

Lecture 1. BIOINFORMATICS

fatchiyah, dept bio UB 2/21/2012

Fatchiyah, M.Kes.,Ph.D

Asc. Prof of Molecular Genetics

Dept of Biology, Brawijaya University

Email: fatchiya@ub.ac.id

Website: <http://fatchiyah.lecture.ub.ac.id>

What is Bioinformatics

- The use of computers to collect, analyze, and interpret biological information at the molecular level.

"The mathematical, statistical and computing methods that aim to solve biological problems using DNA and amino acid sequences and related information."

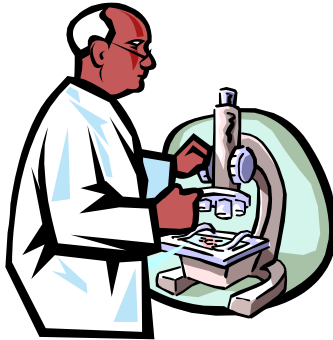
- A set of software tools for molecular sequence analysis

Analysis development

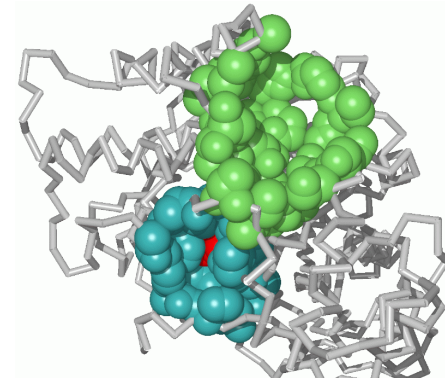
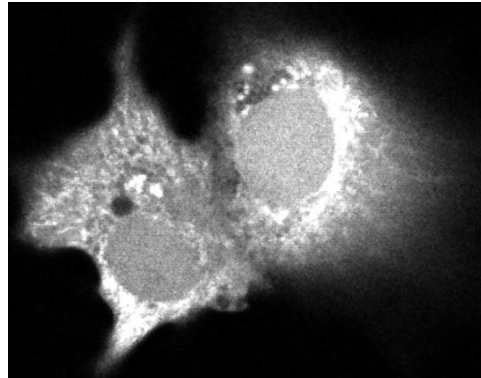
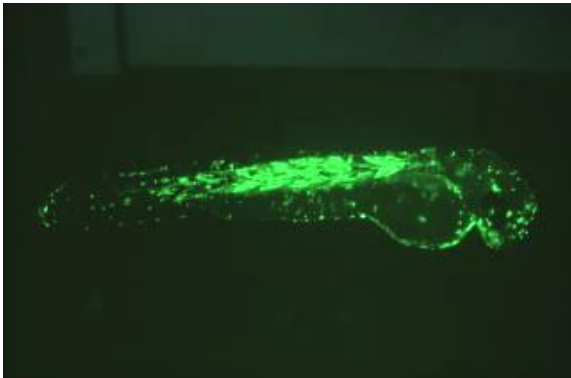
In Vivo



In Vitro



In Silico



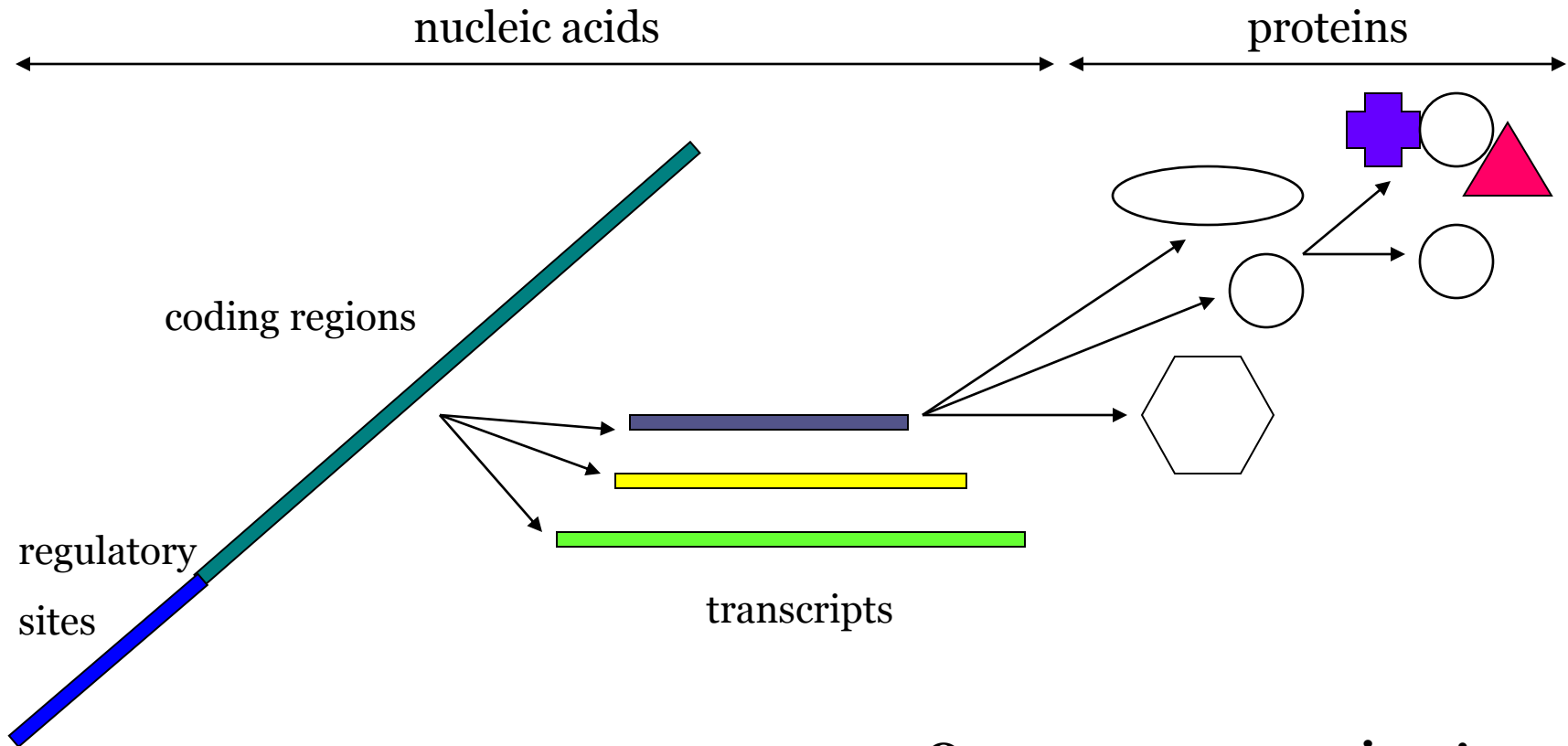
What is bioinformatics?

- **an emerging interdisciplinary research area**
- **deals with the computational management and analysis of biological information: genes, genomes, proteins, cells, ecological systems, medical information, robots, artificial intelligence...**

Basic concepts

- **conceptual foundations of bioinformatics:**
evolution
protein folding
protein function
- **bioinformatics builds mathematical models of these processes -
to infer relationships between components of complex biological systems**

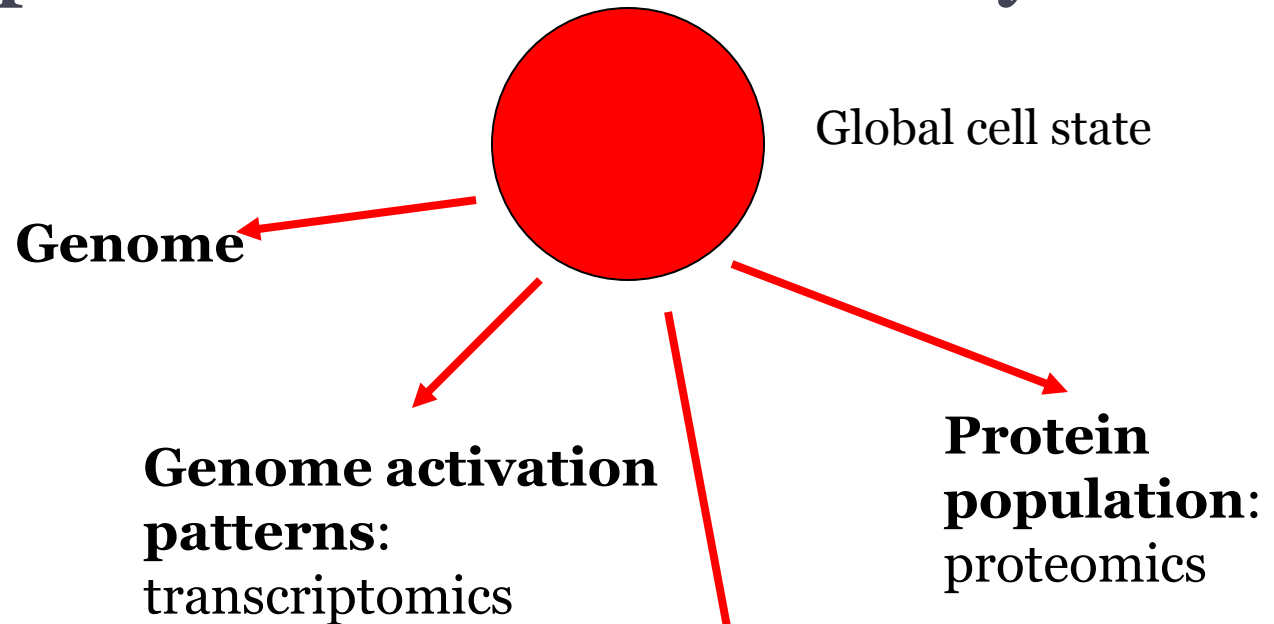
Information processing in cells



One-to-many mappings!

