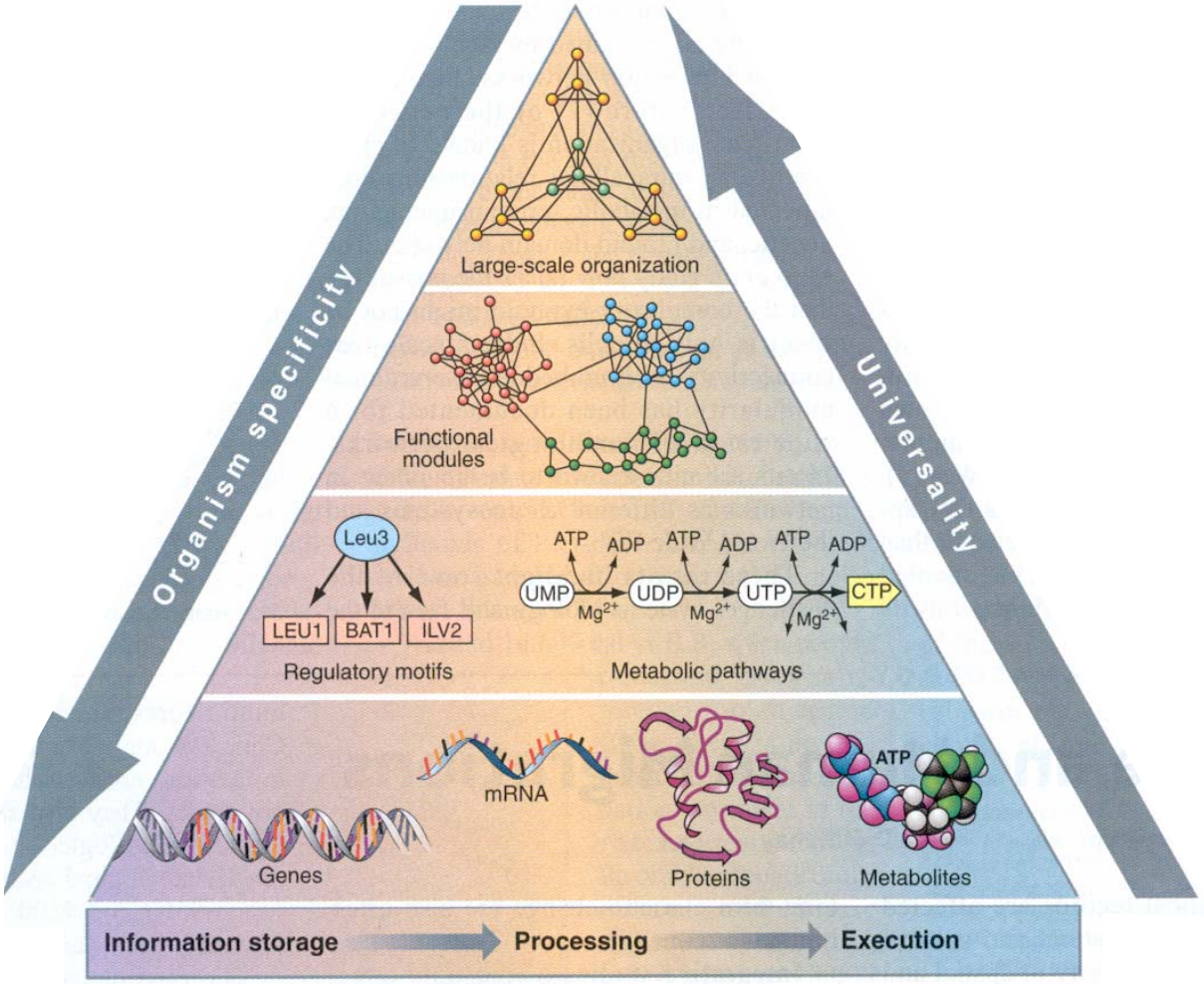
I. Dynamics of living systems

- Understanding the dynamics at the **molecular** level.
- Understanding the dynamics at the **cellular** level
- **Filling the gap** between these two levels

