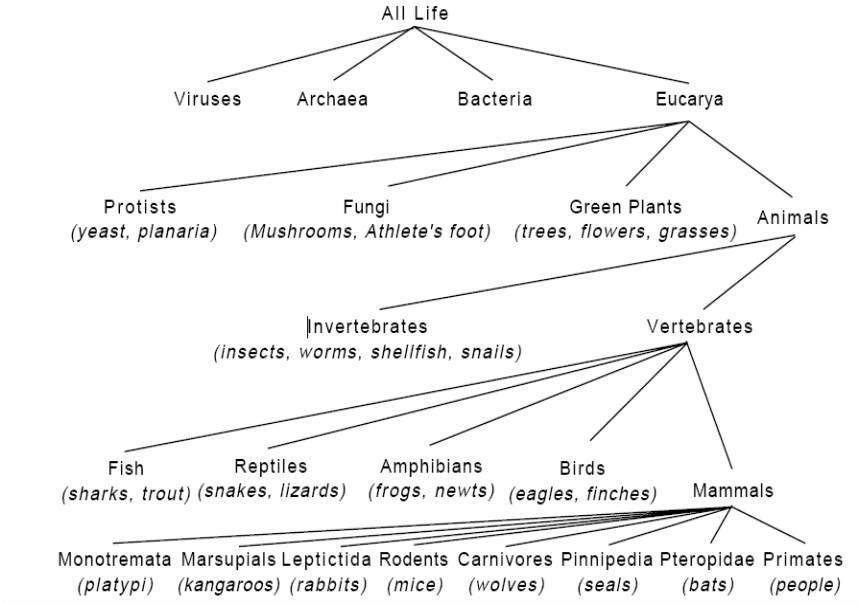
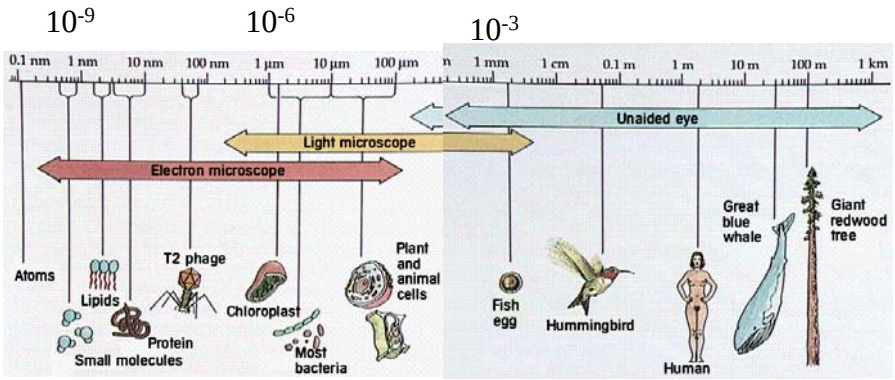
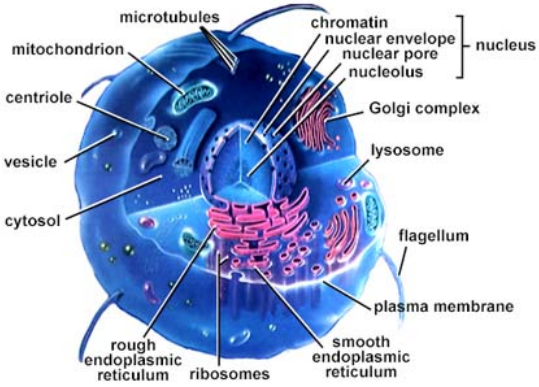
CISC 841 Bioinformatics (Spring 2006)

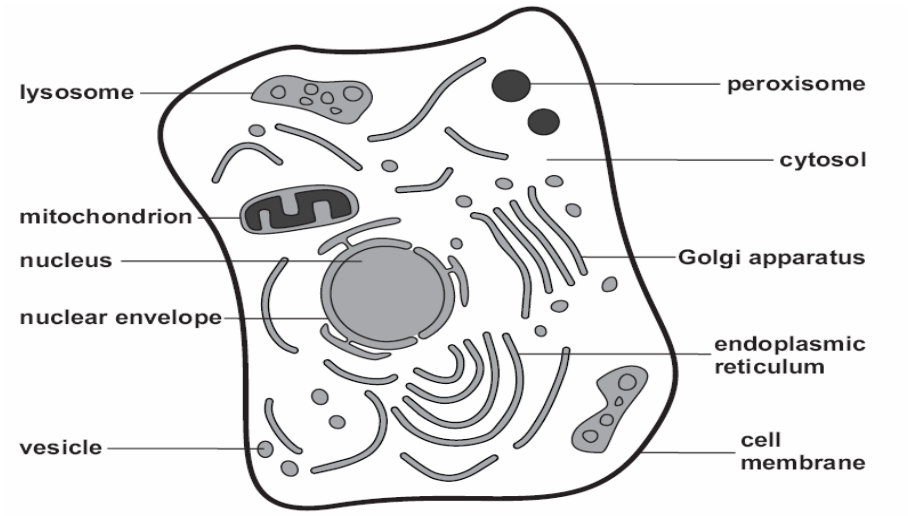
A Primer on Molecular Biology



CISC841, S06, Lec2, Liao


10^{-9}
10^{-6}
10^{-3}
0.1 nm
1 nm
10 nm
100 nm
1 μm
10 μm
100 μm
1 mm
1 cm
0.1 m
1 m
10 m
100 m
1 km
Unaided eye
Light microscope
Electron microscope
Atoms
Lipids
T2 phage
Protein
Small molecules
Chloroplast
Most bacteria
Plant and animal cells
Fish egg
Hummingbird
Human
Great blue whale
Giant redwood tree


