Bioinformatics
(Globex, Summer2015)

Molecular Biology
A Primer

- **What is Life?**

  - Three kingdoms
  - The Cell theory

- **Central Dogma**

  - Genetic code
  - Transcription
  - Translation

- **Molecular biology tools**

  - Clone
  - PCR
  - Sequencing
  - Microarray
  - Yeast 2 hybrid

Organisms: three kindoms of life -- eukaryotes, eubacteria, and archea

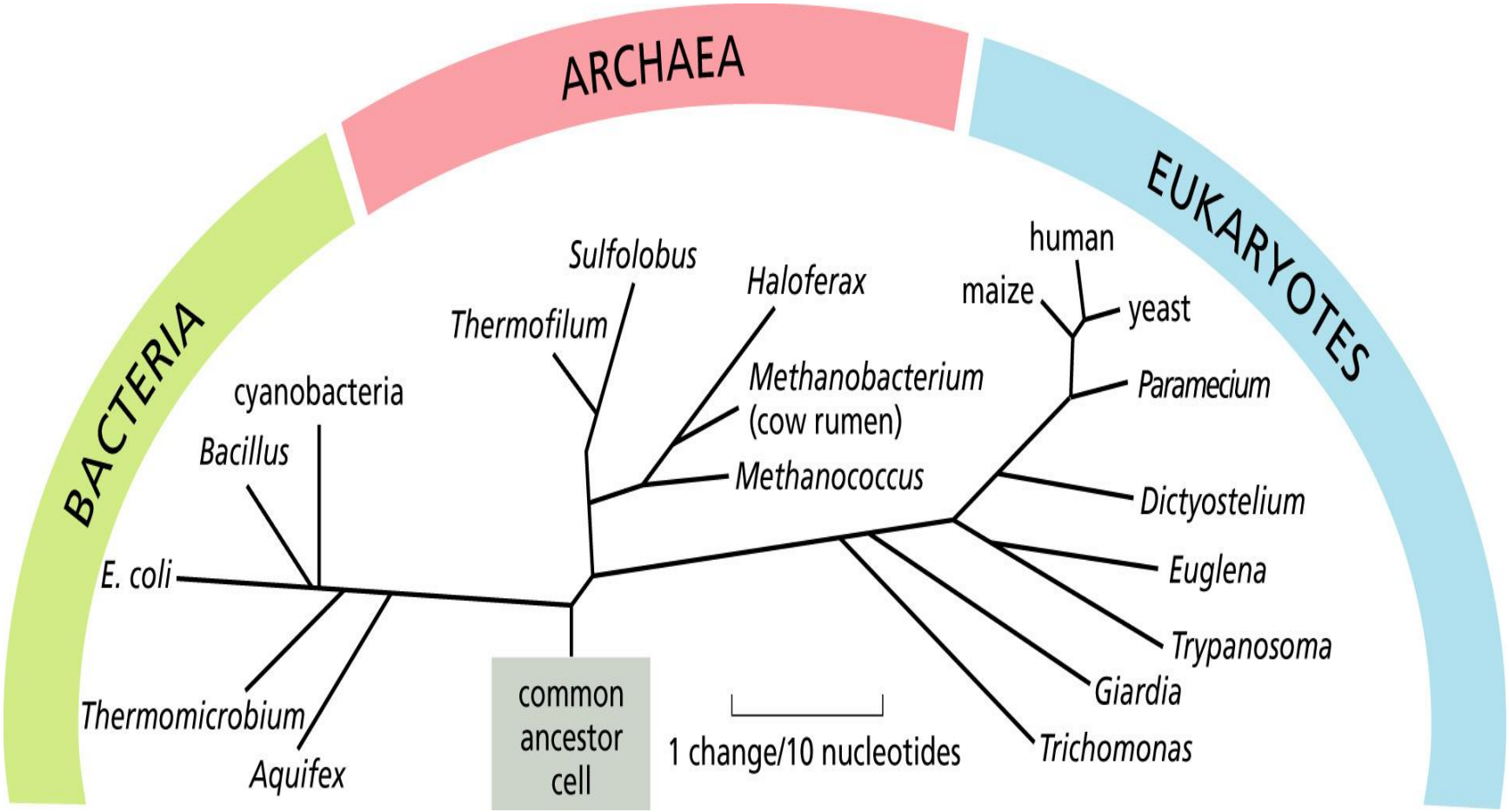
- Observation: a lot of living things
- Why does Mother nature have this biodiversity?
- Answers
  - Simple classification based on morphological features
  - Theory: evolution – mutations, natural selection, …
- Tree of life
  - NCBI Taxonomy
    http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy

Model organisms: E. coli, Drosophila, C. elegans, Yeast, Arabidopsis, Mouse, …

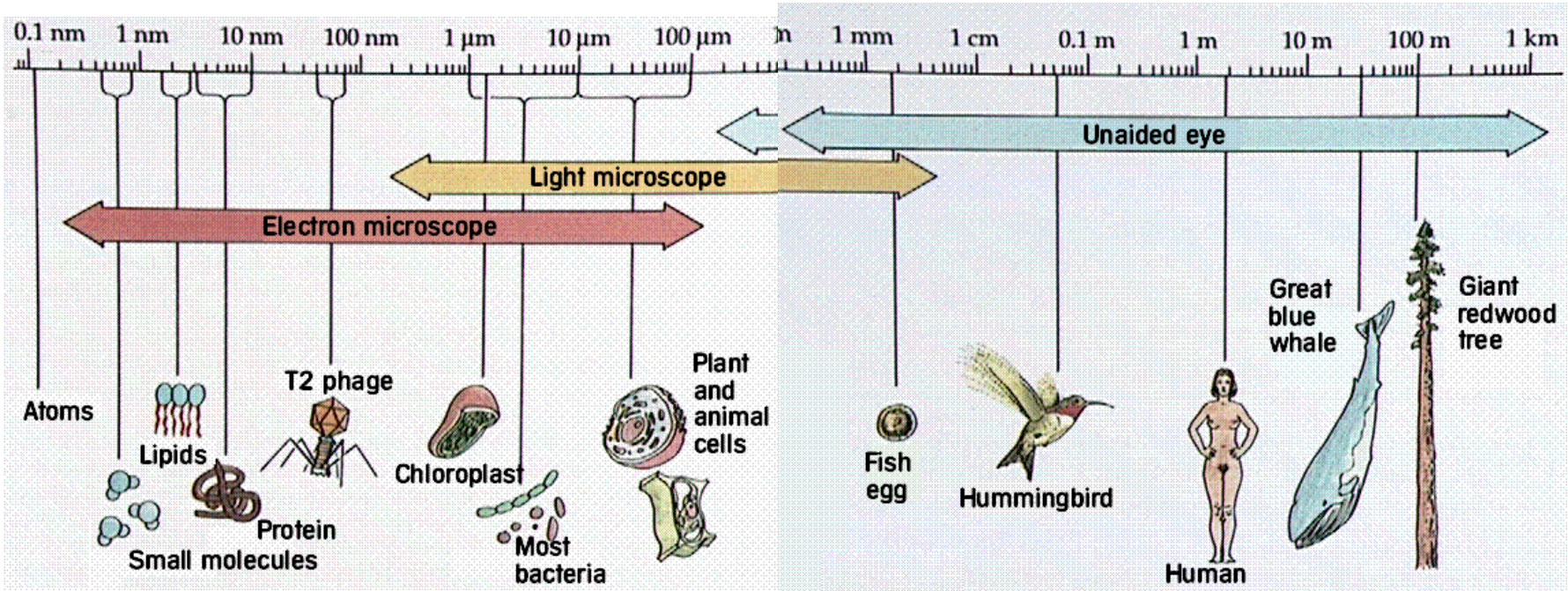
Cell: the basic unit of life

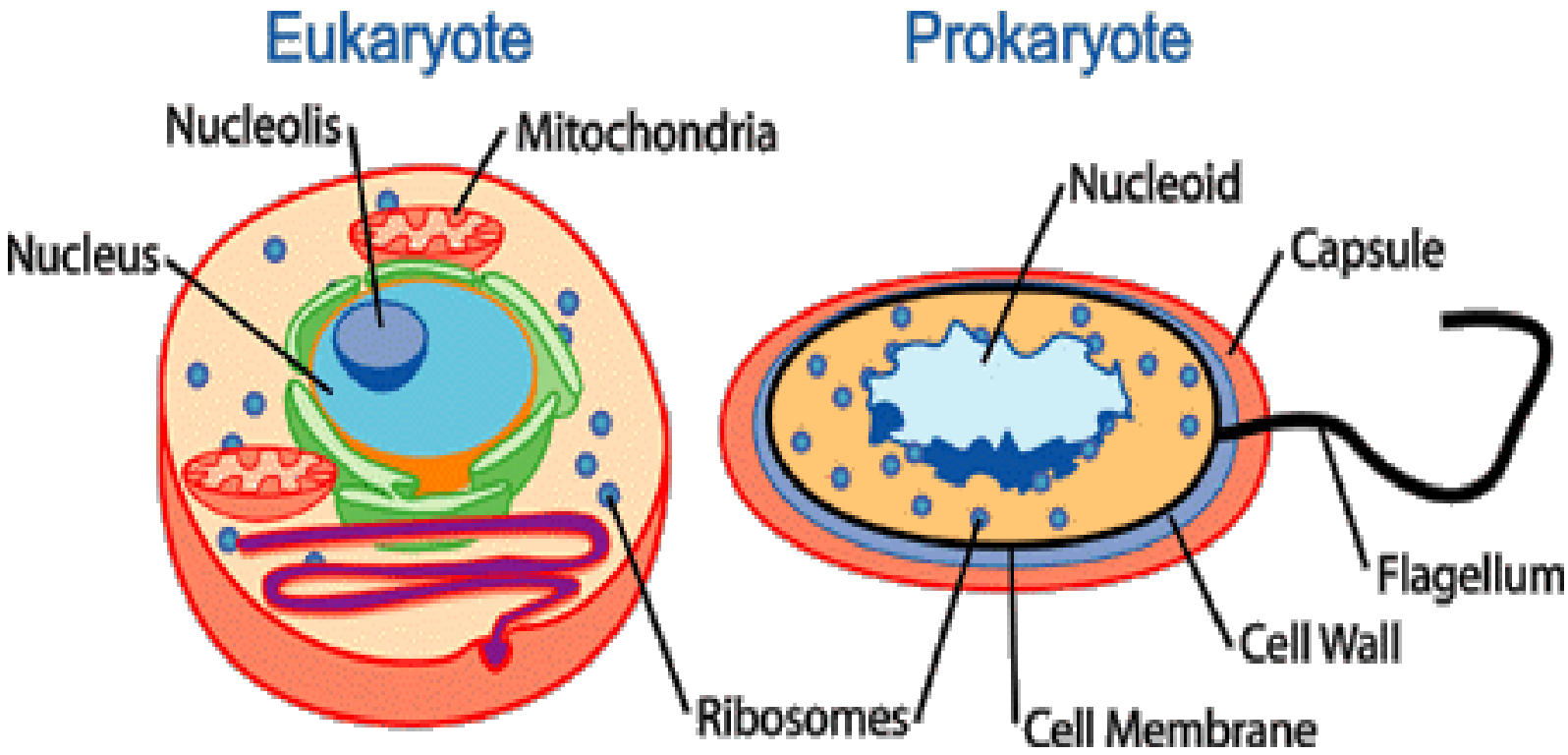
- Every living thing is made of cells.
- Every cell comes from a pre-existing cell.

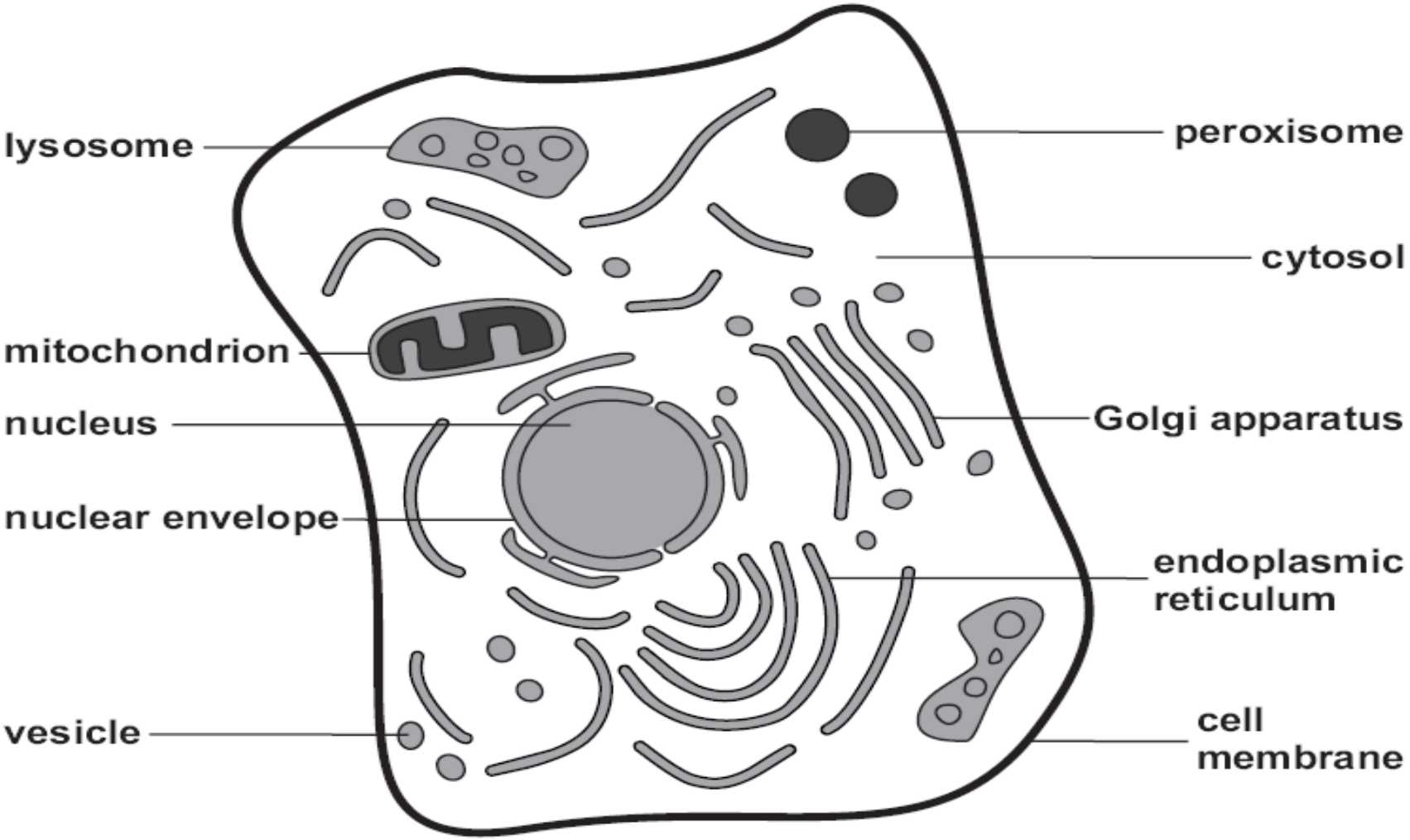
Tree of Life


ARCHAEA
BACTERIA
EUKARYOTES
Sulfolobus
Haloferax
Thermofilum
human
maize
yeast
Paramecium
Methanobacterium
(cow rumen)
cyanobacteria
Bacillus
Methanococcus
Dictyostelium
E. coli
Euglena
Trypanosoma
Giardia
Thermomicrobium
Aquifex
common
ancestor
cell
1 change/10 nucleotides
Trichomonas

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
Small molecules
T2 phage
Protein
Chloroplast
Most bacteria
Plant and animal cells
Fish egg
Hummingbird
Human
Great blue whale
Giant redwood tree


Eukaryote
Prokaryote
Nucleolis
Mitochondria
Nucleus
Ribosomes
Nucleoid
Capsule
Flagellum
Cell Wall
Cell Membrane


lysosome
peroxisome
cytosol
mitochondrion
nucleus
Golgi apparatus
nuclear envelope
endoplasmic reticulum
vesicle
cell membrane

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	Cell Mass in 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%

Macro-molecule	DNA	RNA
Building blocks	nucleotides (A,C,G,T)	nucleotides (A,C,G,U)
Typical length	1000s to 10^9s	100s to 1000s
Structure	double helix, tightly packed and organized in several levels	complex 3D structure, with structural motifs (secondary structure)
Function	storage of (most of) the hereditary information of an organism: the genome, which contains the genes as subsequences	■ *messenger RNA (mRNA)*: serves as the blueprint for protein production ■ *transfer RNA (tRNA)*: connects codons to amino acids (implementing the genetic code); used by the ribosome ■ *ribosomal RNA (rRNA)*: forms part of the ribosome (amounting to ∼90% of the total RNA)
Location	nucleus, mitochondria, chloroplasts	nucleus, cytosol, mitochondria, chloroplasts

Chromosome (DNA)

> circular, also called plasmid when small (for bacteria)
> linear (for eukaryotes)

Genes: segments on DNA that contain the instructions for organism's structure and function

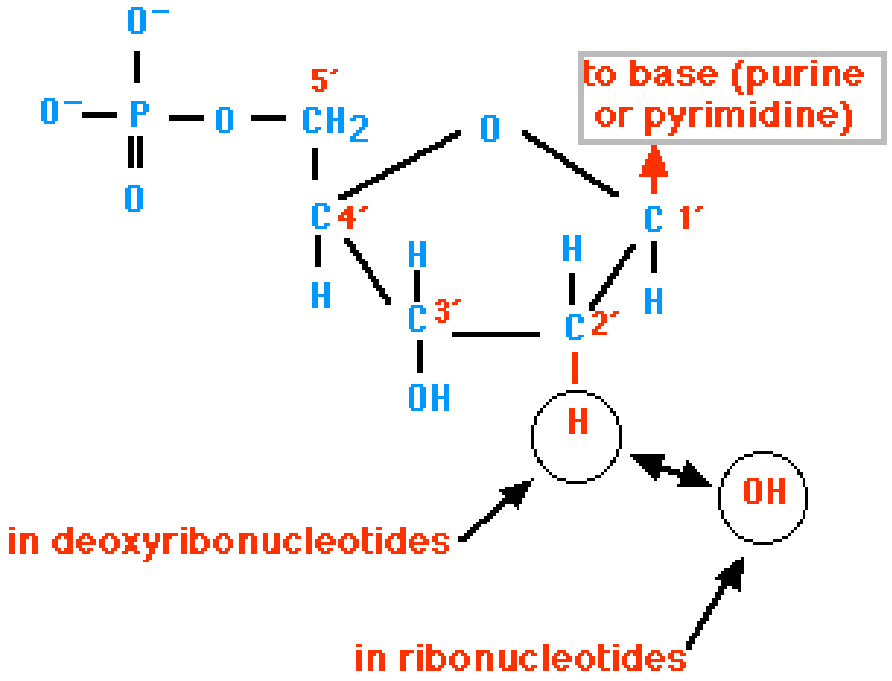
Proteins: the workhorse for the cell.
> establishment and maintenance of structure
> transport. e.g., hemoglobin, and integral transmembrane proteins
> protection and defense. e.g., immunoglobin G
> Control and regulation. e.g., receptors, and DNA binding proteins
> Catalysis. e.g., enzymes

Small molecules:

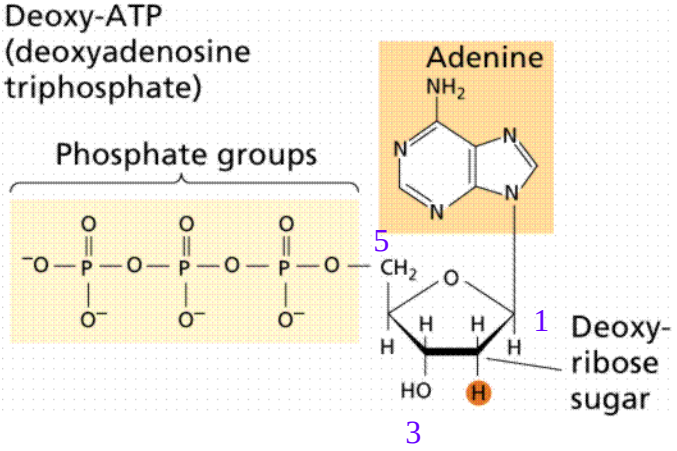
> sugar: carbohydrate

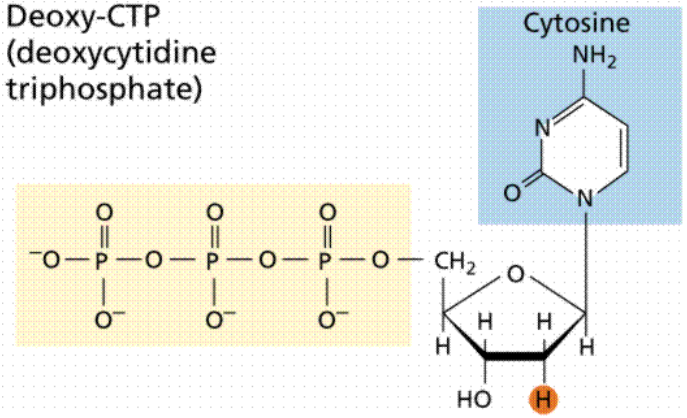
> fatty acids

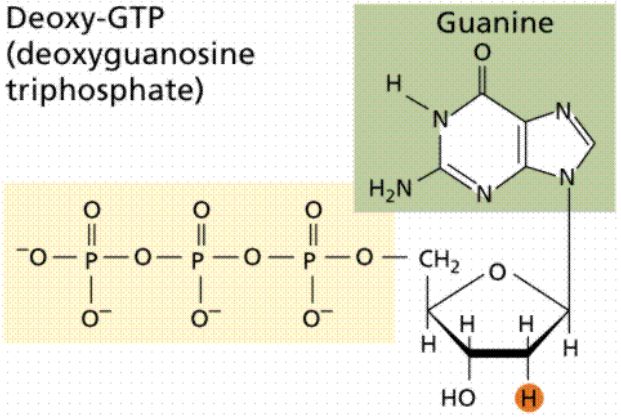
> nucleotides: A, C, G, T (Purines: A and G; Pyrimidines: C and T)



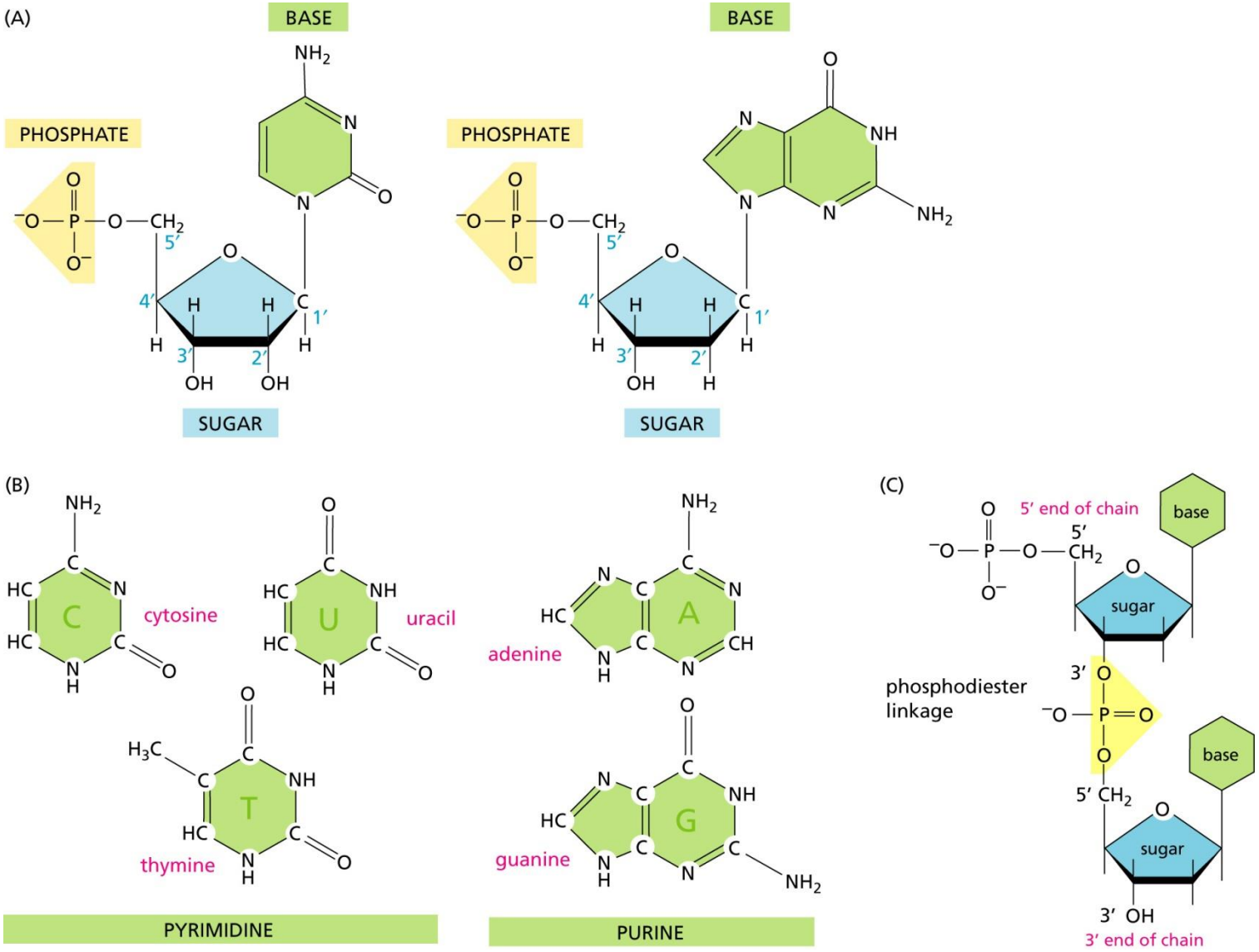
Structure of the bases (Thymine is not shown here)