Context-dependence!

Global approaches: Toward a new Systems Biology



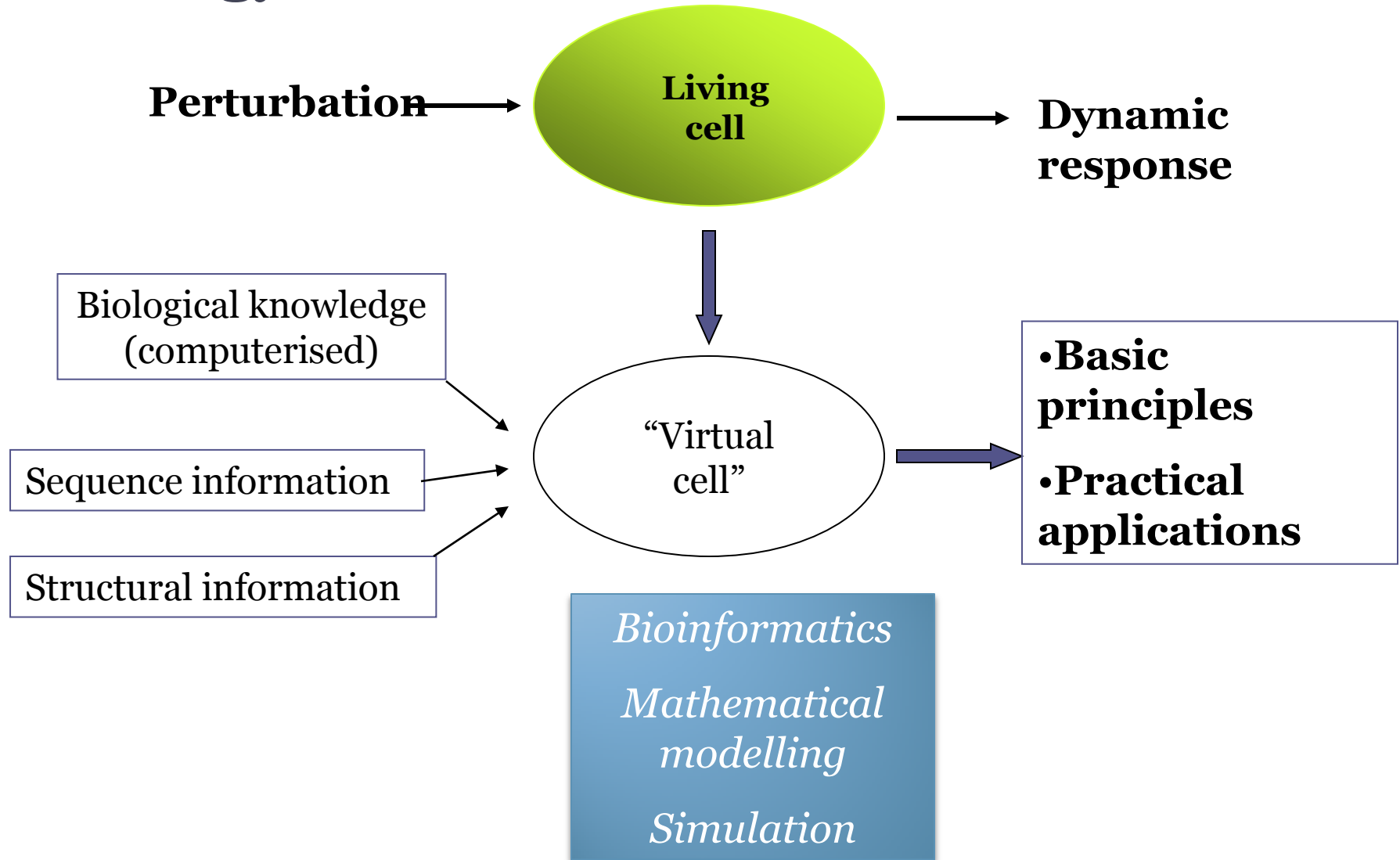
•How does the spatial and temporal organisation of living matter give rise to

Organisation:

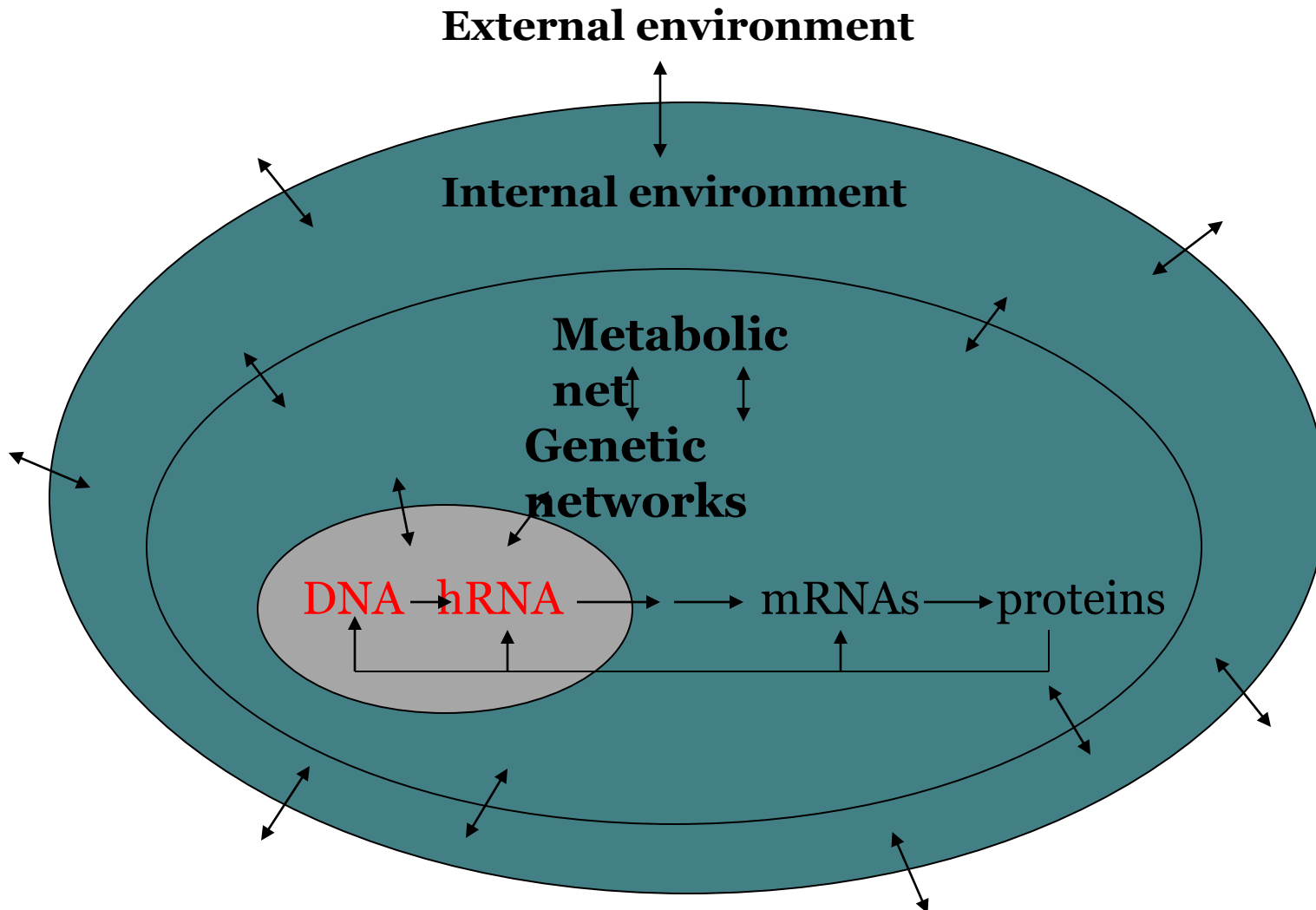
tissue imaging ↔ EM ↔ X-ray, NMR
cells
molecular complexes

