Systems Biology

Polyomics & Systems Biology

Genomics

Transcriptomics

Proteomics

Metabolomics

XXXomics

What is Systems Biology?

• Systems Biology - *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

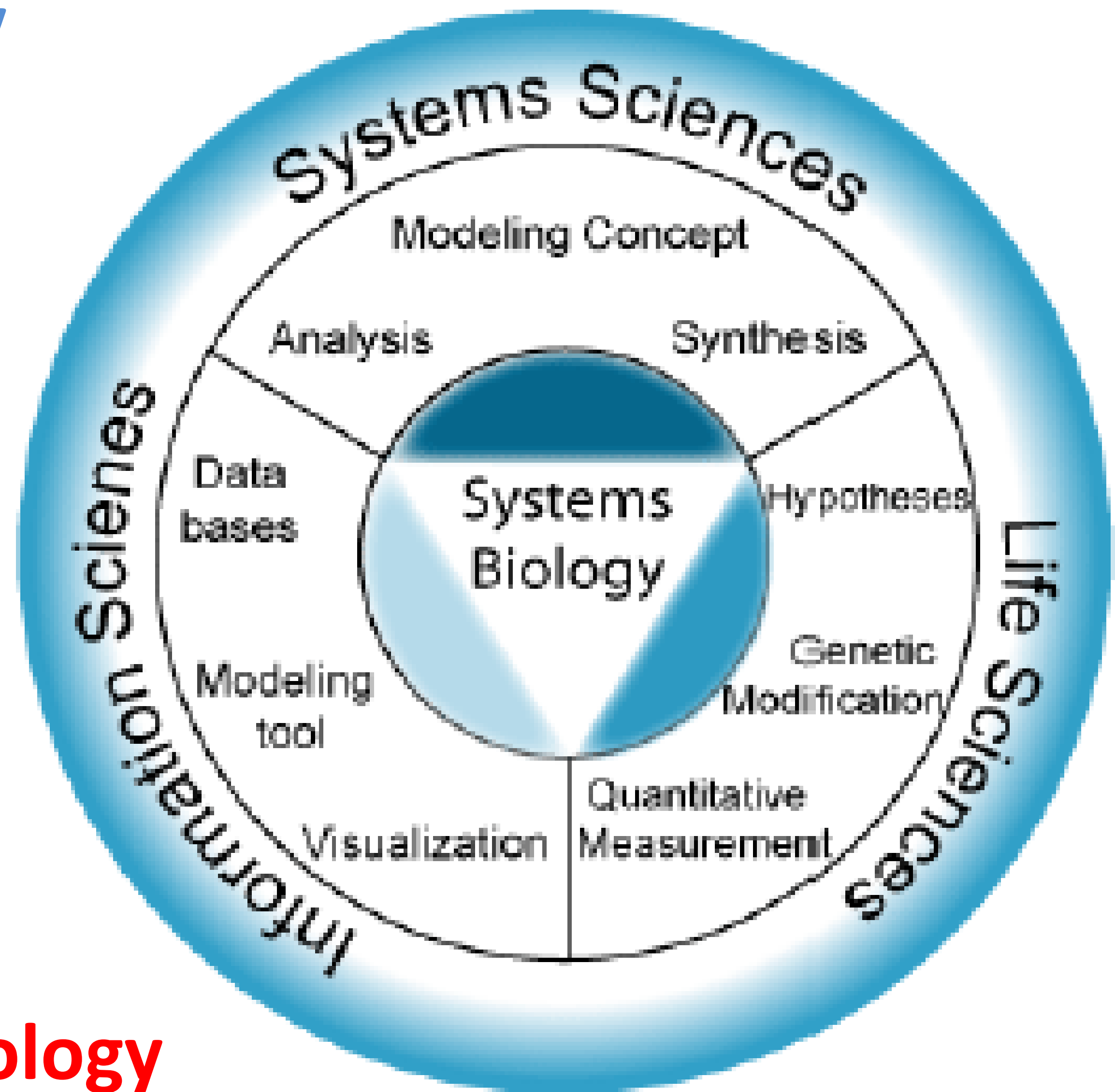
(2) proposal of mathematical models that might account for at least some significant aspects of this data set,

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

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

Systems Biology

Analysis
Understanding

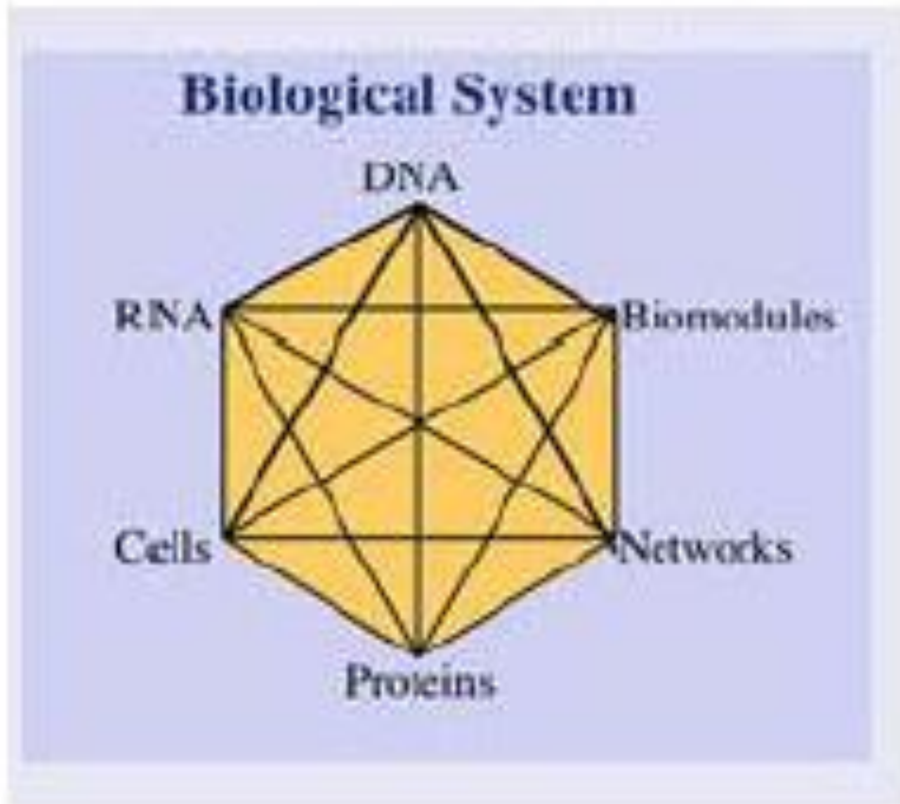


Translation into Applications

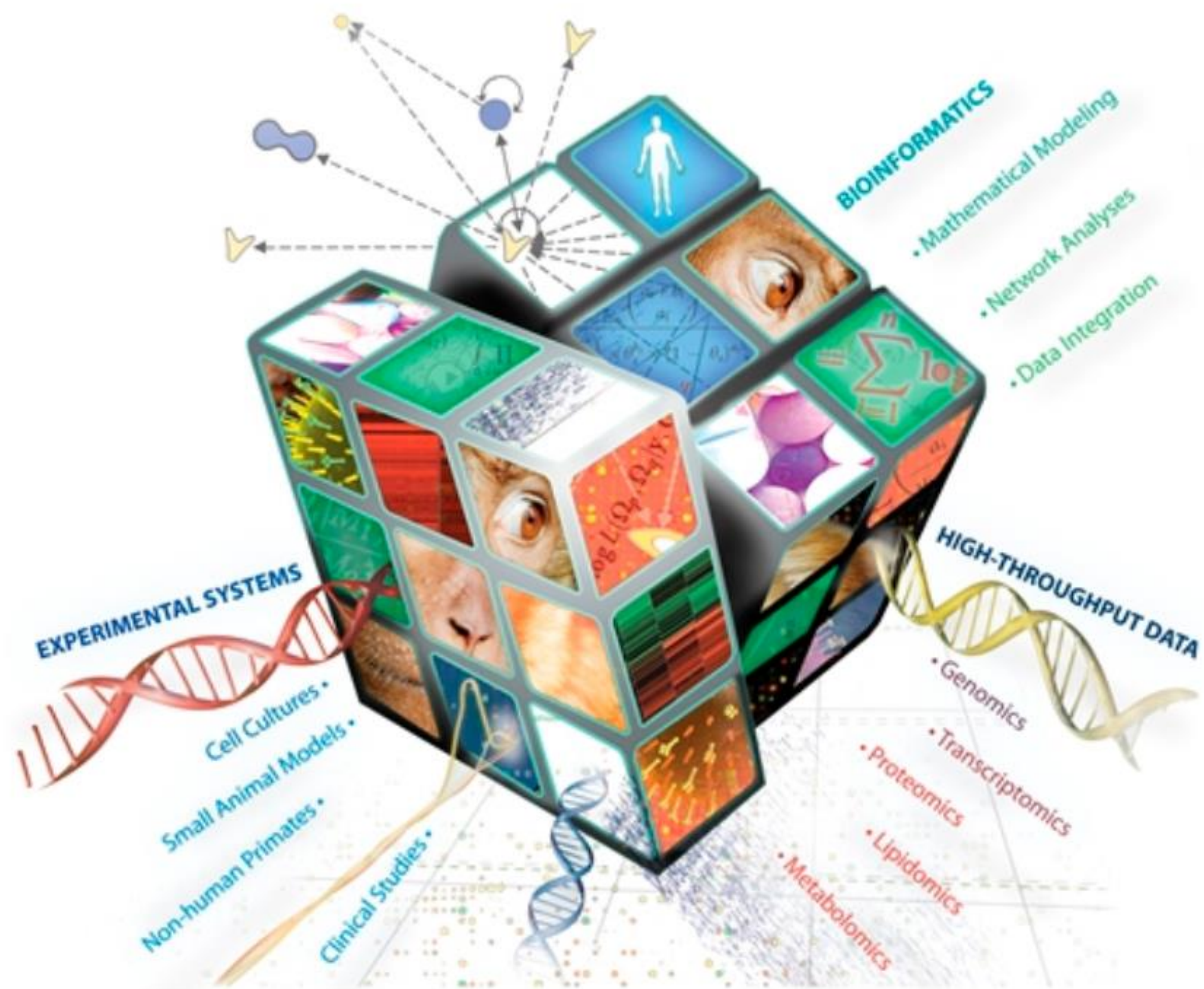
Systems Biotechnology

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Systems Biology



http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf



https://www.systemsvirology.org/sbvhi3/

Systems Biology

What's it good for?

- **Basic Science/"Understanding Life"**
- **Predicting Phenotype from Genotype**
- **Understanding/Predicting Metabolism**
- **Understanding Cellular Networks**
- **Understanding Cell-Cell Communication**
- **Understanding Pathogenicity/Toxicity**
- **"Raising the Bar" for Biologists**

Making Biology a Predictive Science

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Systems Biology

Present Situation

- **100's of completed genomes**
- **1000's of known reactions**
- **10,000's of known 3D structures**
- **100,000's of protein-ligand interactions**
- **1,000,000's of known proteins & enzymes**
- **Decades of biological/chemical know-how**
- **Computational & Mathematical resources**

The Push to Systems Biology

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Systems Biology

The Technologies of Systems Biology

- **Genomics (HT-DNA sequencing)**
- **Mutation detection (SNP methods)**
- **Transcriptomics (Gene/Transcript measurement, SAGE, gene chips, microarrays)**
- **Proteomics (MS, 2D-PAGE, protein chips, Yeast-2-hybrid, X-ray, NMR)**
- **Metabolomics (NMR, X-ray, capillary electrophoresis)**

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Going From Technology to Systems Biology

- **Genomics → Genometrics**
- **Proteomics → Proteometrics**
- **Metabolomics → Metabometrics**
- **Phenomics → Phenometrics**
- **Bioinformatics → Biosimulation**

Quantify, quantify, quantify ……

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Systems Biology

Data aquisition

Single Unit Data ← → Comprising Data Sets

Single Pathways

← → Integrated Networks

Single Regulatory Units

Static Systems ← → Dynamic Systems

Single Systems – examples

Enzyme Catalyzed Reaction Step – Kinetics

Enzyme – catalytic mechanism

Regulatory Protein – Action on Transcription

Receptor – Interaction with Signal Molecule

Genome Analysis

Genome Sequencing

New Technologies – Ultrafast Sequencing

Future Vision: personal Genome → < 1000 €

Sequence Analysis

Annotation

Comparative Genome Analysis

Functional Assays

Genome-wide Gene Knock out Mutagenesis

Genome-wide gene silencing

Next Generation Sequencing

Applied Biosystems Sanger sequencing

Technology : Sanger dideoxy sequencing method

Read Length : up to 1,000bp

Throughput : About 1Mb/day and machine

Roche/454 GS FLX

Technology : Pyrosequencing

Read Length : Average 350-400 bp

Throughput : 400 Mbp/run

Single-read accuracy : > 99.5%

Applied Biosystems SOLiD

As a Certified Sequencing provider, Beckman Coulter Genomics' experience with the SOLiD sequencing platform surpasses that of any other sequencing provider.

Read lengths: 50 bp fragment, 25 bp and 35 bp paired

Throughput (reads): > 160 million reads per slide, fragment

NCBI Data Formats: SOLiD Native or SRF

Robust di-base chemistry for accurate base-calling

Color space analysis with AB proprietary software for high confidence SNP calls

www.beckmangenomics.com

Illumina Genome Analyser

Technology : Reversible terminators and Clonal Single Molecule Array technology

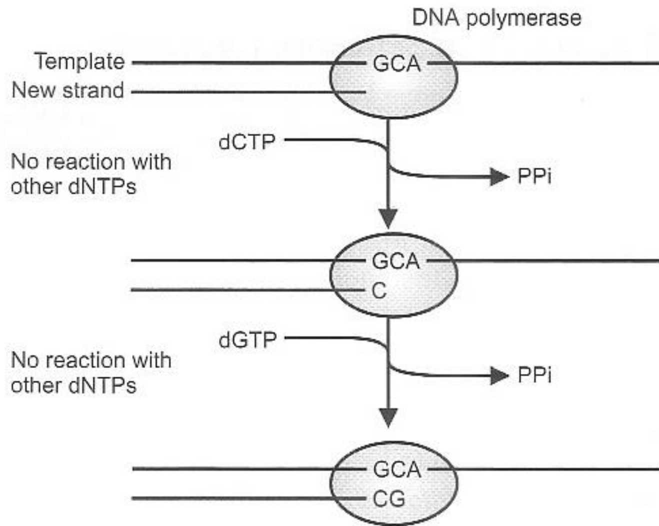
Read Length : 36 bp, up to 75 bp (and growing)

Throughput : 1Gbp per single read, > 3Gbp for paired-end libraries

Per base read accuracy : > 98.5%

http://www.dnavision.com

Pyrosequencing



Pyrophosphate release at each step is detected by a coupled reaction involving ATP sulfurylase, with adenosine 5' phosphosulfate (APS) as substrate, and luciferase

