Introduction to System Biology

Yinghao Wu
Department of Systems and Computational Biology
Albert Einstein College of Medicine
Fall 2014

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Outline

- Basic concepts
- Theoretical principles
- Experimental techniques

Outline

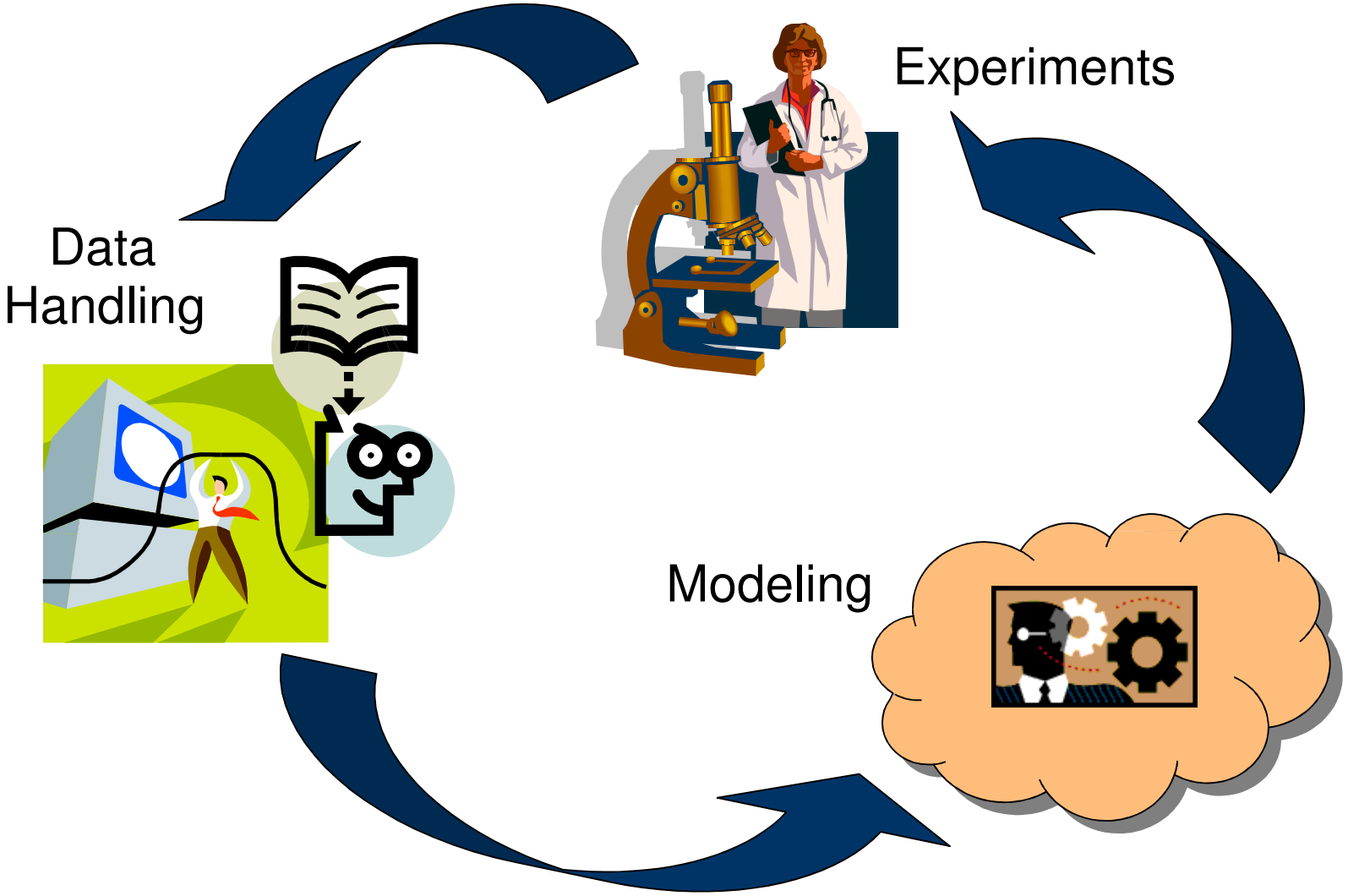
- Basic concepts
- Theoretical principles
- Experimental techniques

What is Systems Biology?

A systems biology approach means

- Investigating the components of cellular networks and their interactions
- Applying experimental high-throughput and whole-genome techniques
- Integrating computational and theoretical methods with experimental efforts

An Iterative Approach



Main Related Disciplines

- Biology
- Biotechnology
- Mathematics and Statistics
- Physics and Chemistry
- Information Science
- Engineering (Biomedical, Chemical, Electronic)
- Computer, Systems & Control,

…

Foundations

- Improved biological knowledge with the prospect of utilization in biotechnology and health care
- New experimental techniques in genomics and proteomics
- Classical mathematical modeling of biological processes
- Computer power for simulation of complex systems
- Storage and retrieval capability in large databases and data mining techniques
- Internet as the medium for the widespread availability from multiple resources of knowledge

Goals

- Models to unveil cellular mechanisms causing altered phenotypes
- Predictive tools to design cells with desired properties
- Individualized and predictive medicine

Models

In almost any case, models are only rough representations of their biological counterparts

Nevertheless, models enable to:

- Elucidate basic properties of modeled systems
- Check the reliability of basic assumptions
- Uncover lack of knowledge and requirements for clarification
- Create large repository of current knowledge

Definition of a Model

It depends on whom you ask…

- Geneticist: the mouse family Ts65DN serves as a model for human trisomy 21
- Chemist: a reaction network, described by dots (for metabolites) and arrows (for reactions)
- Mathematician/Engineer: the same reaction network can be modeled by a system of nonlinear ODEs

Abstractive representation of objects or processes that explains features of these objects or processes

Mathematical Models

Biological processes can be described in mathematical terms, however

- A biological object can be investigated by means of different experimental methods
- Each biological process can be described through different (mathematical) models
- The choice of a mathematical model or an algorithm depends on the problem, the purpose, and the intention of the investigator
- Modeling has to reflect essential properties of the system: Different models may highlight different aspects of the same instance

Model Development

Formulation of the problem:

- Identify the specific questions that shall be answered, along with background, problem and hypotheses

Available Knowledge: Check and collect quantitative and structural knowledge

- Components of the system
- Interaction map and kind of interactions
- Experimental results with respect to phenotypic responses against different stimuli (gene knockout, RNAi, environmental conditions)

Model Development

Selection of model structure:

- Level of description (atomistic, molecular, cellular, physiological)
- Deterministic or stochastic model
- Discrete or continuous variables
- Static, dynamical, spatio-temporal dynamical

Robustness/Sensitivity Analysis:

Test the dependence of the system behavior on changes of the parameters

Model Development

Predictive results from models

Experimental Tests

Assessment of the agreement and divergences between experimental results and model behavior

Iterative refinement of the hypotheses (and of the model)

Data Integration

Observation of biological phenomena is restricted to the granularity and precision of the available experimental techniques

A strong impulse to the development of a systematic approach in the last years has been given by the new high-throughput biotechnologies

- Sequencing of human and other genomes (genomics)
- Monitoring genome expression (transcriptomics)
- Discovering protein-protein and -DNA interactions (proteomics)

Different types of information need to be integrated

Issues in Data Integration

Data representation and storage:

- (too) Many databases (GO, KEGG, PDB, Reactome…)
- XML-like annotation languages (SBML, CellML)

Information retrieval

- Tools for retrieving information from multiple remote DBs

Data correlation

- Find the correlation between phenotypes and genomic/proteomic profiles
- Statistics, data mining, pattern analysis, clustering, PCA, …

Information Exchange

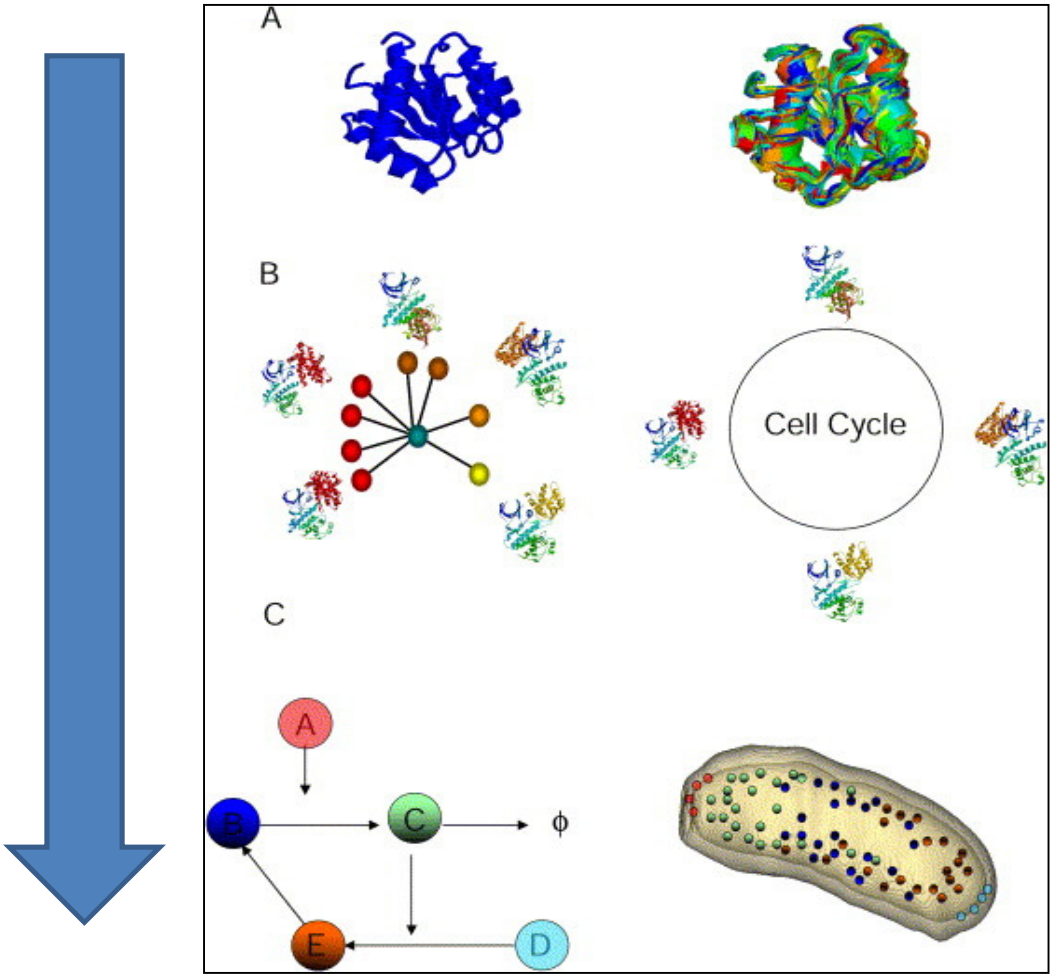
- The interdisciplinary nature of systems biology requires the exchange of information among scientists from different fields
- Mathematical formulas have to be made understandable for biologists
- People acquainted with the rigid rules of mathematics and computers have to understand the diversity of biological objects and the uncertainty in the outcome of experiments

Outline

- Basic concepts
- Theoretical principles
- Experimental techniques

General Aims

- Quantitative analysis of components and dynamics of complex biological systems



Features of complex systems

- Nonlinearity

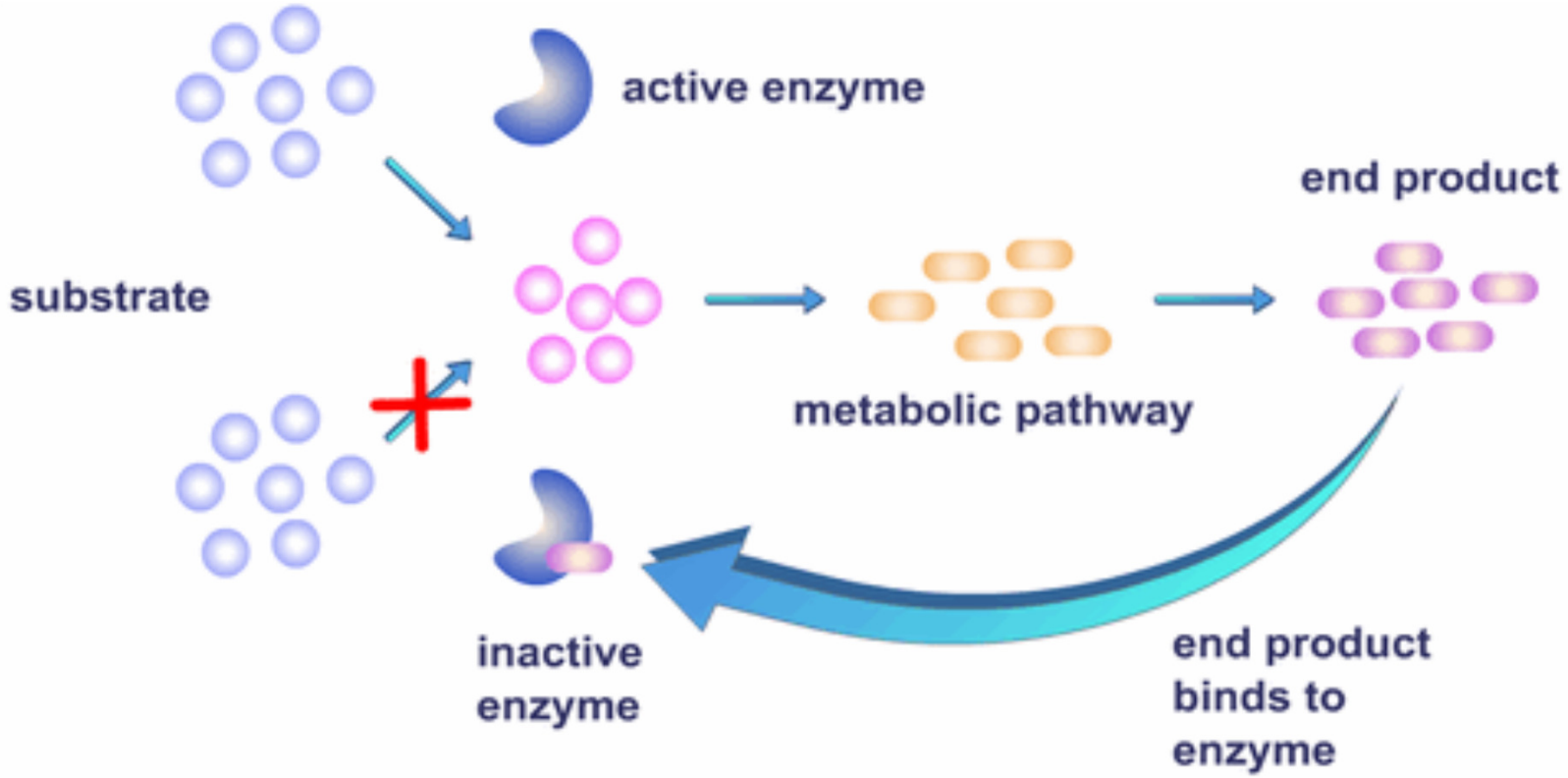
$$A + A \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} B$$