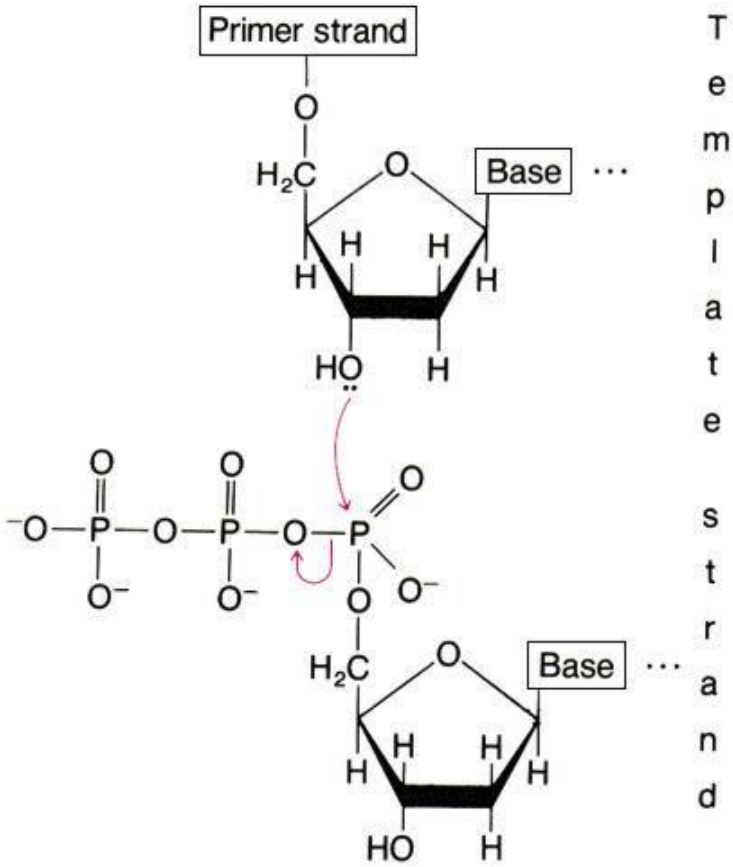






- Purines: A and G
- Pyrimidines: C and T
- Oligonucleotide: a DNA of a few tens of nucleotides
- ATP, ADP, AMP


(A)
BASE
PHOSPHATE
SUGAR
BASE
PHOSPHATE
SUGAR
(B)
cytosine
uracil
thymine
adenine
guanine
PYRIMIDINE
PURINE
(C)
5′ end of chain
base
sugar
phosphodiester linkage
base
sugar
3′ end of chain


Primer strand
Base
Base
Template strand

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

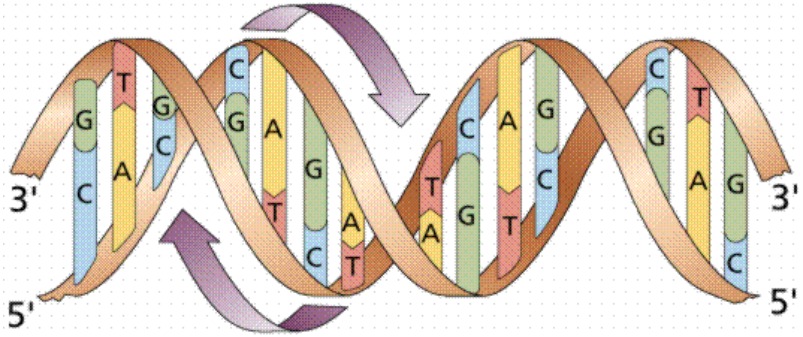
double strand DNA sequences form a helix via hydrogen bonds between complementary bases

hydrogen bond:

- weak: about 3~5 kJ/mol (A covalent C-C bond has 380 kJ/mol), will break when heated

- saturation:

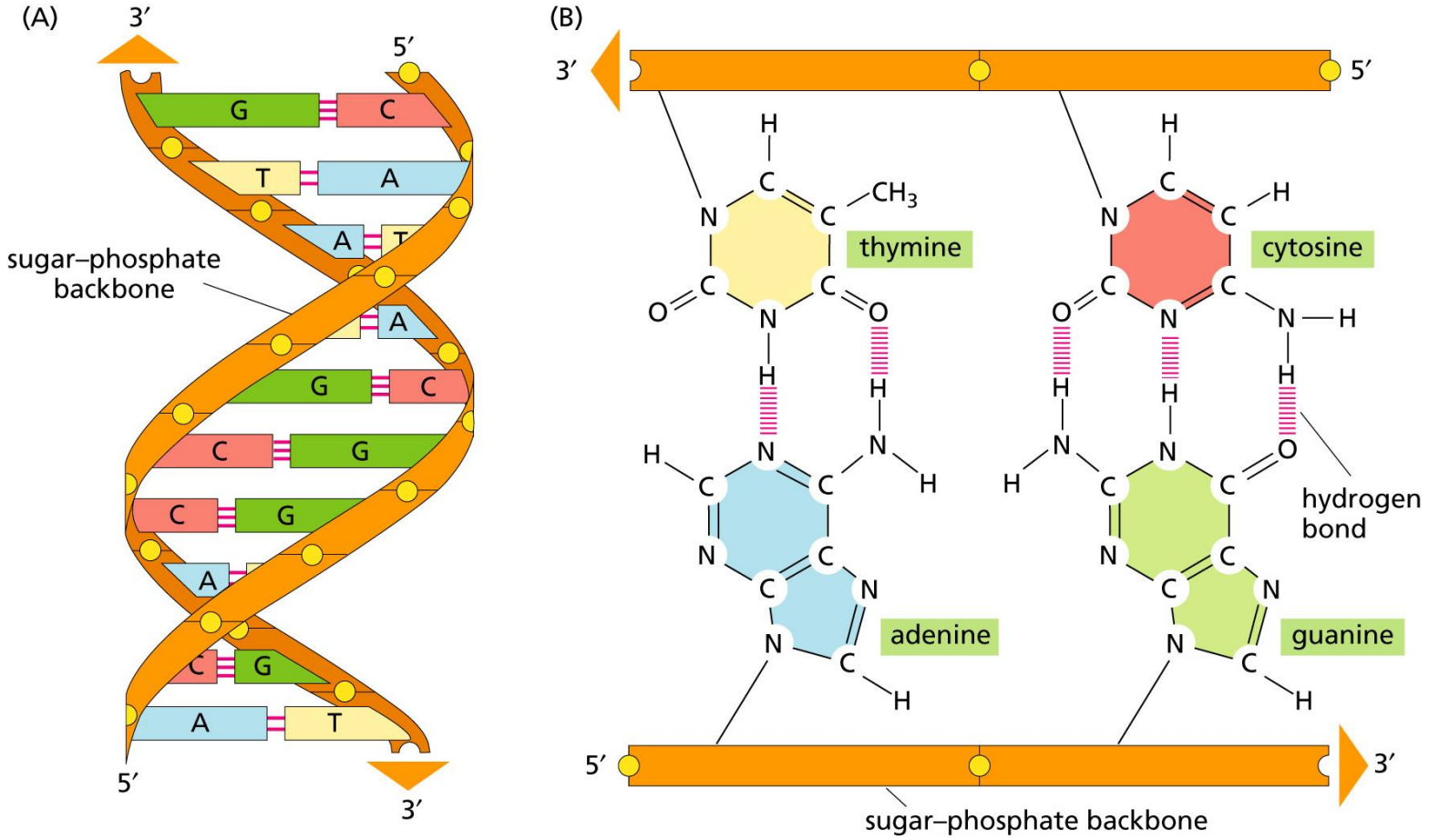
- specific:



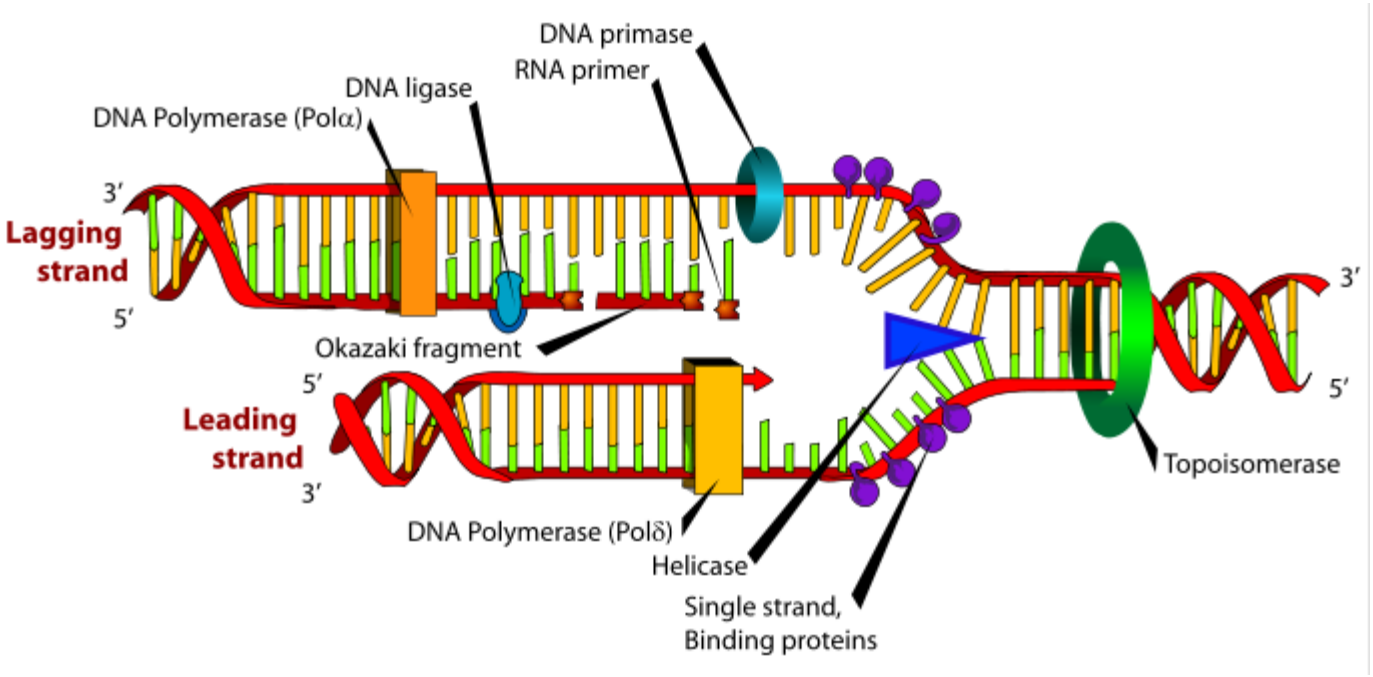
The rules for base pairing (**Watson-Crick** base pairing) :

A with **T**: the purine **adenine** (A) always pairs with the pyrimidine **thymine** (T)

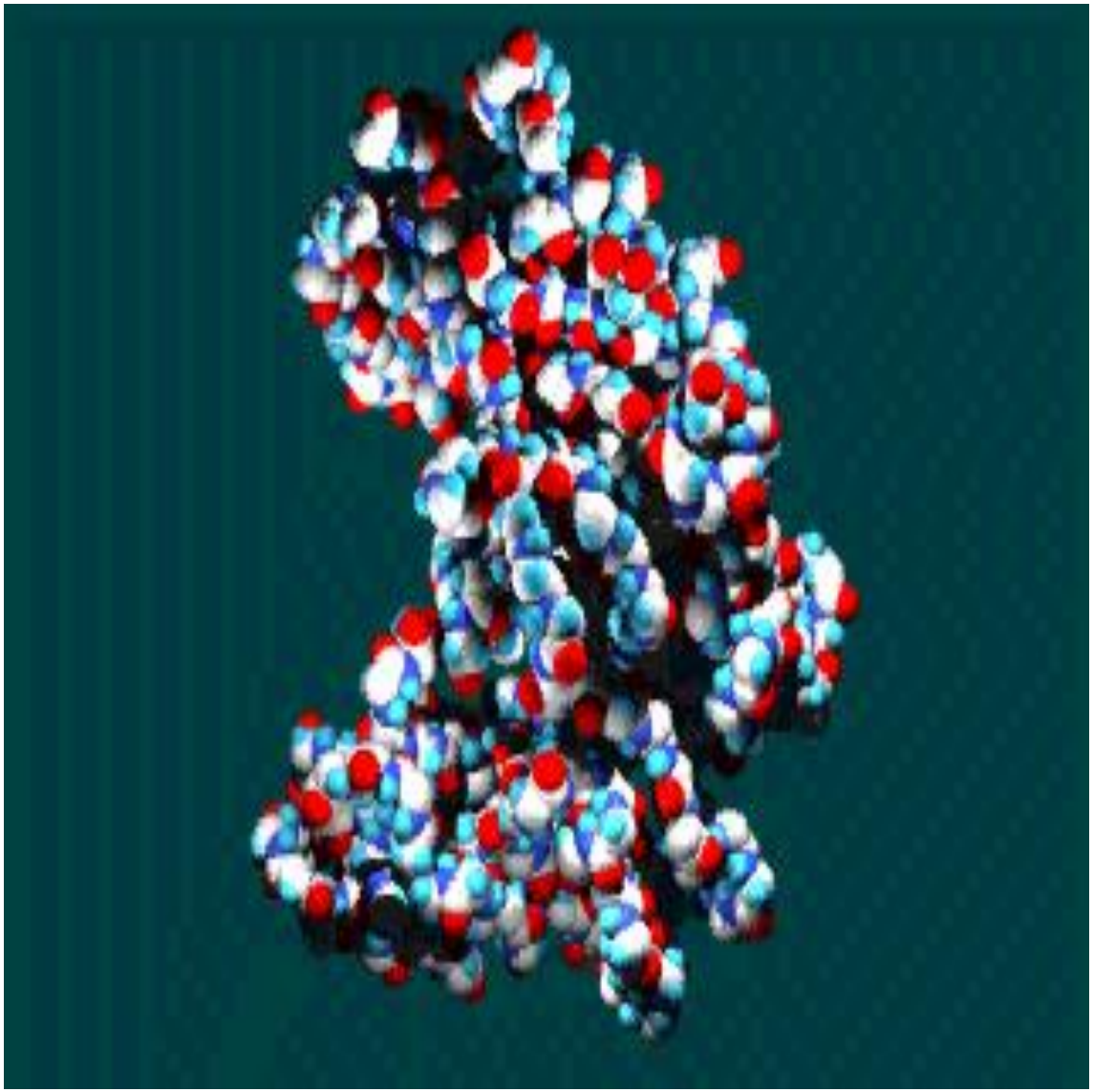
C with **G**: the pyrimidine **cytosine** (C) always pairs with the purine **guanine** (G)



DNA replication



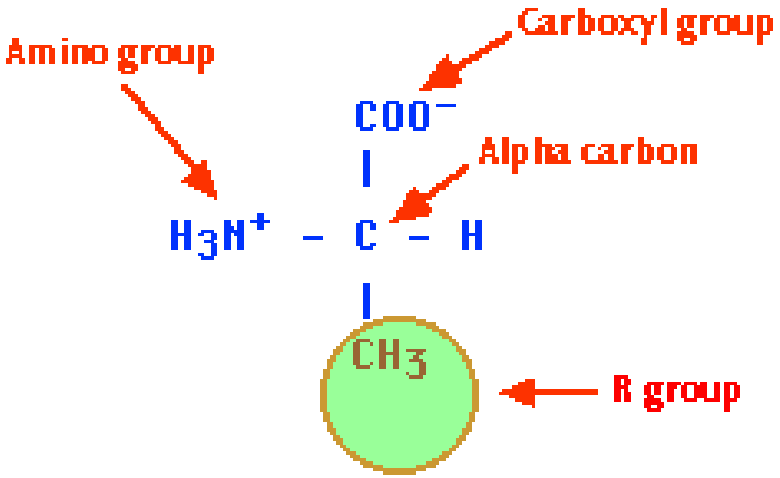
Macro-molecule	**Protein**	**Polysaccharides**
Building blocks	amino acids (20 different types)	monosaccharides (several types)
Typical length	10s to 1000s	up to 10^9 (e.g., starch)
Structure	complex and versatile, with structural motifs (secondary structure, domains, etc.)	often not linearly bonded but tree-like
Function	Extremely diverse. For example, ■ *enzymes* catalyze reactions of other molecules; ■ *structural proteins* build and stabilize the structure of the cell; ■ *receptors*, *kinases*, and other proteins receive, transport, and process signals from the exterior; ■ *transcription factors (TF)* regulate the production of all proteins.	■ modification of proteins and their properties ■ storage of energy (e.g., in starch) ■ structural stability (e.g., in chitin) ■ storage of water (e.g., in extracellular matrix in cartilage)
Location	everywhere in- and outside cell; dissolved in water or embedded in a membrane	everywhere in- and outside cell; often bound to proteins



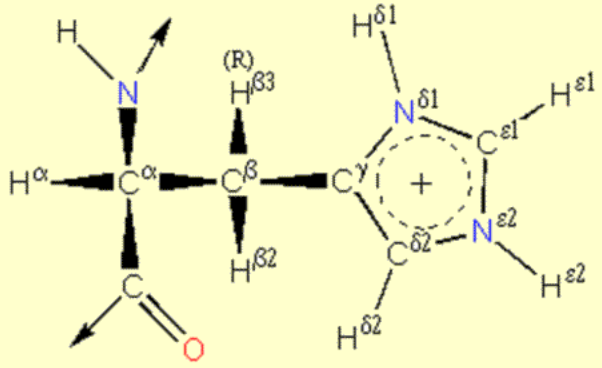
The Amino Acids

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

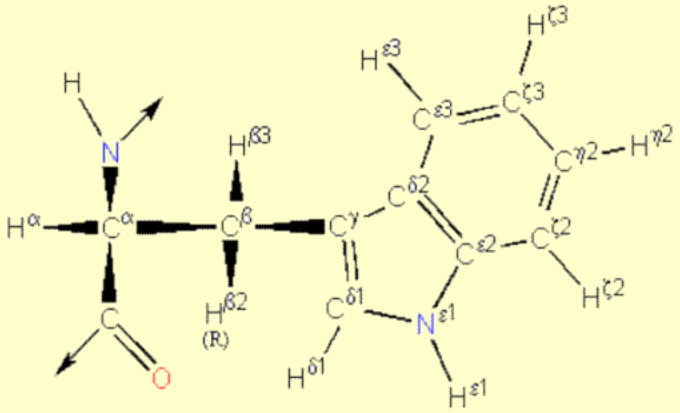
Alanine	Ala	A	hydrophobic
Arginine	Arg	R	free amino group makes it basic and hydrophilic
Asparagine	Asn	N	carbohydrate can be covalently linked ("N-linked) to its -NH
Aspartic acid	Asp	D	free carboxyl group makes it acidic and hydrophilic
Cysteine	Cys	C	oxidation of their sulfhydryl (-SH) groups link 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


Amino group
Carboxyl group
COO⁻
Alpha carbon
H3N+ – C – H
CH3
R group

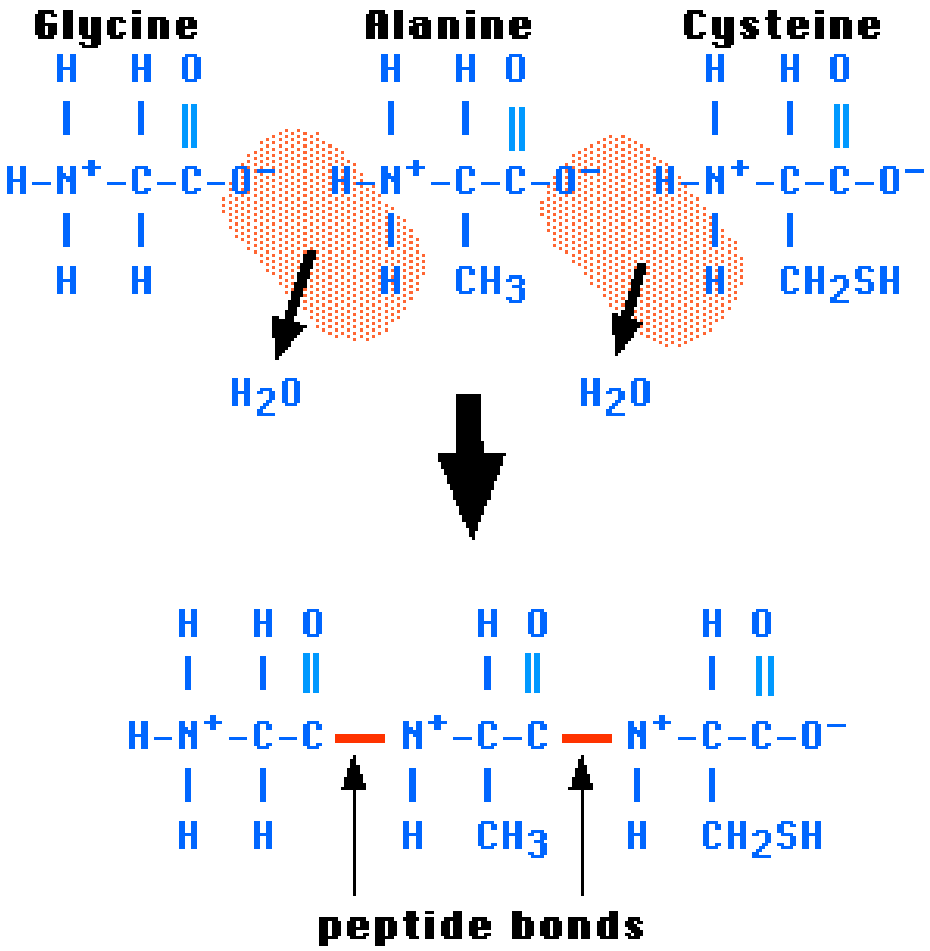
L-Histidine
(His)



L-Tryptophan
(Trp)