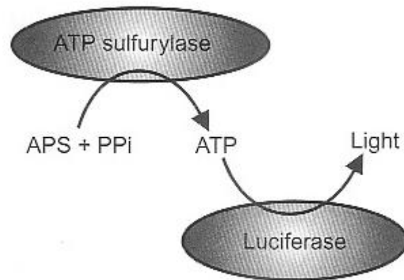


Figure 12.1 Pyrosequencing

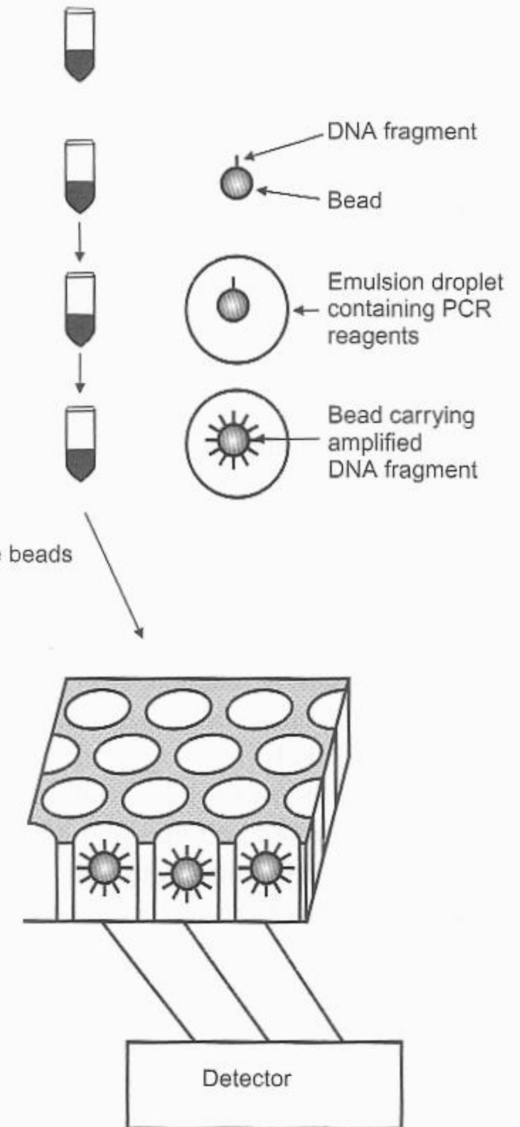
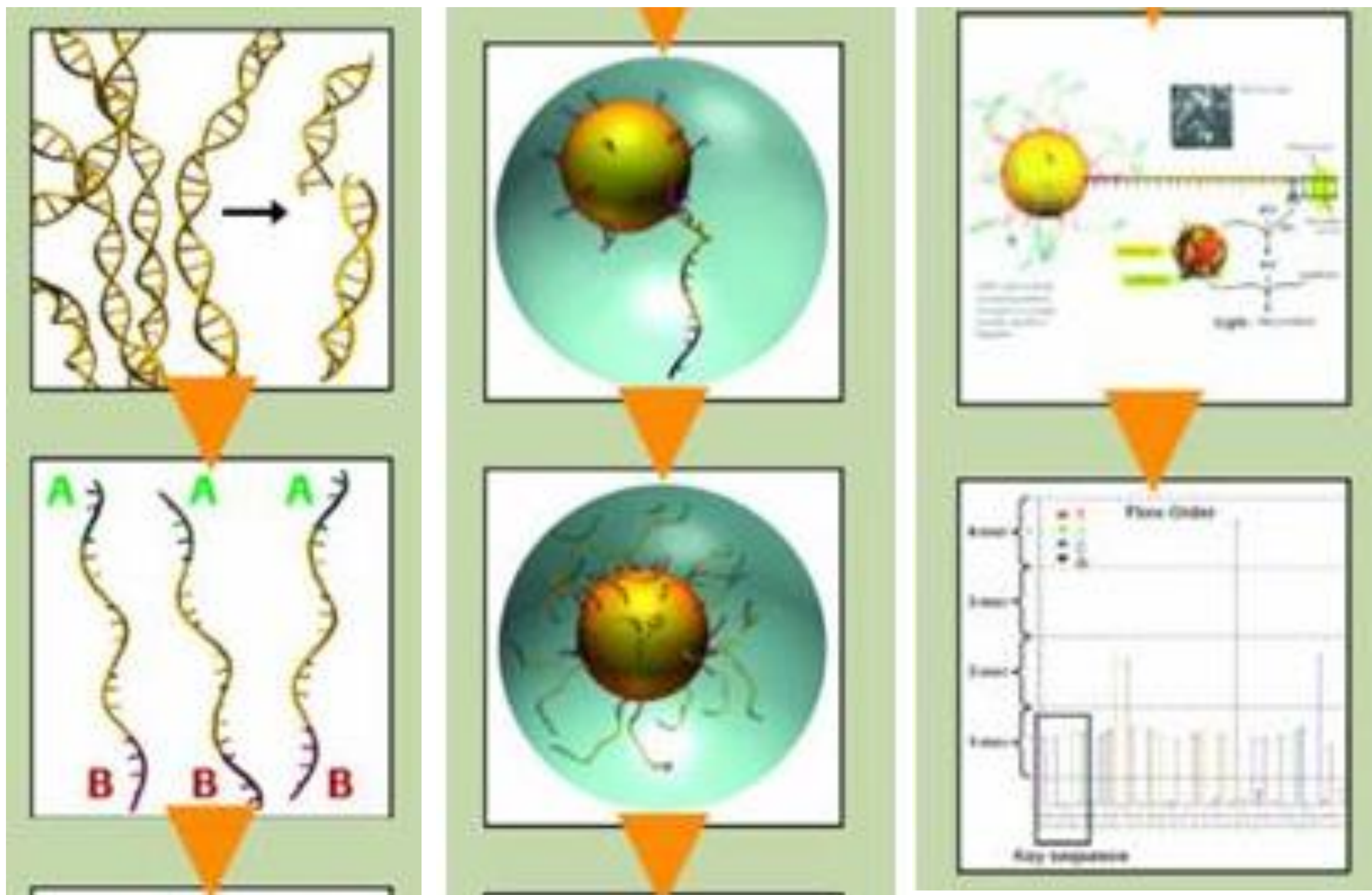
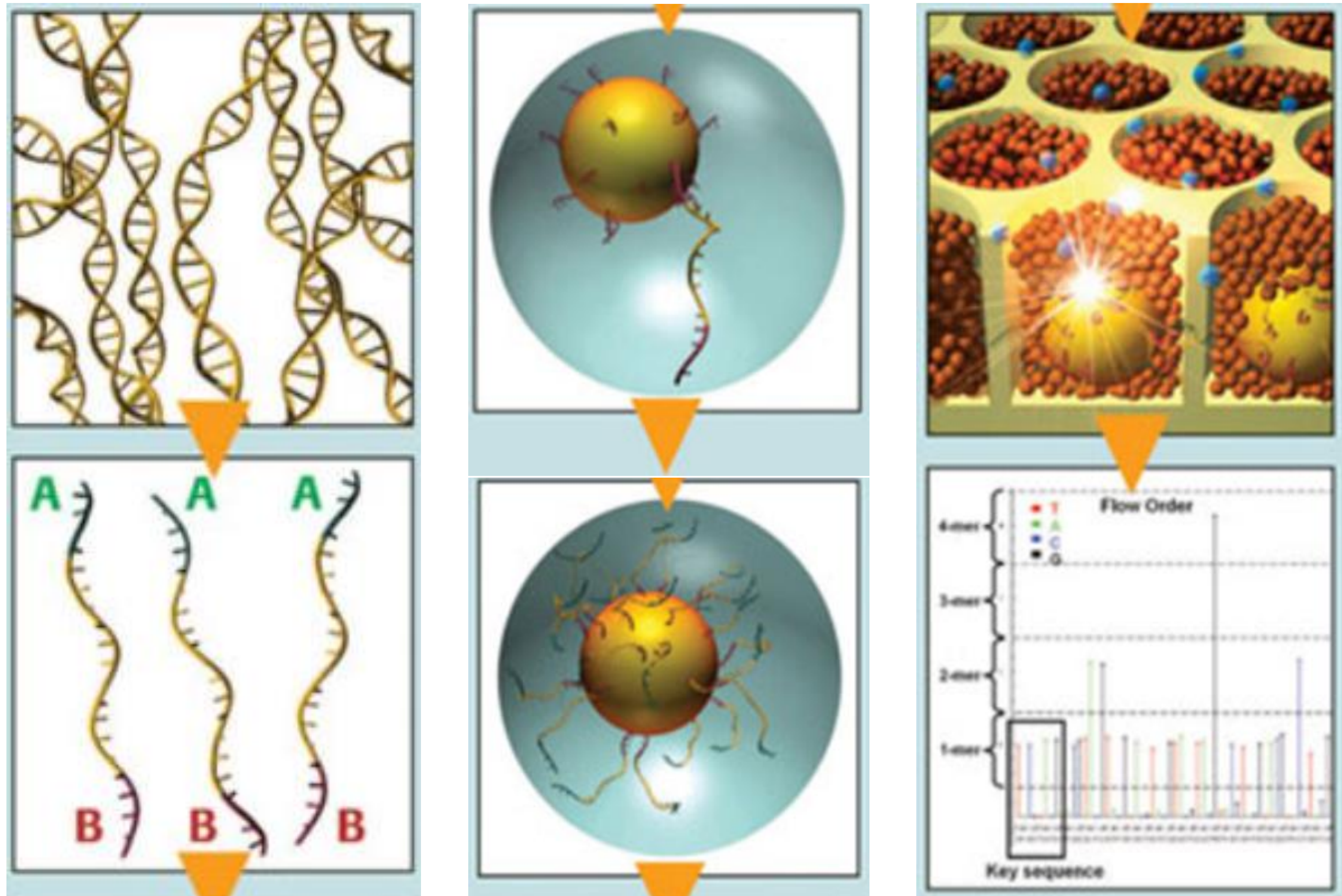


Figure 12.2 Automated pyrosequencing

The Roche 454/GS FLX Sequencing Technology

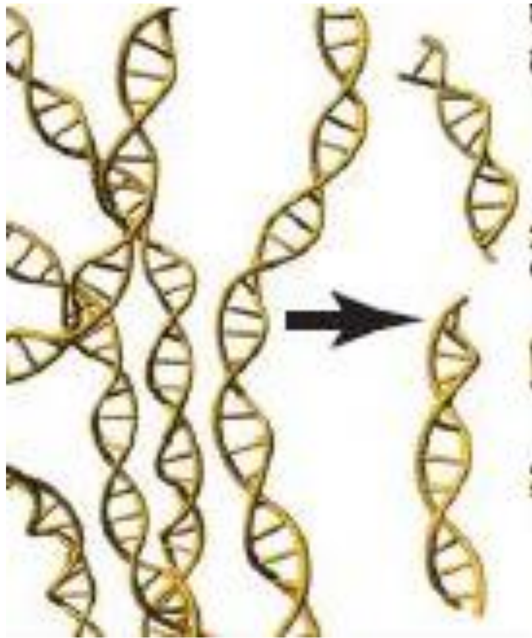


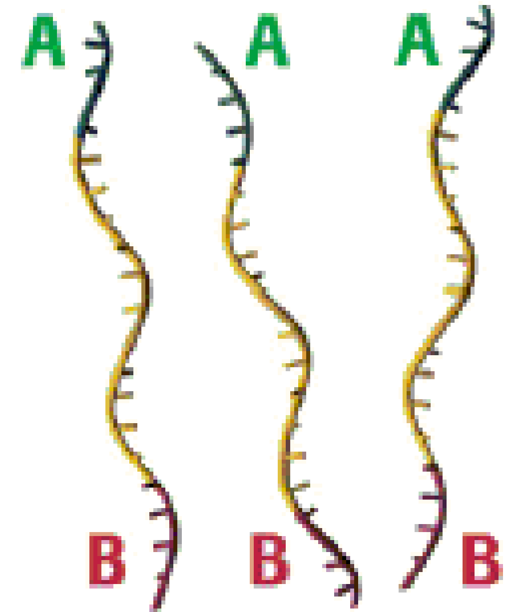
The Roche 454/GS FLX Sequencing Technology



http://www.genengnews.com/gen-articles/third-generation-sequencing-debuts/3257/?page=2

The Roche 454/GS FLX Sequencing Technology

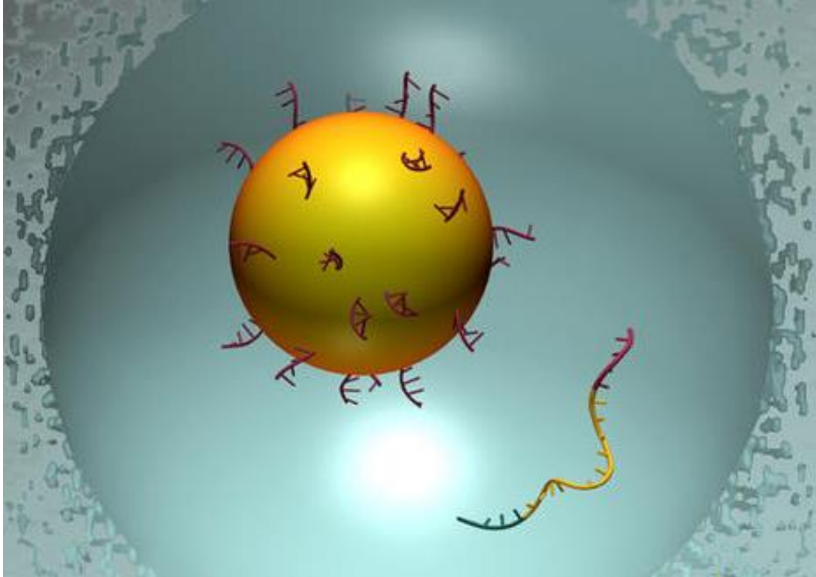




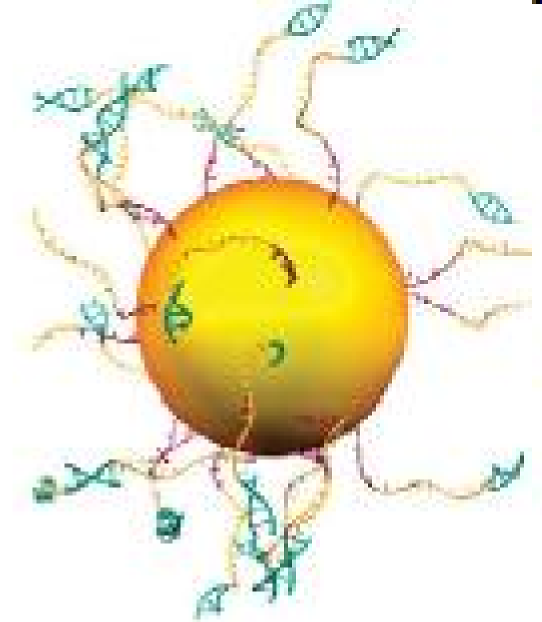
The GS FLX sequencer supports sequencing of various different nucleic acid starting materials such as genomic DNA, PCR products, BACs and cDNA. Samples consisting of longer sequences are first sheared into a random library of 300-800 base-pair long fragments.

http://www.454.com/

The Roche 454/GS FLX Sequencing Technology



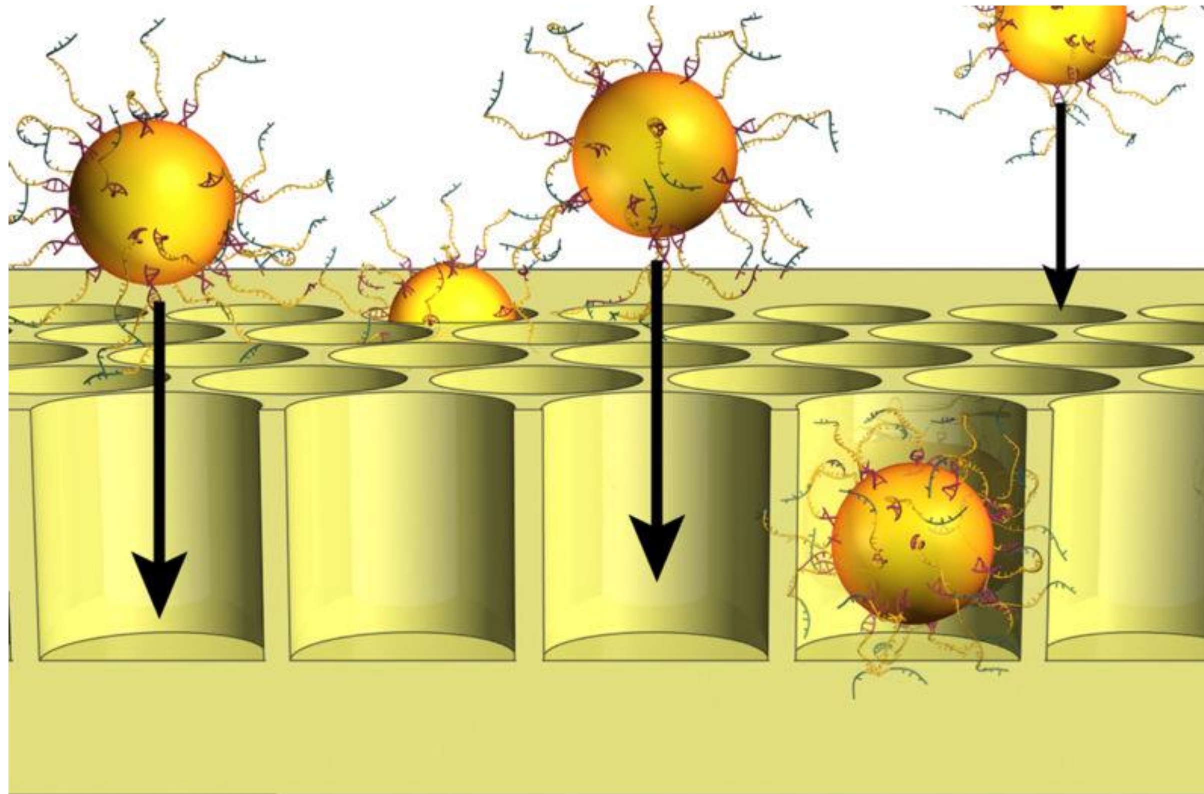
Aided by the adaptors individual fragments are captured on their own unique beads. A bead and the bound fragment together with a water-in-oil emulsion form a microreactor so that each fragment can be amplified without contamination via the so called emulsion PCR (emPCR). The entire fragment collection is amplified in parallel.