Global approaches: Toward a new Systems Biology

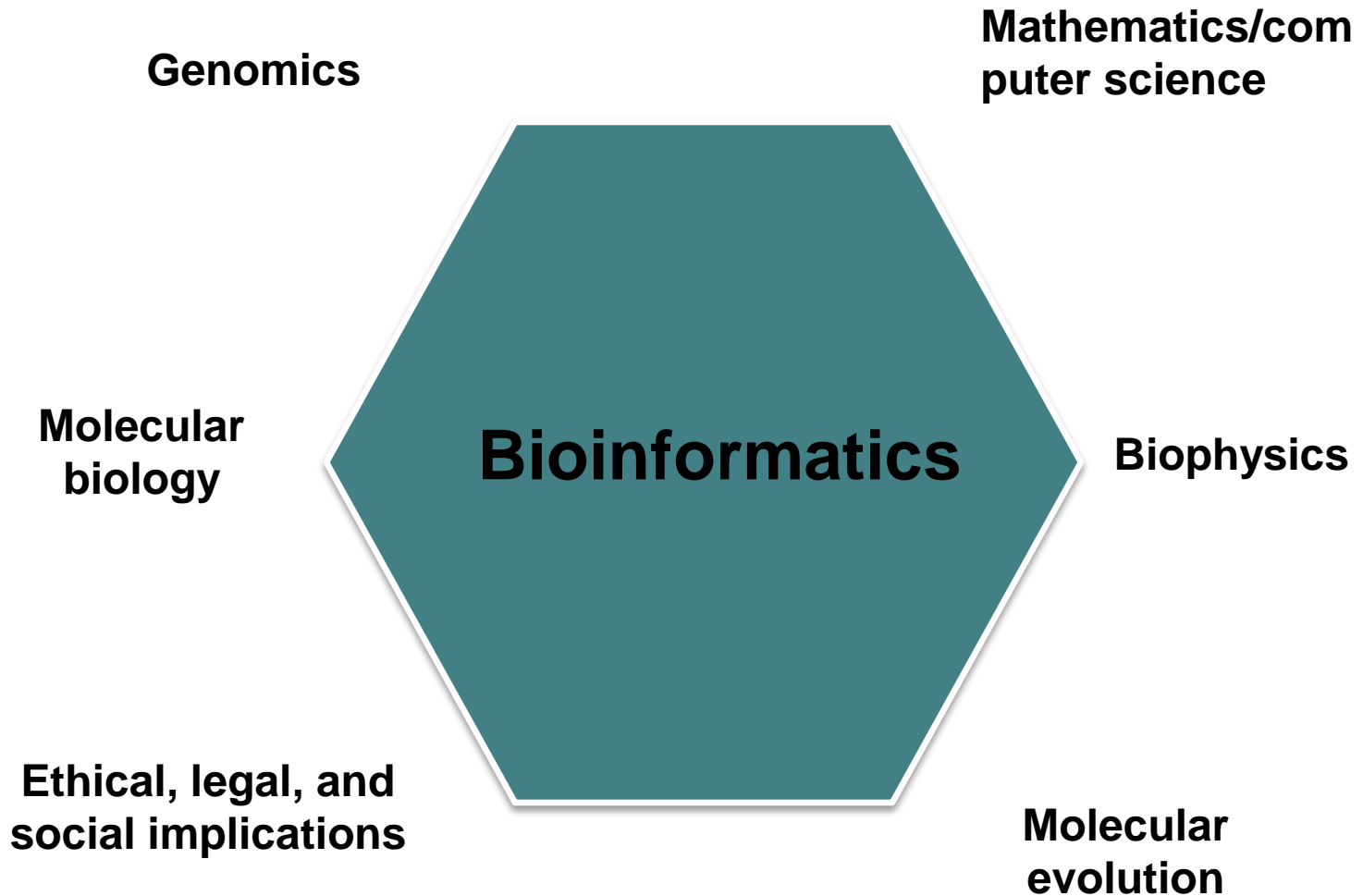
fatchiyah, dept bio UB 2/21/2012



To explore the pathway networking inside and/or among cells or tissues to communicate in between.



Bioinformatics in context

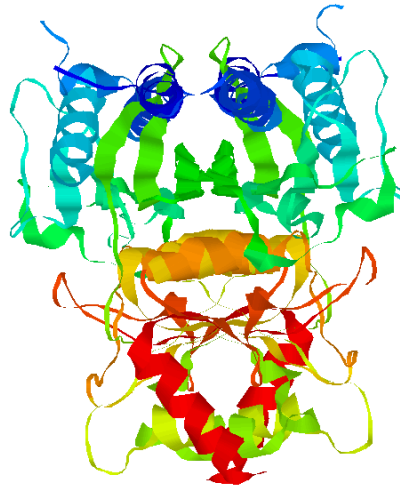


The Core of Bioinformatics to date

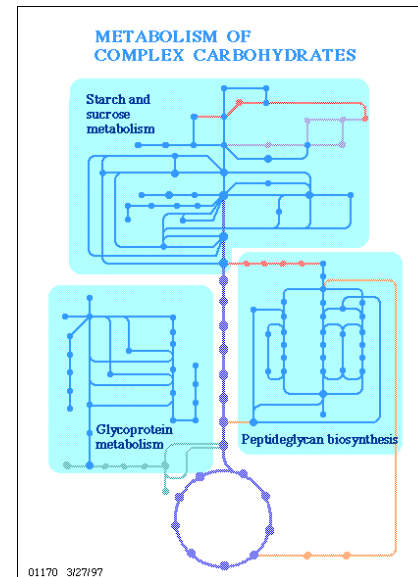
- Relationships between

```
TDQAAFDTNIVTLTRFVMEQGR  
KARGTGEMTQLLSLCTAVKAI  
STAVRKAGIAHLYGIAGSTNVT  
GDQVKKLDVLSNDLVINVLKSS  
FATCVLVTEEDKNAIIVEPEKR  
GKYVVCFDPLDGSSNIDCLVSI  
GTIFGIYRKNSTDEPSEKDALQ  
PGRNLVAAGYALYGSATMLV
```

sequence



3D structure



protein functions

- Properties and evolution of genes, genomes, proteins, metabolic pathways in cells

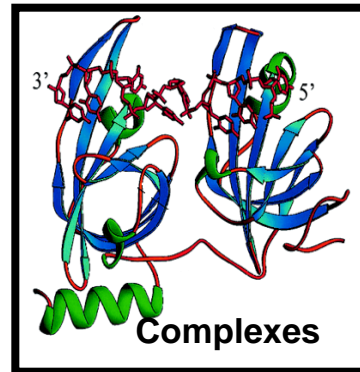
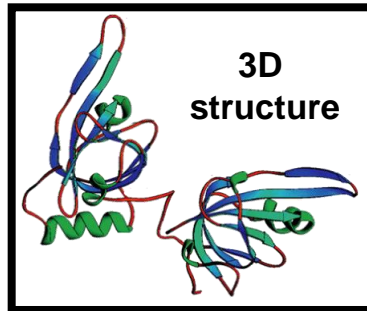
- Use of this knowledge for prediction, modelling, and design

BIOLOGICAL STRUCTURE

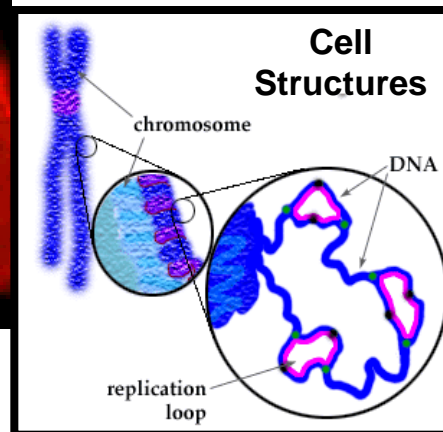
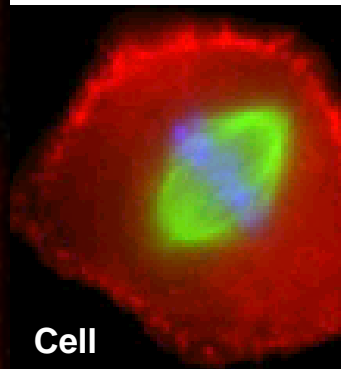
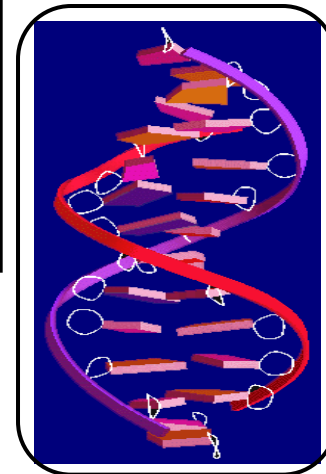
faculty and staff UB 2/21/2012