$$\frac{d[B]}{dt} = k_1[A]^2 - k_{-1}[B]$$

global properties not simple sum of parts

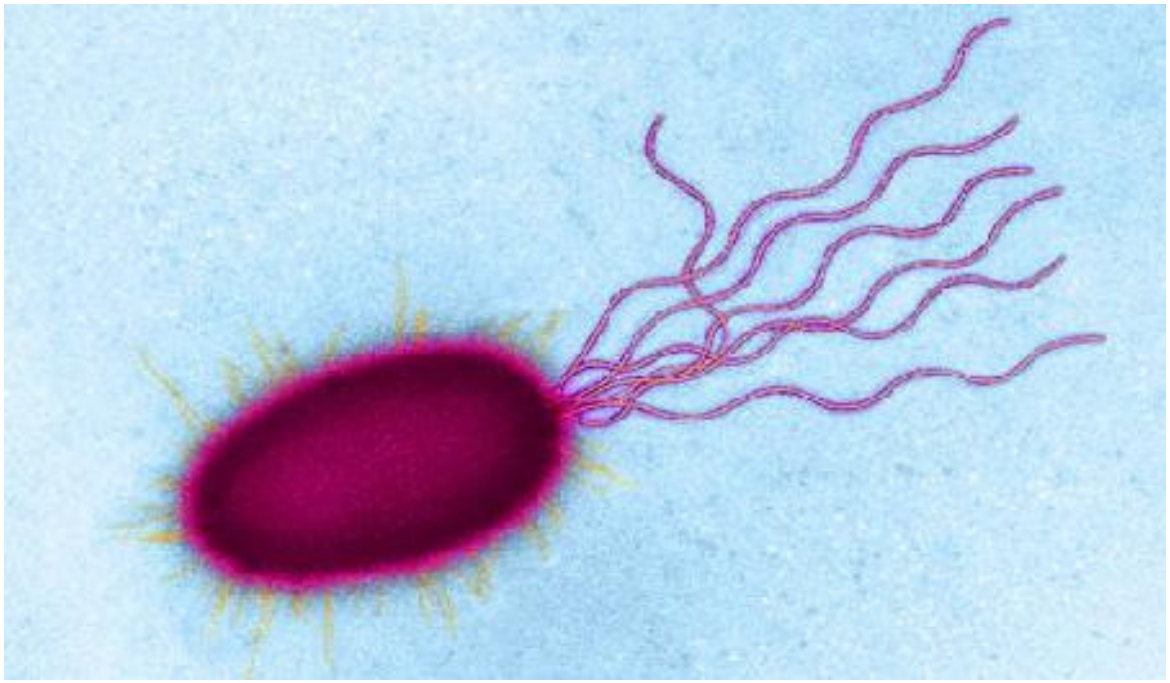
Features of complex systems

- Feedback loops



Features of complex systems

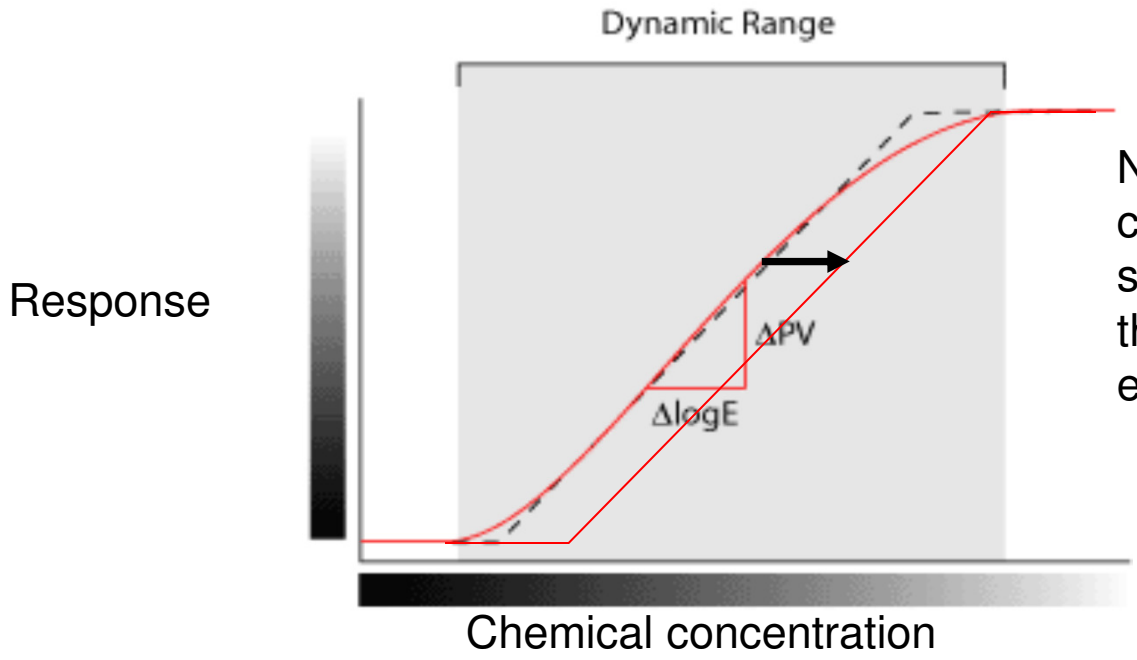
- Open systems (dissipation of energy)



Flagella uses energy: $\frac{dE}{dt} < 0$

Features of complex systems

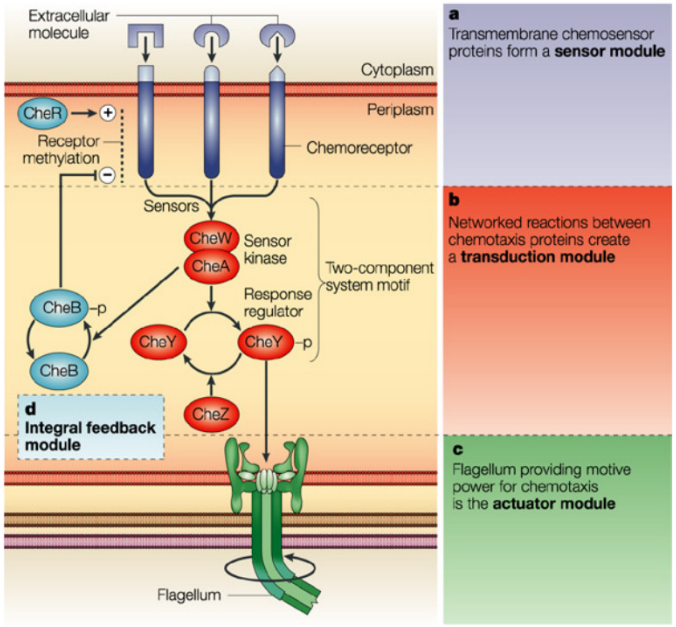
- Can have memory (response history dependent)



New protein may remain in cell after initial response, shifting the rate of reaction the next time the cell is exposed to a chemical

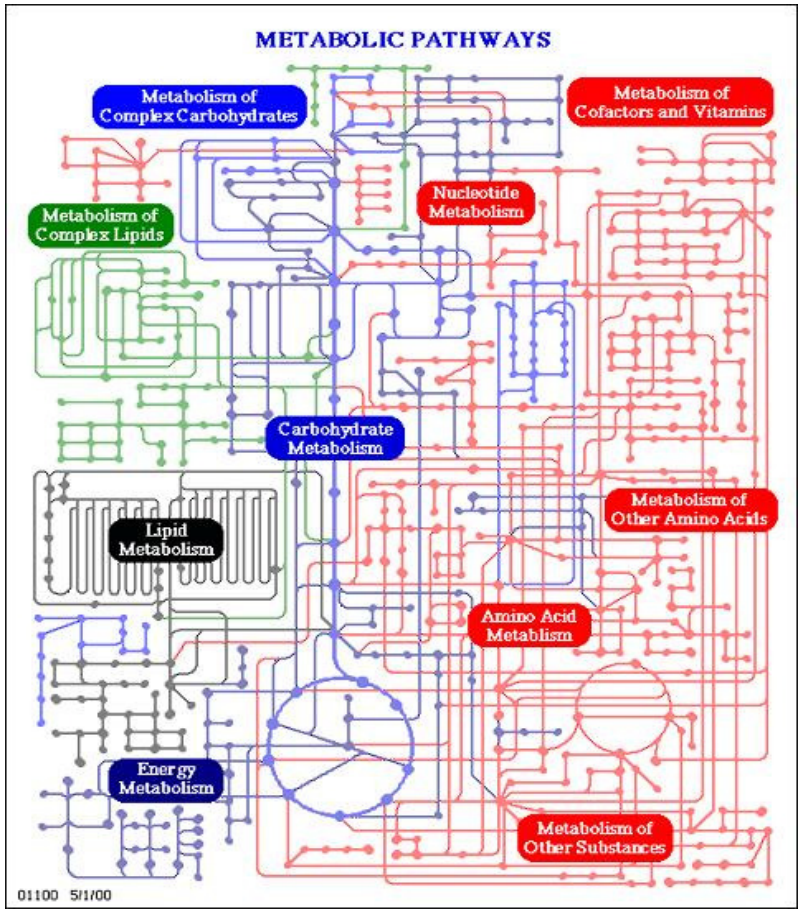
Features of complex systems

- Modularity
  - Interacting nodes w/ common function
  - Common modules frequently used in different networks



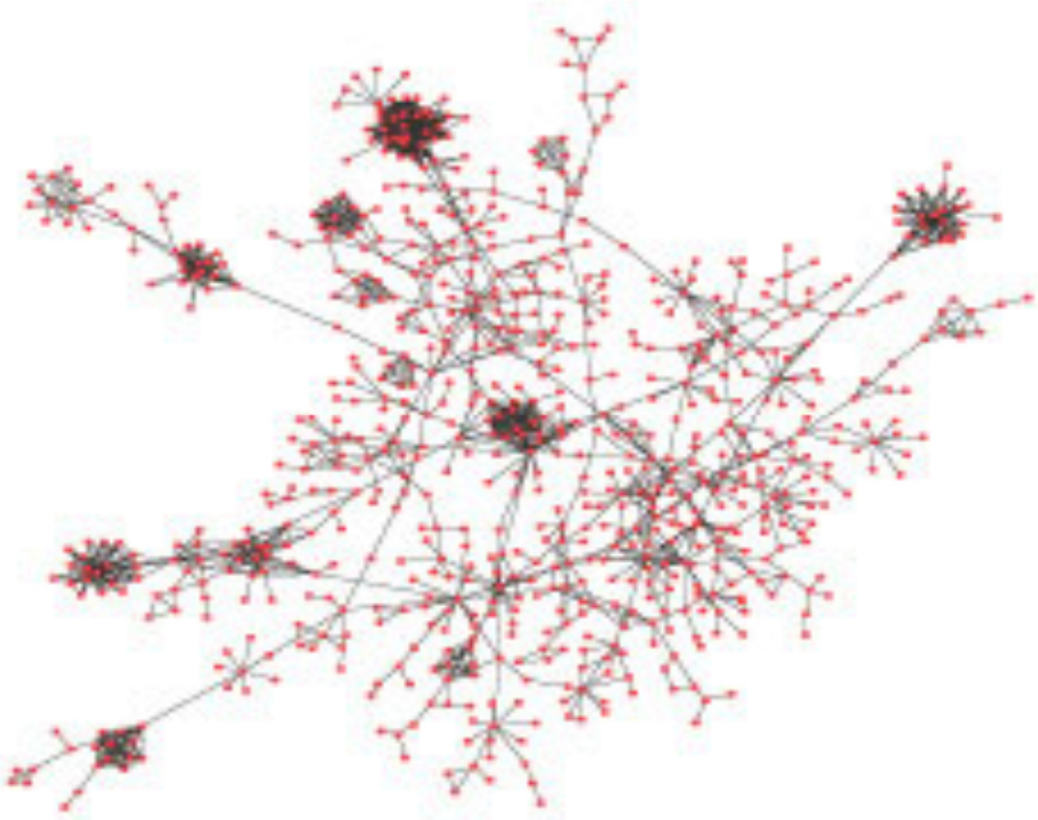
Features of complex systems

- Robustness
  - Insensitivity to parameter variation



Features of complex systems

- There are no *precise* boundaries
  - Crosstalks between different subnetworks



Different levels of theoretical models

- Quantitatively account for these properties
  - Different levels of modeling
- Three tiers
  - Static interactions
  - Deterministic
  - Stochastic
- Models which transcend tiers…

Different levels of theoretical models

- Tier 1: Interactome
  - Which molecules talk to each other in networks?
- Tier 2: Deterministic
  - What is the average case behavior?
- Tier 3: Stochastic
  - What is the variance of the system?

Different levels of theoretical models

- Tier 1
  - Get parts list

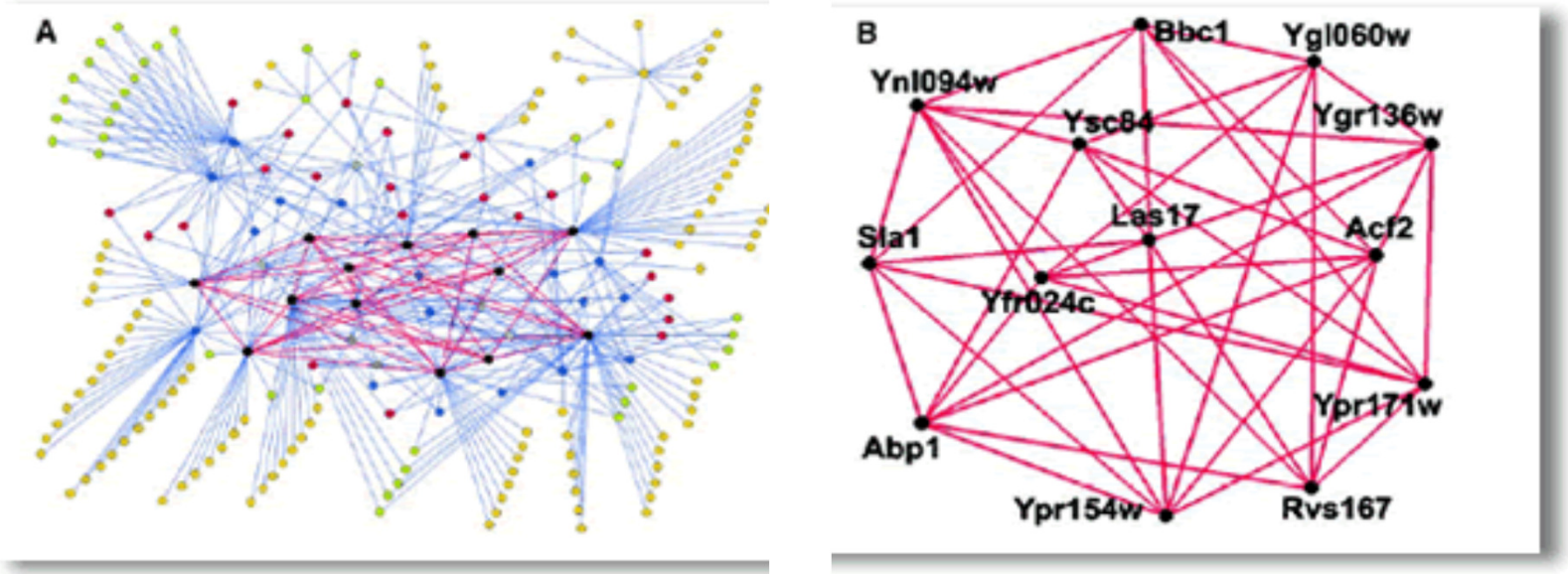


Table 1. Examples of interlinked positive feedback loops in biological regulation.

System	Positive feedback loops
Mitotic trigger	Cdc2 → Cdc25 → Cdc2 Cdc2 -| Wee1 -| Cdc2 Cdc2 -| Myt1 -| Cdc2
p53 regulation	p53 → PTEN -| Akt → Mdm-2 -| p53 p53 → p21 -| CDK2 -| Rb -|Mdm-2 -| p53
Xenopus oocyte maturation	Cdc2 → Mos → Cdc2 Cdc2 → Cdc25 → Cdc2 Cdc2 → Myt1 → Cdc2

Different levels of theoretical models

- Tier 2 & 3
  - Enumerate biochemistry
  - Define network/mathematical relationships
  - Compute numerical solutions

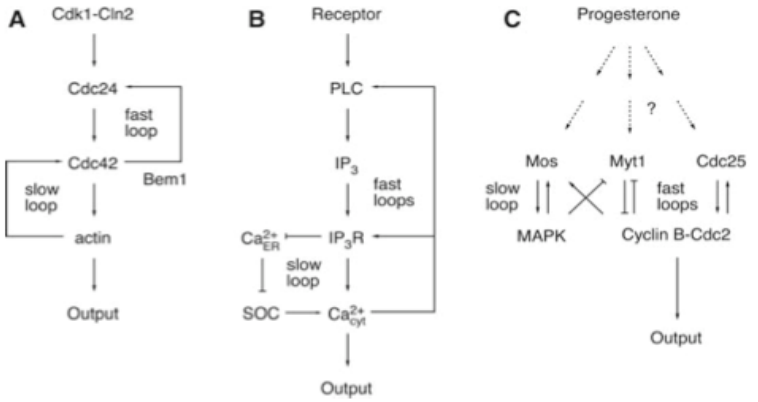


Table 1. Examples of interlinked positive feedback loops in biological regulation.

