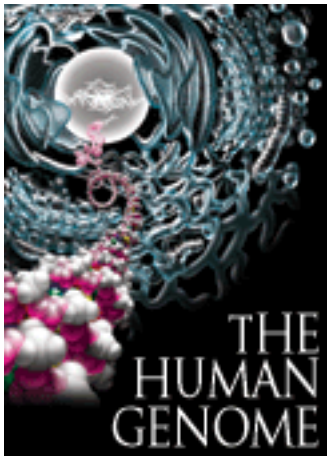



# HUMAN GENOME PROJECT (HGP)

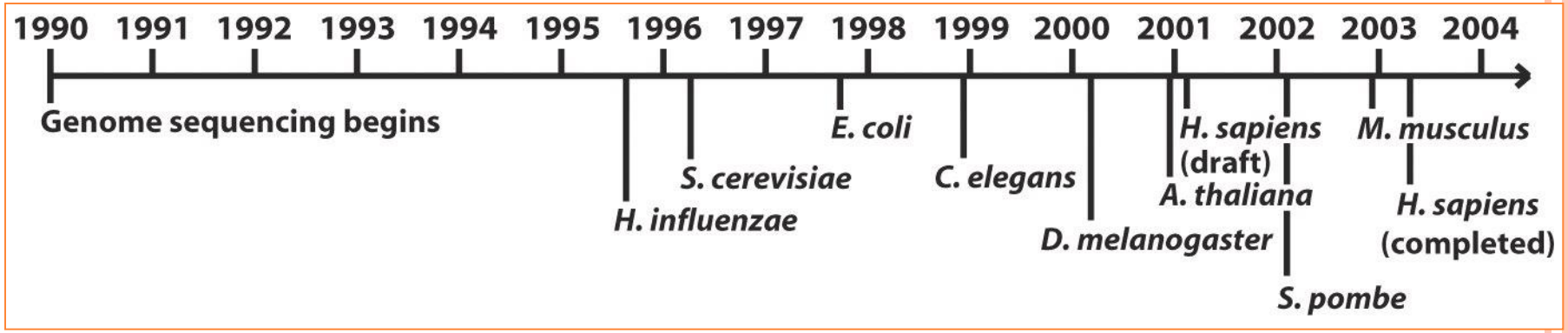
**Fatchiyah**

**http://fatchiyah.lecture.ub.ac.id**

# Human genome project

- The human genome project was initiated in 1989 with the goal of sequencing the 3 billion base-pair human genome in 15 years.The National Institutes of Health and the Department of Energy instituted the joint project. 20 centers contributed.
- There was great skepticism that this could be accomplished in a reasonable amount of time.
- In 1998, the company Celera genomics formed to sequence the human genome.
- Celera and the HGP concurrently announced the human genome draft in 2001. The genome was completed in 2004.

06.3.01



# Human genome project results

- Estimated 27,894 genes
- ~1.1% in exons.
- 1/1000 bp differ between individual humans: SNPs (single nucleotide polymorphisms
- From SNPs arise human variety.

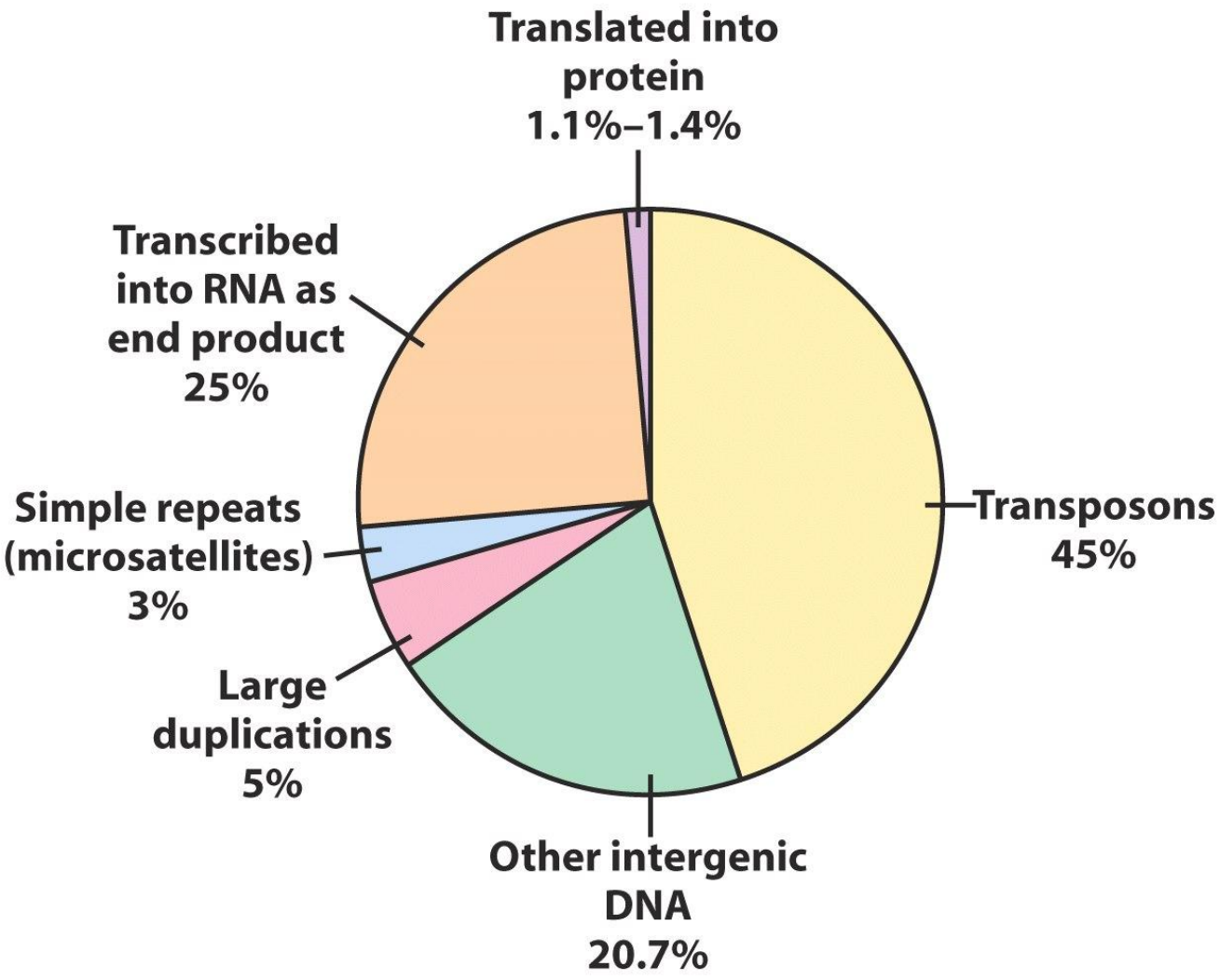



Fig. 9-19

# Human genome project results

- <1% of SNPs are expected to impact protein function.
- Thus, thousands of genetic variations contribute to human diversity (not millions!)

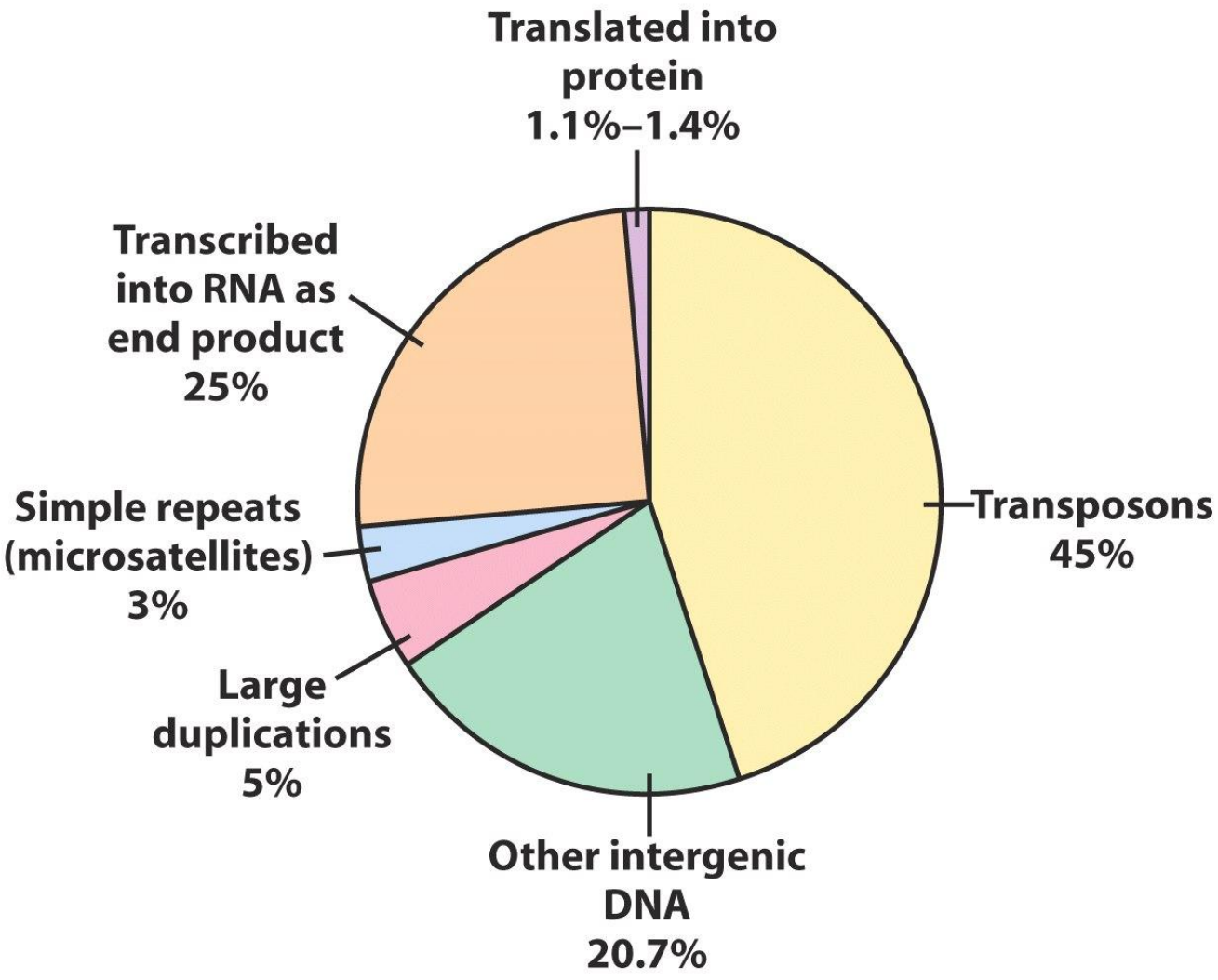
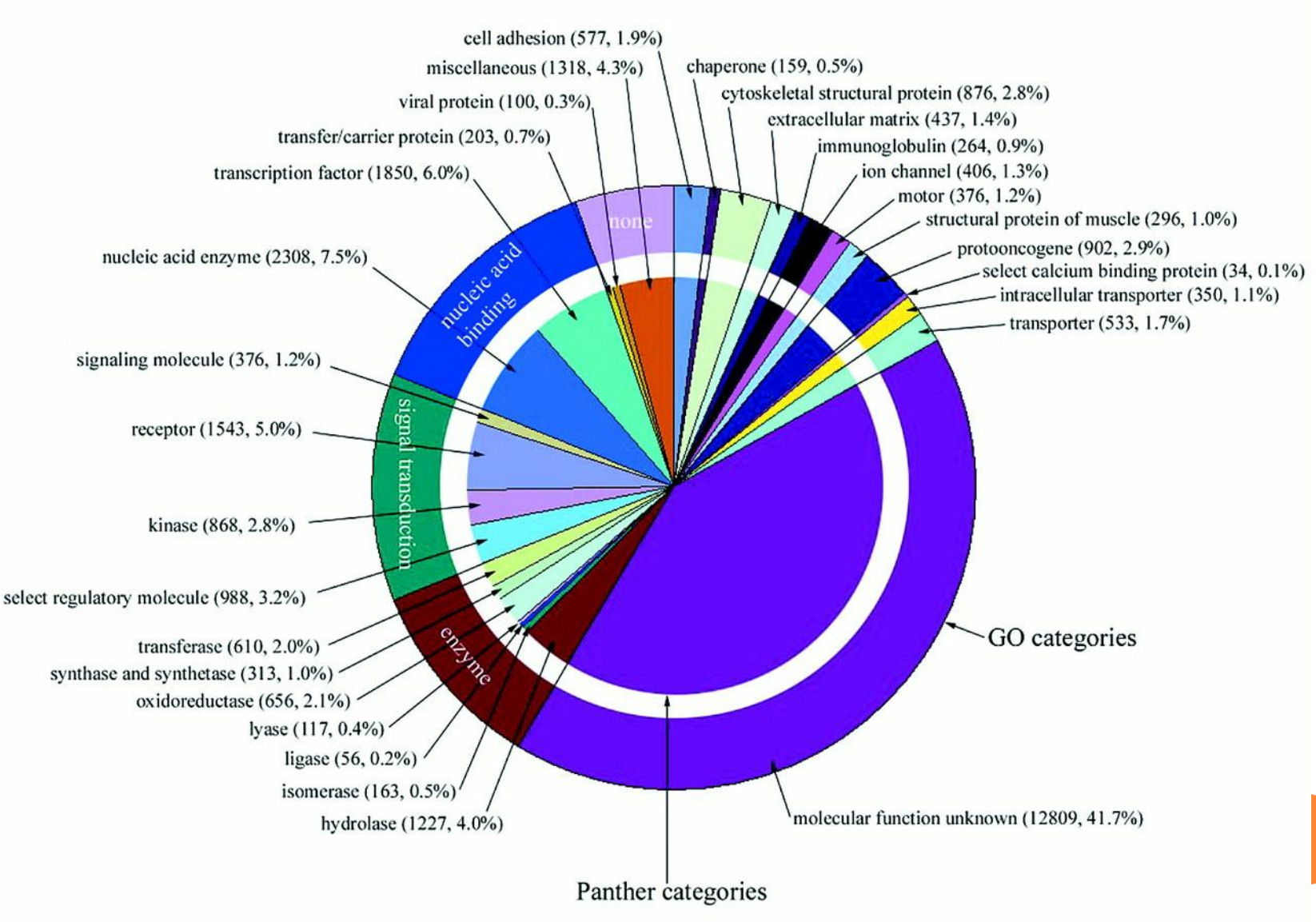



