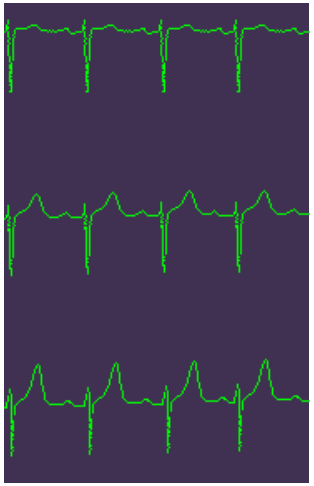
Echocardiogram

Stress Test

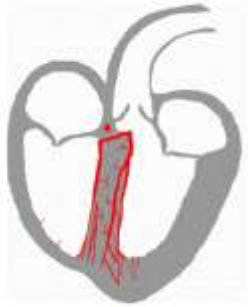
Blood pressure

Conductance

Heart



Electrocardiogram



Navigating across multi-scale multi-level biological systems

Structuring Complexity in Engineering

Early models of engineered system behaviour are cognitive models of the system as described by experts

Engineering has used- Tacit knowledge, Rules, Experience

These help in defining physical behaviour to some extent- however knowledge in rules not sufficient

Need physical laws to be obeyed and tested with experiments

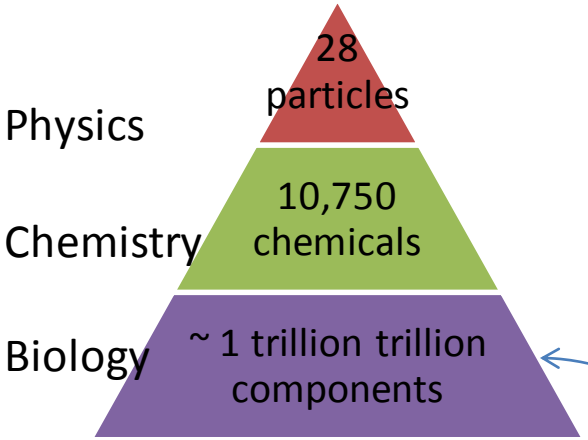
Therefore quantitative -- measurable mathematical models and simulations are needed at various fidelity levels

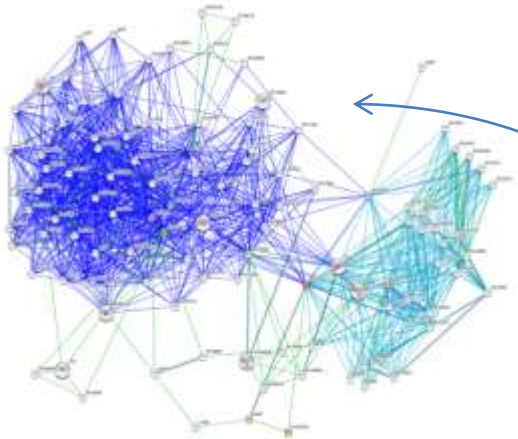
Courtesy: Chandra S; National Aerospace Laboratories

Learning from Engineering

Has a well-defined blue-print

Provides clues to Structuring Complexity





Need to obtain a blue-print for biological systems as well

for

(a) Basic understanding of the living systems &

(b) Application in Medicine & Biotechnology

BUT

Problems associated are

- Too many players
- Highly complex interactions

Challenges in studying biological systems

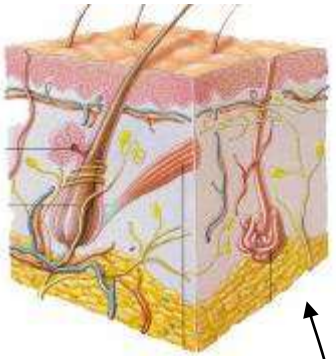
Several challenges must be met however, in order to study biological systems.

- formulate biological questions as network amenable problems,
- reconstruct networks with appropriate resolution from available data.
- establish relationships within each layer of data but more importantly among different levels of data,
- **to identify and understand the flow of information in terms of biochemical, biophysical structural and molecular signals within a cell, leading to various biological events.**

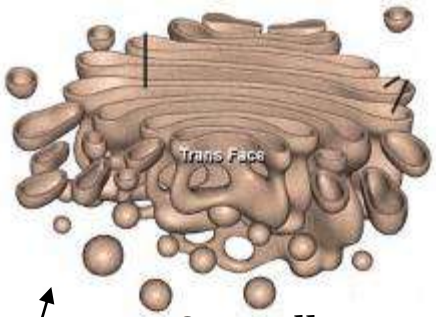
Hierarchical structures in living systems