Sequence

MESDAMESETMESSRSMYN
AMEISWALTERYALLKINCAL
LMEWALLYIPREFERDREVIL
MYSELFIMACENTERDIRATV
ANDYINTENNESSEEILIKENM
RANDDYNAMICSRPADNAPRI
MASERADCALCYCLINNDRKI
NASEMRPCALTRACTINKAR
KICIPCDPKIQDENVSDETVS
WILLWINITALL



Structural Scales

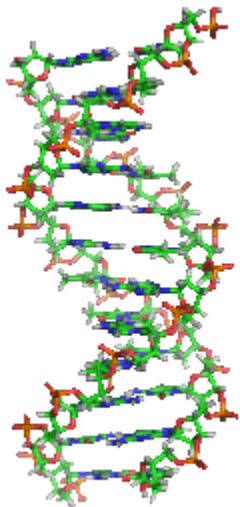


System Dynamics

Blue print of gene

- Genome sequence: for the first time there is a blueprint of the activity of a cell
- Gene expression, in the form of cDNA array, and proteomic studies: how these genes interact, interfacing with each other, and how they form networks.
- On structural level, the mechanism how these molecules work.
- Major impact on diagnosis, treatment, drug discovery, regulation and metabolism, biodegradation

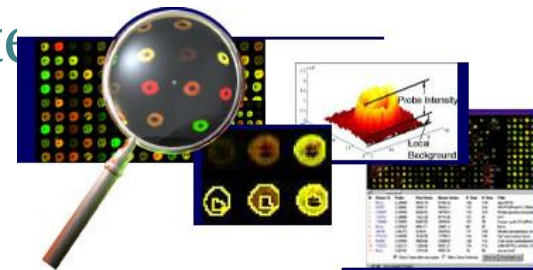
Challenges in Computational Biology



1. Obtain the genome of an organism.
2. Identify and annotate genes.
3. Find the sequences, three dimensional structures, and functions of proteins.
4. Find sequences of proteins that have desired three dimensional structures.
5. Compare DNA sequences and proteins sequences for similarity.
6. Study the evolution of sequences and species.

Bioinformatics

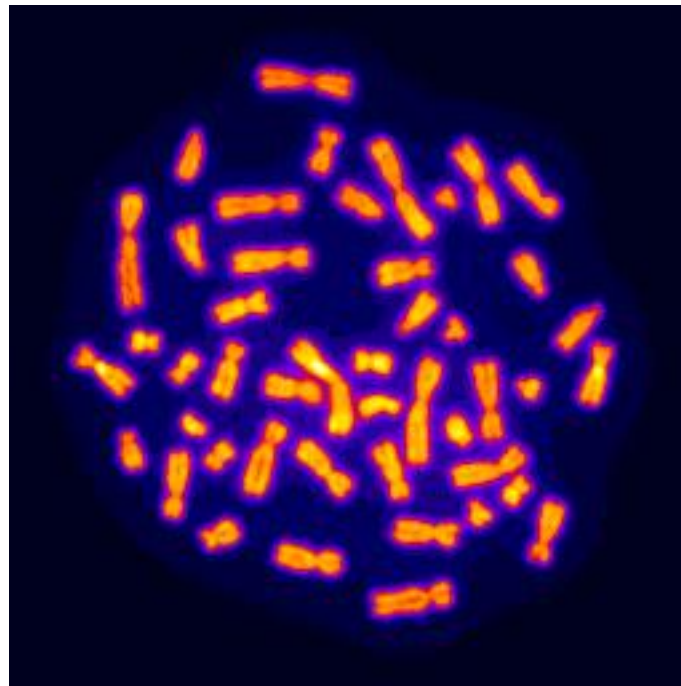
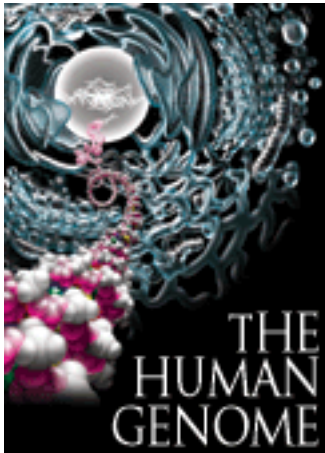
- Computational analysis of high-throughput biological data
 - Whole genome sequencing.
 - Global genomic expression & profiling.
 - Functional genomics.
 - Structural genomics/proteomics.
 - Comparative genomics.



I. The Human Genome Project

The genome sequence is complete - almost!

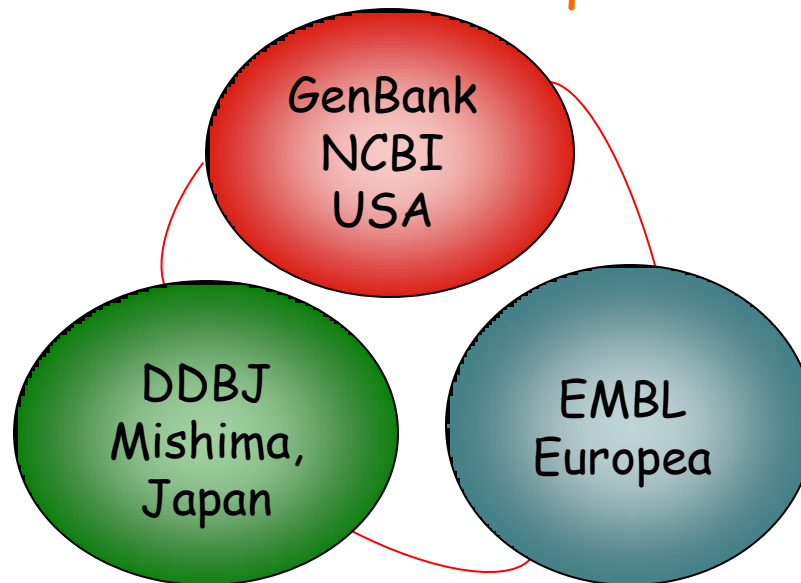
- approximately 3.2 billion base pairs.



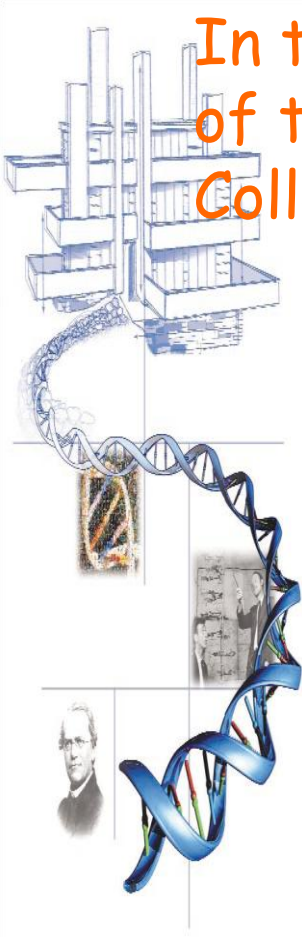
Internationally Networking Collaboration



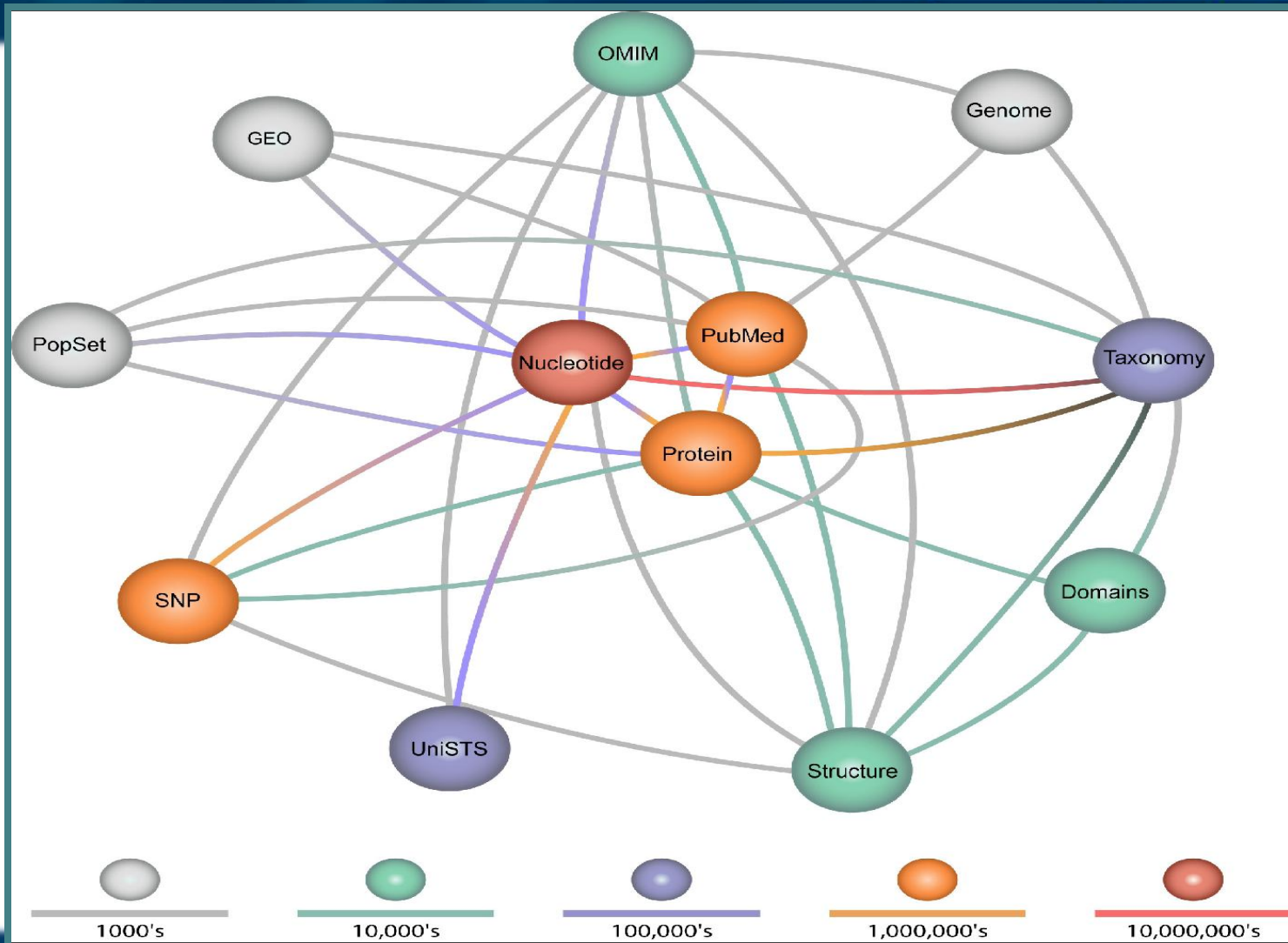
In the mid-1990s, the GenBank database became part of the International Nucleotide Sequence Database Collaboration:

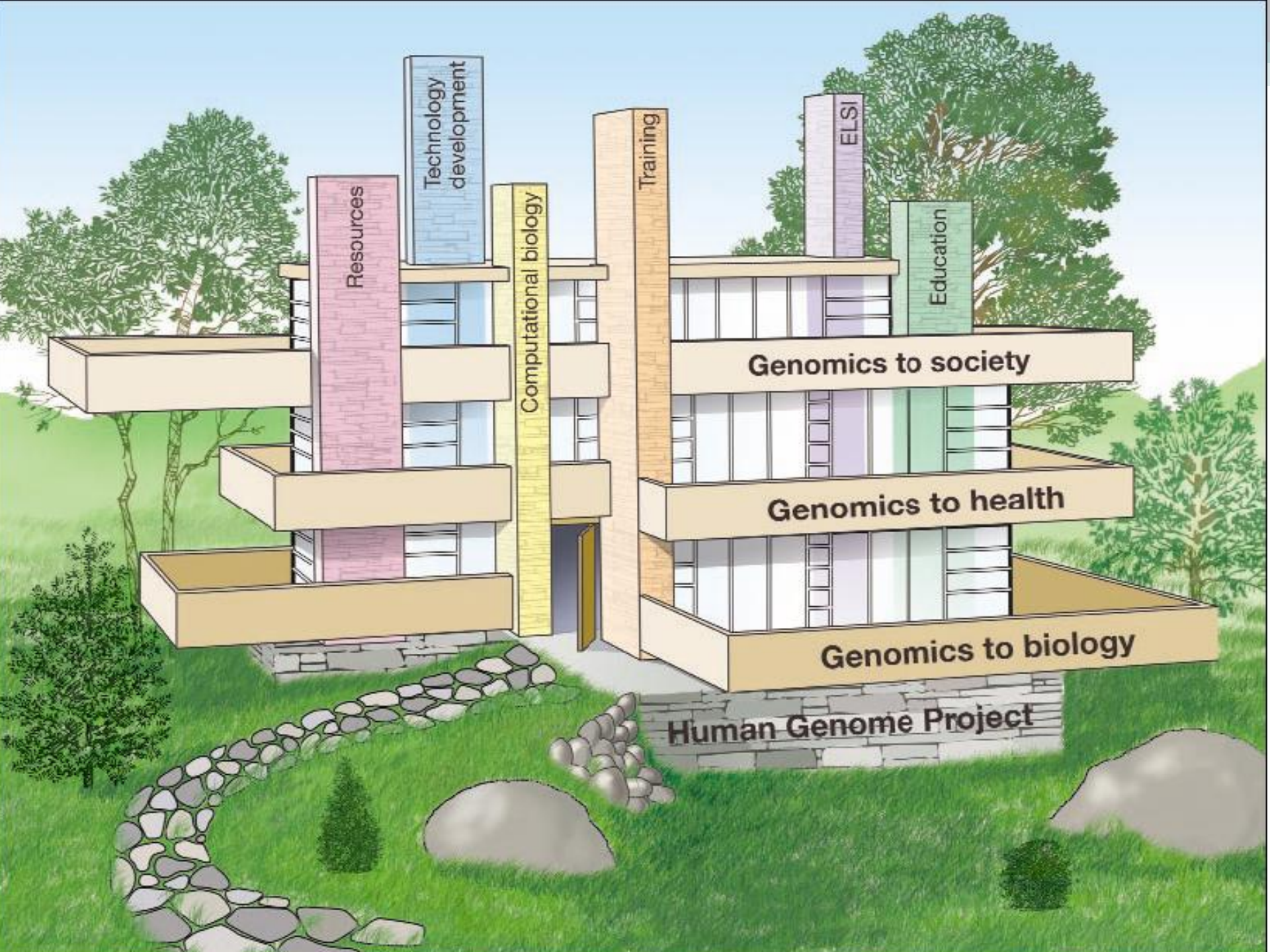


NCBI investigators maintain on going collaborations with several institutes within NIH and also with numerous academic and government research laboratories



The original version of Entrez had just 3 nodes: nucleotides, proteins, and PubMed abstracts.





Resources

Technology
development

Computational biology

Training

ELSI

Education

Genomics to society

Genomics to health

Genomics to biology

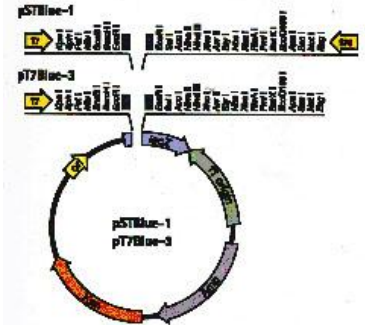
Human Genome Project

Free unrestricted access for all



The Wellcome Trust

The door to discovery is wide open



Genome browsers

Ensembl

www.ensembl.org

University of California Santa Cruz

<http://genome.cse.ucsc.edu>

MGD the Jackson Laboratory

www.informatics.jax.org

Genome Databases

European Bioinformatics Institutes

www.ebi.ac.uk

GenBank

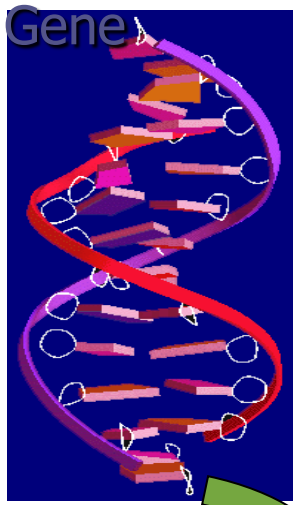
www.ncbi.nlm.nih.gov

DNA Data Bank of Japan

www.ddbj.nig.ac.jp



The Flow of Biotechnology Information



> DNA sequence

AATTCATGAAAATCGTATACTGGTCTGGTACCGGCAACAC
 TGAGAAAATGGCAGAGCTCATCGCTAAAGGTATCATCGAA
 TCTGGTAAAGACGTCAACACCATCAACGTGTCTGACGTTA
 ACATCGATGAACTGCTGAACGAAGATATCCTGATCCTGGG
 TTGCTCTGCCATGGGCGATGAAGTTCTCGAGGAAAGCGAA
 TTTGAACCGTTCATCGAAGAGATCTCTACCAAATCTCTG
 GTAAGAAGGTTGCGCTGTTTCGGTTCTTACGGTTGGGGCGA
 CGGTAAGTGGATGCGTGACTTCGAAGAACGTATGAACGGC
 TACGGTTGCGTTGTTGTTGAGACCCCGCTGATCGTTCAGA
 ACGAGCCGGACGAAGCTGAGCAGGACTGCATCGAATTTGG
 TAAGAAGATCGCGAACATCTAGTAGA

**> 500, 000 genes
sequenced to date**



> Protein sequence

MKIVYWSGTGNTEKMAELIAKGI
 IESGKDVNTINVSDVNI
 DELLNEDILILGCSAMGDEVLEESE
 FEPFIEEISTKISGK
 KVALFGSYGWGDGKWMRDFEERM
 NGYGCVVVETPLIVQNE
 PDEAEQDCIEFGKKIANI