Fig. 9-19

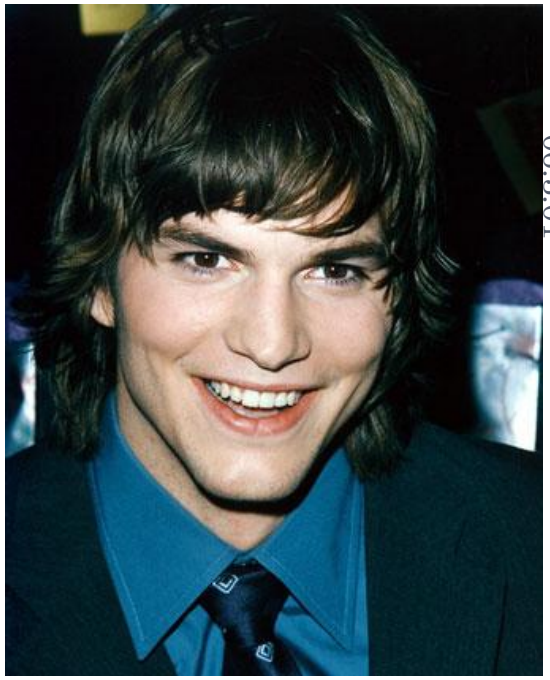
# Human genome project - gene function




Venter et al., *Science* (2001)*291*, 1304-1351.

06.3.01

# Human genome project



06.3.01

Fruit fly: 13,000 genes

Human: 35,000 genes.

The surprisingly small number of genes in the human genome (~ 100,000 expected; < 30,000 identified) was a major surprise from the project.

# Human genome project

The modest number of genes indicates we must look elsewhere to explain the human complexity.

06.3.01

- Gene regulation, modification (i.e. methylation)
- Chromosomal modifications
- Location, quantity, timing of transcription
- Tissue-specific protein expression
- Roles (regulatory, other) of intronic DNA
- RNA splicing
- RNA roles in gene expression
- RNA editing (changes made to mRNA)
- Translational control (at ribosome)
- Alterations in protein-protein interactions



Venter et al., *Science* (2001)*291,* 1304-1351.

# Proteomics

Proteomics is the determination and analysis of the complete complement of proteins expressed by a genome.

06.3.01

- There remain thousands of proteins in each eukaryotic cell about which we know nothing.
- Characterizing the proteome is a much larger task than the genome. This links genes to function:
  - Phenotypic function: effects of a protein on an entire organism
  - Cellular function: the network of interactions with other proteins in the cell
  - Molecular function: the biochemical activity of a protein

# Proteomics strategies

- Comparative genomics: Compare with genes and proteins of known function. Uses sequence and structural relationships. The increasing availability of genomics data greatly aids this approach.
  - Orthologs: Genes of different species but possessing a clear sequence and functional relationship to each other.
  - Paralogs: Genes within an organism with a sequence and structural relationship.

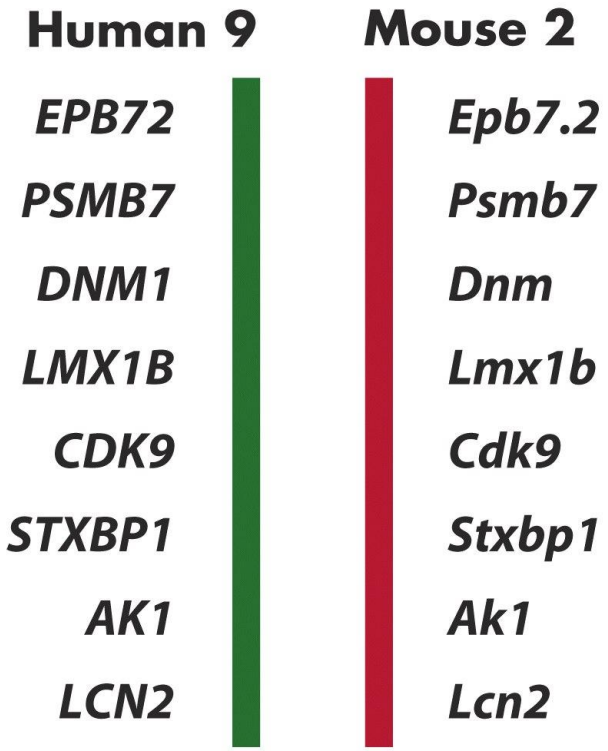



Fig. 9-20

# Proteomics strategies

For genes with no identifiable relationships to known genes, other approaches need to be applied.

06.3

2-D gel electrophoresis and mass spectrometry: Analyze the appearance or particular proteins from different tissues, as a function of development, or from tissues treated in different ways.


Fig. 3-22

# Probing Protein Interactions

Analysis of protein-protein interactions also can reveal important information about a protein's function and its role in the cell.

The yeast two-hybrid system allows for detection of protein-protein interactions by bringing together the DNA binding domain and the activation domain of the yeast Gal4 protein via interaction of two proteins and expression of a reporter gene.

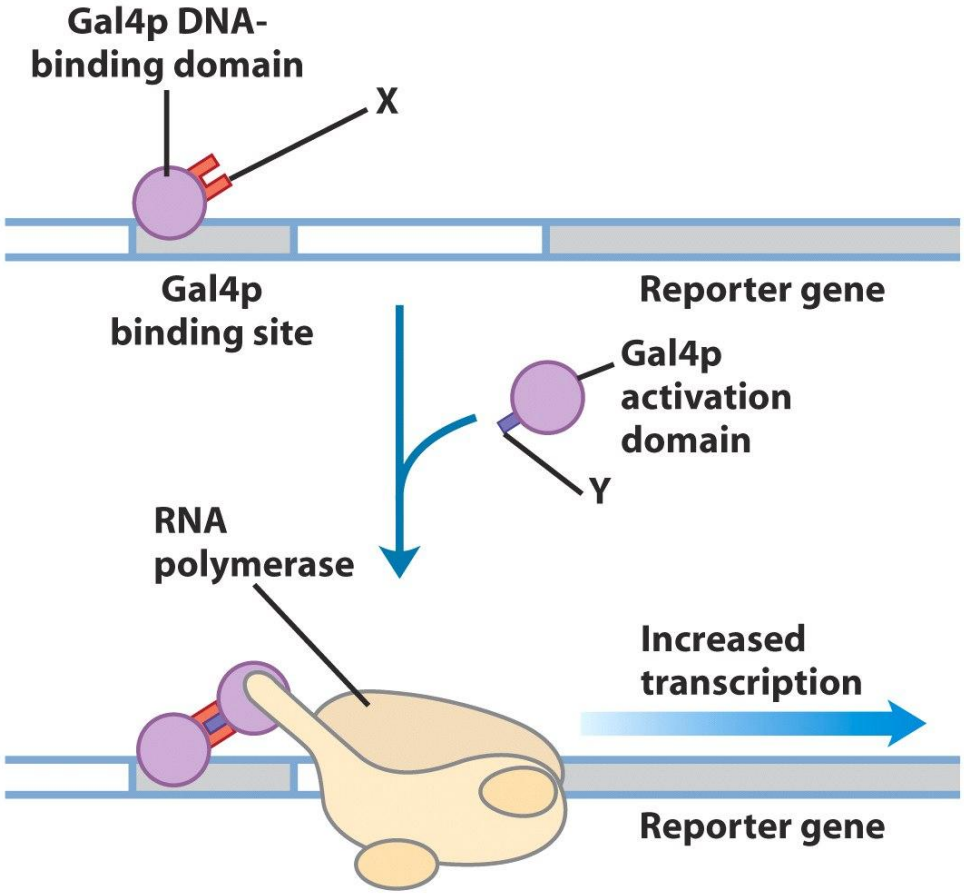



Fig. 9-25

06.3.01

# Proteomics strategies

Analysis of protein-protein interactions also can reveal important information about a protein's function and its role in the cell.