The emPCR amplifies each fragment several million times. After amplification the emulsion shell is broken and the clonally amplified beads are ready for loading onto the fibre-optic PicoTiterDevice for sequencing

http://www.454.com/

Pyro-Sequencing – 454 Technology



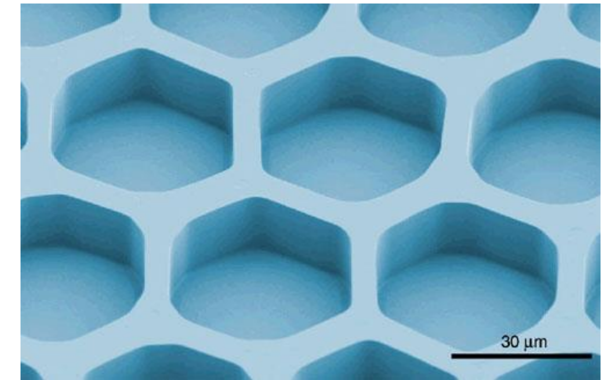
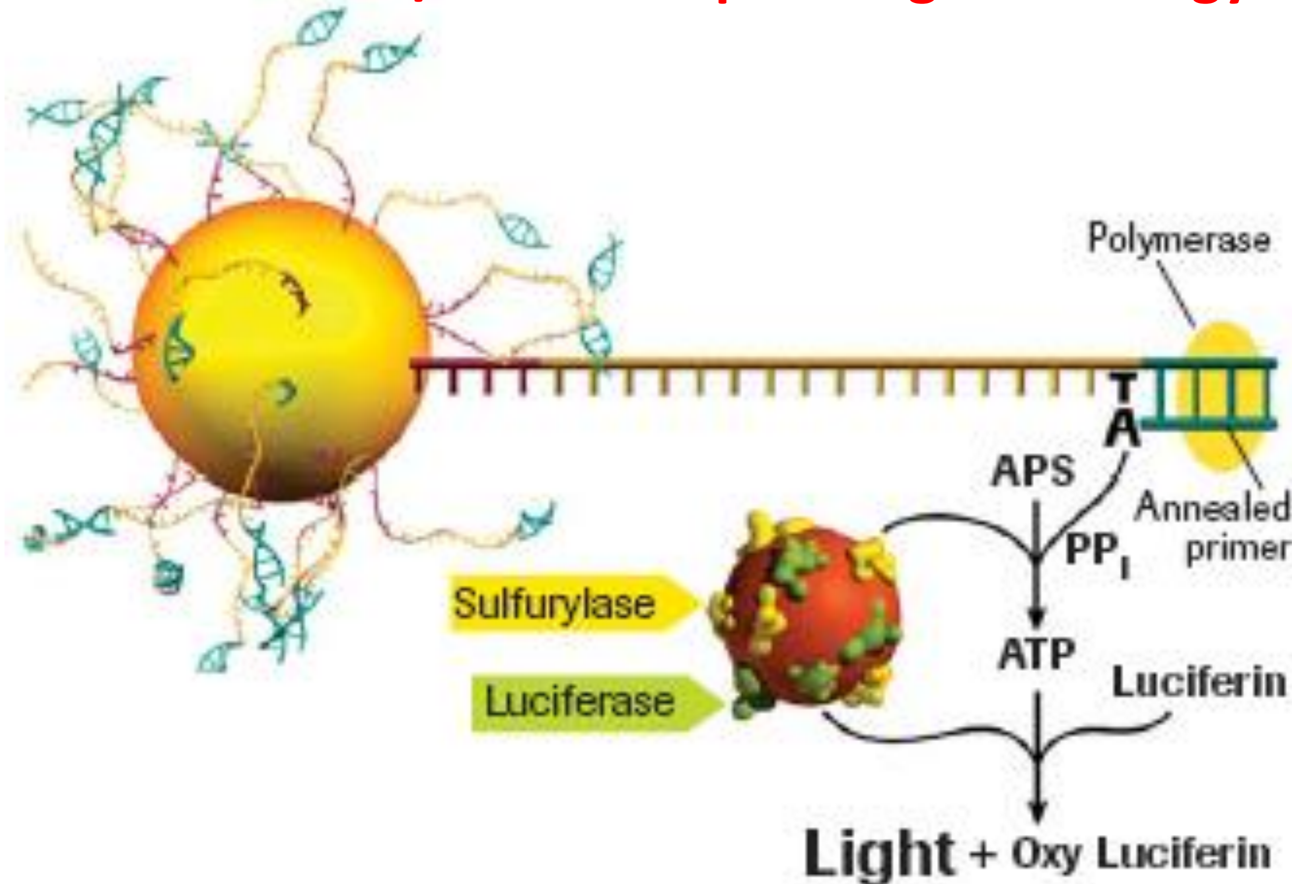


Figure 1. Scanning electron micrograph of etched well in 454 PicoTiter sequencing plate. 454's technology is based on performing hundreds of thousands of simultaneous sequencing reactions in 75 picoliter (44 µm) wells. All molecular biology reactions—DNA amplification, sequencing by synthesis, and signal light generation—occur in a single well.

http://www.454.com/

The Roche 454/GS FLX Sequencing Technology



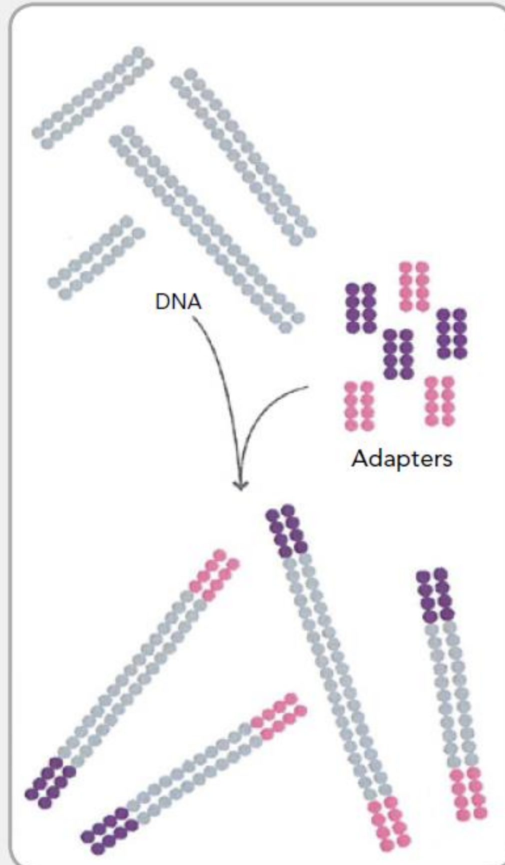
Sequencing is accomplished by synthesizing the complementary strands of the bead attached templates. In a number of cycles the four bases (ATGC) are sequentially washed over the PicoTiterPlate. The incorporation of a new base is associated with the release of inorganic pyrophosphate starting a chemical cascade. This results in the generation of a light signal which is captured by a CCD camera.

http://www.454.com/

DNA Sequencing with Illumina (Solexa®) Technology

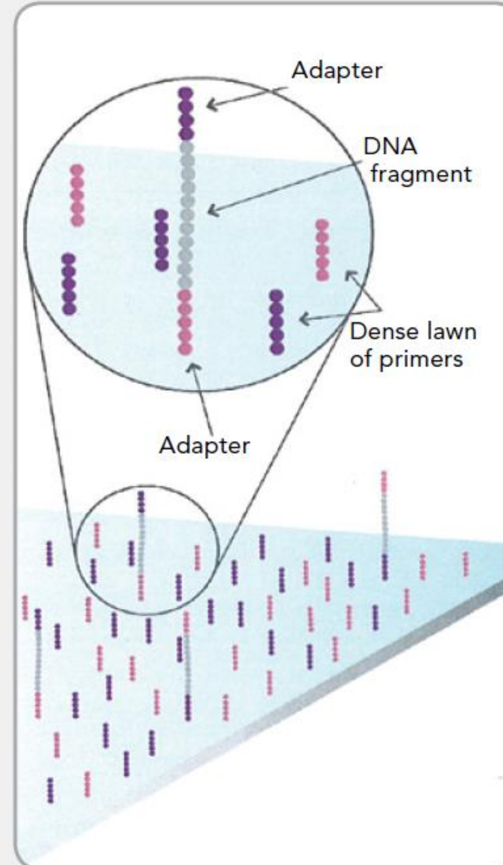
FIGURE 2: SEQUENCING TECHNOLOGY OVERVIEW

1. PREPARE GENOMIC DNA SAMPLE



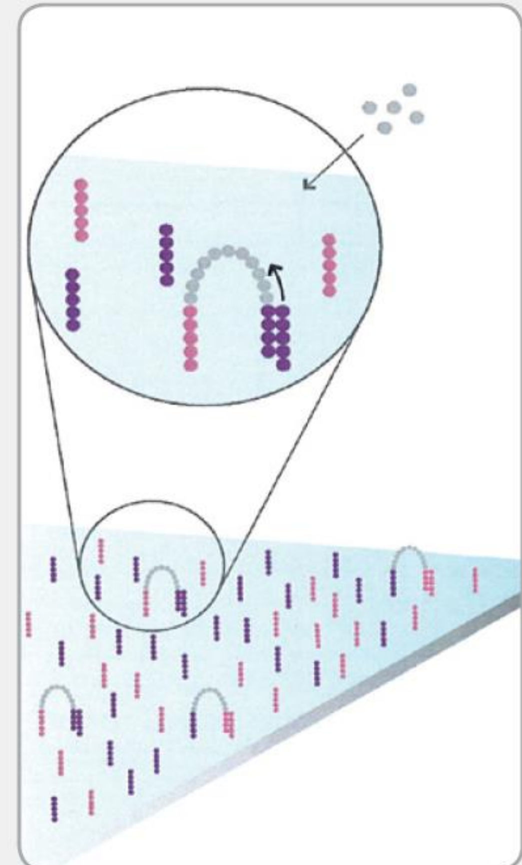
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION

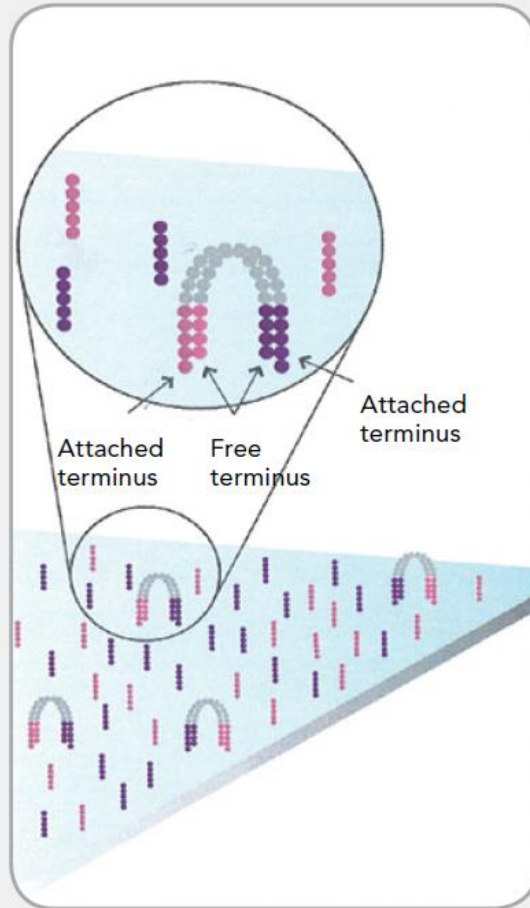


Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

www.illumina.com/documents/products/.../techspotlight_sequencing.pdf

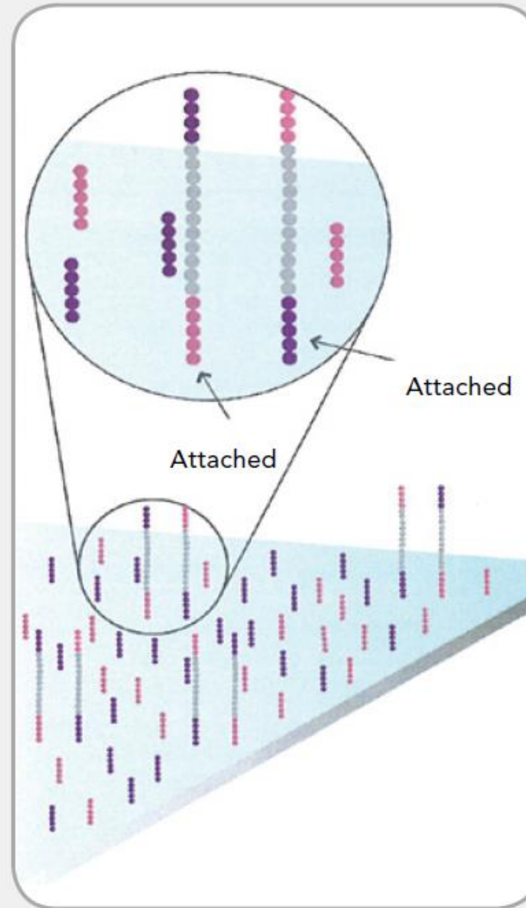
DNA Sequencing with Illumina (Solexa®) Technology

4. FRAGMENTS BECOME DOUBLE STRANDED



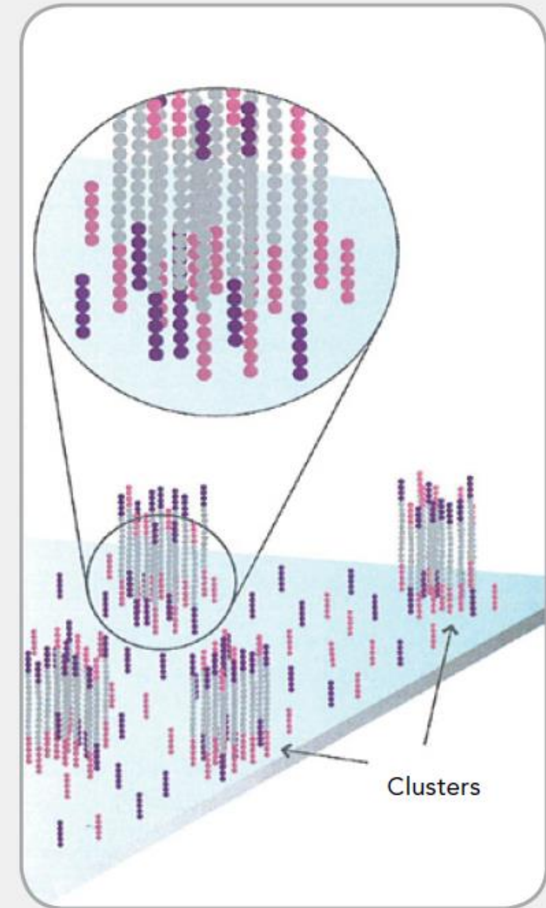
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



Denaturation leaves single-stranded templates anchored to the substrate.

