INTRODUCTION TO SYSTEMS BIOLOGY

Andrea Giansanti
Dipartimento di Fisica, Sapienza Università di Roma
(Ed. CU013, st.211 0649914367)
Andrea.giansanti@uniroma1.it

CB_23_24_L31, Roma 12 dec 2023 .

SAPIENZA
UNIVERSITÀ DI ROMA
Dipartimento di Management

DIPARTIMENTO DI FISICA

COMPUTATIONAL BIOLOGY
“Bioinformatics”
“machine learning
“Systems biology”
“Cellular networks”
“Computational cell biology”

v. corso di biofisica computazionale CB_intro.pdf, CB_quick.pptx

Organization of biological cells

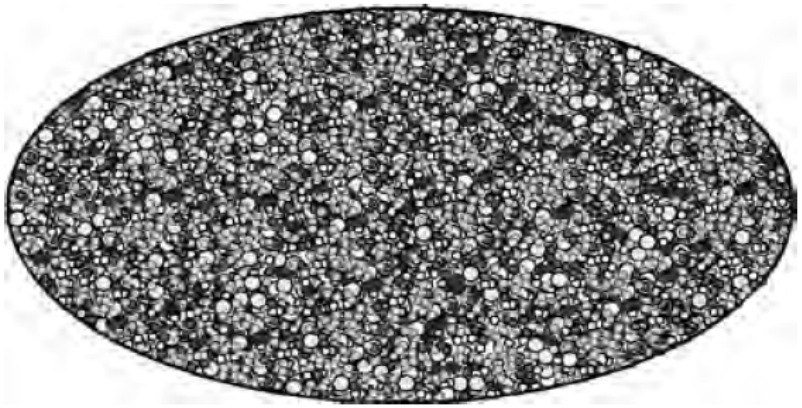


Figure 1.1 Is this how we should view a biological cell? This schematic picture makes an important point: about 30% of the volume of a biological cell is taken up by millions of individual proteins. Thus, biological cells are really "full". However, such pictures do not tell us much about the organization of biological processes and, as we will see later in this textbook, there are many different hierarchies of order in such a cell.

What is Systems Biology?

- *To understand biology at the system level, we must examine the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism. Properties of systems, such as robustness, emerge as central issues, and understanding these properties may have an impact on the future of medicine.*

Hiroaki Kitano

SBI The Systems Biology Institute

What is Systems Biology?

- *addresses the analysis of **entire** biological systems*
- ***interdisciplinary** approach to the investigation of all the components and networks contributing to a biological system*
- *[involves] new **dynamic computer modeling** programs which ultimately might allow us to simulate entire organisms based on their individual cellular components*
- *Strategy of Systems Biology is dependent on **interactive cycles of predictions and experimentation**.*
- *Allow[s Biology] to move from the ranks of a descriptive science to an **exact science**.*

(Quotes from SystemsX.ch website)

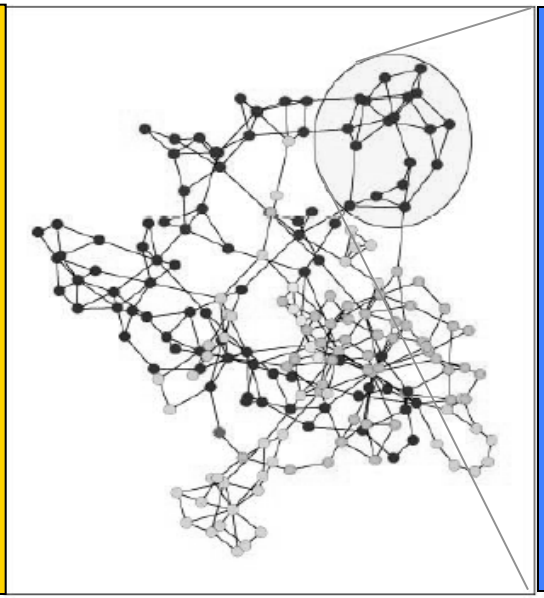
What is Systems Biology?

Ø **identify *elements* (genes, molecules, cells, …)**

Ø **ascertain their *relationships* (co-expressed, interacting, …)**

Ø ***integrate* information to obtain view of system as a *whole***

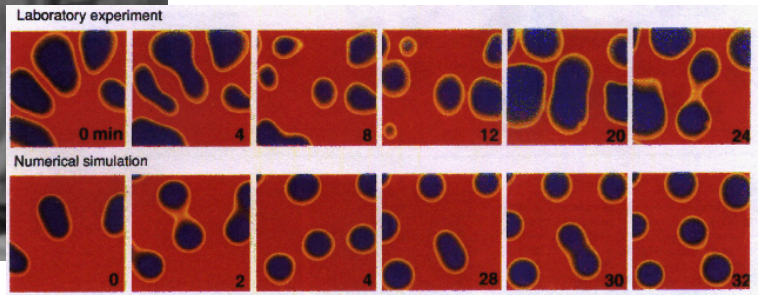
Large (genomic) systems
- many uncharacterized elements
- relationships unknown
- *computational analysis* should:
  - § improve annotation
  - § reveal relations
  - § reduce complexity



Small systems
- elements well-known
- many relationships established
- *quantitative modeling* of systems properties like:
  - § Dynamics
  - § Robustness
  - § Logics

Alan Turing: 1952

- "The Chemical Basis of Morphogenesis," 1952, *Phil. Trans. Roy. Soc. of London*, Series B: Biological Sciences, **237**:37—72.
- ***A reaction-diffusion model for development*.**



What is Systems Biology?