The two fusions are created in separate yeast strains which are mated. The mated mixture is grown under conditions on which the yeast cannot survive unless the reporter gene is expressed. Surviving colonies have interacting protein fusion pairs.

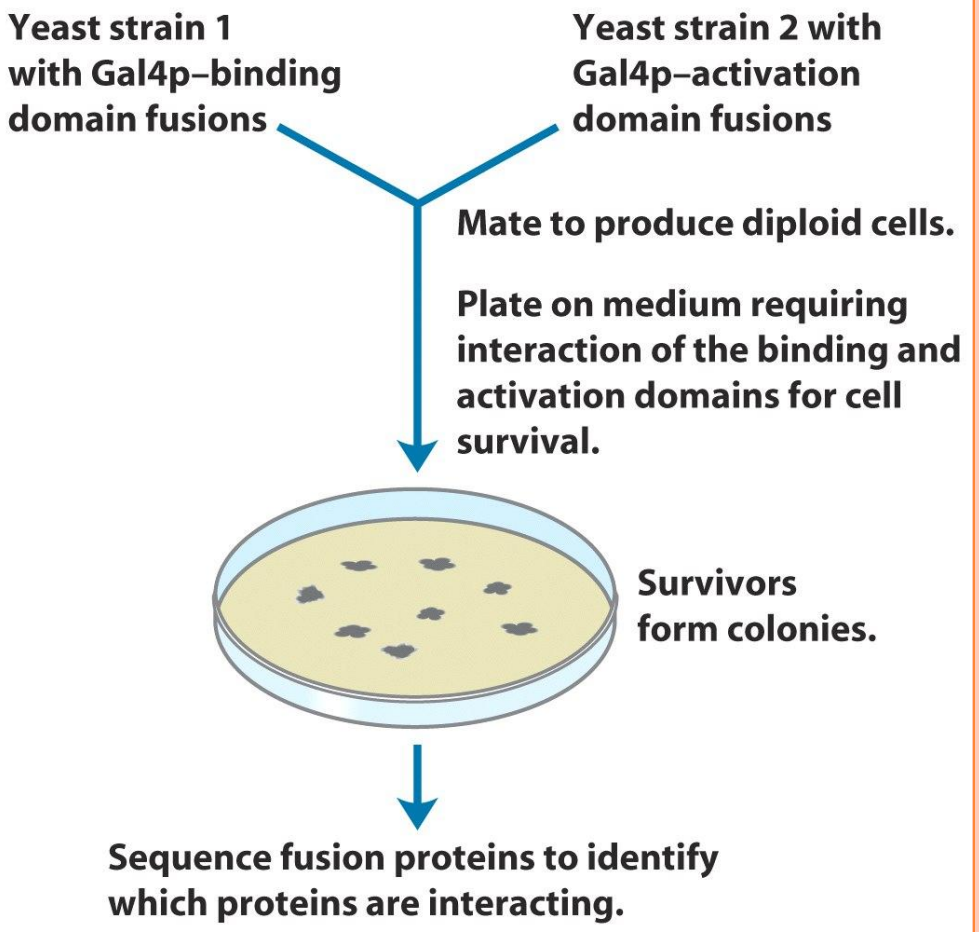



Fig. 9-25

# Gene cloning and expression plasmids

- Ori: where plasmid replication is initiated by cellular enzymes. This is required to propagate the plasmid within the cell.
- *tet*$^R$ and *amp*$^R$: genes that confer resistance of the antibiotics tetracycline and ampicillin.
- *Eco*RI, *Bam*HI, … unique sequences that are targets for endonucleases. Provide sites for cutting the plasmid.

A common *E. coli* cloning plasmid, pBR322:

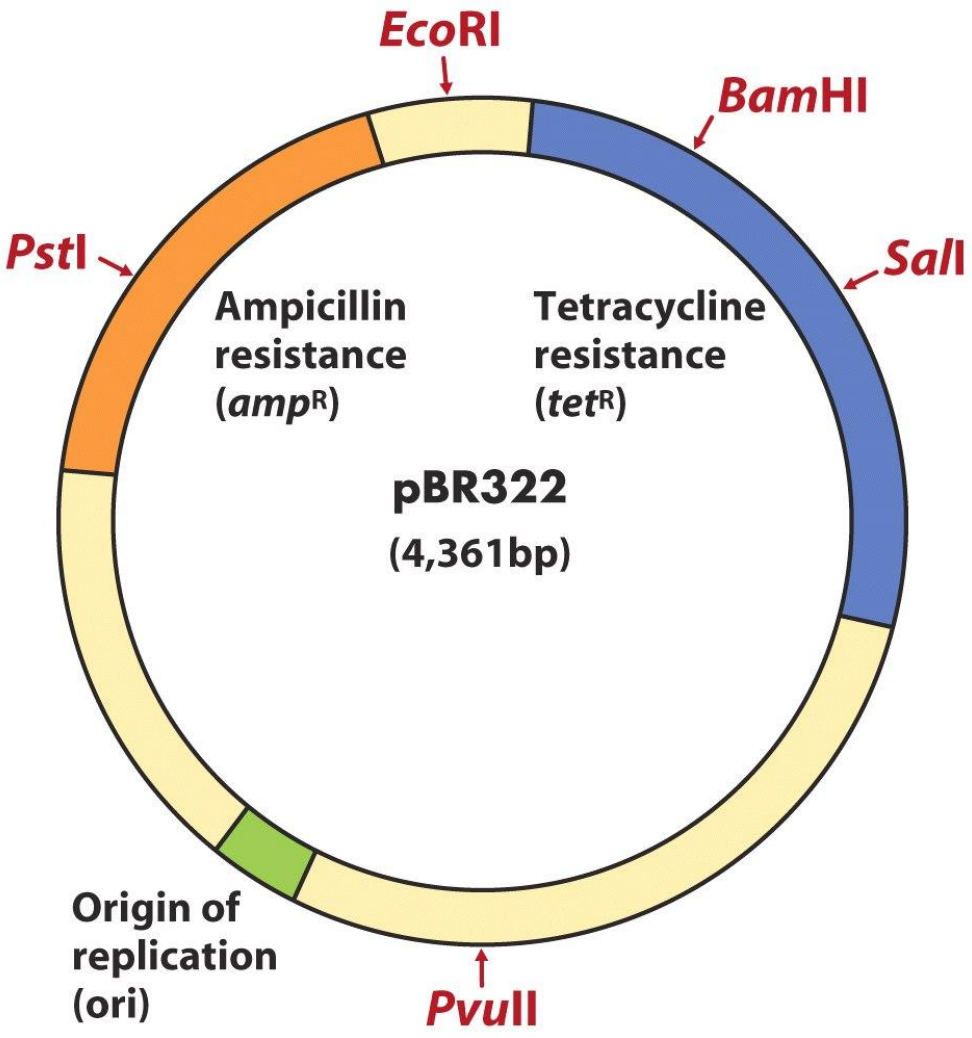



Fig. 9-4

# Micro Arrays

- **Tests for the presence of a nucleic acid sequence by hybridizing a probe bound to a matrix to the target sequence.**
- **Many different probes can be bound to the same matrix.**
- **Therefore, a single sample can be evaluated for many different target sequences simultaneously.**

# Micro Arrays

- **Expression Arrays - tests for mRNA expressed in a tissue.**
- **Sequencing Arrays - tests for nucleotide sequence in a fragment of DNA (sequencing by hybridization - ideal for detection of single nucleotide polymorphisms[snps]).**