Dynamics ←→ Function

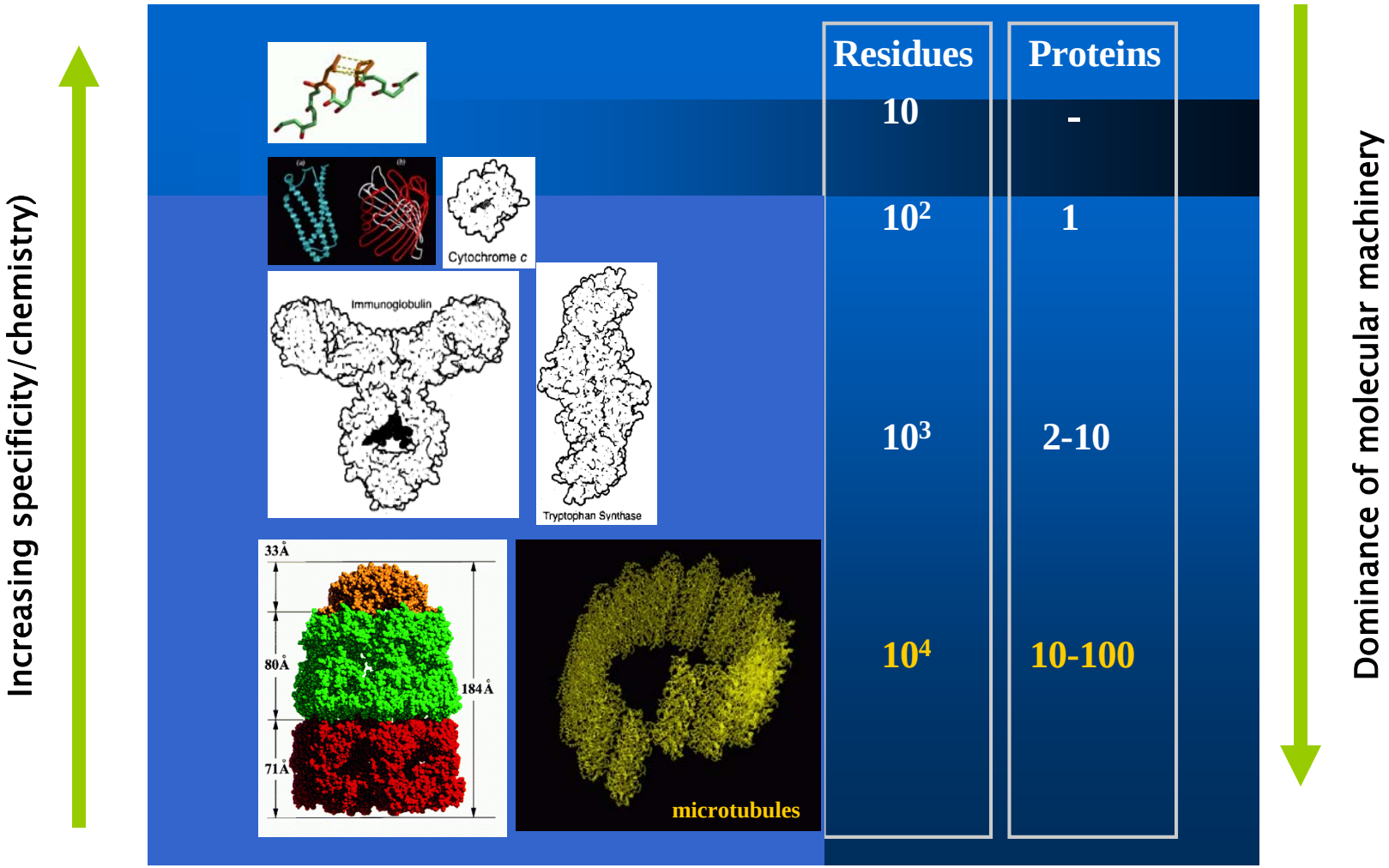
Life's complexity pyramid



Oltvai & Barabasi, Science 2002, 298, 763-764

"The complexity pyramid might not be specific only to cells"

Different levels of structural organization:

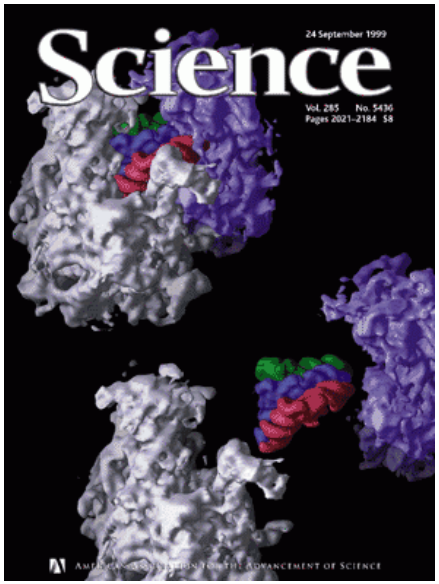


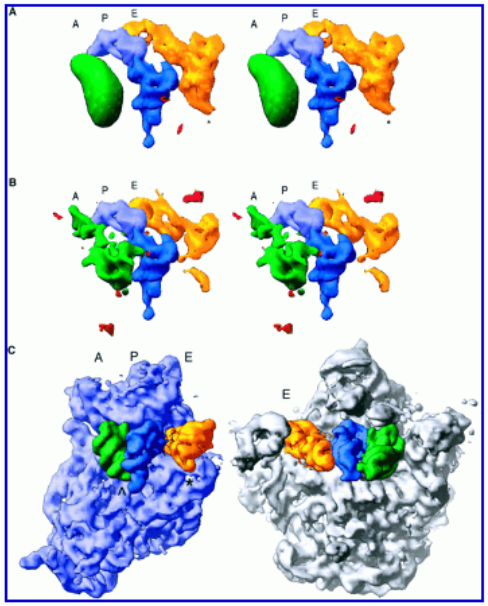
Large structures

- **Ribosomal functional complexes**

  (Cate, Yusupov, Yusupova,Earnest & Noller, Science 1999, 285, 2095).

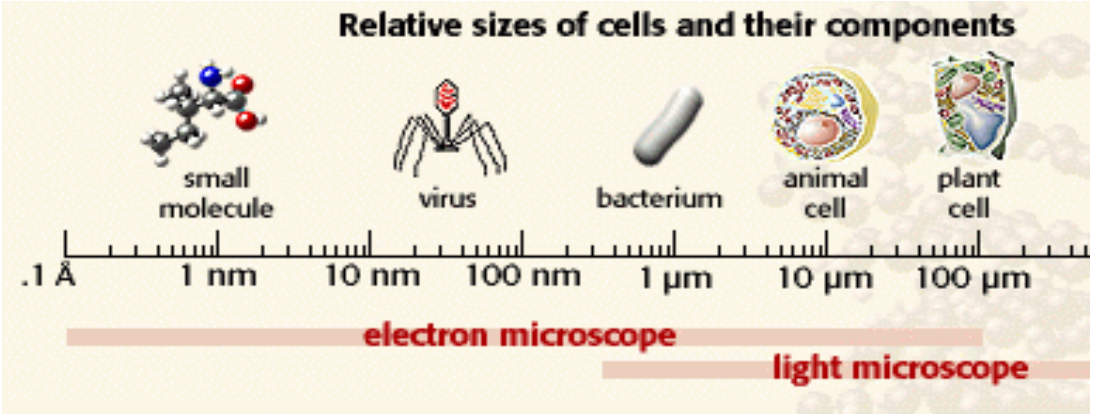
- **Complexes formed by several proteins, cascades of interactions**





Computational models & methods for;

- Large assemblies, complexes of proteins
- Membranes and cytosolic fibrous systems
- Cellular pathways. Signaling & regulation of cell cycle