- *To me, systems biology seeks to explain biological phenomenon not on a gene by gene basis, but through the interaction of all the cellular and biochemical components in a cell or an organism. Since, biologists have always sought to understand the mechanisms sustaining living systems, solutions arising from systems biology have always been the goal in biology. Previously, however, we did not have the knowledge or the tools.*

Edison T Liu
Genome Institute of Singapore

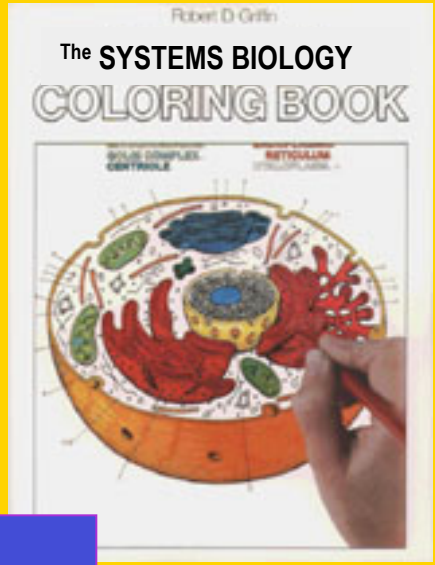
Systems Biology

Combining the mathematical rigor of numerology with the predictive power of astrology.