microtubules
mitochondrion
centriole
vesicle
cytosol
chromatin
nuclear envelope
nuclear pore
nucleolus
nucleus
Golgi complex
lysosome
flagellum
plasma membrane
rough endoplasmic reticulum
ribosomes
smooth endoplasmic reticulum



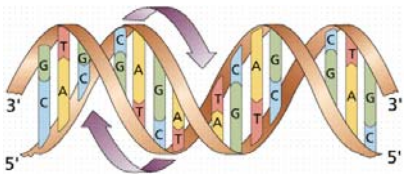
Compartment	Function(s)	Membrane
Cytosol	protein synthesis, general metabolism, etc.	single
Nucleus	■ storage of main genome (DNA molecules) ■ RNA synthesis ■ ribosome synthesis (in the nucleolus)	double
Endoplasmatic reticulum (ER) (inner space of nuclear membrane, extending throughout the cell)	■ synthesis of most lipids (membrane) ■ synthesis of proteins for single-membrane organelles (rough ER) ■ post-translational processing of those proteins	single
Golgi apparatus	■ post-translational processing of proteins ■ distribution of proteins and lipids to single-membrane organelles	single
Vesicles (mobile bubbles)	transport of proteins and membrane between single-membrane organelles and to/from cell exterior	single
Endosomes	■ contain material taken up from the exterior; or ■ secrete contents (mainly proteins) to cell exterior	single
Lysosomes/vacuoles (plants, fungi)	digest of molecules, organelles, etc. / store waste and nutrients, control cell size	single
Peroxisomes	carry out oxidative (dangerous) reactions	single
Cell exterior / extracellular matrix	■ extracellular matrix connects cells, stabilizes the organism, contains nutrients, etc. ■ in polarized cells (e.g., nerve cells), the exterior is divided into basolateral and apical parts	single
Mitochondria	generate ATP by oxidizing nutrients	double
Chloroplasts (in plants)	generate energy-rich molecules from sunlight	double

Molecule type	Cell Mass in Bacteria	Mammals
H_2O (water)	70%	70%
DNA	1%	0.25%
RNA	6%	1%
proteins	15%	18%
lipids (fat)	2%	5%
polysaccharides (sugar)	2%	2%
metabolites and inorganic ions	4%	4%

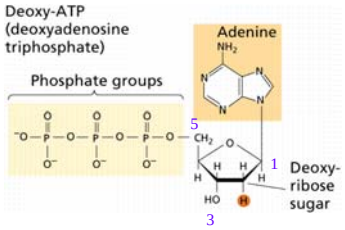
DNA (double helix, hydrogen bond, complementary bases A-T, G-C)

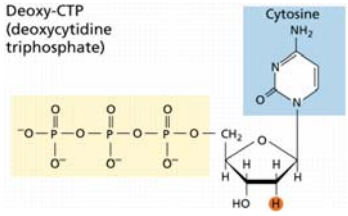
5' end phosphate group
3' end is free
1' position is attached with the base

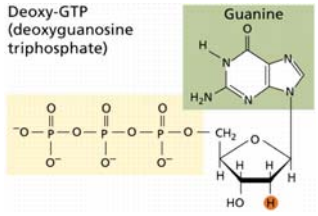
double strand DNA sequences form a helix via hydrogen bonds between complementary bases
hydrogen bond:
- weak: about 3~5 kJ/mol (A covalent C-C bond has 380 kJ/mol), will break when heated
- saturation:
- spefic:



Structure of the bases (Thymine is not shown here)







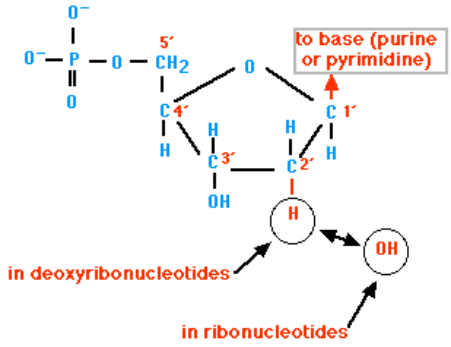
- Purines:A and G
- Pyrimidines: C and T
- Oligonucleotide: a DNA of a few tens of nucleotides
- ATP, ADP, AMP

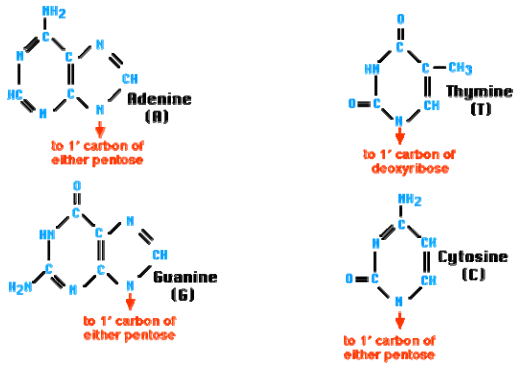
Small molecules:

> sugar: carbohydrate

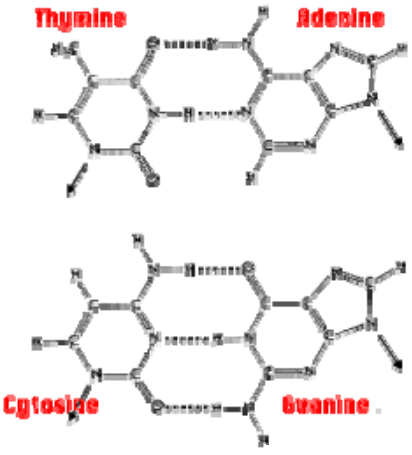
> fatty acids

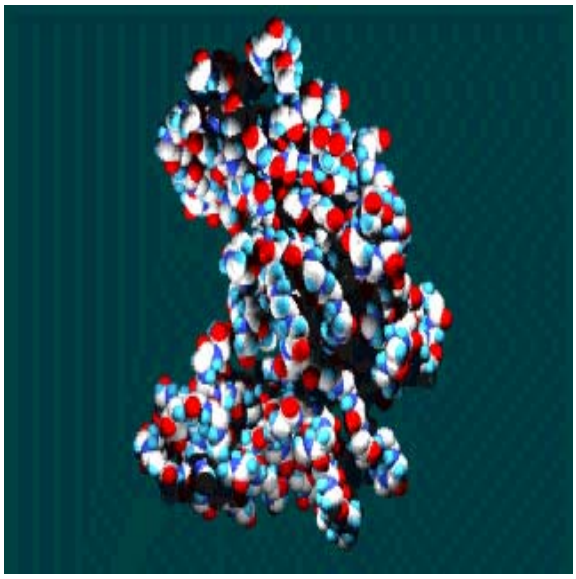
> nucleotides: A, C, G, T (Purines: A and G; Pyrimidines: C and T)





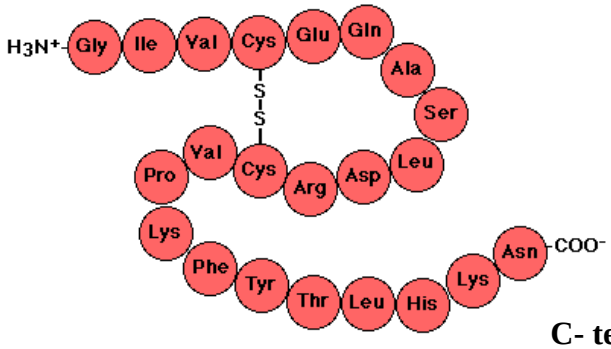
The rules for base pairing (**Watson-Crick** base pairing) :
A with **T**: the purine **adenine** (A) always pairs with the pyrimidine **thymine** (T)
C with **G**: the pyrimidine **cytosine** (C) always pairs with the purine **guanine** (G)





Polypeptide

N-terminal

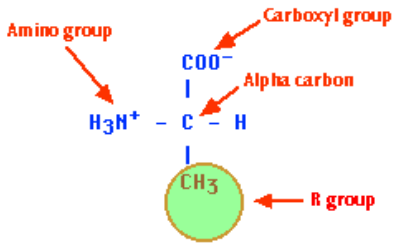


C- terminal

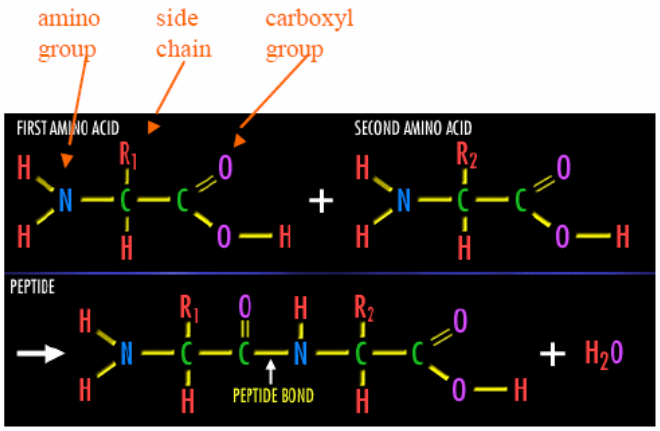
The Amino Acids

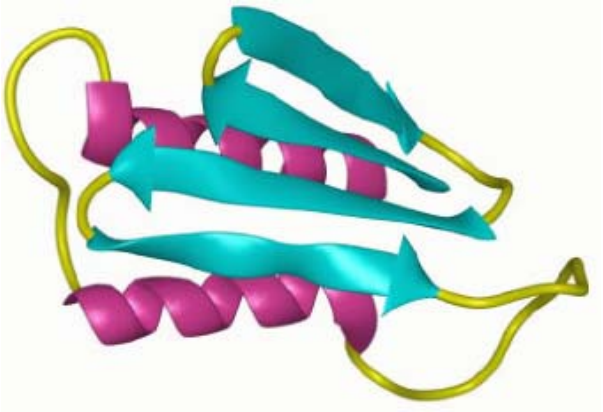
(For each amino acid, both the three-letter and single-letter codes are given. CLICK the NAME to see the structural formula)