System	Positive feedback loops
Mitotic trigger	Cdc2 → Cdc25 → Cdc2 Cdc2 -| Wee1 -| Cdc2 Cdc2 -| Myt1 -| Cdc2
p53 regulation	p53 → PTEN -| Akt → Mdm-2 -| p53 p53 → p21 -| CDK2 -| Rb -|Mdm-2 -| p53
Xenopus oocyte maturation	Cdc2 → Mos → Cdc2 Cdc2 → Cdc25 → Cdc2 Cdc2 → Myt1 → Cdc2

1) One loop

$$\frac{dOUT}{dt} = k_{out\_on} * A * (1 - OUT) - k_{out\_off} * OUT + k_{out\_min}$$

$$\frac{dA}{dt} = [stimulus * \frac{OUT^n}{OUT^n + ec_{50}^n} * (1 - A) - A + k_{min}] * \tau_A$$

2) Two loops

$$\frac{dOUT}{dt} = k_{out\_on} * (A + B) * (1 - OUT) - k_{out\_off} * OUT + k_{out\_min}$$

$$\frac{dA}{dt} = [stimulus * \frac{OUT^n}{OUT^n + ec_{50}^n} * (1 - A) - A + k_{min}] * \tau_A$$

$$\frac{dB}{dt} = [stimulus * \frac{OUT^n}{OUT^n + ec_{50}^n} * (1 - B) - B + k_{min}] * \tau_B$$

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Different levels of theoretical models

- **Tier 2 & 3**
  - **Deterministic: Behavior of system with respect to time is predicted with certainty given initial conditions**
  - **Stochastic: Dynamics cannot be predicted with certainty given initial conditions**

Different levels of theoretical models

- Deterministic
  - Ordinary differential equations (ODE's)
    - Concentration as a function of time only
  - Partial differential equations (PDE's)
    - Concentration as a function of space and time
- Stochastic
  - Stochastic update equations
    - Molecule numbers as random variables
    - functions of time

$$\frac{d\vec{x}}{dt} = f(\vec{x})$$

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - \nu\frac{\partial C}{\partial x} + R$$

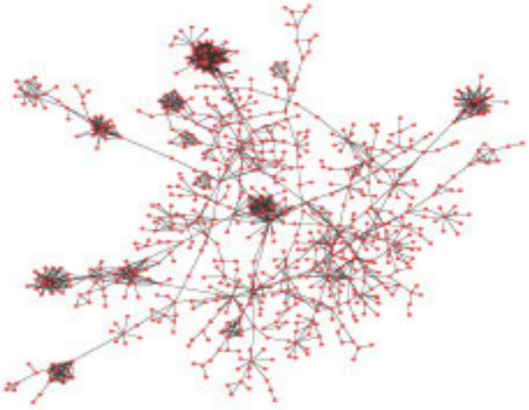
$$\frac{\partial}{\partial t}P(Y,t|Y_0,t_0) = \sum_{\mu}^{M}[c_\mu h_\mu(Y-\alpha_\mu)$$

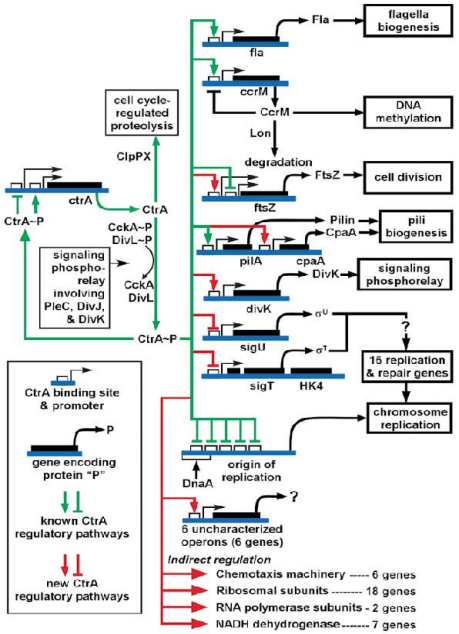
$$\times P(Y-\alpha_\mu,t|Y_0,t_0) - c_\mu h_\mu(Y)P(Y,t|Y_0,t_0)]$$

Y = # molecules at time t

Tier 1: Static interactome analysis

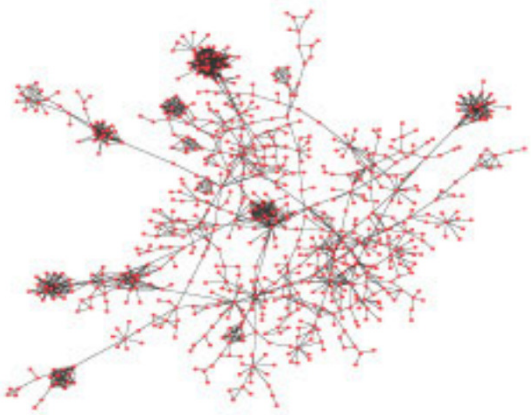
- **Protein-protein**
  - **Signal transduction**
  - **Cell cycle**
- **Protein-DNA**
  - **Gene regulation**
- **Metabolic pathways**
  - **Respiration**
  - **cAMP**

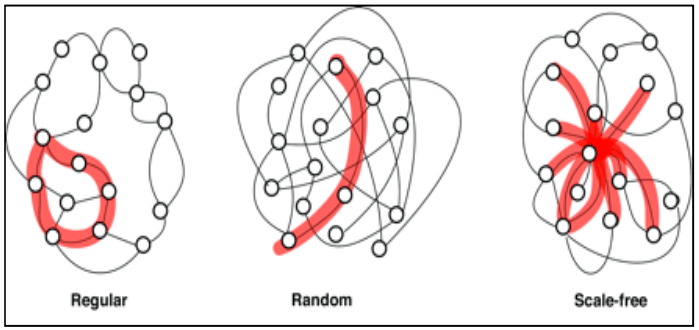


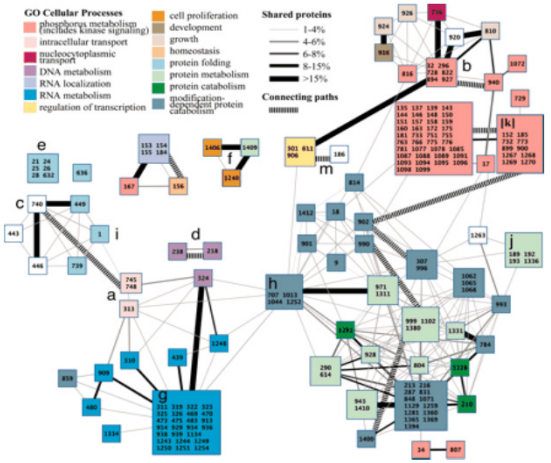


Tier 1: Static interactome analysis

- Goals
  - Determine network topology
  - Network statistics
  - Analyze modular structure

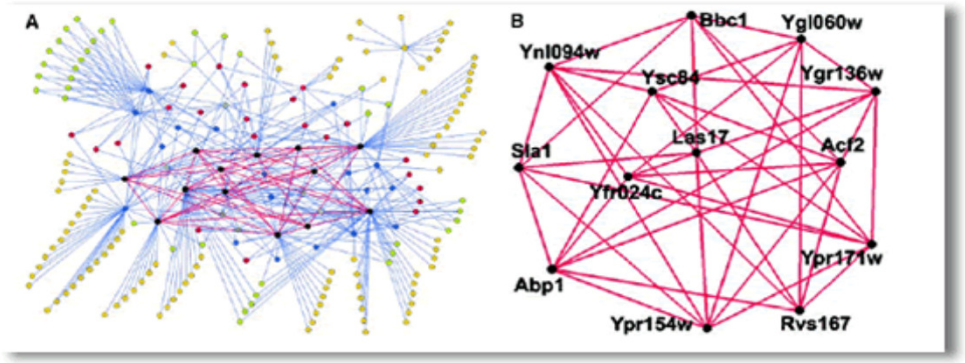




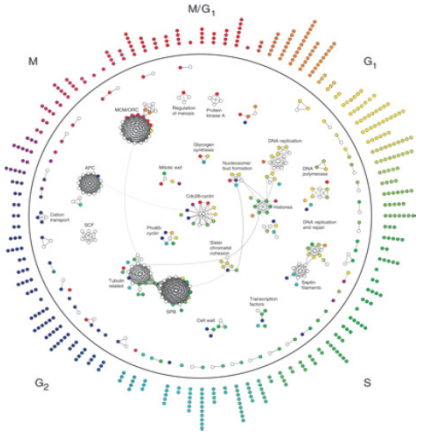


Tier 1: Static interactome analysis

- Limitations:
  - Time, space, population average
  - Crude interactions
    - strength
    - types
  - Global features
    - starting point for Tier 2 & 3



typical interactome



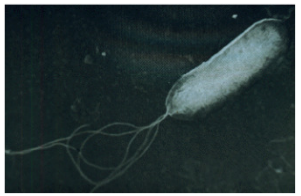
first time-varying yeast interactome (Bork 2005)

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Tier 2: Deterministic Models

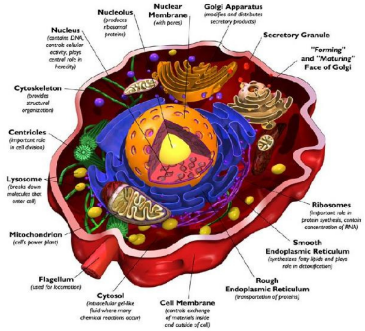
- Goal
  - model mesoscale system
  - average case behavior
- Three levels
  - ODE system
  - ODE compartment system
  - PDE

$$\frac{dC}{dt} = (\text{generation}) - (\text{consumption})$$



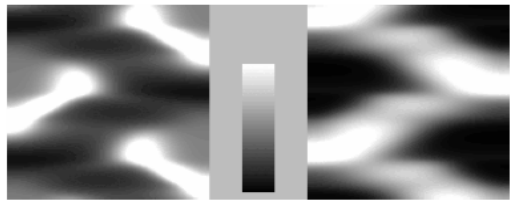
lumped cell

$$\frac{dC_i}{dt} = (\text{generation}) + (\text{flux in}) - (\text{consumption}) - (\text{flux out})$$



cell compartments

$$\begin{aligned}
\frac{\partial \rho_D}{\partial t} &= D_D \frac{\partial^2 \rho_D}{\partial x^2} - \frac{\sigma_1 \rho_D}{1+\sigma_1' \rho_e} + \sigma_2 \rho_e \rho_d && (1)\\
\frac{\partial \rho_d}{\partial t} &= \frac{\sigma_1 \rho_D}{1+\sigma_1' \rho_e} - \sigma_2 \rho_e \rho_d && (2)\\
\frac{\partial \rho_E}{\partial t} &= D_E \frac{\partial^2 \rho_E}{\partial x^2} - \sigma_3 \rho_D \rho_E + \frac{\sigma_4 \rho_e}{1+\sigma_4' \rho_D} && (3)\\
\frac{\partial \rho_e}{\partial t} &= \sigma_3 \rho_D \rho_E - \frac{\sigma_4 \rho_e}{1+\sigma_4' \rho_D}. && (4)
\end{aligned}$$



continuous time & space
(MinCDE oscillation)

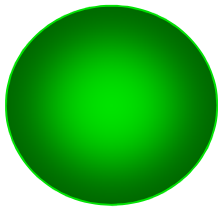
EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Tier 2: Deterministic Models

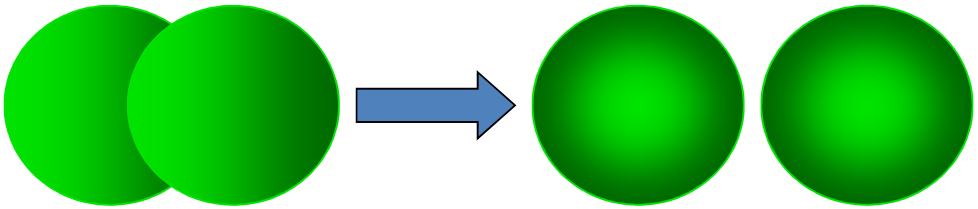
A simple example:

- One of simplest experiments in biology: Tracking cell divisions (eg, bacteria) over time.
- Analogous dynamics for tumor cell divisions (what they learn in med school):

A tumor starts as one cell

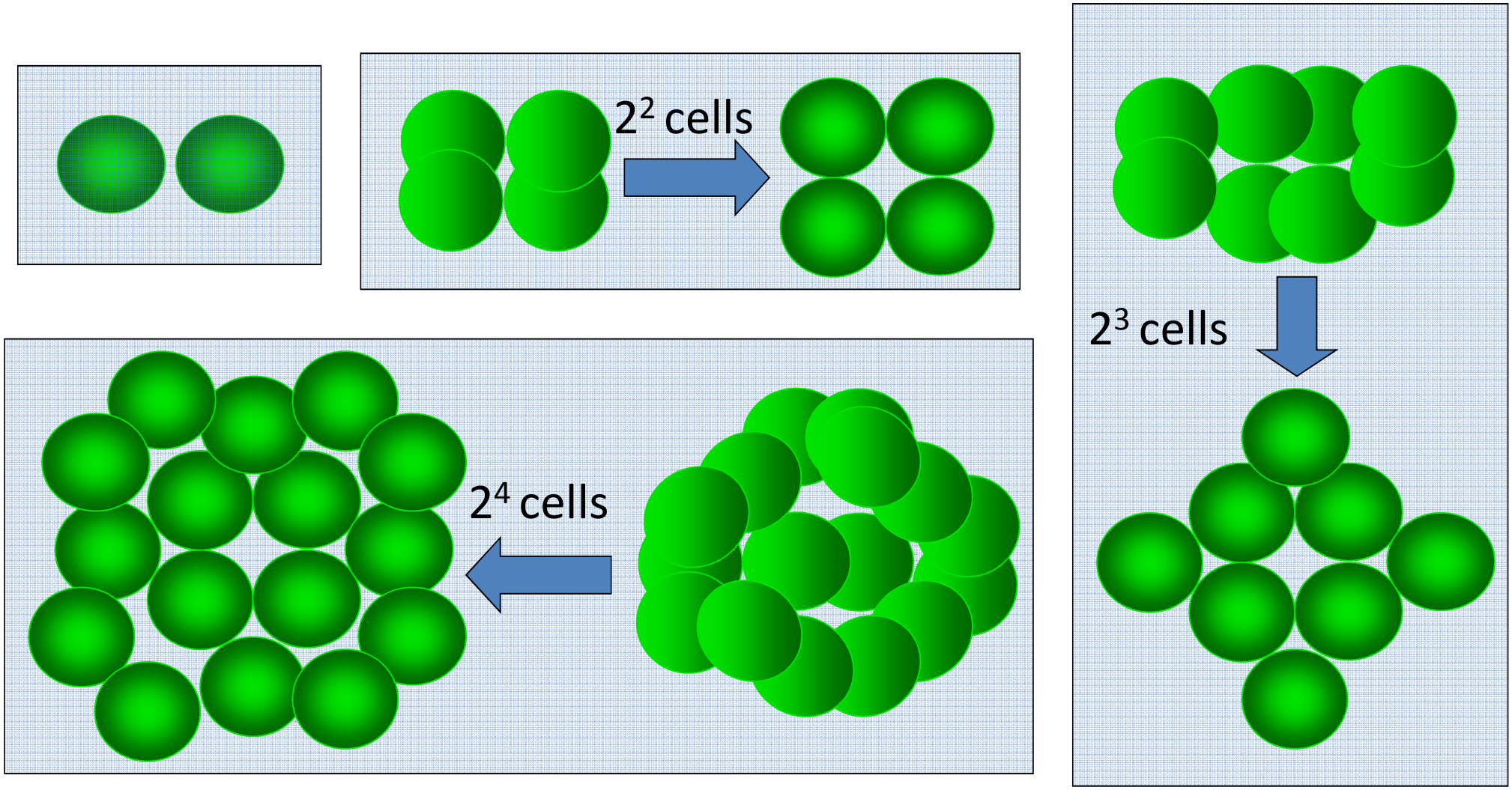


The cell divides and becomes two cells



Tier 2: Deterministic Models

Cell divisions continue…



Tier 2: Deterministic Models

- Ex: If *N* (representing, eg, bacterial density, or number of tumor cells) is a continuous function of *t* (time), then the derivative of *N* with respect to *t* is another function, called *dN/dt*, whose value is *defined* by the limit process

$$\frac{dN}{dt} = \lim_{t \to 0} \frac{N(t + \Delta t) - N(t)}{\Delta t}$$

- This represents the change is *N* with respect to time.

