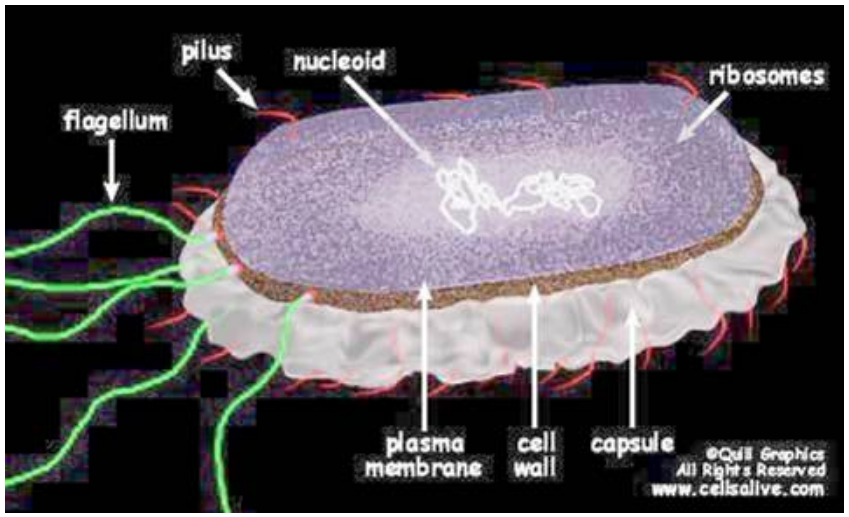
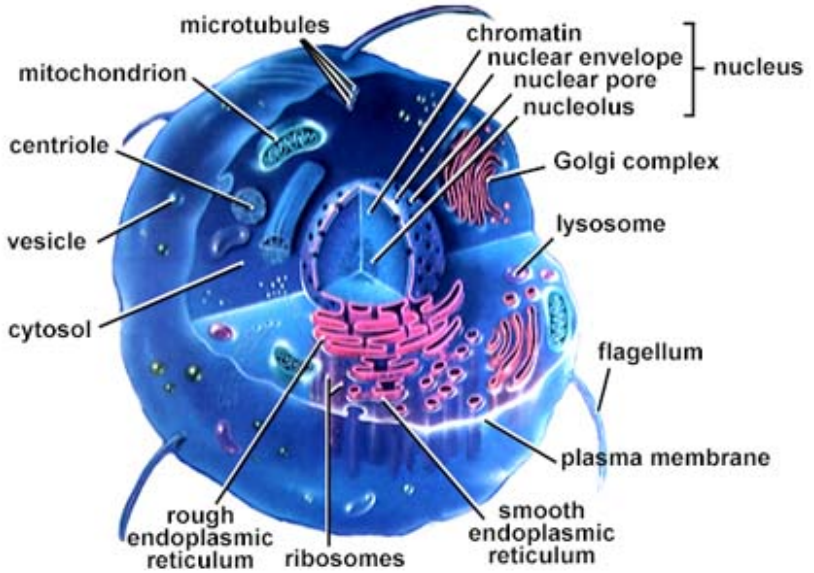
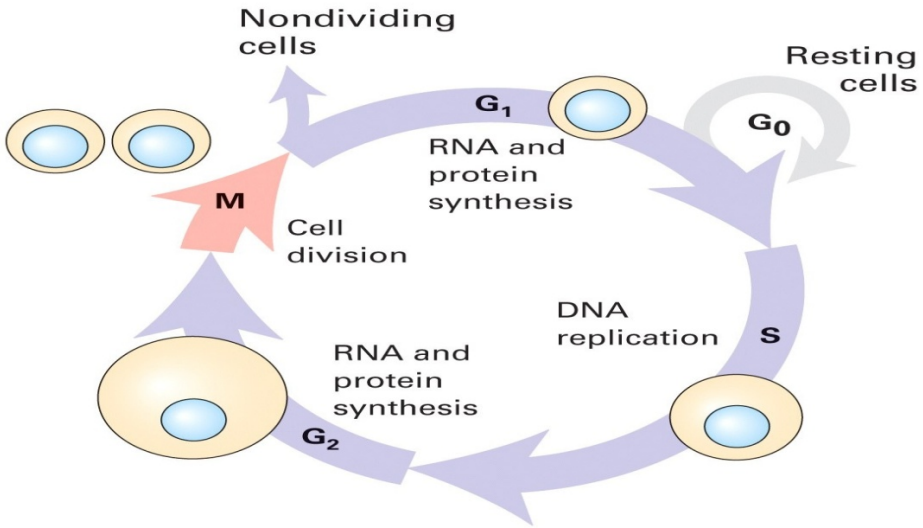
SUMMARY