Numerlogy

Numeristan

Cyberia



HOTzone

Astrolog y

Astrostan

Infostan

Interpretive Biology

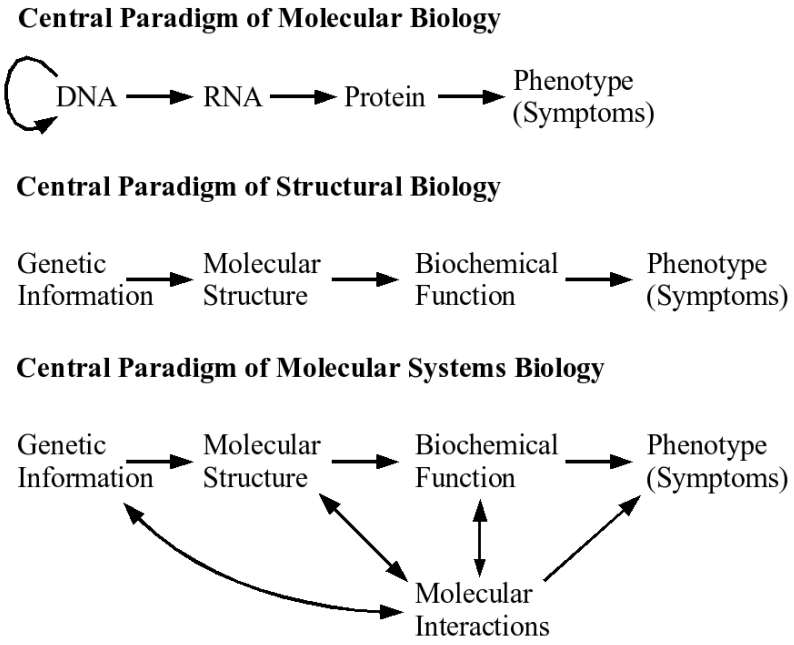
Integrative Biology

Computational Biology

Bioinformatics

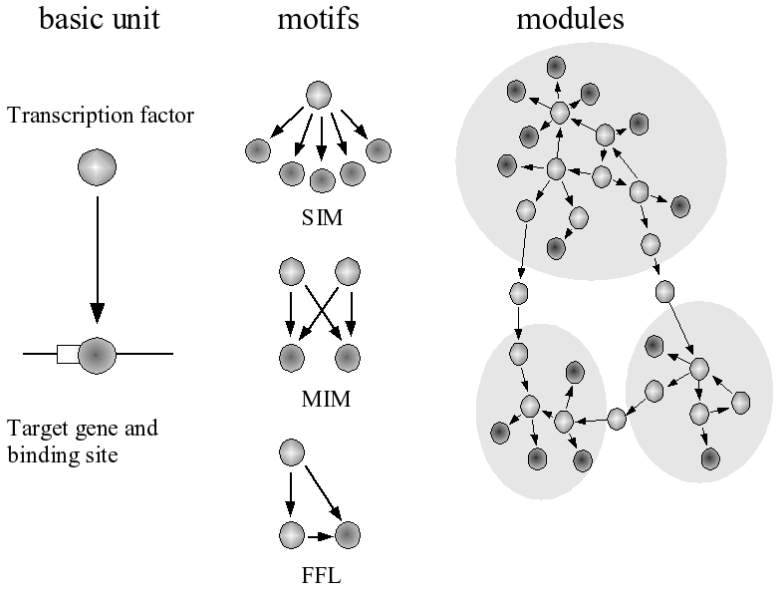
BioSpice

Organization of biological cells



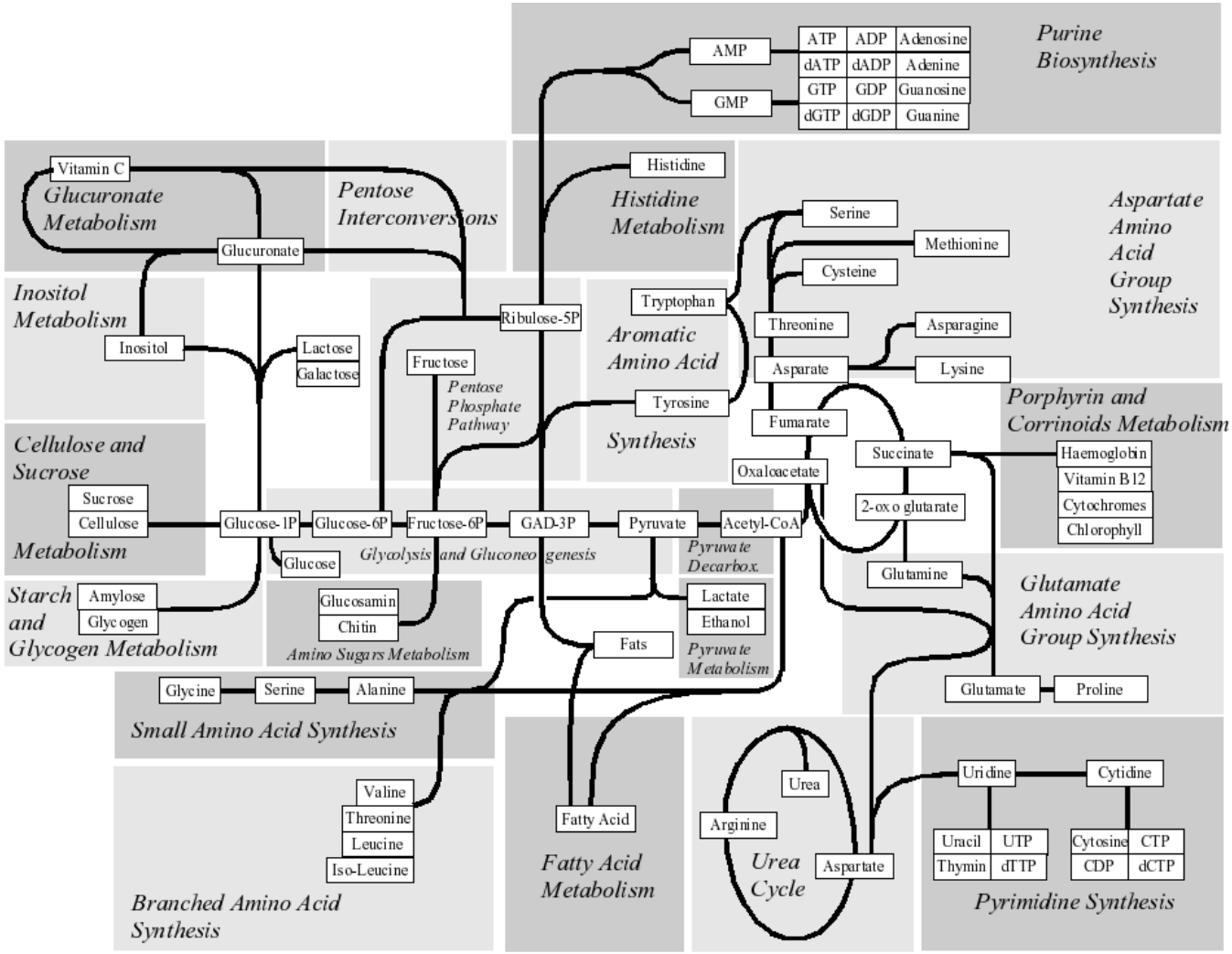
(Top) since the 1950ies, a paradigm became established that information flows from DNA over RNA to protein synthesis which then gives rise to particular phenotypes.
(Middle) The upcome of structural biology - the first crystal structure of the protein Myoglobin was determined in 1960 - emphasized the importance of the three-dimensional structures of proteins determining their function.
(Bottom) Today, we have realized the central role played by molecular interactions that influence all other elements.

Organization of transcriptional regulatory networks



(left) The 'basic unit' comprises the transcription factor, its target gene with DNA recognition site and the regulatory interaction between them.
(middle) Units are often organized into network 'motifs' which comprise specific patterns of inter-regulation that are over-represented in networks. Examples of motifs include single input multiple output (SIM), multiple input multiple output (MIM), and feed-forward loop (FFL) motifs.
(right) Network motifs can be interconnected to form semi-independent 'modules' many of which have been identified by integrating regulatory interaction data with gene expression data, and imposing evolutionary conservation. The next level consists of the entire network (not shown).

Major metabolic pathways



Content (ca.)