Tier 2: Deterministic Models

- Let $N(t)$ = bacterial density over time
- Let K = the reproduction rate of the bacteria per unit time ($K > 0$)
- Observe bacterial cell density at times t and $(t+\Delta t)$.

  Then

$$N(t+\Delta t) \approx N(t) + K\,N(t)\,\Delta t$$

Total density at time $t+\Delta t$ $\approx$ Total density at time t + increase in density due to reproduction during time interval Δt

- Rewrite: $(N(t+\Delta t) - N(t))/\Delta t \approx KN(t)$

Tier 2: Deterministic Models

- Take the limit as $\Delta t \rightarrow 0$

$$dN/dT = KN$$

"Exponential growth" (Malthus:1798)

- Analytic solution possible here.

$$N(t) = N_0 e^{Kt}$$

$$N_0 = N(0)$$

- Implication: Can calculate doubling time

Tier 2: Deterministic Models

- Find "population doubling time" t:

$$N(\tau)/N_0 = 2 \quad \text{and} \quad N(t) = N_0 e^{Kt}$$

$$\text{imply} \quad 2 = e^{K\tau}$$

Taking logs and solving for t gives

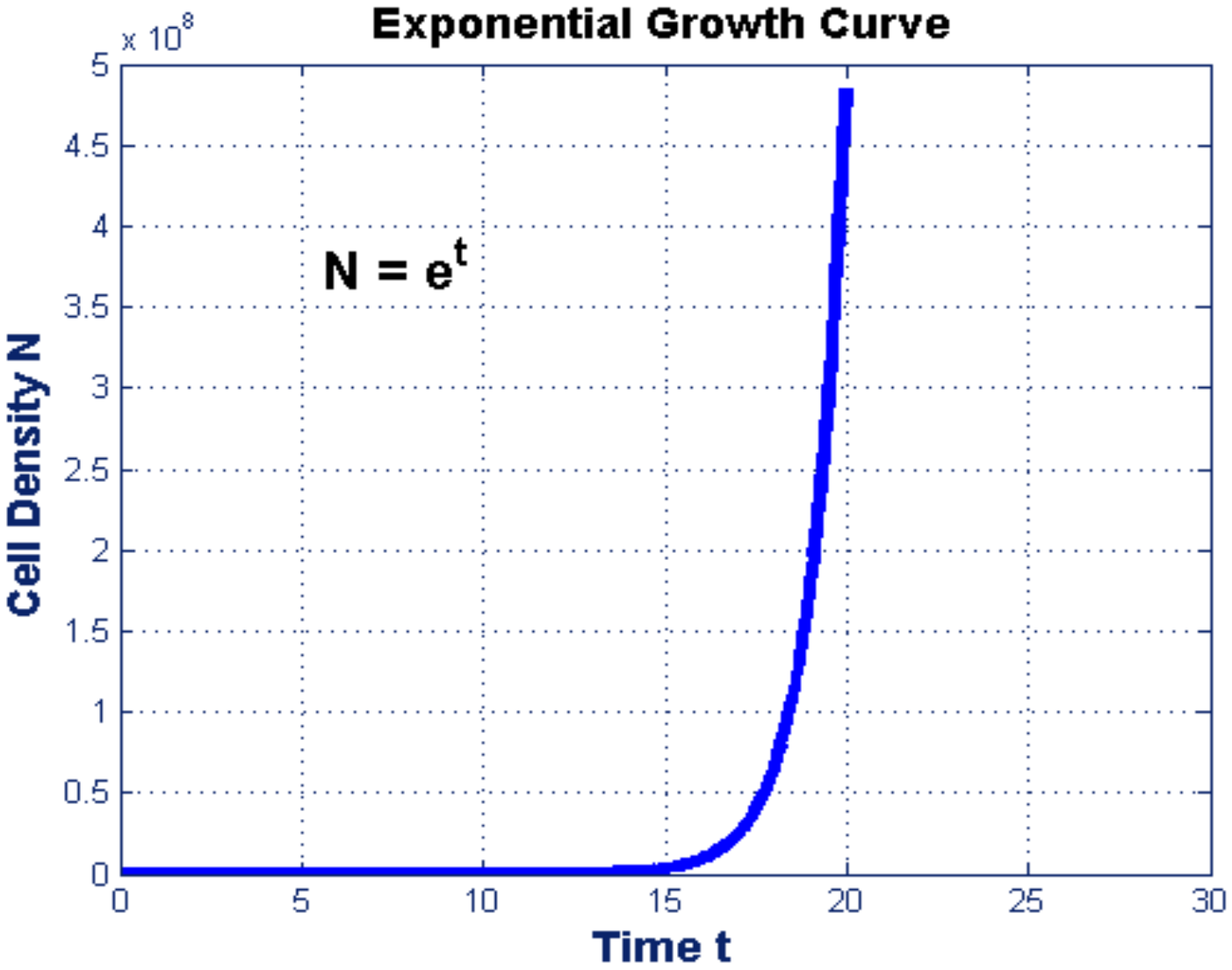
$$\ln(2) = K\tau \longrightarrow \tau = \ln(2)/K$$

- Point: doubling time inversely proportional to reproductive constant K

Tier 2: Deterministic Models

- Doubling time t=ln(2)/K
- Suppose K=ln(2), so t=1, ie, cell population doubles in 1 day.
- $2^{30} \approx 10^9$. In 30 days, 1 cell →→ detectable population
- Tumor will reach 100 grams between days 36 and 37.
- One week later, tumor weighs a kilo and is lethal.
- Every cancer cell must be killed to eliminate the tumor

Exponential Growth: Realistic?



Extending the Growth Model: Additional Assumptions + New System

- Reproductive rate K is proportional to the nutrient concentration, $C(t)$: so $K(C)=\mathrm{k}C$
- a units of nutrient are consumed in producing 1 unit of pop'n increment → system of equations:

$$dN/dt = \kappa CN$$

$$dC/dt = -\alpha\left(dN/dt\right) = -\alpha\kappa CN$$

- Further simplify the system of ODEs:

$$dN/dt = \kappa\left(C_0 - \alpha N\right)N$$

- Logistic Growth Law!

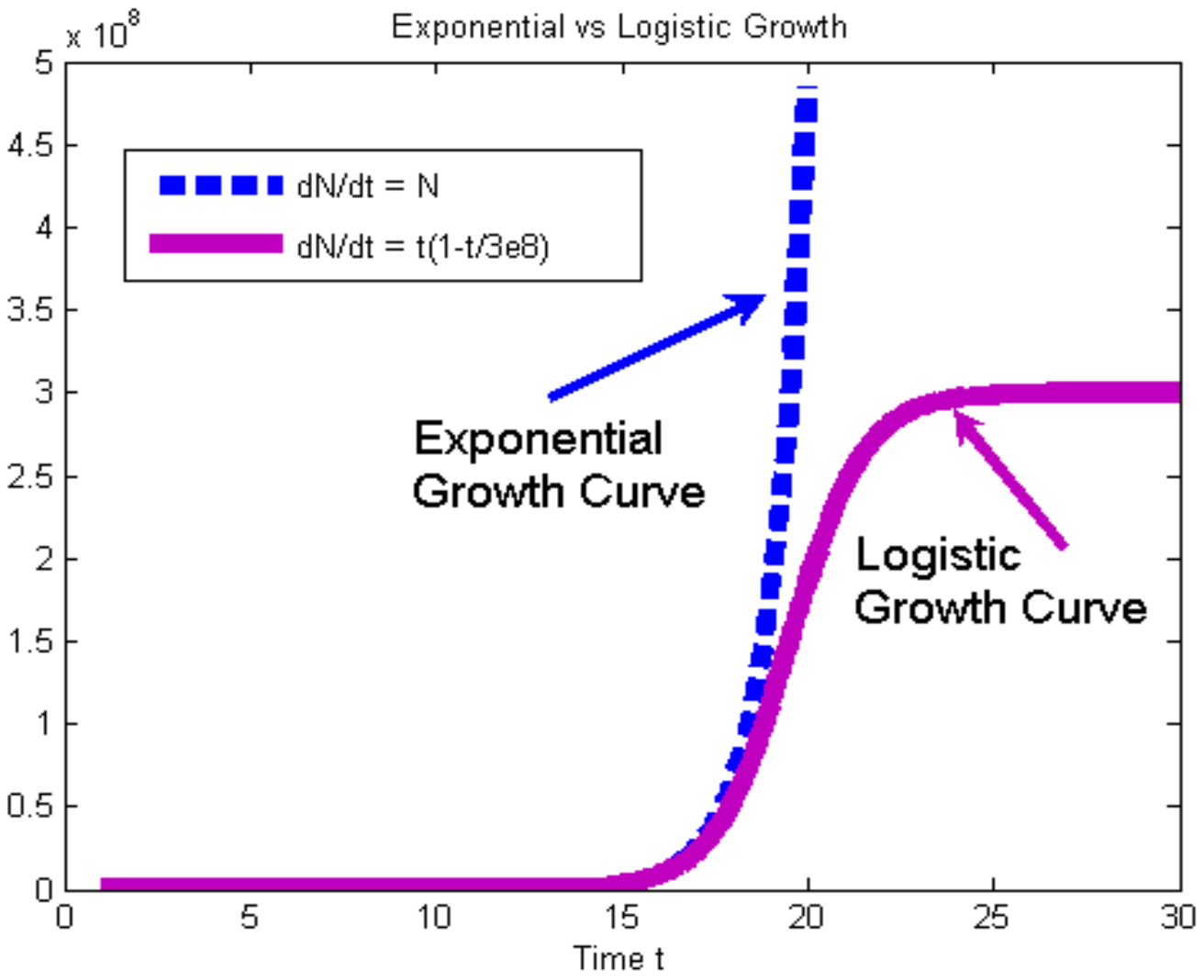
Analysis of Logistic Model for Cell Growth

- Solution:

$$N(t) = N_0\left(\frac{C_0}{\alpha}\right) \Big/ \left(N_0 + \left[\left(\frac{C_0}{\alpha}\right) - N_0\right] e^{-\kappa C_0 t}\right)$$

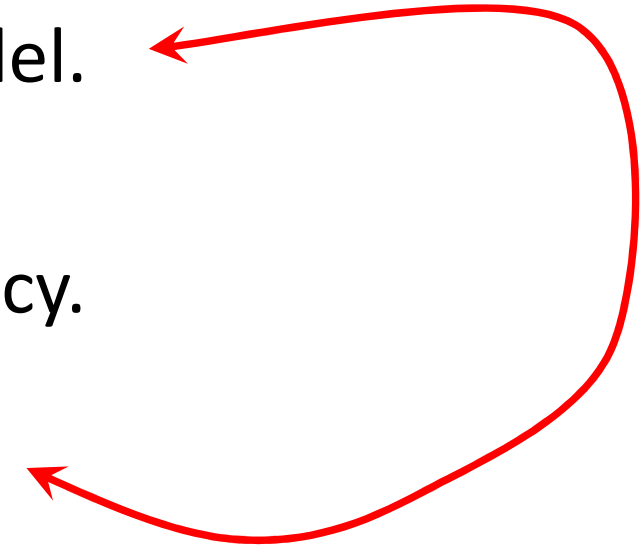
- N_0 = initial population
- ${}_kC_0$ = intrinsic growth rate
- $C_0/_a$ = carrying capacity
- For small popn levels N, N grows about "exponentially", with growth rate r ≈ ${}_kC_0$
- As time $t \to \infty$, $N \to N(\infty)=C_0/_a$
- This "self limiting" behavior may be more realistic for longer times

Exponential versus Logistic Growth



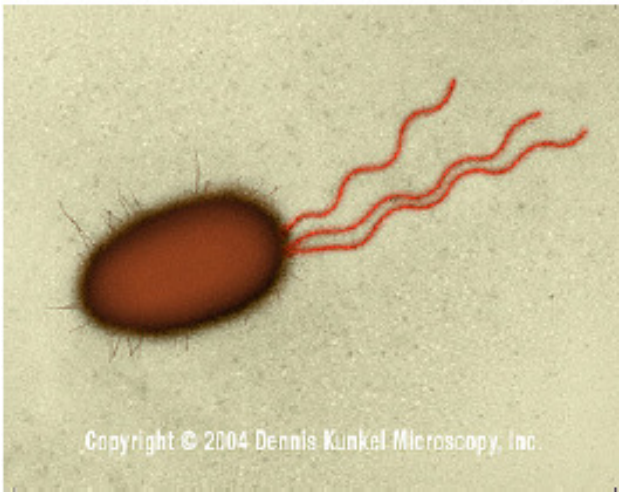
Tier 2: Deterministic Models

1. Ask the question.
2. Select the modeling approach.
3. Formulate the model.
4. Solve the model.
5. Validate the accuracy.
6. Revise the model.



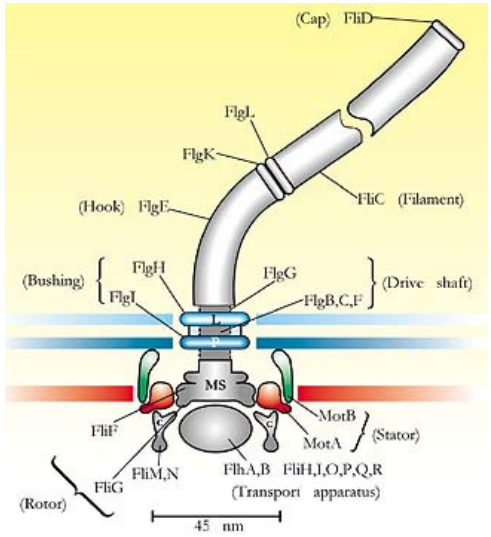
Tier 2: Deterministic Models

- More compliated example
  - Robustness in bacterial chemotaxis
- Chemotaxis: bacterial migration towards/away from chemicals

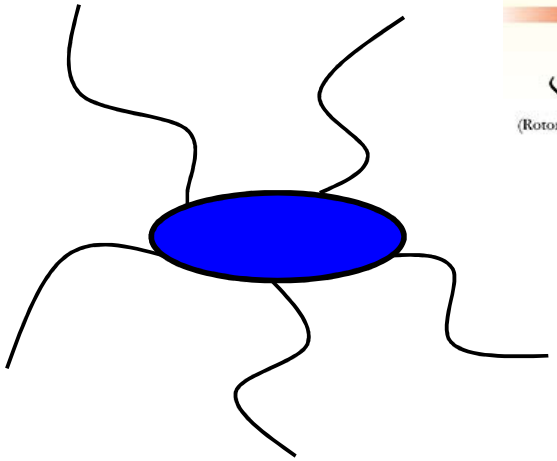


Tier 2: Deterministic Models

Clockwise



Counter-clockwise



Tumble

Run

Tier 2: Deterministic Models

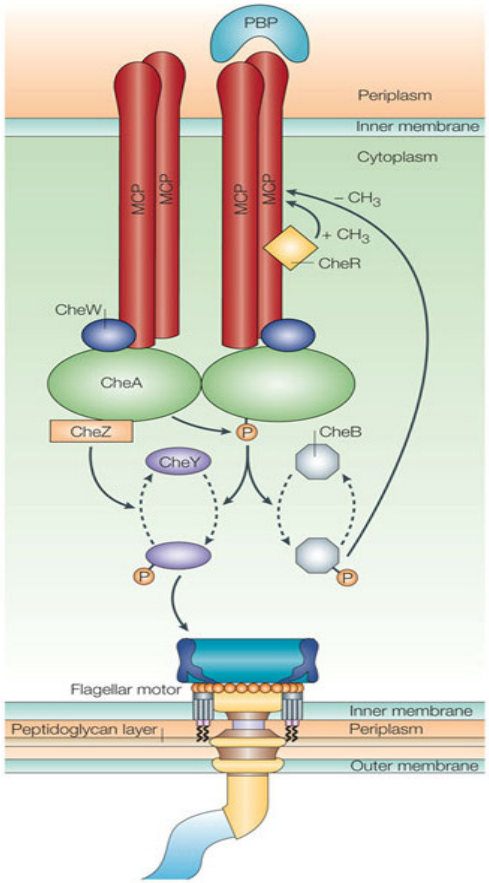
Attractant
(e.g. nutrient)