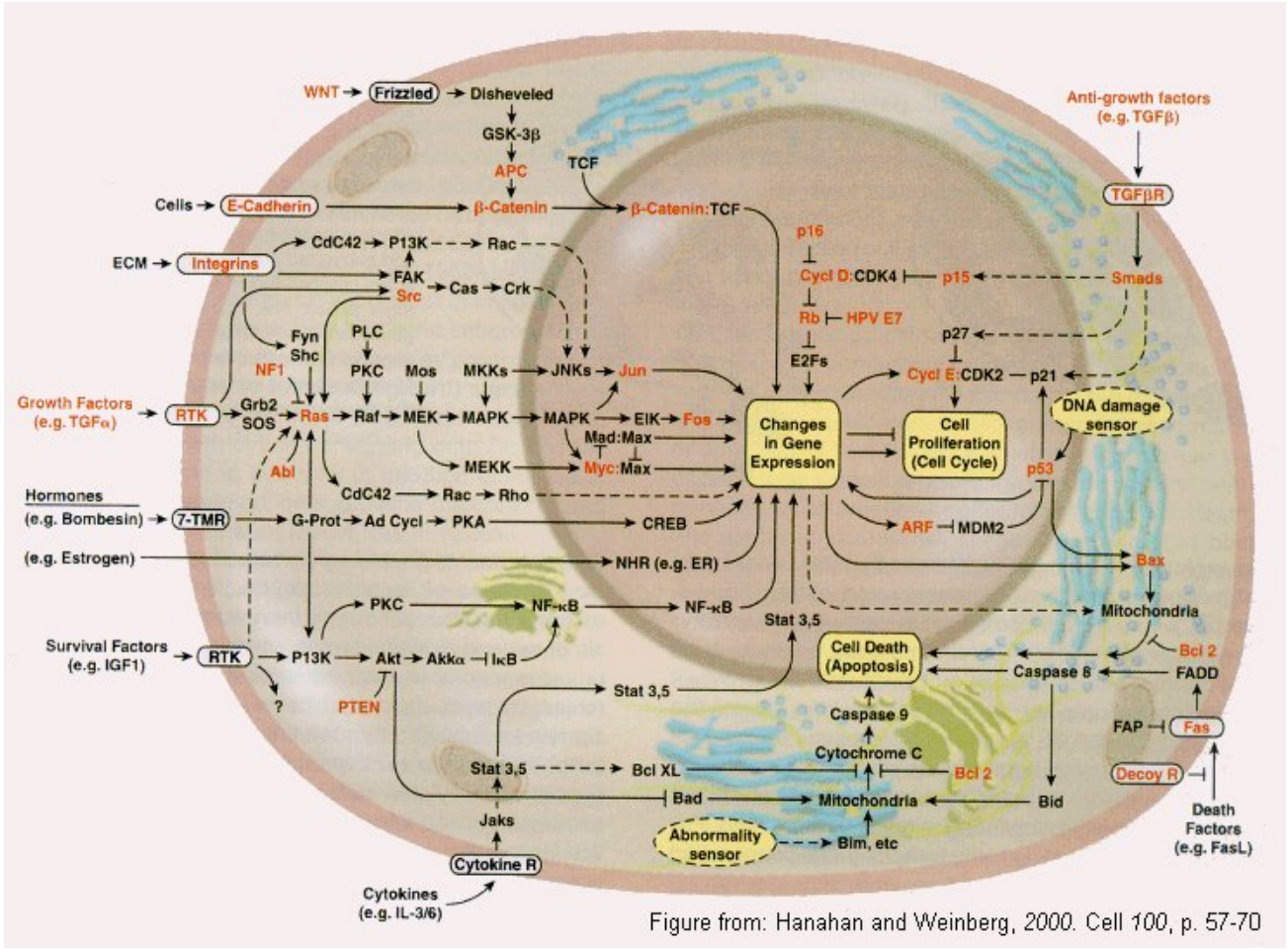


Figure from: Hanahan and Weinberg, *2000.* Cell *100*, p. 57-70

Bridging the gap between Molecular and Cellular scales

- What is the **optimal** (realistic, but computationally efficient) model **for a given scale** (length and time) of representation?

- Which level of **details** is needed for representing **global** (collective) motions?

- How much **specificity** we need for modeling **large** scale systems and/or motions?

- What should be the **minimal** **ingredients** of a **simplified** (reductionist) model?

Computational challenges

- Realistic modeling of the space-time dependence of cellular processes
- Multiscale representations of structure/dynamics
- Extracting rules from microscopic simulations, for macroscopic approaches
- Bridging between continuous and discrete models

Systems Biology

There is not yet consensus on what systems biology actually is!

"The analysis of networks, regulation, how the things work from a whole system point of view" *Sauro*

"physiology of cells" *Adam Arkin*

"mathematical modeling of biological systems" *Schneider*

Extensive usage of math and computations

An old concept that became popular with

- the sequencing of the human genome
- genomics, proteomics, metabolomics concepts
- microarray technologies, instrumentation that allow for high throughput measurement of DNA, RNA and proteins – global data sets

Mathematical Modeling of Cellular Networks

Numerous approaches have been used starting from the work of Jacob & Monod. On one end is **to include as many details as possible**. However, in many cases, the rates of reactions and the original concentrations are not known, nor are all the intermediate states and connectivities *(state/block diagram)*; and the complexity of the system makes it difficult to study sensitivity to parameters and initial conditions. On the other extreme is the abstract approach taken by Glass and Kaufmann where the individual components are taken to be **Boolean variables** (either on or off) and the behavior is completely determined by the topology of the interactions and the switching rules. While this greatly simplifies the models, the graded nature of responses is often important.