Protein interaction networks:

- different topologies (random networks, scale-free networks)
- intro of protein complexes: exp. data
- computational analysis (mathematical graphs)
- graphical layout (force minimization)
- quality check (Bayesian analysis)
- modularity
- network flow

Transcriptional regulatory networks, motifs

Signal transduction networks

Metabolic networks: metabolic flux analysis, extreme pathways, elementary modes

FFT protein-protein docking, fitting into EM maps, tomography

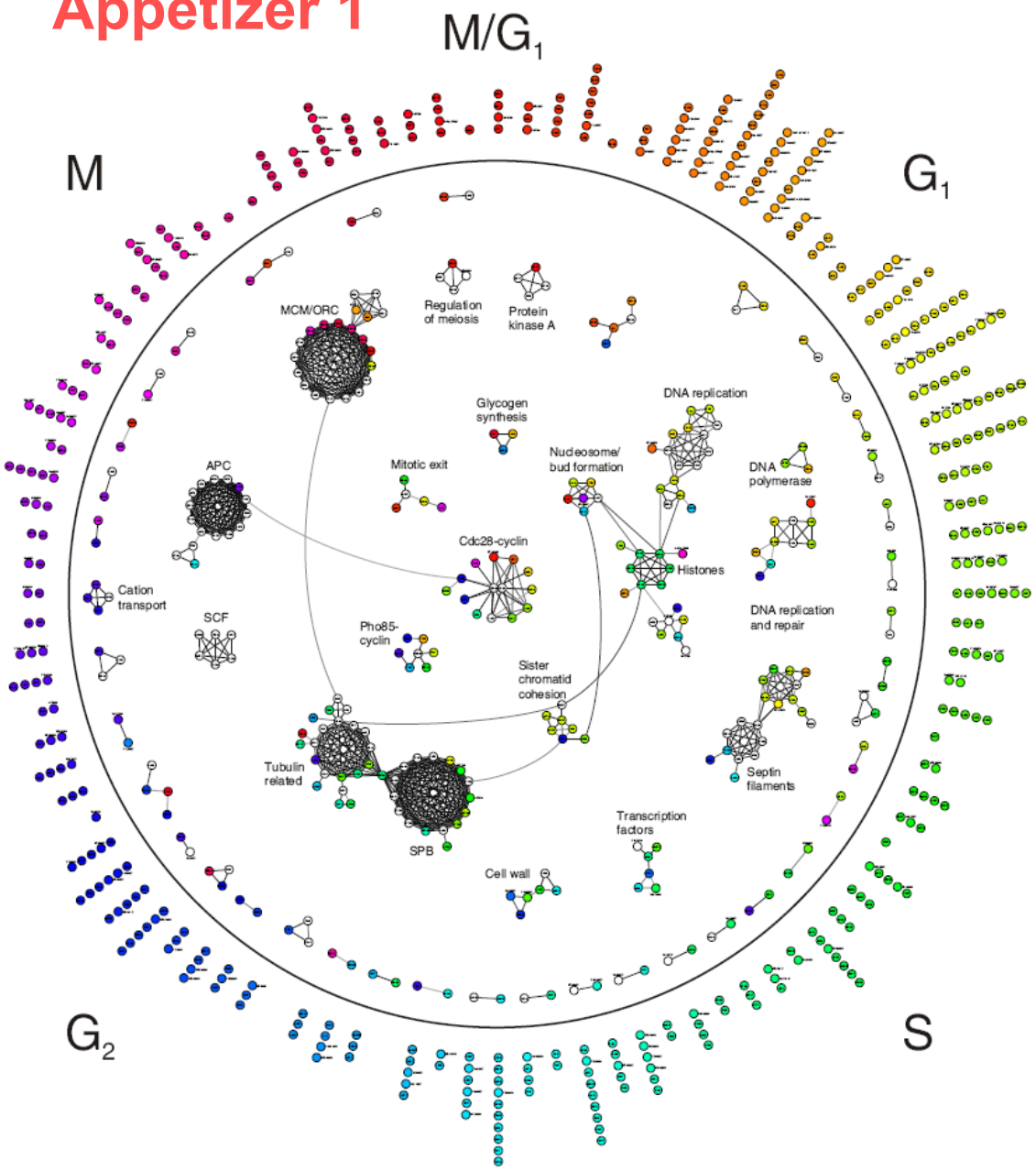
integration of interactome and regulome (Lichtenberg),
integration of protein networks with metabolic pathways

Appetizer 1

Cell cycle proteins that are part of complexes or other physical interactions are shown within the circle.

For the dynamic proteins, the time of peak expression is shown by the node color; static proteins are represented as white nodes.

Outside the circle, the dynamic proteins without interactions are positioned and colored according to their peak time.