Repellent
(e.g. toxin)

Chemotaxis: reduction in tumbling frequency to drive swimming toward attractant

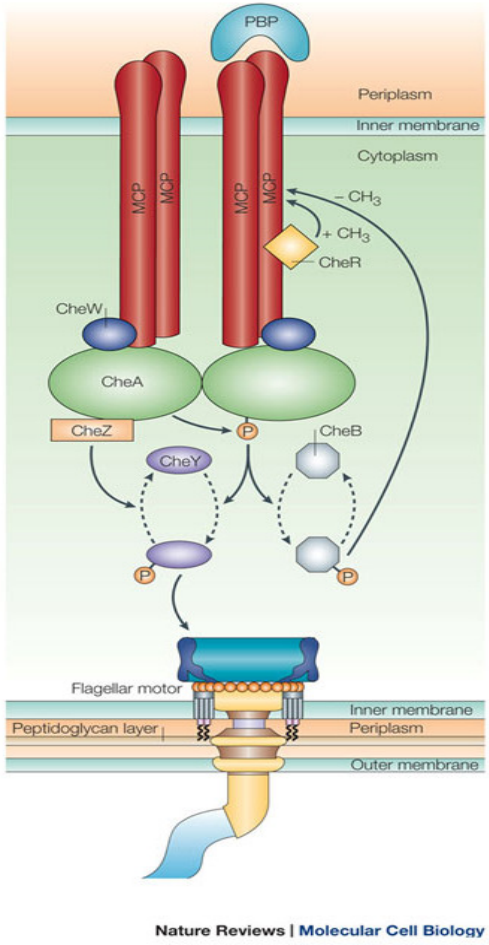
Tier 2: Deterministic Models

- Model: three modules
  - Parameters
    - concentrations
    - Binding affinities



Nature Reviews | Molecular Cell Biology

Tier 2: Deterministic Models



Nature Reviews | Molecular Cell Biology

- **Ligand binding**

$$L + T_{n(p)}(CheR) \underset{km5\sim km7}{\overset{k5\sim k7}{\rightleftarrows}} LT_{n(p)}(CheR)$$

- **Methylation**

$$(L)T_{n(p)} + CheR \xrightarrow{k1c\sim k4c} (L)T_{n+1(p)} + CheR$$

$$(L)T_{n(p)} + CheB_p \xrightarrow{k1m\sim k4m} (L)T_{n-1(p)} + CheB_p$$

- **Phosphorylation**

$$(L)T_n(CheR) + ATP \xrightarrow{k7\sim k9} (L)T_{np}(CheR) + ADP$$

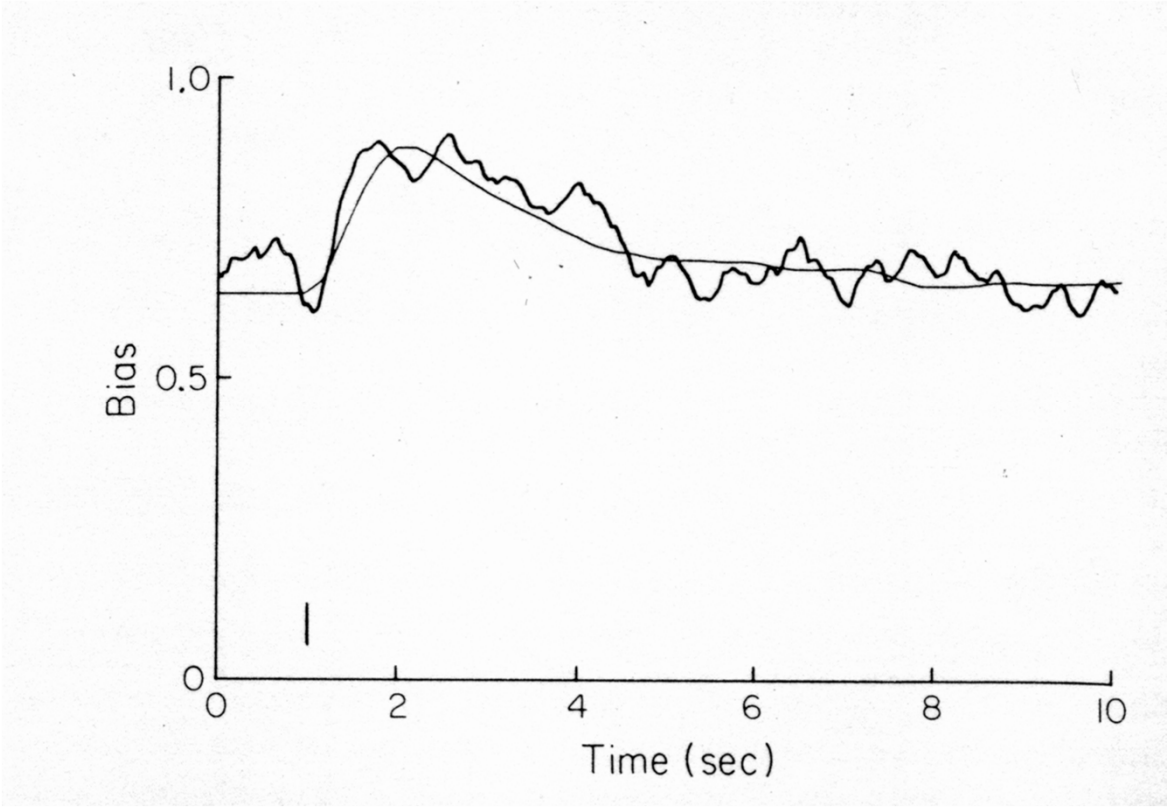
$$(L)T_{np}(CheR) + CheY \xrightarrow{ky} T_n(CheR) + CheY$$

$$(L)T_{np}(CheR) + CheB \xrightarrow{kb} T_n(CheR) + CheB_p$$

$$CheY_p + CheZ \xrightarrow{kmy} CheZ + CheY + P$$

$$CheB_p \xrightarrow{kmb} CheB + P$$

Tier 2: Deterministic Models



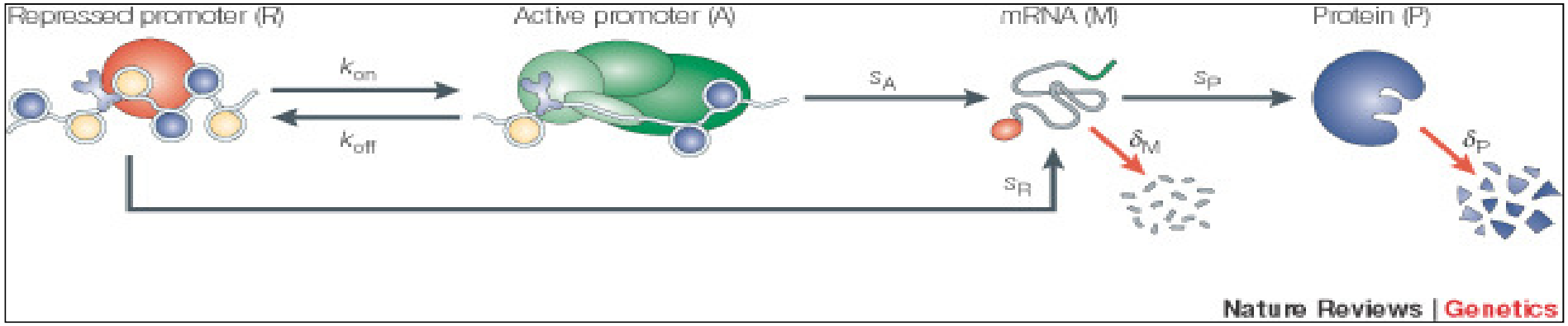
Perfect Adaptation in Bacterial Chemotaxis Signaling

Tier 2: Deterministic Models

- **The adaptation precision of the E. coli chemotaxis network is highly robust to perturbations**
- **but adaptation time and steady-state behavior are fine-tuned**

Tier 3: Stochastic analysis

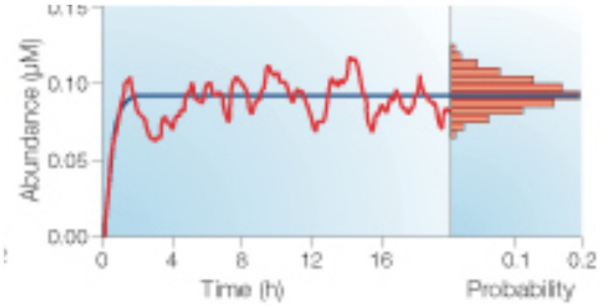
- Fluctuations in abundance of expressed molecules at the single-cell level
  - Leads to non-genetic individuality of isogenic population



Tier 3: Stochastic Analysis

- When stochasticity is negligible, use deterministic modeling…
- Molecular “noise” is low:
  - System is large
    - molar quantities
  - Fast kinetics
    - reaction time negligible
  - Large cell volume
    - infinite boundary conditions

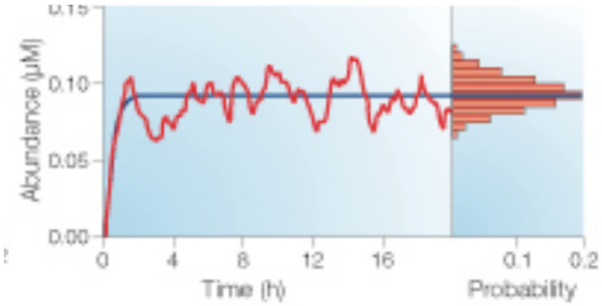
$$\text{Variance: } \frac{1}{\sqrt{N}}$$



Tier 3: Stochastic Analysis

- Molecular “noise” is high:
  - System is small
    - finite molecule count matters
  - Slow kinetics
    - relative to movement time
  - Large cell volume
    - relative to molecule size
- Need explicit stochastic modeling!

$$\text{Variance: } \frac{1}{\sqrt{N}}$$



Tier 3: Ensemble Noise

- Transcriptional bursting
  - Leaky transcription
  - Slow transitions between chromatin states
- Translational bursting
  - Low mRNA copy number

Tier 3: Temporal Noise

Chemical reactions are described by the *law of mass action*

- the speed of a reaction is proportional to the concentrations of the individual reactants involved
- **However, a specific reaction between two molecules depends on their random collisions.**

Gillespie's simulation algorithm

- stochastic method based on the theory of collisions
- each reaction takes a (continuous) random time which is exponentially distributed

Canonical way of modeling molecular stochasticity

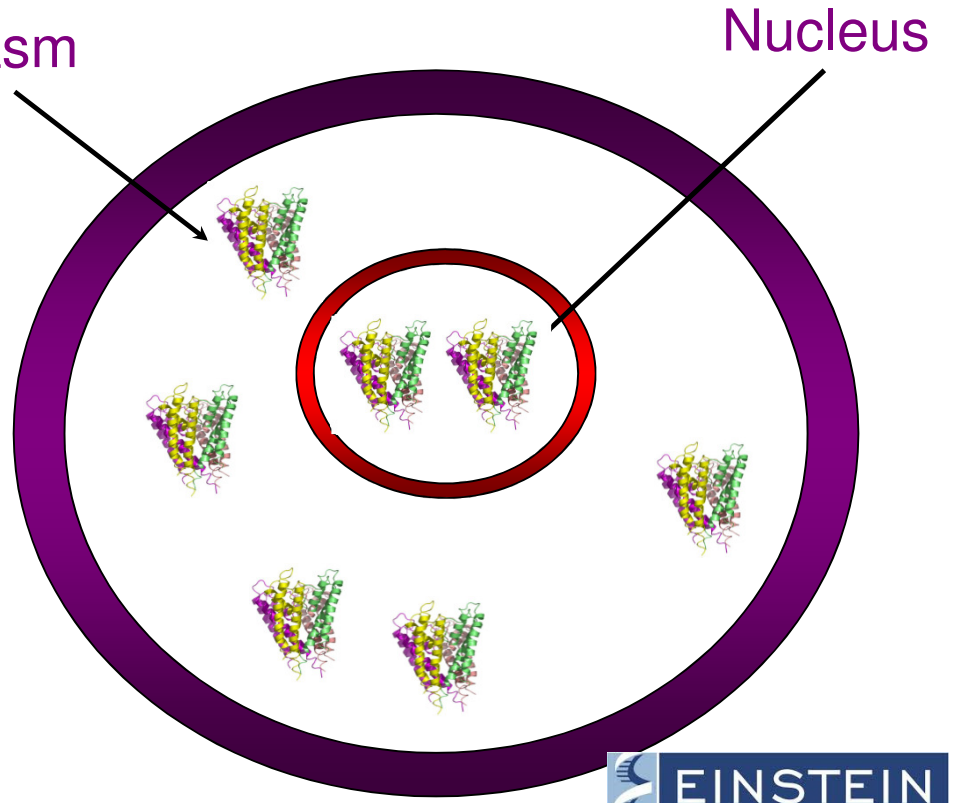
Tier 3: Spatial Noise

Finite number effect: translocation of molecules from the nucleus to the cytoplasm have a large effect on nuclear concentration

N = average molecular abundance

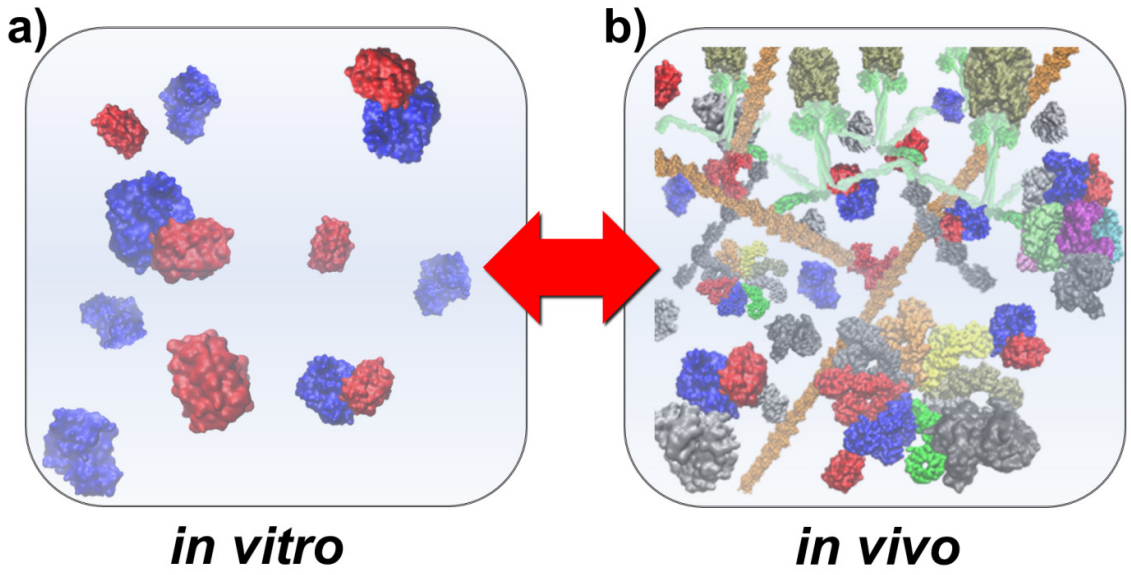
η (coefficient of variation) = σ/N

- Decrease in abundance results in a $1/\sqrt{N}$ scaling of the noise ($\eta=1/\sqrt{N}$)



Tier 3: Stochastic Analysis

- Measurement of chemical kinetics parameters and molecular concentrations *in vivo*
  - Differences between *in vitro* and *in vivo* data



Summary

- Each biological process can be described through different (mathematical) models
- The choice of a mathematical model or an algorithm depends on the problem, the purpose, and the intention of the investigator
- Modeling has to reflect essential properties of the system: Different models may highlight different aspects of the same instance