Methods for simplifying and reducing complex models essentially exploit differences in time and spatial scales, when these scales are separable. Many of the systems are inherently stochastic and we utilize a *master equation* formalism where transition (or jump) probabilities between states control the probabilistic evolution of states. Closely associated are the hybrid models that involve stochastically forced differential equations (Langevin dynamics), or the time evolution of probability density functions (Fokker-Planck formalism). These approaches have been successfully used in other disciplines, but have not yet been exploited by theoretical biologists. Deterministic differential equations for mean concentrations can be derived directly from the stochastic formulations.

The differential equation description can be further reduced to produce simpler models that still capture the essential properties of the system. An example is the *pseudo-steady state approximation* for the reactants and products of the fast steps in serial reactions (e.g. Michealis-Menten mechanism).

The solution of averaged differential equations in terms of the slow variables, as used in neural networks or weakly coupled oscillators, is another mathematical tool for model reduction.

Mathematical tools such as PCA for decomposing the dynamics into its different modes, filtering out the noise or reconstructing the dominant pathways are also useful. For example, the eigenvector corresponding to the zero eigenvalue of the transition matrix in the master equation formalism yields the steady state probabilities of the individual components of the system, while the eigenvector associated with the smallest eigenvalue extracts the slowest (or least probable) passage.

.

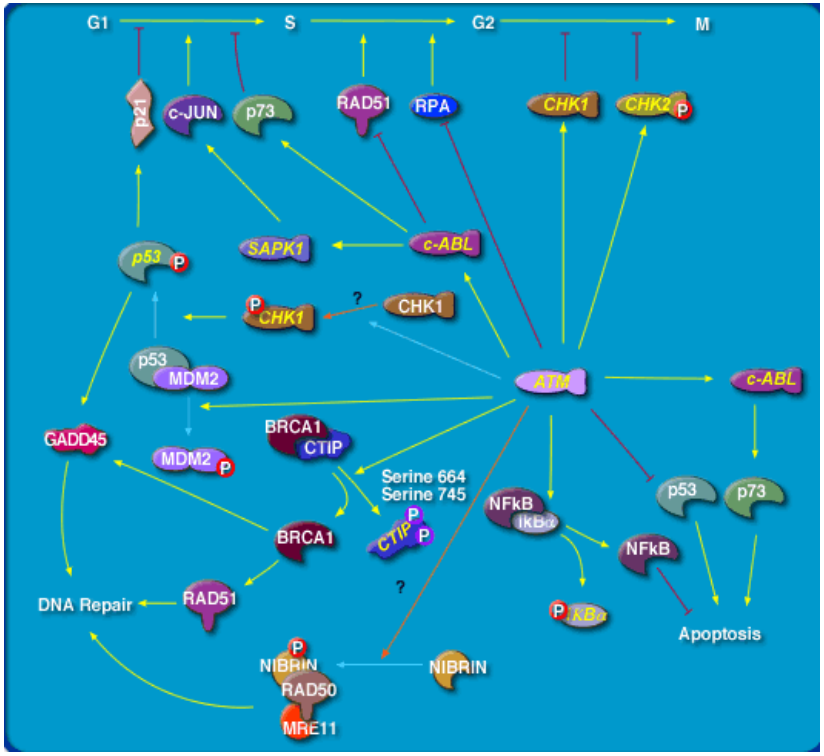
Two challenges:

- To ensure that the important details are not being neglected,
- to be able to construct simplified models that are *quantitative*, rather than just *qualitative*.

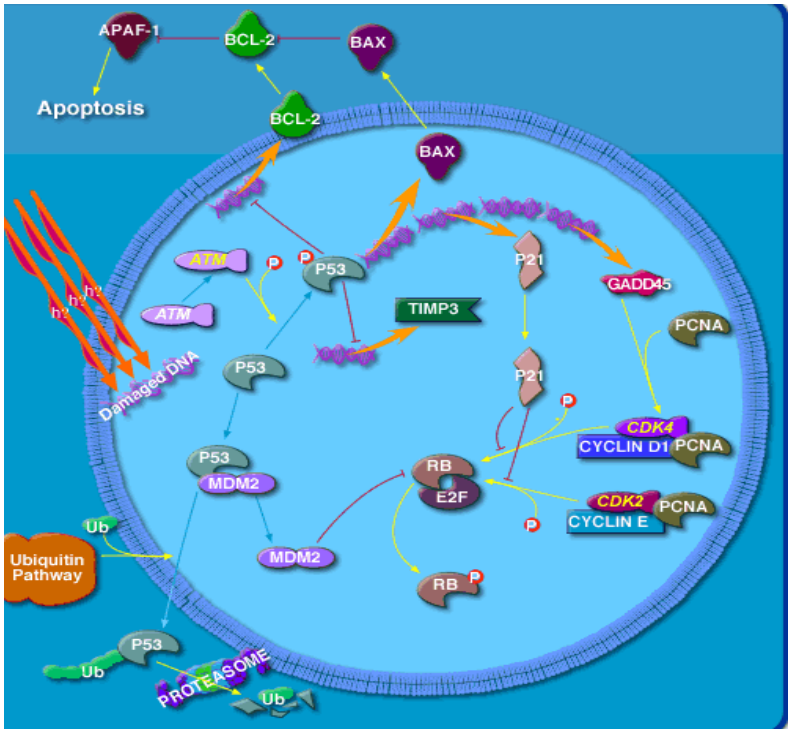
Celllular pathways are usually described by simple