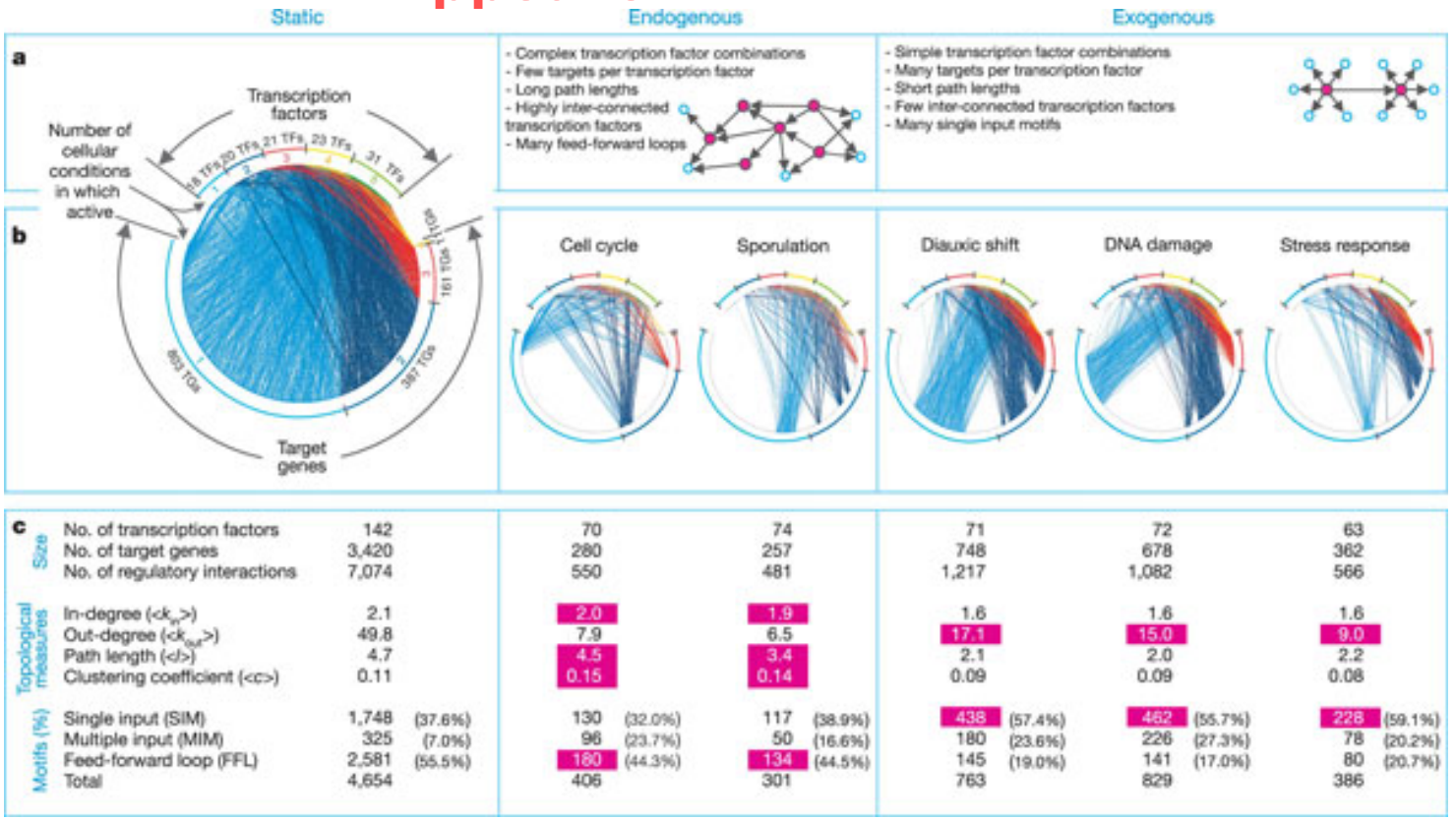


Lichtenberg *et al.* Science 307, 724 (2005)

Appetizer 2

a, Schematics and summary of properties for the endogenous and exogenous sub-networks.

b, Graphs of the static and condition-specific networks. Transcription factors and target genes are shown as nodes in the upper and lower sections of each graph respectively, and regulatory interactions are drawn as edges; they are coloured by the number of conditions in which they are active. Different conditions use distinct sections of the network.



c, Standard statistics (global topological measures and local network motifs) describing network structures. These vary between endogenous and exogenous conditions; those that are high compared with other conditions are shaded. (Note, the graph for the static state displays only sections that are active in at least one condition, but the table provides statistics for the entire network including inactive regions.)

Luscombe, Babu, … Teichmann, Gerstein, Nature 431, 308 (2004)

Mathematical techniques covered

Mathematical graphs – classification of protein-protein interaction networks,
– algorithms on graphs
– regulatory networks

Bayesian networks – protein interaction networks

Boolean networks – transcriptional networks

Fourier transformation – protein/protein-docking, pattern matching

Linear and convex algebra
– metabolic networks

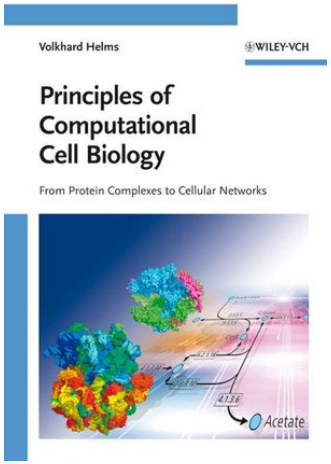
Ordinary and **stochastic differential equations**
– kinetic modelling of signal transduction pathways

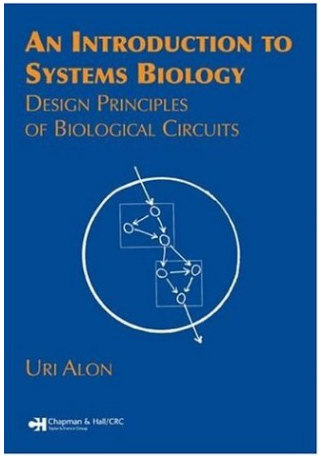
Literature

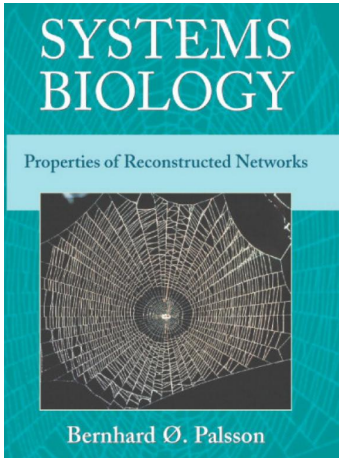
lecture slides are available already; final version 3 days before lecture

suggested reading: links will be put up on course website
http://gepard.bioinformatik.uni-saarland.de/teaching...