Alanine	Ala	A	hydrophobic
Arginine	Arg	R	free amino group makes it basic and hydrophilic
Asparagine	Asn	N	carbohydrate can be covalently linked ("N-linked) to its -NH
Aspartic acid	Asp	D	free carboxyl group makes it acidic and hydrophilic
Cysteine	Cys	C	oxidation of their sulfhydryl (-SH) groups link 2 Cys (S-S)
Glutamic acid	Glu	E	free carboxyl group makes it acidic and hydrophilic
Glutamine	Gln	Q	moderately hydrophilic
Glycine	Gly	G	so small it is amphiphilic (can exist in any surroundings)
Histidine	His	H	basic and hydrophilic
Isoleucine	Ile	I	hydrophobic
Leucine	Leu	L	hydrophobic
Lysine	Lys	K	strongly basic and hydrophilic
Methionine	Met	M	hydrophobic
Phenylalanine	Phe	F	very hydrophobic
Proline	Pro	P	causes kinks in the chain
Serine	Ser	S	carbohydrate can be covalently linked ("O-linked") to its -OH
Threonine	Thr	T	carbohydrate can be covalently linked ("O-linked") to its -OH
Tryptophan	Trp	W	scarce in most plant proteins
Tyrosine	Tyr	Y	a phosphate or sulfate group can be covalently attached to its -OH
Valine	Val	V	hydrophobic

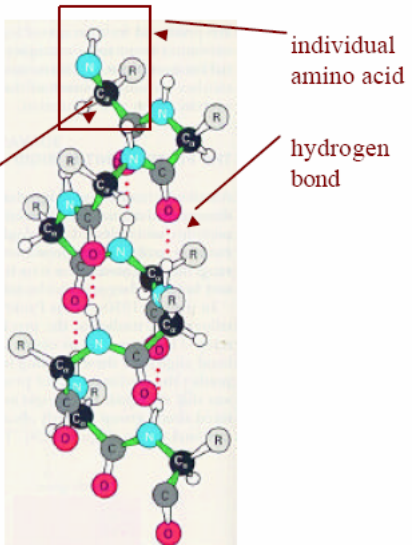


Peptide Bonds



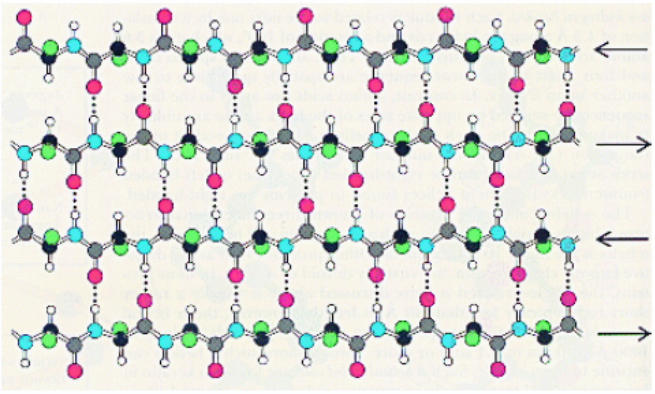


α Helices



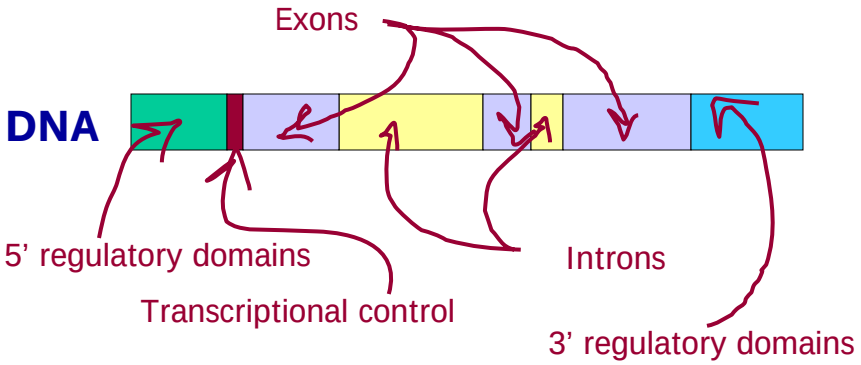
- Helix complete turn every 3.6 AAs
- Hydrogen bond between (-C=O) of one AA and (-N-H) of its 4th neighboring AA