# DNA microarray

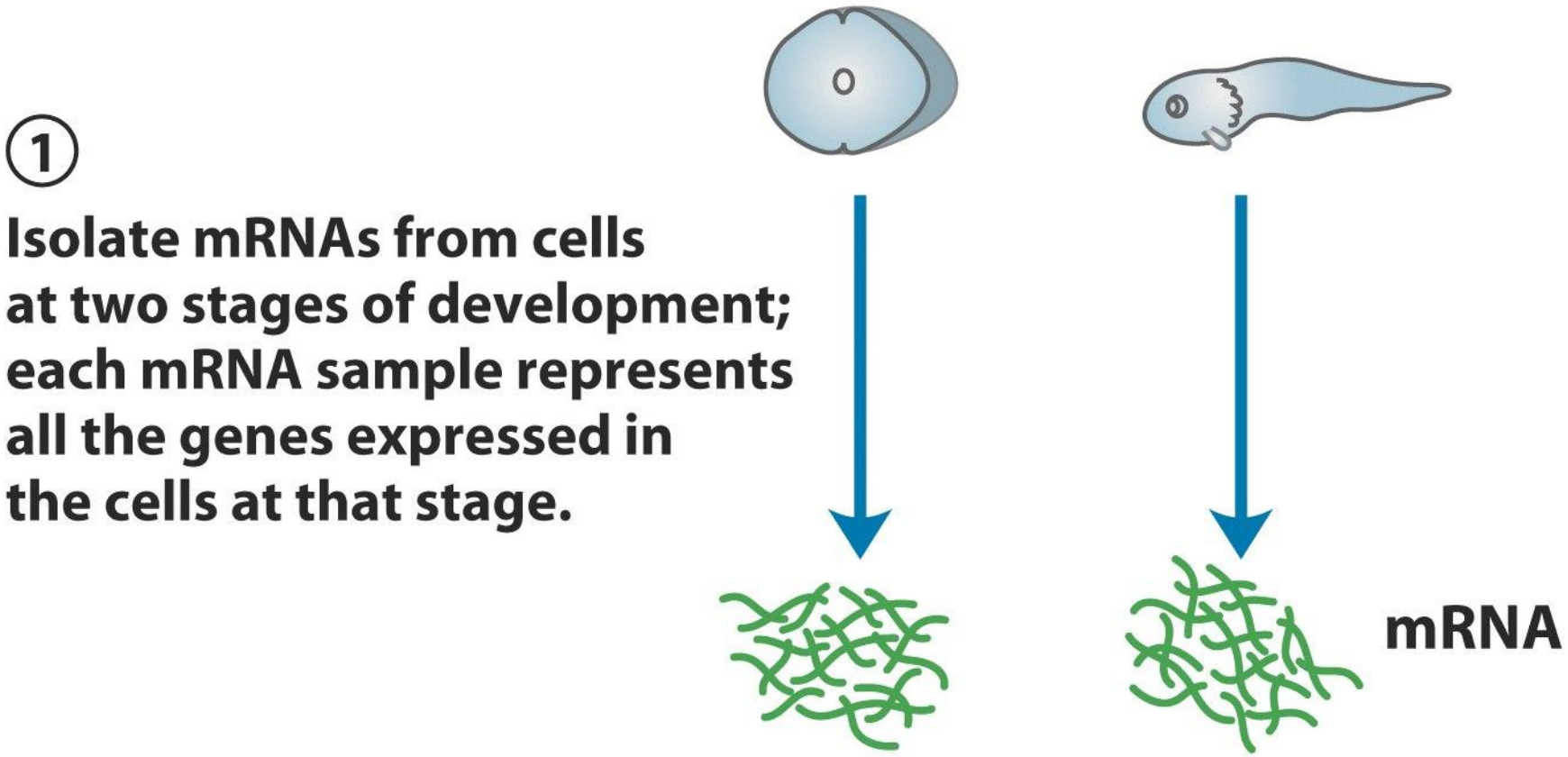
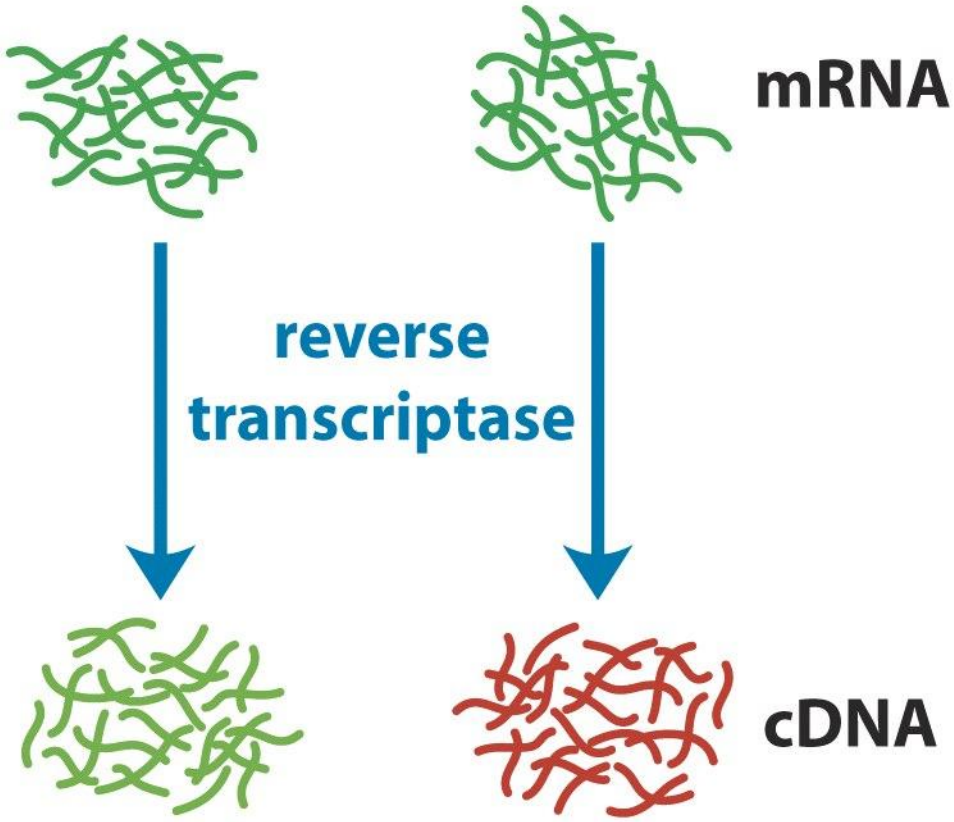





Fig. 9-22

# DNA microarray

② **Convert mRNAs to cDNAs by reverse transcriptase, using fluorescently labeled deoxyribonucleotide triphosphates.**

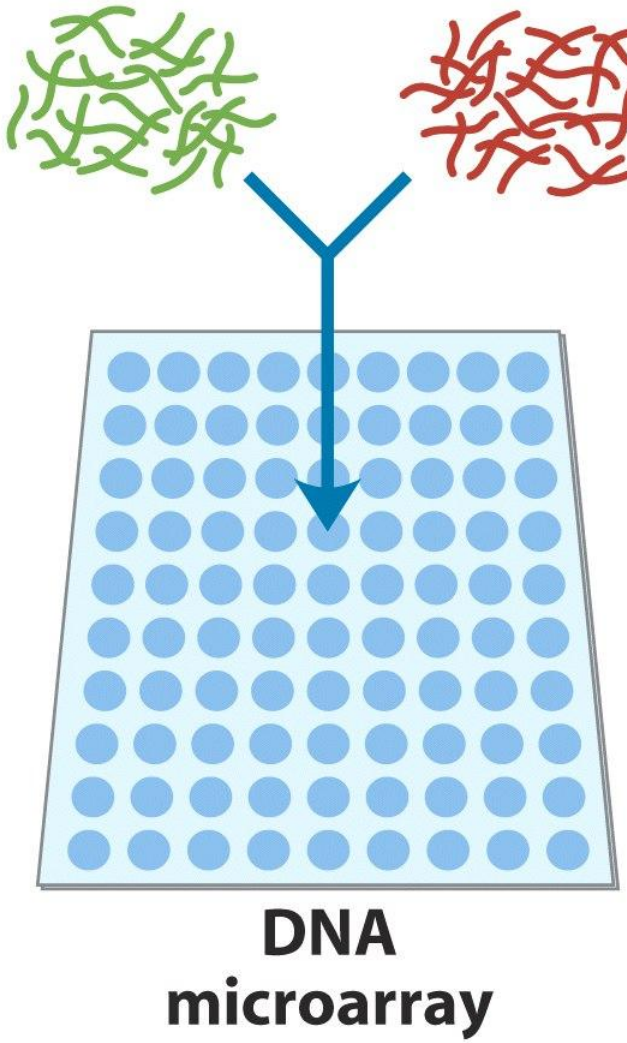



cDNA: complementary DNA (prepared from mRNA).



Fig. 9-22

# DNA microarray




③

**Add the cDNAs to a microarray; fluorescent cDNAs anneal to complementary sequences on the microarray.**

06.3.01



Fig. 9-22

# DNA microarray

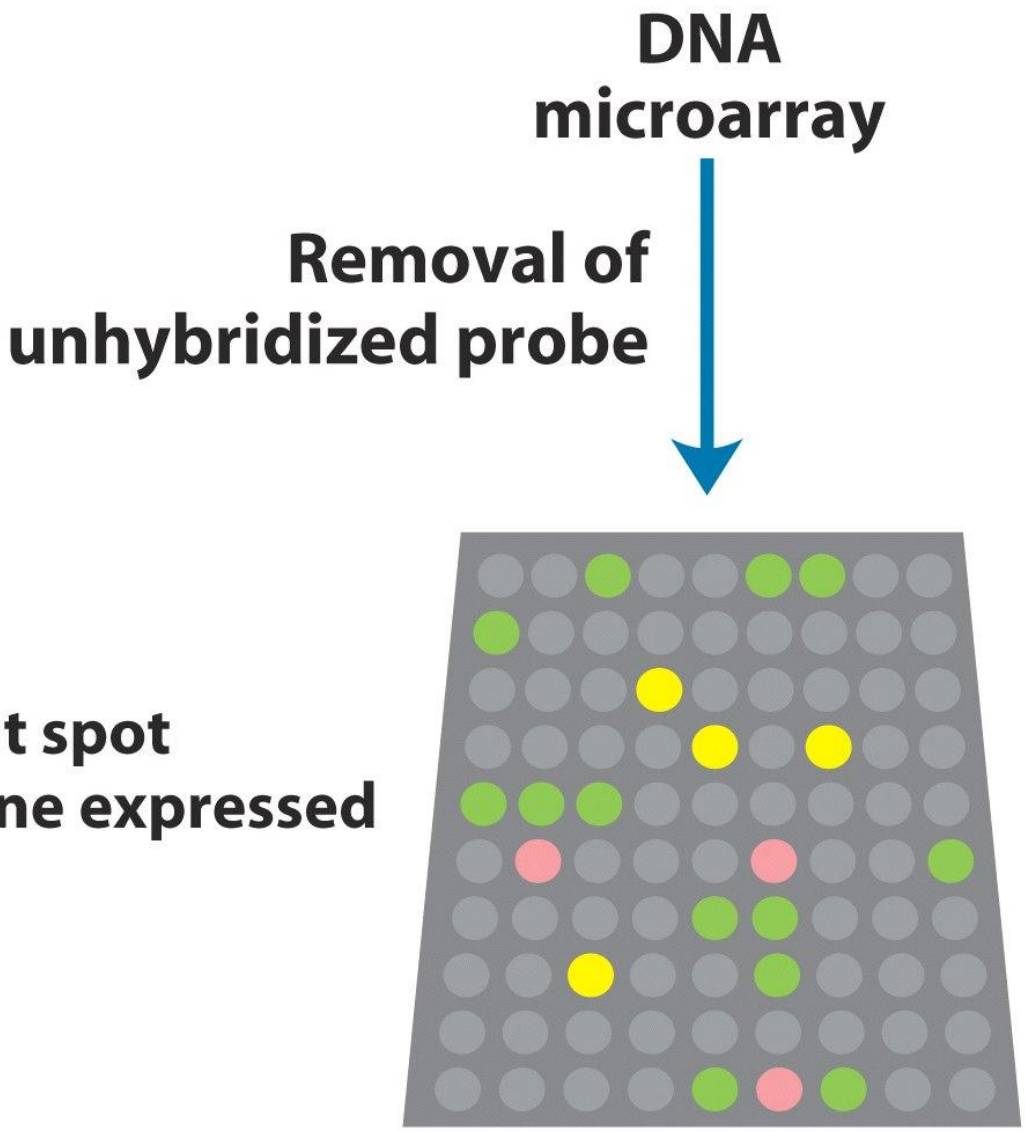



Fig. 9-22

# Gene Expression Profile

Each spot in this microarray contains DNA from one of the 6,200 genes in the yeast genome.

The different colors indicate conditions under which the genes are expressed. Here, green spots represents mRNAs abundant early in development, red RNAs are abundant later in development.