Outline

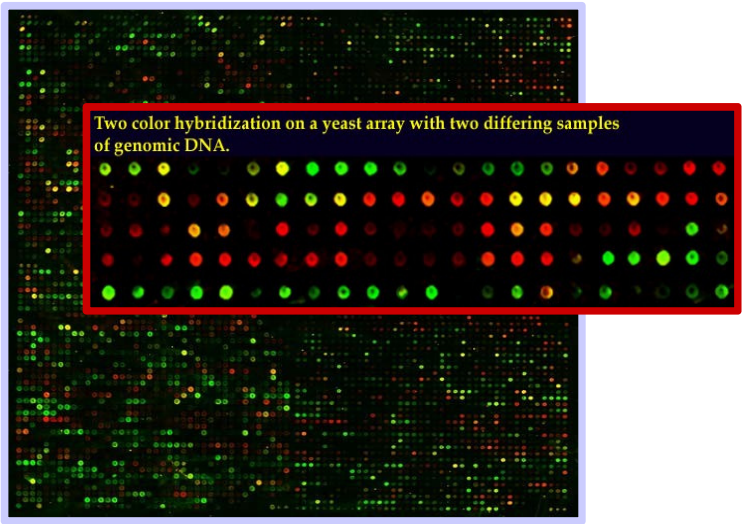
- Basic concepts
- Theoretical principles
- Experimental techniques

Getting Experimental Data

- Systems scientists are typically not concerned about the difficult and often ignore processes by which biological relationships and interactions are identified
- However, it is worth having a glimpse at the basic techniques used in experimental biology
- Two main issues arise when using experimental techniques:
  - How to get quantitative data out of experiments designed to give qualitative answers
  - How to get time-sampled (synchronized) measurements of many biological objects (possibly cheaply and in a suitable time)

DNA Microarrays

- DNA chips, also called DNA microarrays, are method for the high-throughput analysis of gene expression
- Instead of looking at the expression of a single gene, microarrays allow one to monitor the expression of several thousand genes in a single experiment, resulting in a global picture of cellular activity
- Hence they represent a key tool for implementing a systems biology approach



Single Feature

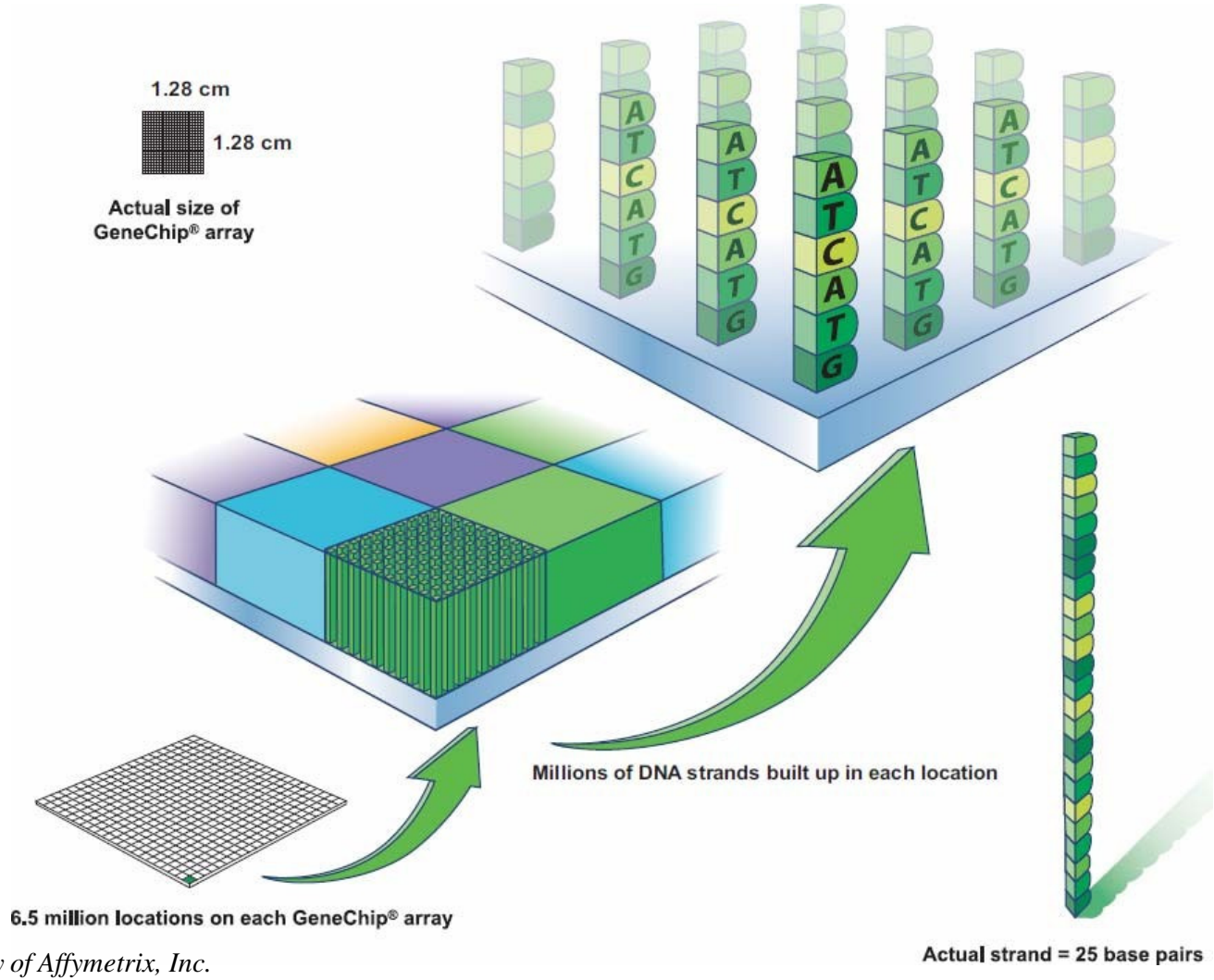


Image courtesy of Affymetrix, Inc.

Probes Hybridization

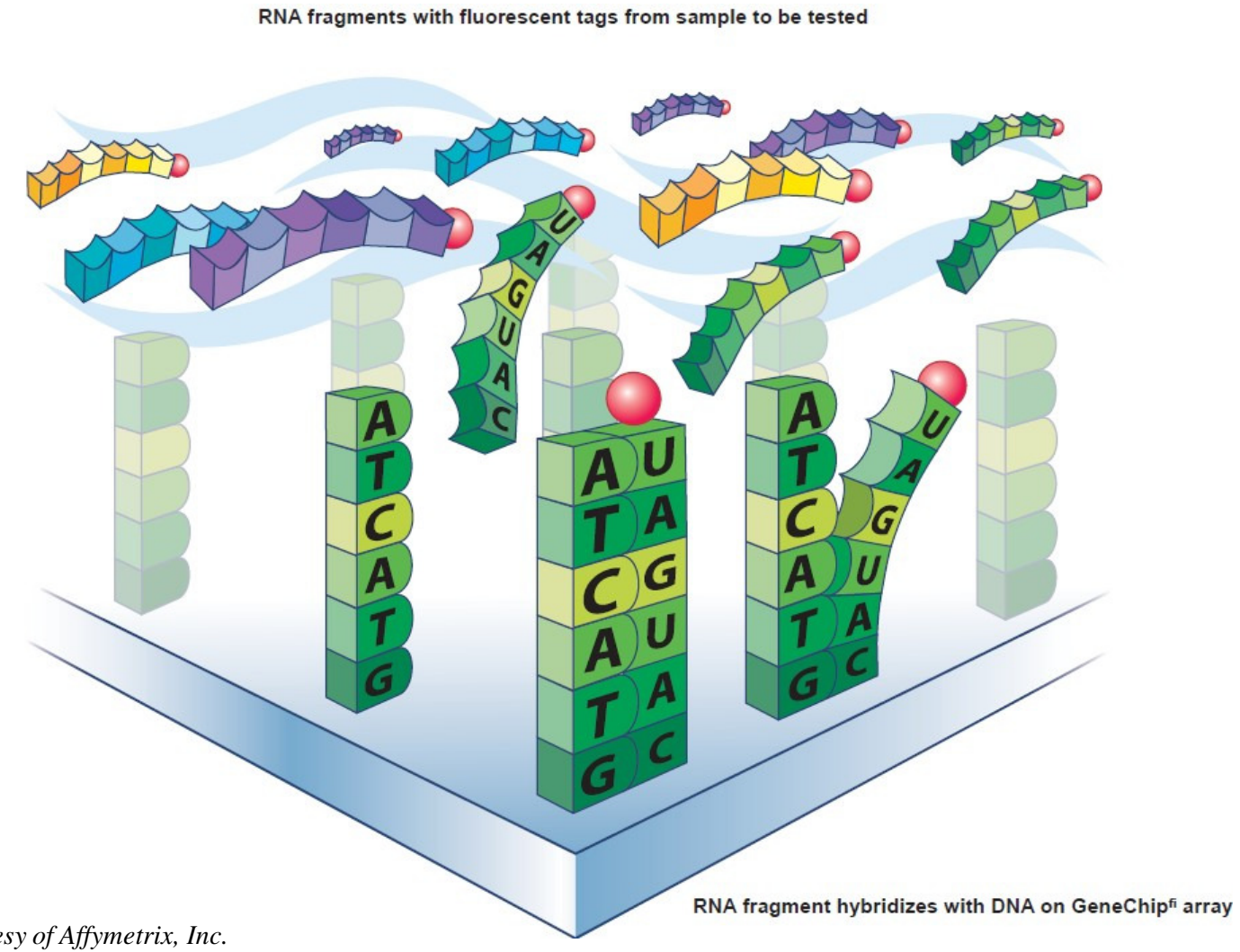


Image courtesy of Affymetrix, Inc.

Hybridized DNA Visualization

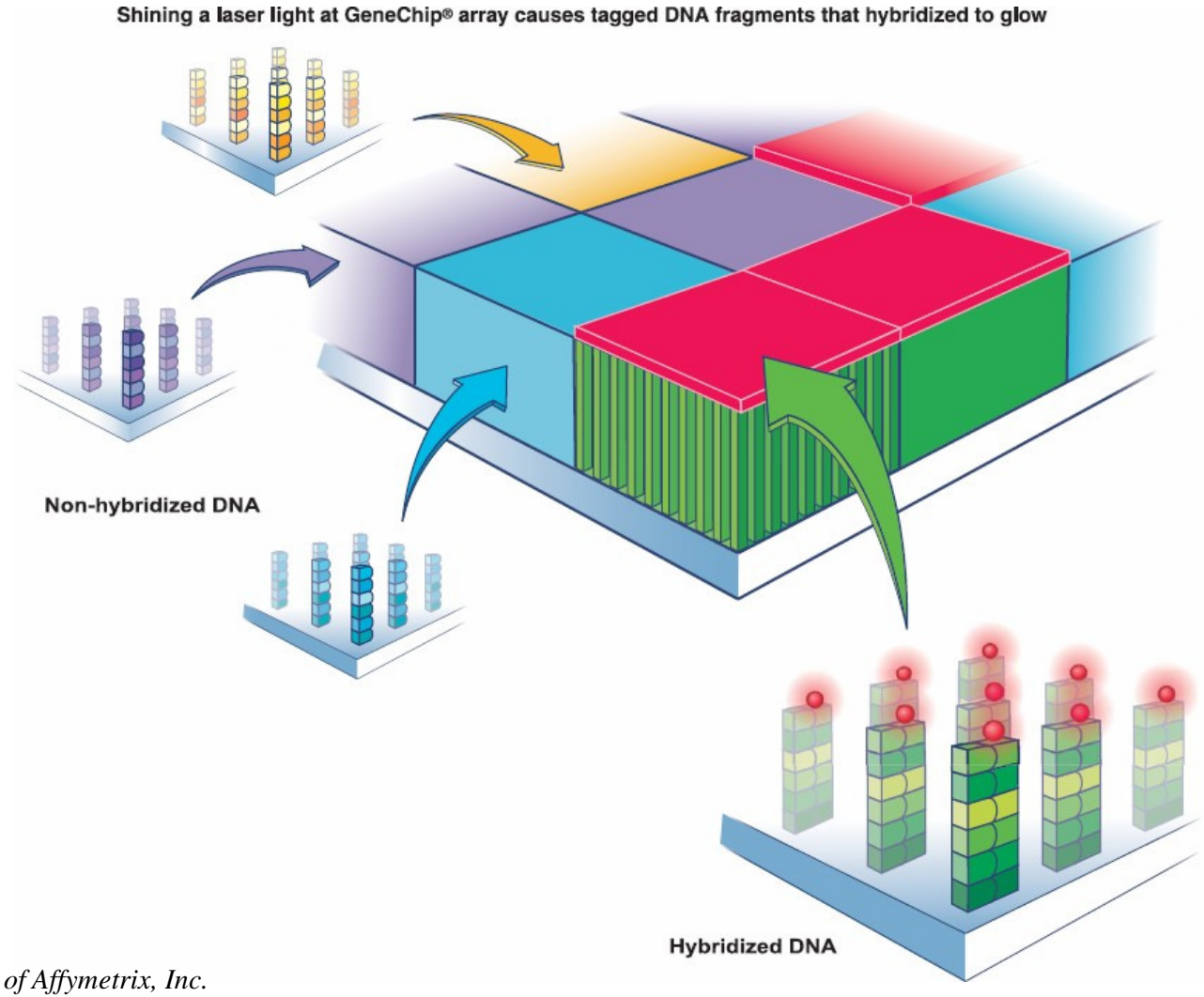
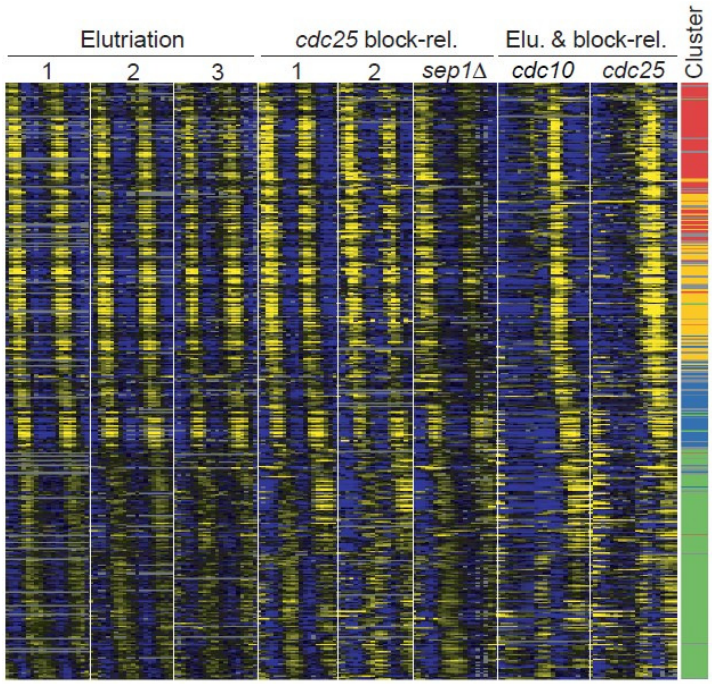
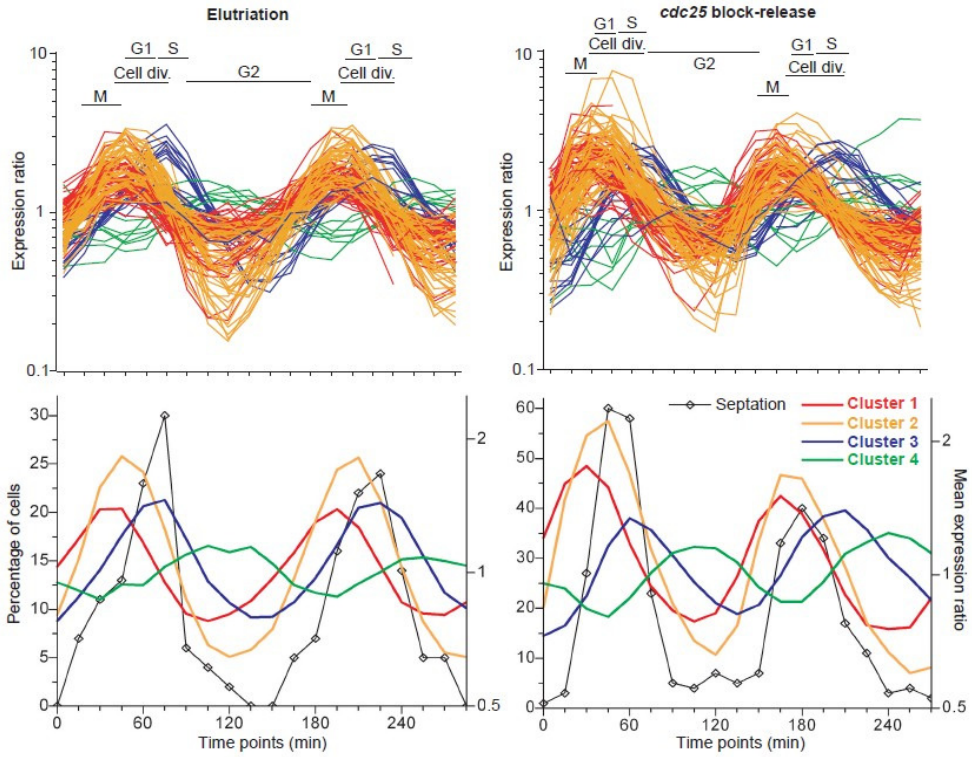


Image courtesy of Affymetrix, Inc.