Cell

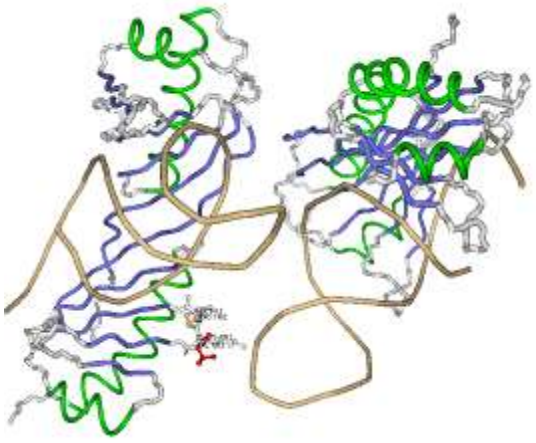
Tissue



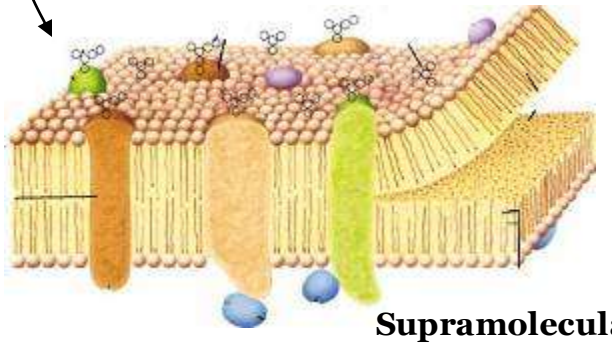
Organ



Organelle

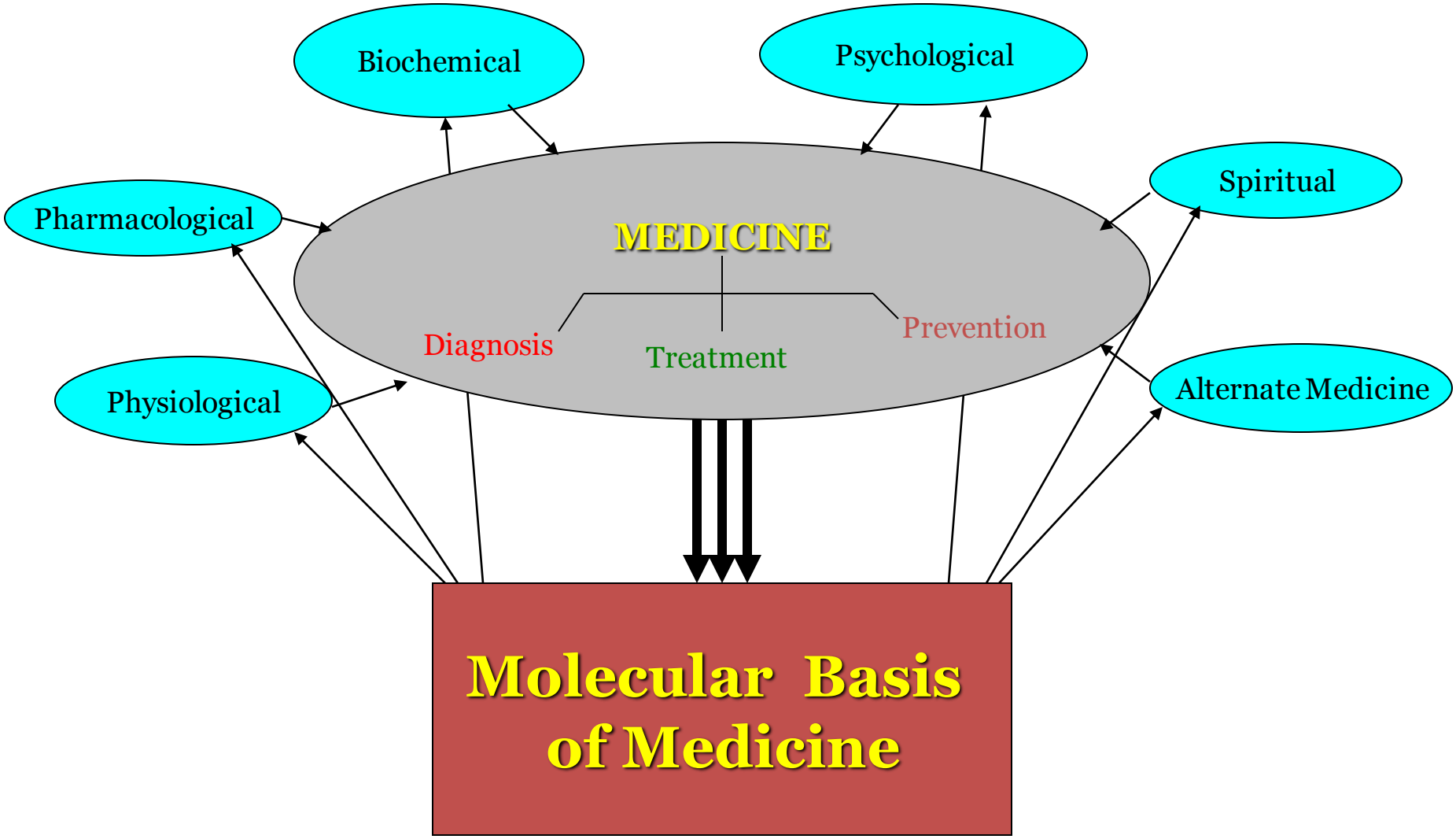


Macromolecule

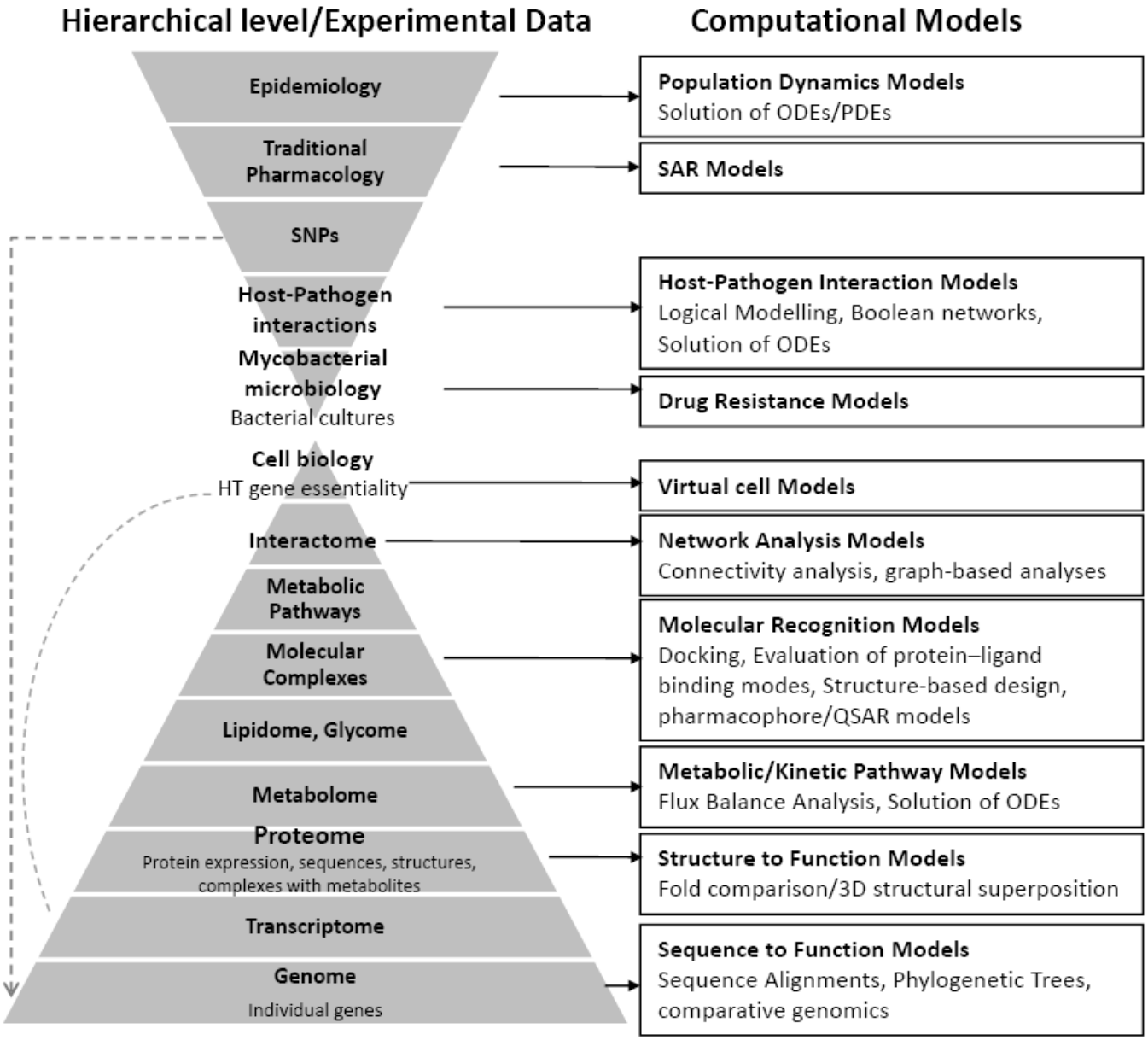


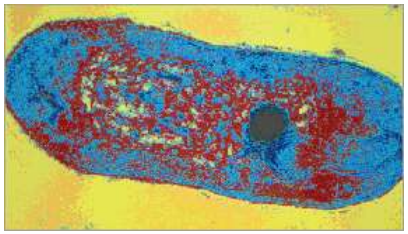
Supramolecular assembly

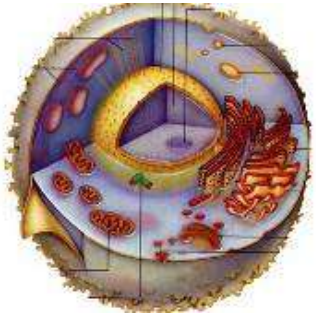
Organism


Biochemical
Psychological
Pharmacological
Spiritual
MEDICINE
Diagnosis
Treatment
Prevention
Physiological
Alternate Medicine
Molecular Basis of Medicine