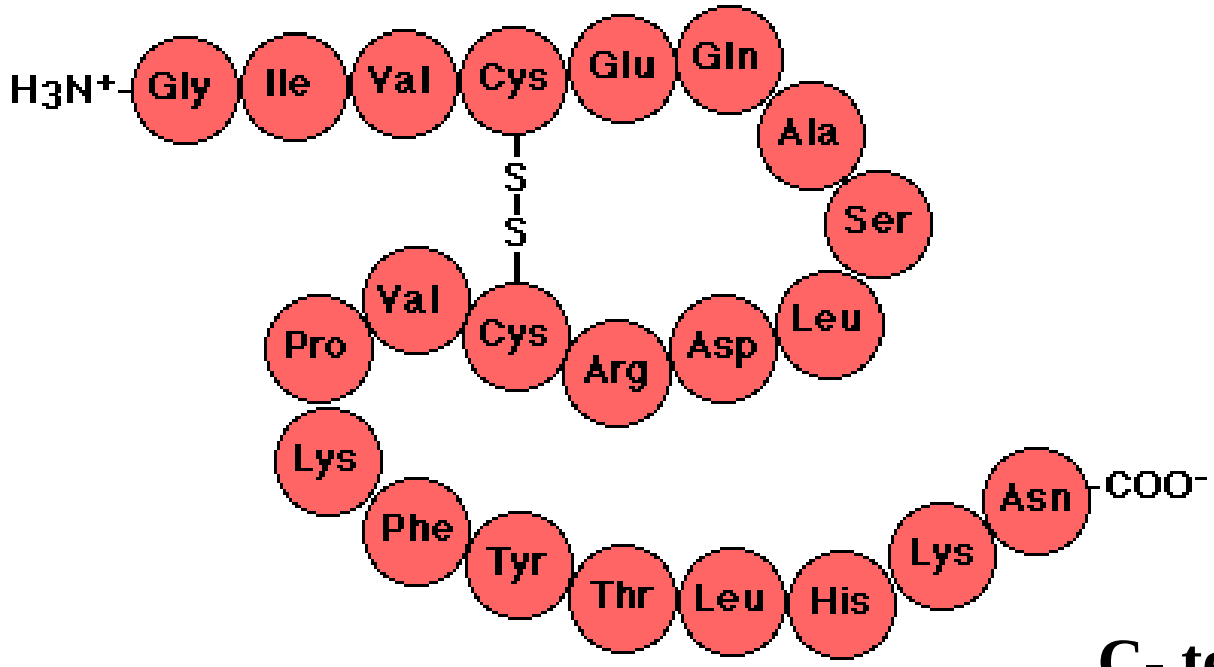


Peptide bond

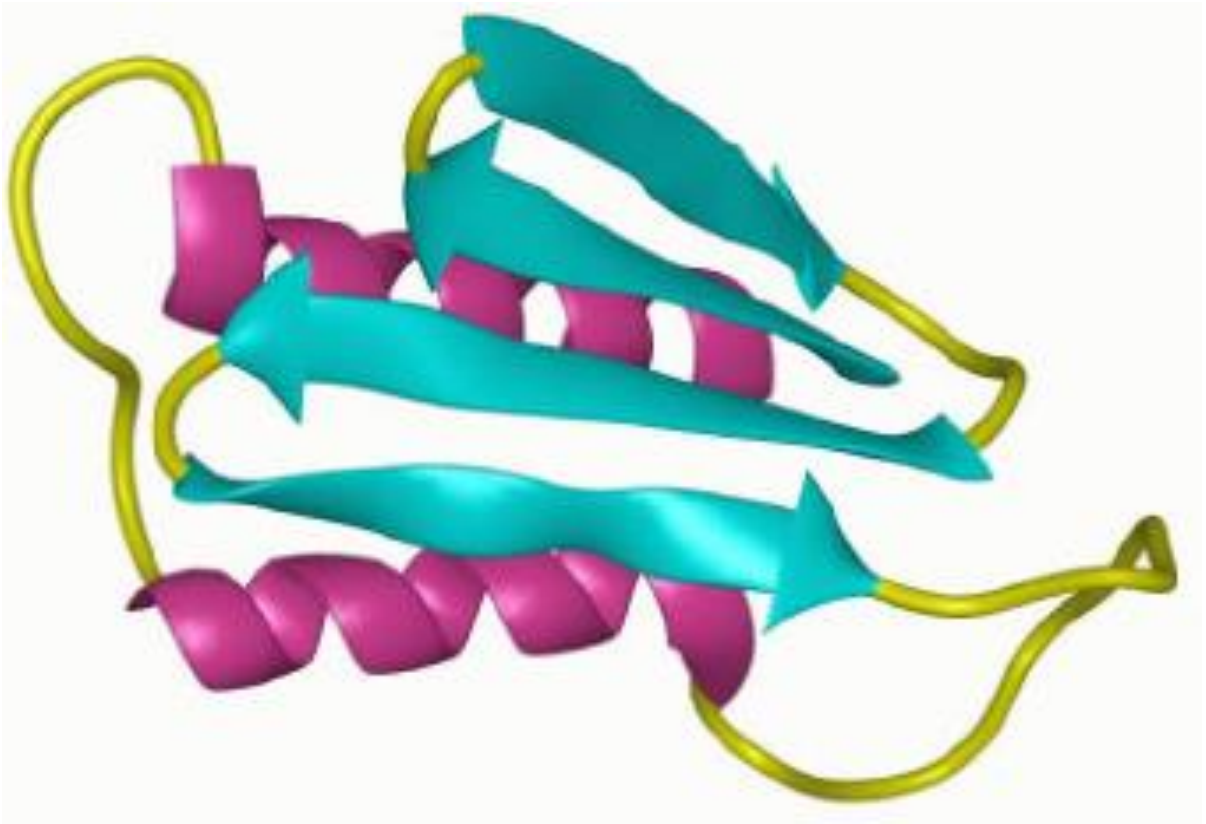


Polypeptide

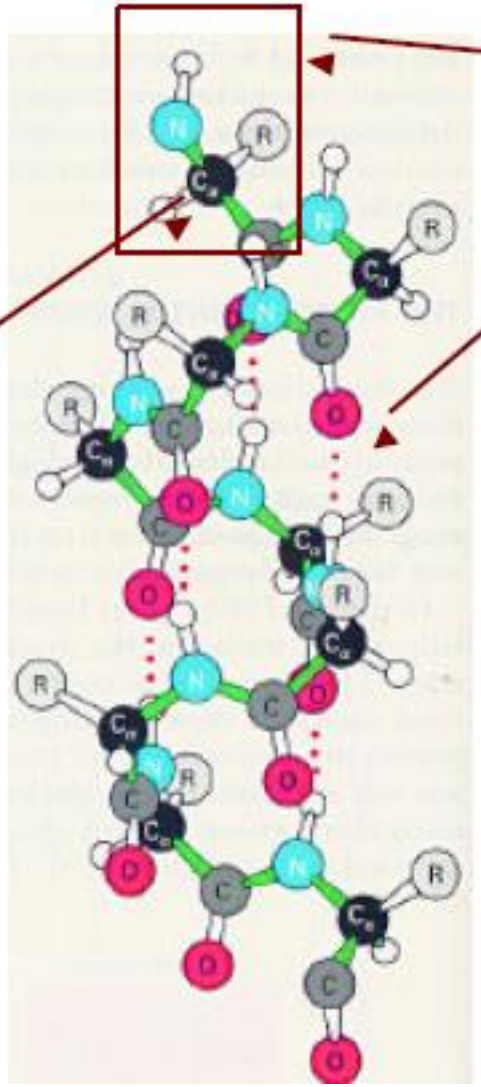
N-terminal



C- terminal

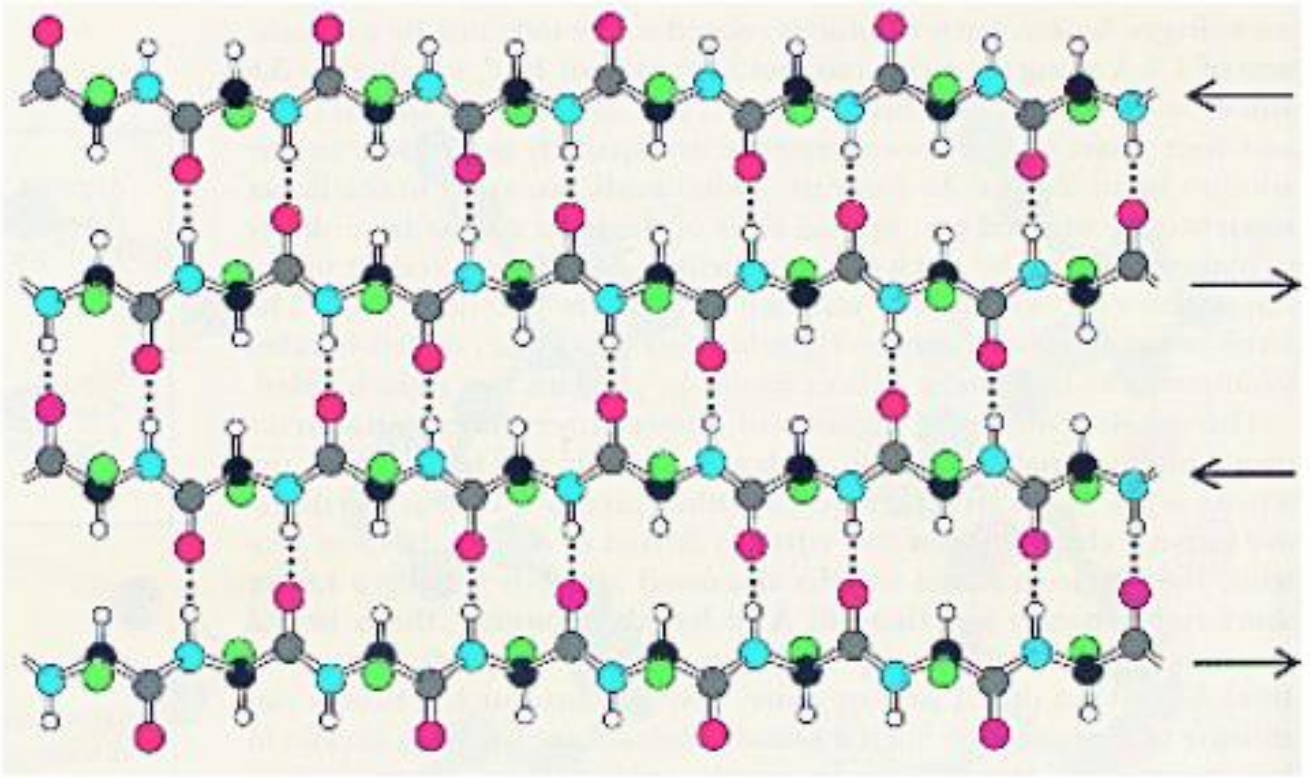


α Helices



- 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



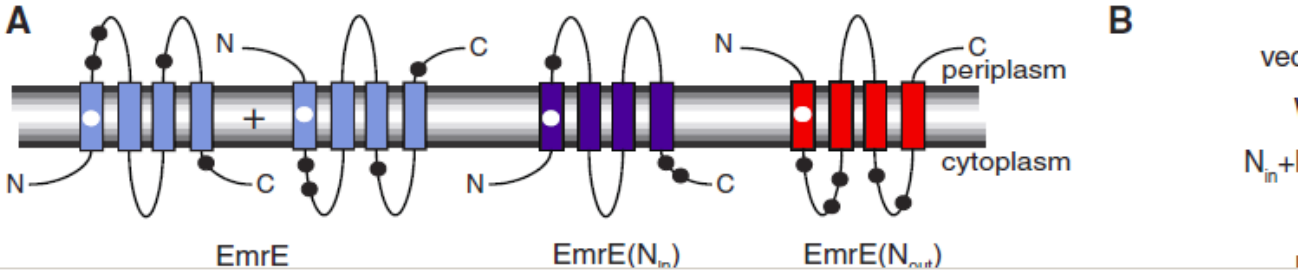
Hydrogen bond b/w carbonyl oxygen atom on one chain and NH group on the adjacent chain

Control of Membrane Protein Topology by a Single C-Terminal Residue

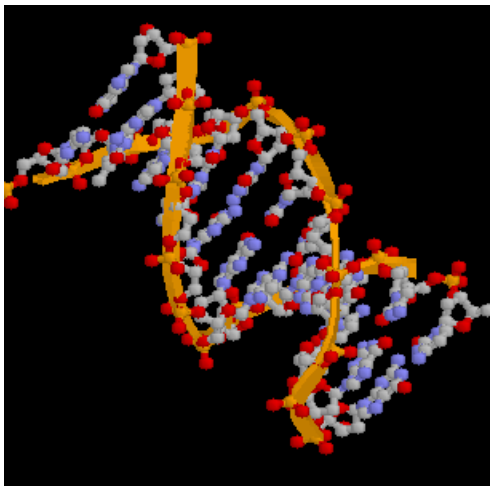
Susanna Seppälä,[1] Joanna S. Slusky,[1] Pilar Lloris-Garcerá,[1] Mikaela Rapp,[1]* Gunnar von Heijne[1,2]†

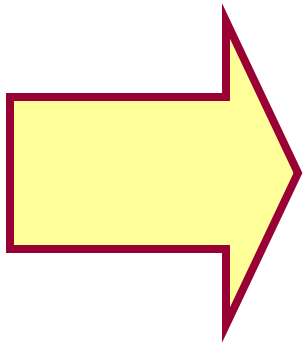
The mechanism by which multispanning helix-bundle membrane proteins are inserted into their target membrane remains unclear. In both prokaryotic and eukaryotic cells, membrane proteins are inserted cotranslationally into the lipid bilayer. Positively charged residues flanking the transmembrane helices are important topological determinants, but it is not known whether they act strictly locally, affecting only the nearest transmembrane helices, or can act globally, affecting the topology of the entire protein. Here we found that the topology of an *Escherichia coli* inner membrane protein with four or five transmembrane helices could be controlled by a single positively charged residue placed in different locations throughout the protein, including the very C terminus. This observation points to an unanticipated plasticity in membrane protein insertion mechanisms.

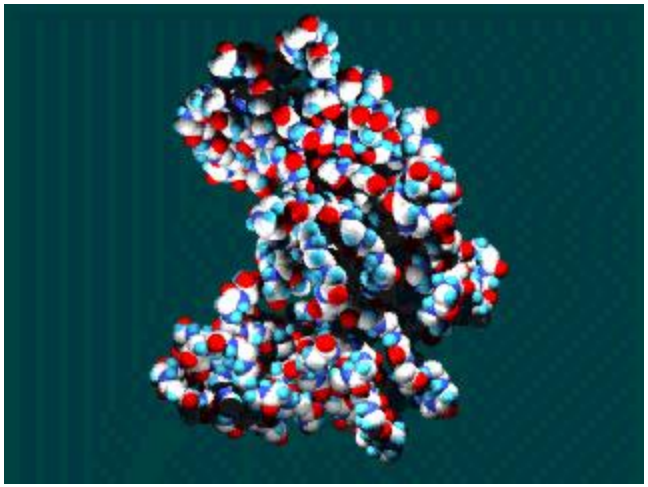
Integral α-helical membrane proteins carry out a wide range of central biological functions. They have two conspicuous structural features: hydrophobic transmembrane α helices and a strong bias in the distribution of positively charged arginine (Arg) and lysine (Lys) residues



Information Expression





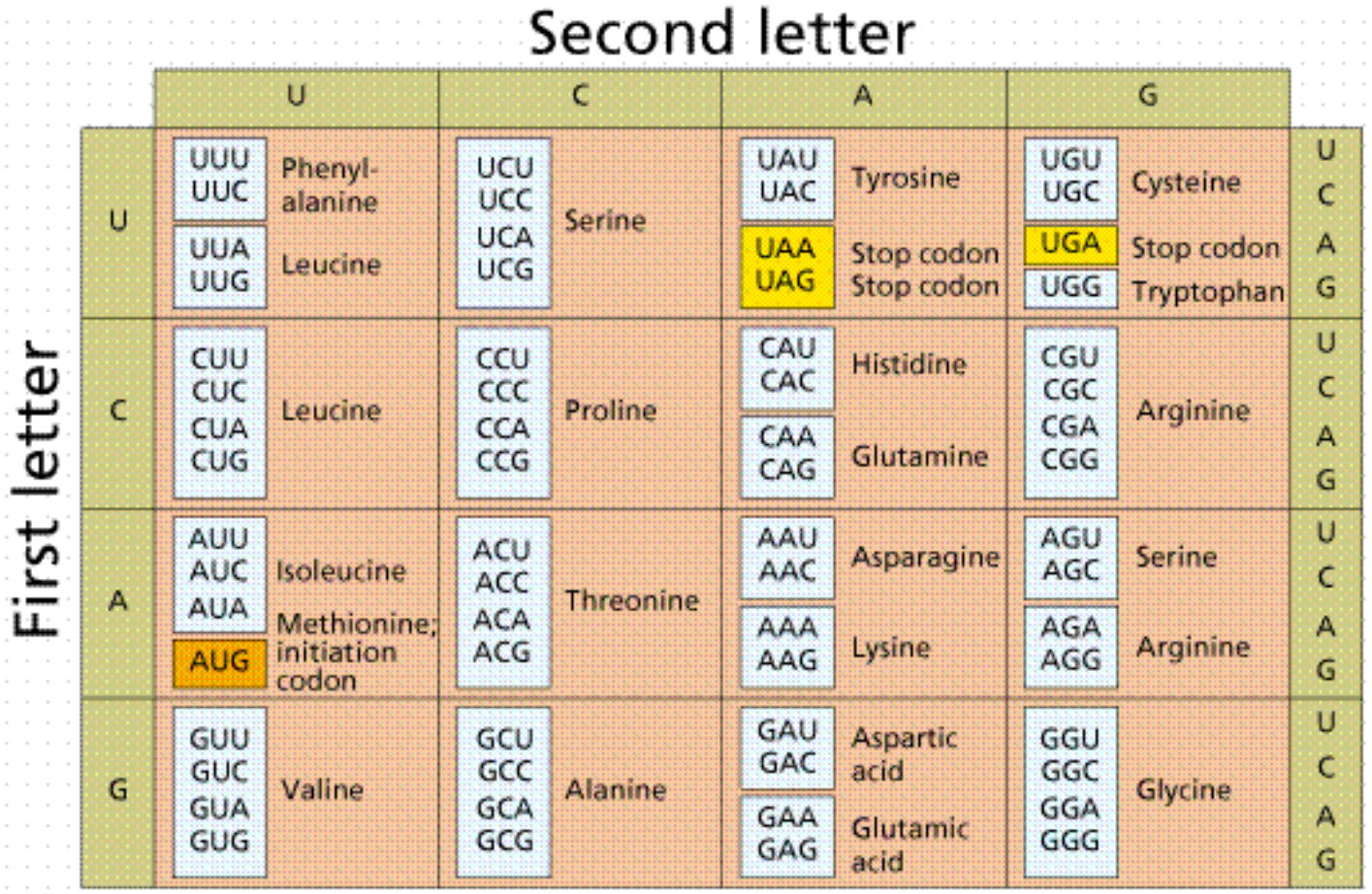


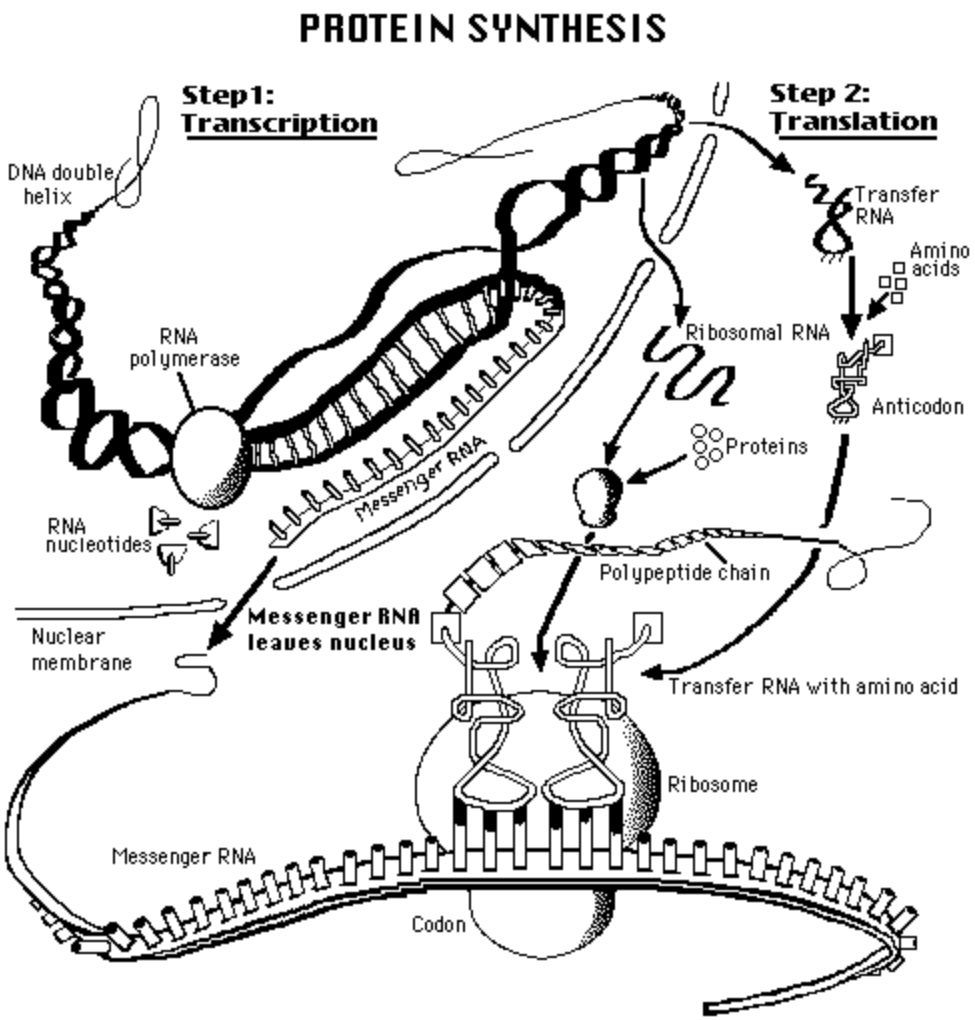
1-D information array

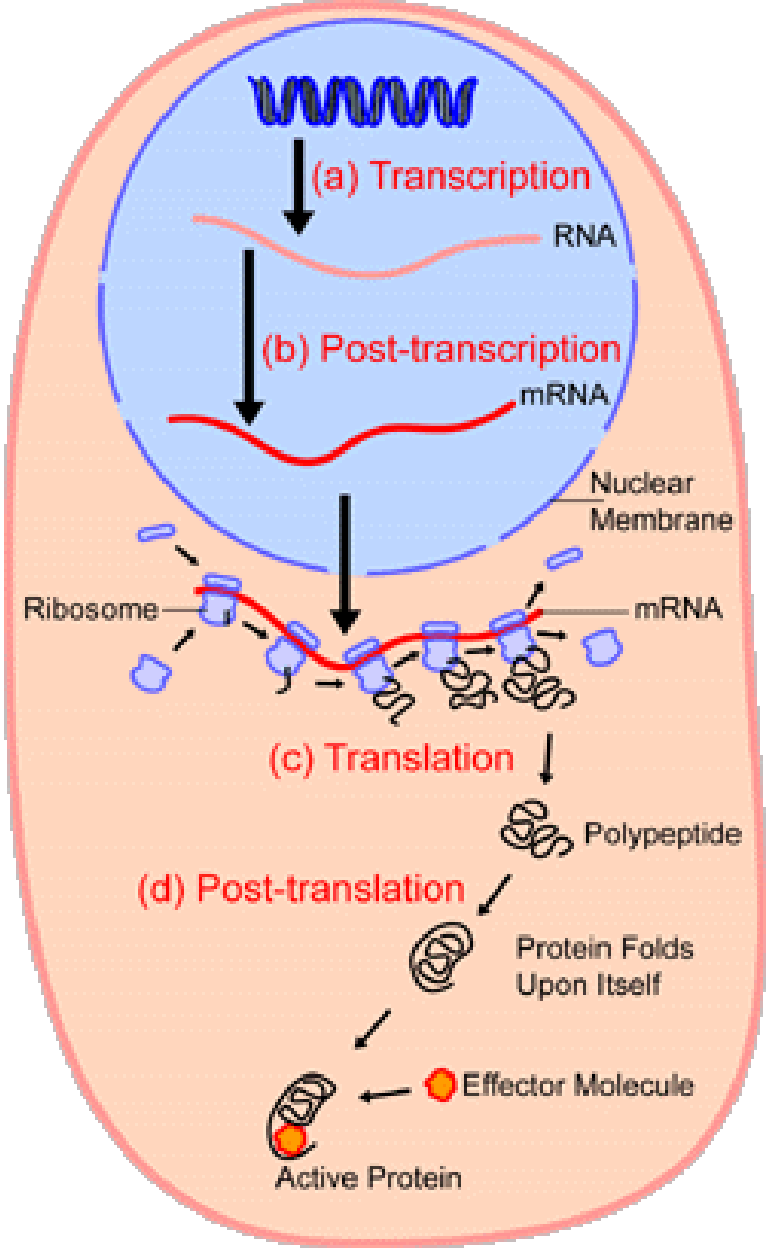
3-D biochemical structure

Central Dogma: DNA $\rightarrow$ RNA $\rightarrow$ Protein

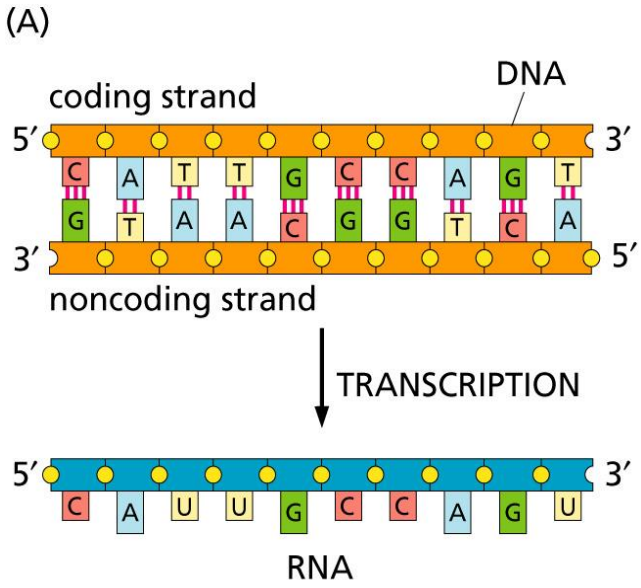
Genetic Code: codons

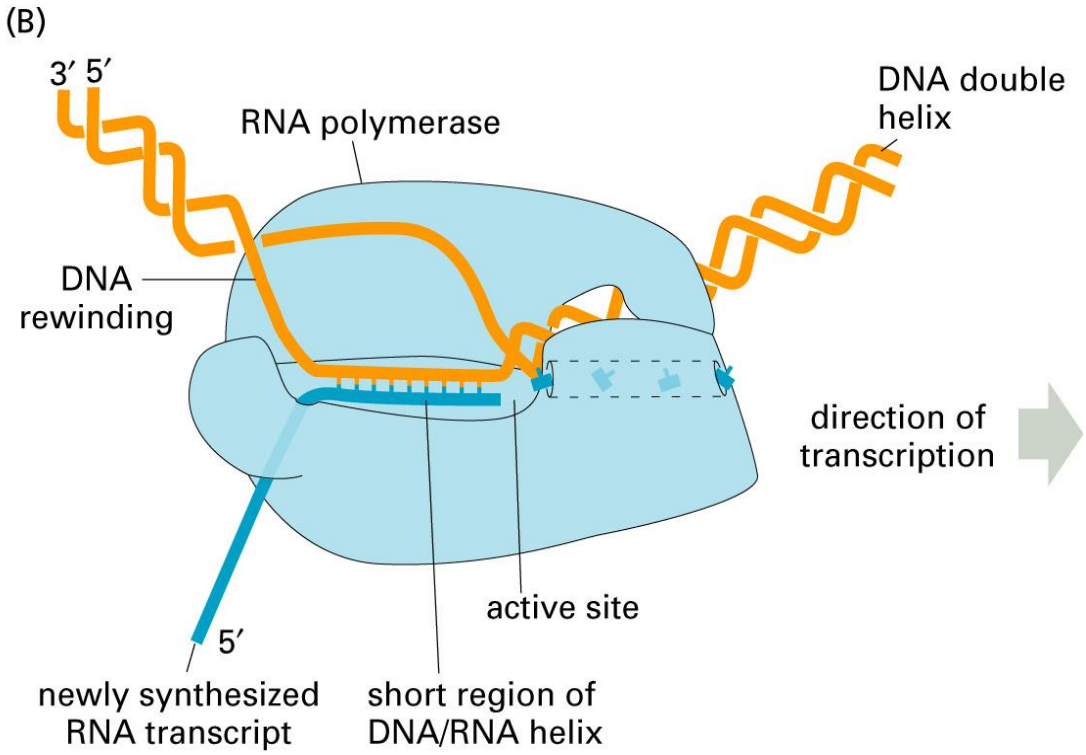



PROTEIN SYNTHESIS
Step1:
Transcription
Step 2:
Translation
DNA double helix
RNA polymerase
Messenger RNA
RNA nucleotides
Transfer RNA
Amino acids
Ribosomal RNA
Anticodon
Proteins
Polypeptide chain
Nuclear membrane
Messenger RNA leaves nucleus
Transfer RNA with amino acid
Ribosome
Messenger RNA
Codon