β Strands



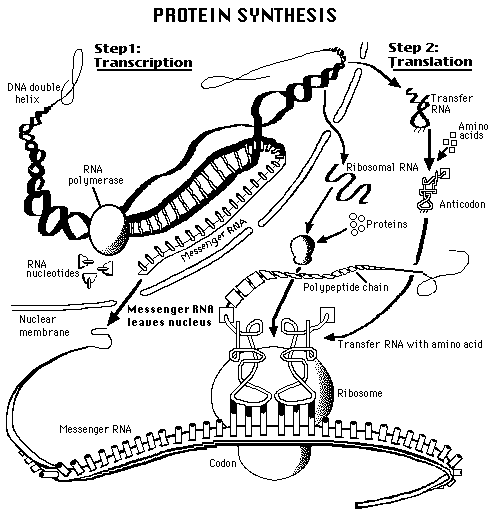
Hydrogen bond b/w carbonyl oxygen atom on one chain and NH group on the adjacent chain

Gene Structure

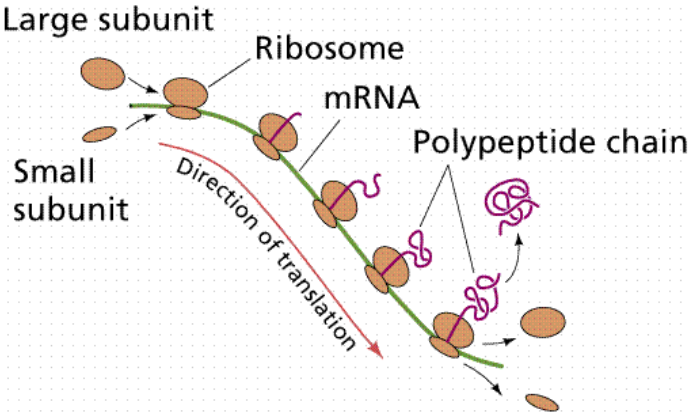


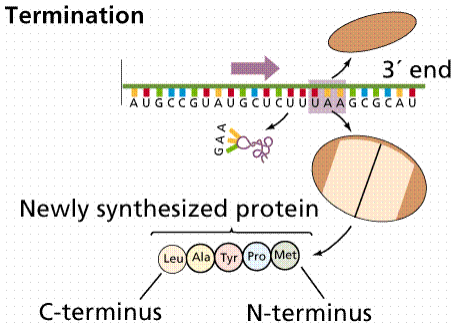
Post-transcriptional processing: hnRNA to mRNA