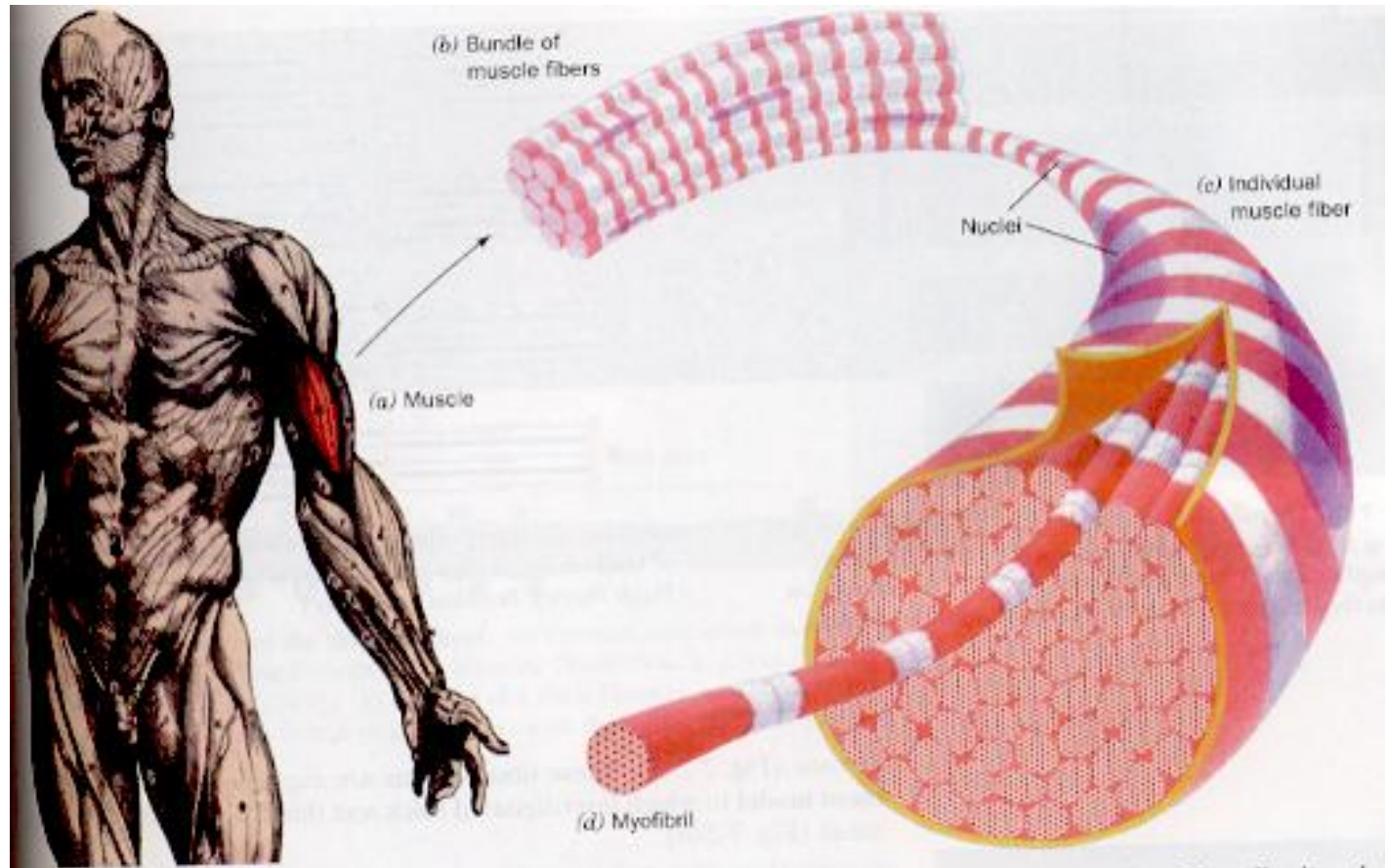
**Expected number of
unique protein
structures:**

~ 700-1, 000

Proteins: Molecular Machines

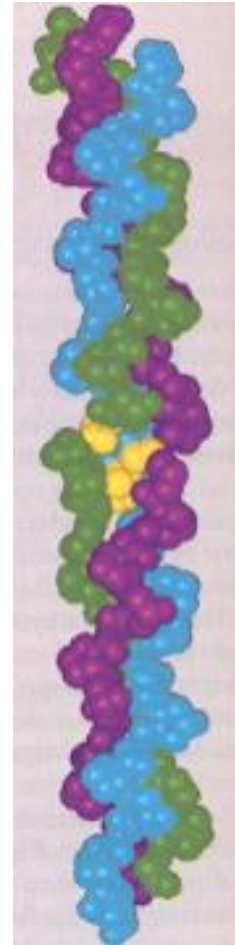
- Proteins in your muscles allows you to move:

myosin
and
actin



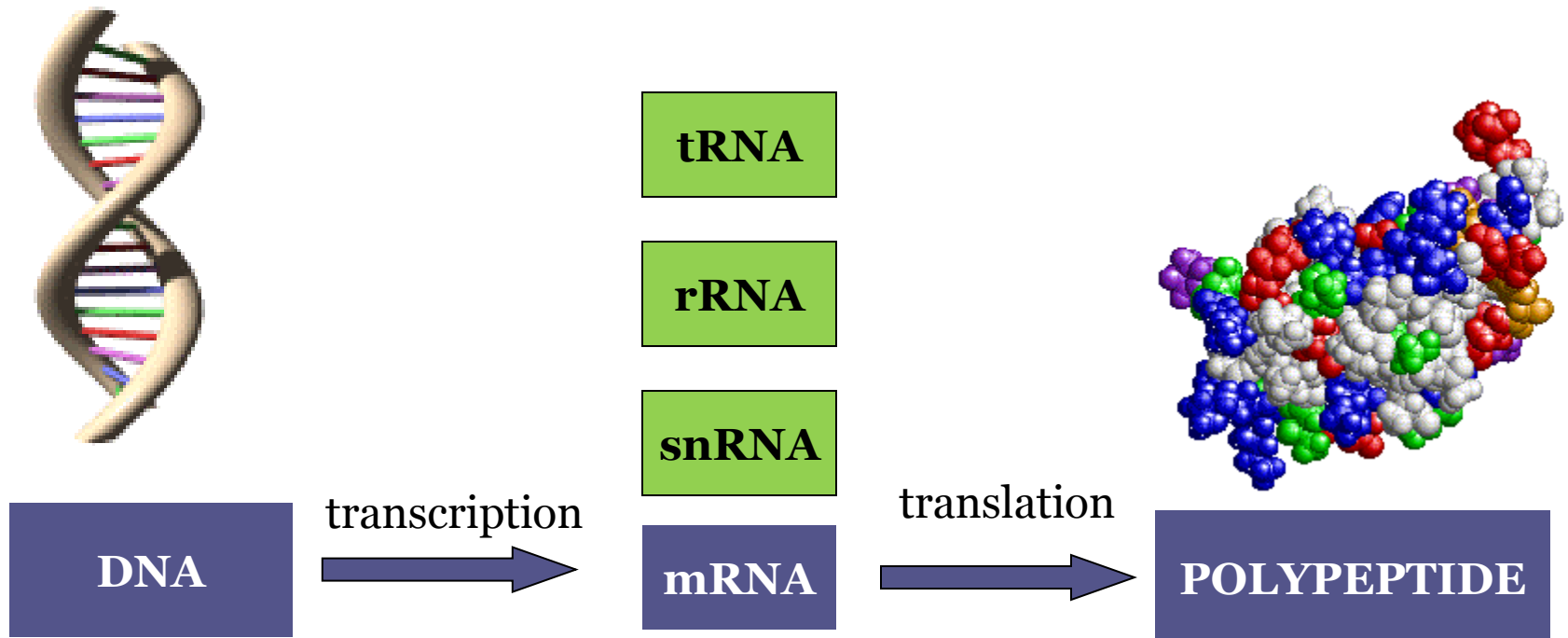
Proteins: Molecular Machines

- Enzymes
(digestion, catalysis)
- Structure (collagen)



Central Dogma of Molecular Biology

- DNA → RNA → Protein → Phenotype
- Transcription : DNA → RNA
- Translation : RNA → Protein