(a) Transcription
RNA
(b) Post-transcription
mRNA
Nuclear
Membrane
Ribosome
mRNA
(c) Translation
Polypeptide
(d) Post-translation
Protein Folds
Upon Itself
Effector Molecule
Active Protein

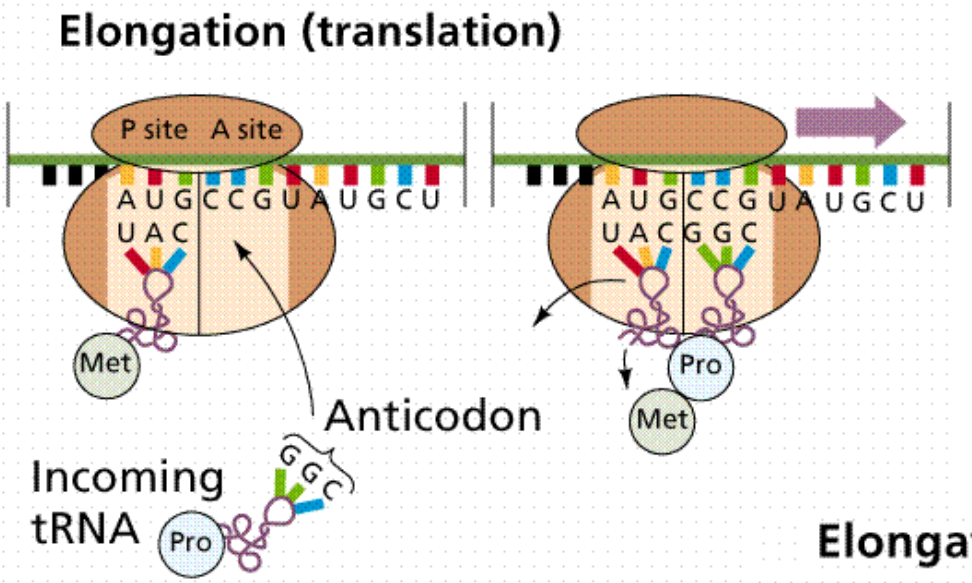
Transcription

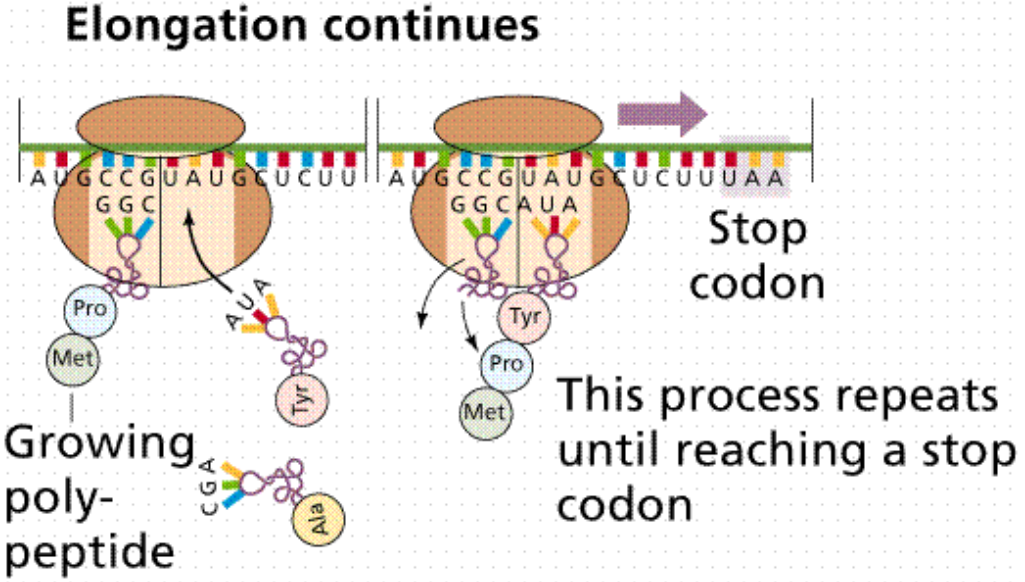


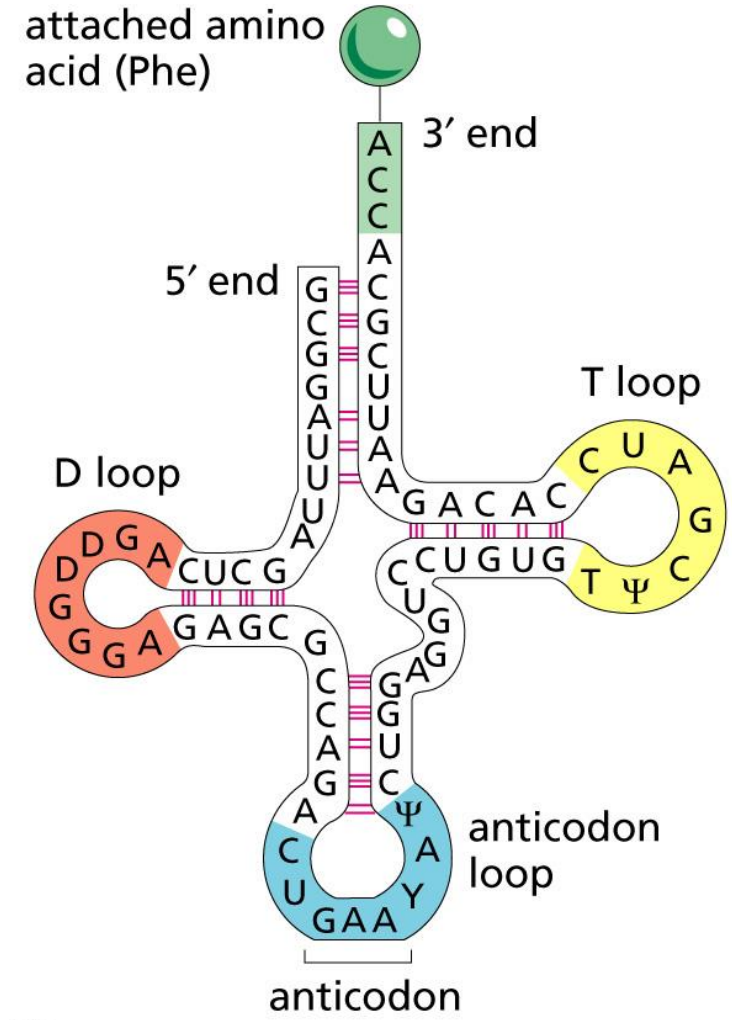


Translation

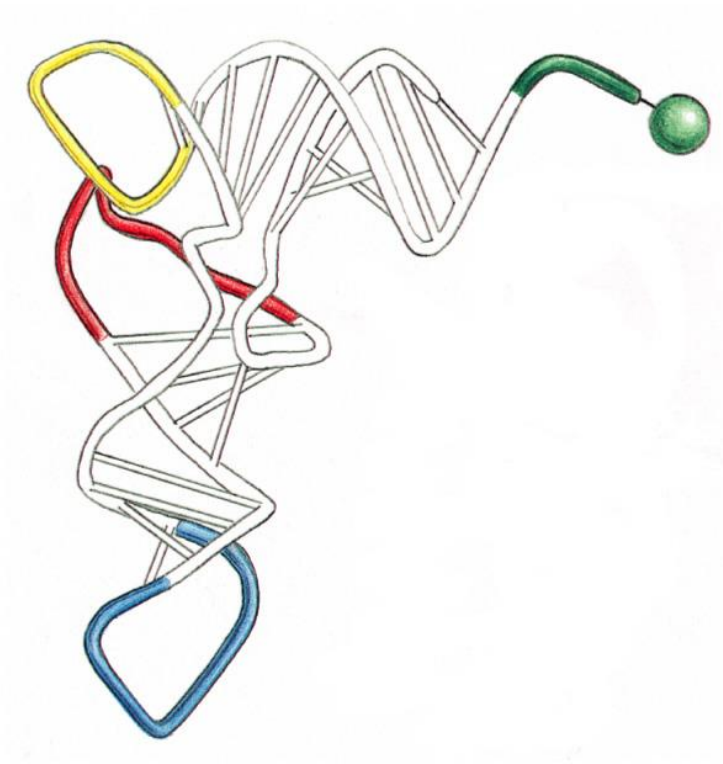
http://www.youtube.com/watch?v=B6O6uRb1D38





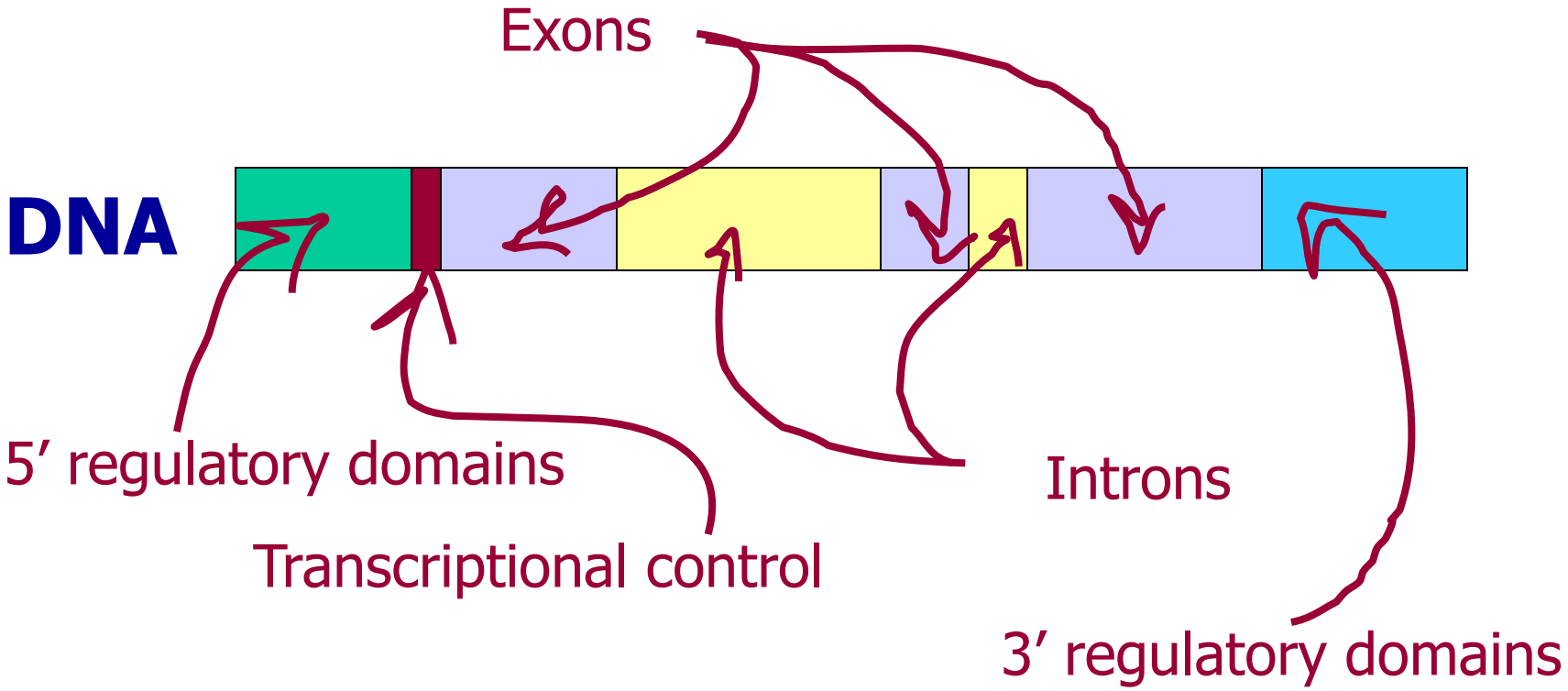

attached amino acid (Phe)
3′ end
5′ end
T loop
D loop
anticodon loop
anticodon

(A)

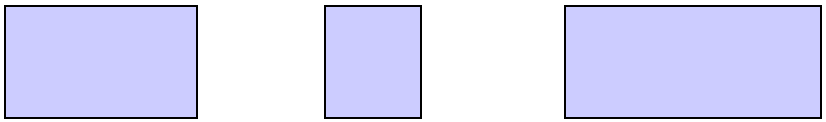


(B)

Gene Structure



Post-transcriptional processing: hnRNA to mRNA



Translation: mRNA to protein

How complex can a 4 letter code really be?

atcgggctatcgatagctatagcgcgatatatcgcgcgtatatgcgcgcatattag
tagctagtgctgattcatctggactgtcgtaatatatacgcgcccggctatcgcgct
atgcgcgatatcgcgcggcgctatataaatattaaaaaataaaatatatatatatgc
tgcgcgatagcgctataggcgcgctatccatatataggcgctcgcccgggcgcga
tgcatcggctacggctagctgtagctagtcggcgattagcggcttatatgcggcga
gcgatgagagtcgcggctataggcttaggctatagcgctagtatatagcggctagc
cgcgtagacgcgatagcgtagctagcggcgcgcgtatatagcgcttaagagcca
aaatgcgtctagcgctataatatgcgctatagctatatgcggctattatatagcgca
gcgctagctagcgtatcaggcgaggagatcgatgctactgatcgatgctagagca
gcgtcatgctagtagtgccatatatatgctgagcgcgcgtagctcgatattacgcta
cctagatgctagcgagctatgatcgtagca.....

- Alternative splicing
  - Exception to the “One gene one protein” rule.
- Codon usage
  - http://www.kazusa.or.jp/codon/

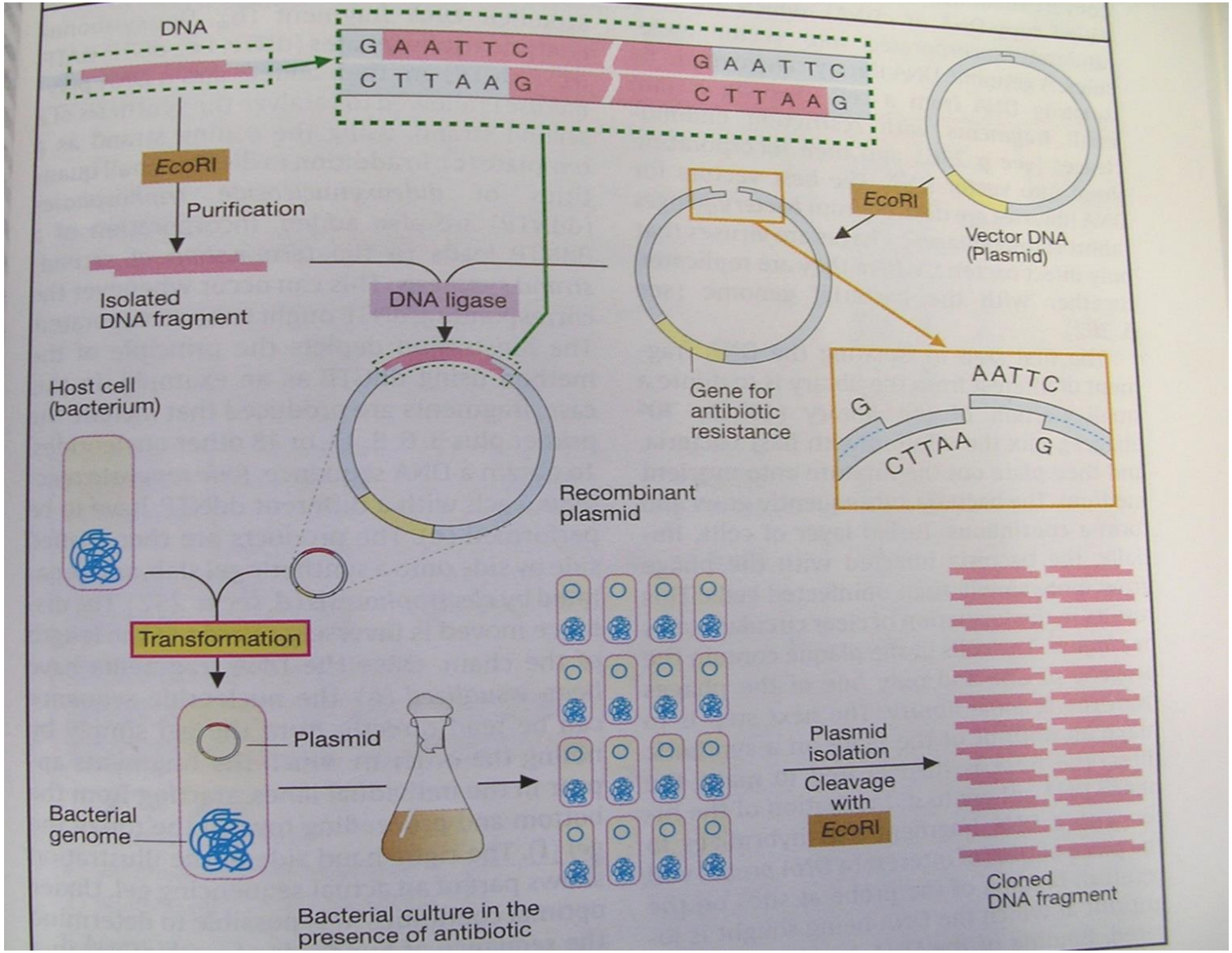
Given a DNA sequence, we like to computationally

- Identify genes,
  - introns, exons, alternative splicing sites, promoters, …
- Determine the functions of the protein that a gene encodes
- Identify functional signatures, e.g., motifs
- Determine the structure of proteins

Molecular biology tools

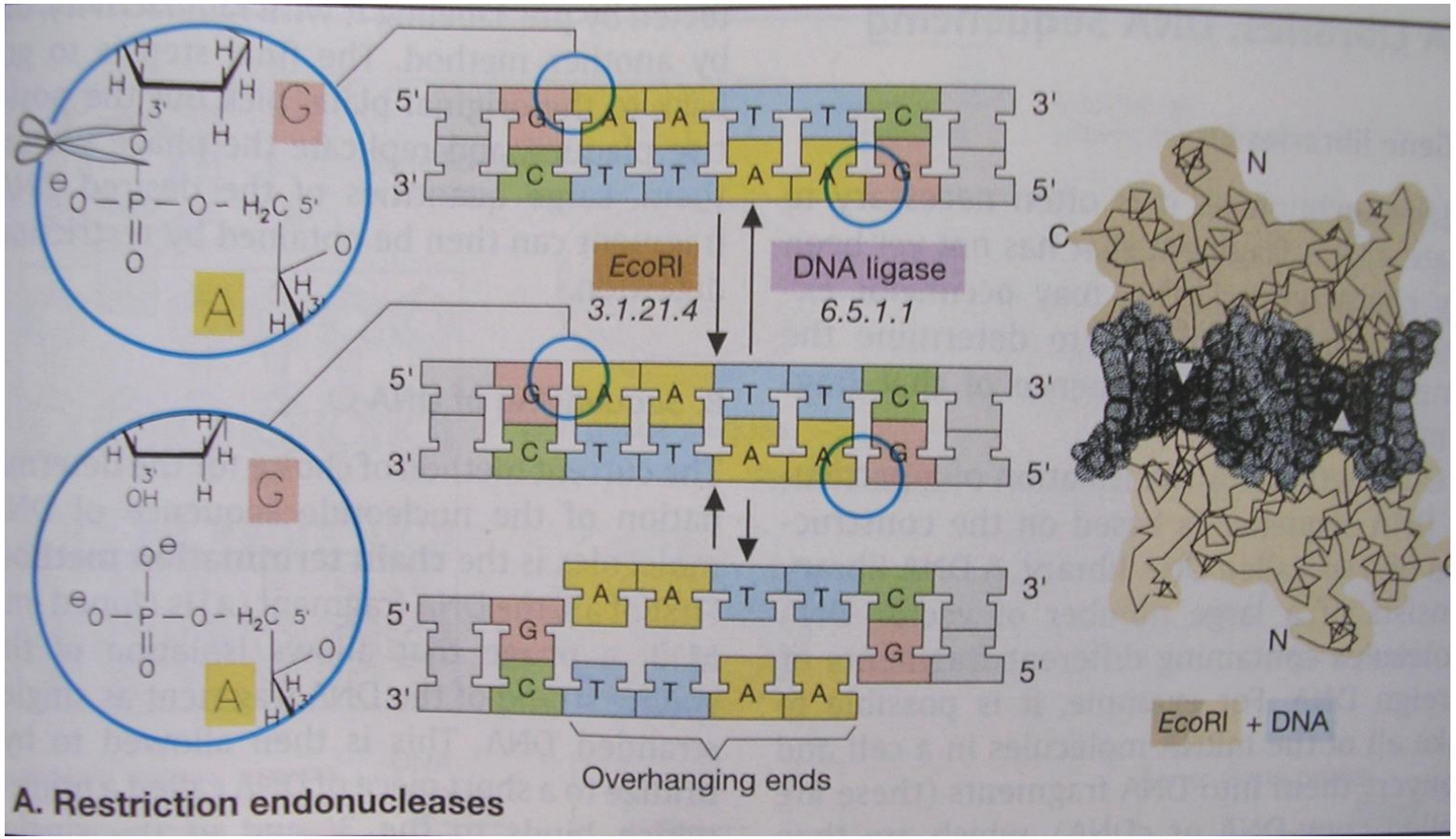
- Clone
- PCR
- Sequencing
- Microarray
- Yeast 2 hybrid