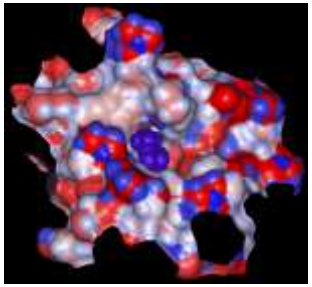
Integrated Systems Approach

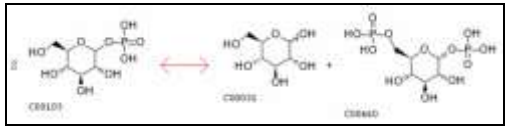


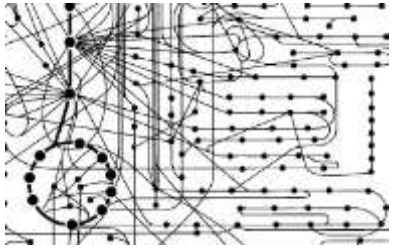




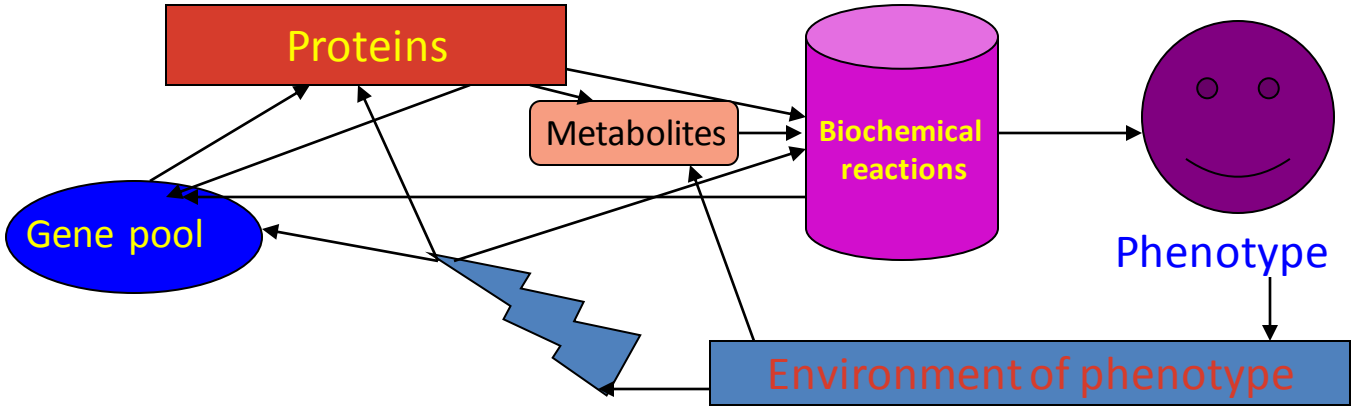
MSFVVTIPEALAAVATDLAGIGSTIGTAN
AAAVPTTTVLAAAADEVSAAMAALFSGHA
QLAYQALSAQAALFHEQFVRALTAGAGSY







Understanding a cell



Descriptions:

- Sequences (f_n implicit)
- Structures
- Metabolites
- Biochemical Reactions (f_n explicit)
- Regulation elements
- Network of proteins/metabolites/reactions

Systems Biology

Genomics

Proteomics

Genome sequence

Network

Functional Annotation

Stoichiometric Matrix

R1	R2	R3	R4	R5	R6	R7	
1	−1	−1	0	0	0	0	A
0	1	0	−1	−1	0	0	B
0	0	1	0	1	−1	0	C
0	0	0	0	0	1	−1	D

Temporal Profiling

Bioinformatics

Structure

Differential Equation

Binding Site

Schrodinger Equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\psi = E\psi$$

Mathematical biology

Applications

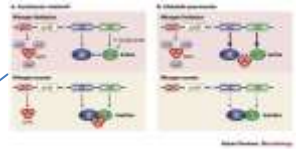
Resolution of information

Model Scale / Complexity

Taxonomy of Models

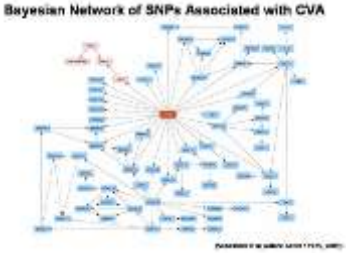
Low granularity

High granularity



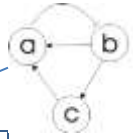
Qualitative Models

Machine Learning
Statistical Models

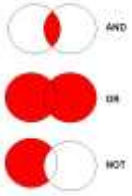


Bayesian Networks

Network
Topological Analysis



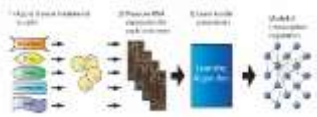
		0;0	0;1	1;0	1;1
a	b;c	0	1	1	0
b	a;a	1	0	0	1
c	b;b	0	1	1	1



Boolean Modelling

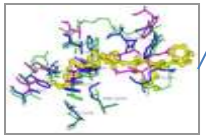
Stoichiometric Matrix
Flux Balance Analysis

$$\begin{array}{c} R1\ R2\ R3\ R4\ R5\ R6\ R7 \\ \begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \end{matrix} \end{array}$$



Stochastic Modelling

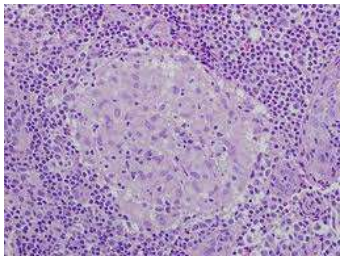
Differential Equations

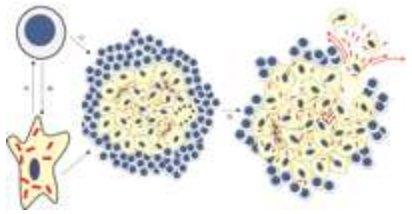


Molecular Recognition
Models

Atomistic Models

M.tuberculosis: A successful pathogen





Tuberculosis has been present in humans since antiquity
(Earliest evidence in prehistoric humans – 7000 to 18000 BC)

4000 proteins

1000+ biochemical reactions; 100s signalling / regulatory events

High redundancy in the genome

Robust system

Contains several immune evasion mechanisms

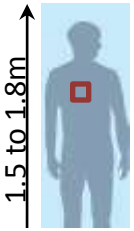
Dormant state that can reactivate after decades to cause active disease

Many people infected but do not contract the disease

→ A multi-level view necessary

COMMON DESIGN PRINCIPLES ACROSS DIVERSE EXPRESSIONS OF LIFE

•100 trillion cells
•3 billion base pairs
•20000-25000 proteins
•210 distinct cell types
•400 billion chemical reactions every secon

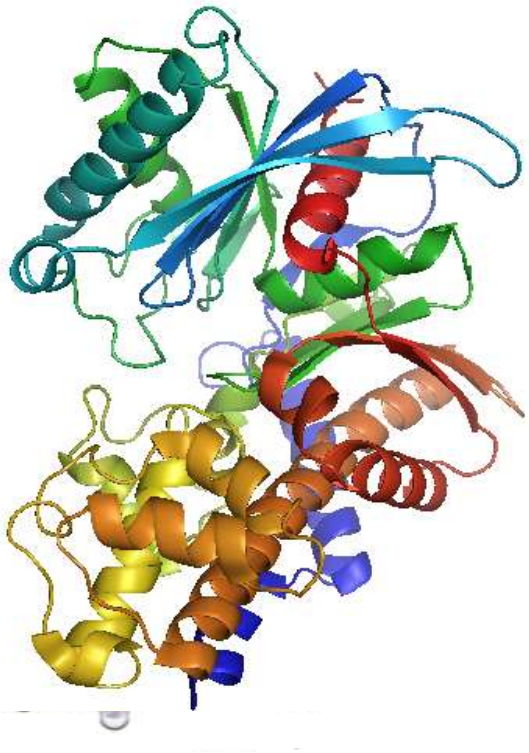


•2-4µ in length
•0.2-0.5µ in width

•Each cell
•4 million base pairs
•4000 proteins
•1000 metabolites
4000 X 10000 atoms