2 types of cells: Prokaryotes v.s. Eukaryotes

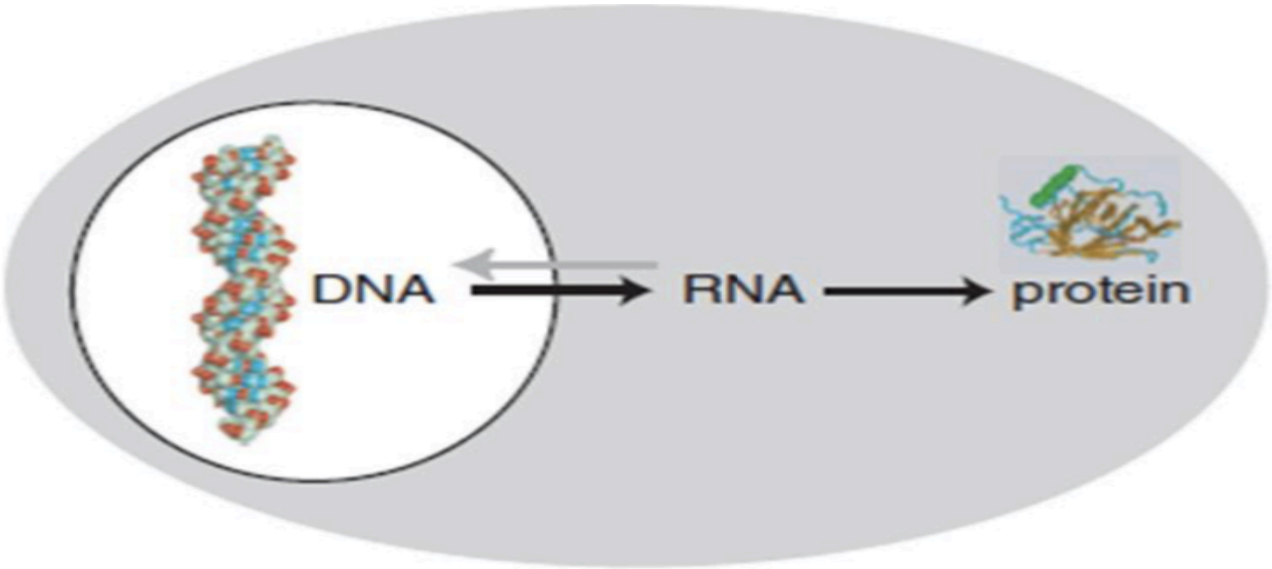


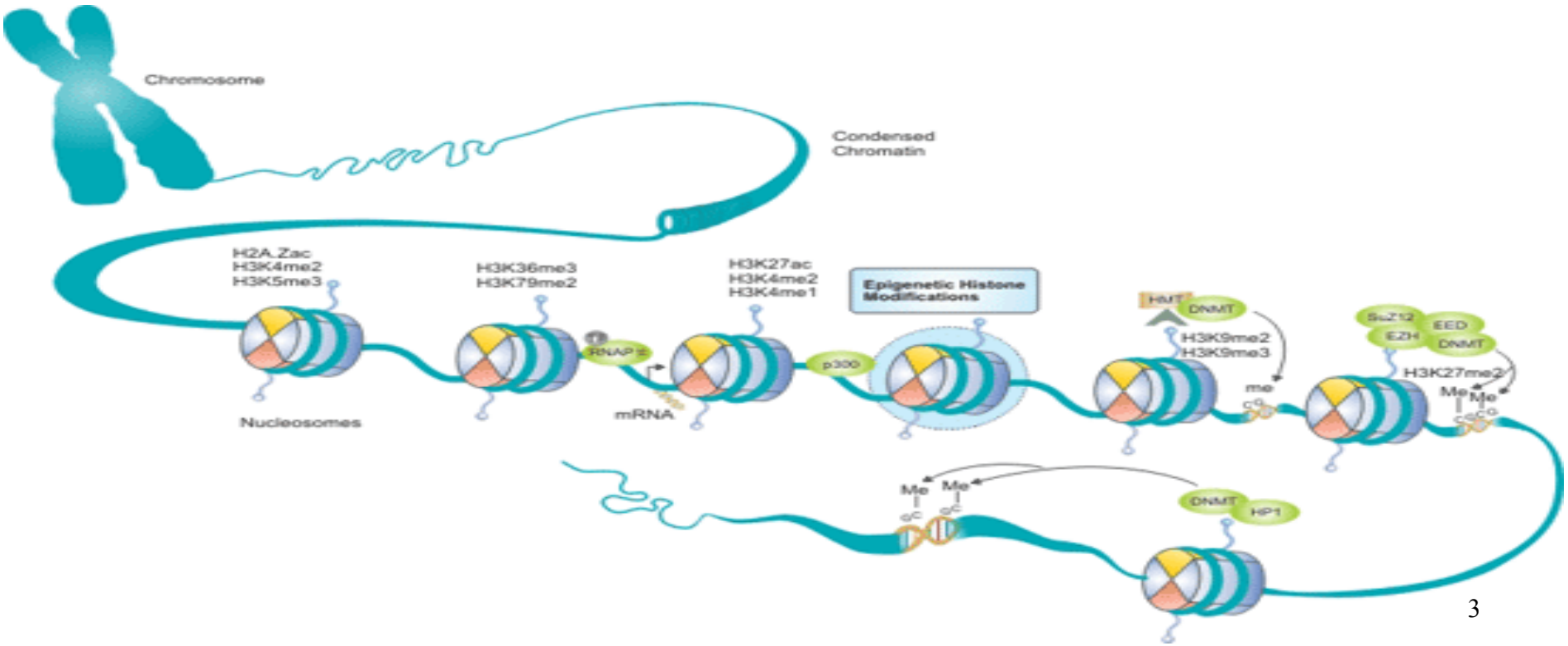


All cells have common cycles



- Born, eat, replicate, and die


DNA
RNA
protein

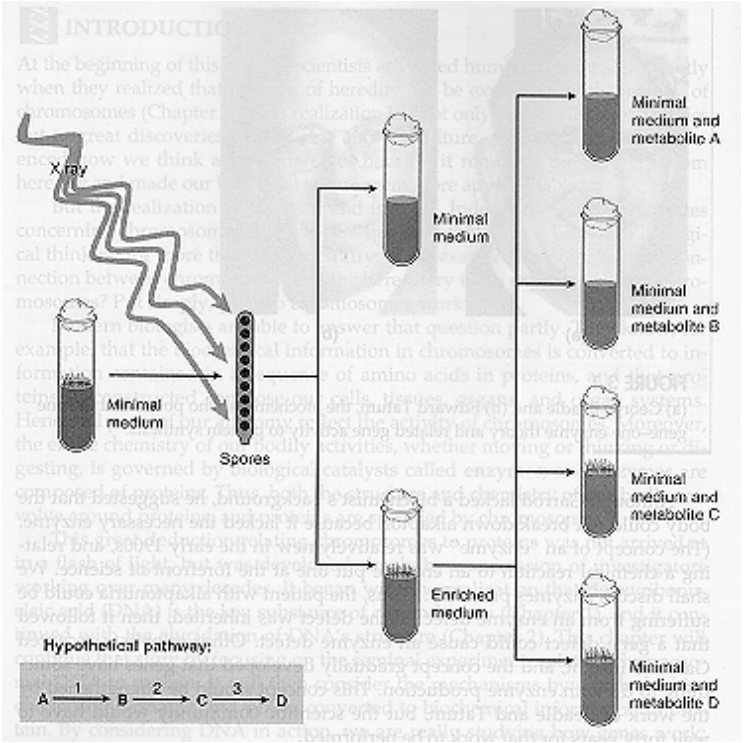

Chromosome
Condensed Chromatin
H2A.Zac
H3K4me2
H3K5me3
H3K36me3
H3K79me2
H3K27ac
H3K4me2
H3K4me1
Epigenetic Histone Modifications
RNAP II
mRNA
p300
HMT
DNMT
H3K9me2
H3K9me3
me
SuZ12
EED
EZH
DNMT
H3K27me2
Me
Me
Nucleosomes
DNMT
HP1

What Do Genes Do?

- ***Beadle and Tatum Experiment***
- ***Design of Life*** (gene->protein)
- protein synthesis
  - Central dogma of molecular biology

Beadle and Tatum Experiment

- Experiment done at Stanford University 1941
- The hypothesis: One gene specifies the production of one enzyme
- They chose to work with bread mold (*Neurospora)* biochemistry already known (worked out by Carl C. Lindegren)
  - easy to grow, maintain
  - short life cycle
  - easy to induce mutations
  - easy to identify and isolate mutants

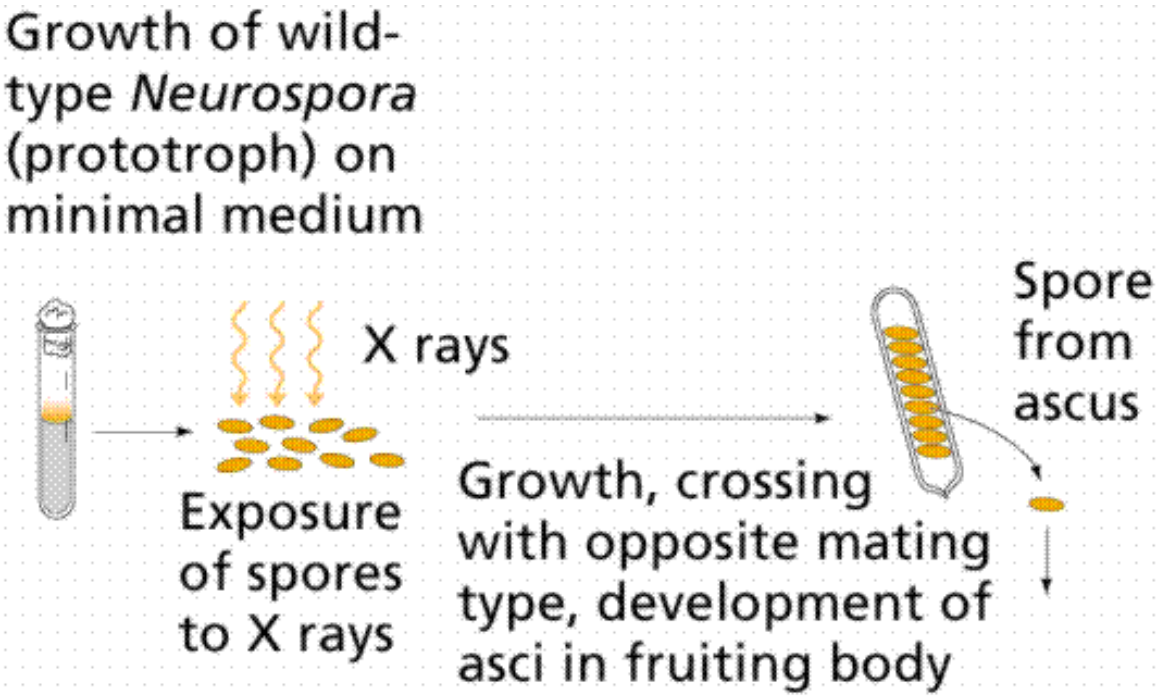


Beadle and Tatum Experiment Procedure

- 2 different growth media:
  - Complete - consists of agar, inorganic salts, malt & yeast extract, and glucose
  - Minimal - consists of agar, inorganic salts, biotin, disaccharide and fat

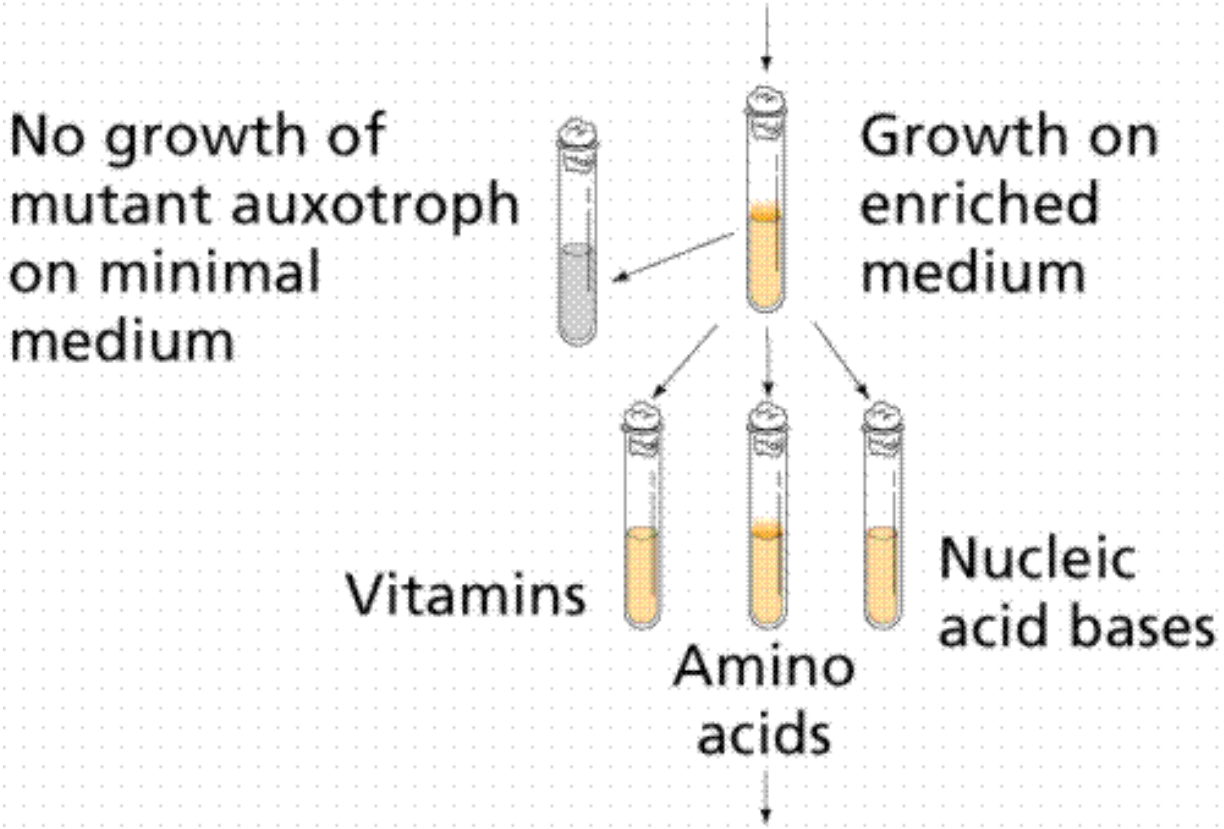
- X-ray used to irradiate Neurospora to induce mutation
- Mutated spores placed onto minimal medium

Beadle and Tatum Experiment Procedure



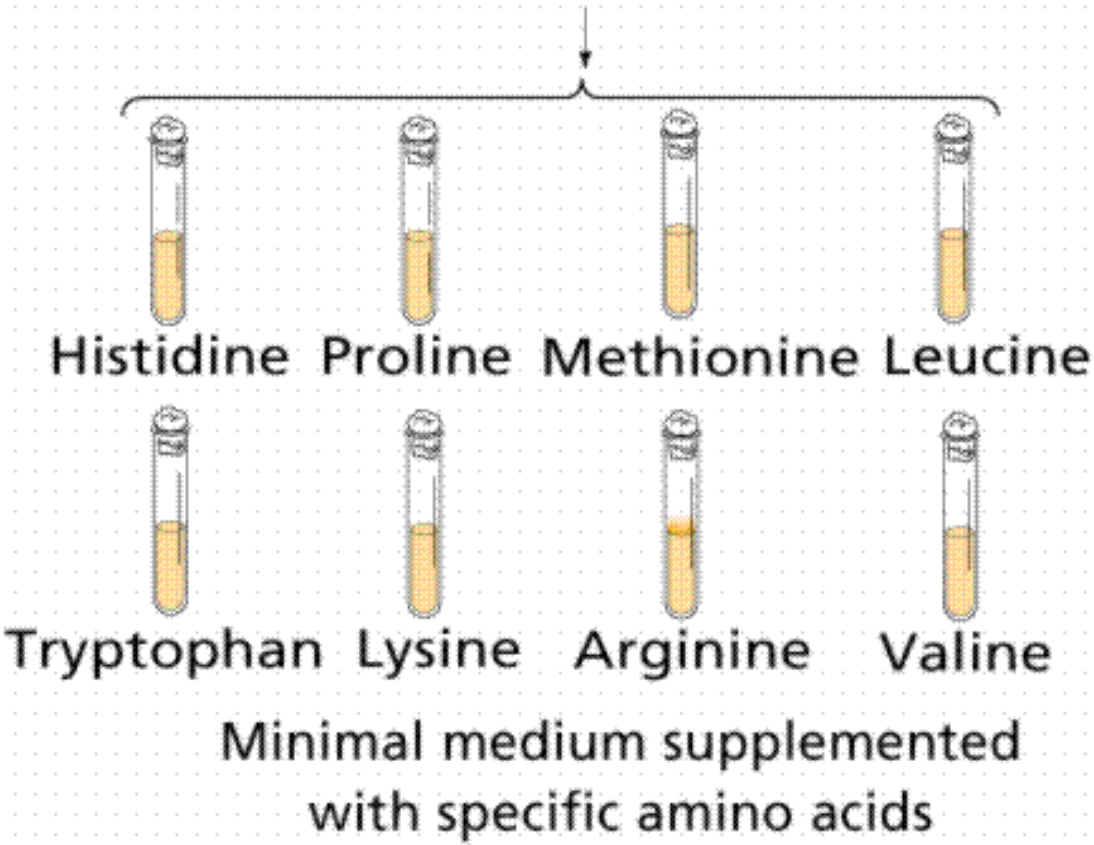
Images from Purves et al., Life: The Science of Biology, 4th Edition, by Sinauer Associates