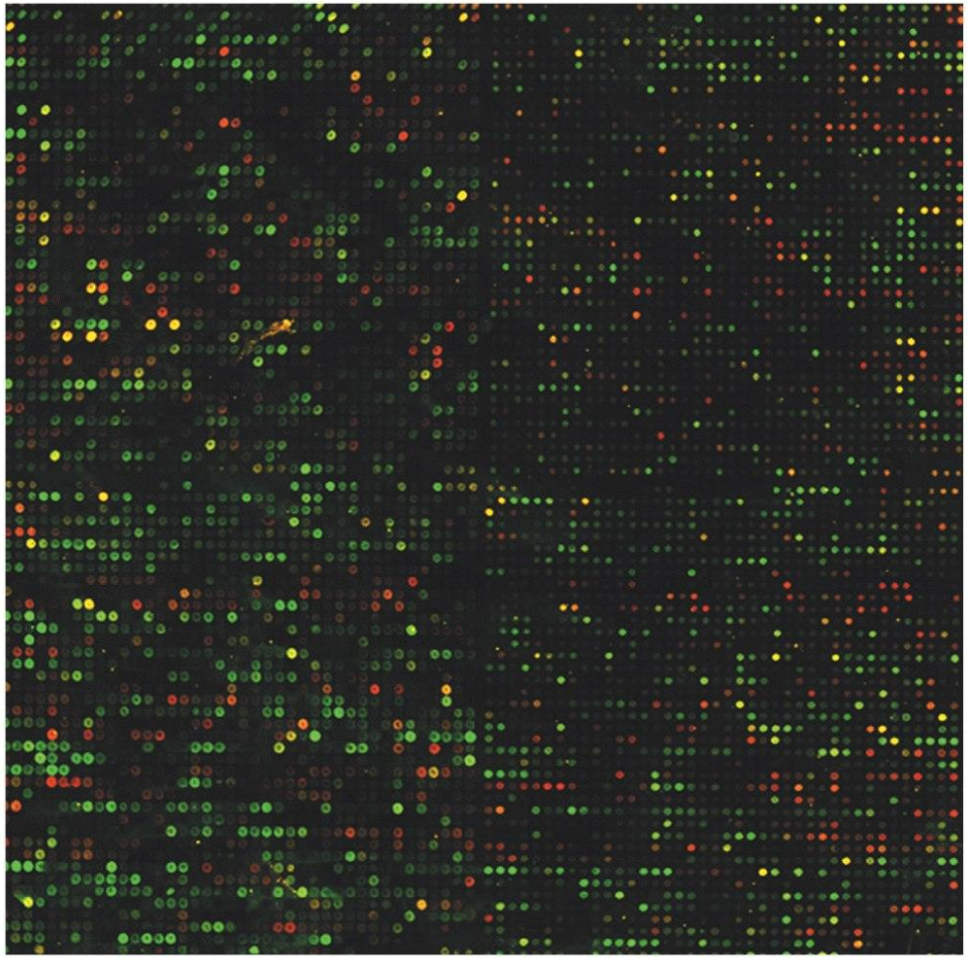
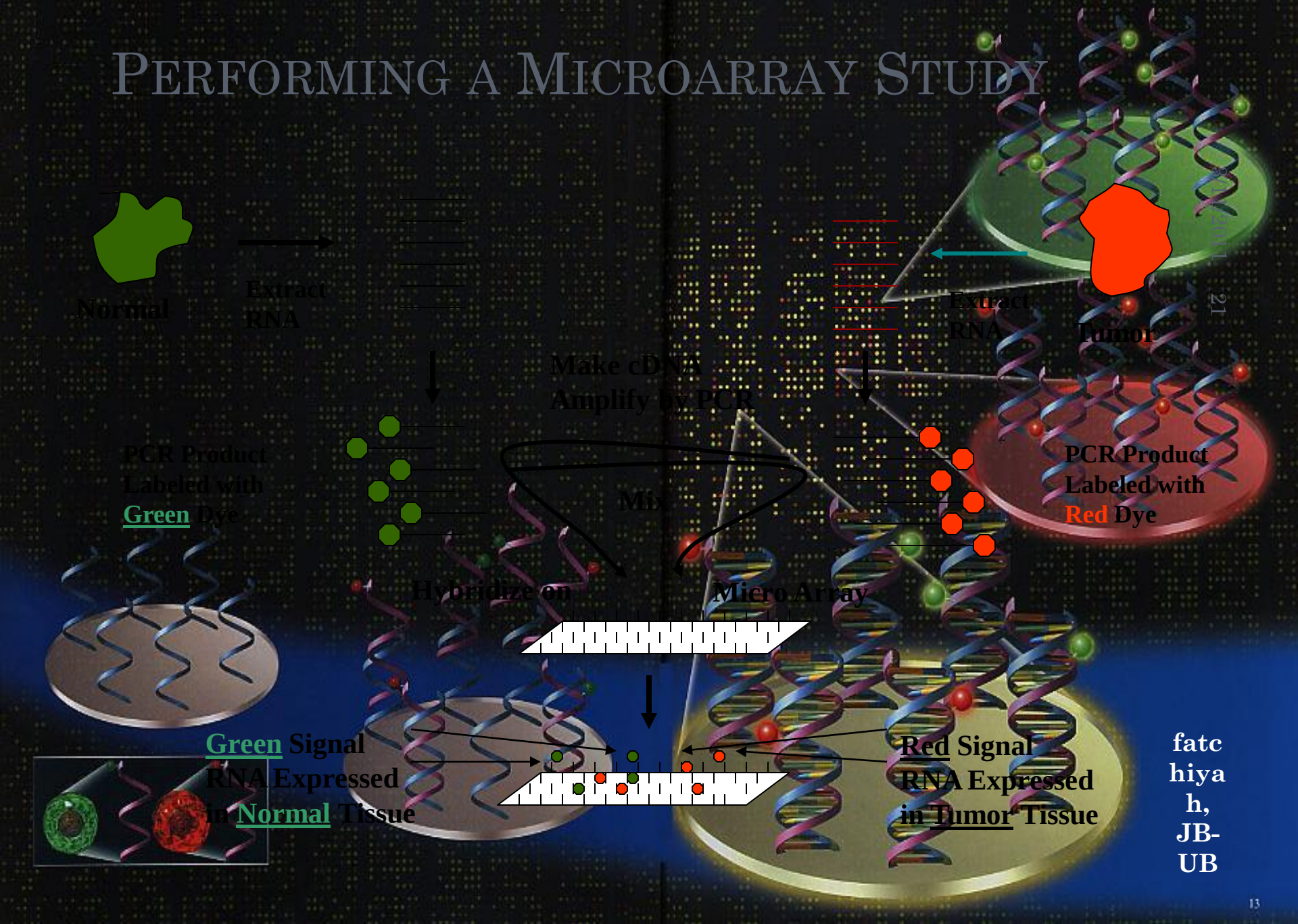