lei

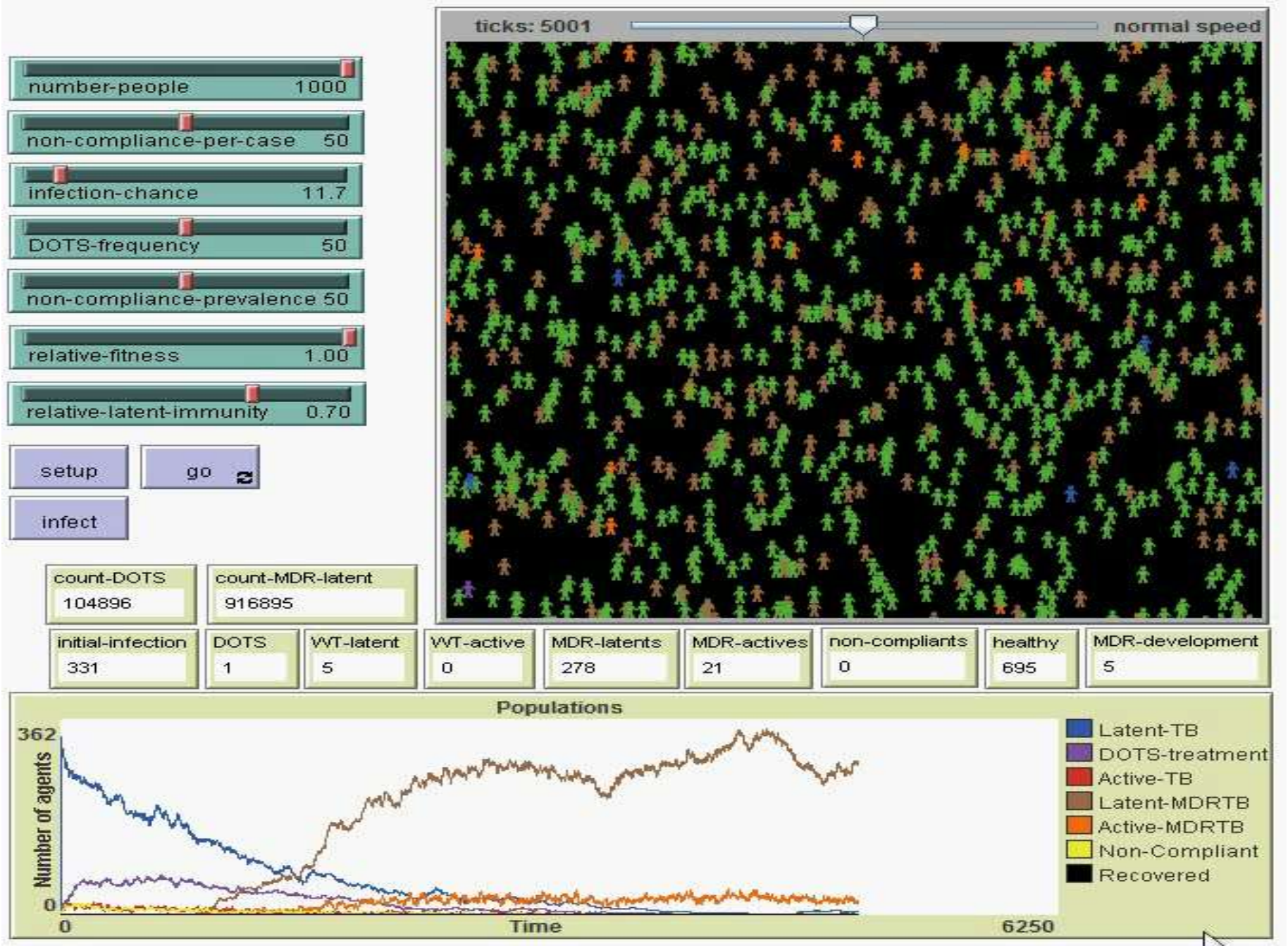
Three dimensional structure of one protein
Contains ~10,000 atoms