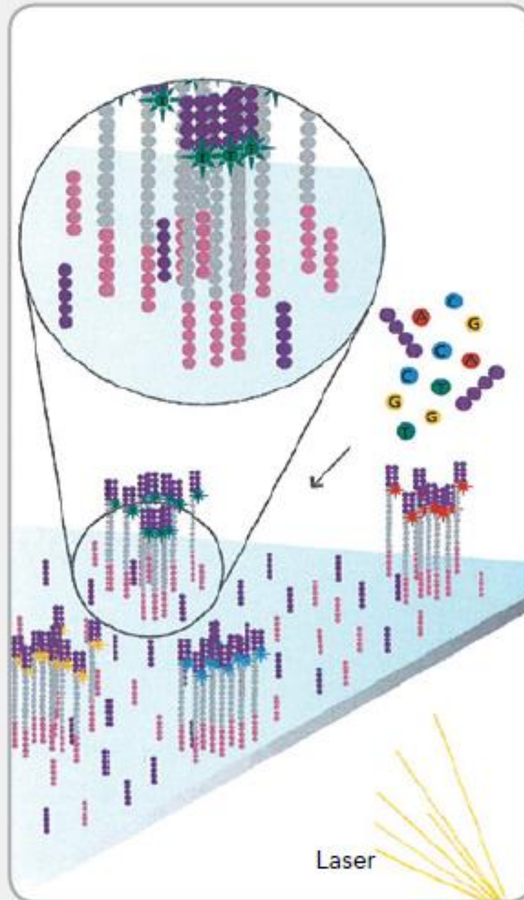
6. COMPLETE AMPLIFICATION



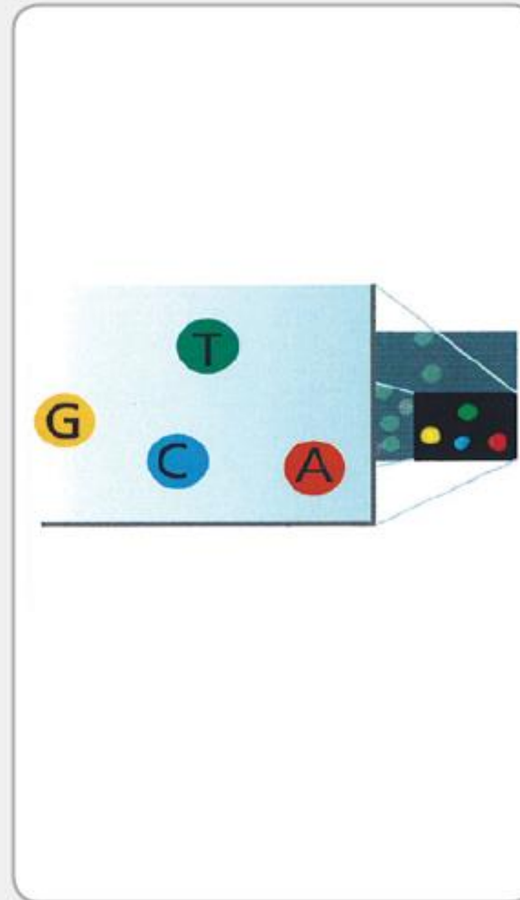
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

www.illumina.com/documents/products/.../techspotlight_sequencing.pdf

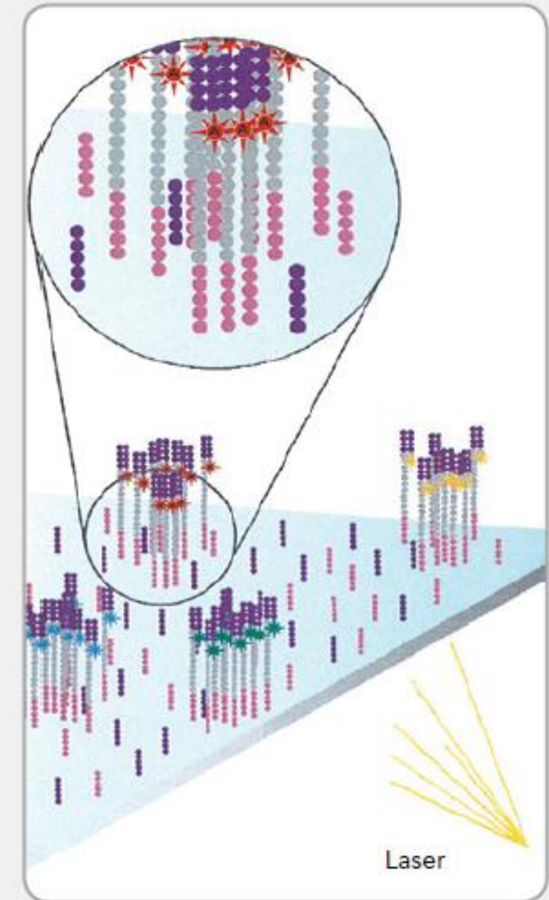
DNA Sequencing with Illumina (Solexa®) Technology



First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.



After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

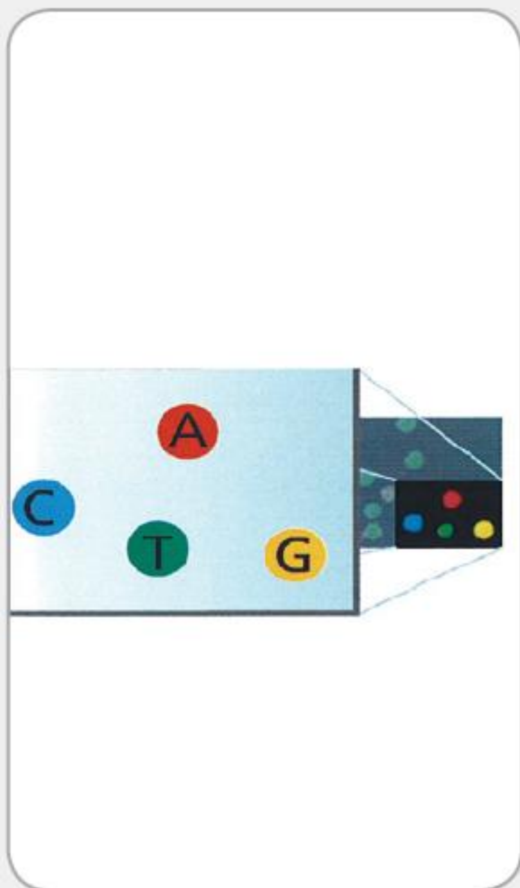


Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

www.illumina.com/documents/products/.../techspotlight_sequencing.pdf

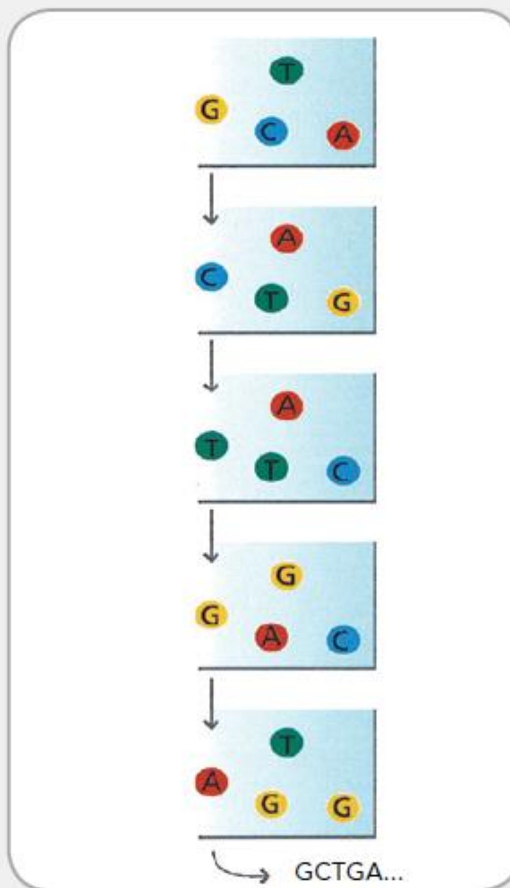
DNA Sequencing with Illumina (Solexa®) Technology

10. IMAGE SECOND CHEMISTRY CYCLE



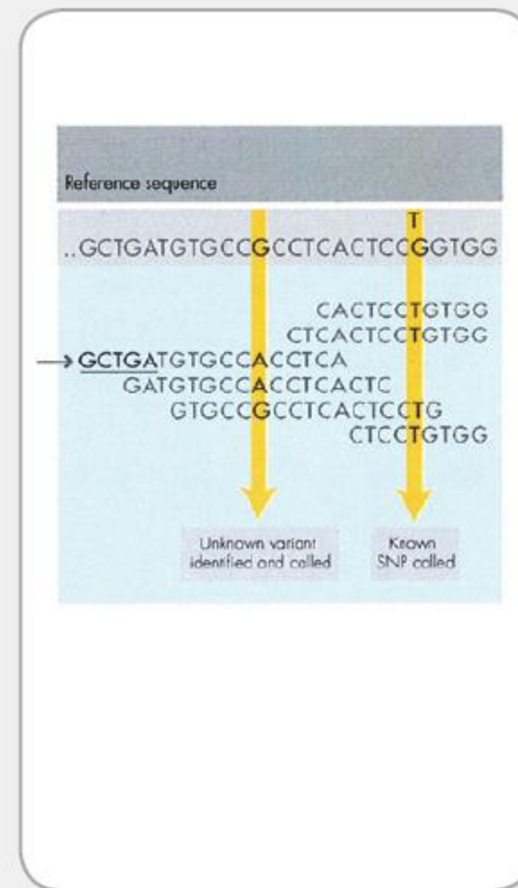
After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES



Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at time.

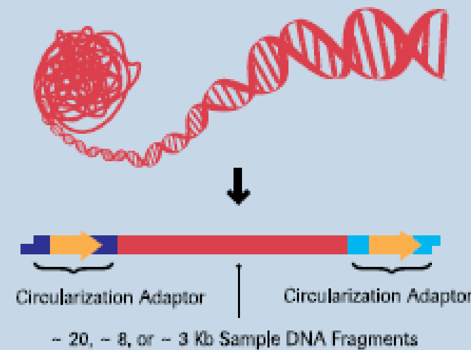
12. ALIGN DATA



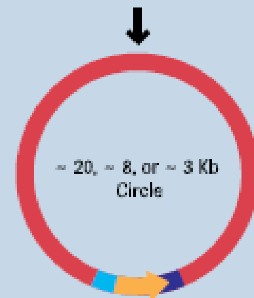
Align data, compare to a reference, and identify sequence differences.

www.illumina.com/documents/products/.../techspotlight_sequencing.pdf

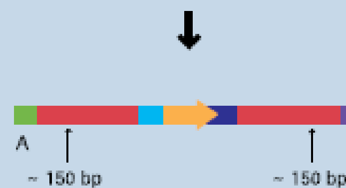
The GS FLX Titanium Series Paired End Protocol



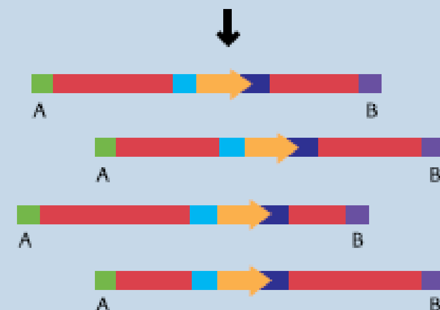
1. **Circularization-Ready Fragments.** Genomic DNA is sheared into 20 Kb, 8 Kb or 3 Kb fragments and adaptors are added to the end of each fragment.



2. **Circularized DNA.** DNA is circularized.



3. **Paired End Library Construct.** The circularized DNA is fragmented and fragments containing the added adaptors are isolated and amplified for sequencing.



4. **Paired End Library.** Resulting library consists of true paired end reads with two end tags averaging over 150 bp and separated by 20 Kb, 8 Kb or 3 Kb.

http://www.454.com/

Genome Analysis

Genome Sequencing

New Technologies – Ultrafast Sequencing

Future Vision: personal Genome → < 1000 €

Sequence Analysis

Annotation

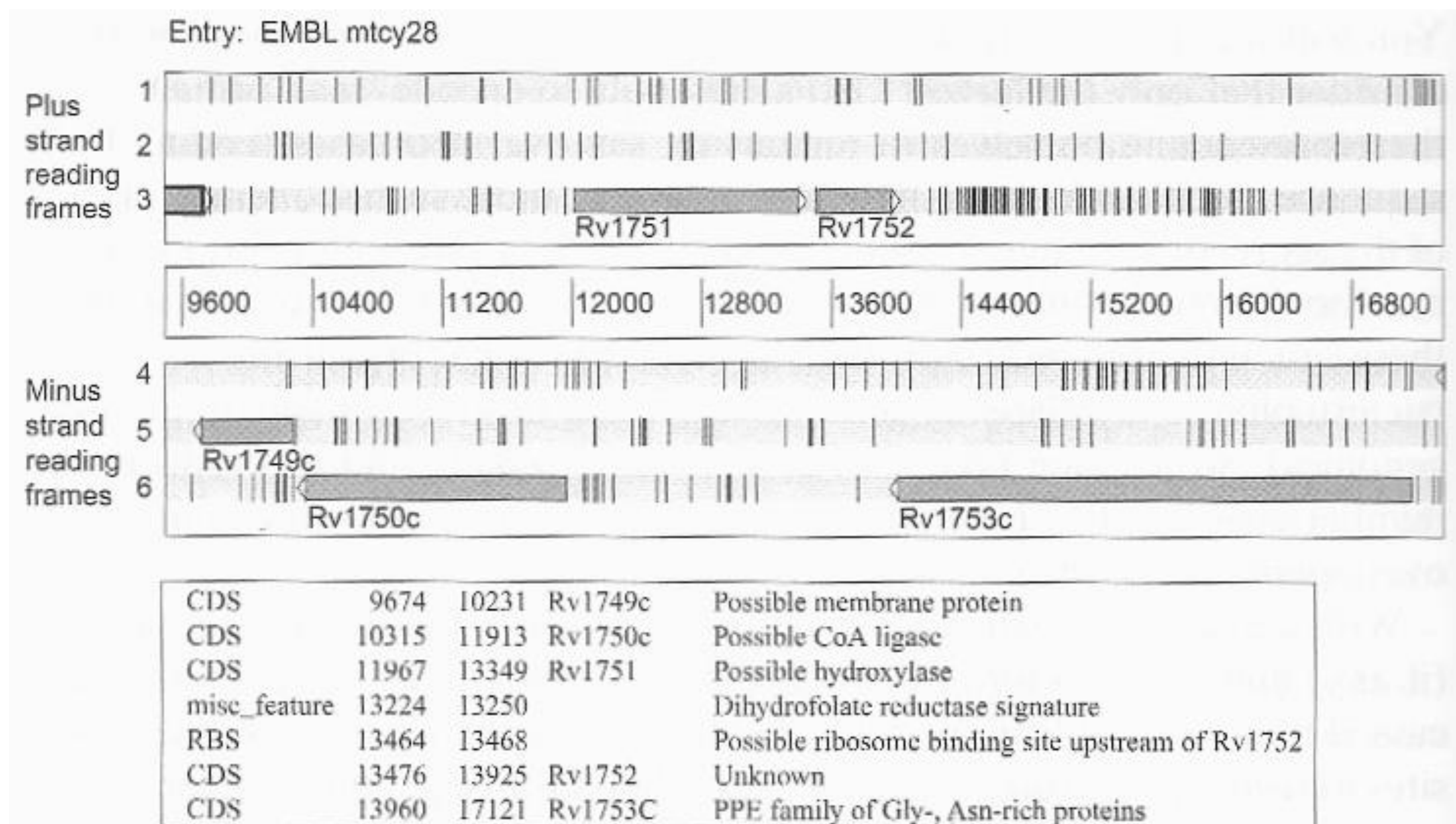
Comparative Genome Analysis

Functional Assays

Genome-wide Gene Knock out Mutagenesis

Genome-wide gene silencing

Sequence Analysis - Annotation



CDS	9674	10231	Rv1749c	Possible membrane protein
CDS	10315	11913	Rv1750c	Possible CoA ligase
CDS	11967	13349	Rv1751	Possible hydroxylase
misc_feature	13224	13250		Dihydrofolate reductase signature
RBS	13464	13468		Possible ribosome binding site upstream of Rv1752
CDS	13476	13925	Rv1752	Unknown
CDS	13960	17121	Rv1753C	PPE family of Gly-, Asn-rich proteins