Beadle and Tatum Experiment Procedure



Images from Purves et al., Life: The Science of Biology, 4th Edition, by Sinauer Associates

Beadle and Tatum Experiment Procedure



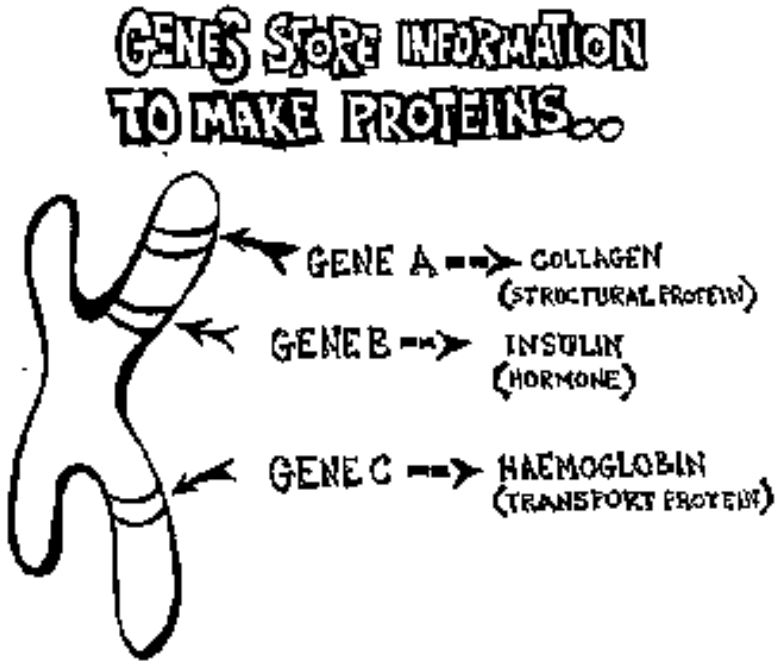
Images from Purves et al., Life: The Science of Biology, 4th Edition, by Sinauer Associates

Beadle and Tatum Experiment Conclusions

- Irradiated Neurospora survived when supplemented with Vitamin B6
- X-rays damaged genes that produces a protein responsible for the synthesis of Vitamin B6
- three mutant strains - substances unable to synthesize (Vitamin B6, Vitamin B1 and Para-aminobenzoic acid) essential growth factors
- crosses between normal and mutant strains showed differed by a single gene
- hypothesized that there was more than one step in the synthesis of Vitamin B6 and that mutation affects only one specific step
- Evidence: One gene specifies the production of one enzyme!

Genes Make Proteins

- genome-> genes ->protein(forms cellular structural & life functional)->pathways & physiology



Proteins: Workhorses of the Cell

- 20 different **amino acids**
  - different chemical properties cause the protein chains to fold up into specific three-dimensional structures that define their particular functions in the cell.
- Proteins do all essential work for the cell
  - build cellular structures
  - digest nutrients
  - execute metabolic functions
  - Mediate information flow within a cell and among cellular communities.
- Proteins work together with other proteins or nucleic acids as "molecular machines"
  - structures that fit together and function in highly specific, lock-and-key ways.

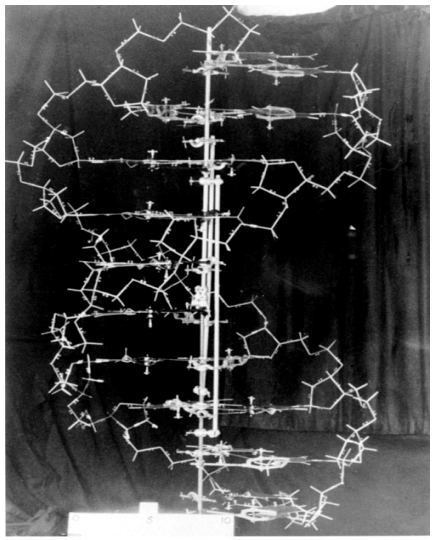
What Molecule Codes For Genes?

- ***Discovery of the Structure of DNA***
  - ***Watson and Crick***

- ***DNA Basics***

Discovery of DNA

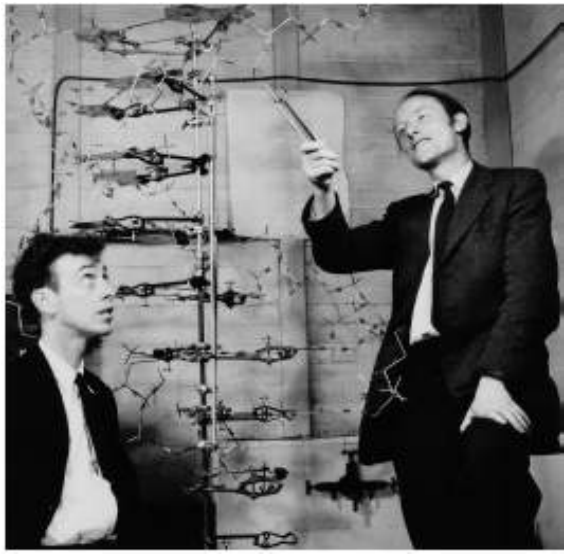
- DNA Sequences
  - Chargaff and Vischer, 1949
    - DNA consisting of A, T, G, C
      - Adenine, Guanine, Cytosine, Thymine
  - Chargaff Rule
    - Noticing #A≈#T and #G≈#C
      - A "strange but possibly meaningless" phenomenon.
- Wow!! A Double Helix
  - Watson and Crick, *Nature,* April 25, 1953
  - 1 Biologist
    1 Physics Ph.D. Student
    \+ 900 words
    = Nobel Prize
  - Rich, 1973
    - Structural biologist at MIT.
    - DNA's structure in atomic resolution.



Crick Watson

Watson & Crick – *"…the secret of life"*

- Watson: a zoologist, Crick: a physicist
- "*In 1947 Crick knew no biology and practically no organic chemistry or crystallography..*" – www.nobel.se
- Applying Chagraff's rules and the X-ray image from Rosalind Franklin, they constructed a "tinkertoy" model showing the double helix
- Their 1953 *Nature* paper: *"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."*

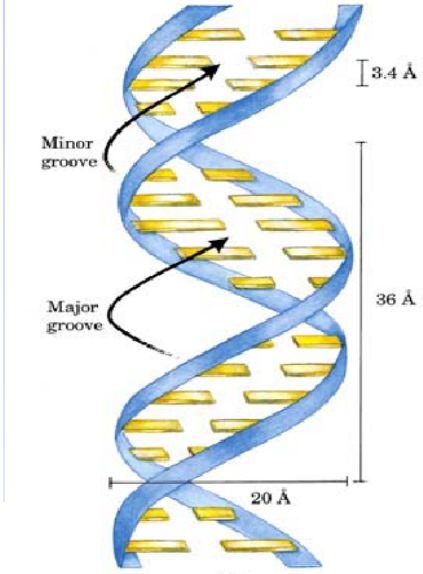


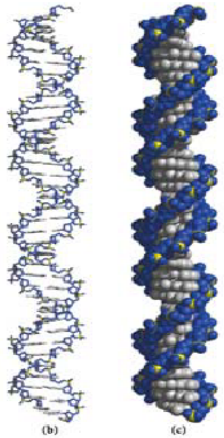
Watson & Crick with DNA model

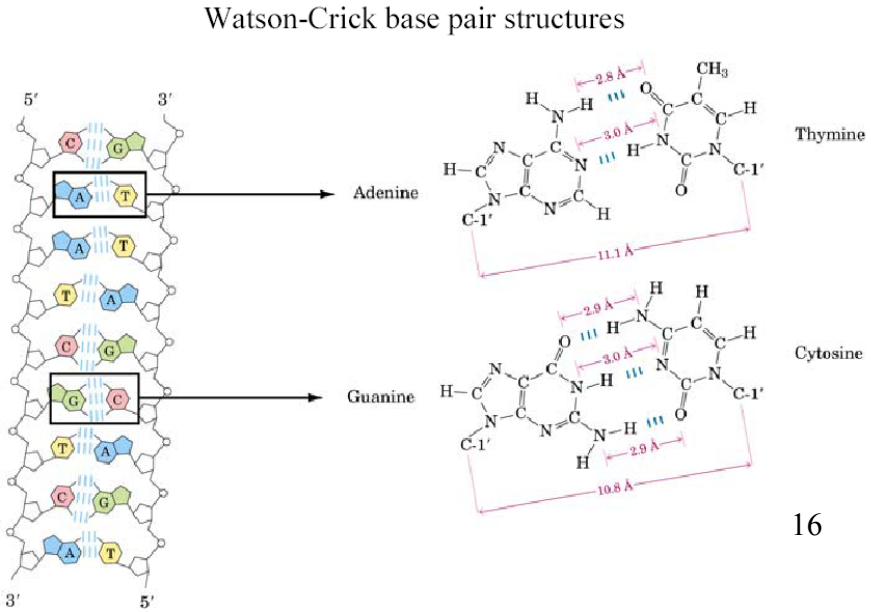
Rosalind Franklin with X-ray image of DNA