DNA Cloning



Courtesy of Color Atlas of Biochemistry

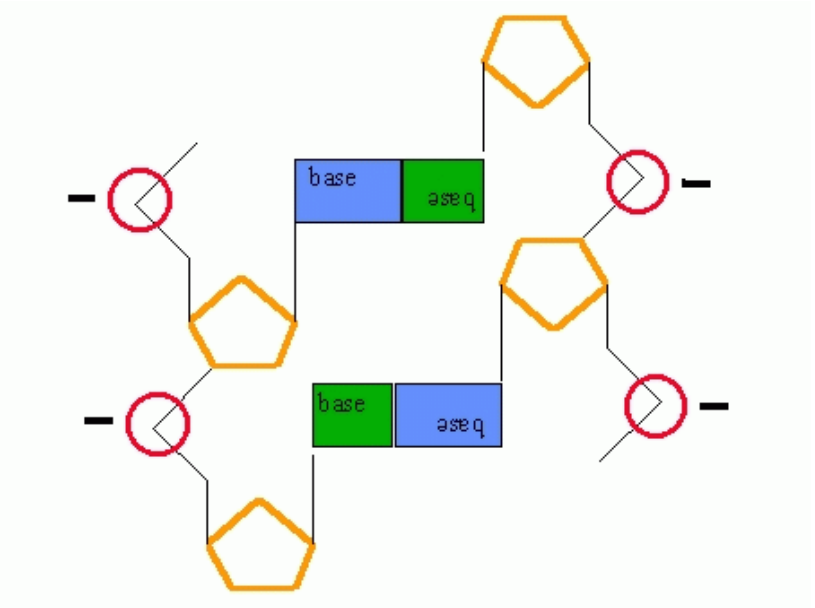
Restriction endonucleases

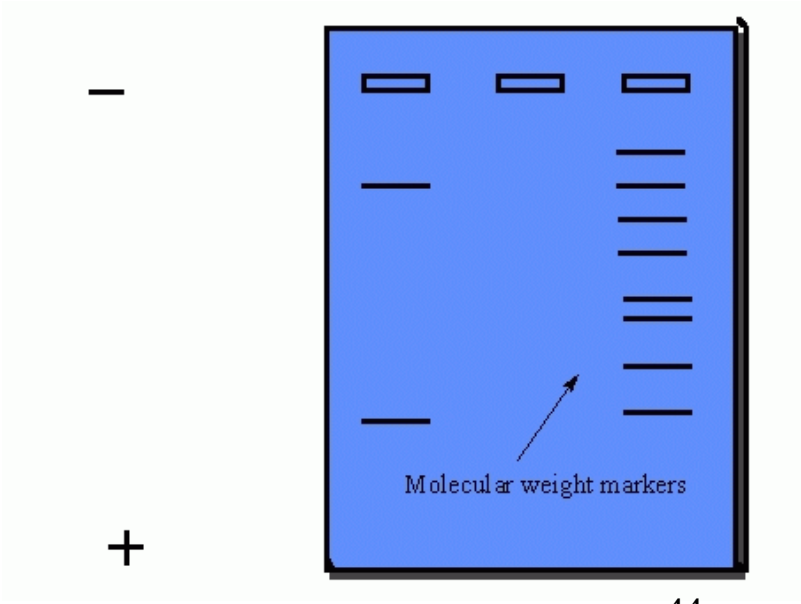


Gel electrophoresis

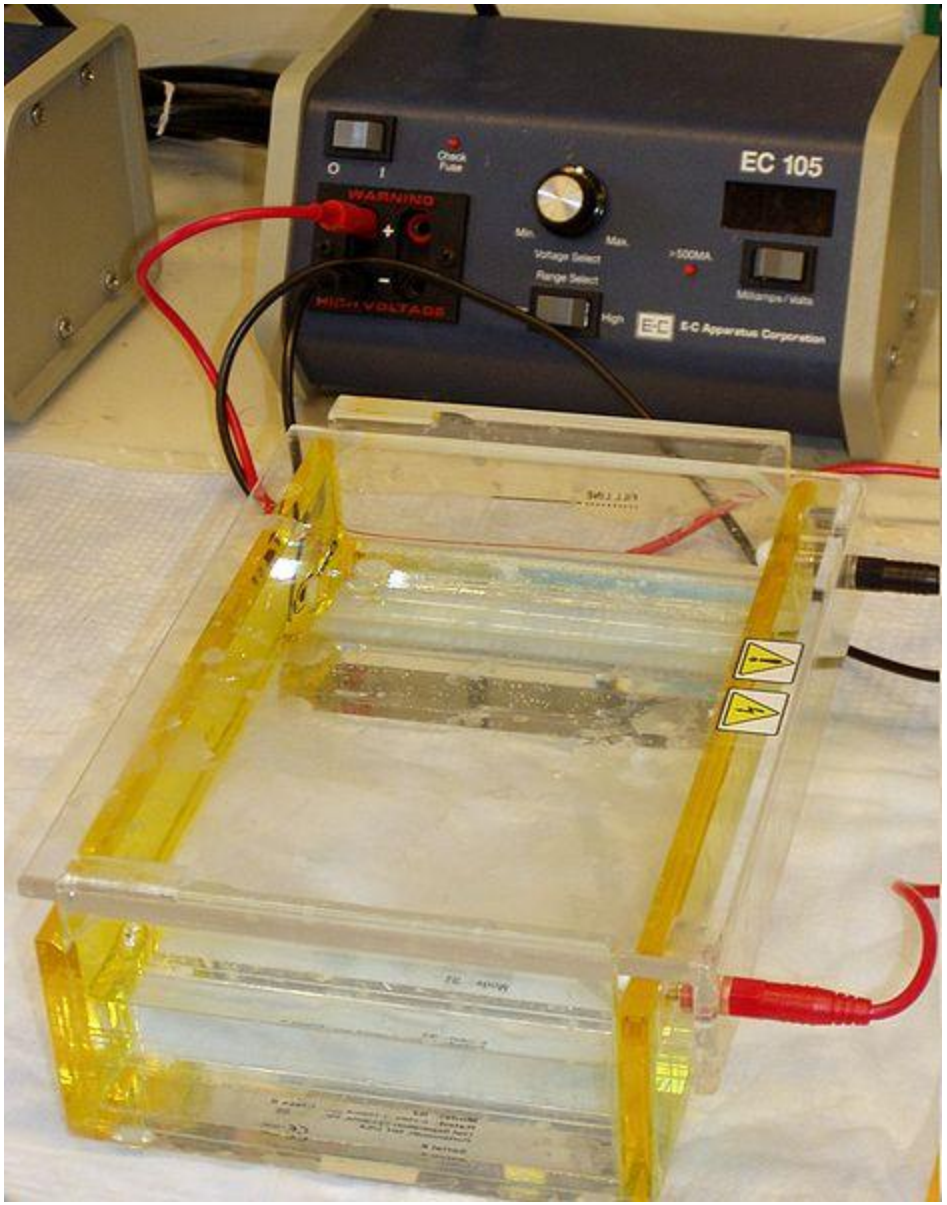
a procedure for separating a mixture of molecules through a stationary material (gel) in an electrical field.

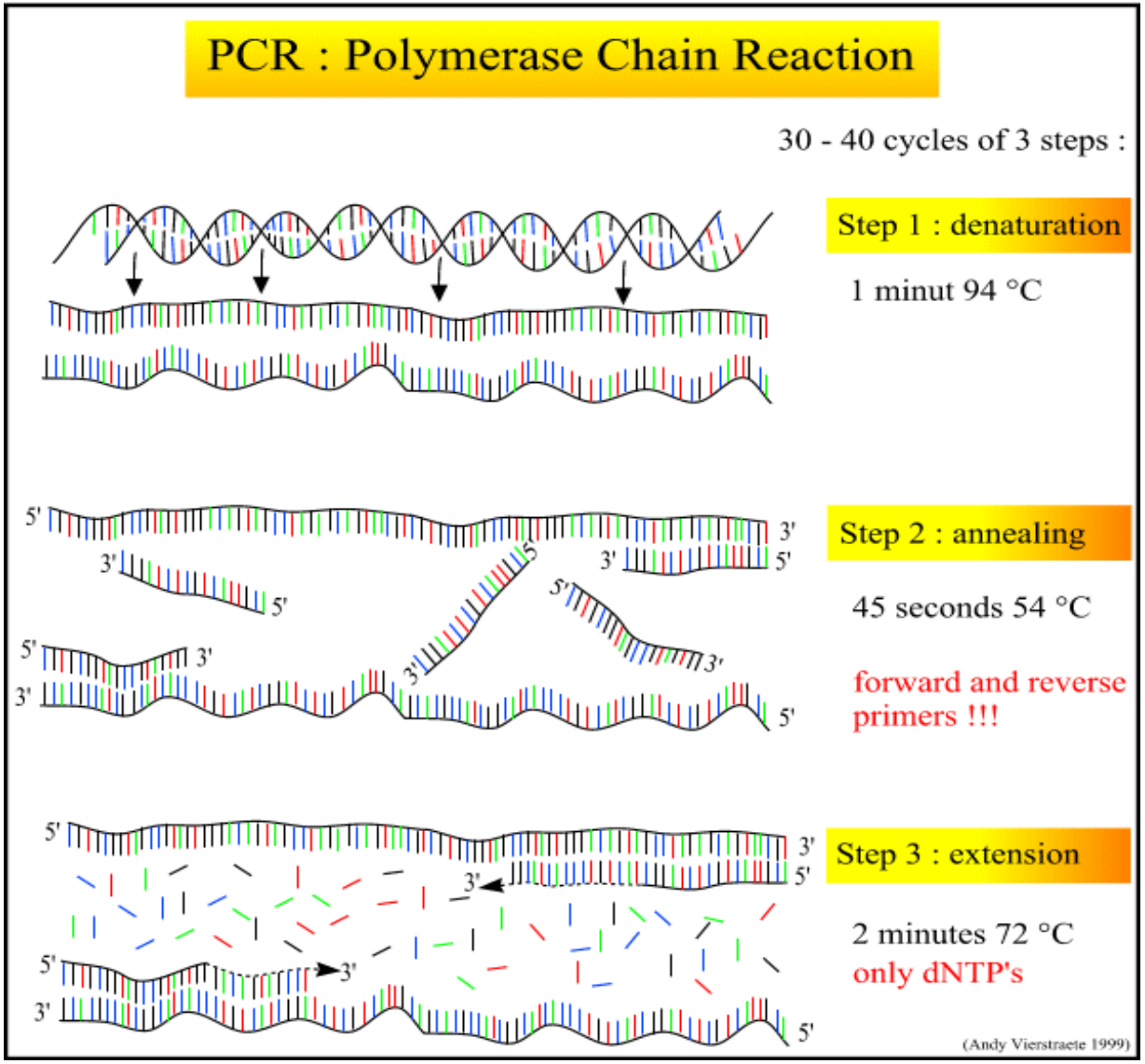
- **Organic molecules such as DNA are charged. DNA is negatively charged because the phosphates (red circles) that form the sugar-phosphate backbone of a DNA molecule have a negative charge.**
- **A gel is prepared which will act as a support for separation of the fragments of DNA. The gel is a jello-like material, usually agarose, a substance derived from seaweed.**
- **Holes are created in the gel. These will serve as a reservoir to hold the DNA solution.**
- **Large molecules have difficulty getting through the holes in the matrix. Small molecules move easily through the holes**
- **Because of this, the distance moved is inversely related to the weight, and therefore the length of DNAs.**
- **Molecular weight markers, usually a mixture of DNAs with known molecular weights, are often electrophoresed along with DNAs to estimate the sizes of DNA fragments in the sample.**



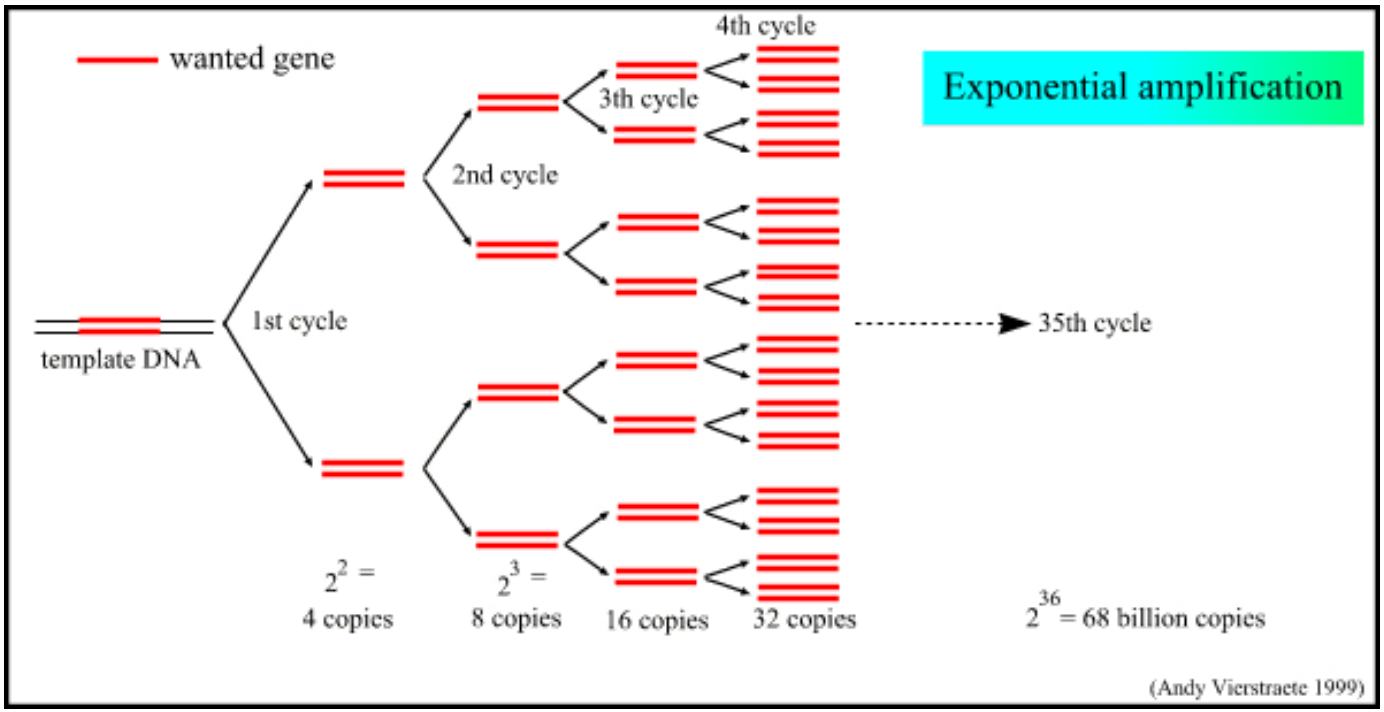


Gel electrophoresis apparatus - An agarose gel is placed in this buffer-filled box and electrical field is applied via the power supply to the rear. The negative terminal is at the far end (black wire), so DNA migrates toward the camera.





The bases (complementary to the template) are coupled to the primer on the 3' side (the polymerase adds dNTP's from 5' to 3', reading the template from 3' to 5' side, bases are added complementary to the template)

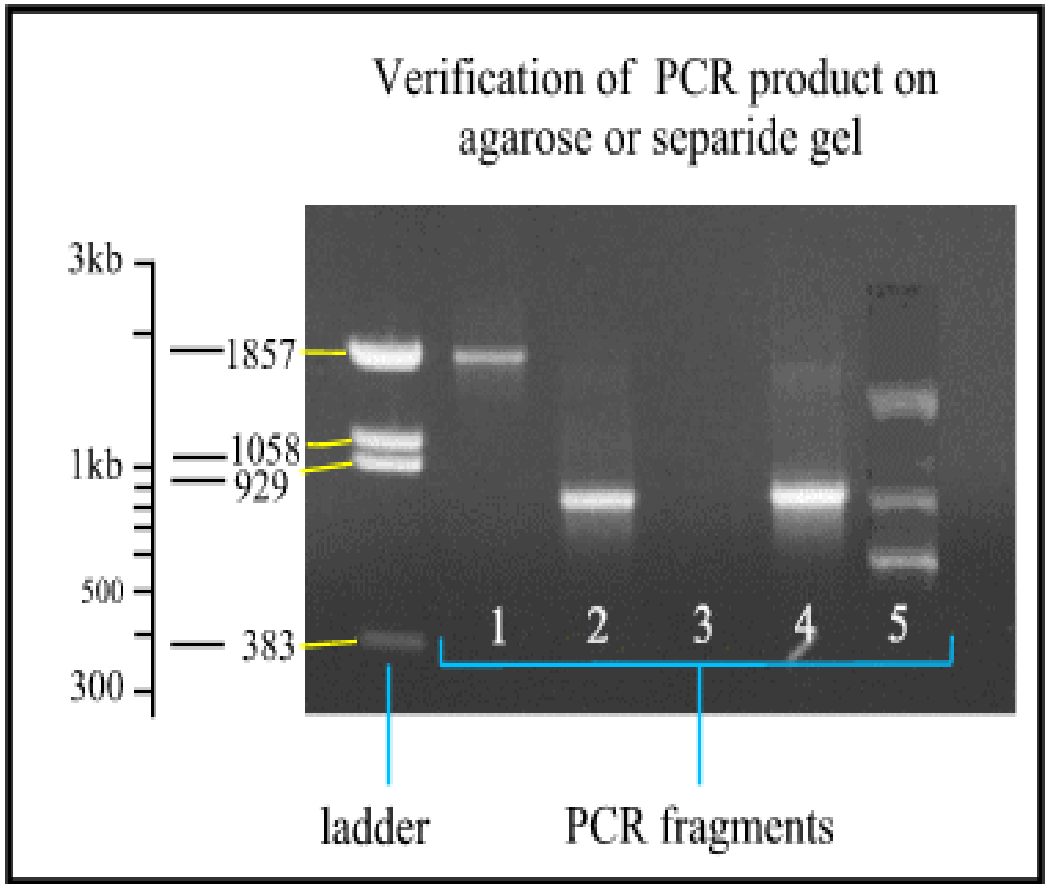


• No need for restriction enzymes, vectors, and host cells.

• Only need: two primers, sufficient amounts of the four deoxylribonulceoside triphosphates, and heat-tolerant DNA polymerase.

•Primer design is computationally aided.

- Is it the right size?
- Is there a product?
- Is it a mixture?



How PCR is used in forensic science

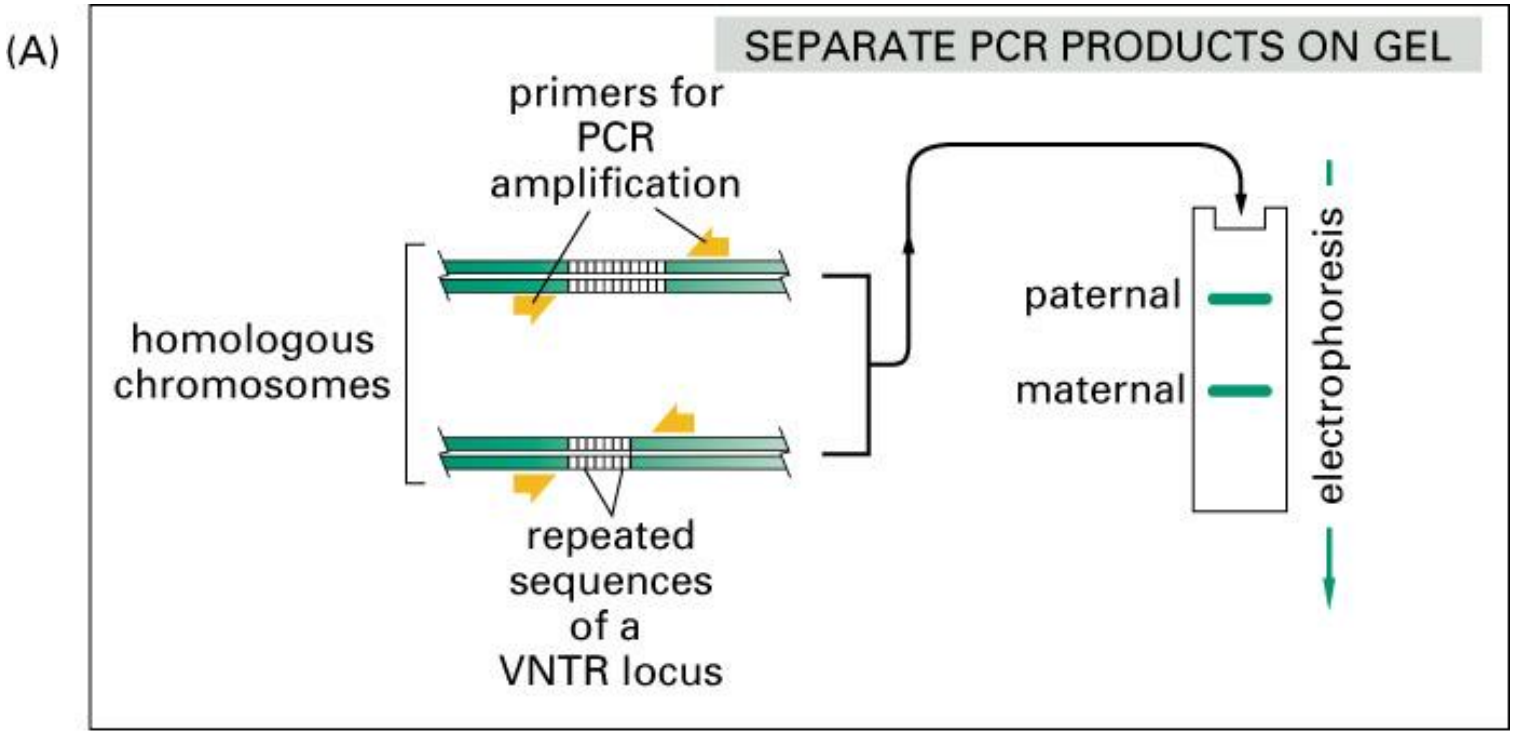
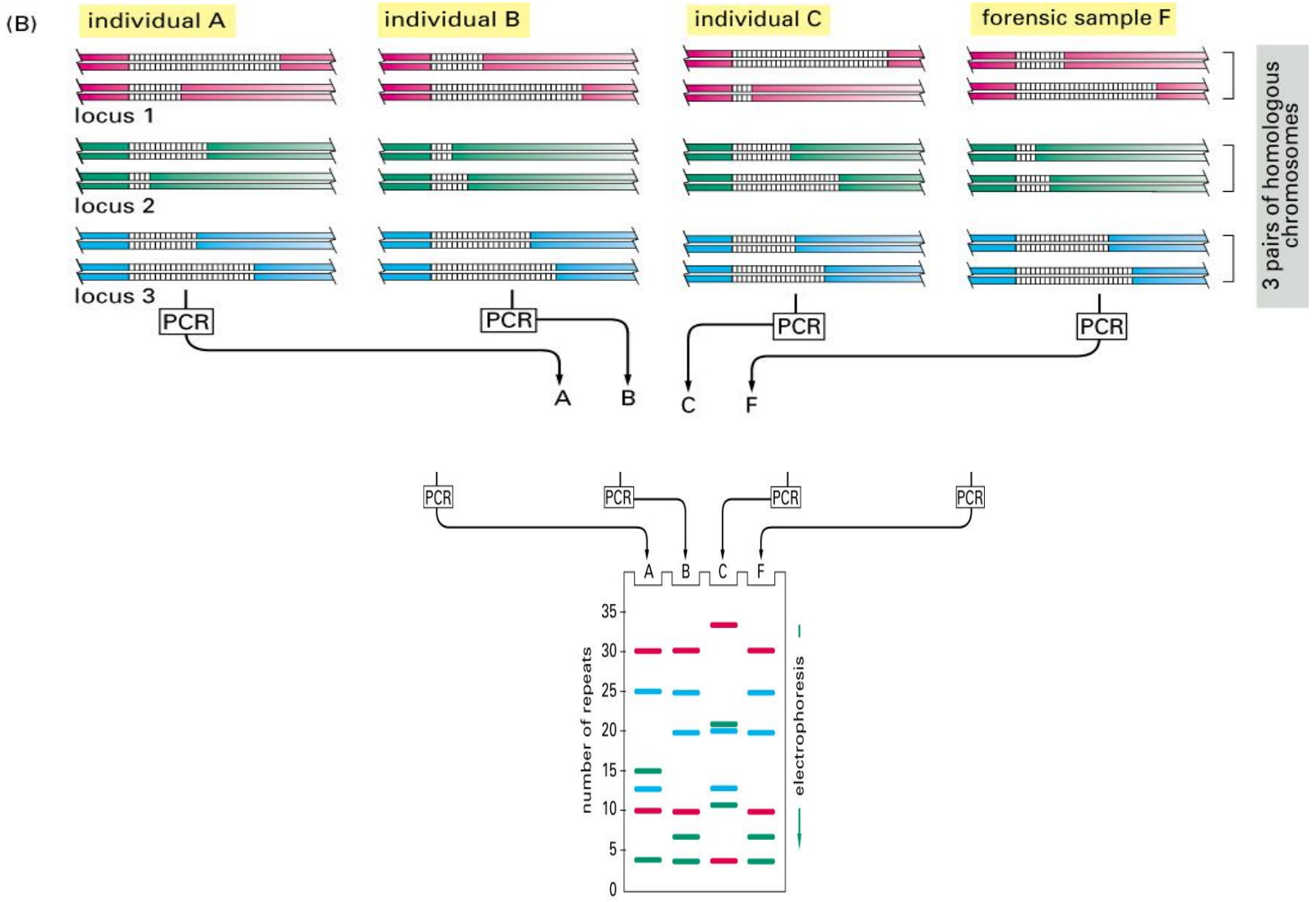


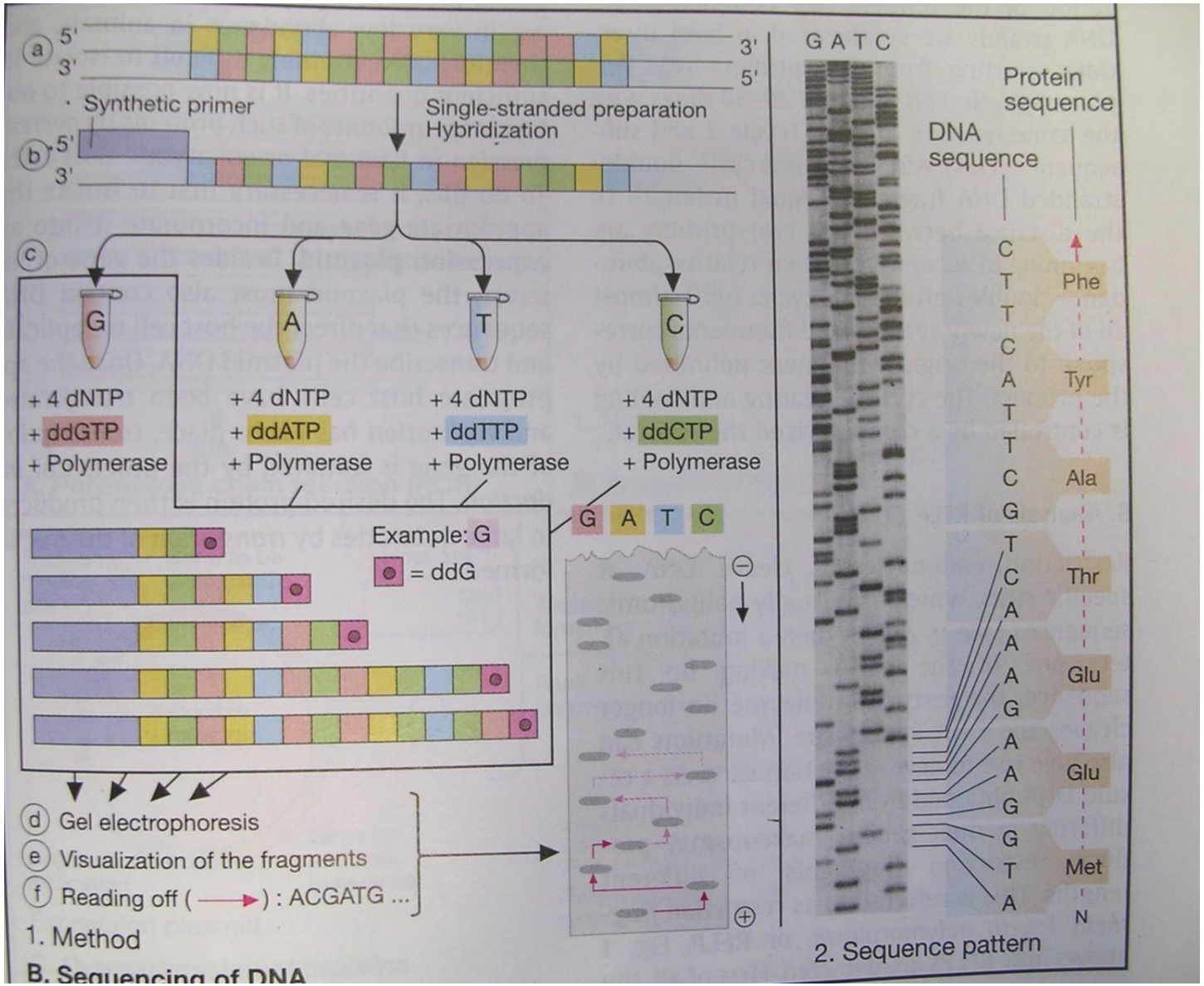
Figure 8–41 part 1 of 4. Molecular Biology of the Cell, 4th Edition.