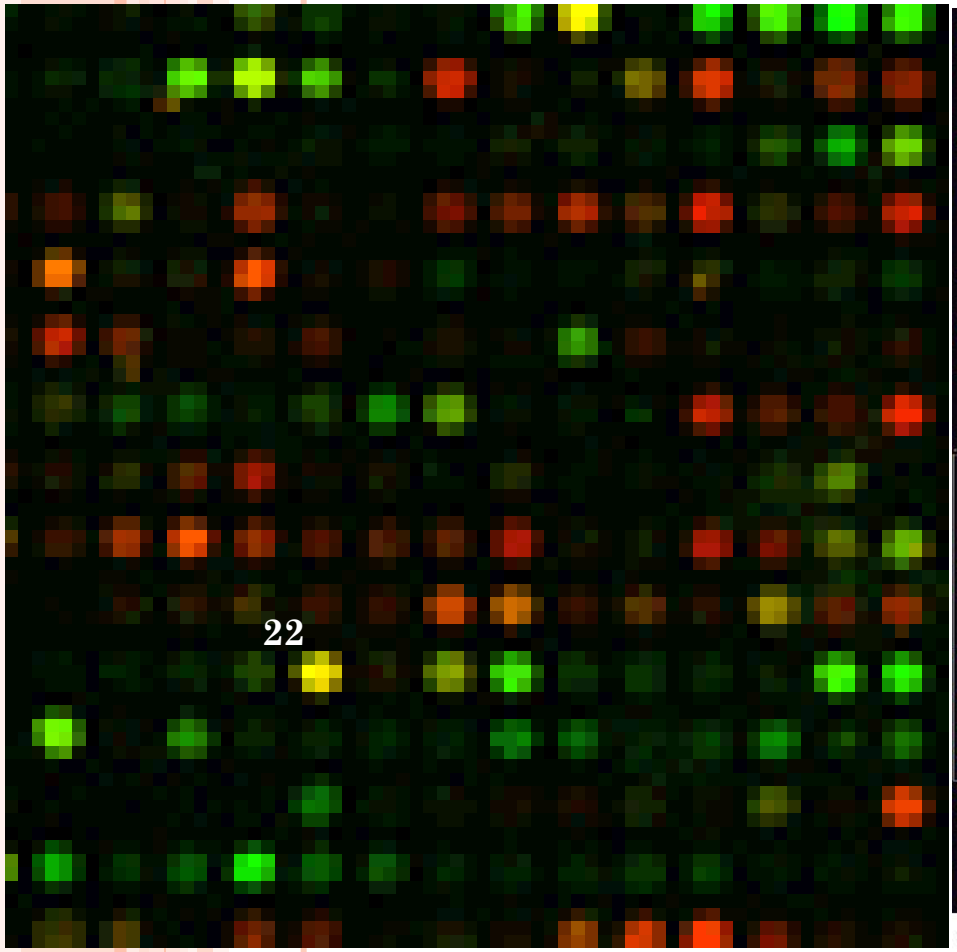


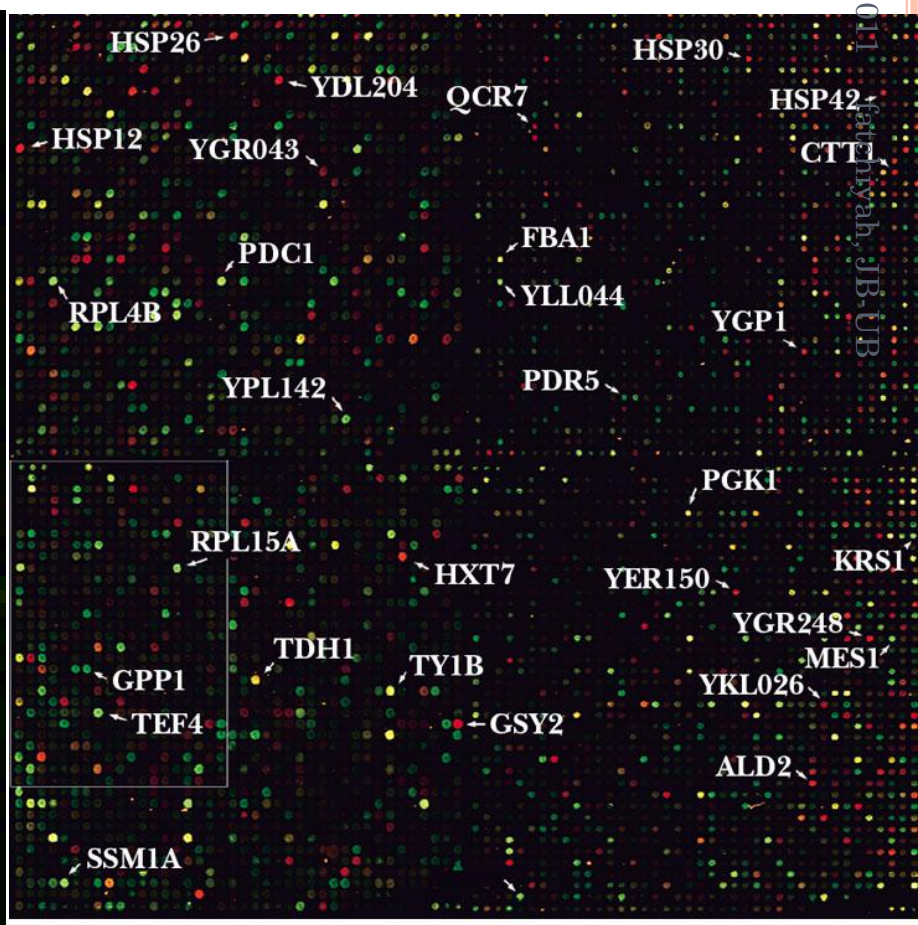
Fig. 9-23

# Performing a Microarray Study

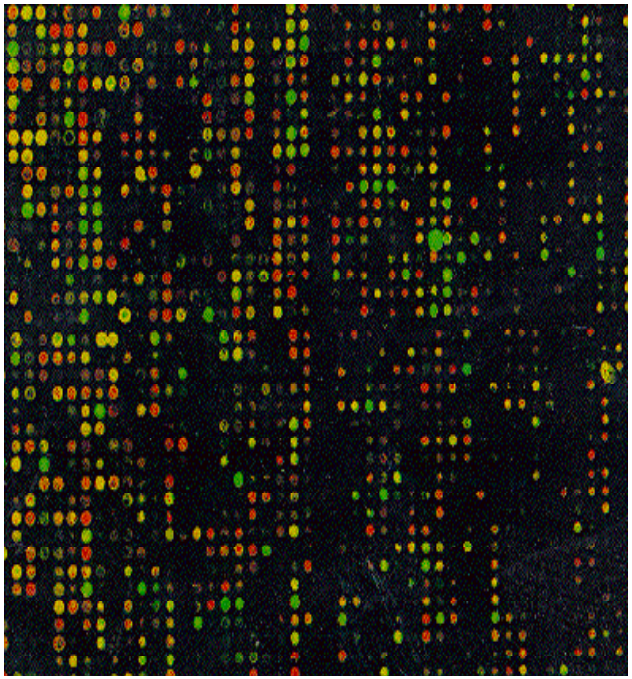


# Microarray result

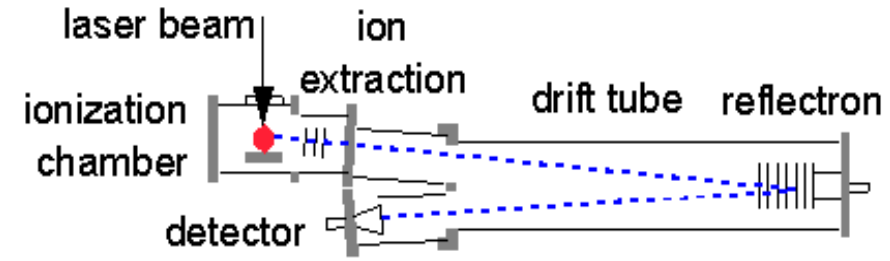


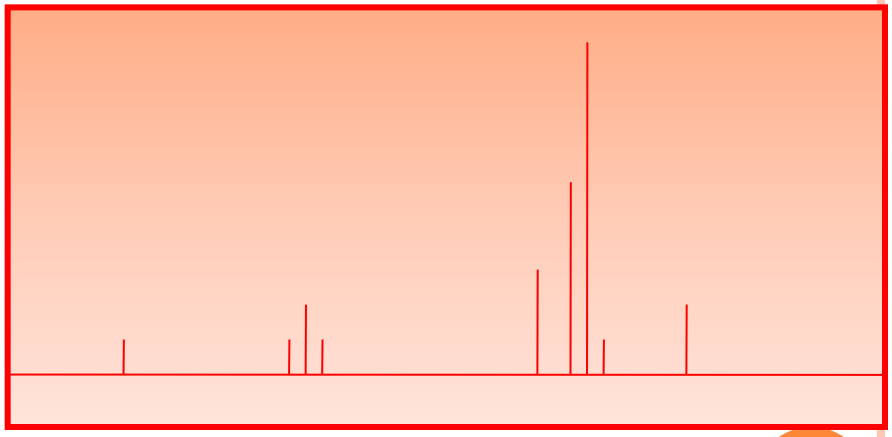



# EXPRESSIONAL PROTEOMICS

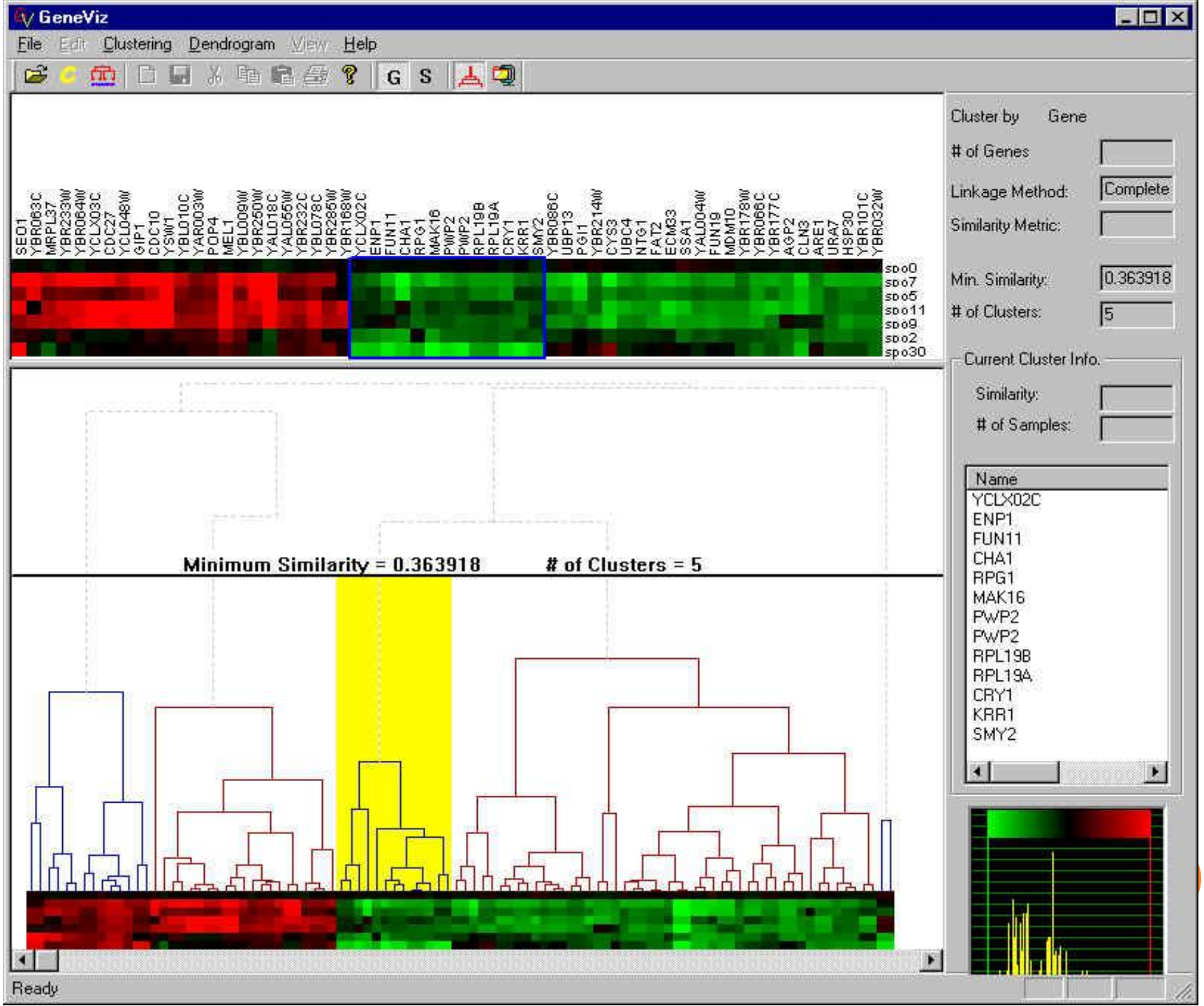


**Microarray**



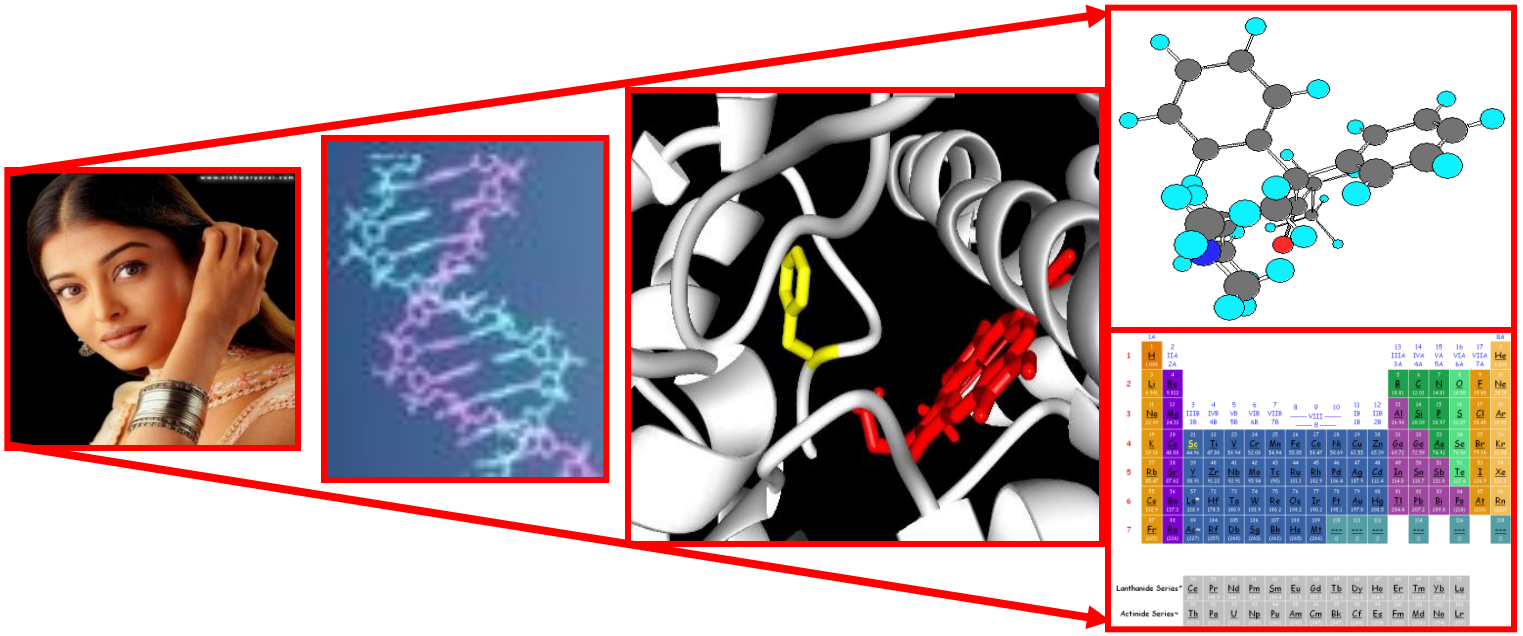



**QTOF Mass Spectrometry**


GeneViz
File Edit Clustering Dendrogram View Help
G S
SEO1 YBR063C MRPL37 YBR233W YBR064W YCLX03C CDC27 YCL048W GIP1 CDC10 YSW1 YBL010C YAR003W POP4 MEL1 YBL009W YBR250W YAL018C YAL055W YBR232C YBL078C YBR285W YBR168W YCLX02C ENP1 FUN11 CHA1 RPG1 MAK16 PWP2 PWP2 RPL19B RPL19A CRY1 KRR1 SMY2 YBR086C UBP13 PGI1 YBR214W CYS3 UBC4 NTG1 FAT2 ECM33 SSA1 YAL004W FUN19 MDM10 YBR178W YBR066C YBR177C AGP2 CLN3 ARE1 URA7 HSP30 YBR101C YBR032W
spo0
spo7
spo5
spo11
spo9
spo2
spo30
Minimum Similarity = 0.363918
# of Clusters = 5
Cluster by Gene
# of Genes
Linkage Method: Complete
Similarity Metric:
Min. Similarity: 0.363918
# of Clusters: 5
Current Cluster Info.
Similarity:
# of Samples:
Name
YCLX02C
ENP1
FUN11
CHA1
RPG1
MAK16
PWP2
PWP2
RPL19B
RPL19A
CRY1
KRR1
SMY2
Ready