Text books







Systems biology

Biological research in the 1900s followed a reductionist approach:

detect unusual phenotype -> isolate/purify 1 protein/gene, determine its function

However, it is increasingly clear that discrete biological function can only rarely be attributed to an individual molecule.

->new task of understanding the structure and dynamics of the complex intercellular web of interactions that contribute to the structure and function of a living cell.

Systems biology

Development of high-throughput data-collection techniques,
e.g. microarrays, protein chips, yeast two-hybrid screens
allow to simultaneously interrogate all cell components at any given time.

-> there exists various types of interaction webs/networks
- protein-protein interaction network
- metabolic network
- signalling network
- transcription/regulatory network ...

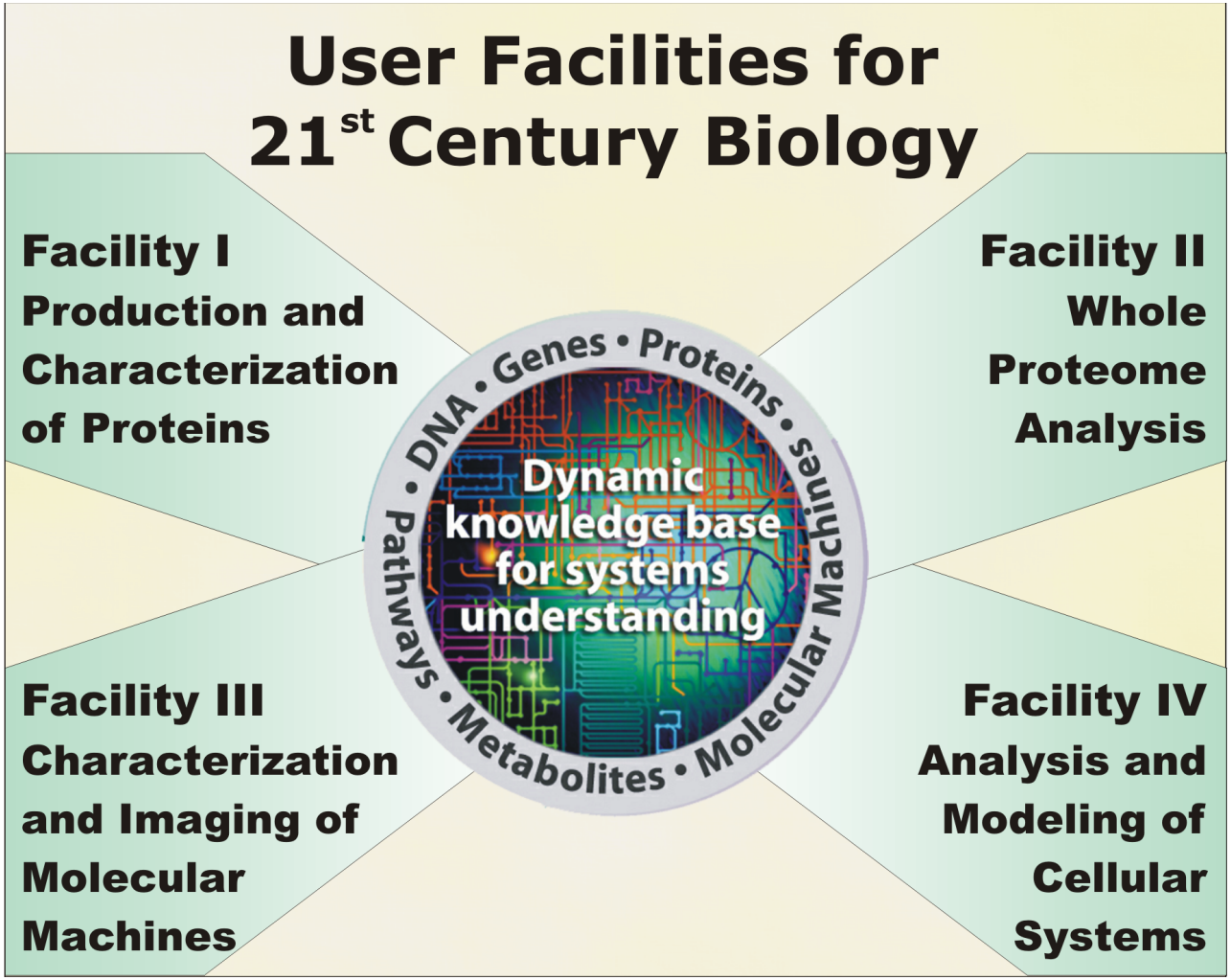
These networks are not independent but form „network of networks“.