(B)
individual A
individual B
individual C
forensic sample F
locus 1
locus 2
locus 3
3 pairs of homologous chromosomes
PCR
PCR
PCR
PCR
A
B
C
F
PCR
PCR
PCR
PCR
A
B
C
F
35
30
25
20
15
10
5
0
number of repeats
electrophoresis

Evolution of Sequencing Technology

- ABI Platform
- 1988
  - 16 samples per/day ~250 bp/sample
- 1992
  - 48 samples 3 times/day ~450 bp/sample
- 1996-8
  - 96 samples 4 times/day ~600 bp/sample
- 2006
  - 96 samples 4 times/day ~700-800 bp/sample
- 454
  - 300,000 – 400,000 samples 2 times per day 100 bp/sample
- Solexa
  - 2,000,000 – 3,000,000 samples once every 3 days 35-50 bp/sample
  - 40,000,000 samples once every 3 days 50 bp/sample

Sequencing DNA (Sanger's method)



Courtesy of Color Atlas of Biochemistry

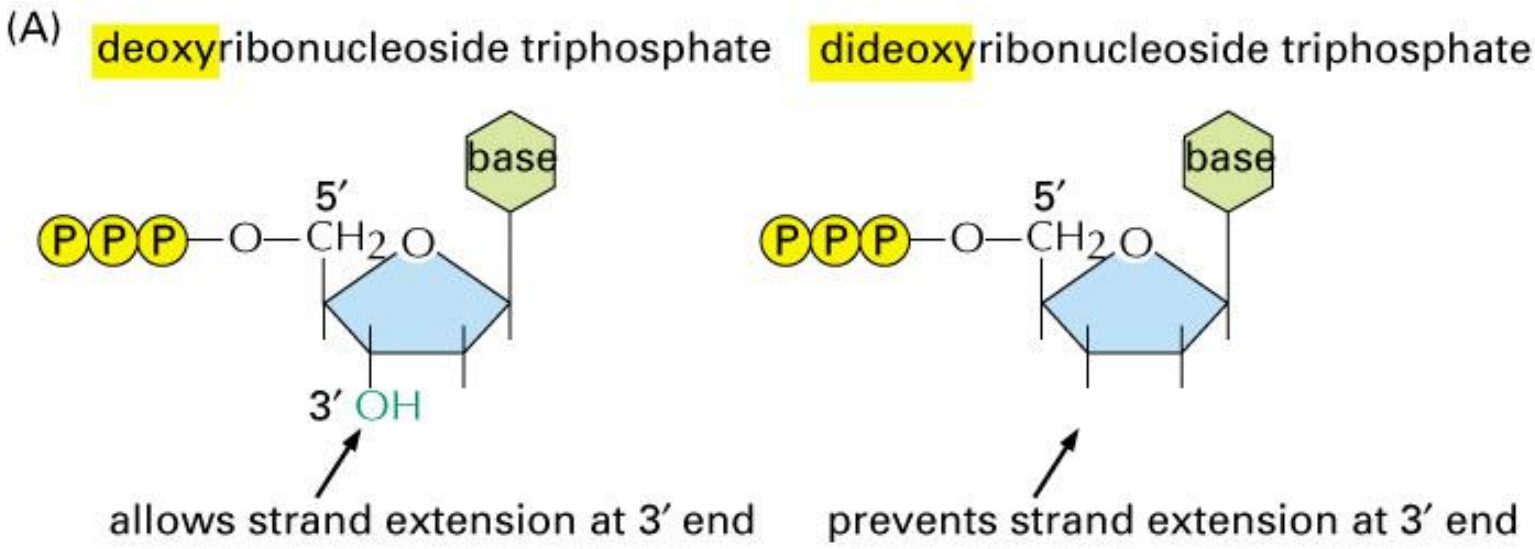
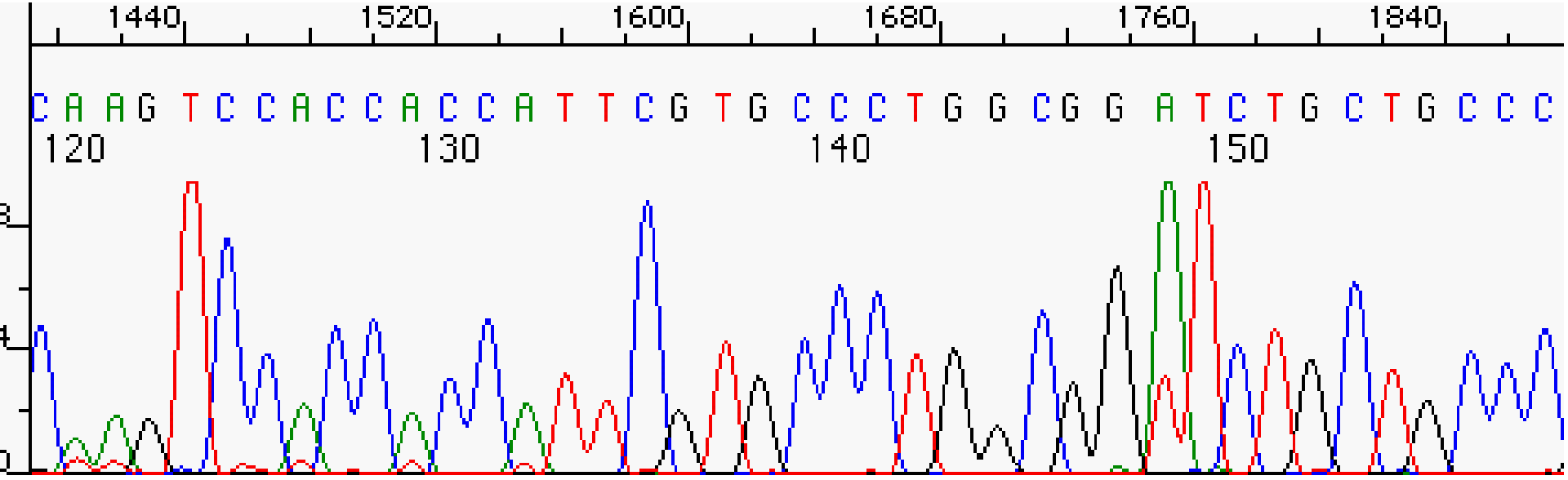
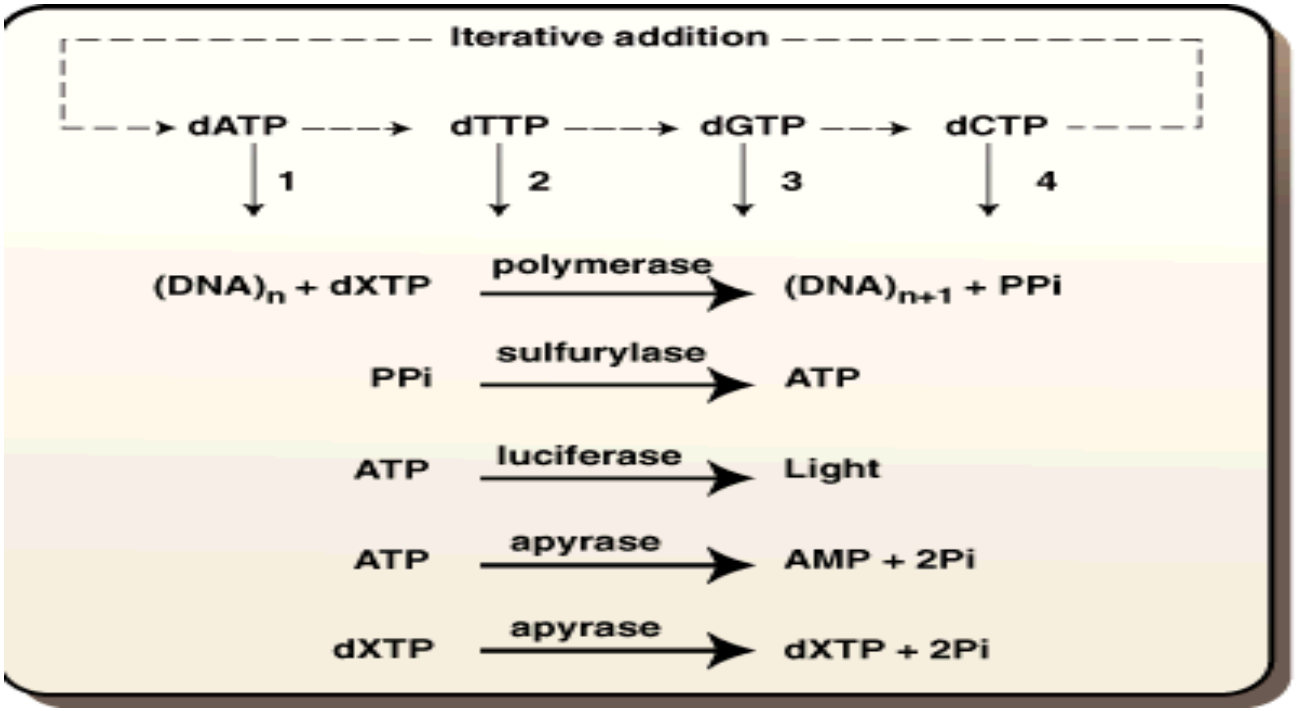


Figure 8–36 part 1 of 4. Molecular Biology of the Cell, 4th Edition.

Chromatograph of Sequence data



Sequencing by synthesis

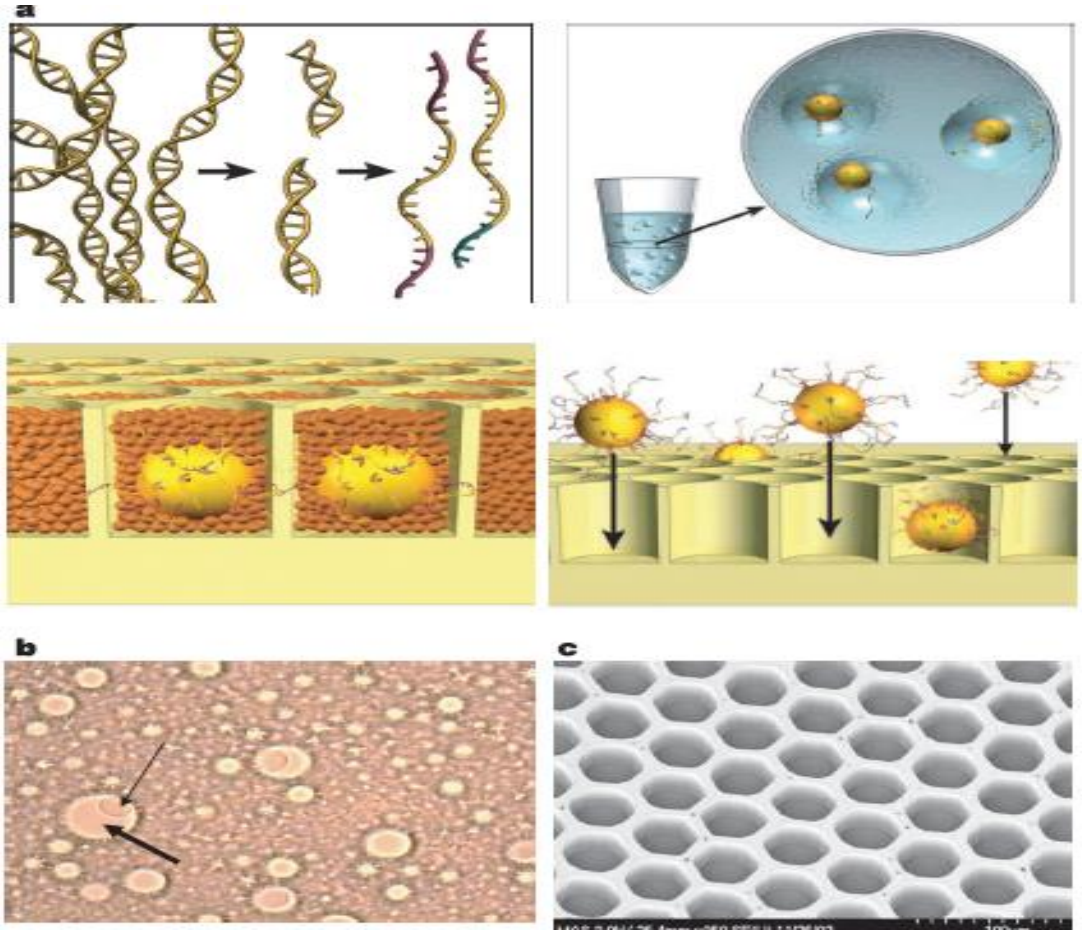


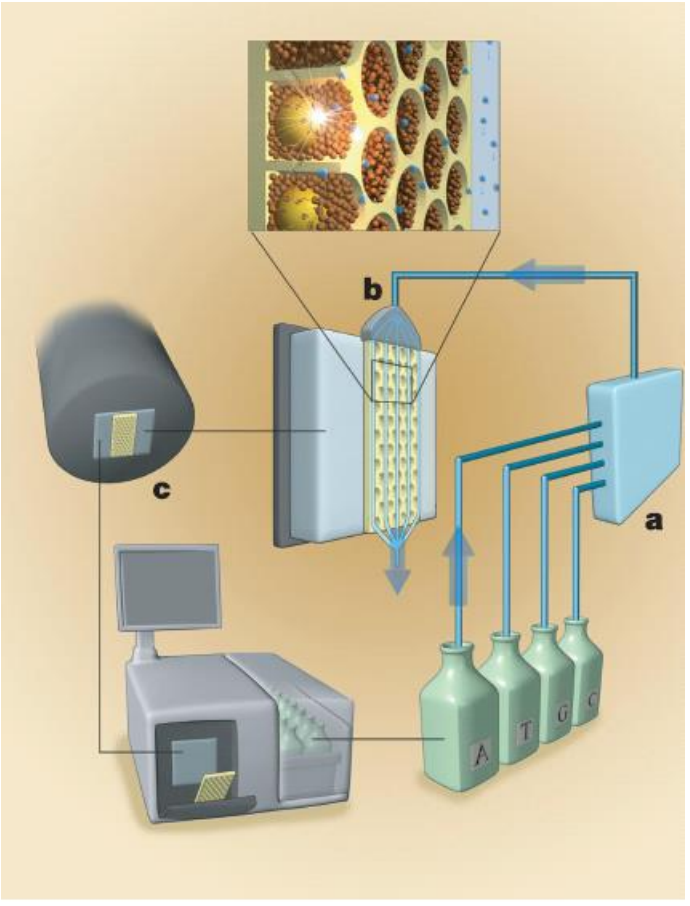
• Four nucleotides are added stepwise to the template hybridized to a primer.

• Incorporation of a deoxynucleotide, determined by complementing with the template, will release PPi and light, which can be detected by a CCD (charge-coupled device) camera

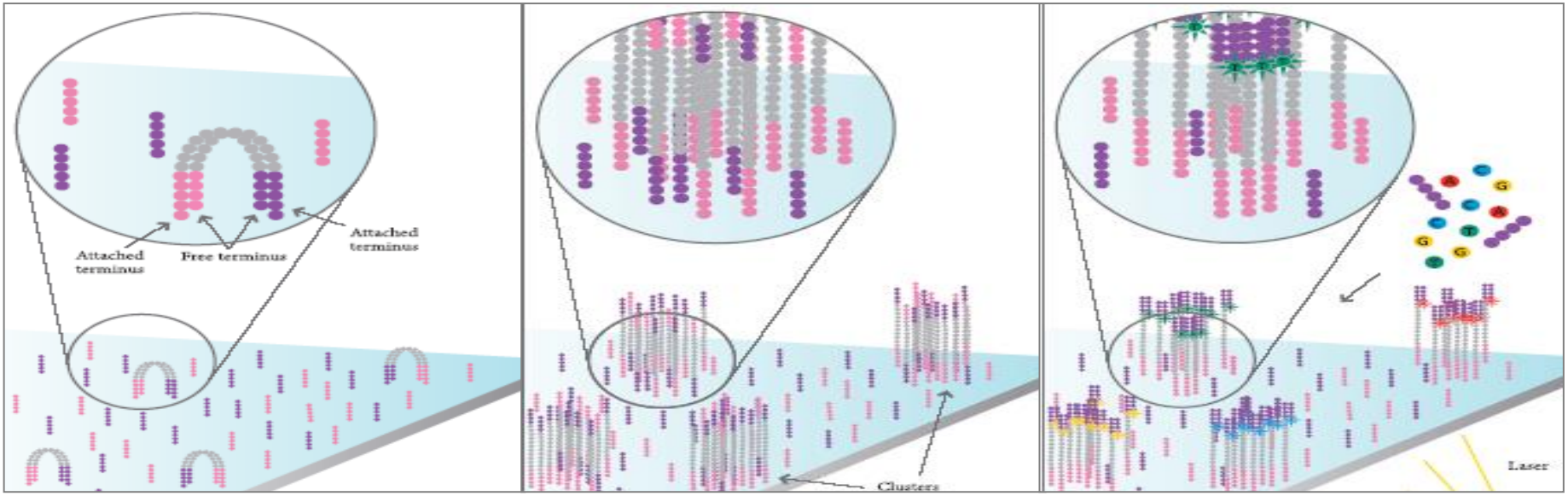
• Unincorporated deoxynucleotides and the produced ATP are degraded between each cycle by the nucleotide-degrading enzyme.

454 Sequencing





SBS Sequencing



SBS Sequencing**.** cDNA or genomic template (nebulized DNA) is prepared by the addition of 5' and 3' adapters. This is diluted and applied to the surface of a chip pre-coated with a dense lawn of primers **(A)**; "bridge-PCR" with unlabelled nucleotides is used to create a PCR colony covalently linked to the surface. This creates millions of dense clusters of dsDNA template **(B)**. Sequencing is performed by addition of four dye-labeled *reversible* nucleotide terminators along with the same primers and a custom polymerase. Laser excitation and image capture is used to determine the first nucleotide for all clusters in parallel **(C)**. The 3' blocked terminus is removed along with the fluorophore (dye), and the second bases are determined the same way. The process continues for 25-35 nucleotides. (Figure Modified from www.solexa.com).

The New York Times
nytimes.com

July 18, 2006

The Quest for the $1,000 Human Genome

By NICHOLAS WADE

As part of an intensive effort to develop a new generation of machines that will sequence DNA at a vastly reduced cost, scientists are decoding a new human genome — that of James D. Watson, the co-discoverer of the structure of DNA and the first director of the National Institutes of Health's human genome project.

The New York Times
nytimes.com

October 5, 2006

$10 Million Prize Set Up for Speedy DNA Decoding

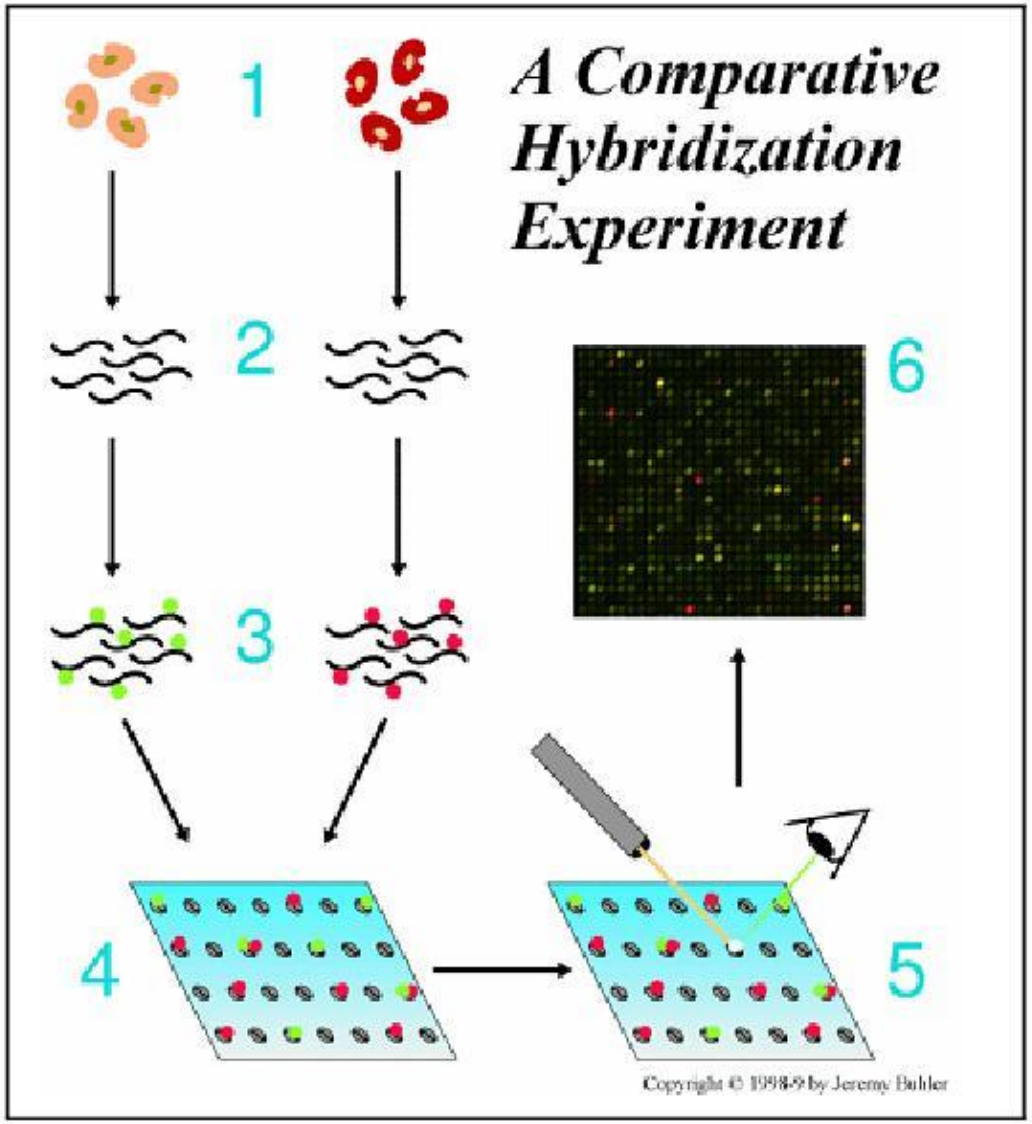
By NICHOLAS WADE

A $10 million prize for cheap and rapid sequencing of the human genome was announced today by the X Prize Foundation of Santa Monica, Calif.