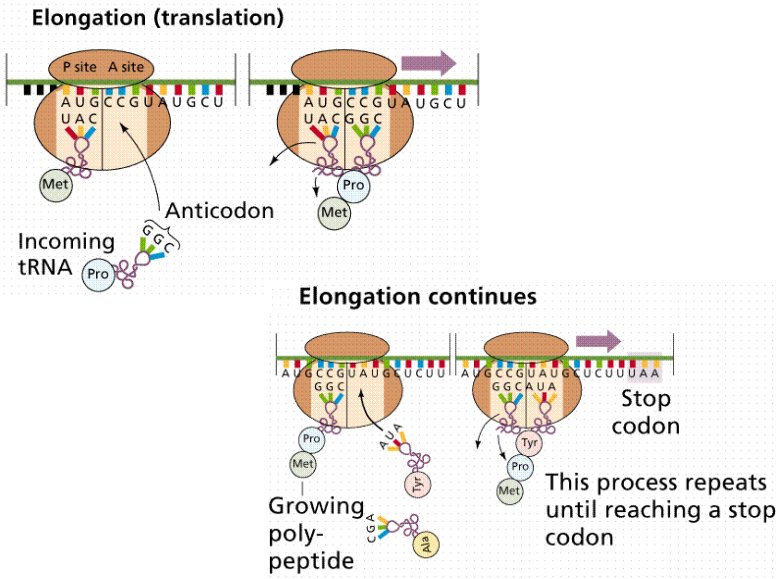
Translation: mRNA to protein

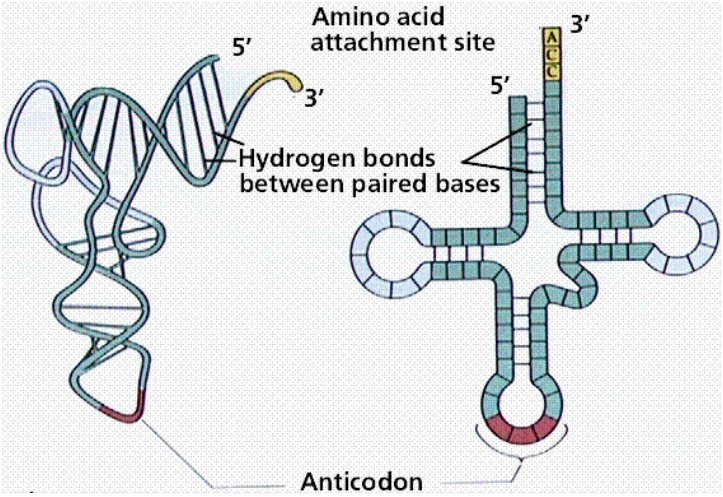


CISC841, S06, Lec2, Liao

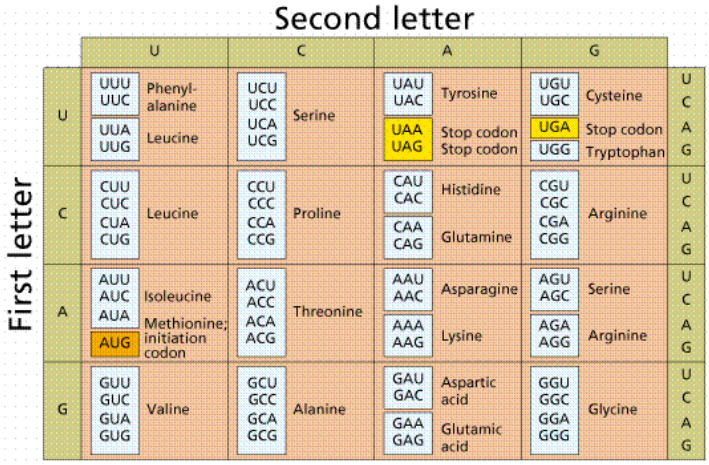

Large subunit
Ribosome
mRNA
Polypeptide chain
Small subunit
Direction of translation


Termination
3´ end
AUGCCGUAUGCUCUUUAAGCGCAU
GAA
Newly synthesized protein
Leu
Ala
Tyr
Pro
Met
C-terminus
N-terminus


Elongation (translation)
P site
A site
Met
Anticodon
Incoming tRNA
Pro
Elongation continues
Stop codon
Growing poly-peptide
Tyr
Ala
This process repeats until reaching a stop codon

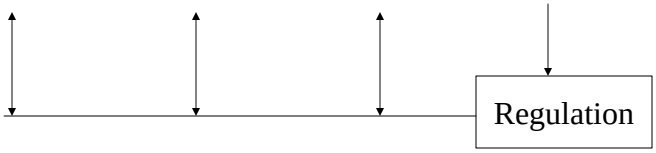

Amino acid attachment site
5′
3′
Hydrogen bonds between paired bases
Anticodon

Genetic Code: codons

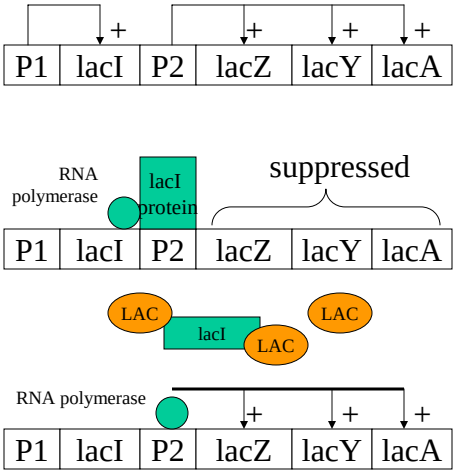


Regulation

Genes (DNA) → Message (RNA) → Proteins → Function/ Environment ↔ Other Cells



Operon



lac operon on *E. coli*

Repressor protein coded by lacI, bind to P2 preventing transcription of lacZ, lacY and lacA

Lactose binds with lacI, allowing RNA polymerase to bind to P2 and transcribe the structural genes

Boolean Networks: An example

X0	X1	X2	X3	
1	1	1	0	P0
-	1	0	1	P1
1	-	0	0	P2
1	1	-	1	P3
1	1	1	+	P4

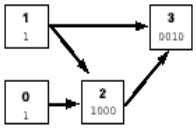
1: induced

0: suppressed

-: forced low

+: forced high

Interpreting data ⇕ Reverse Engineering



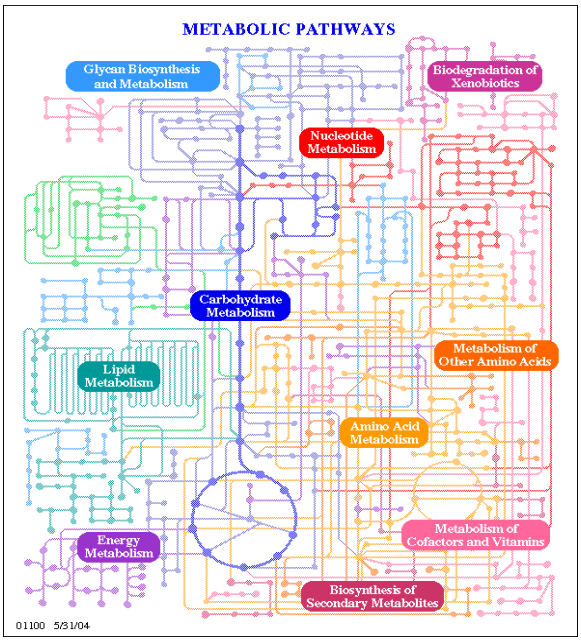
A *A directed graph structure with numbered nodes connected by edges*

x1	1	0	1	0
x2	1	1	0	0
x3	**0**	**0**	**1**	**0**

B *The truth table (shown for node 3 only)*

```
x0 := 1
x1 := 1
x2 := x0 and x1
x3 := x1 and not x2
```