DNA: The Basis of Life

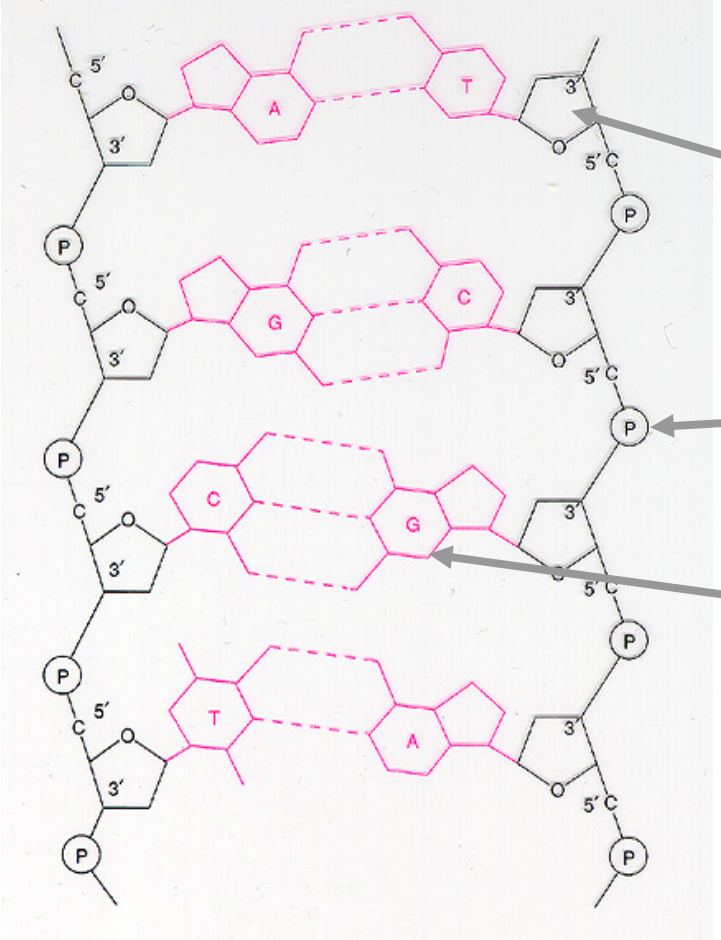
- Stores all information of life
- Deoxyribonucleic Acid (DNA)
  - Double stranded with complementary strands A-T, C-G
- DNA is a polymer
  - Sugar-Phosphate-Base
  - Bases held together by H bonding to the opposite strand







DNA, continued



Sugar

Phosphate

Base (A,T, C or G)

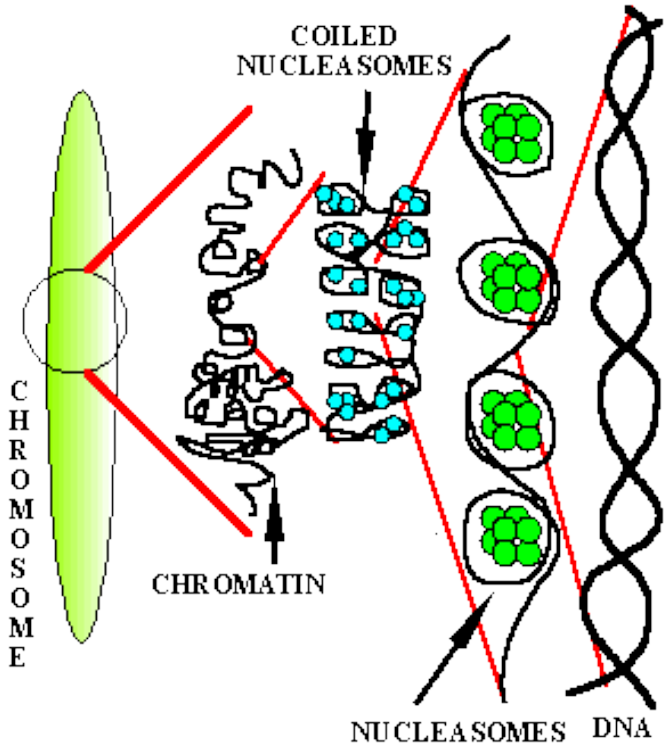
http://www.bio.miami.edu/dana/104/DNA2.jpg

Double helix of DNA

- James Watson and Francis Crick proposed a model for the structure of DNA.
  - Utilizing X-ray diffraction data, obtained from crystals of DNA
- This model predicted that DNA
  - as a helix of two complementary anti-parallel strands,
  - wound around each other in a rightward direction
  - stabilized by H-bonding between bases in adjacent strands.
  - The bases are in the interior of the helix
    - Purine bases form hydrogen bonds with pyrimidine.

DNA: The Basis of Life

- Humans have about 3 billion base pairs.
  - How do you package it into a cell?
  - How does the cell know where in the highly packed DNA where to start transcription?
    - Special regulatory sequences
  - DNA size does not mean more complex
- Complexity of DNA
  - Eukaryotic genomes consist of variable amounts of DNA
    - Single Copy or Unique DNA
    - Highly Repetitive DNA



DNA, continued

- DNA has a double helix structure. However, it is not symmetric. It has a "forward" and "backward" direction. The ends are labeled 5' and 3' after the Carbon atoms in the sugar component.

  5' AATCGCAAT 3'

  3' TTAGCGTTA 5'

DNA always reads 5' to 3' for transcription and replication

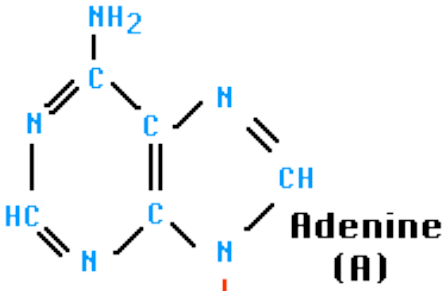
DNA Components

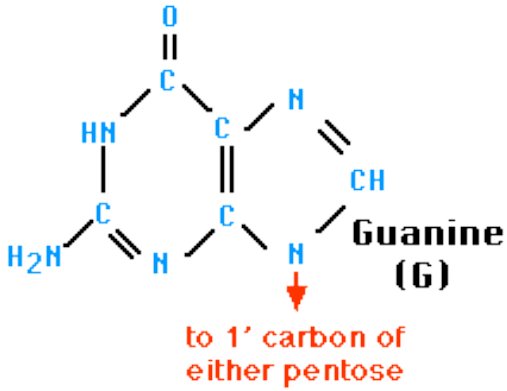
- **Nitrogenous Base:**
  N is important for hydrogen bonding between bases
  A – adenine with T – thymine (double H-bond)
  C – cytosine with G – guanine (triple H-bond)

- **Sugar:**
  Ribose (5 carbon)
  Base covalently bonds with 1' carbon
  Phosphate covalently bonds with 5' carbon
  Normal ribose (OH on 2' carbon) – RNA
  deoxyribose (H on 2' carbon) – DNA
  dideoxyribose (H on 2' & 3' carbon) – used in DNA sequencing

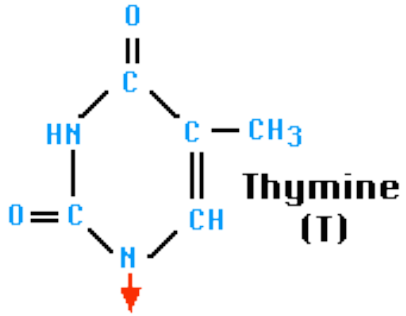
- **Phosphate:**
  negatively charged

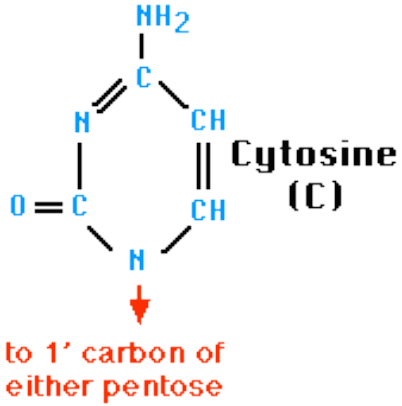
The Purines





The Pyrimidines

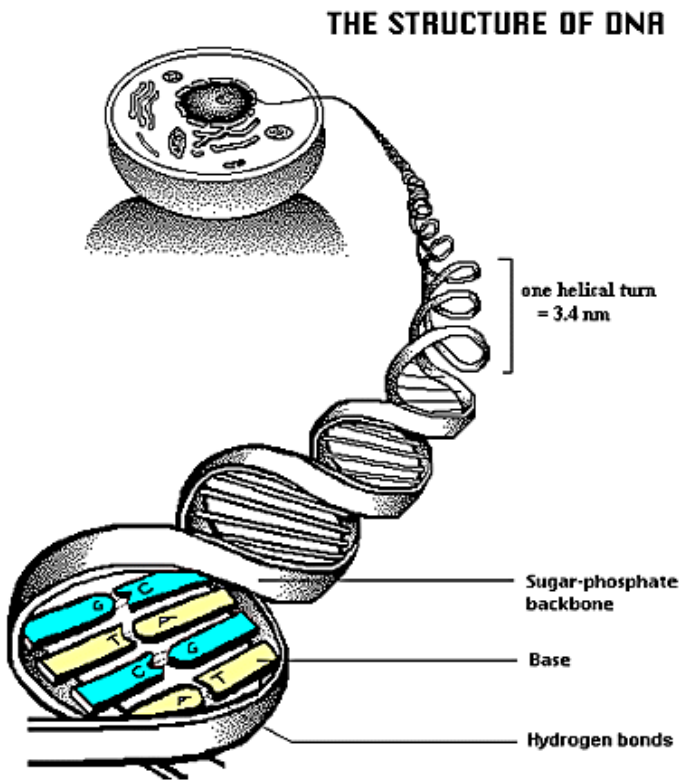




Double helix of DNA

- The double helix of DNA has these features:
  - Concentration of adenine (A) is equal to thymine (T)
  - Concentration of cytidine (C) is equal to guanine (G).
  - Watson-Crick base-pairing A will only base-pair with T, and C with G
    - base-pairs of G and C contain three H-bonds,
    - Base-pairs of A and T contain two H-bonds.
    - G-C base-pairs are more stable than A-T base-pairs
  - Two polynucleotide strands wound around each other.
  - The backbone of each consists of alternating deoxyribose and phosphate groups

Double helix of DNA

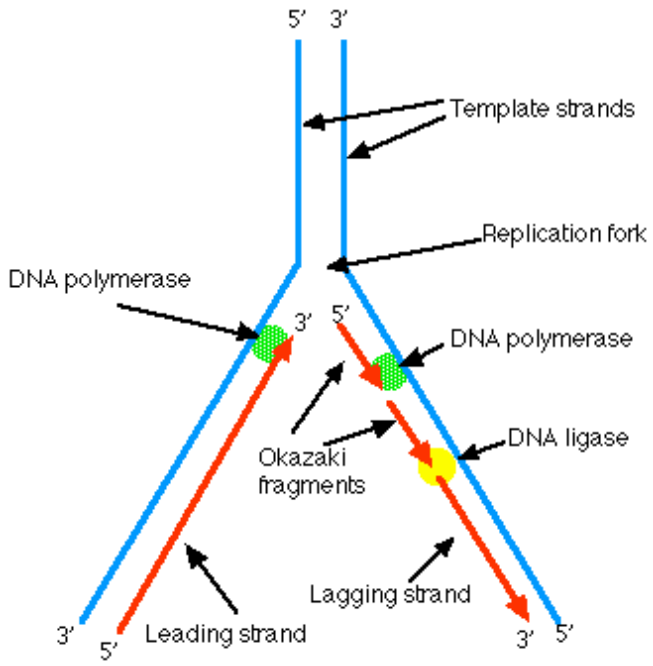


Double helix of DNA

- The DNA strands are assembled in the 5' to 3' direction
  - by convention, we "read" them the same way.
- The phosphate group bonded to the 5' carbon atom of one deoxyribose is covalently bonded to the 3' carbon of the next.
- The purine or pyrimidine attached to each deoxyribose projects in toward the axis of the helix.
- Each base forms hydrogen bonds with the one directly opposite it, forming base pairs (also called nucleotide pairs).