Time-Course Experimental Data

Microarrays enable to derive time-course experimental data, with a desired sampling time





Rustici et al, Nature Genetics 36(8), 2004

Protein Microarrays

The function of a gene is realized by the coded protein, not by its mRNA

This fact addresses a main flaw in the DNA microarray approach

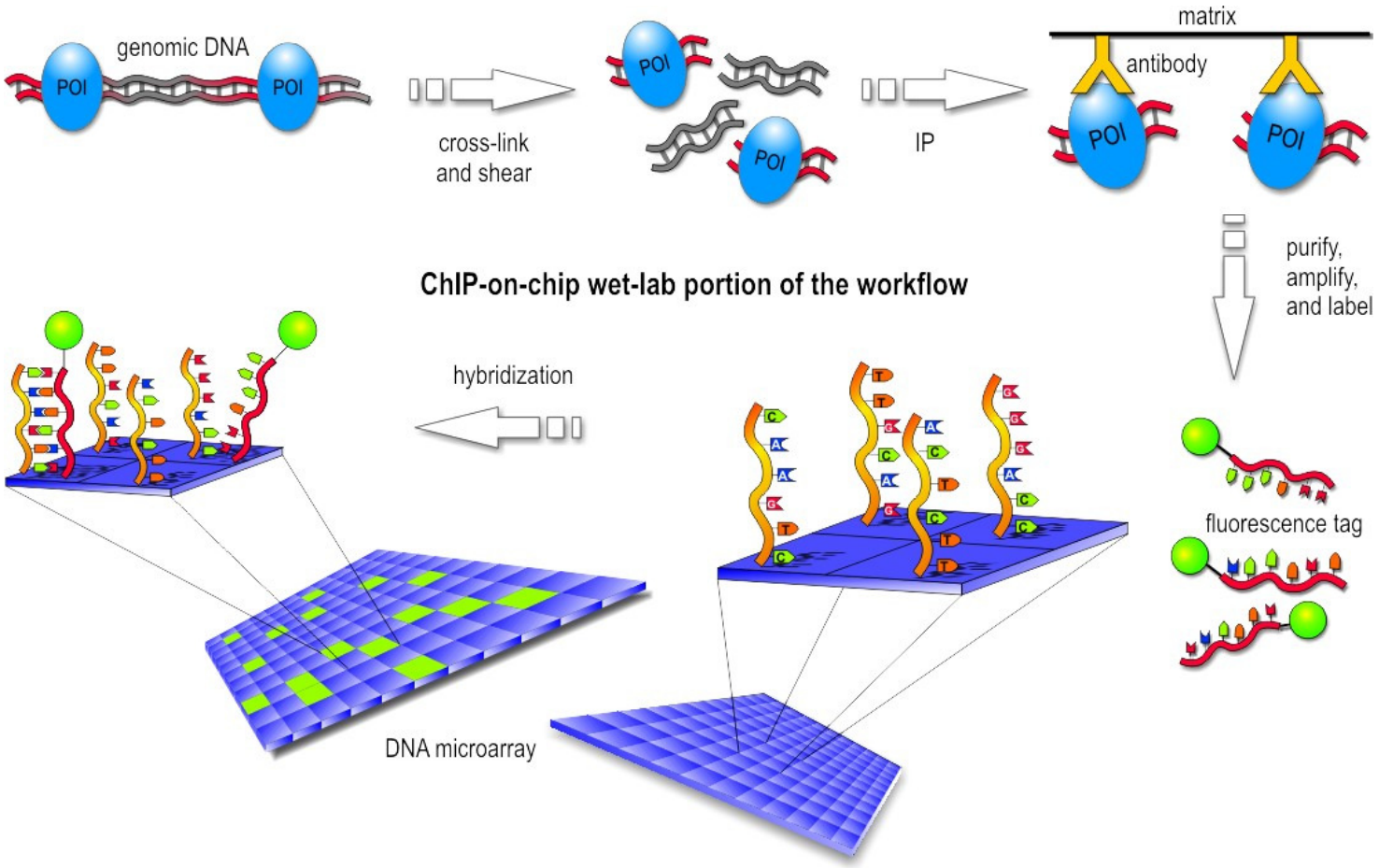
Solution: protein microarrays! Unfortunately…

- Proteins are not as uniform as DNA
- It is not (yet) possible to generate the amount of recombinant protein needed for high-throughput experiments
- Optimal interaction conditions (in terms of temperature, ionic strenght, pH) are largely varying among proteins

Nonetheless, research on protein chips is rapidly progressing

ChIP on Chip

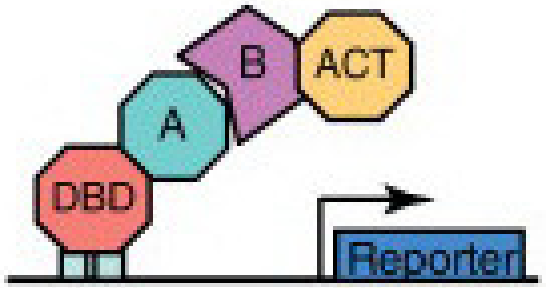
Chromatin Immuno–Precipitation is used to discover transcription factors (protein-DNA interactions)



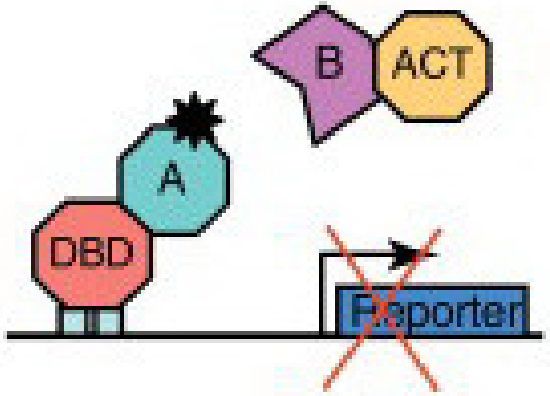
Yeast Two-Hybrid (Y2H)

- The two-hybrid system is a molecular genetic tool which facilitates the study of protein-protein interactions.
- If two proteins interact, then a reporter gene is transcriptionally activated.
- A color reaction can be seen on specific media.
- You can use this to
  - Study the interaction between two proteins which you expect to interact
  - Find proteins (prey) which interact with a protein you have already (bait).

Yeast Two-Hybrid (Y2H)



Growth
blue
luminescence



No growth
white
no light

Key:

DBD — DNA binding domain

ACT REP — Transcription activation or repressor domain

A DBD — Bait protein

B ACT — Prey protein

Yeast two hybrid advantages and disadvantages

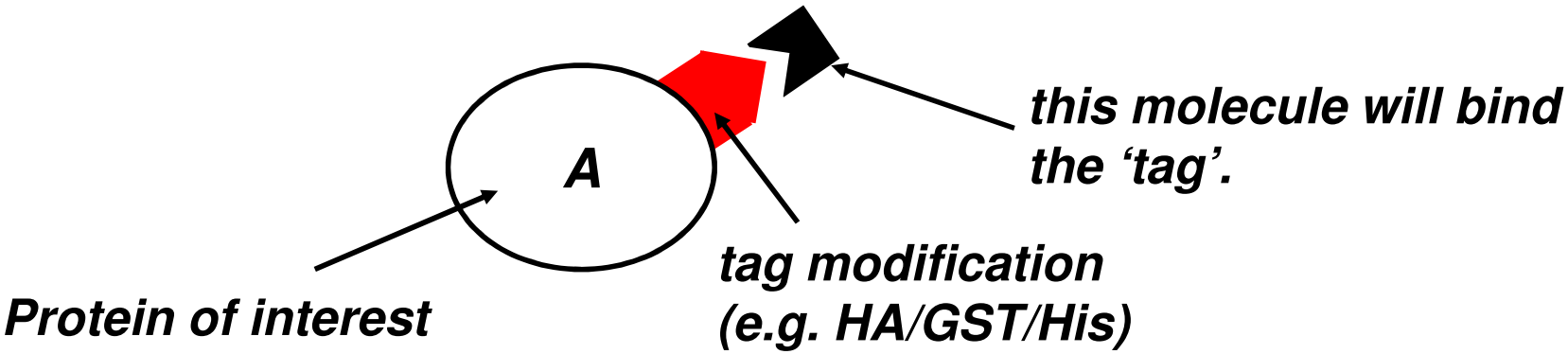
Advantages

- Y2H is relatively sensitive – can detect even transient and unstable interactions (e.g. signal transduction interactions)
- can be used to map binary interactions (e.g. detail interactions within a complex of proteins)
- can be used to assess protein interactions for non-native yeast proteins

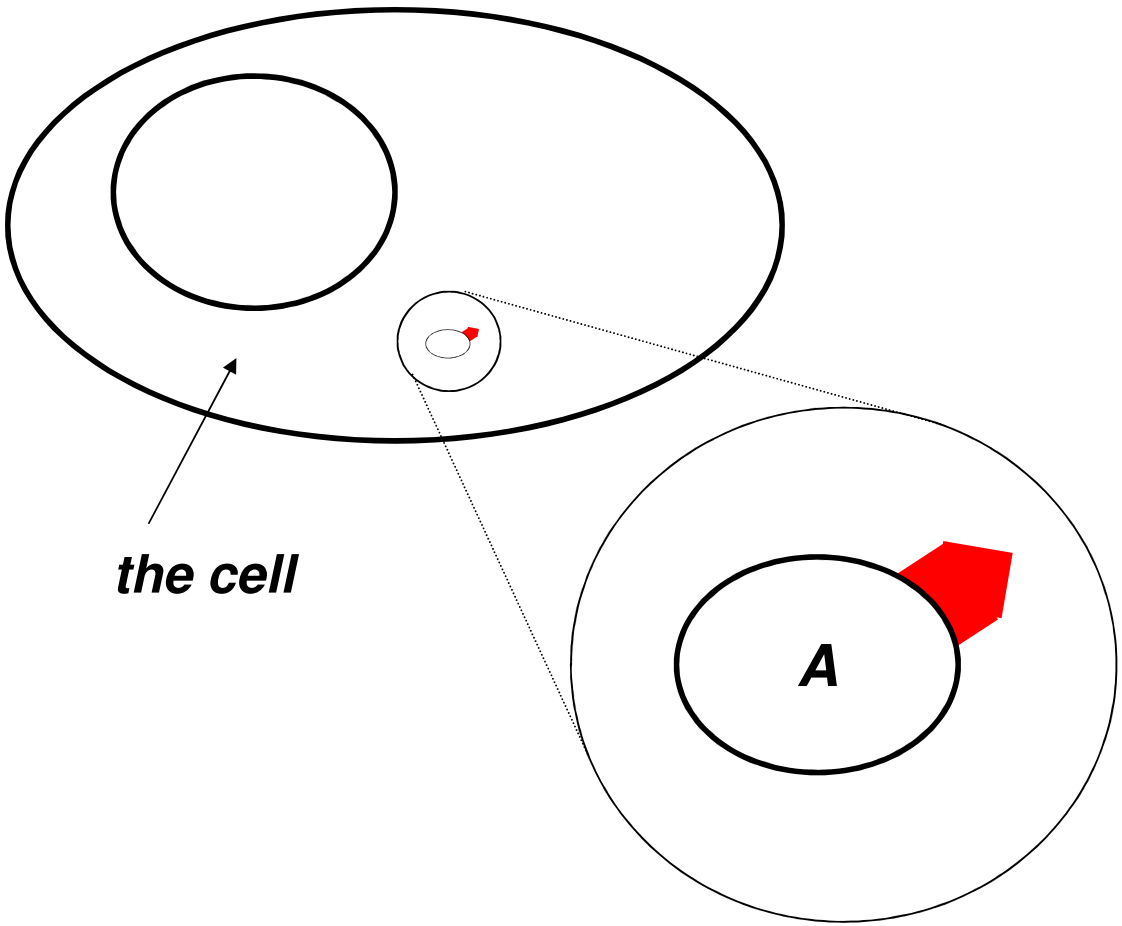
Disadvantages

- Interactions must occur within the nucleus to be detected (e.g. will miss membrane-bound proteins, etc.)
- Interactions may not be biologically relevant (e.g. proteins are not normally co-localized, or need additional co-factors for proper folding)
- Condition-specifity?