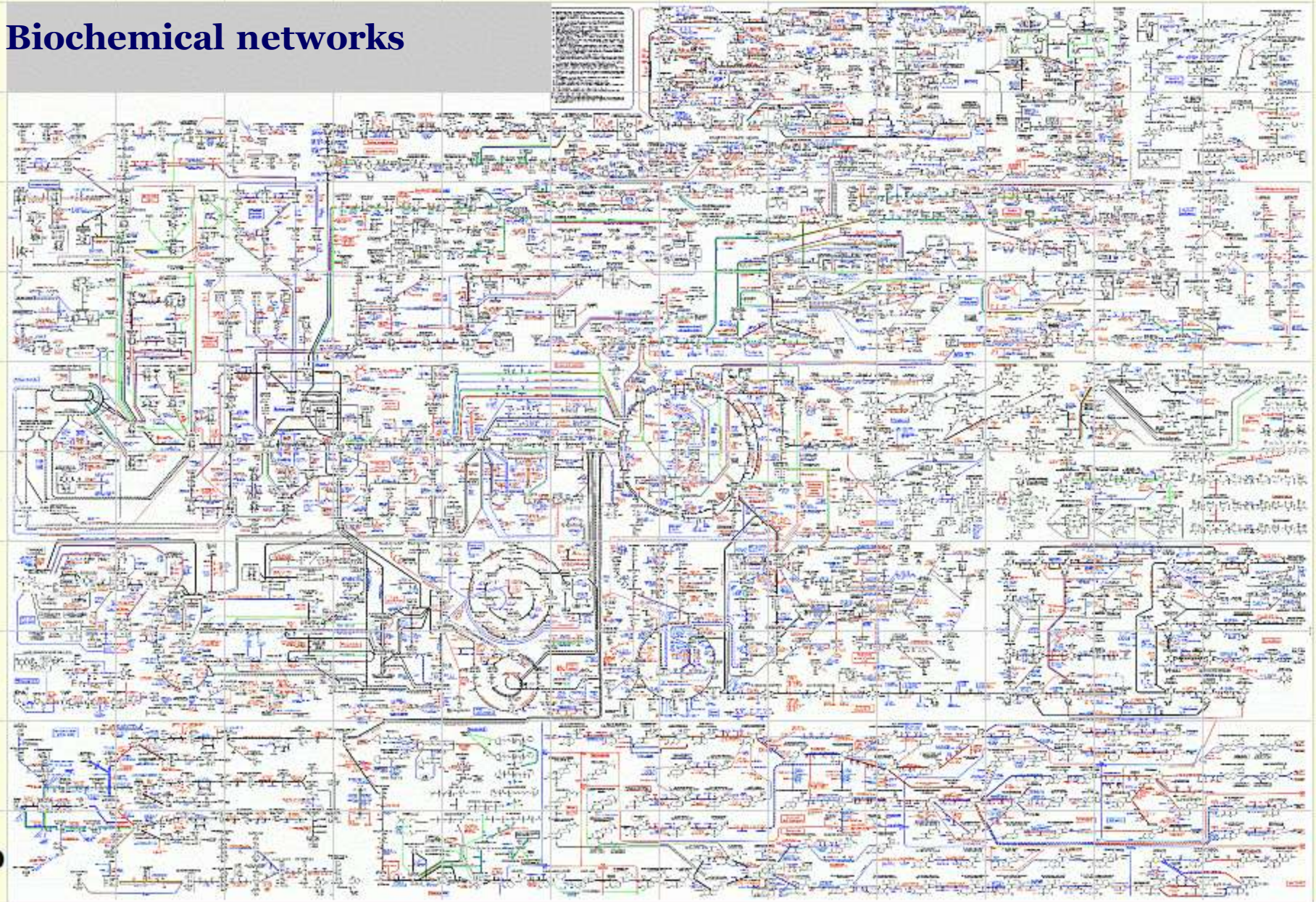
cytosis of the bacteria
ıman macrophage

Pathway- ~ 15 proteins

Protein-Protein interactior network

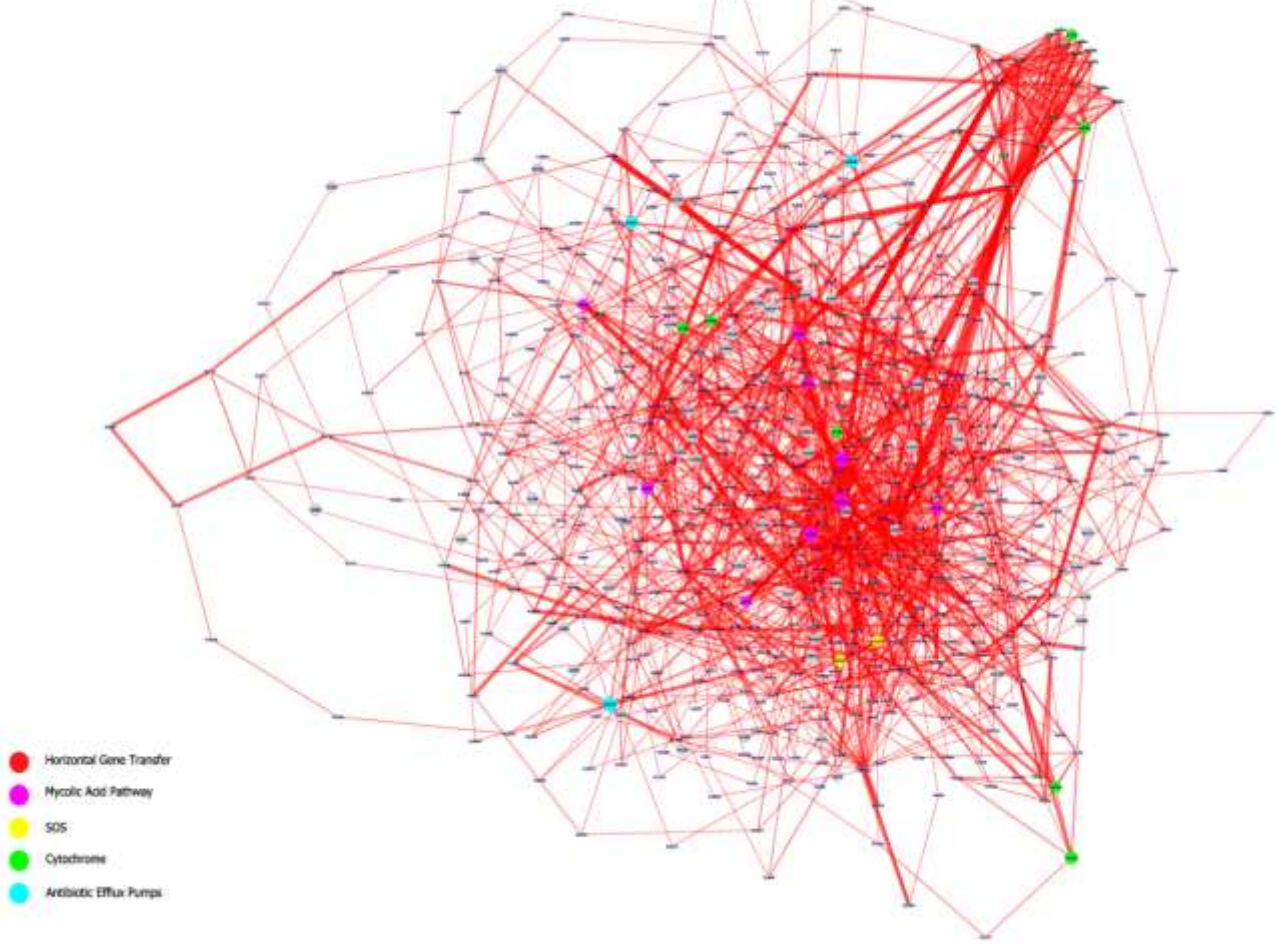

ticks: 5001
normal speed
number-people 1000
non-compliance-per-case 50
infection-chance 11.7
DOTS-frequency 50
non-compliance-prevalence 50
relative-fitness 1.00
relative-latent-immunity 0.70
setup
go
infect
count-DOTS
104896
count-MDR-latent
916895
initial-infection
331
DOTS
1
WT-latent
5
WT-active
0
MDR-latents
278
MDR-actives
21
non-compliants
0
healthy
695
MDR-development
5
Populations
362
Number of agents
0
0
Time
6250
Latent-TB
DOTS-treatment
Active-TB
Latent-MDRTB
Active-MDRTB
Non-Compliant
Recovered

Biochemical networks



Cellular Networks

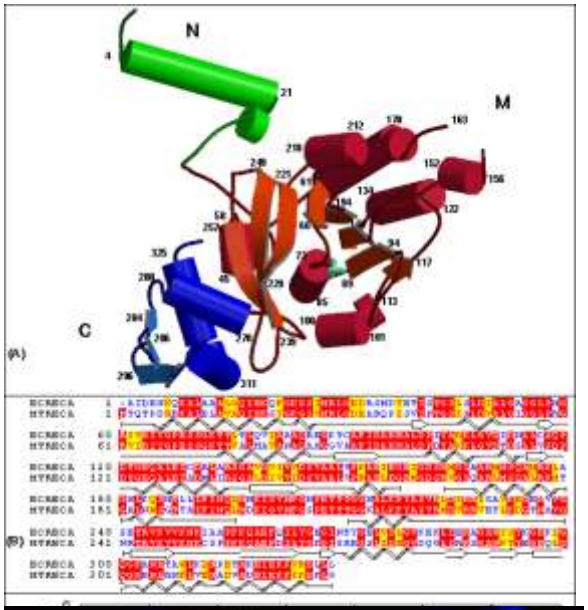
Abstraction of the flow of information that leads to Drug Resistance in TB bacilli

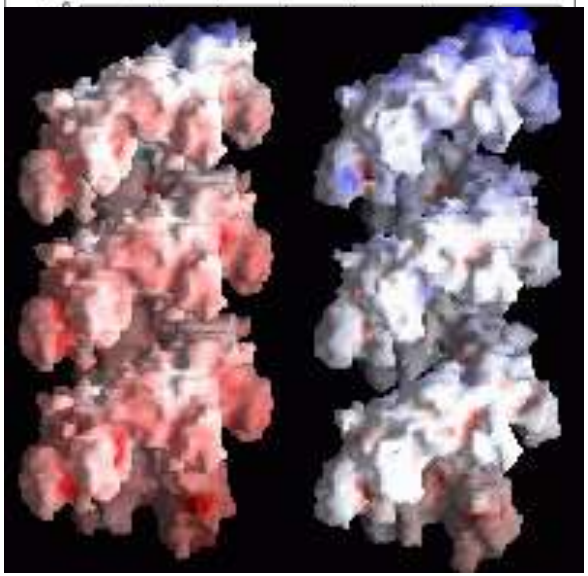


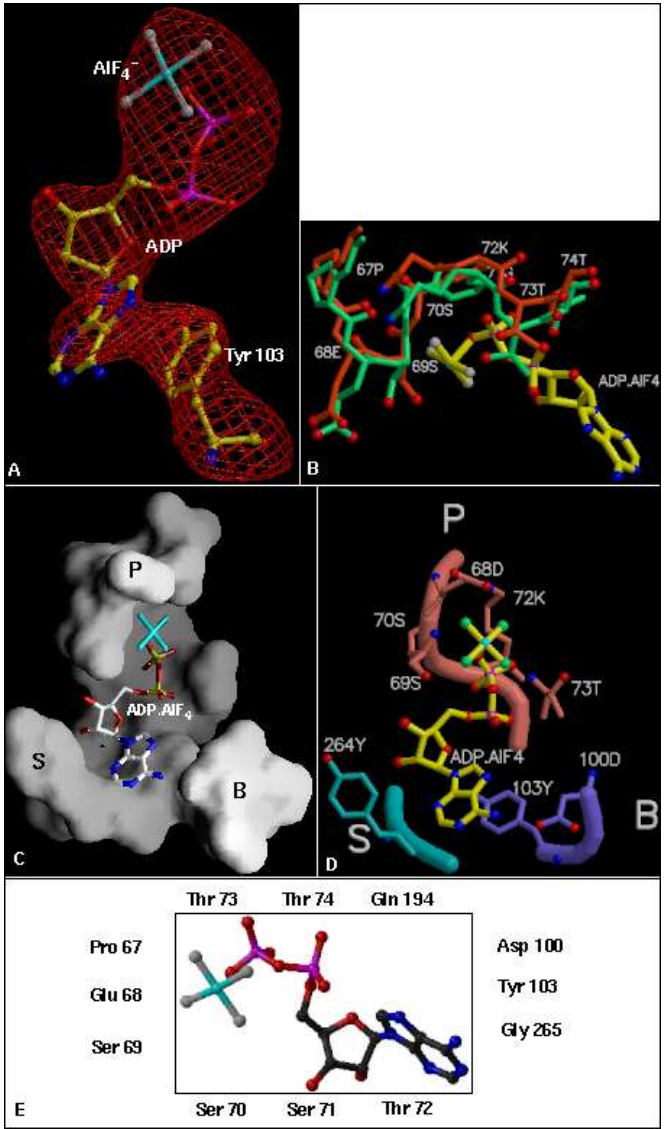
- Network of shortest paths from MAP to Resistance Genes
- 616 nodes and 1,683 edges
- Paths scored based on edge frequency, up-regulation of source and target nodes