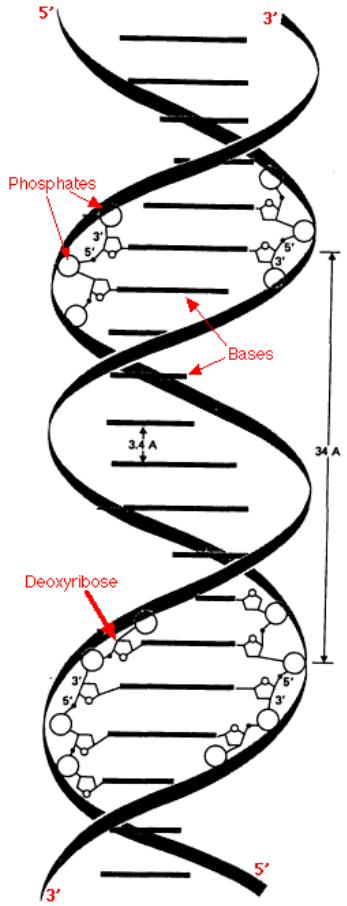
DNA - replication

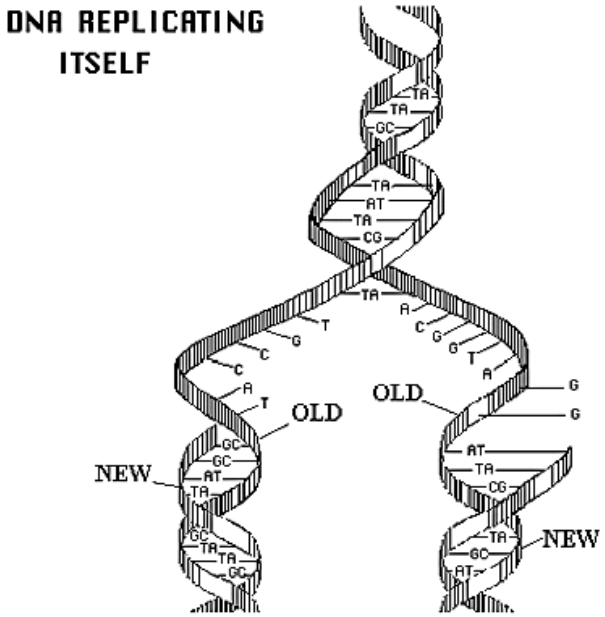
- DNA can replicate by splitting, and rebuilding each strand.
- Note that the rebuilding of each strand uses slightly different mechanisms due to the 5' 3' asymmetry, but each daughter strand is an exact replica of the original strand.



http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNAReplication.html

DNA Replication



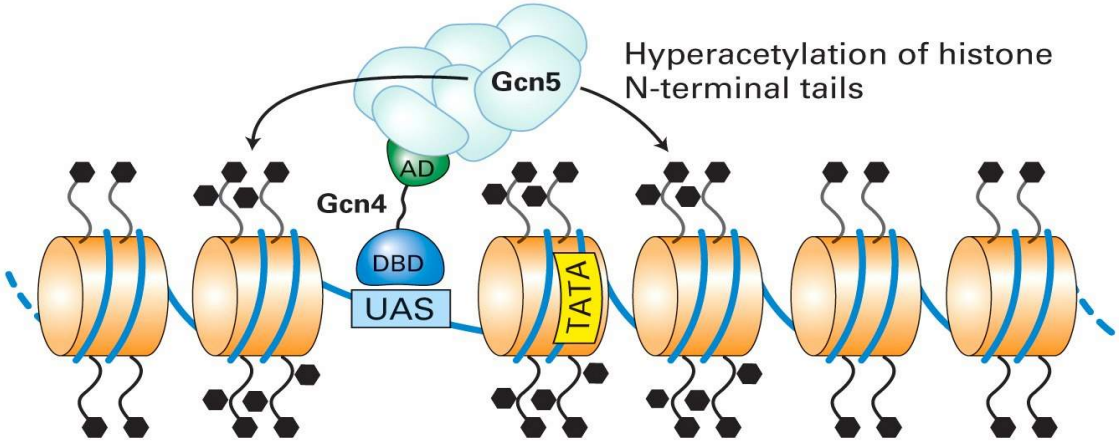


Superstructure Implications

- DNA in a living cell is in a highly compacted and structured state
- Transcription factors and RNA polymerase need ACCESS to do their work
- Transcription is dependent on the structural state – SEQUENCE alone does not tell the whole story

The Histone Code

- State of histone tails govern TF access to DNA
- State is governed by amino acid sequence and modification (acetylation, phosphorylation, methylation)



Lodish et al. *Molecular Biology of the Cell* (5th ed.). W.H. Freeman & Co., 2003.

What carries information between DNA to Proteins

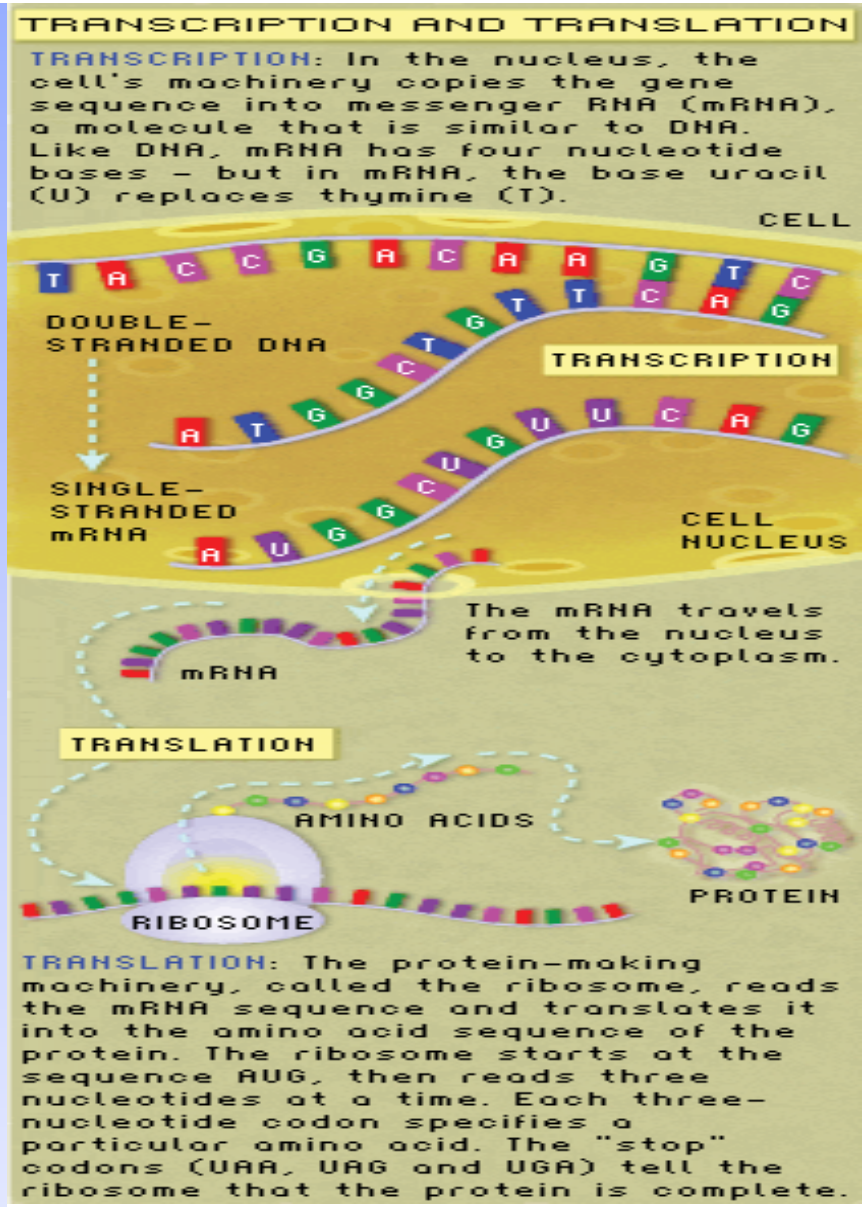
- *Central Dogma Of Biology*
- *RNA*
- *Transcription*
- *Splicing hnRNA-> mRNA*

- **Central Dogma**

  (DNA→RNA→protein) The paradigm that DNA directs its transcription to RNA, which is then translated into a protein. By understanding this process and how it is regulated, we can make predictions and models of cells.
- **Transcription**

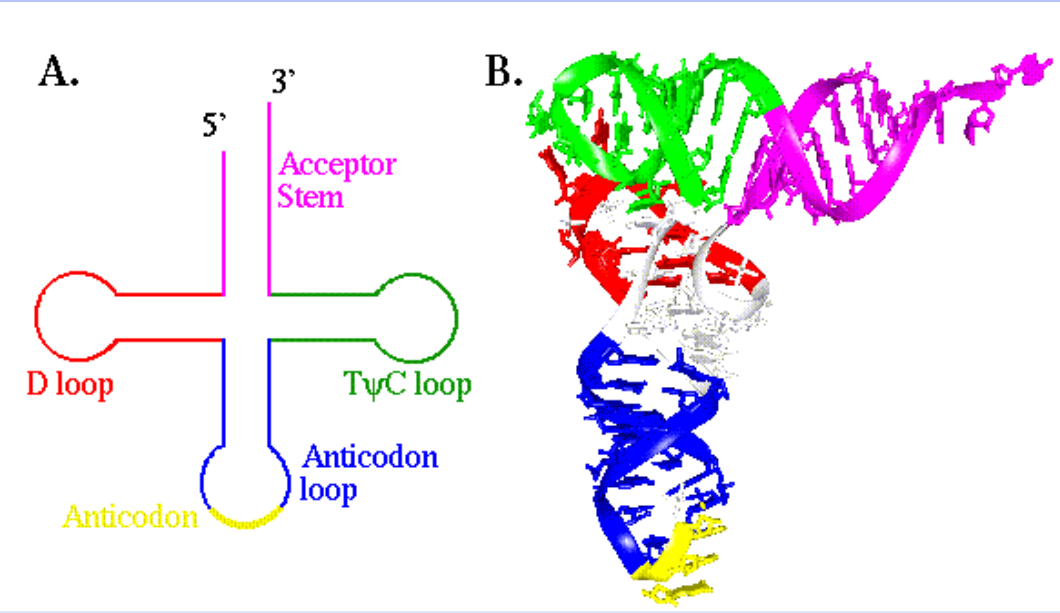
  (DNA→RNA) The process which transfers genetic information from the DNA to the RNA.
- **Translation**

  (RNA→protein) The process of transforming RNA to protein as specified by the genetic code.