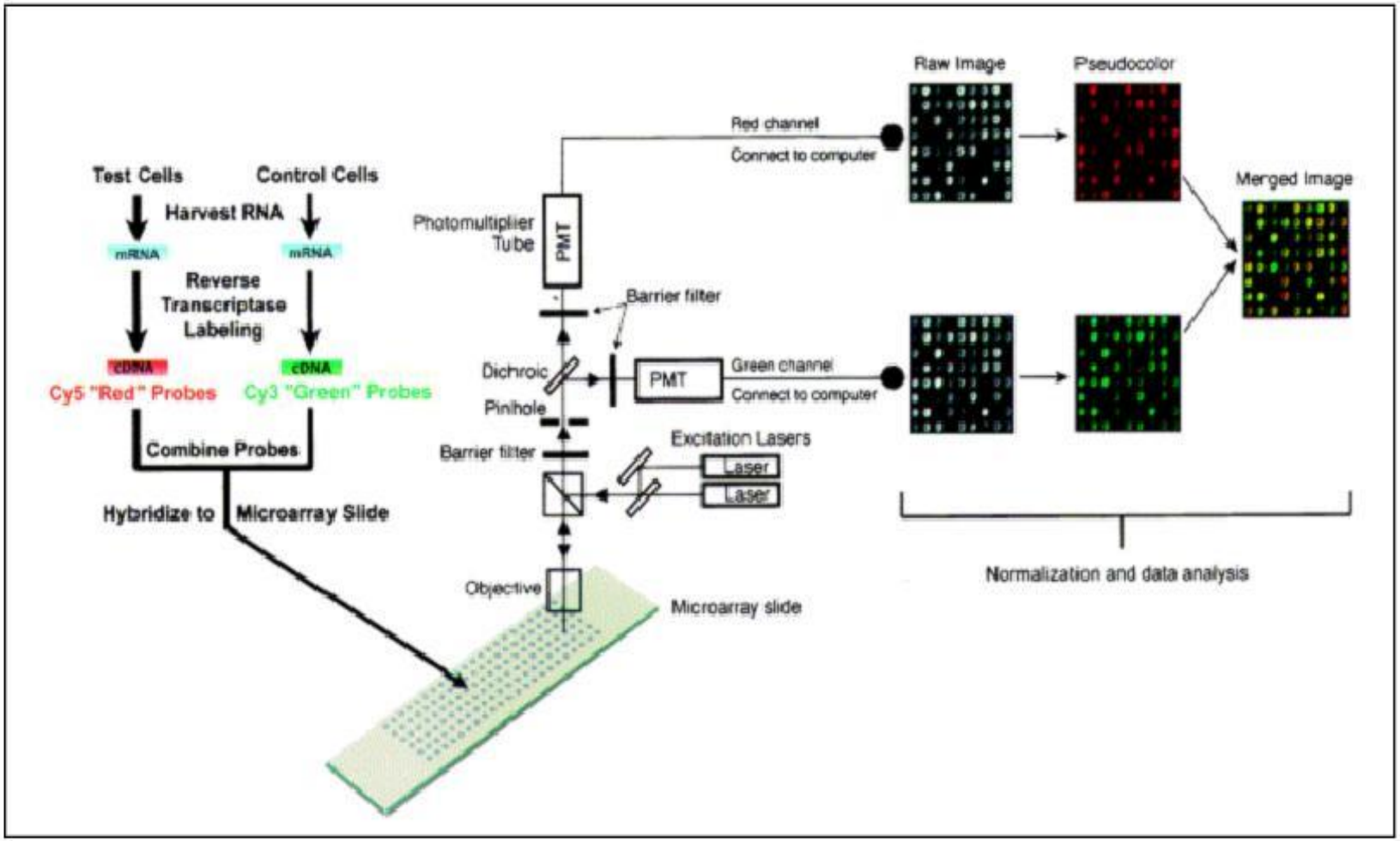
DNA Microarray, 2d gel, MSMS, yeast 2-hybrid.

Gene expression

- How many copies of a gene (its product) is present in the cell?
- For experimental reasons, gene expressions are measured by numbers of mRNAs, not directly by proteins. (See Proteomics)
- Various cell types are due to different genes expressed.
- The difference between diseased (e.g., cancerous) and non-diseased
- Diseased cells are often resulted from the abnormal levels of expression of key genes.

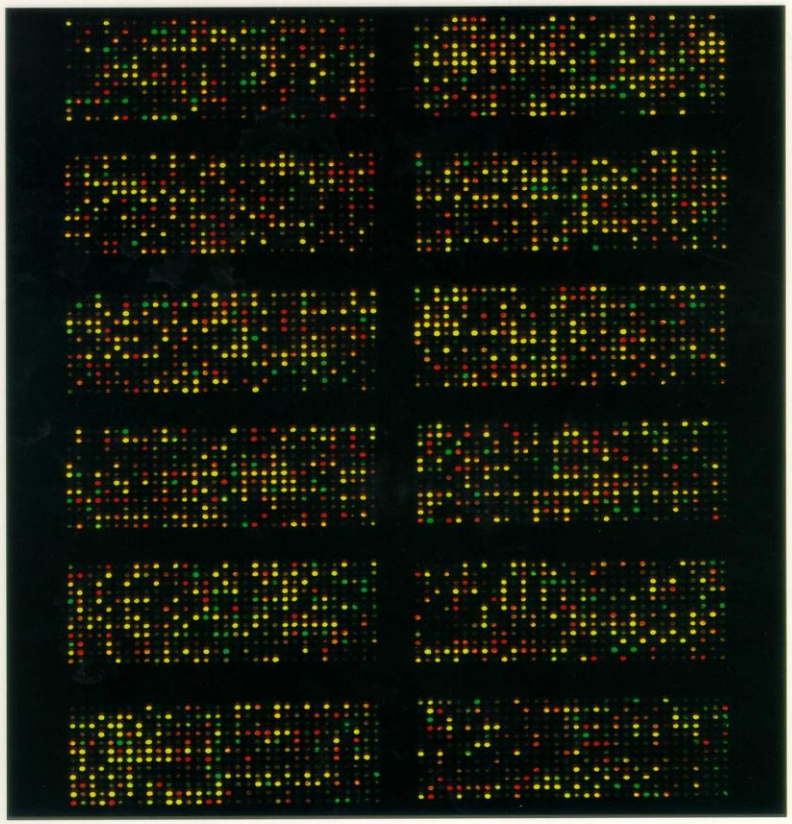

A Comparative Hybridization Experiment
1
2
3
4
5
6
Copyright © 1998-9 by Jeremy Buhler

- Microarray
  - Oligonucleotide (Affymetrix) array
    - Oligo (~ 25 bases long)
    - High density (1cm^2 contain 100k oligos)
  - cDNA array
    - cDNA (RT-PCR), much longer (> 1000 bases)
    - Varied density of cDNA on each spot, hybridization depends on length
    - Less possibility for false positives
  - Image processing
  - Background subtraction
  - Normalization


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

AUGUST 2000, VOLUME 182, NUMBER 16, PAGES 4375–4676



Journal of Bacteriology

AMERICAN SOCIETY FOR MICROBIOLOGY

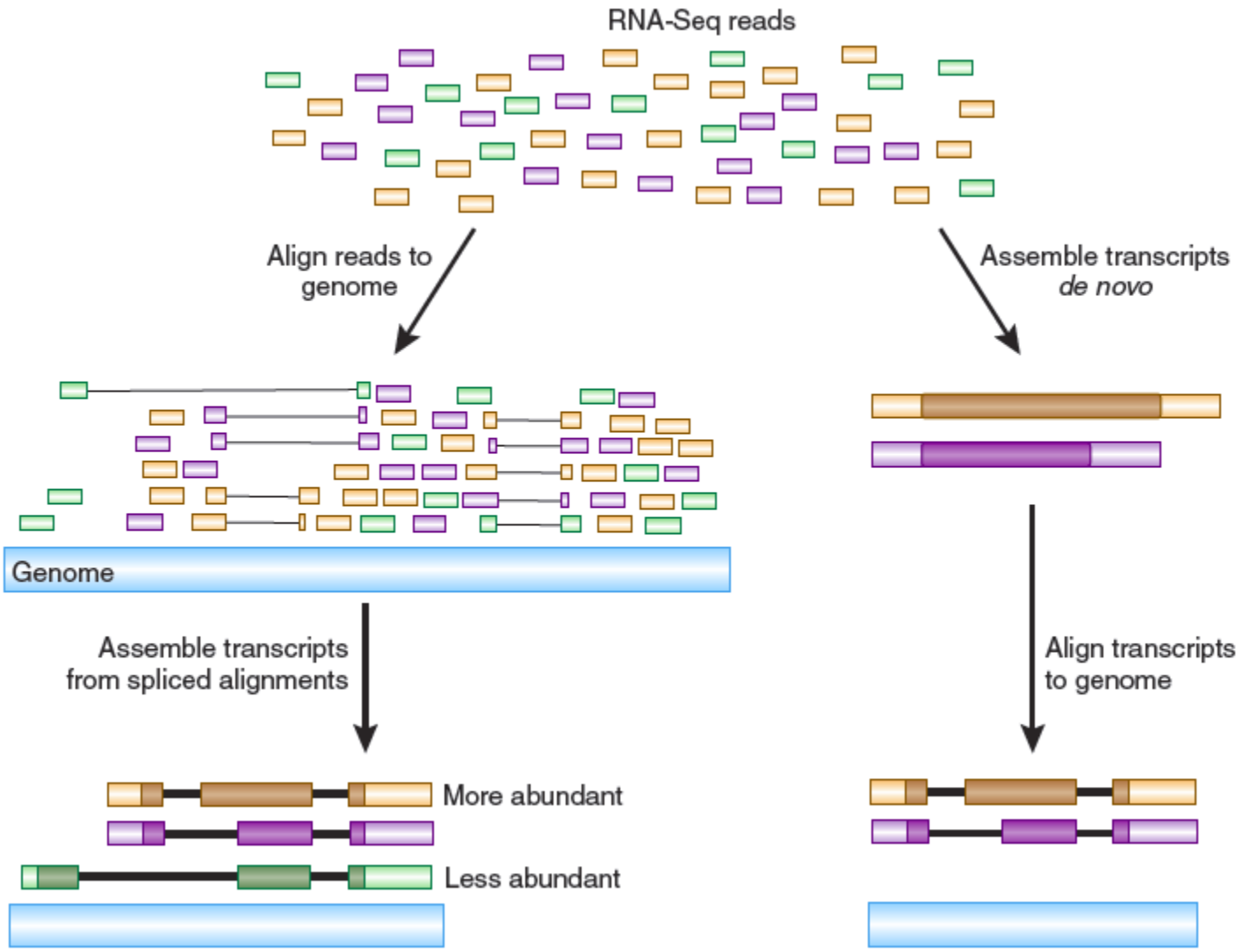
AUGUST 2000, VOLUME 182, NUMBER 16

Published Twice Monthly

Applications

- Inferring transcription regulatory networks
- Understanding correlation between genotype and phenotype
- predicting genotype <=> phenotype
- Phenotypes:
  - drug/therapy response
  - drug-drug interactions for expression
  - drug mechanism
  - interacting pathways of metabolism

Transcriptome by RNA-seq



What is proteomics?

- Like genomics is the study of all genes in a genome, proteomics is the study of all proteins of a cell at a given time.
- Three aspects
  - Biological process (why is this being done? e.g. movement of cell)
  - Molecular function (what kind of molecule is this? e.g., ATPase)
  - Cellular component (where is this located? e.g., ribosome)

Why is it difficult?

- Moving target
  - Cell-to-cell variations
  - Cell behavior changes with time
- Lack of high throughput technology
  - Protein chips? Protein sequences do not have hybridization that DNA sequences have.

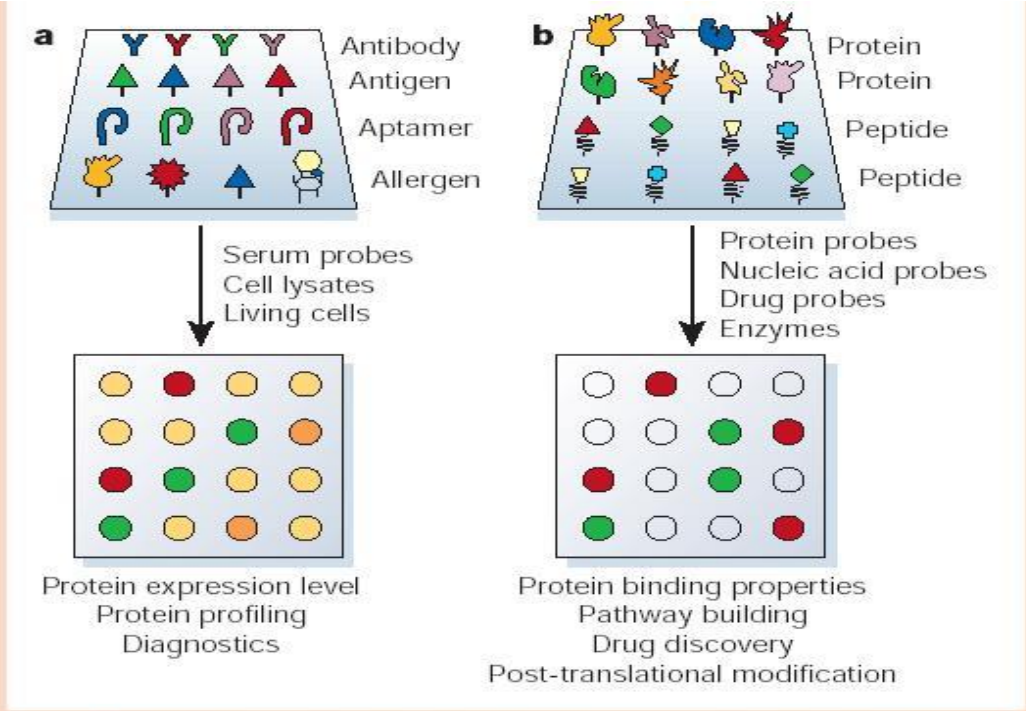


Figure 1 Analytical versus functional protein microarrays. **a**, Analytical protein microarray. Different types of ligands, including antibodies, antigens, DNA or RNA aptamers, carbohydrates or small molecules, with high affinity and specificity, are spotted down onto a derivatized surface. These chips can be used for monitoring protein expression level, protein profiling and clinical diagnostics. Similar to the procedure in DNA microarray experiments, protein samples from two biological states to be compared are separately labelled with red or green fluorescent dyes, mixed, and incubated with the chips. Spots in red or green colour identify an excess of proteins from one state over the other. **b**, Functional protein microarray. Native proteins or peptides are individually purified or synthesized using high-throughput approaches and arrayed onto a suitable surface to form the functional protein microarrays. These chips are used to analyse protein activities, binding properties and post-translational modifications. With the proper detection method, functional protein microarrays can be used to identify the substrates of enzymes of interest.

2D gel electrophoresis

- Isoelectric points (first dimension)
- Molecular weights (second dimension)

Both pI and MW are functions of amino acid sequence of a protein.

Some proteins do not resolve well by 2D gels.

Issues:

- Detection of spots (image processing)
- Quantification of each spot
- Identification of each spot (Mass Spectrometry)

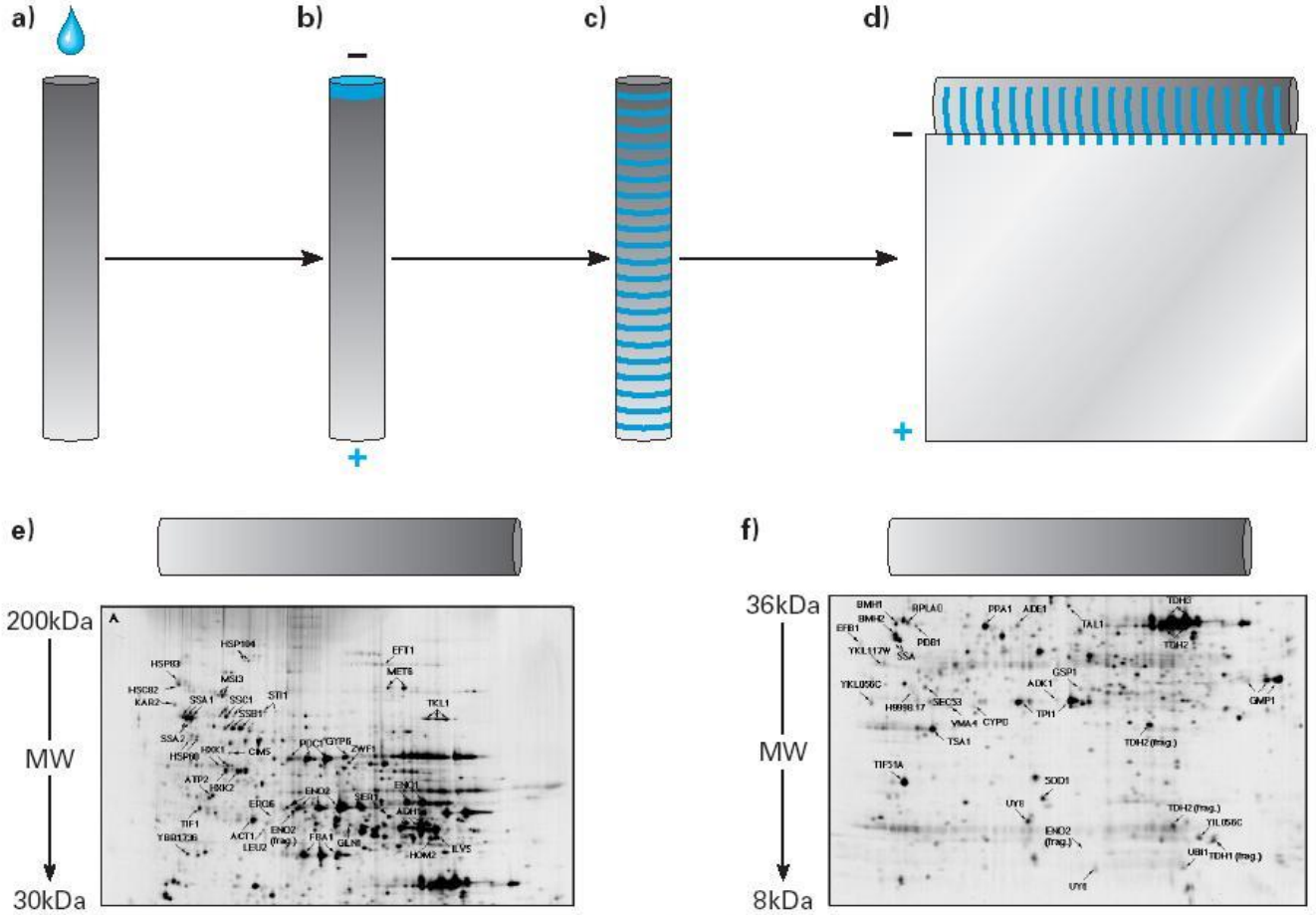
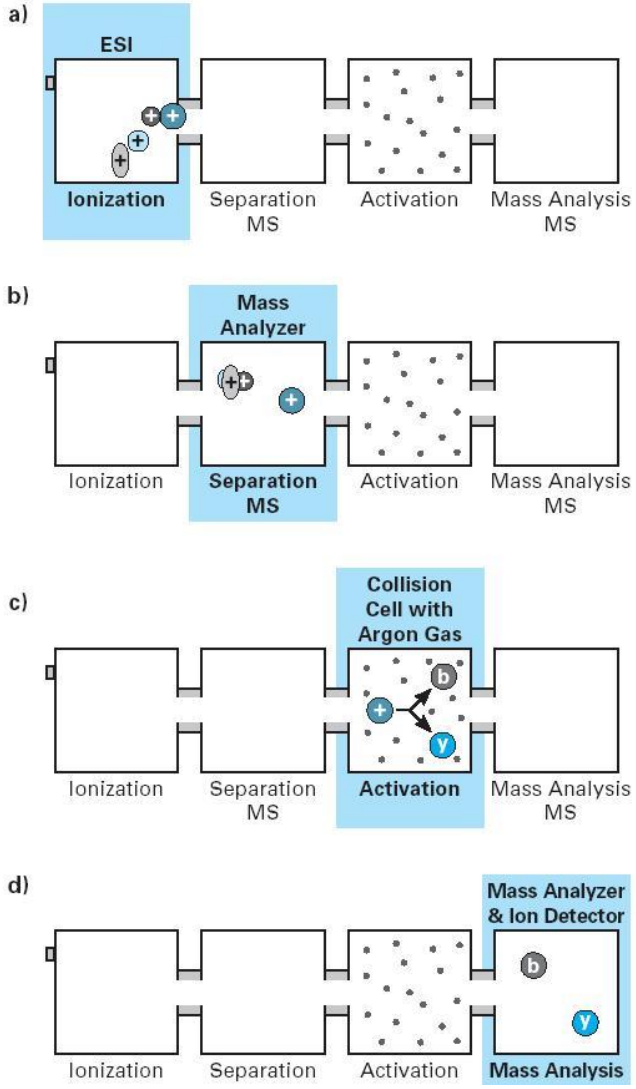


FIGURE 6.18 • Two-dimensional (2D) gel electrophoresis. Each column (a–c) with a gradation of gray shading represents the isoelectric focusing gel with a pH gradient. **a)** A mixture of proteins (blue drop) is applied to the isoelectric focusing gel and **b)** exposed to an electrical current. **c)** Proteins migrate to their isoelectric points (pI) and stop moving. **d)** This tubular gel is placed on top of a slab polyacrylamide gel that contains SDS and is subjected to electrophoresis (SDS-PAGE). Proteins migrate into the slab gel according to their molecular weights. Yeast cells were grown in rich media and subjected to 2D gel analysis. Using duplicate isoelectric focusing gels, large **e)** and small **f)** proteins were analyzed on separate gels. The same spots appear at the bottom of e) and the top of f). Molecular weights are resolved on the Y-axis and pIs on the X-axis. Panels e) and f) are from the Swiss 2D database at ExPASy.


a)
ESI
Ionization
Separation MS
Activation
Mass Analysis MS
b)
Mass Analyzer
Ionization
Separation MS
Activation
Mass Analysis MS
c)
Collision Cell with Argon Gas
b
y
Ionization
Separation MS
Activation
Mass Analysis MS
d)
Mass Analyzer & Ion Detector
b
y
Ionization
Separation
Activation
Mass Analysis

a)

S–P–A–F–D–S–I–M–A–E–T–L–K
(protonated mass 1410.6)

Mass$^+$	b-ions	y-ions	Mass$^+$
81.1	S	PAFDSIMAETLK	1323.6
185.2	SP	AFDSIMAETLK	1226.4
256.3	SPA	FDSIMAETLK	1155.4
403.5	SPAF	DSIMAETLK	1008.2
518.5	SPAFD	SIMAETLK	893.1
605.6	SPAFDS	IMAETLK	806.0
718.8	SPAFDSI	MAETLK	692.3
850.0	SPAFDSIM	AETLK	561.7
921.1	SPAFDSIMA	ETLK	490.6
1050.2	SPAFDSIMAE	TLK	361.5
1151.3	SPAFDSIMAET	LK	260.4
1264.4	SPAFDSIMAETL	K	147.2

b)

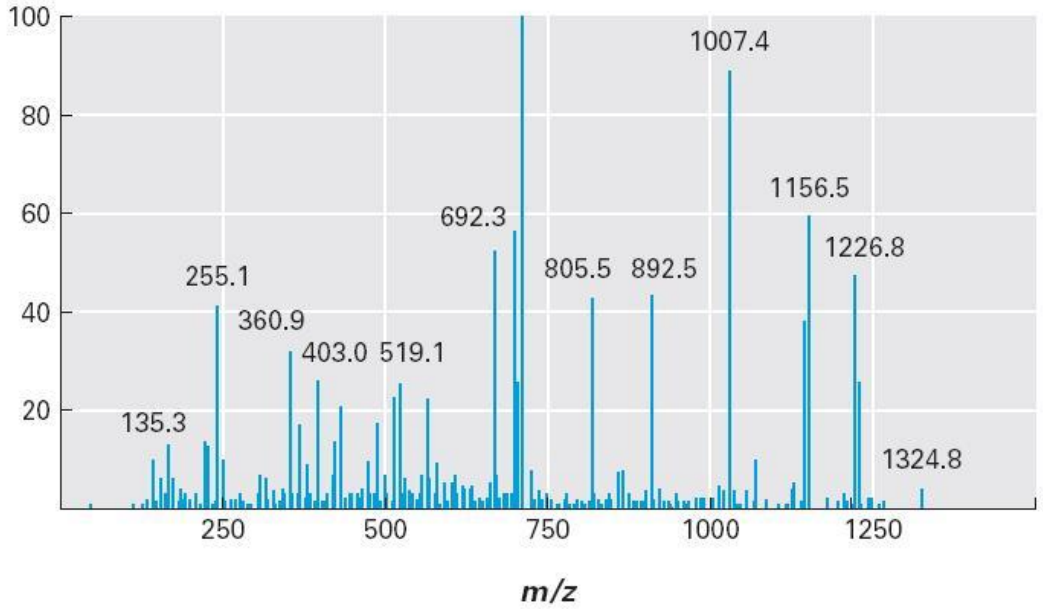


FIGURE 6.20 • Protein identification through peptide fragment formation and separation. When a group of identical proteins is broken into its peptide pieces, more than one pair of b and y peptides will be formed. **a)** One protein sequence and its calculated mass on top, with the b peptides/masses (gray) and the y peptides/masses (blue) below. **b)** An experimentally determined mass/charge spectrum from the peptide in a). Notice that some peaks are higher than others, which means that some b/y peptide pieces were more abundant than others. The spectrum is used to determine the peptide's amino acid sequence and protein identity.