Mass-Action Kinetics

ATM signaling

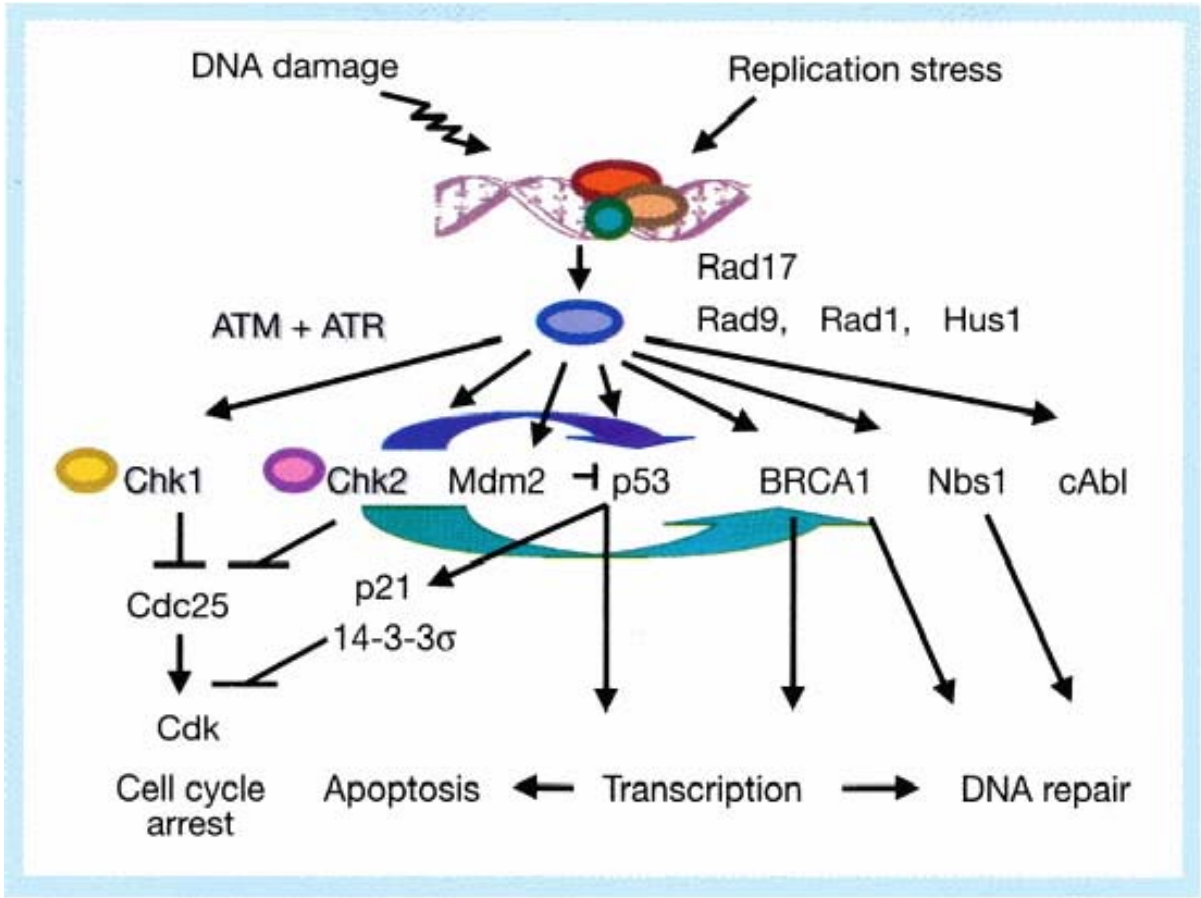


P53 regulation



http://www.biocarta.com/

DNA damage signaling pathways



A shift from

“ list of genes” or “proteins”

to

“structure” and “dynamics”

Three components of the computational approach:

- System structure (network of gene interactions, biochemical pathways, etc.)
- Properties of the components (reactivities, binding affinities, etc.)
- System dynamics (sensitivity analysis, bifurcation analysis)

System = Genome, proteome, metabolome, etc.