RNA

- RNA is similar to DNA chemically. It is usually only a single strand. T(hyamine) is replaced by U(racil)
- Some forms of RNA can form secondary structures by "pairing up" with itself. This can have change its properties dramatically. DNA and RNA can pair with each other.



tRNA linear and 3D view: http://www.cgl.ucsf.edu/home/glasfeld/tutorial/trna/trna.gif

RNA, continued

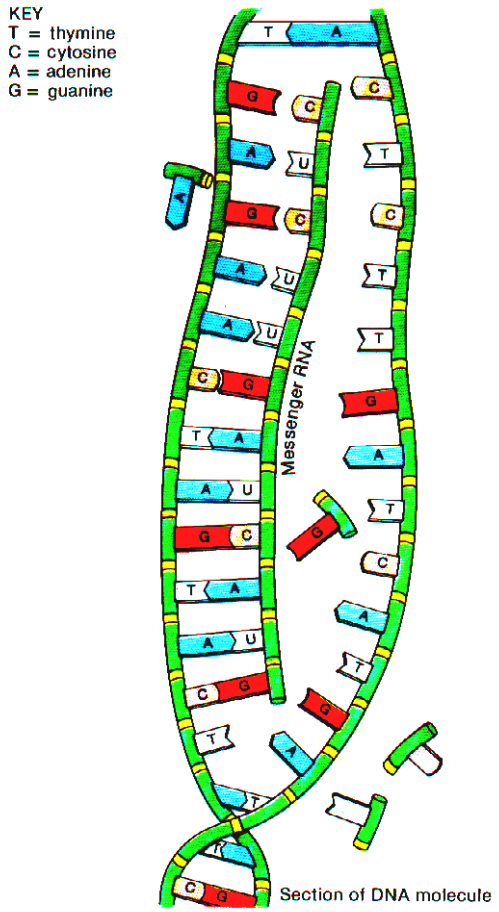
- Several types exist, classified by function
- mRNA – this is what is usually being referred to when a Bioinformatician says "RNA". This is used to carry a gene's *m*essage out of the nucleus.
- tRNA – *t*ransfers genetic information from mRNA to an amino acid sequence
- rRNA – *r*ibosomal RNA. Part of the ribosome which is involved in translation.

Terminology for Transcription

- hnRNA (heterogeneous nuclear RNA): Eukaryotic mRNA primary transcipts whose introns have not yet been excised (pre-mRNA).
- Phosphodiester Bond: Esterification linkage between a phosphate group and two alcohol groups.
- Promoter: A special sequence of nucleotides indicating the starting point for RNA synthesis.
- RNA (ribonucleotide): Nucleotides A,U,G, and C with ribose
- RNA Polymerase II: Multisubunit enzyme that catalyzes the synthesis of an RNA molecule on a DNA template from nucleoside triphosphate precursors.
- Terminator: Signal in DNA that halts transcription.

Transcription

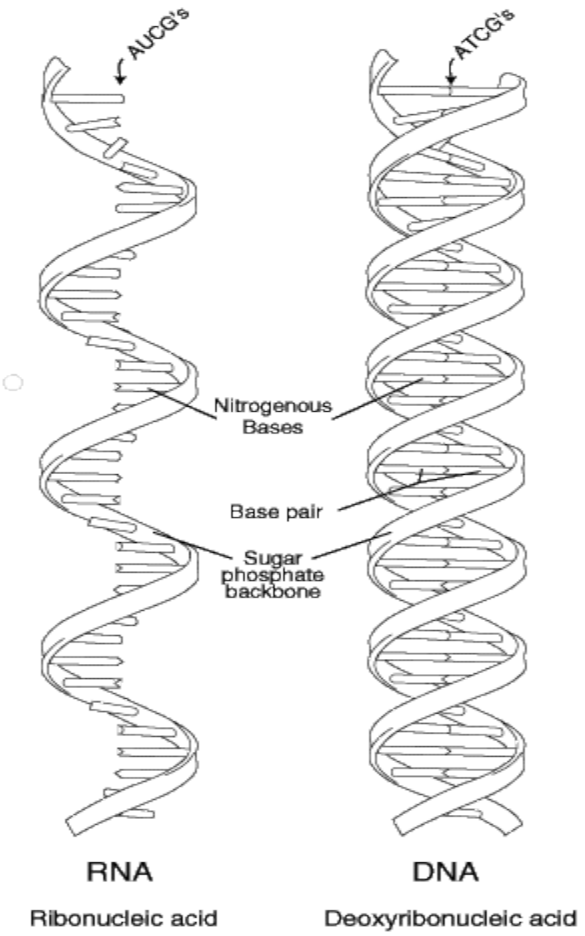
- The process of making RNA from DNA
- Catalyzed by "transcriptase" enzyme
- Needs a promoter region to begin transcription.
- ~50 base pairs/second in bacteria, but multiple transcriptions can occur simultaneously



http://ghs.gresham.k12.or.us/science/ps/sci/ibbio/chem/nucleic/chpt15/transcription.gif

DNA → RNA: Transcription

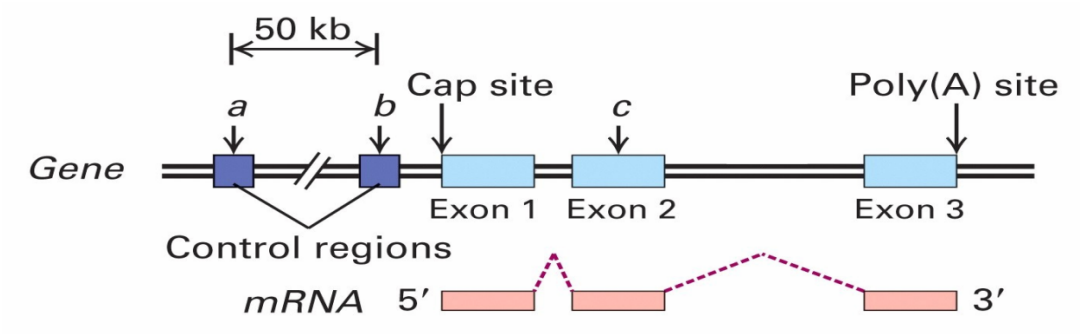
- DNA gets transcribed by a protein known as *RNA-polymerase*
- This process builds a chain of bases that will become mRNA
- RNA and DNA are similar, except that RNA is single stranded and thus less stable than DNA



Transcription, continued

- Transcription is highly regulated. Most DNA is in a dense form where it cannot be transcribed.
- To begin transcription requires a promoter, a small specific sequence of DNA to which polymerase can bind (~40 base pairs "upstream" of gene)
- Finding these promoter regions is a partially solved problem that is related to motif finding.
- There can also be repressors and inhibitors acting in various ways to stop transcription. This makes regulation of gene transcription complex to understand.

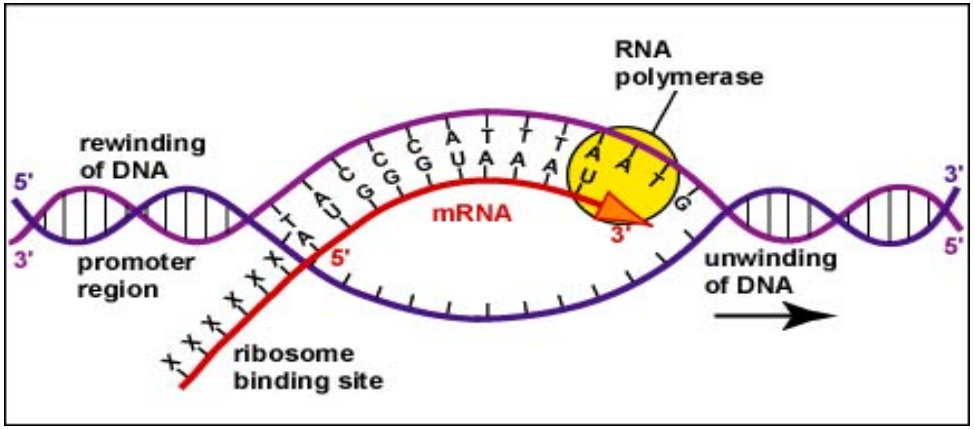
Definition of a Gene



- Regulatory regions: up to 50 kb upstream of +1 site
- Exons: protein coding and untranslated regions (UTR)
  1 to 178 exons per gene (mean 8.8)
  8 bp to 17 kb per exon (mean 145 bp)
- Introns: splice acceptor and donor sites, junk DNA
  average 1 kb – 50 kb per intron
- Gene size: Largest – 2.4 Mb (Dystrophin). Mean – 27 kb.