DOE initiative: Genomes to Life
a coordinated effort

slides borrowed
from talk of

Marvin Frazier

Life Sciences Division
U.S. Dept of Energy

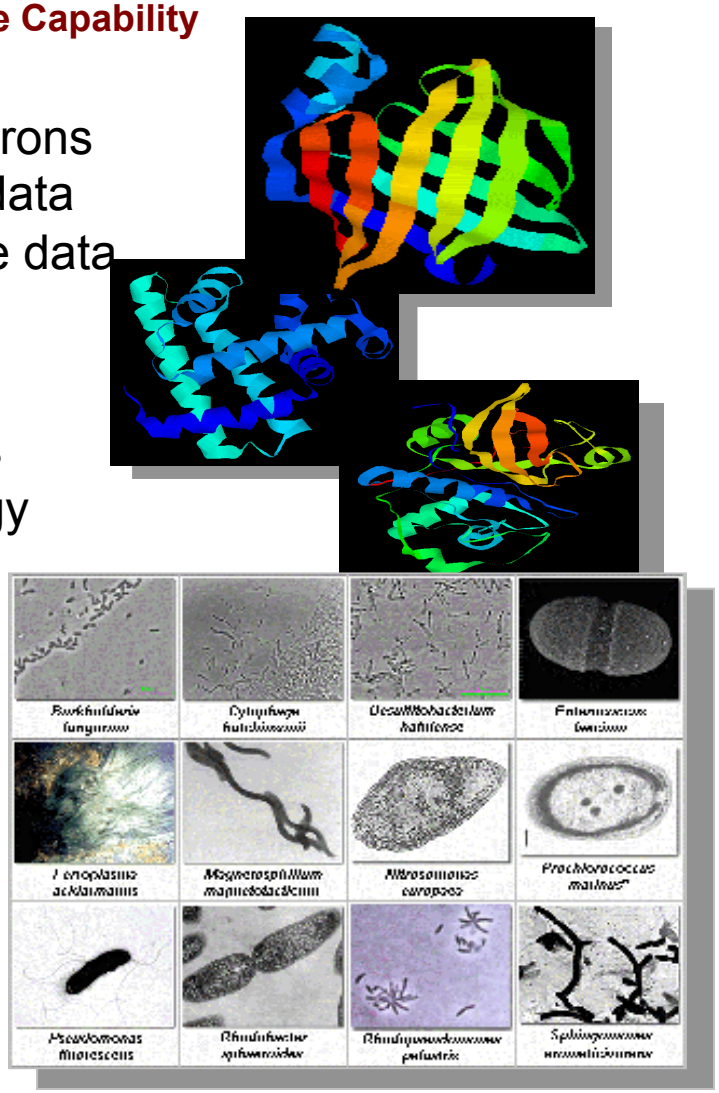


Facility I

Production and Characterization of Proteins

Estimating Microbial Genome Capability

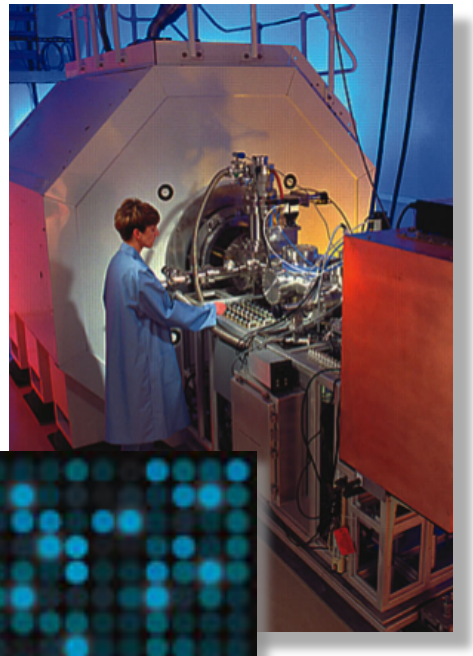
- Computational Analysis
  - Genome analysis of genes, proteins, and operons
  - Metabolic pathways analysis from reference data
  - Protein machines estimate from PM reference data
- Knowledge Captured
  - Initial annotation of genome
  - Initial perceptions of pathways and processes
  - Recognized machines, function, and homology
  - Novel proteins/machines (including prioritization)
  - Production conditions and experience



Facility II

Whole Proteome Analysis

Modeling Proteome Expression, Regulation, and Pathways



- Analysis and Modeling
  - Mass spectrometry expression analysis
  - Metabolic and regulatory pathway/ network analysis and modeling

- Knowledge Captured
  - Expression data and conditions
  - Novel pathways and processes
  - Functional inferences about novel protein machines
  - Genome super annotation: regulation, function, and processes (deep knowledge about cellular subsystems)