The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 19, 524-531, 2003.

Control and Dynamical Systems, MC 107-81, California Institute of Technology, Pasadena, CA 91125, USA. sysbio-team@caltech.edu

Existing problems

1. Lack of information on biochemical networks or cell signaling and regulation pathways.
   (Expression patterns indicate co-expressed genes, that are not necessarily co-regulated, or involved in the same pathway or function. Correlation between genes does not always provide information on causality
2. Define the scope and abstraction level of the model, which depends on the available knowledge. Lack of quantitative data on rate constants, concentrations)
3. Robustness is an essential property, which may be induced by adaptation, parameter insensitivity, negative feedback loops or feed-forward control mechanisms, structural stability or modularity.

Kirchhoff/connectivity matrix is analogous to the

Transition Rate Matrix

of

Master Equation Formalism

Applications to macromolecular dynamics:

- Rotational dynamics of polymers (DRIS)
  (Bahar, Erman & Monnerie, Adv Polym Sci. 1994)
- Folding dynamics of model proteins
  (Ozkan, Bahar & Dill, Nature Struct Biol. 2001)

How to analyze the time evolution of macrostates?

- **Master Equation formalism**
- $d\mathbf{P(t)}/dt = \mathbf{A}P(t)$

- Formal solution:

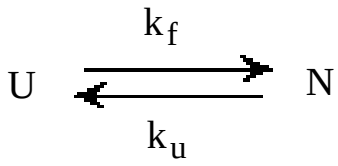
  $P(t) = \exp\{-\mathbf{A}t\}\, P(0) = \mathbf{B}\, \exp\{\Lambda^{-1}\}\, \mathbf{B}^{-1}\, \mathbf{P}(0)$

  (transition probability matrix)

Classical kinetic modelling of protein folding/unfolding

Two-state transition

The simplest type of transition between states U and N is a two-state process, given by the scheme



The differential rate expressions holding in this case are

$$d[U]/dt = - k_f [U] + k_u [N]$$

$$d[N]/dt = + k_f [U] - k_u [N]$$

where [U] and [I] are the instantaneous (time-dependent) concentrations of the unfolded and folded conformations, respectively, and k_f and k_u are the folding and unfolding rate constants

Initial concentrations = $[U]_0$ and $[N]_0$.

In folding experiments, we take $[N]_0 = 0$, and the instantaneous concentration [N] is given by

$$[N] = [U]_0 - [U]$$

such that the differential folding equation reduces to a non homogeneous, first order differential equation

$$d[U]/dt = - (kf + ku) [U] + ku [U]_0$$

The solution is:

$$[U] / [U]_0 = \frac{k_u}{k_f + k_u} + \frac{k_f}{k_f + k_u} \exp \{-(k_f + k_u)t\}$$

The equilibrium concentrations define the equilibrium constant for the folding reaction

$$K_{UN} = \frac{[N]_\infty}{[U]_\infty} = \frac{k_f}{k_u}$$

The equilibrium constant is related to the free energyof unfolding by the equation

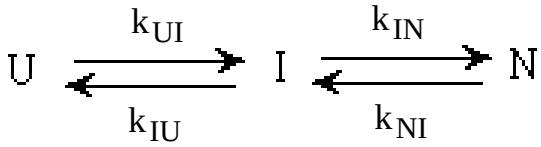
$$\Delta G_{UN} = - RT \ln K_{UN}$$

Several proteins have been observed to obey such a two-state transitions.

Sequential transition from U to N

The transition from U to N has been shown in numerous examples above to proceed through the formation of one or more intermediates.