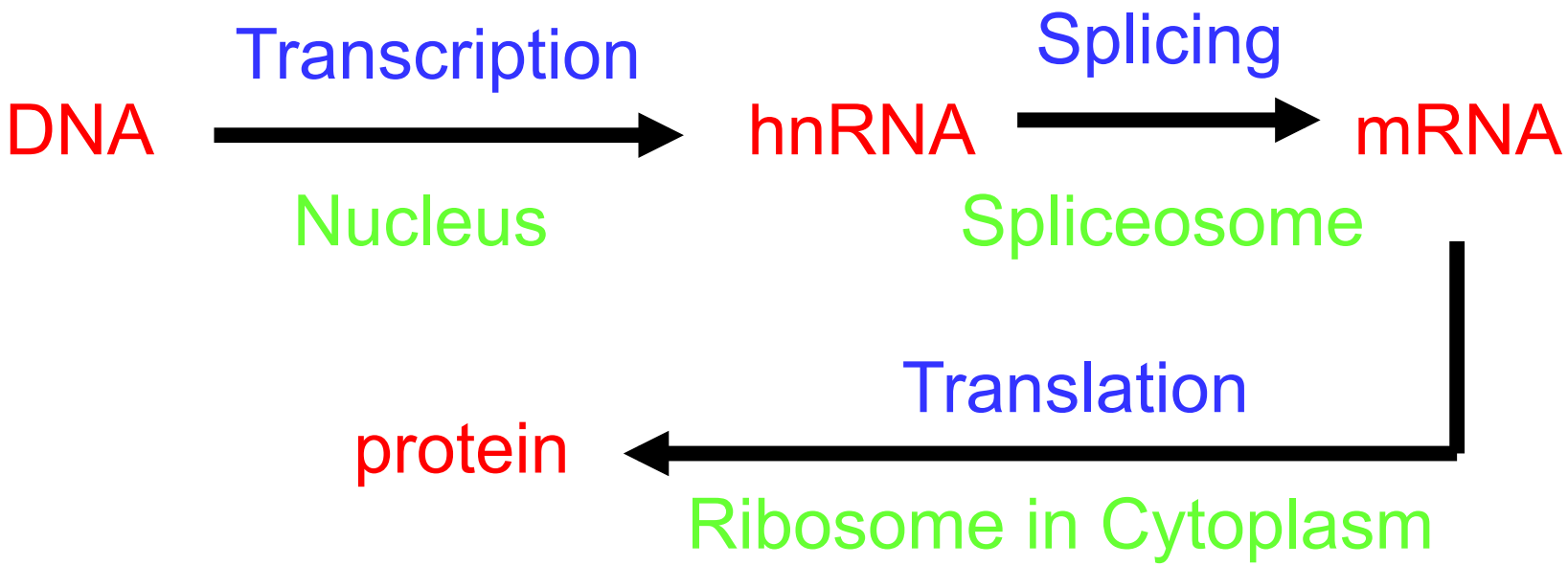
Transcription: DNA → hnRNA

- Transcription occurs in the nucleus.
- σ factor from RNA polymerase reads the promoter sequence and opens a small portion of the double helix exposing the DNA bases.



- RNA polymerase II catalyzes the formation of phosphodiester bond that link nucleotides together to form a linear chain from 5' to 3' by unwinding the helix just ahead of the active site for polymerization of complementary base pairs.
- The hydrolysis of high energy bonds of the substrates (nucleoside triphosphates ATP, CTP, GTP, and UTP) provides energy to drive the reaction.
- During transcription, the DNA helix reforms as RNA forms.
- When the terminator sequence is met, polymerase halts and releases both the DNA template and the RNA.

Central Dogma Revisited

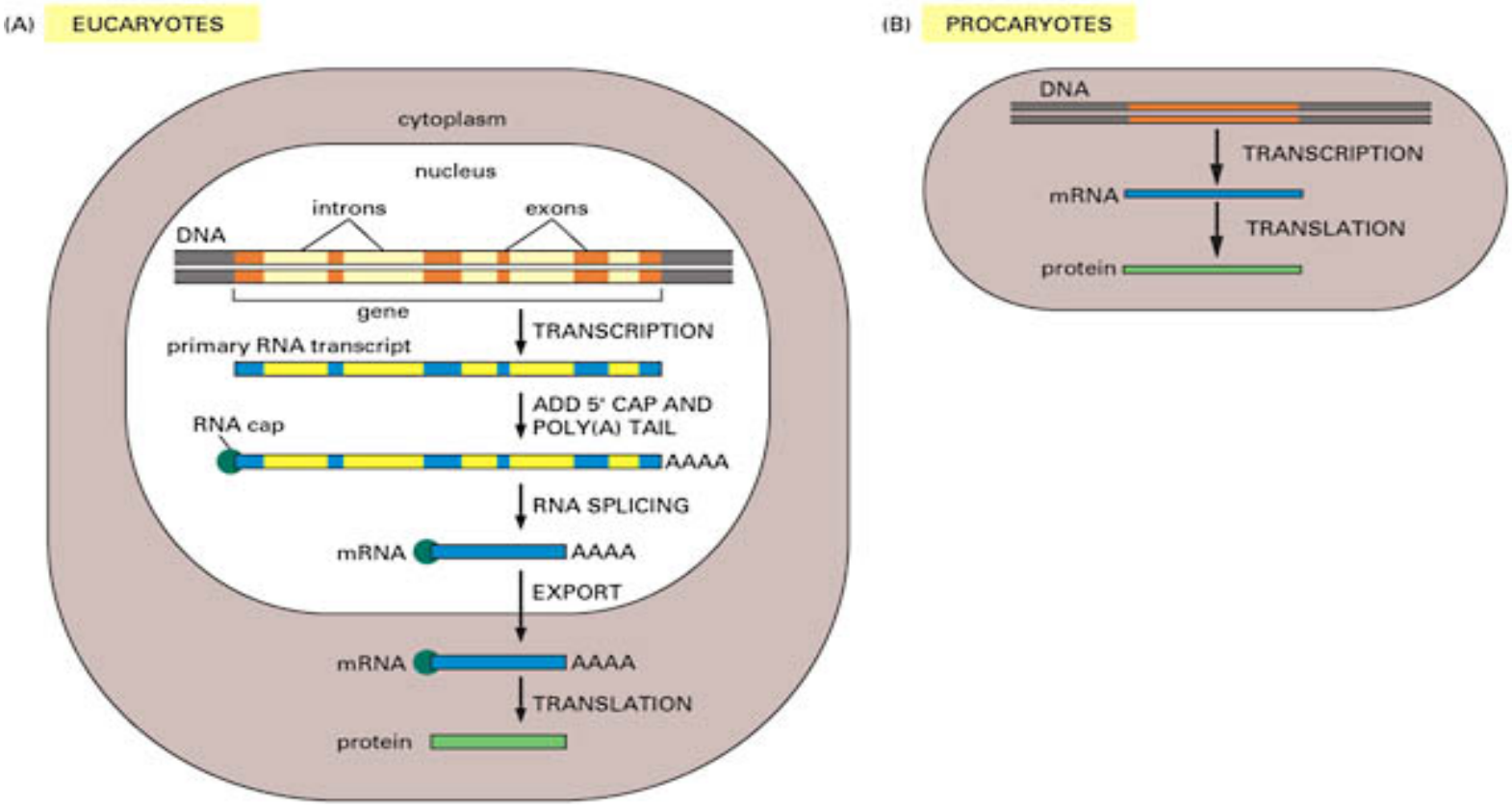


- Base Pairing Rule: A and T or U is held together by 2 hydrogen bonds and G and C is held together by 3 hydrogen bonds.
- Note: Some mRNA stays as RNA (ie tRNA,rRNA).

Terminology for Splicing

- Exon: A portion of the gene that appears in both the primary and the mature mRNA transcripts.
- Intron: A portion of the gene that is transcribed but excised prior to translation.
- Lariat structure: The structure that an intron in mRNA takes during excision/splicing.
- Spliceosome: An organelle that carries out the splicing reactions whereby the pre-mRNA is converted to a mature mRNA.

Splicing


(A) EUCARYOTES
cytoplasm
nucleus
introns
exons
DNA
gene
TRANSCRIPTION
primary RNA transcript
ADD 5' CAP AND
POLY(A) TAIL
RNA cap
AAAA
RNA SPLICING
mRNA
AAAA
EXPORT
mRNA
AAAA
TRANSLATION
protein
(B) PROCARYOTES
DNA
TRANSCRIPTION
mRNA
TRANSLATION
protein

Splicing: hnRNA → mRNA

- Takes place on spliceosome that brings together a hnRNA, snRNPs, and a variety of pre-mRNA binding proteins.
- 2 transesterification reactions:

1. 2',5' phosphodiester bond forms between an intron adenosine residue and the intron's 5'-terminal phosphate group and a lariat structure is formed.
2. The free 3'-OH group of the 5' exon displaces the 3' end of the intron, forming a phosphodiester bond with the 5' terminal phosphate of the 3' exon to yield the spliced product. The lariat formed intron is the degraded.