Figure 12.6 Open reading frames: display of coding sequences; edited display from analysis of a DNA sequence and databank annotations using Artemis

Sequence Analysis - Comparison

Figure 12.11 Genome structure comparison using gMAP. (a) Comparison of entire genomes; (b) comparison of selection region 3, from (a); (c) detailed comparison of selected region 3, from (b)

Genome Analysis

Genome Sequencing

New Technologies – Ultrafast Sequencing

Future Vision: personal Genome → < 1000 €

Sequence Analysis

Annotation

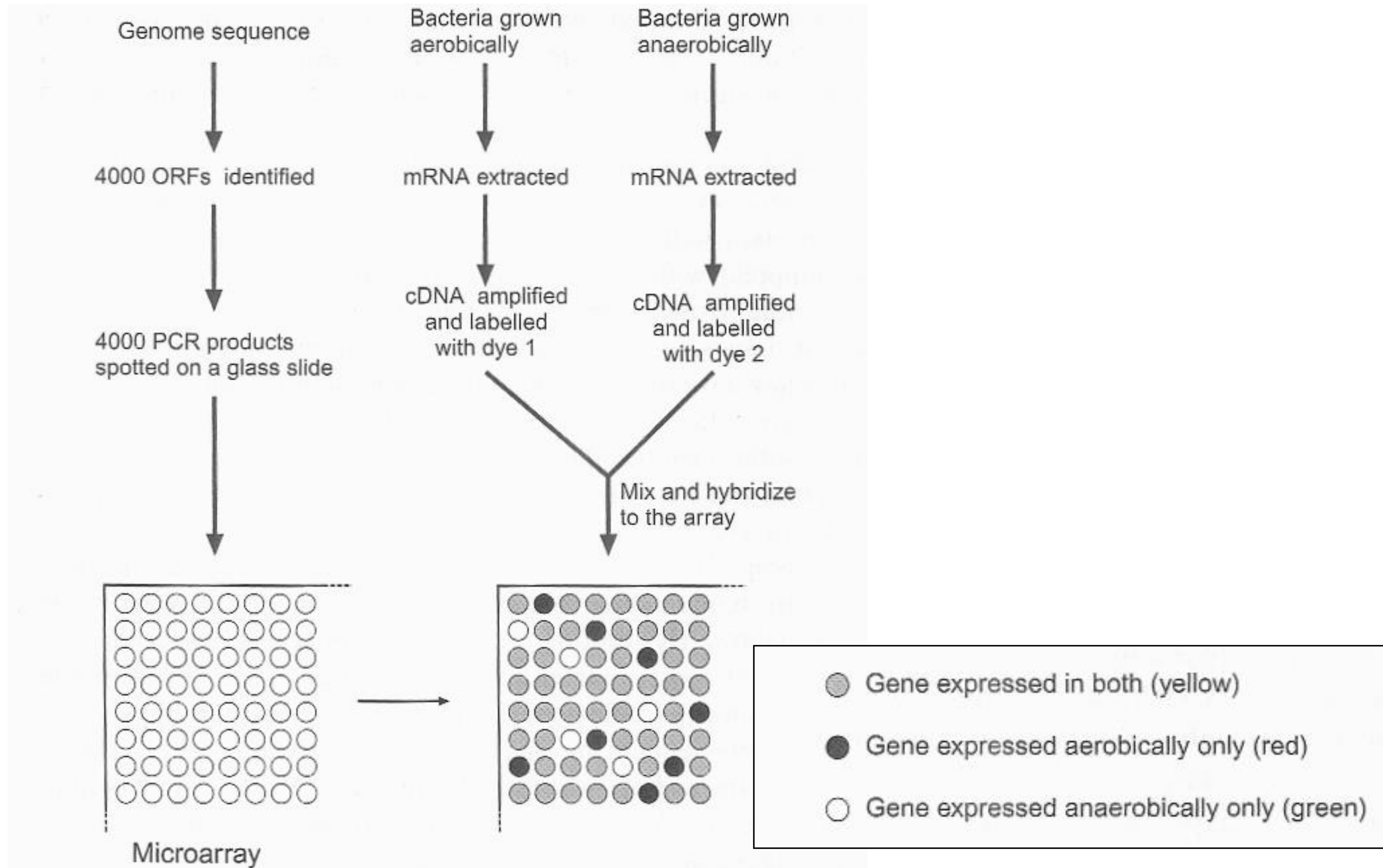
Comparative Genome Analysis

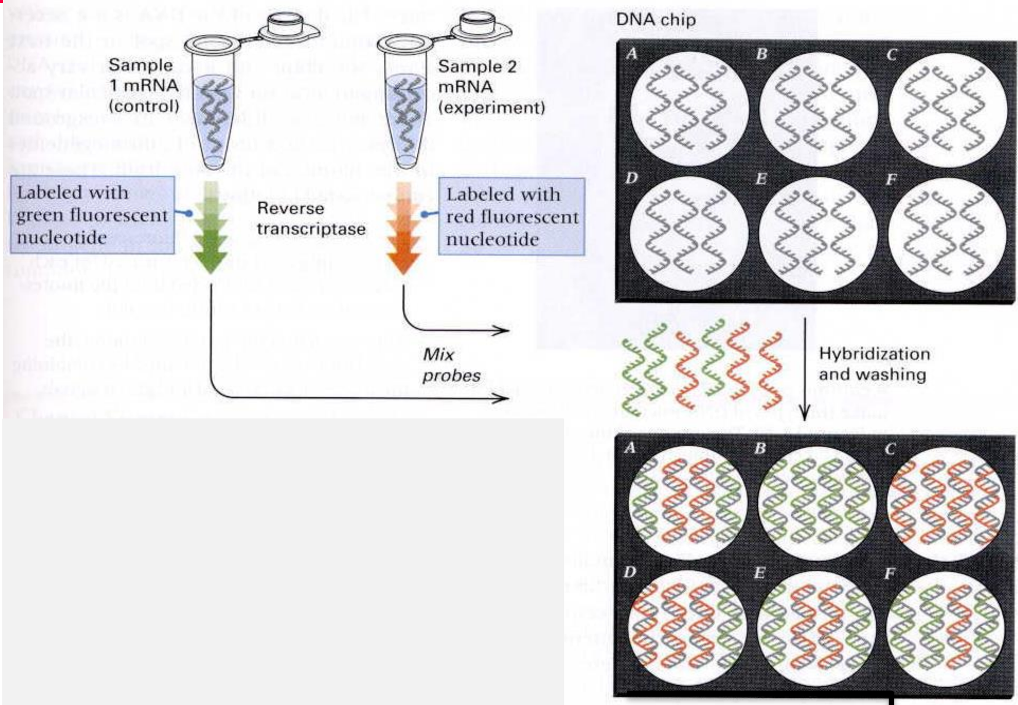
Functional Assays

Genome-wide Gene Knock out Mutagenesis

Genome-wide gene silencing

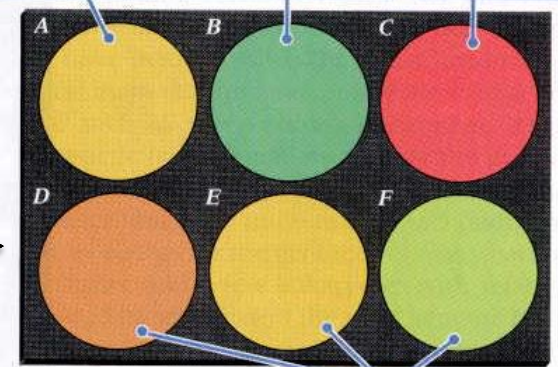
Transcriptome Analysis



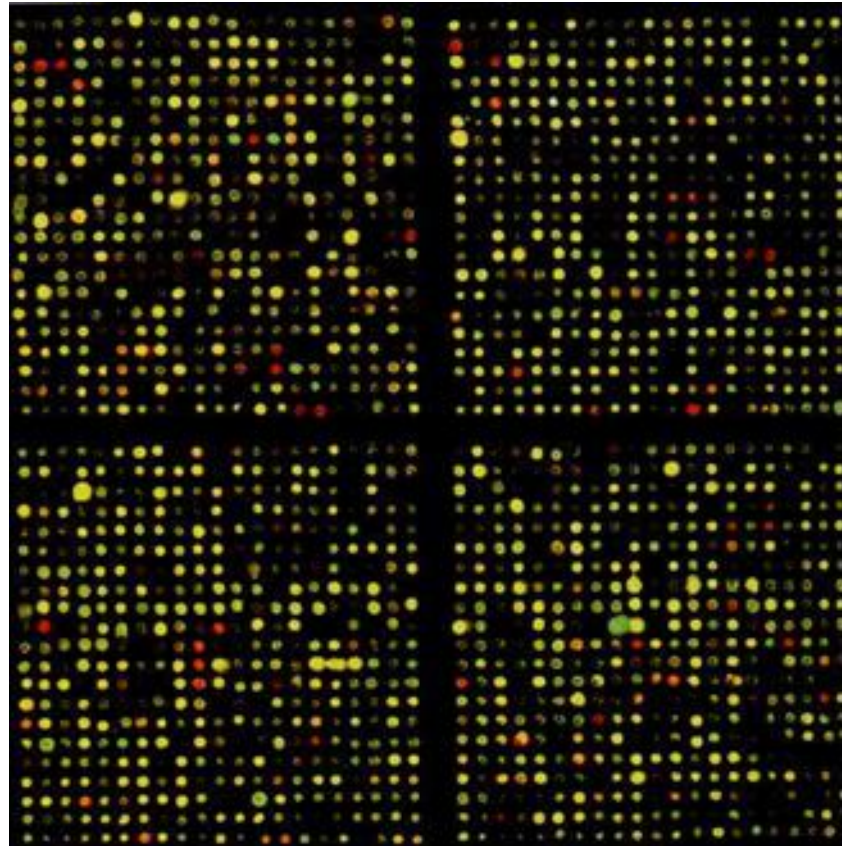


Principle of operation of one type of DNA chip.

Confocal microscope fluorescence scanning



At the top are dried microdrops, each of which contains immobilized DNA strands from a different gene (*A*-*F*). These are hybridized with a mixture of fluorescence-labeled DNA samples obtained by reverse transcription of cellular mRNA. Competitive hybridization of red (experimental) and green (control) label is proportional to the relativeabundance of each mRNA species in the samples. The relative levels or red and green fluorescence of each spot are assayed by microscopic scanning and displayed as a single color. Red or orange indicates overexpression in the experimental sample, green or yellow-green underexpression in the experimental sample and yellow equal expression.



Small part of a yeast DNA chip showing 1764 spots, each specific for hybridization with a different mRNA sequence. The color of each spot indicates the relative level of gene expression in experimental and control samples. The complete chip for allyeast open reading frames includes over 6200 spots. [Courtesy of Jeffrey P. Townsend, Duccio Cavalieri and the Harvard Center for Genomics Research]