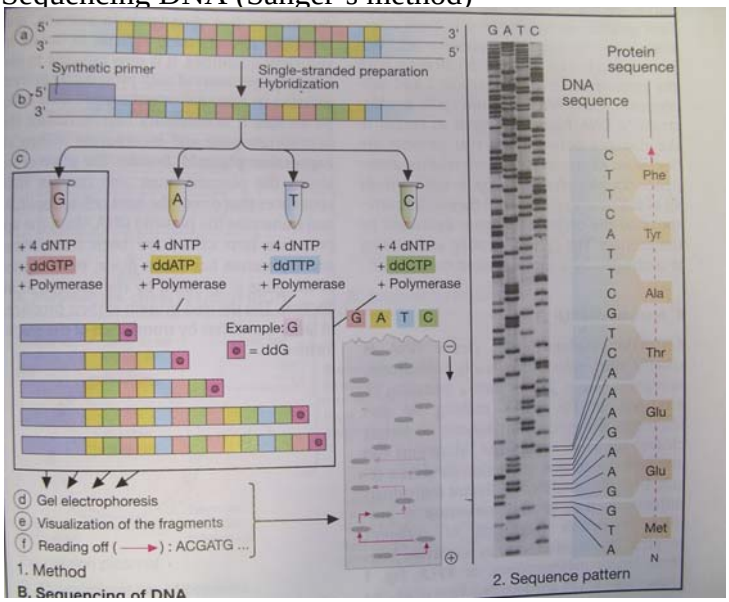
C *The logic equations for each node*



High throughput technologies

- Genome sequencing
- DNA microarray
- Yeast 2 hybrid system

Sequencing DNA (Sanger's method)

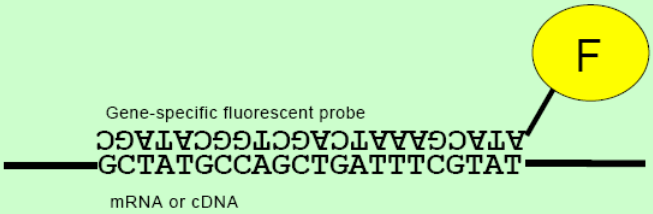


CISC841, S06, Lec2, Liao

Courtesy of Color Atlas of Biochemistry

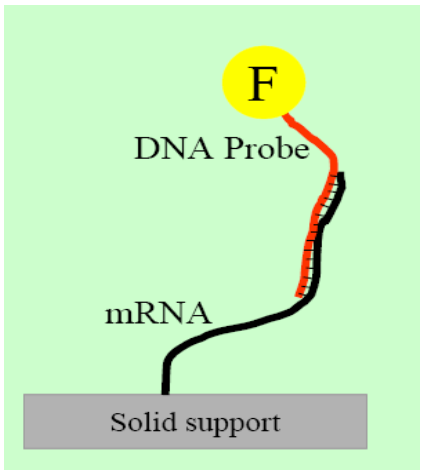
Hybridization-based methods

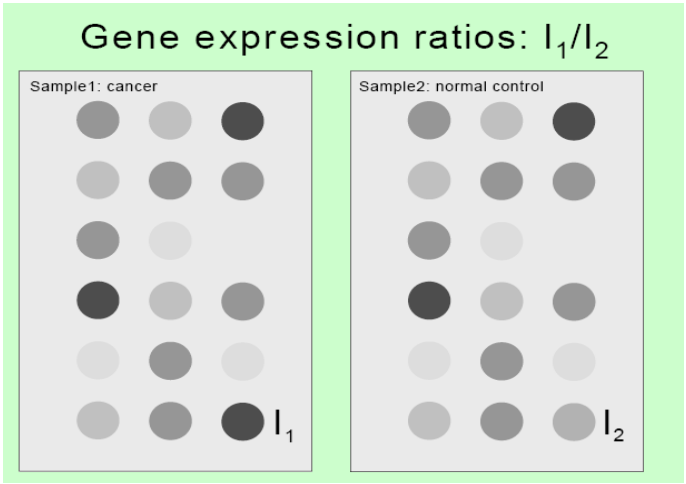
A cell may express perhaps 15,000 mRNA molecules - to measure one at a time we use probes which are sequence-specific