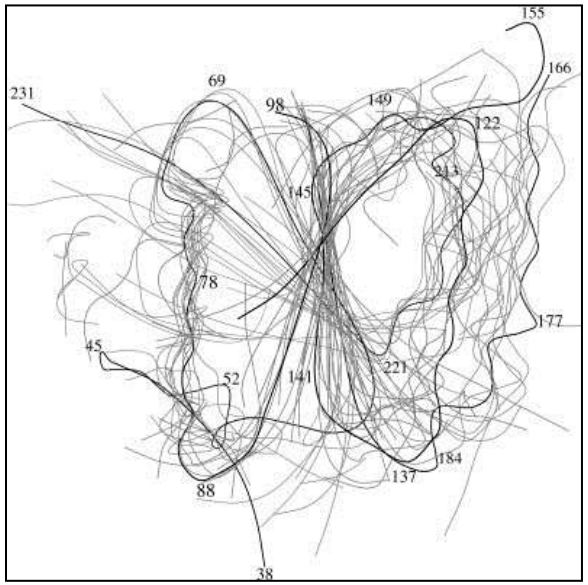
Raman and Chandra, 2008, BMC Microbiology

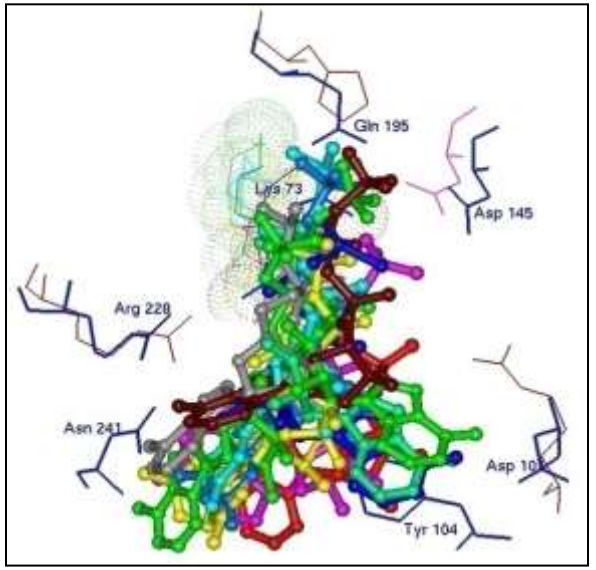
Structural Bioinformatics











Virtual screening in drug discovery/design

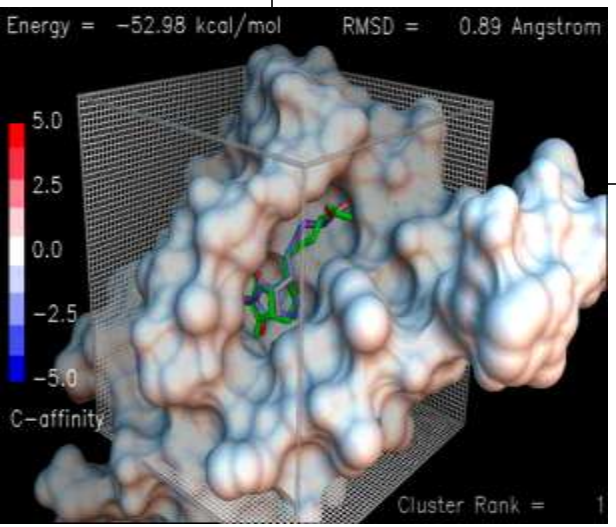
Ligand size: ~ 10-50 atoms (flexible)

Protein size: ~ 2000-5000 atoms (rigid)

1 docking run: ~ 10^6 energy evaluations

on a high-end PC

=> CPU time ~ 30min.



Modest Ligand database: ~ 10^6 compounds

Protein molecules are NOT rigid & multiple conformations must be sampled

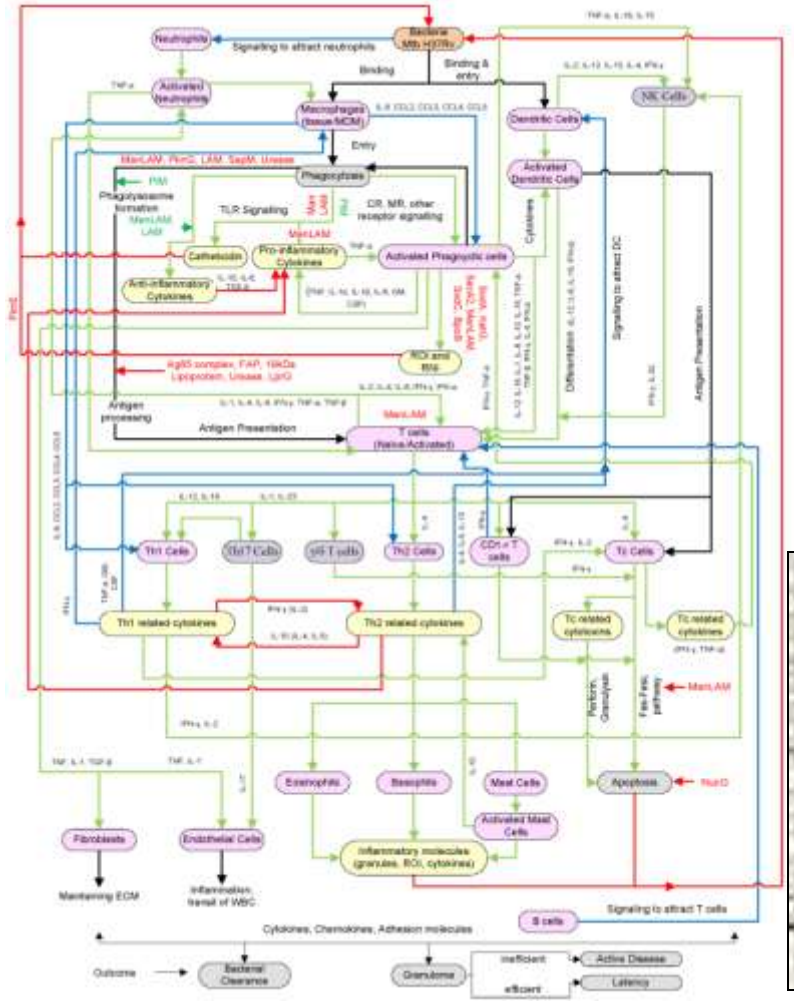
A database search requires:

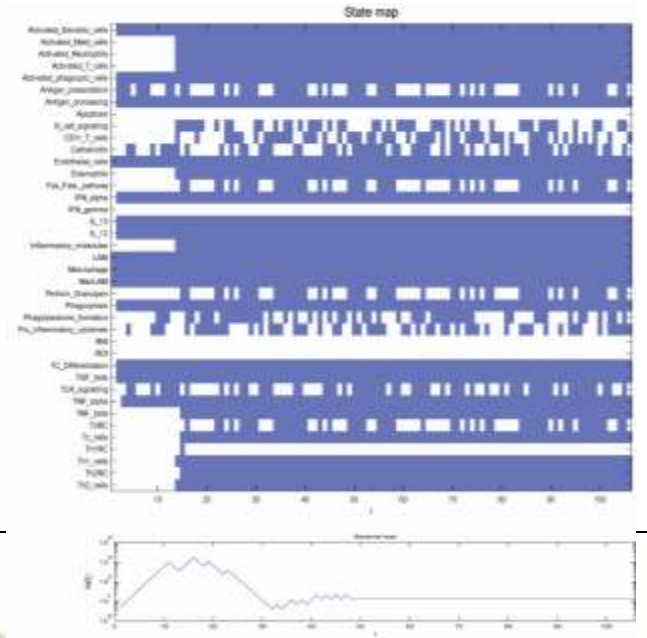
~30min * 10 (protein conformations) * 1 million (ligands in database)