# What is Molecular Modeling?

- A science that elucidates and validates experimental evidence through imagination, visualization, and rationalization
- Applied in many areas of research (Academic/Industrial)

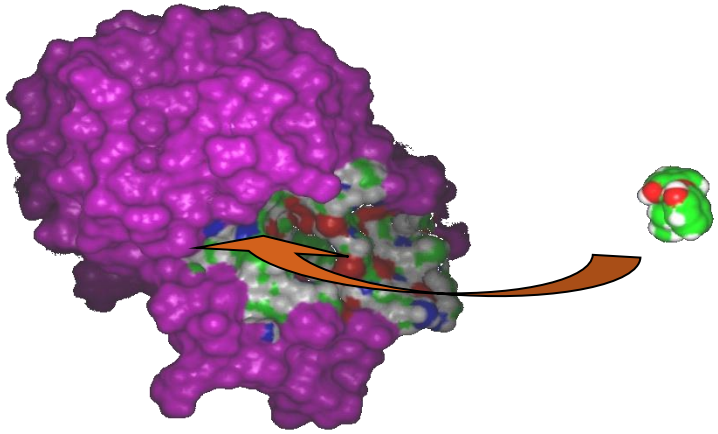


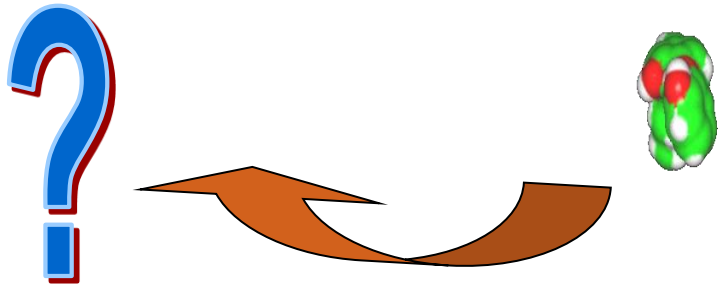
Caveat: Is the interpolation and extrapolation reliable?

# Drug Design

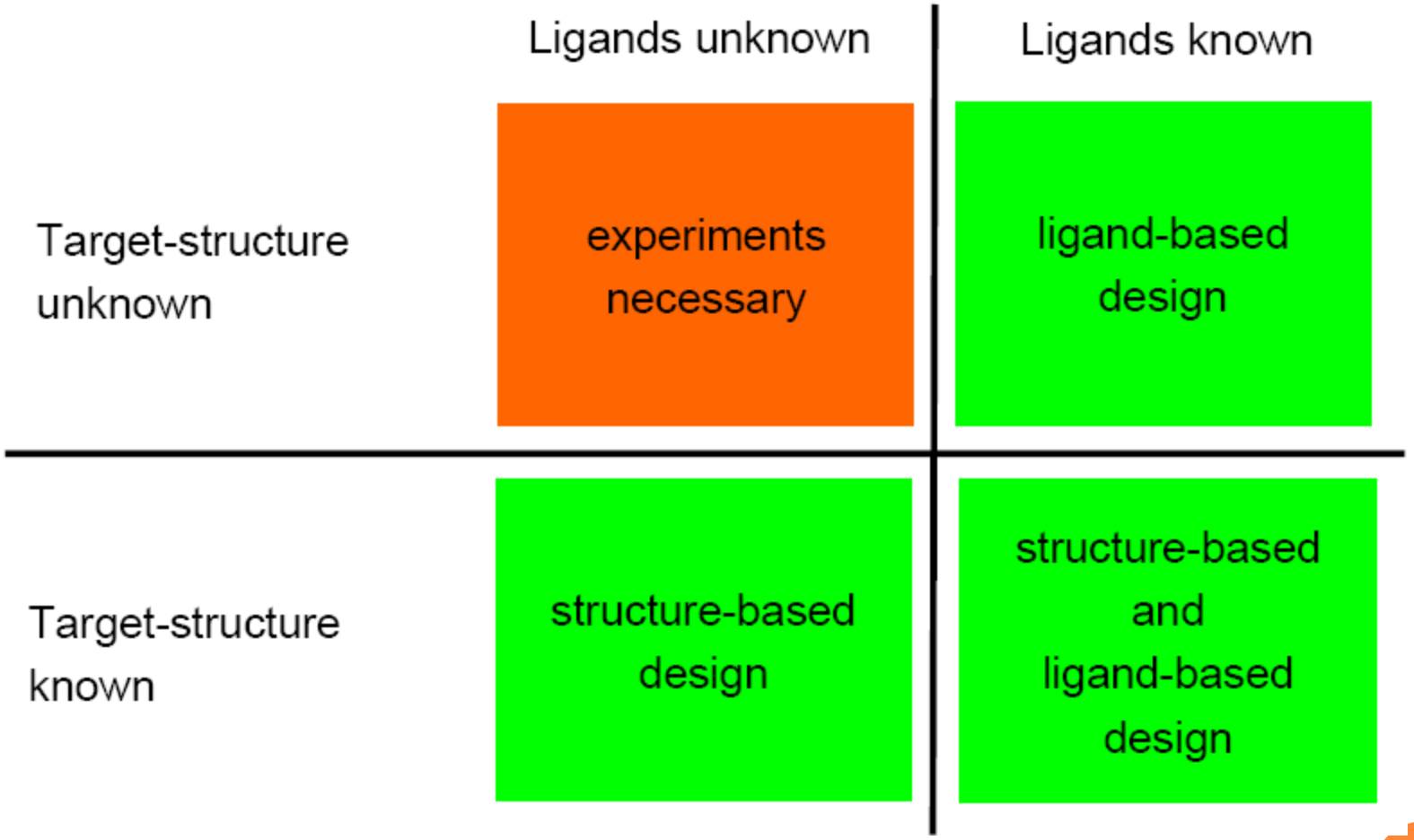
Structure based

Ligand based





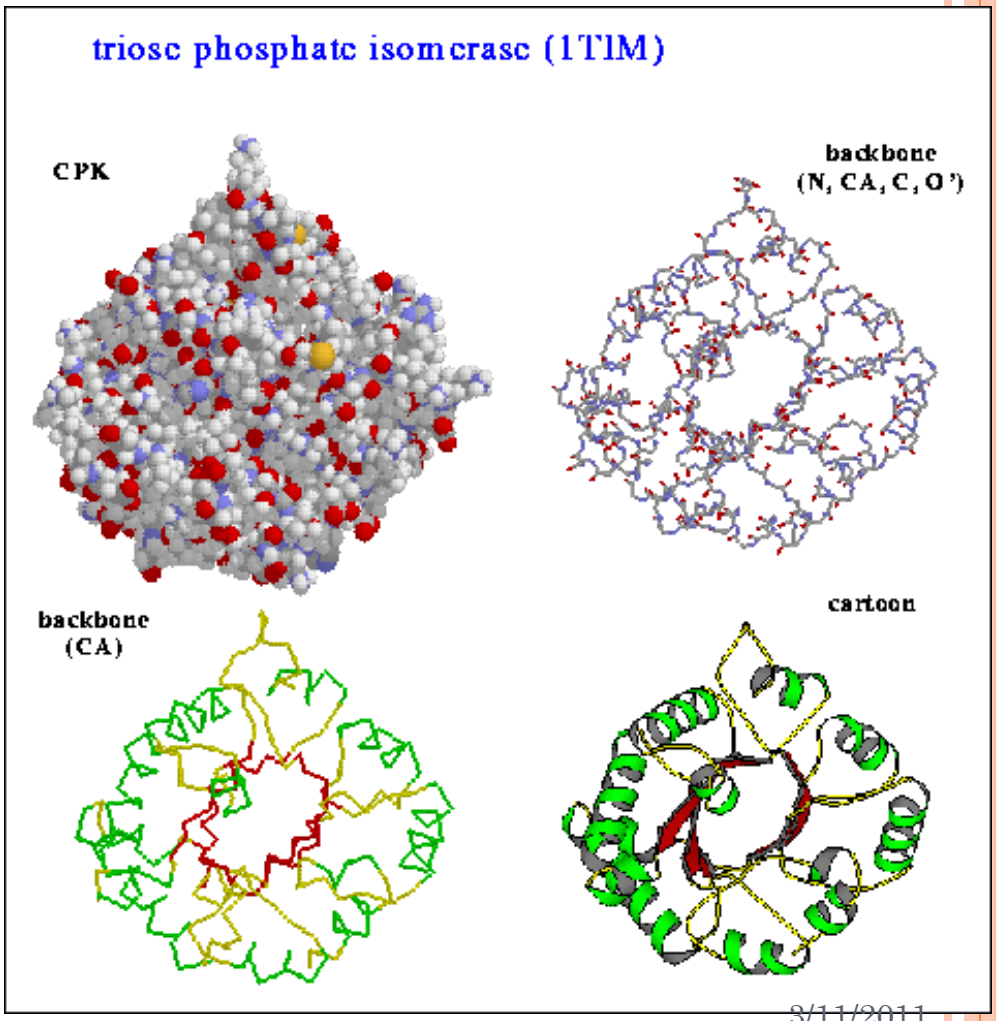
## Structure and Ligand Based Design



# From protein sequence to Structure

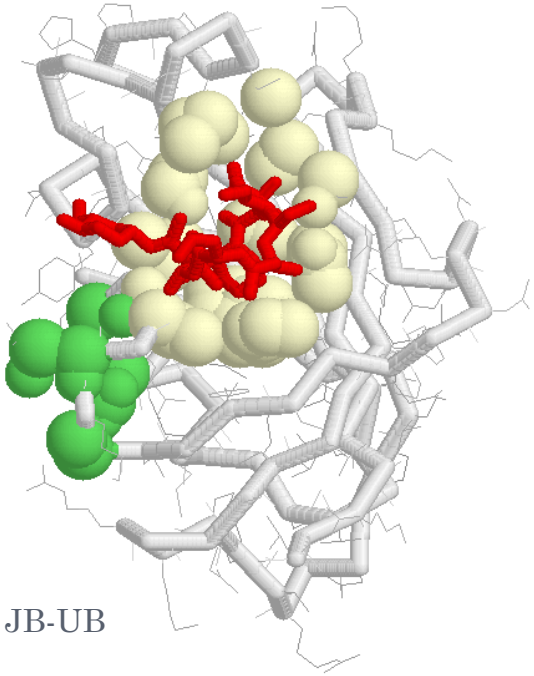
APRKFFVGGNWKMNGD
KKSLGELIHTLNGAKL
SADTEVVCGAPSIYLD
FARQKLDAKIGVAAQN
CYKVPKGAFTGEISPA
MIKDIGAAWVILGHSE
RRHVFGESDELIGQKV
AHALAEGLGVIACIGE
KLDEREAGITEKVVFE
QTKAIADNVKDWSKVV
LAYEPVWAIGTGKTAT
PQQAQEVHEKLRGWLK
SHVSDAVAQSTRIIYG
GSVTGGNCKELASQHD
VDGFLVGGASLKPEFV
DIINAKH