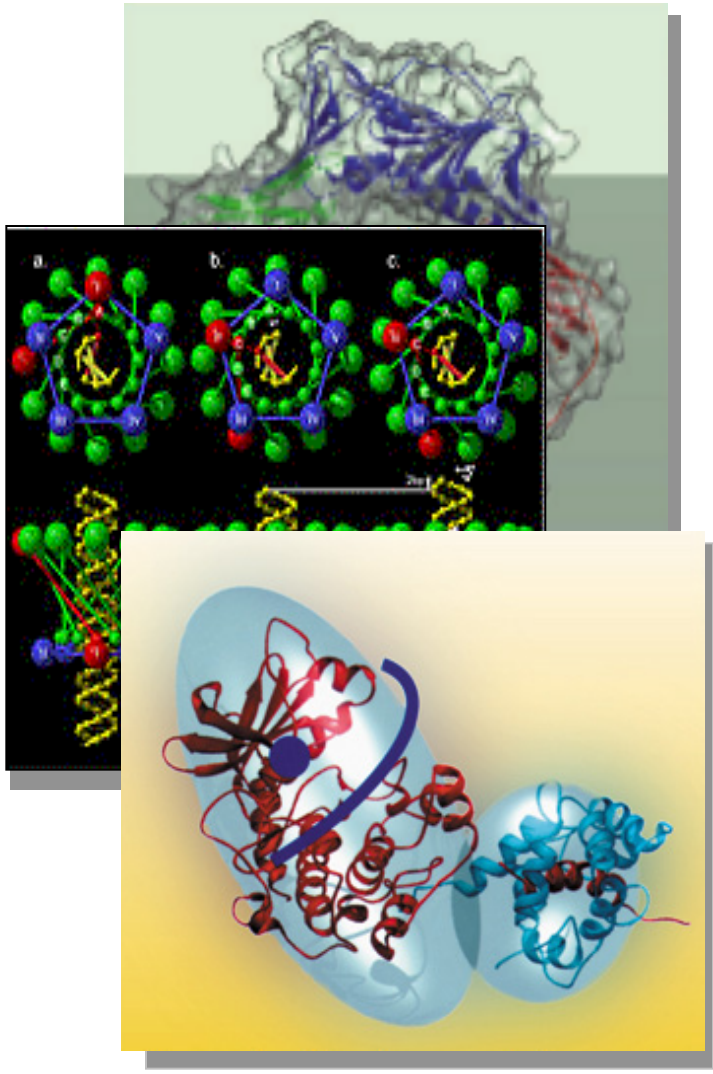
Facility III

Characterization and Imaging of Molecular Machines

Exploring Molecular Machine Geometry and Dynamics

- Computational Analysis, Modeling and Simulation
  - Image analysis/cryoelectron microscopy
  - Protein interaction analysis/mass spec
  - Machine geometry and docking modeling
  - Machine biophysical dynamic simulation

- Knowledge Captured
  - Machine composition, organization, geometry, assembly and disassembly
  - Component docking and dynamic simulations of machines

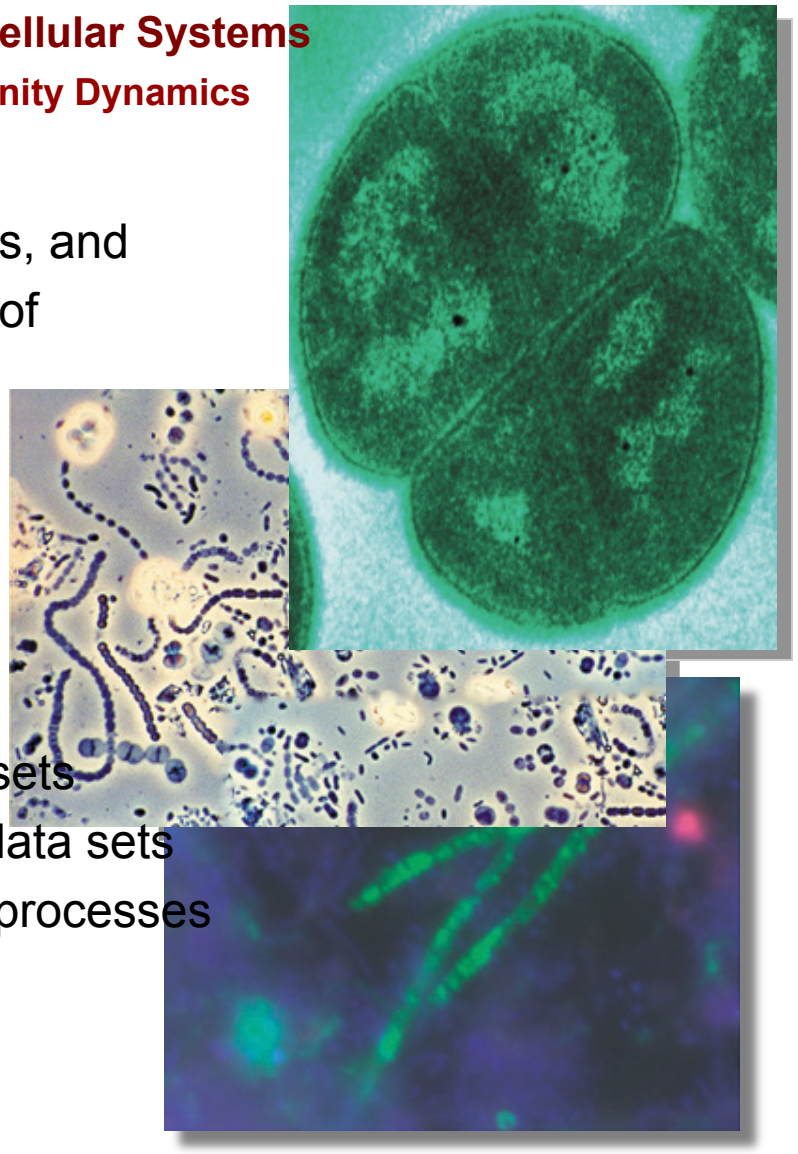


Facility IV

Analysis and Modeling of Cellular Systems

Simulating Cell and Community Dynamics

- Analysis, Modeling and Simulation
  - Couple knowledge of pathways, networks, and machines to generate an understanding of cellular and multi-cellular systems
  - Metabolism, regulation, and machine simulation
  - Cell and multicell modeling and flux visualization
- Knowledge Captured
  - Cell and community measurement data sets
  - Protein machine assembly time-course data sets
  - Dynamic models and simulations of cell processes



“Genomes To Life” Computing Roadmap