Gene Silencing

RNAi – RNA Interference

RISC → RNA Induced Silencing Complex

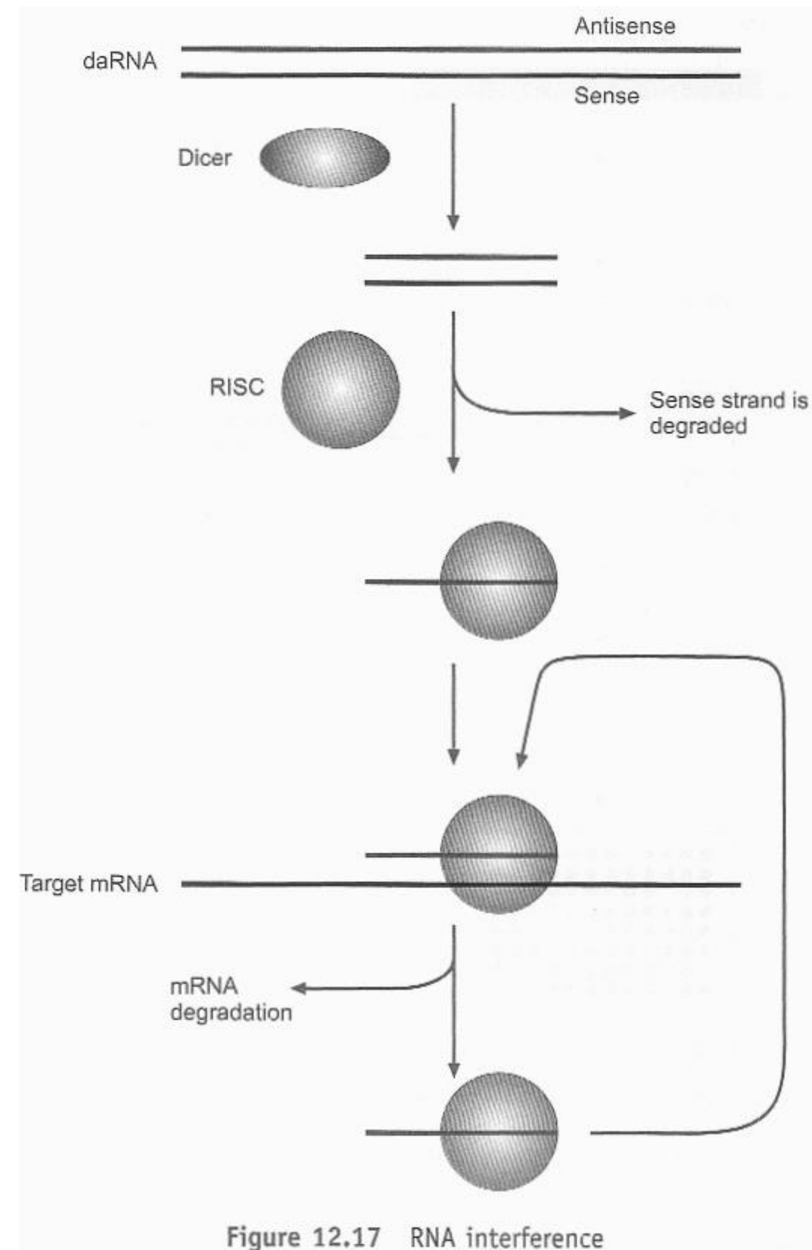


Figure 12.17 RNA interference

RNA Interference

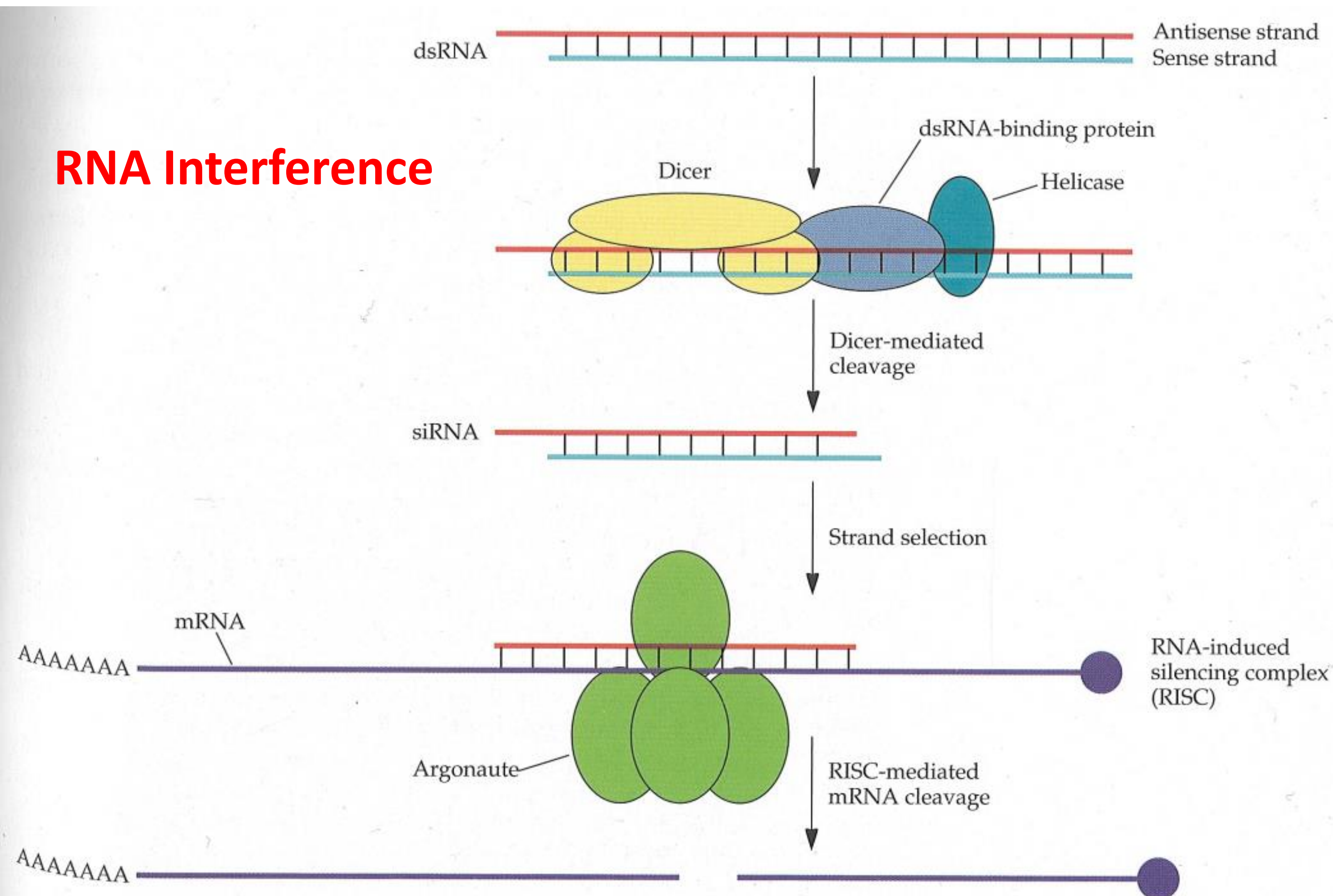


FIGURE 11.13 Overview of the process of RNA interference. Following introduction of double-stranded RNA (dsRNA) into a cell, the Dicer complex binds to the RNA and cleaves it into an siRNA containing approximately 21 bp. The antisense strand (red) becomes part of the RISC, directing the cleavage of the complementary mRNA.

Glick et al., Molecular Biotechnology – Principles and Applications of Recombinant DNA, 4th ed., ASM Press

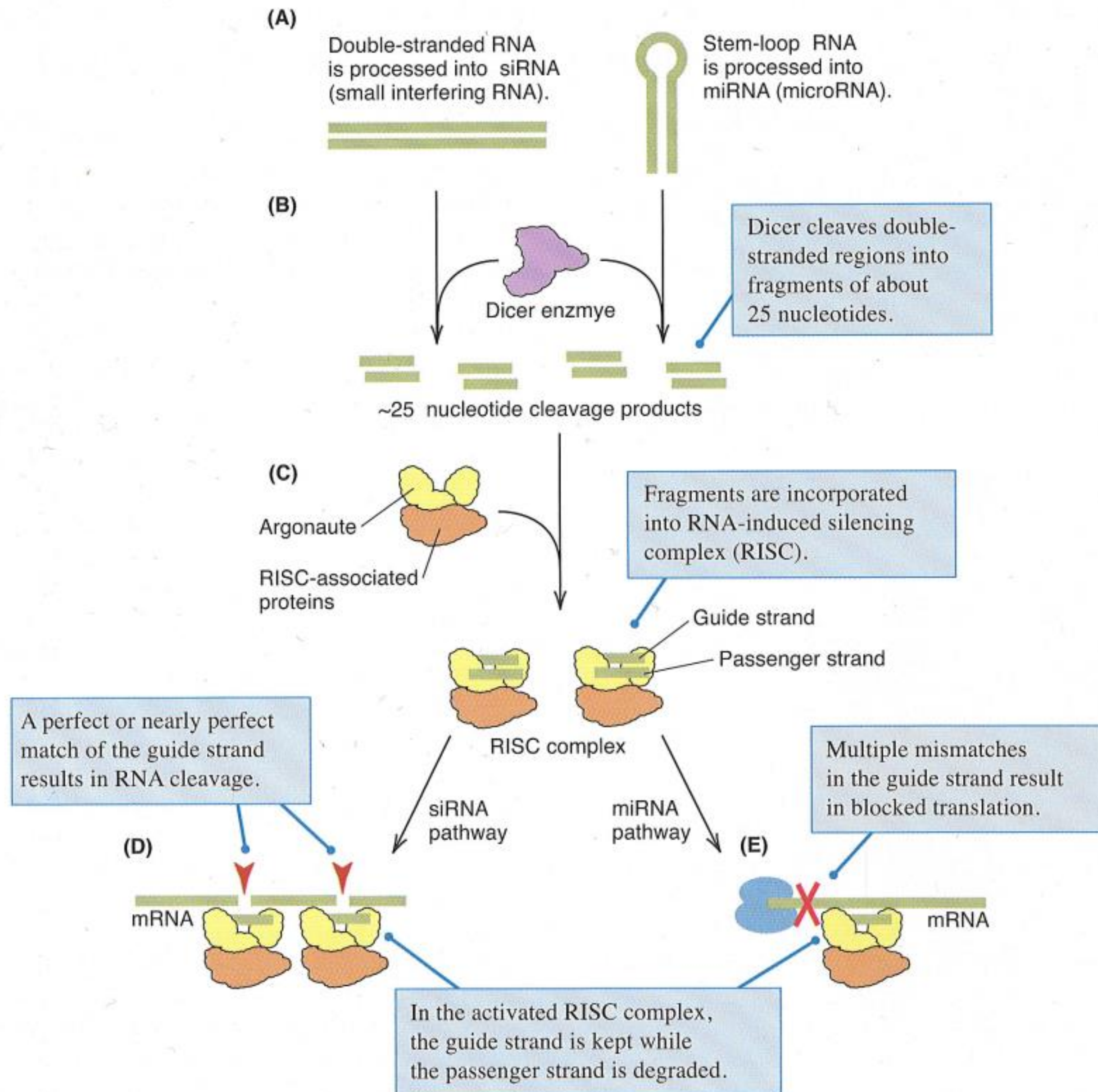


FIGURE 6.22 RNA interference using siRNA (left) and miRNA (right).

Molecular Biotechnology, Carolyn A. Dehlinger, Jones and Bartlett Learning

Proteome Analysis

2-D denaturing Gels

HPLC – MS Technology

Protein Modifications

Functional Analysis
Enzyme function
(Catalom)
Protein Interaction
(Interactome)

Protein Microarrays

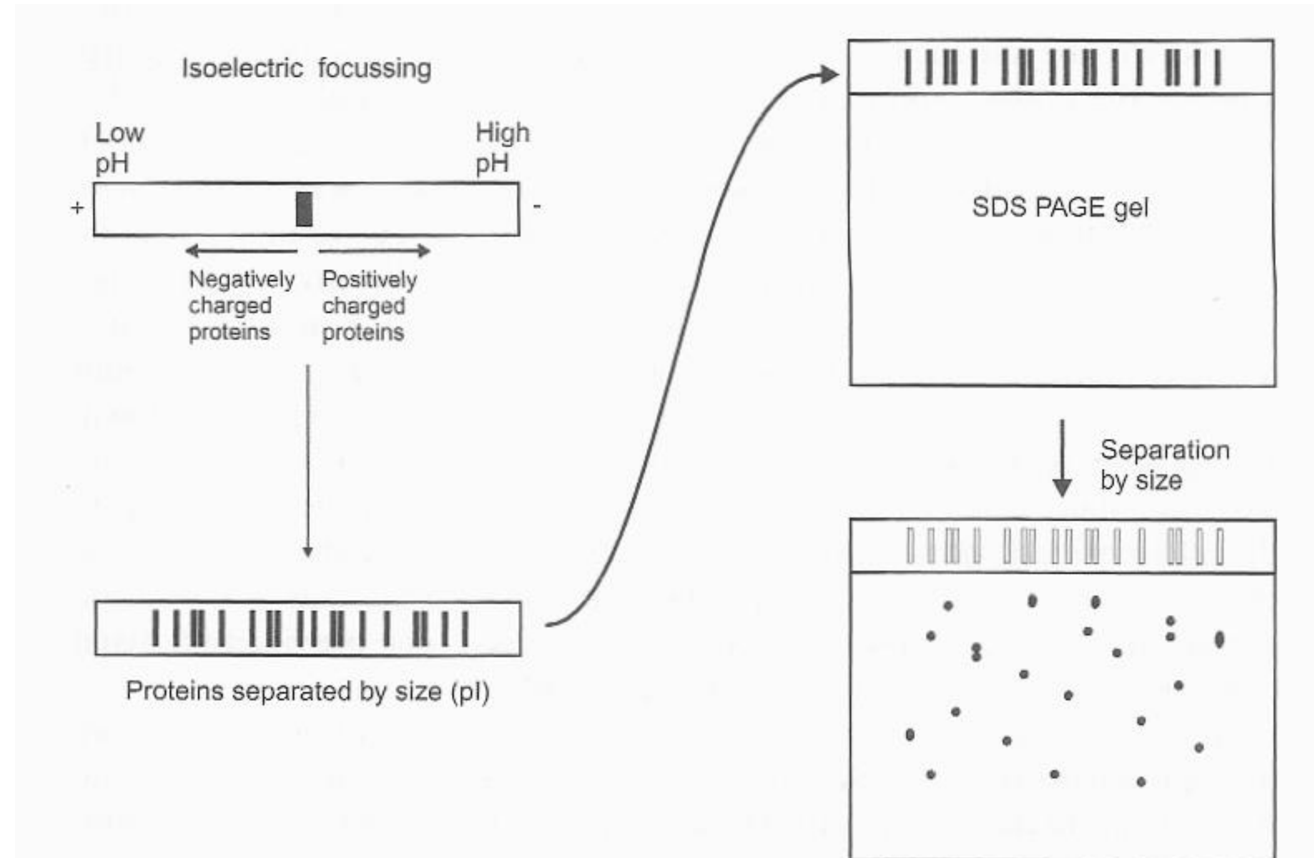
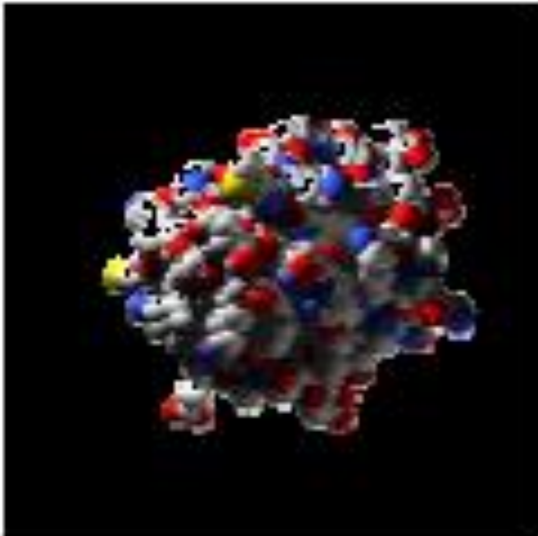


Figure 14.6 Two-dimensional gel electrophoresis

How to Do it?

Three Types of Simulation



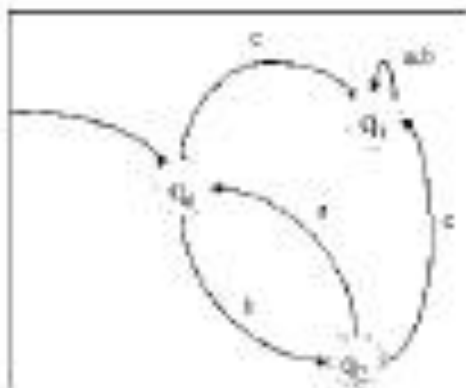
Atomic Scale
0.1 - 1.0 nm
Coordinate data
Dynamic data
0.1 - 10 ns
Molecular dynamics

Meso Scale
1.0 - 10 nm
Interaction data
Kon, Koff, Kd
10 ns - 10 ms
Mesodynamics

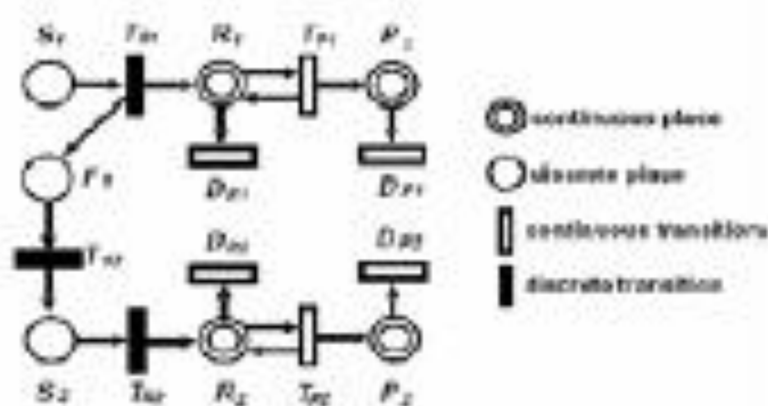
Continuum Model
10 - 100 nm
Concentrations
Diffusion rates
10 ms - 1000 s
Fluid dynamics

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

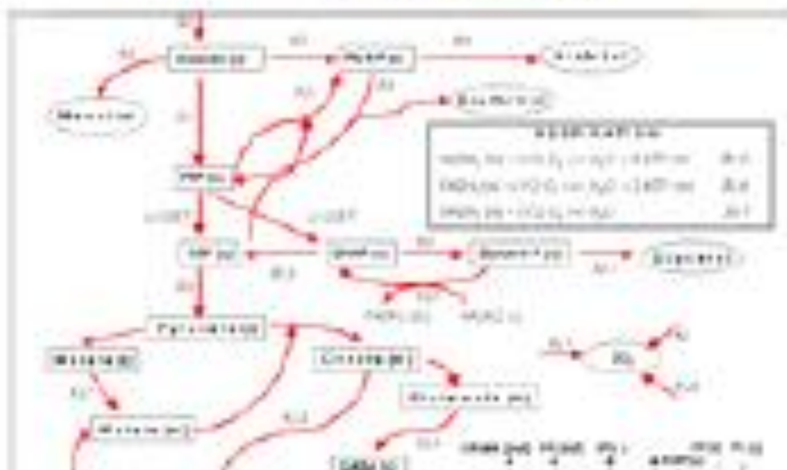
How To Do it? (Computationally)



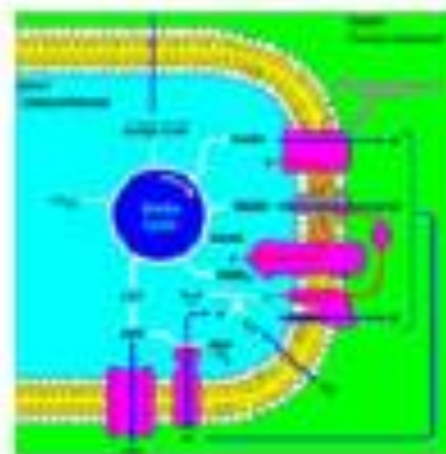
Pi Calculus



Petri Nets



Flux Balance Analysis



Differential Eqs

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Some Problems…

- **Almost all simulation systems are ultimately based on solving either:**

- ordinary differential equations (ODEs),
- partial differential equations (PDEs)
- or stochastic differential equations (SDEs)

- **Differential equations are "hard" to work with**

- when simulating spatial phenomena,
- when dealing with discrete events (binding, switching), non continuous variables (low copy number) or
- when key parameters are unknown or unknowable

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Some Problems…

- **DEs are notorious for instabilities or situations where small rounding errors lead to singularities or chaotic behavior**
- **DE methods are not conducive to visualization or interactive “movies”**
- **DE methods require considerable mathematical skill and understanding (not common among biologists)**
- **DE methods don’t easily capture stochasticity or noise (common in biology)**
- ***Issue of realism – cells don’t do calculus***

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Is There a Better Way?