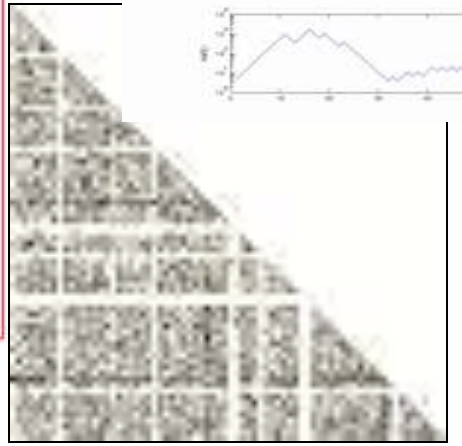
~ $5*10^6$ hrs.

How about studying several proteins?

Host-Pathogen Interaction Modelling: Predicting disease outcome





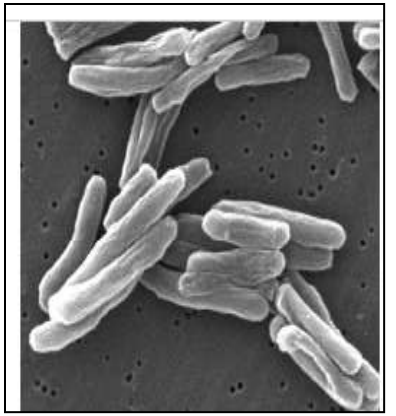


A Boolean model of HPIs developed, Simulations to capture a variety of scenarios

Raman, Bhat & Chandra, Mol. Biosyst, 2010

Biological Design-

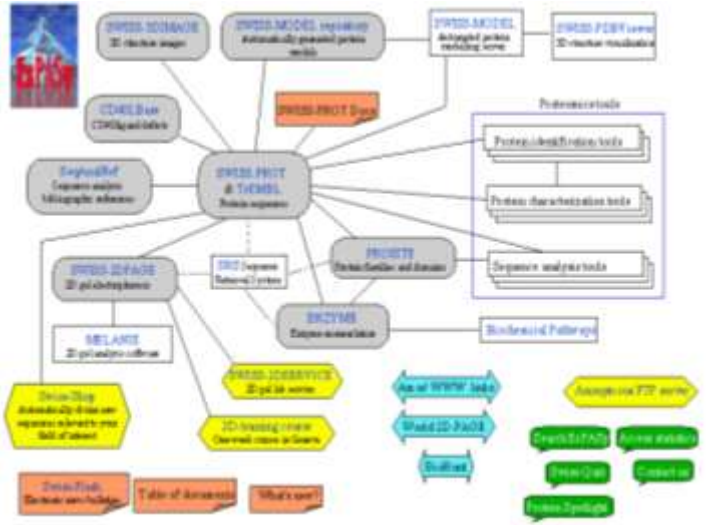
Outcome of a random 'tinkering' process (Evolution)



Engineering – Built on purpose with a pre-designed blue-print

From - Chandra N & Chandra S; Microsoft e-science workshop, October 2009

Data Integration



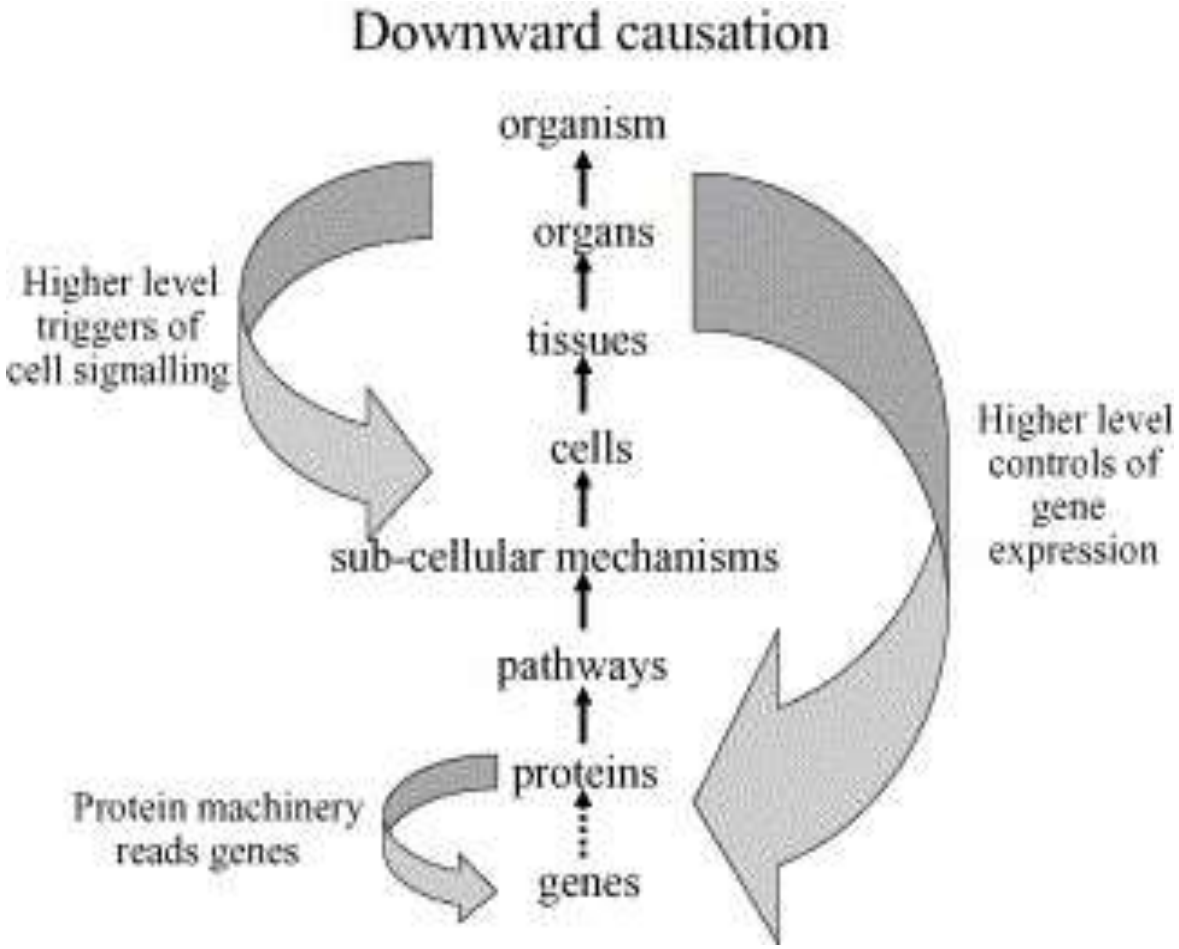
- Data Resources-

  Primary & Derived Databases

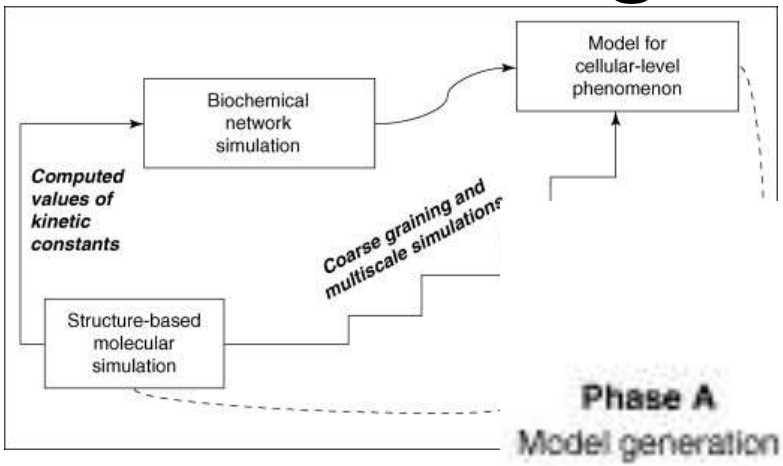
  » cross mapping- across databases
- Data descriptions
- Data representation-Data structures, Syntaxes
- Data Integration
  - Data InterRelationships-Ontologies
  - Data flow pipeline
  - A biology workbench??