Transcription - key steps

- Initiation
- Elongation
- Termination

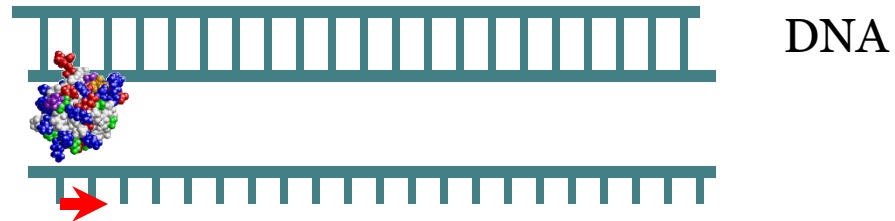


+



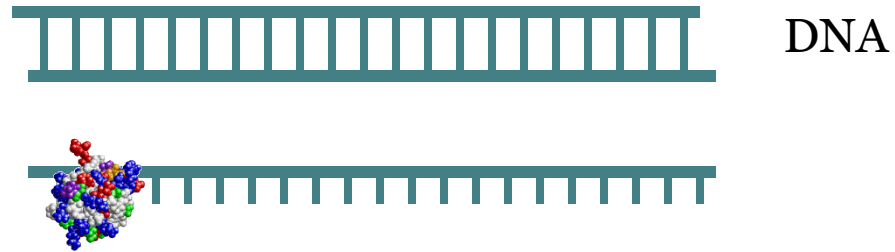
Transcription - key steps

- Initiation
- Elongation
- Termination



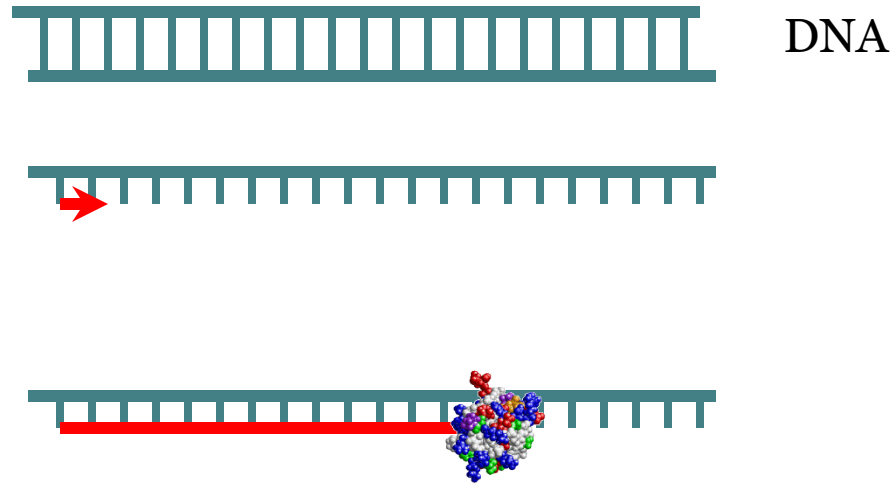
Transcription - key steps

- Initiation
- Elongation
- Termination



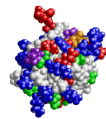
Transcription - key steps

- Initiation
- Elongation
- Termination



Transcription - key steps

- Initiation
- Elongation
- Termination



+



RNA

Promoters

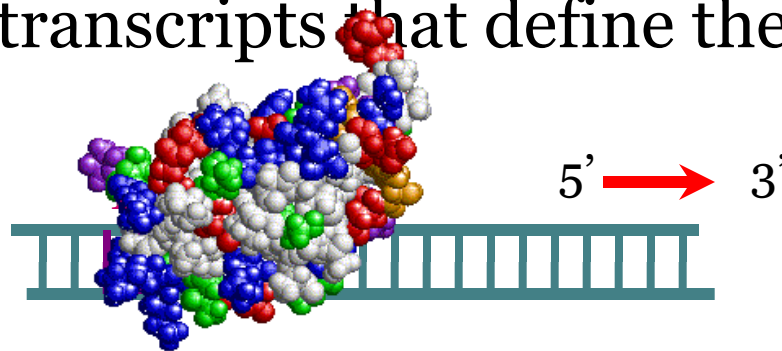
- Promoters are sequences in the DNA just upstream of transcripts that define the sites of initiation.



- The role of the promoter is to *attract RNA polymerase* to the correct start site so transcription can be initiated.

Promoters

- Promoters are sequences in the DNA just upstream of transcripts that define the sites of initiation.



- The role of the promoter is to *attract RNA polymerase* to the correct start site so transcription can be initiated.

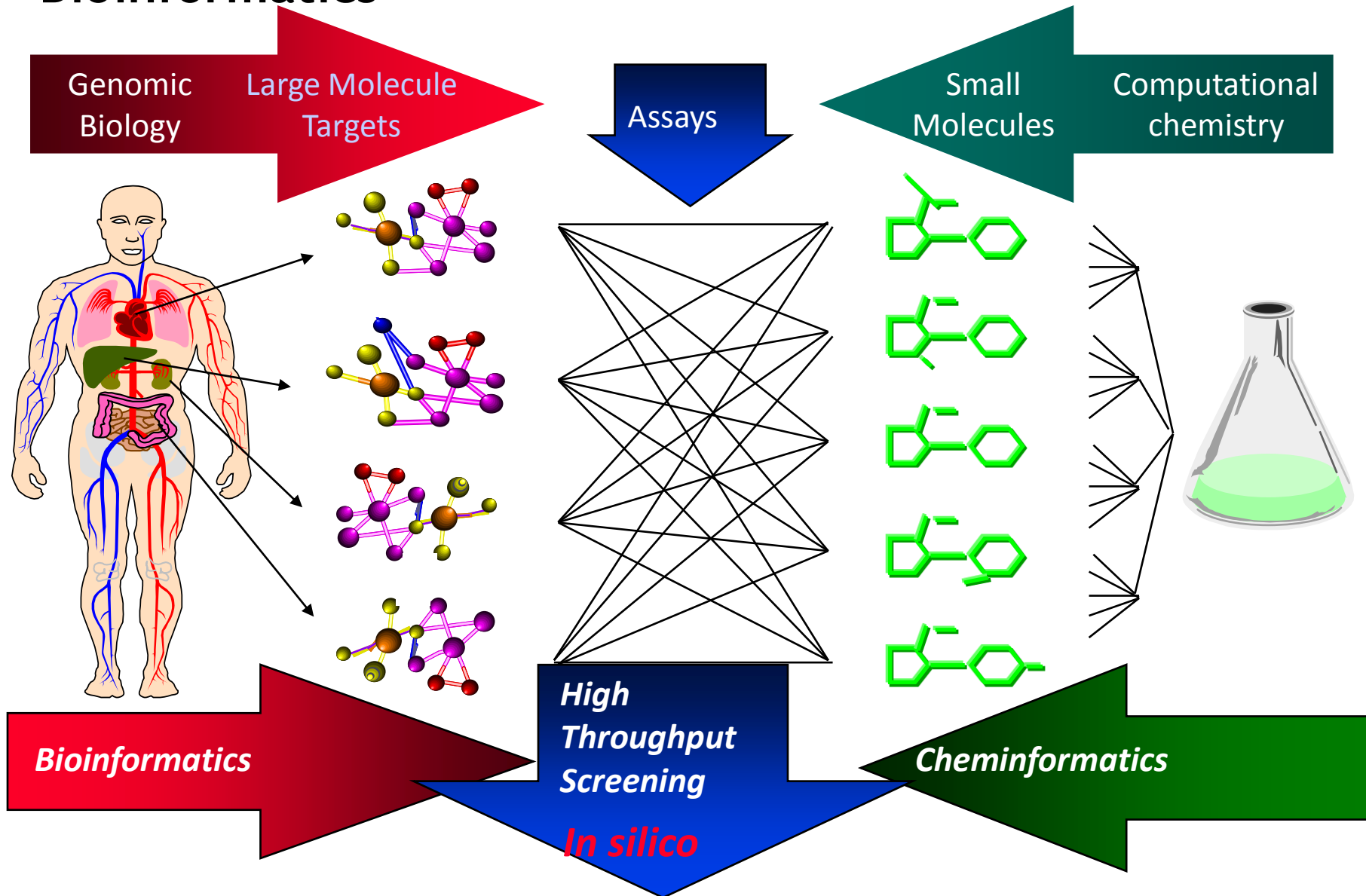
Chemoinformatics

fatchiyah, dept bio UB 2/21/2012

- ⦿ Kombinasi dari sintesis kimia, penyaringan biologis, dan pendekatan *data-mining* yang digunakan untuk penemuan dan pengembangan obat
- ⦿ Ruang lingkup akademis dari *cheminformatics* ini sangat luas. Contoh bidang minatnya antara lain: *Synthesis Planning, Reaction and Structure Retrieval, 3-DStructure Retrieval, Modelling, Computational Chemistry, Visualisation Tools and Utilities.*