- **Sidney Brenner calls it "biological arithmetic – not calculus"**

- **Needs to accommodate the discrete (binding, signaling) and continuous (substrate concentration) nature of many cellular phenomena**

- **Two new approaches which avoid DEs**

– Petri Nets (stochastic and hybrid)
– Cellular automata or agent based methods

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Petri Nets

- **A directed, bipartite graph in which nodes are either "places" (circles) or "transitions" (rectangles)**
- **A Petri net is marked by placing "tokens" on linked or connected places**
- **When all the places have a token, the transition "fires", removing a token from each input place and adding a token to each place pointed to by the transition (its output places)**
- **Petri nets are used to model concurrent systems, particularly network protocols w/o differential eqs.**
- **Hybrid petri nets allow modelling of continuous and discrete phenomena**

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

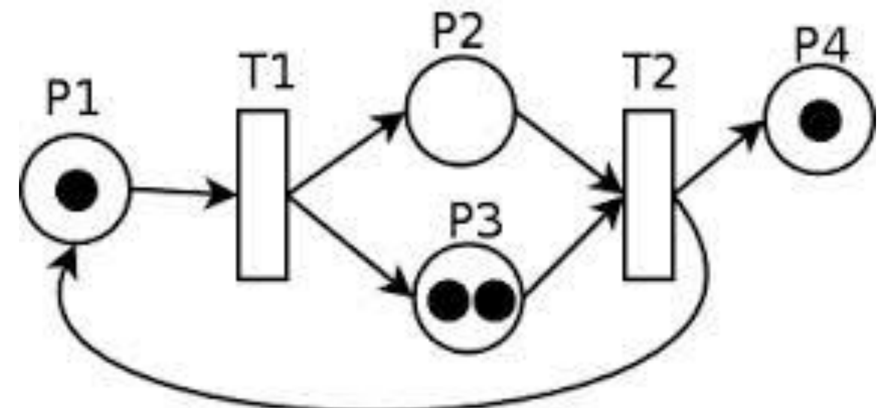
Petri Nets

Petri net (also known as a **place/transition net** or **P/T net**) is one of several mathematical modeling languages for the description of distributed systems.

A Petri net consists of places **(P)**, transitions **(T)**, and arcs.
Arcs run from a place to a transition or vice versa, never between places or between transitions

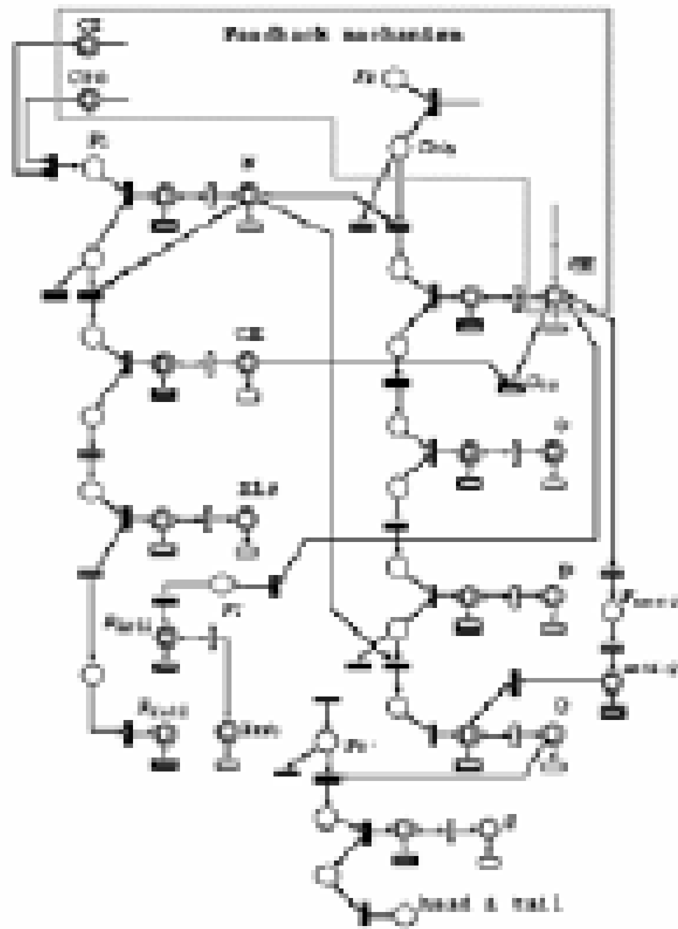
→ Petri nets offer a graphical notation for stepwise processes that include choice, iteration, and concurrent execution.
→ Petri nets have an exact mathematical definition of their execution semantics, with a well-developed mathematical theory for process analysis.

Graphically, places in a Petri net may contain a discrete number of marks called tokens. Any distribution of tokens over the places will represent a configuration of the net called a marking

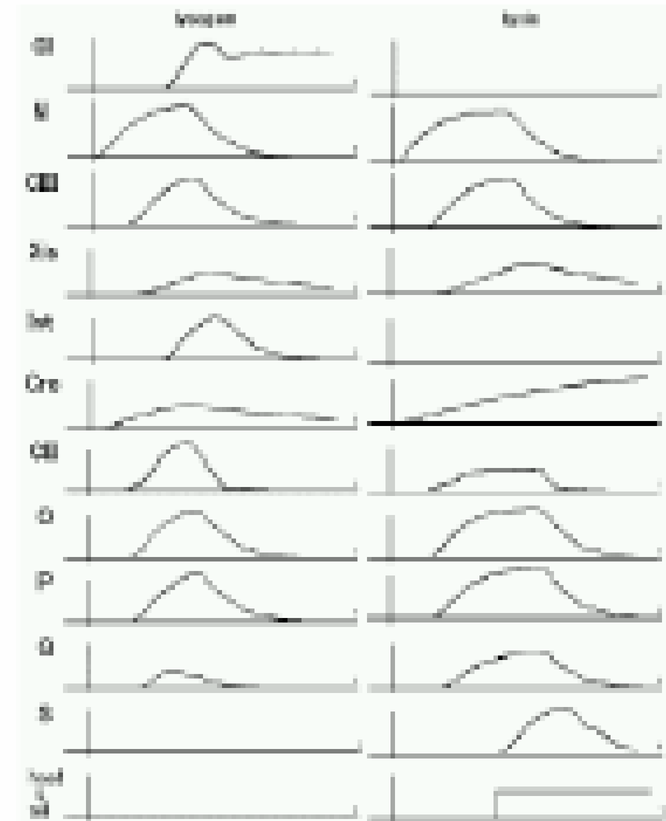


Detailed_petri_net.pngen.wikipedia.org

Petri Nets



λ phage control circuit



Predicted protein expression

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Petri Nets – Limitations

- Not designed to handle spatial events or spatial processes easily

- Stochasticity is "imposed", it does not arise from underlying rules or interactions

- Does not reproduce physical events (brownian motion, collisions, transport, binding, etc.) that might be seen in a cell

– Petri Nets are more like a plumbing and valving control system

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

TU Graz

Cellular Automata

- **Computer modelling method that uses lattices and discrete state "rules" to model time dependent processes**

– a way to animate things

- **No differential equations to solve, easy to calculate, more phenomenological**

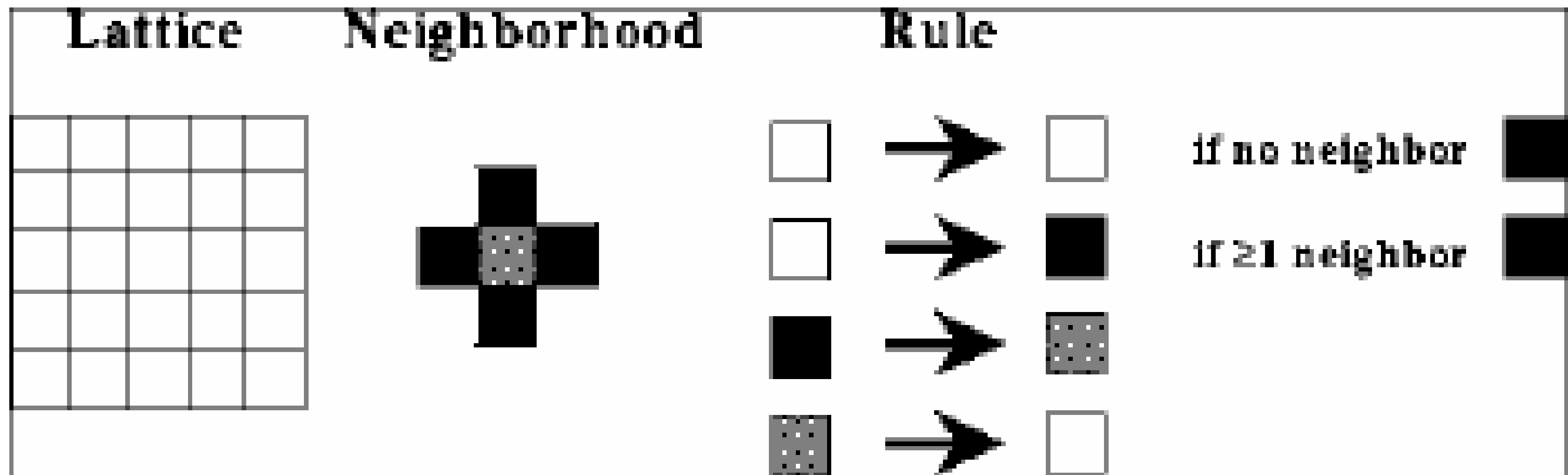
- **Simple unit behavior -> complex group behavior**

- **Used to model fluid flow, percolation, reaction + diffusion, traffic flow, pheromone tracking, predator-prey models, ecology, social nets**

- ***Scales from 10^{-12} to 10^{+12}***

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Cellular Automata



Can be extended to 3D lattice

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Dynamic Cellular Automata

• A novel method to apply Brownian motion to objects in the Cellular Automata lattice (mimics collisions)

• Takes advantage of the scale-free nature of Brownian motion and the scale-free nature of heterogeneous mixtures to

- allow simulations to span many orders of
  time (nanosec to hours) and
  space (nanometers to meters)

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Sim Cell

- **Java application that uses Dynamic Cellular Automata (DCA) to model motions, interactions, transport and transformations at the meso-scale (10^{-8} to 10^{-6} m)**

- **Uses a square, 2D lattice to model processes, lattice squares are equivalent to 3x3 nm regions**

- **Molecular objects are moved randomly and interactions determined according to a set of interaction rules that are only applied when objects are in contact (collision detection)**

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Sim Cell Interactions

- **Five different types of molecules or objects allowed in SimCell:**

1) small molecules,
2) soluble proteins,
3) membrane proteins,
4) DNA molecules, and
5) membranes

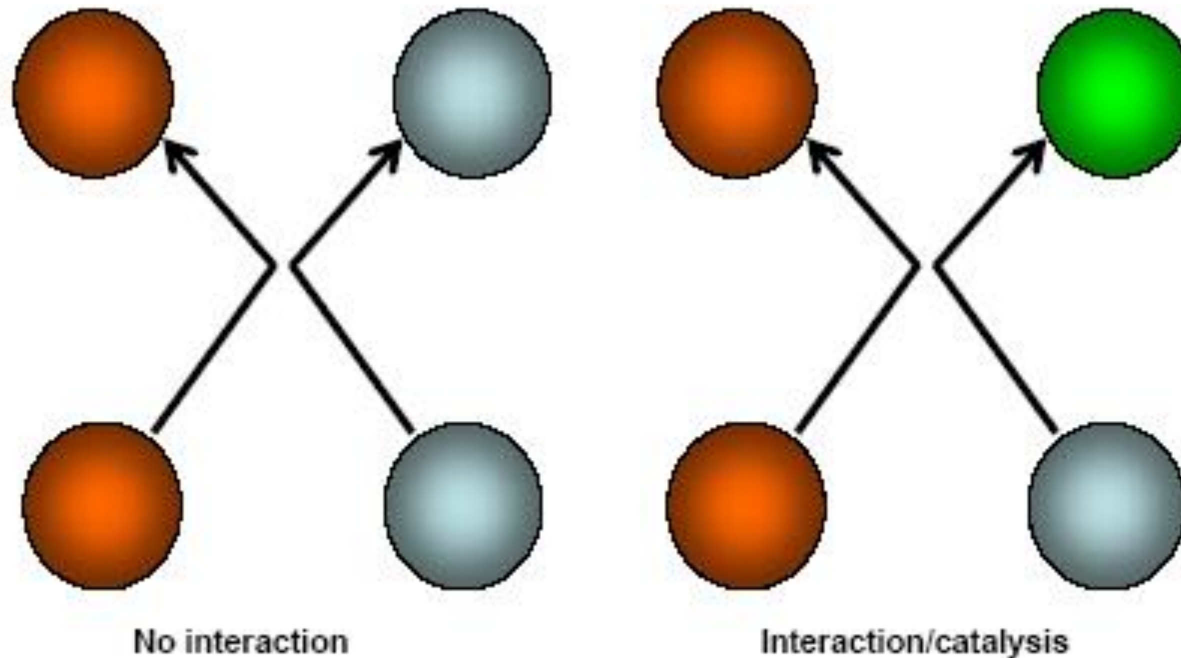
- **Protein-ligand interactions reduced to relatively small number of possibilities**

– Touch and Go (T&G)
– Bind and Stick (B&S)
– Transport (TRA)