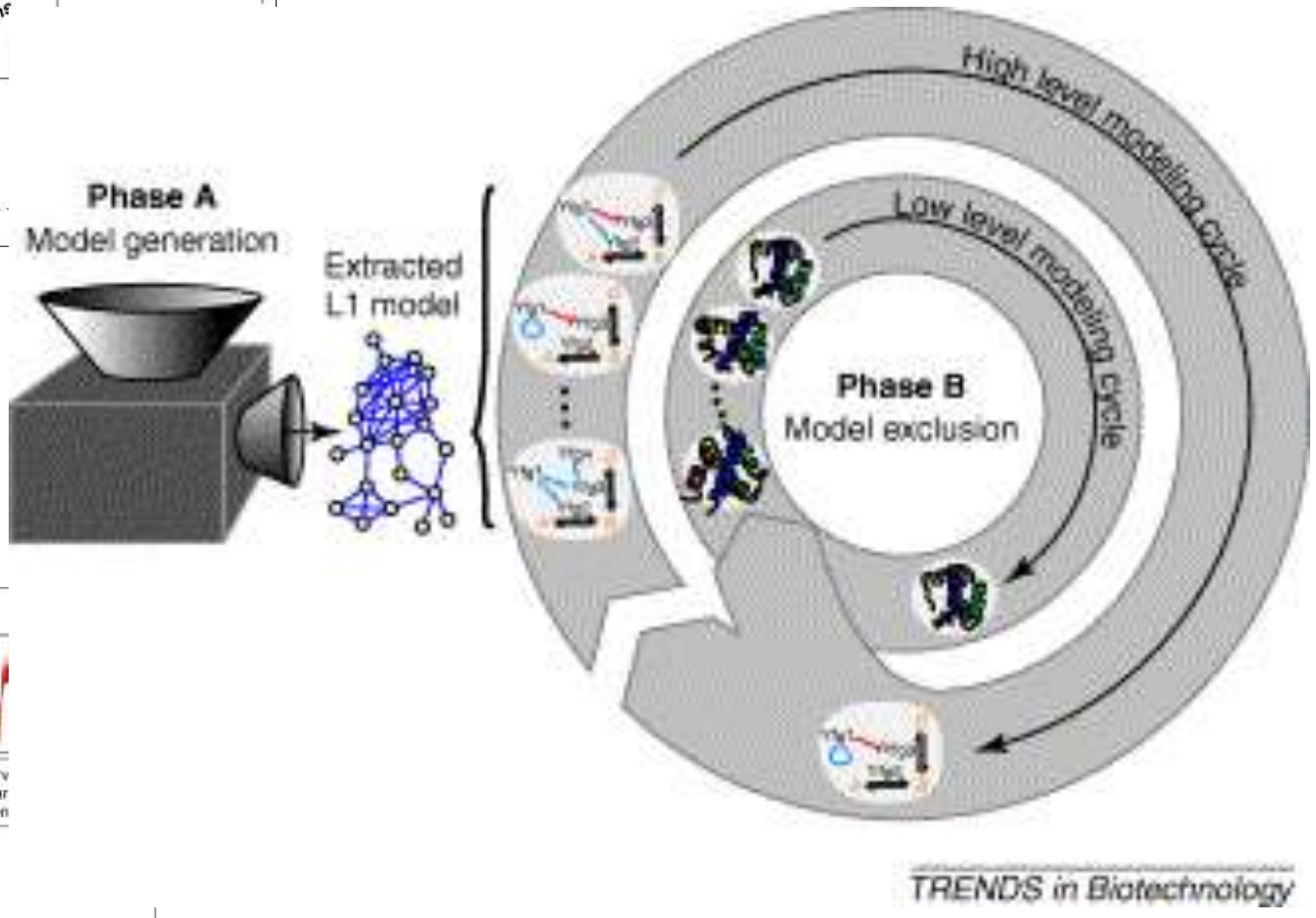
- Data Visualization
- Simulation tools- Iterative with model development

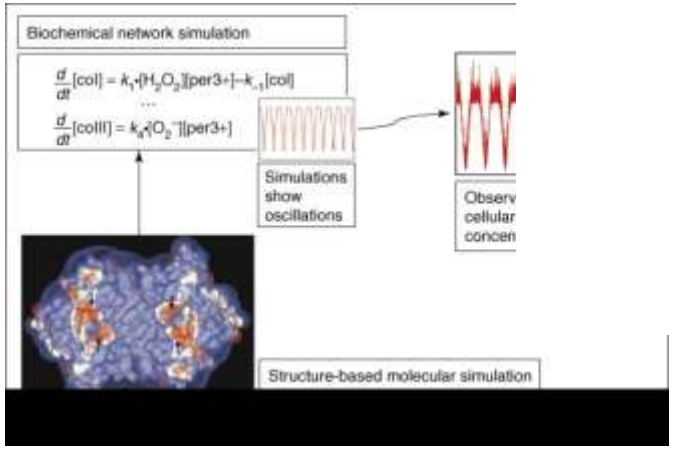
Encoding Inter-relationships



Integrated modelling







Abstraction of the flow of information that leads to Drug Resistance in TB bacilli

Problem Definition

Genome → Proteome

Network reconstruction

Source/ Sink identification

Concept of emergence of drug resistance

Ontologies
Protocols
Syntaxes
Data cross-mapping

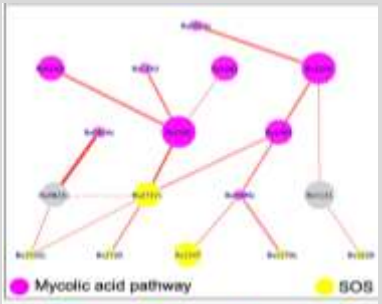
Math Description

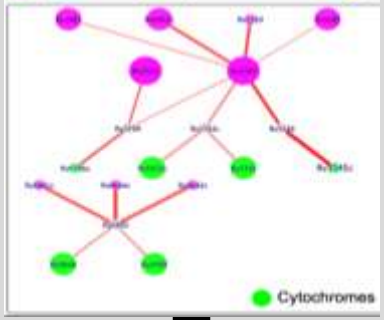
Simulations

Timestep, Mesh Size, Parameter Searches, Sensitivity

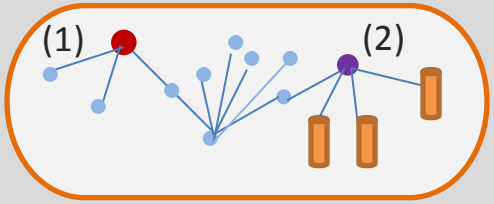
Results

Integrated models





Biological Insights



New Clues for Drug Discovery

Potential of Translation Systems Biology

- Patient GENOTYPE profile – molecular, Biochemical, life style, gene history
- Informatics- Identify disease, disease type, patient type, predict risk of patient for a given disease
- Understand molecular basis of disease, identify causative proteins, biochemical, signalling pathways
- Identify optimal strategy for tackling disease
- Choose best therapeutic intervention tool, drug, vaccine, other clinical tools
- Predict outcome of therapy with the chosen agent- PHENOTYPE
- Pharmacogenomcis, Personalized prescriptions
- Monitor patient, populations, Learn…