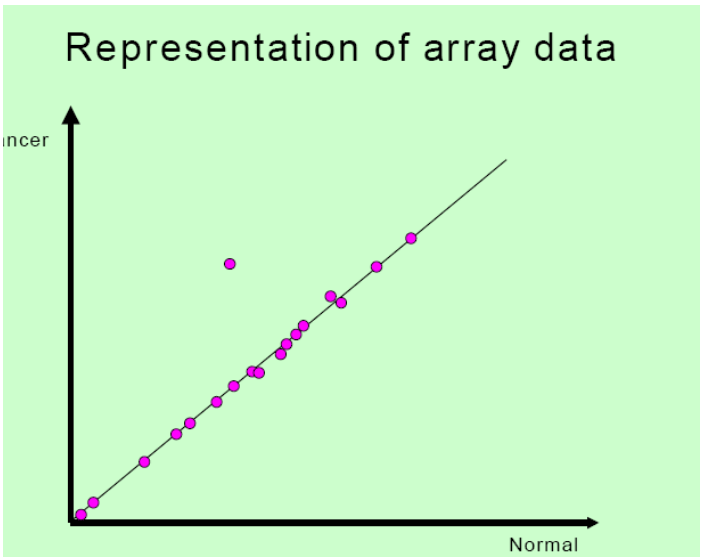


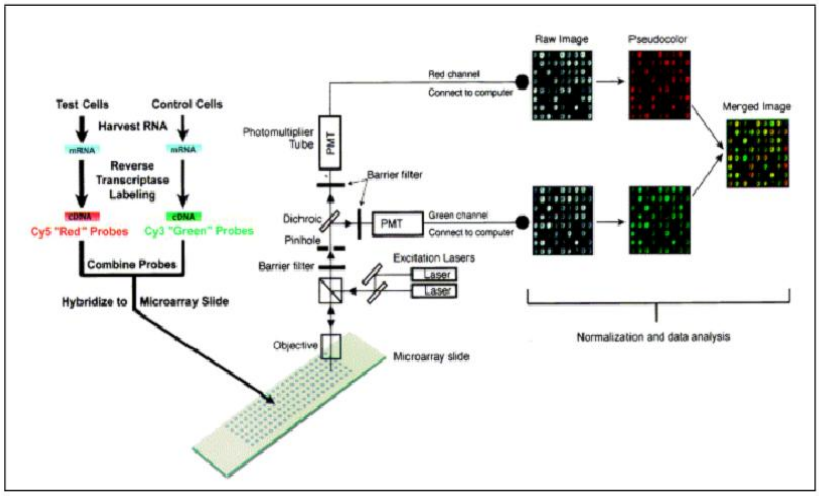
- Watson-Crick base pairs: G-C, A-T

CISC841, S06, Lec2, Liao

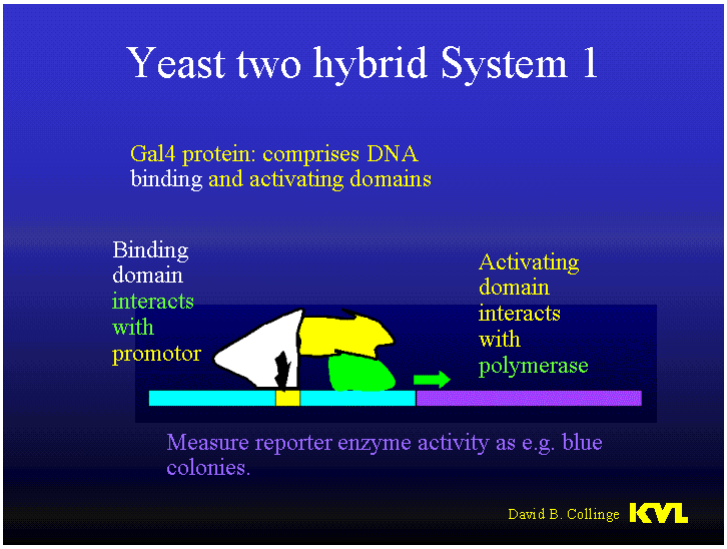

F
DNA Probe
mRNA
Solid support


Gene expression ratios: I_1/I_2
Sample1: cancer
Sample2: normal control
I_1
I_2


Representation of array data
ıncer
Normal


Test Cells
Control Cells
Harvest RNA
mRNA
mRNA
Reverse Transcriptase Labeling
cDNA
cDNA
Cy5 "Red" Probes
Cy3 "Green" Probes
Combine Probes
Hybridize to Microarray Slide
Photomultiplier Tube
PMT
Barrier filter
Dichroic
PMT
Pinhole
Barrier filter
Excitation Lasers
Laser
Laser
Objective
Microarray slide
Red channel
Connect to computer
Green channel
Connect to computer
Raw Image
Pseudocolor
Merged Image
Normalization and data analysis


JOURNAL OF BACTERIOLOGY
Journal of Bacteriology
AMERICAN SOCIETY FOR MICROBIOLOGY
AUGUST 2000, VOLUME 182, NUMBER 16
Published Twice Monthly

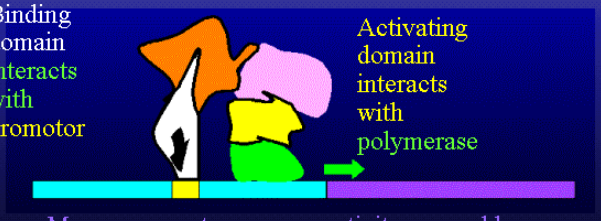

Yeast two hybrid System 1
Gal4 protein: comprises DNA binding and activating domains
Binding domain interacts with promotor
Activating domain interacts with polymerase
Measure reporter enzyme activity as e.g. blue colonies.
David B. Collinge
KVL

Yeast two hybrid System 2

- Gal4 protein: the two domains of the protein do not need to be transcribed in a single protein
- Just as long as they come to interact

Two other protein domains interact

Binding domain interacts with promotor



Activating domain interacts with polymerase

Measure reporter enzyme activity as e.g. blue colonies.

David B. Collinge KVL

CISC841, S06, Lec2, Liao

Yeast two hybrid System (3)

- This is achieved using gene fusions:
- Plasmids carrying different constructs can be expressed in yeast.

Binding domain as a translational fusion with the gene encoding another protein in one plasmid

Activating domain as a translational fusion with the gene encoding a third protein in a second plasmid

David B. Collinge KVL