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

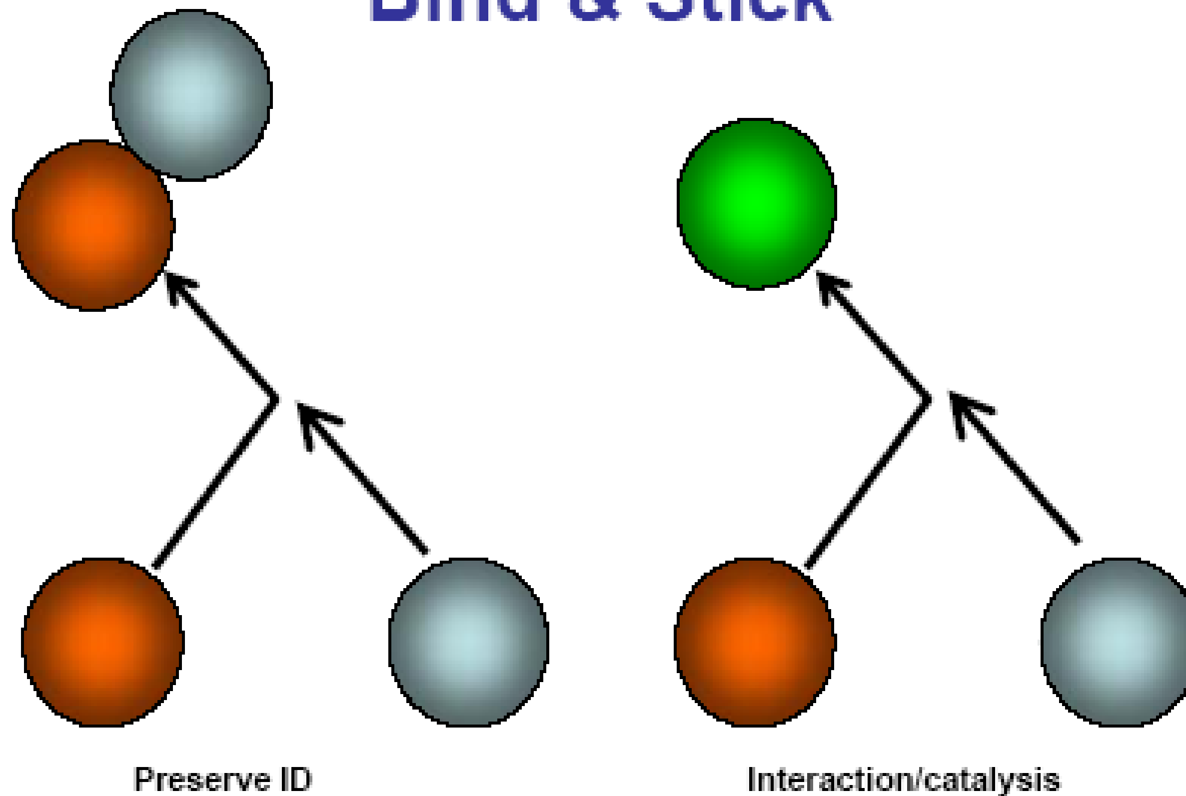
Sim Cell Interactions

Touch & Go



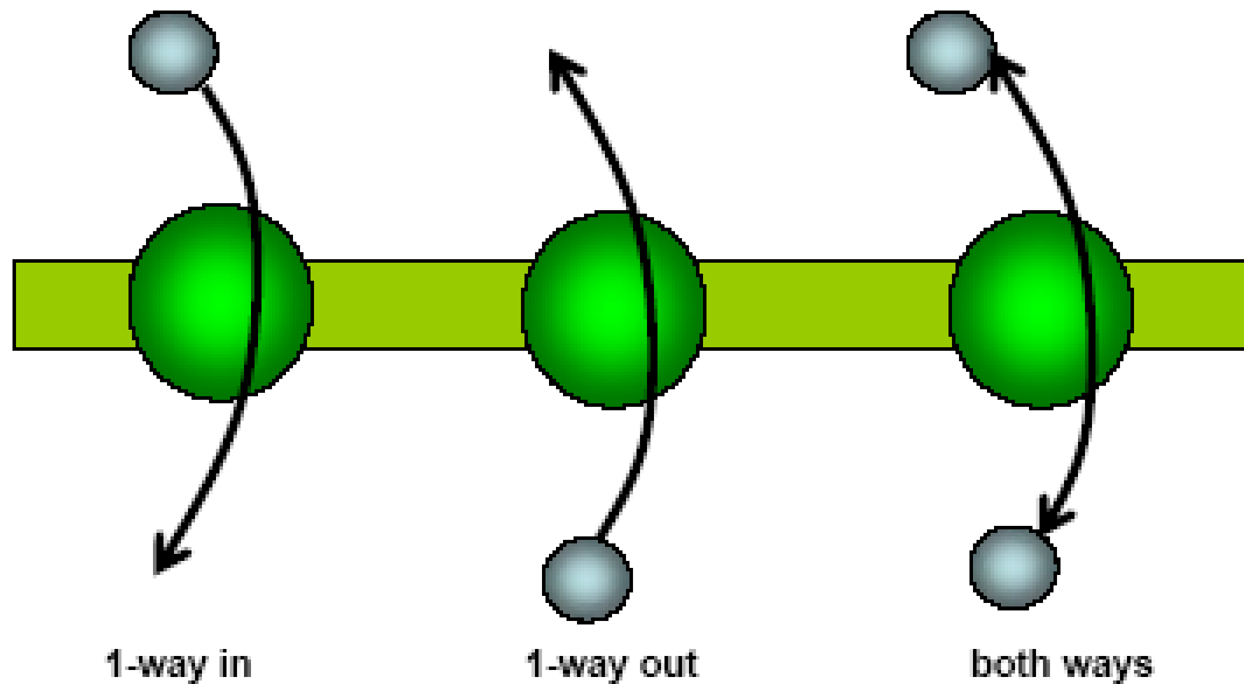
http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Sim Cell Interactions



http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Sim Cell Interactions



http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Sim Cell and Cell Simulation

- **Ideal for model checking and validation**

- **Conceptually equivalent to spatially dependent stochastic Petri nets**

- **Universally applicable: Enzyme kinetics, diffusion, excluded volume, binding, vesicle fusion, osmotic lysis, osmotic pressure, genetic circuits, metabolism, transport, repression, signalling, cell division, embryo gene expression…**

All from one tool!

http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Systems Biology

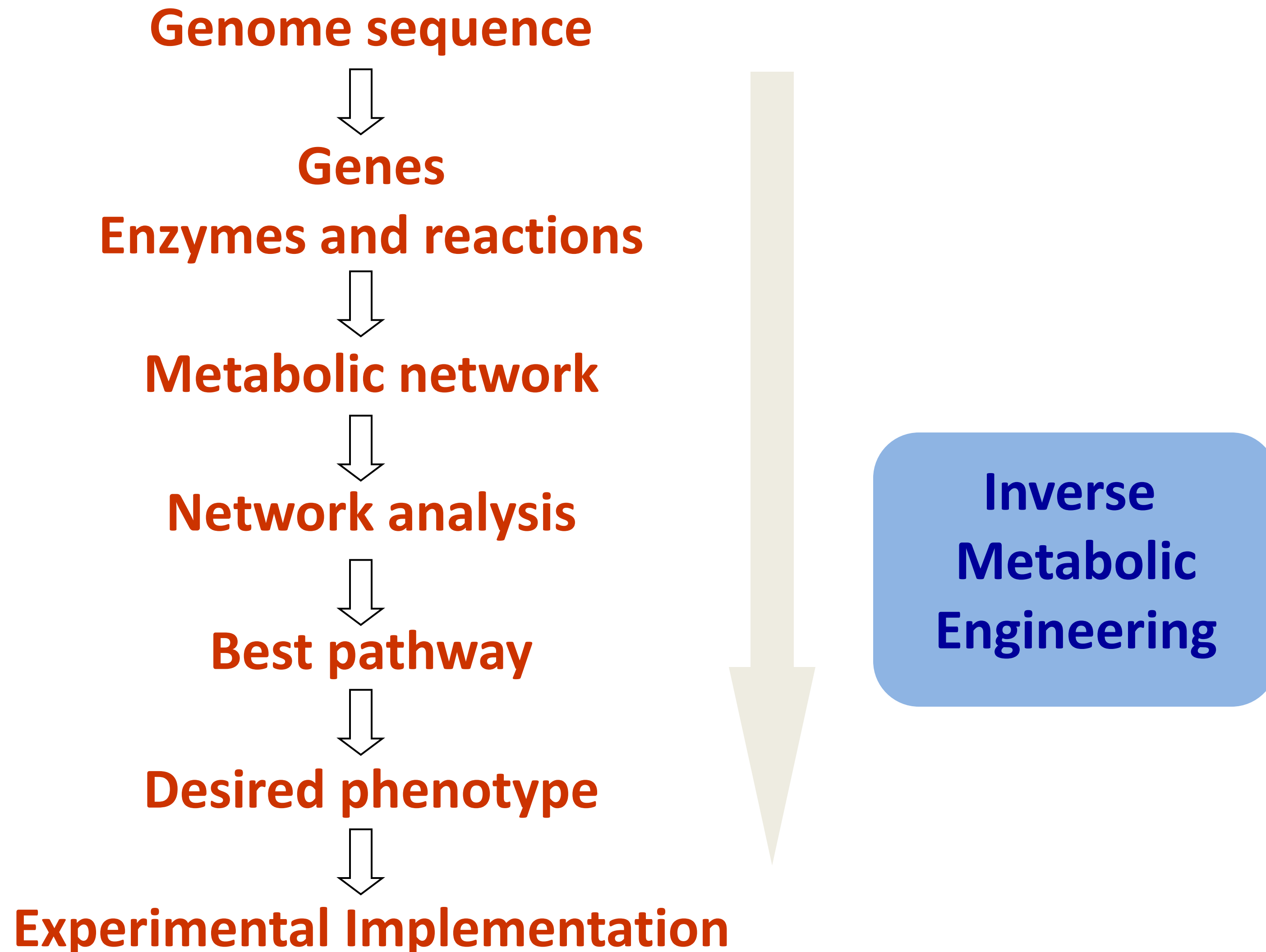
- **Systems Biology requires the integration of**
  - **data archiving,**
  - **experimentation and**
  - **novel computational approaches**

- **There is a clear need for bioinformatics to step up from the static "stamp collection" phase to thinking about systems in dynamic/interactive/integrated terms**

- **New tools are needed to make this possible**
  - **– consider DCA & Petri Nets**

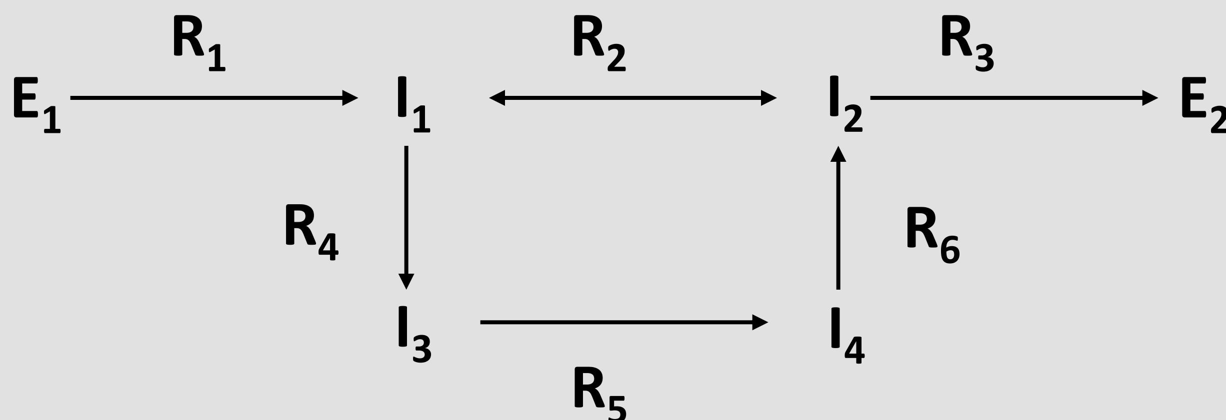
http://redpoll.pharmacy.ualberta.ca/bioinfo301/Systemsbiology_6.pdf

Can we 'compute' a cell?



F.Srienc, University of Minnesota

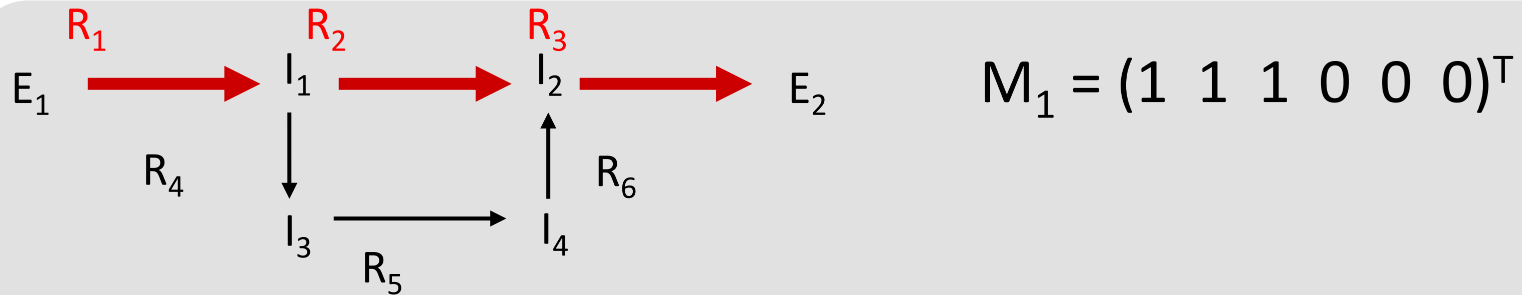
Elementary Modes: Example

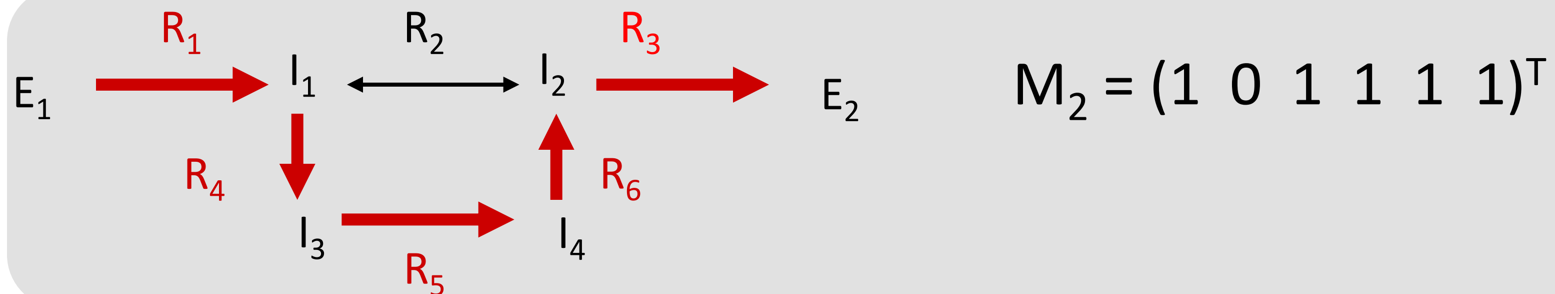


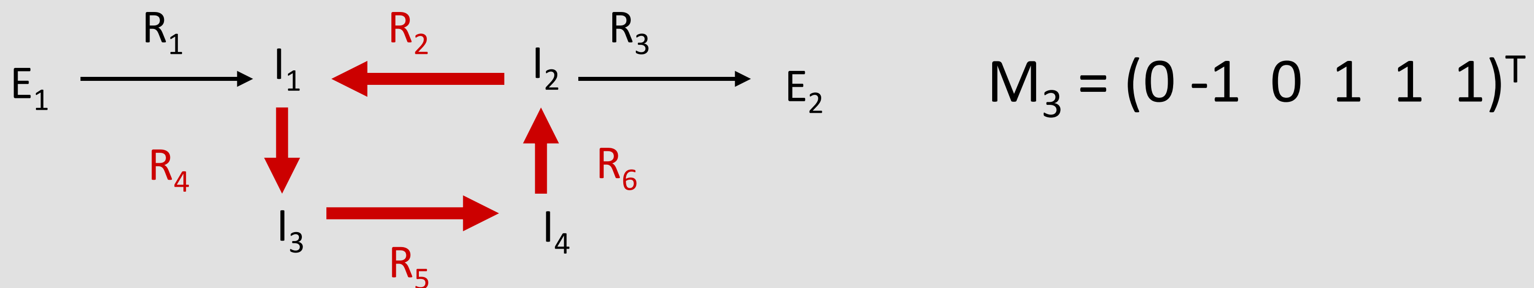
Schuster, S., Hilgetag, C., Woods, J.H. & Fell, D.(1996)

F.Srienc, University of Minnesota

Elementary Modes: Example







F.Srienc, University of Minnesota

Systems Biology

Data aquisition

Single Unit Data ← → Comprising Data Sets

Single Pathways
← → Integrated Networks
Single Regulatory Units

Static Systems ← → Dynamic Systems