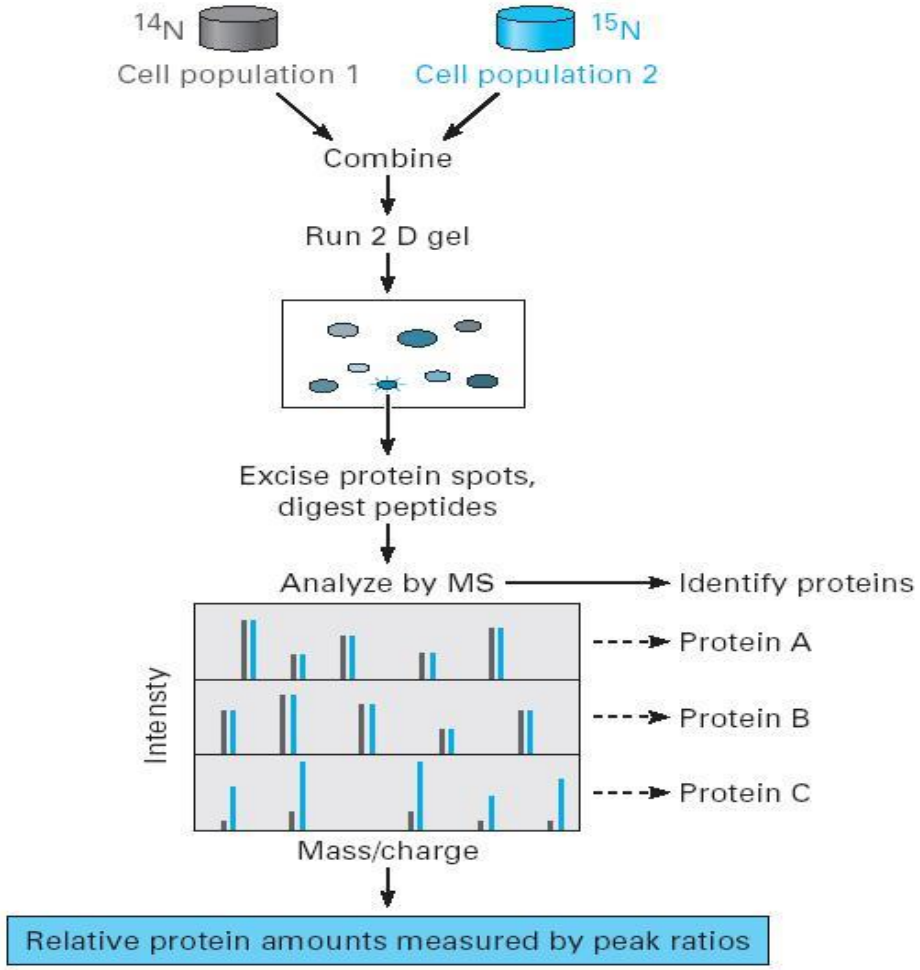


FIGURE 6.24 • Quantifying differences in proteomes. Two cell pools are grown in the presence (blue = cell population 2) or absence (gray = cell population 1) of heavy nitrogen ^{15}N. The proteins are extracted, pooled, and subjected to 2D gel analysis; spots excised; and proteins identified and quantified by MS/MS. The relative areas under the pairs of heavy and light peptide peaks indicate relative abundance of each protein pair.

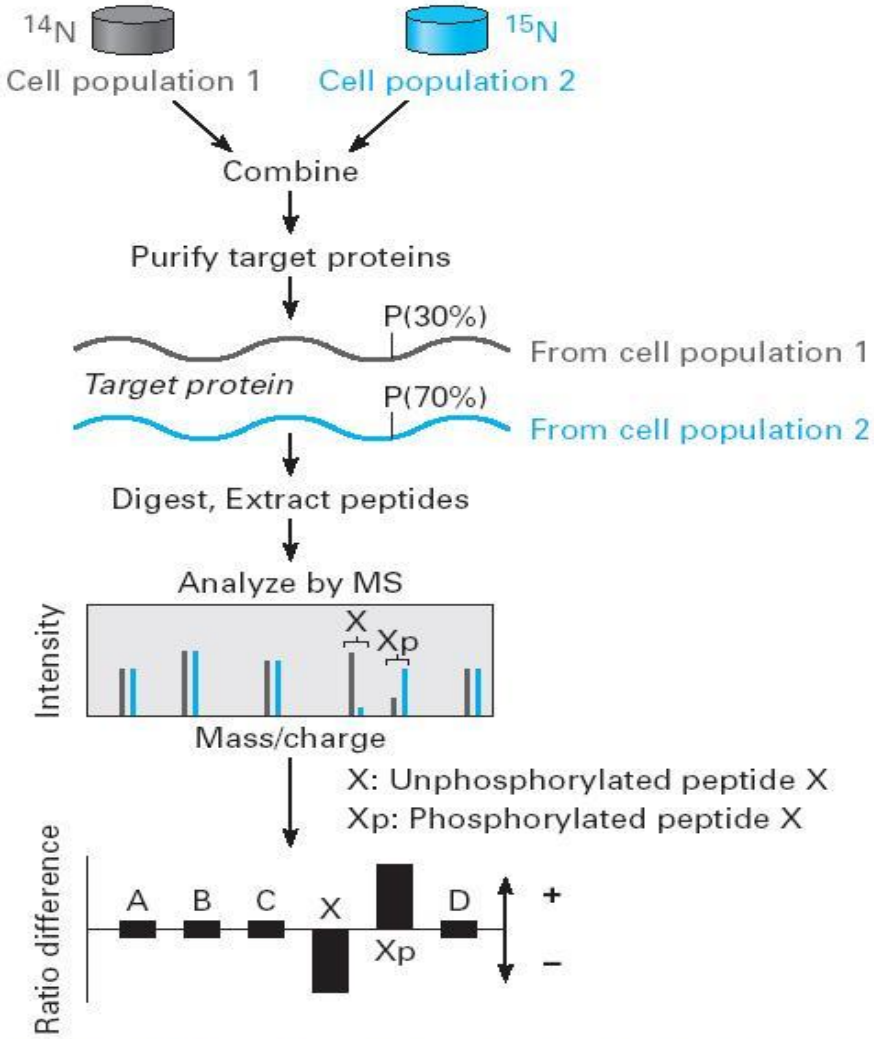


FIGURE 6.25 • Quantifying relative levels of phosphorylation. Bar graphs A, B, C, and D indicate no difference for proteins A–D in the two populations. Compared to population 1, unphosphorylated protein X was reduced in population 2. Phosphorylated protein X (Xp) was increased in population 2.

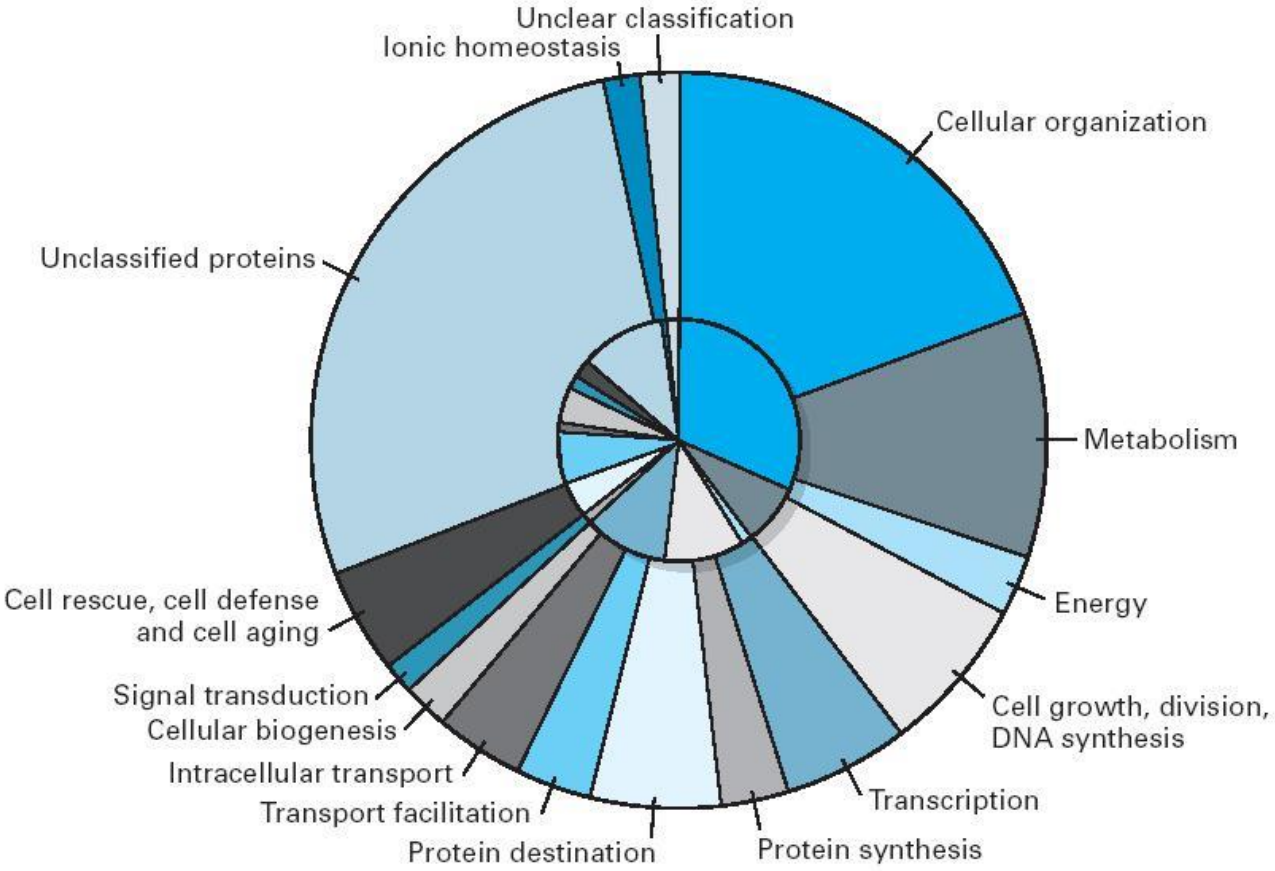


FIGURE 6.9 • Bar code analysis of biological processes. Distribution of functional classes of essential (inner circle) and nonessential (outer circle) genes using criteria from the Munich Information Center for Protein Sequences (MIPS).

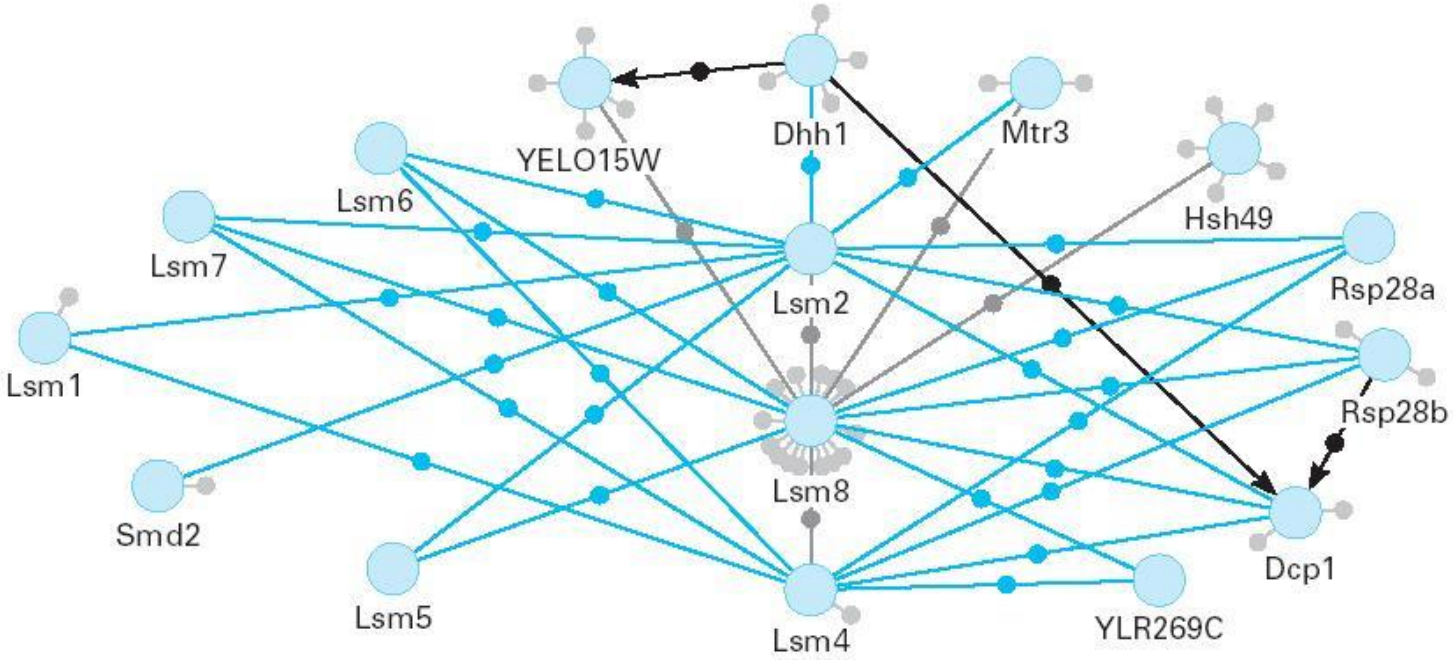


FIGURE 6.15 • Proteomics circuit showing the interactions of RNA splicing proteins. The proteins are indicated by large blue nodes and their interactions with lines. The dots on the lines help you follow each line. Blue lines show interactions that were detected by traditional Y2H array screens, black from multiple high-throughput screens, and gray from literature and array screens. The black arrows point away from the protein used as bait in the screens. Small gray nodes indicate other protein-protein interactions not highlighted here.

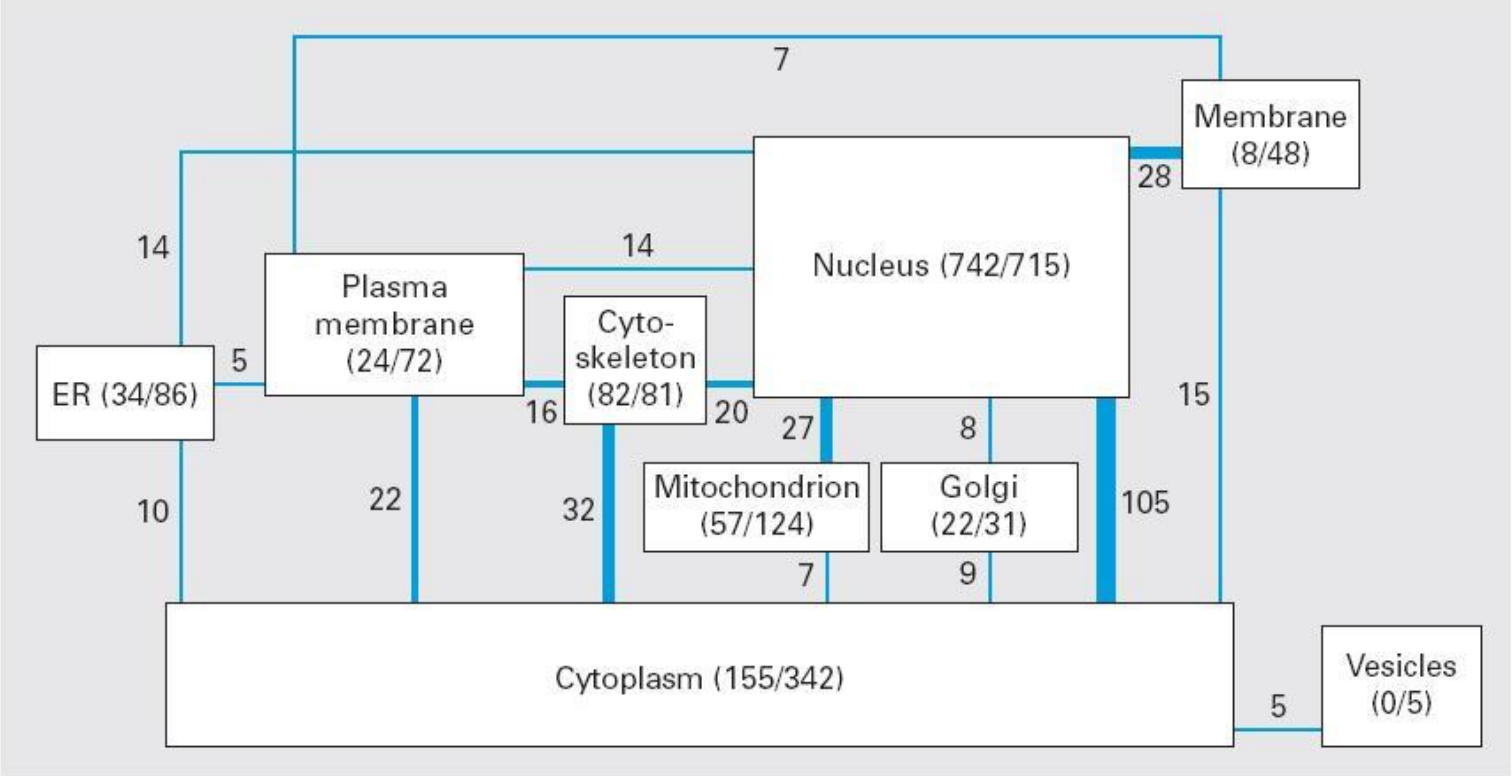


FIGURE 6.17 • Protein interactions grouped by cellular compartments. Numbers in parentheses indicate the number of interactions in the circuit diagram among proteins of this compartment/the total number of proteins in the same compartment. Lines connecting compartments indicate the number of protein interactions by the thickness of each line (numbers of connections are near each line). For example, there are 7 interactions between the 48 membrane proteins and the 72 plasma membrane proteins in "Benno Figure 1."

Yeast two hybrid System 1

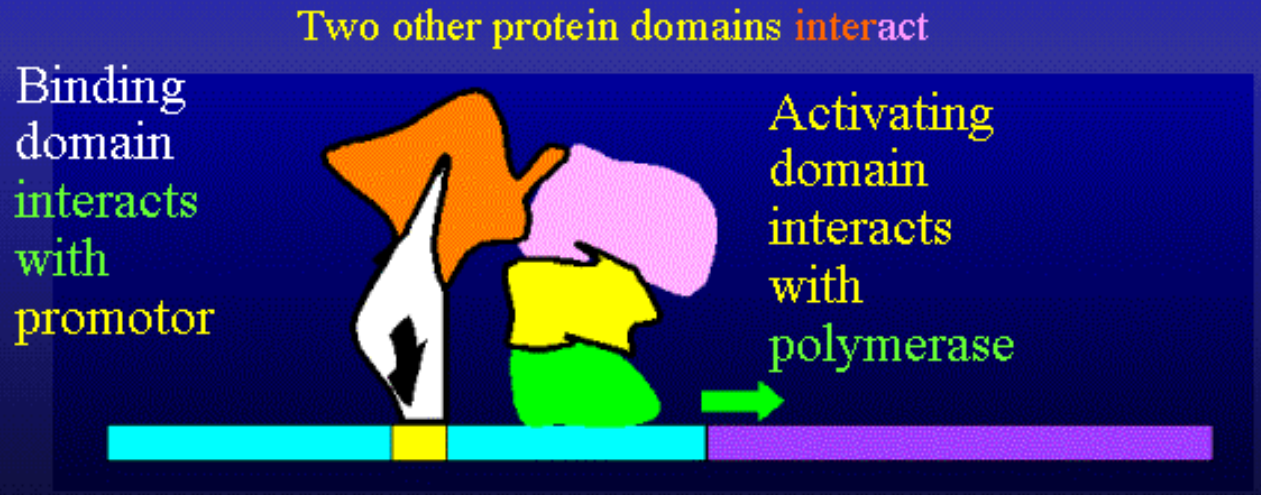
Gal4 protein: comprises DNA binding and activating domains

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



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

David B. Collinge KVL

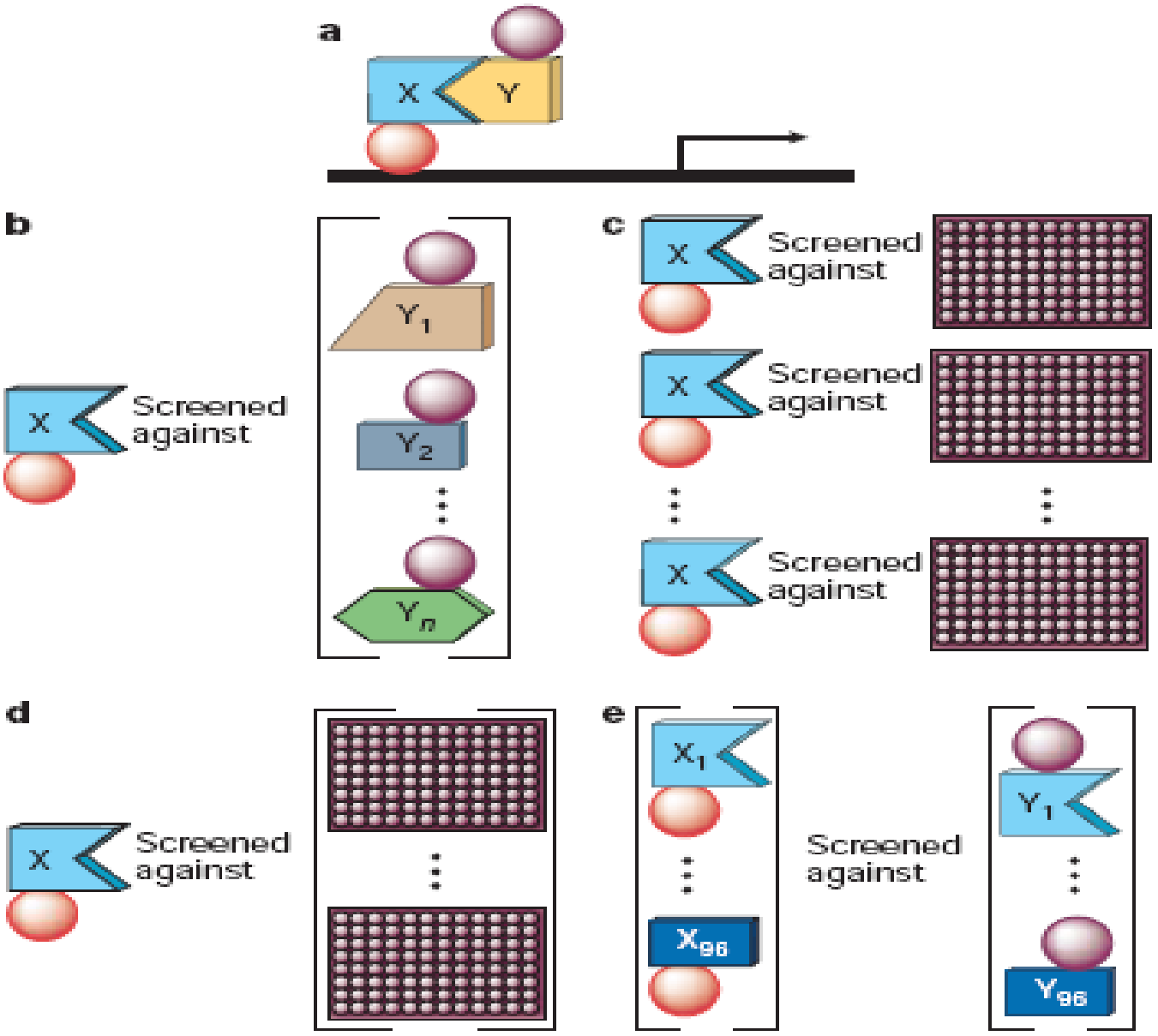
Yeast two hybrid System (3)

- 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


a
X
Y
b
X
Screened
against
Y_1
Y_2
Y_n
c
X
Screened
against
X
Screened
against
X
Screened
against
d
X
Screened
against
e
X_1
X_{96}
Screened
against
Y_1
Y_{96}