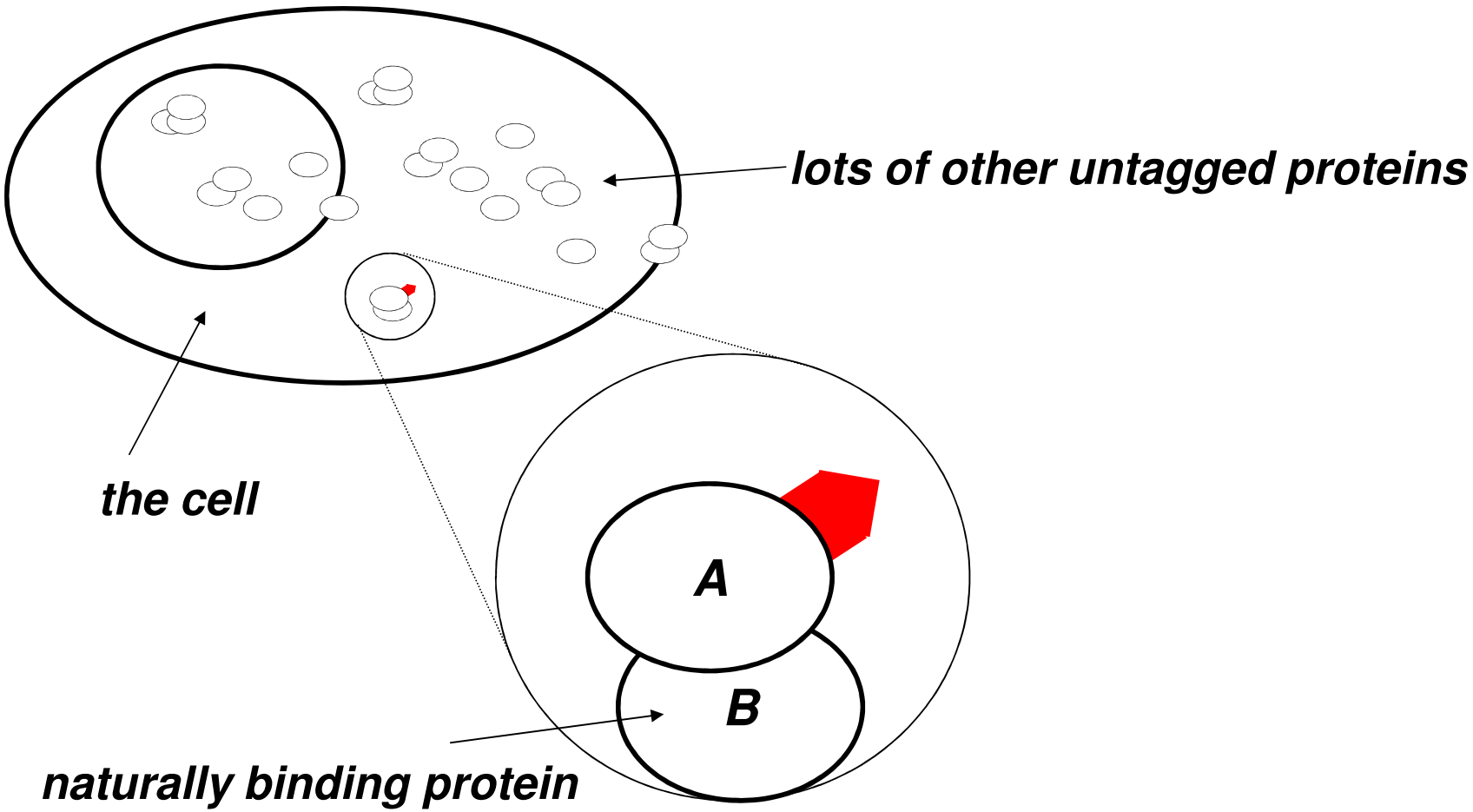
Affinity purification



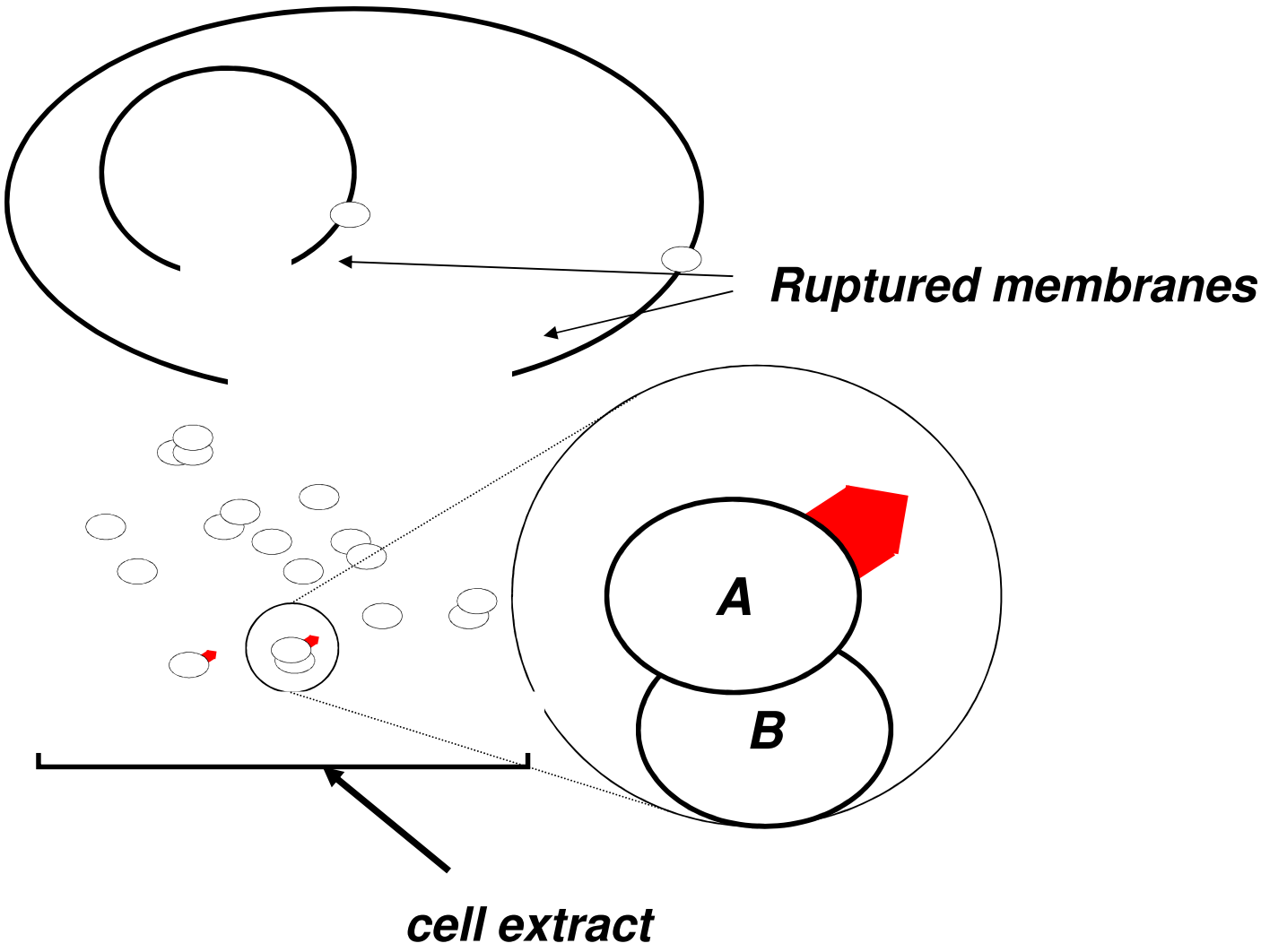
Affinity purification



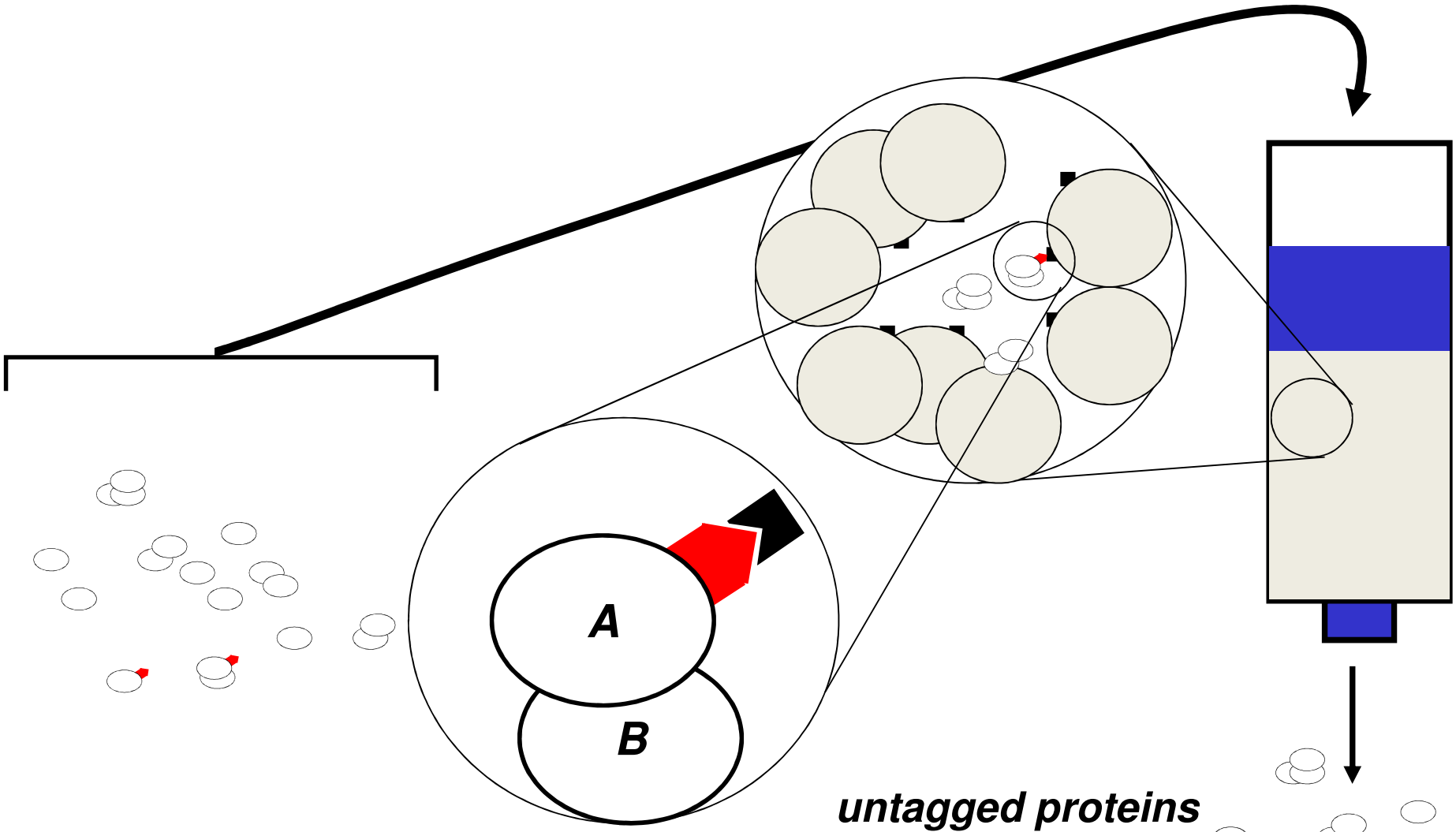
Affinity purification



Affinity purification

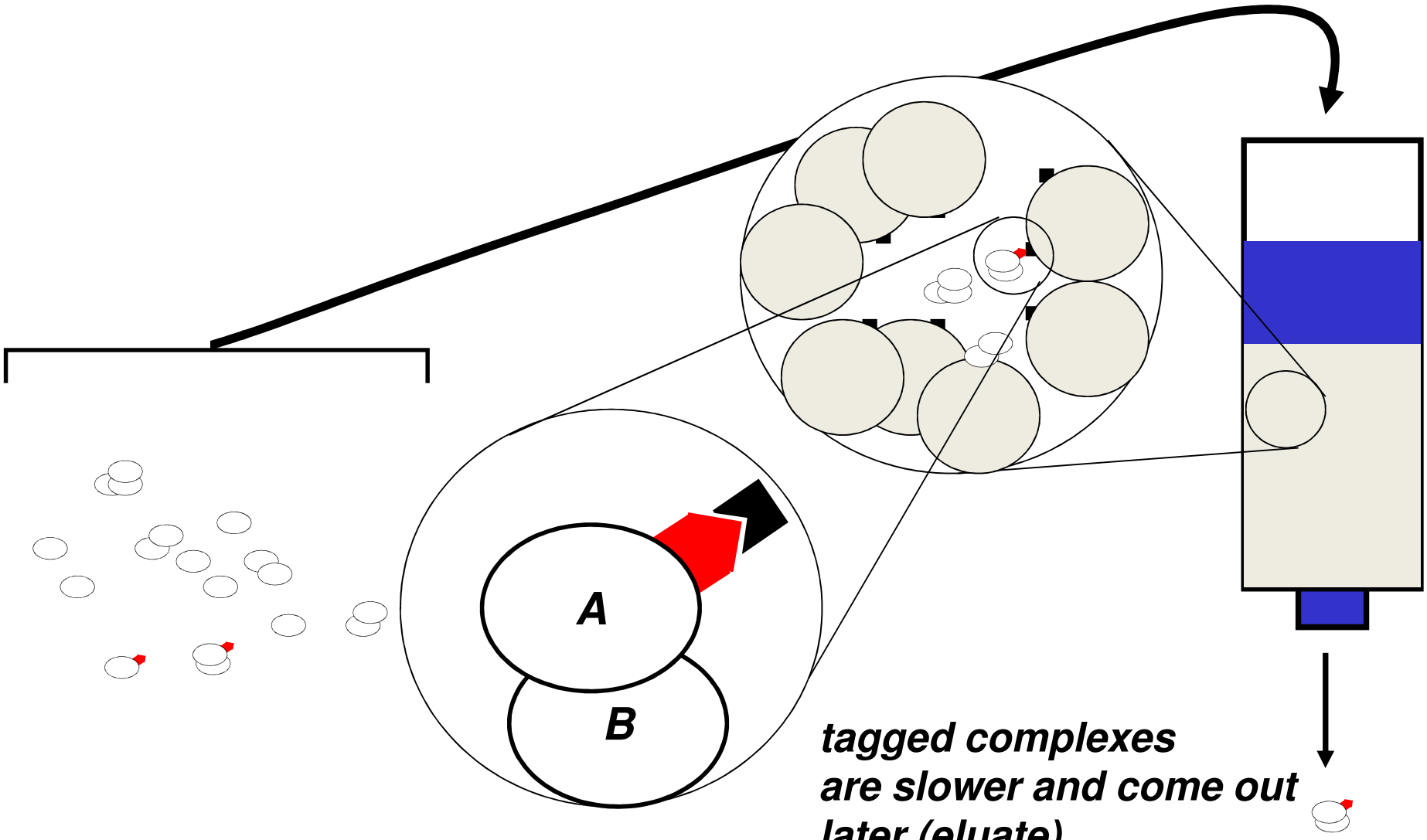


Affinity purification



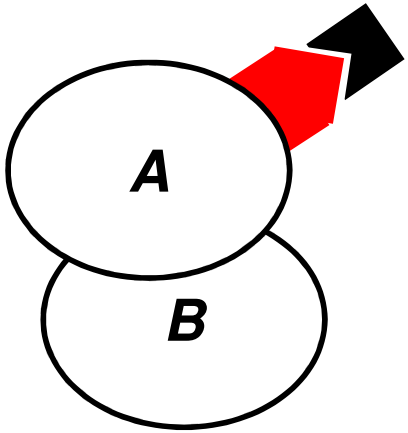
untagged proteins go through fastest (flow-through)

Affinity purification

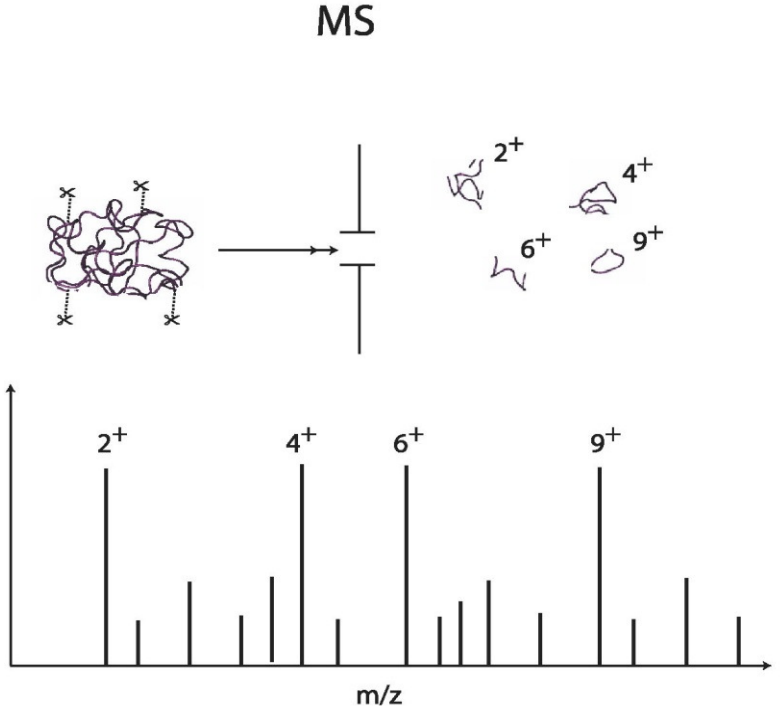


Tandem affinity purification– mass spectrometry (TAP-MS)

Complex pull-down/purification



Mass spec. identification



EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

TAP/MS advantages and disadvantages

Advantages

- detects complexes in their native environment (e.g. not expressed in a yeast cell)
- requires tagging of only one protein at once (not individual pairs as in Y2H)
- ideal for detecting stable protein complexes

Disadvantages

- complex interactions must be extremely robust to be pulled down after lysis, purification, etc. (may miss more transient interactions)
- Interactions are not direct binary relationships– they are inherently complexes (requires “spoke” or “matrix” model for binary transformation)
- Low abundance proteins may be hard to pull-down
- TAP tag may interfere with normal protein interactions

Summary

- Basic concepts
  - An interdisciplinary field
  - Importance of modeling
- Theoretical principles
  - Three levels of theoretical models
- Experimental techniques

Outline

- Modeling gene regulation
- Modeling protein structures
- Modeling protein dynamics
- Modeling protein-protein interactions
- Modeling signaling pathways
- Cell-based simulations
- Multiscale modeling