Integration of Chemoinformatics and Bioinformatics

fatchiyah, dept bio UB 2/21/2012



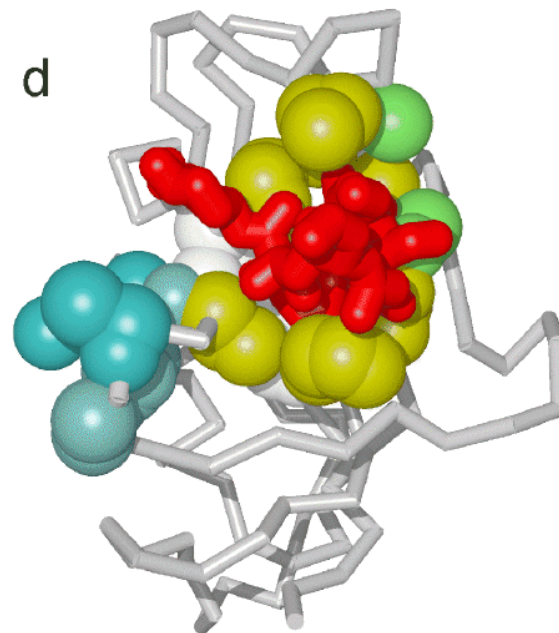
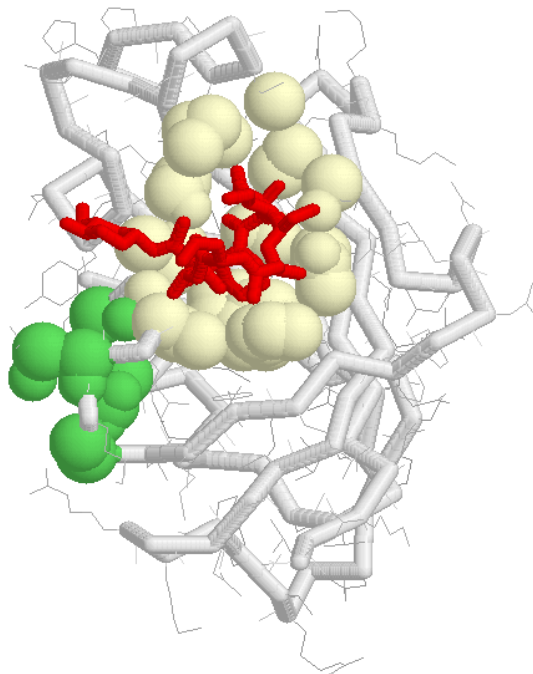
"*proteome*"

- Definisi : "*The PROTEin complement of the genOME*".
- Dan mendefinisikan *proteomics* berkaitan dengan: "**studi kuantitatif dan kualitatif dari ekspresi gen di level dari protein-protein fungsional itu sendiri**".
- Yaitu: "sebuah antarmuka antara **biokimia protein** dengan **biologi molekuler**".

Structural proteomics

fatchiyah, dept bio UB 2/21/2012

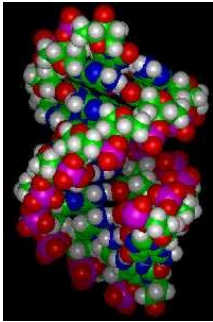
- ⊙ Atlas of Topographic Surfaces of All Known Protein Structures
 - Automatic identification of binding pockets.
 - Measurement size of surface binding pockets.
- ⊙ Drug Discovery
 - Quantifying ligand accessibility.
 - Constructing precise negative imprint or cast of binding site.



Pharmacogenomics

forchuck, dept bio UB 2/21/2012

- The use of DNA sequence information to measure and predict the reaction of individuals to drugs.
- Personalized drugs
- Faster clinical trials
 - Selected trial populations
- Less drug side effects
 - Toxicogenomics



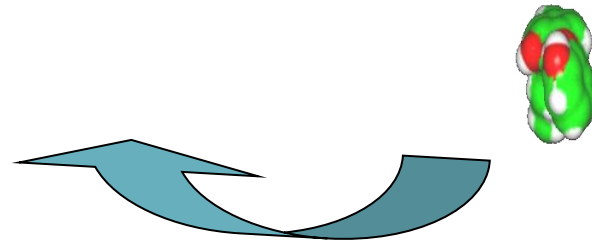
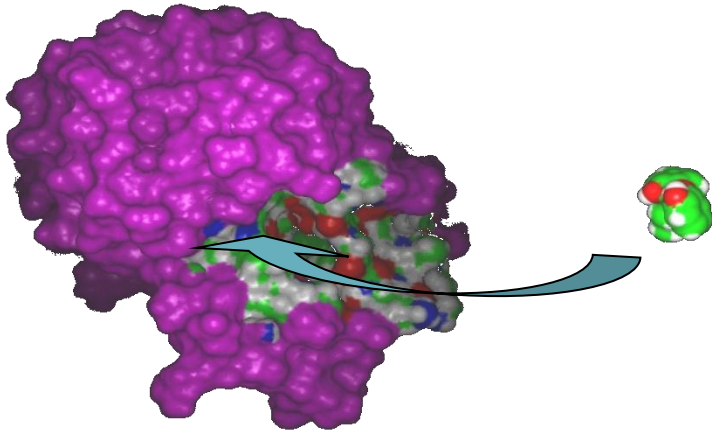
Drug Design

fatchiyah, dept bio UB

2/21/2012

Structure based

Ligand based



Biologists vs Computer scientists

Biologists

- (Almost) Nothing is ever completely true or false
- Biologists strive to understand the very **complicated**, very messy **natural** world.
- more **data driven**
- obsessed with being the first to **discover** something

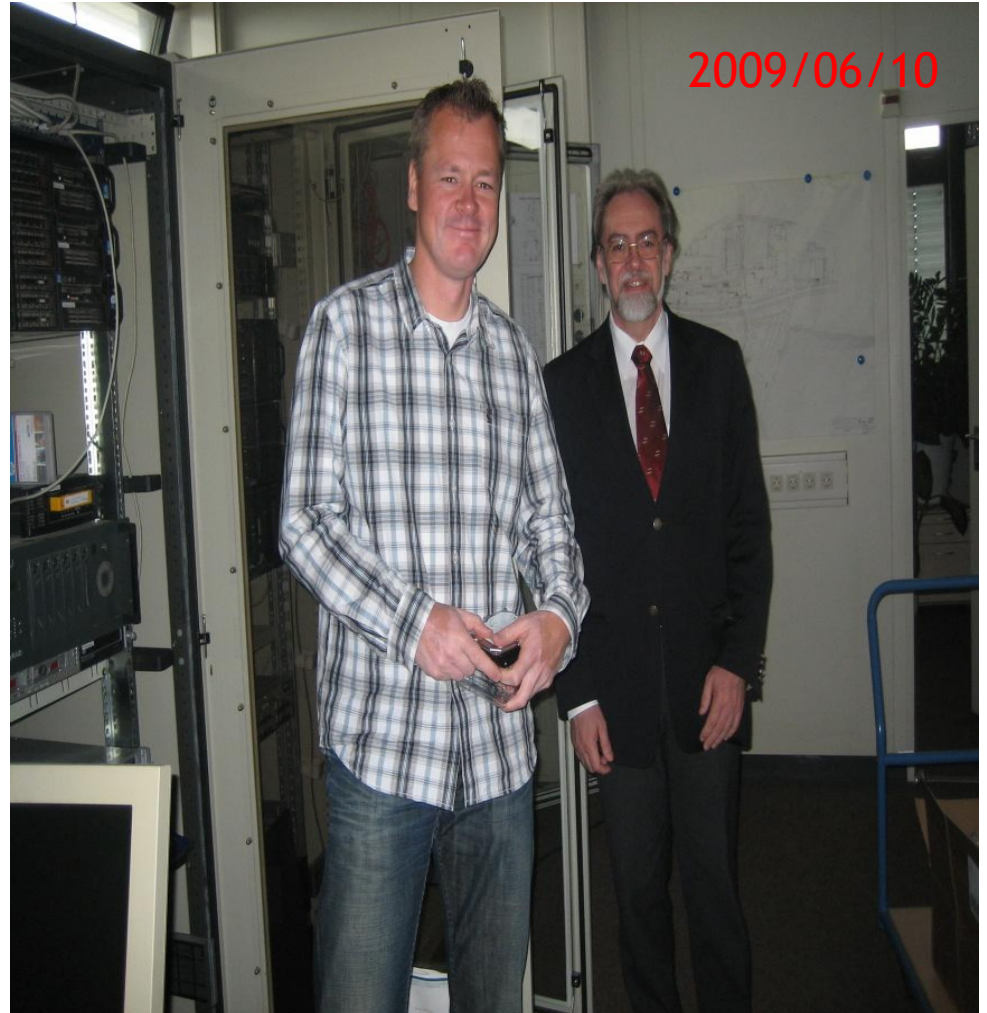
Computer scientists

- Everything is either true or false
- Computer scientists seek to build their own clean and **organized virtual** worlds
- more **algorithm** driven
- obsessed with being the first **to invent or prove** something

Prof Edgar Wagener

fatchiyah, dept bio UB 2/21/2012

President & CEO of BioBase, wolfenbuttel german
Prof of Endocrinology of Medical School of Gouttingen
University, German



Ground rules for bioinformatics

- ⦿ Don't always believe what **programs** tell you

- ⦿ Don't always believe what **databases** tell you

- ⦿ In short, don't be **a naive** user

- computers don't do biology

- ⦿ **Be yourself**