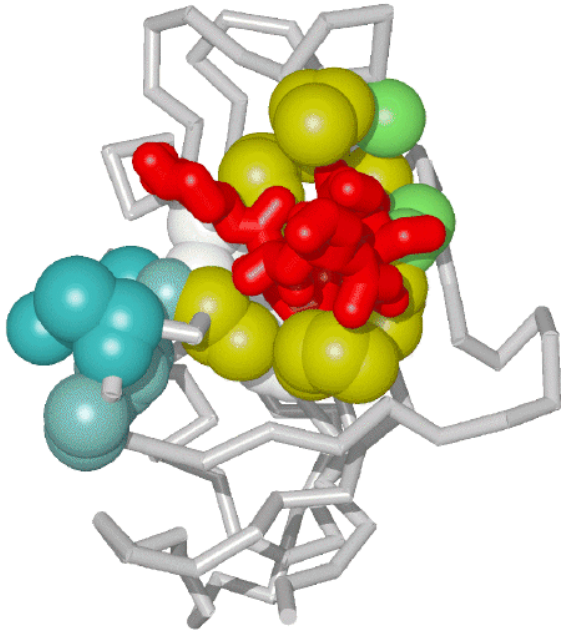
=



# Structural genomics/proteomics

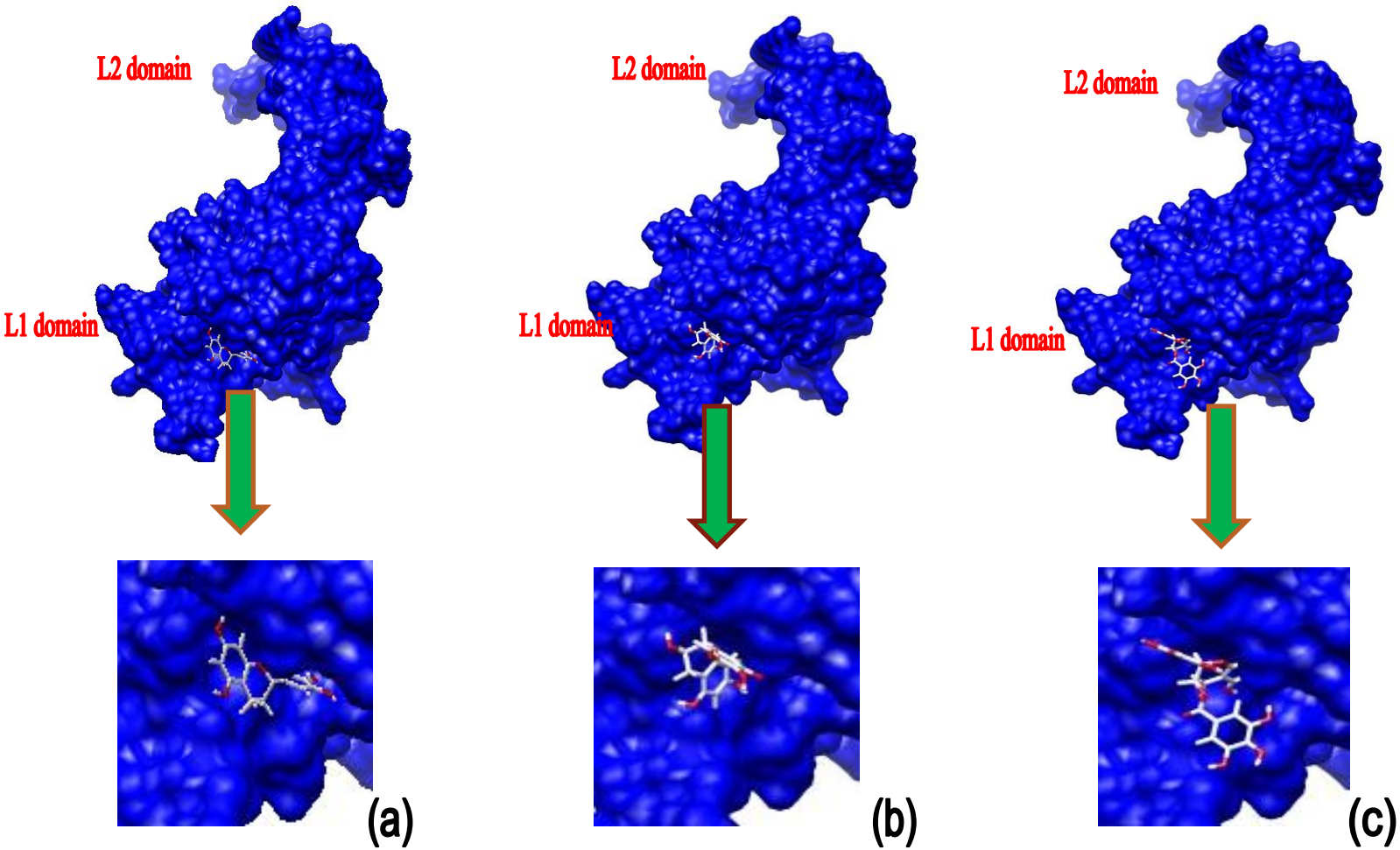
- Atlas of Topographic Surfaces of All Known Protein Structures
  - Automatic identification of binding pockets.
  - Measurement size of surface binding pockets.
- Drug Discovery
  - Quantifying ligand accessibility.
  - Constructing precise negative imprint or cast of binding site.





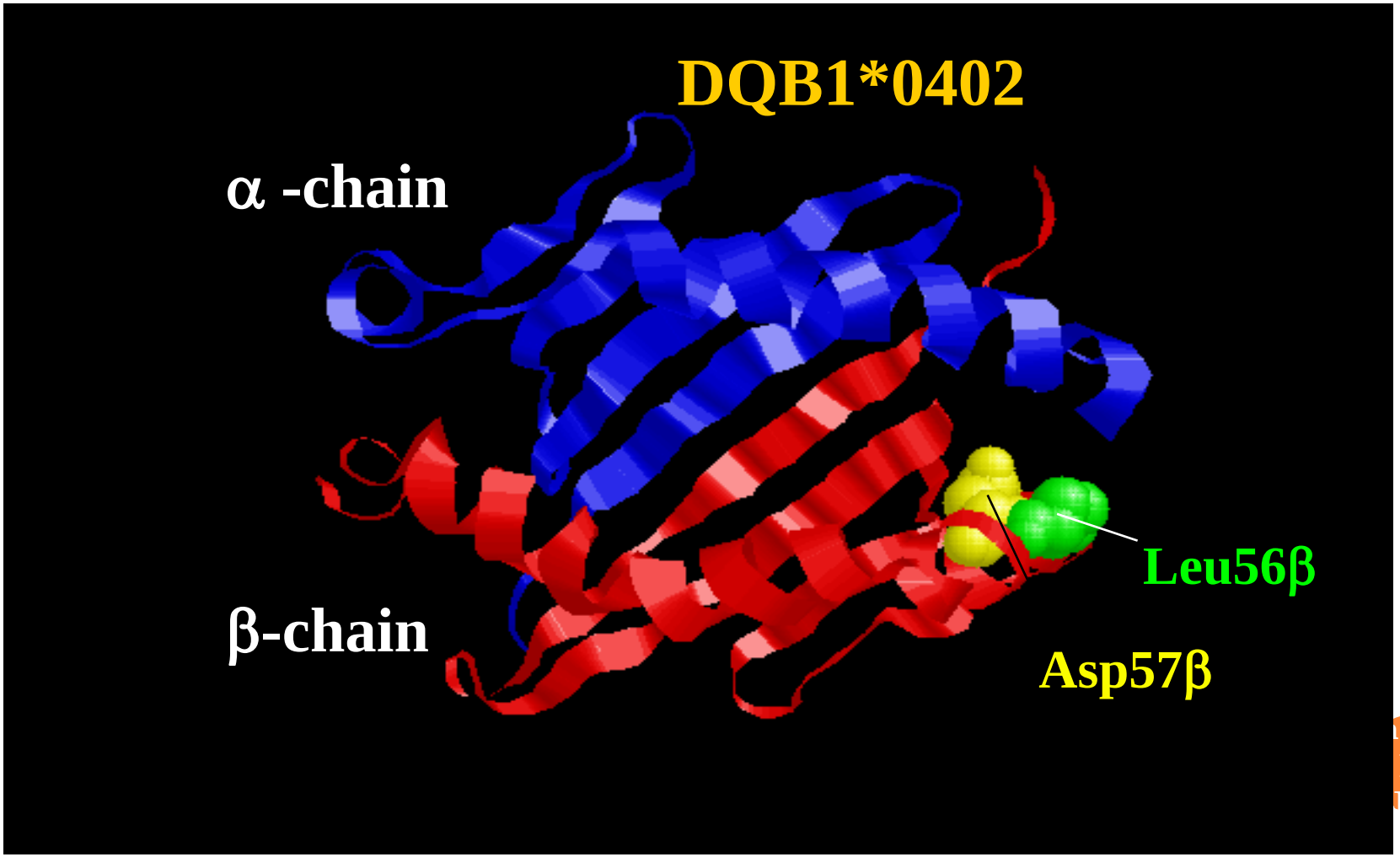
# D. DOCKING IGF-1R DENGAN EC, EGC DAN EGCG

Pembahasan




Gambar 5.6. Struktur 3D IGF-1R (Biru) berikatan dengan EC (a), EGC (b), EGCG (c).

# Protein Molecule:
# HLA-DQB1




Thank you
for
your
attention