Let us consider here the simpler case of a single intermediate. Let k_{XY} designate the rate constant for the passage from state X to state Y. Using this notation,



The set of equations for the differential change

$$\begin{bmatrix} d[U]/dt \\ d[I]/dt \\ d[N]/dt \end{bmatrix} = \begin{bmatrix} -k_{UI} & k_{IU} & 0 \\ k_{UI} & -k_{IU}-k_{IN} & k_{NI} \\ 0 & k_{IN} & -k_{NI} \end{bmatrix} \begin{bmatrix} [U] \\ [I] \\ [N] \end{bmatrix}$$

In concise notation,

$$d\mathbf{X}(t)/dt = \mathbf{A}\,\mathbf{X}(t)$$

where **X**(t) is the vector of the instantaneous concentrations, and **A** is the matrix of rate constants, shortly referred to as *rate matrix*. This matrix equation is similar in form, to *a master equation*, where concentrations are replaced by probabilities.

The set of coupled differential equations is conveniently found by matrix algebra methods, using the similarity transformation

$$\mathbf{A} = \mathbf{B}\,\Lambda\,\mathbf{B}\text{-}1$$

Here **B** is the matrix of eigenvectors of **A**, and Λ is the diagonal matrix of eigenvalues. The instantaneous concentrations/probabilities are controlled by

$$\mathbf{X}(t) = \mathbf{B}\ \exp\{\Lambda t\}\ \mathbf{B}\text{-}1\ \mathbf{X}(0)$$

This equation may be rewritten in explicit notation for each state i (Xi = [U], [I]or [N]) as

$$Xi(t) = \sum_k \sum_j B_{ik}\ \exp\{\lambda_k t\}\ \ B^{-1}{}_{kj}\ X_j(0)$$

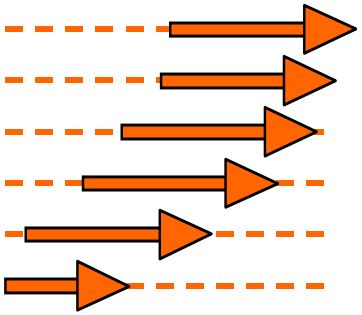
where the subscript denote the particular elements of the matrices, or vectors, and the summations are carried over all elements.

The last equation is similar in form to the *multiexponential form* generally postulated for describing complex processes.

Remarks

- Combination of computations, experiments and theory is vital
- Modeling is an integral part of systems biology → building the networks of interactions and examining its dynamics
- Top-down approach (for modeling diseases, subcellular processes – apoptosis -, or entire ‘silicon cells’)
- most of the info/data needed for modeling is in the text of scientific literature, not in databases or equations
- Has potential to impact drug discovery and development timeline (focus on the connection between molecular and physiological)

Size and Time Scales Applicable to Molecular & Cellular Control Mechanisms

Spatial Scale		Relevant Time Scale
Multicellular	cm	*fs ns μs ms s min hr*
Cellular	mm	
Subcellular	μm	
Multimolecular	nm	
Molecular	Å	
Atomic		

Major challenges of the post-genomic era include a detailed understanding of structure/function relationships and complex interactions for proteins and their assemblies **at various scales**. Computational and mathematical models and simulations become increasingly important to delineate not just the average behavior of biological systems, but also the systems' variability and propensity to switch operating modes and/or to fail.

The range of spatial and temporal scales over which molecular and cellular processes vary is enormous (Table). A variety of theoretical and computational methods have been developed for problems at various scales, and arguably the most mature algorithms underlie software at the lower and upper extremes. At the lower level are molecular dynamics (MD) simulations, which provide information at the atomic scale. In practice, however, their high computational cost precludes space and time scales beyond a few hundreds of residues and nanoseconds. The accuracy of MD simulations is limited by that of the adopted force field, and the simulations usually suffer from incomplete sampling of conformation space. At the higher level of cellular/multicellular processes, on the other hand, are methods based largely on empirical conjugate forces and flows, and involving the simultaneous solution of coupled ordinary or partial differential equations (ODEs or PDEs). In space-free or (relatively) simple compartmental models of biochemical networks, these methods can address biological processes on the time-scale of minutes to hours. Spatial information can be incorporated therein, but at the cost of dramatically increased computation time, because the space must be subdivided into finite elements and the coupled PDEs must be solved for each. Additionally, stochastic effects, which *in vivo* may contribute to the robust nature of the organism, but may also account for switching into disease states, are usually lacking at this level.

Most of functional cell physiology lies in between the two spatio-temporal extremes outlined above, i.e., a finite group of molecules subject to complex structural and spatial arrangements are coupled via stochastic and/or directed interactions driving cellular machinery.