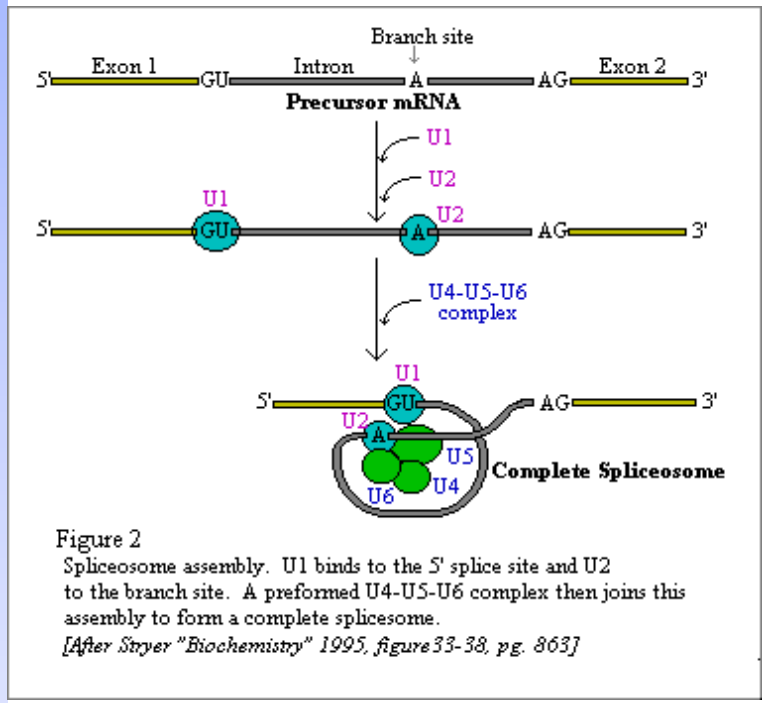


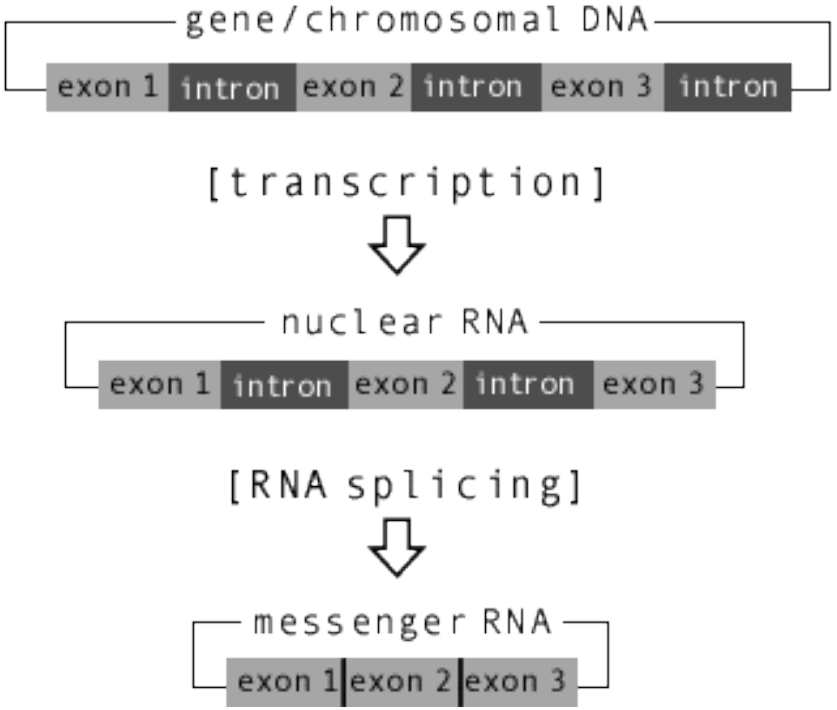
Figure 2
Spliceosome assembly. U1 binds to the 5' splice site and U2 to the branch site. A preformed U4-U5-U6 complex then joins this assembly to form a complete splicesome.
[After Stryer "Biochemistry" 1995, figure 33-38, pg. 863]

Splicing and other RNA processing

- In Eukaryotic cells, RNA is processed between transcription and translation.
- This complicates the relationship between a DNA gene and the protein it codes for.
- Sometimes alternate RNA processing can lead to an alternate protein as a result. This is true in the immune system.

Splicing (Eukaryotes)

- Unprocessed RNA is composed of Introns and Extrons. Introns are removed before the rest is expressed and converted to protein.
- Sometimes alternate splicings can create different valid proteins.
- A typical Eukaryotic gene has 4-20 introns. Locating them by analytical means is not easy.

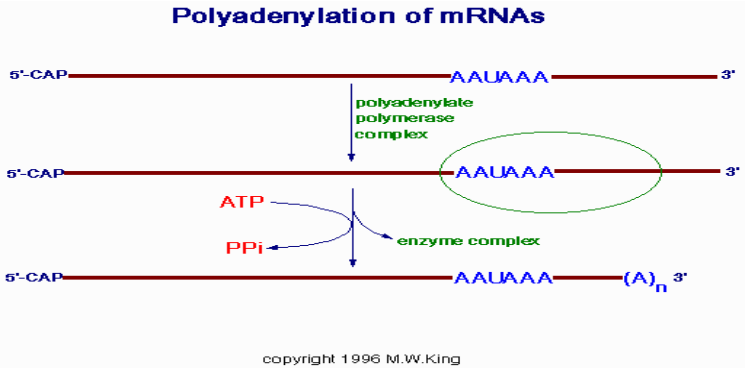


Posttranscriptional Processing: Capping and Poly(A) Tail

Capping

- Prevents 5' exonucleolytic degradation.
- 3 reactions to cap:

1. Phosphatase removes 1 phosphate from 5' end of hnRNA
2. Guanyl transferase adds a GMP in reverse linkage 5' to 5'.
3. Methyl transferase adds methyl group to guanosine.



Poly(A) Tail

- Due to transcription termination process being imprecise.
- 2 reactions to append:

1. Transcript cleaved 15-25 past highly conserved AAUAAA sequence and less than 50 nucleotides before less conserved U rich or GU rich sequences.
2. Poly(A) tail generated from ATP by poly(A) polymerase which is activated by cleavage and polyadenylation specificity factor (CPSF) when CPSF recognizes AAUAAA. Once poly(A) tail has grown approximately 10 residues, CPSF disengages from the recognition site.

How Are Proteins Made? (Translation)

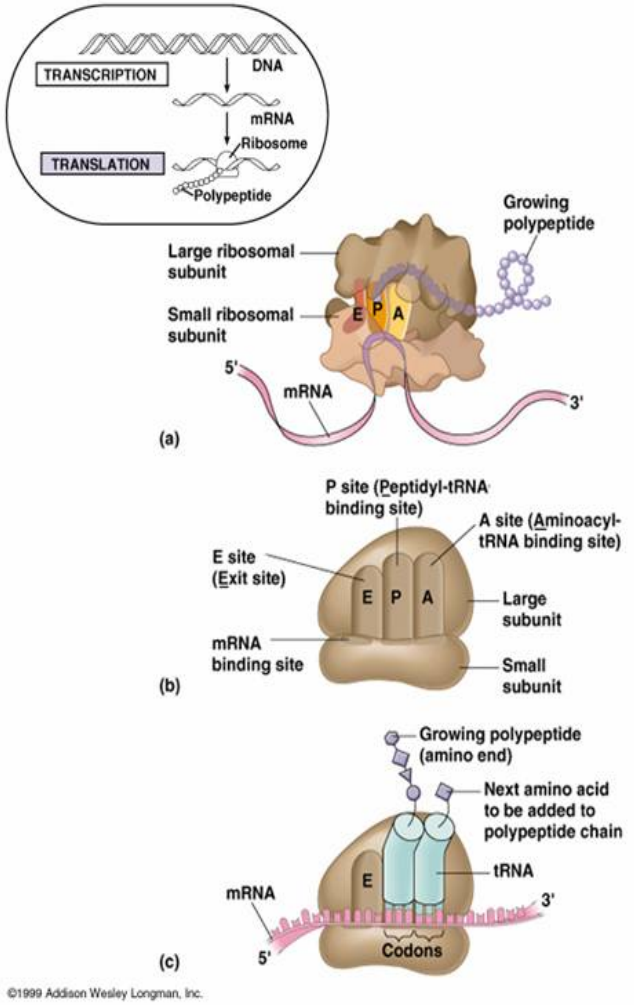
- *mRNA*
- *tRNA*
- *Translation*
- *Protein Synthesis*
- *Protein Folding*

Terminology for Ribosome

- Codon: The sequence of 3 nucleotides in DNA/RNA that encodes for a specific amino acid.
- mRNA (messenger RNA): A ribonucleic acid whose sequence is complementary to that of a protein-coding gene in DNA.
- Ribosome: The organelle that synthesizes polypeptides under the direction of mRNA
- rRNA (ribosomal RNA):The RNA molecules that constitute the bulk of the ribosome and provides structural scaffolding for the ribosome and catalyzes peptide bond formation.
- tRNA (transfer RNA): The small L-shaped RNAs that deliver specific amino acids to ribosomes according to the sequence of a bound mRNA.

mRNA → Ribosome

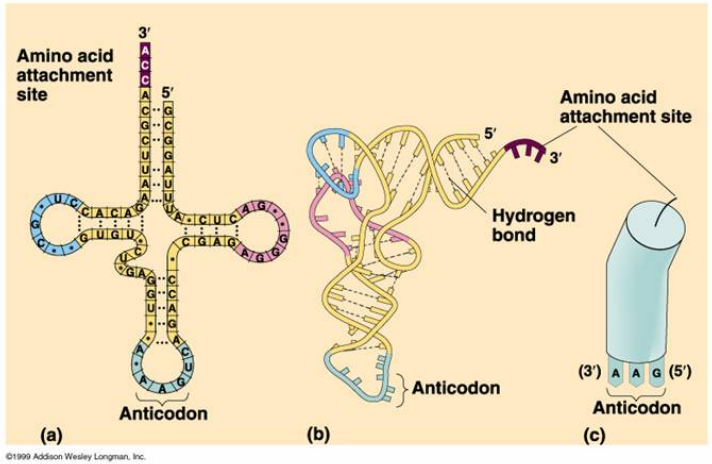
- mRNA leaves the nucleus via nuclear pores.
- Ribosome has 3 binding sites for tRNAs:
  - A-site: position that aminoacyl-tRNA molecule binds to vacant site
  - P-site: site where the new peptide bond is formed.
  - E-site: the exit site
- Two subunits join together on a mRNA molecule near the 5' end.
- The ribosome will read the codons until AUG is reached and then the initiator tRNA binds to the P-site of the ribosome.
- Stop codons have tRNA that recognize a signal to stop translation. Release factors bind to the ribosome which cause the peptidyl transferase to catalyze the addition of water to free the molecule and releases the polypeptide.



Terminology for tRNA and proteins

- Anticodon: The sequence of 3 nucleotides in tRNA that recognizes an mRNA codon through complementary base pairing.
- C-terminal: The end of the protein with the free COOH.
- N-terminal: The end of the protein with the free NH3.

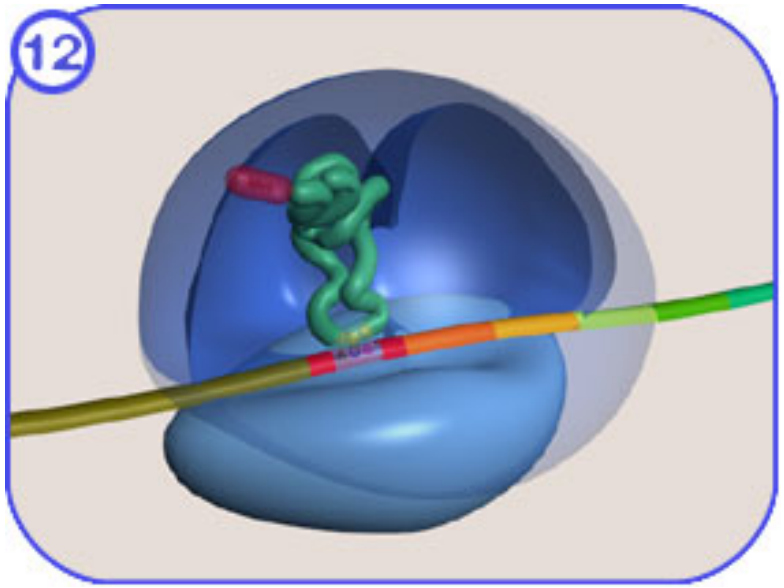
Purpose of tRNA



- The proper tRNA is chosen by having the corresponding anticodon for the mRNA's codon.
- The tRNA then transfers its aminoacyl group to the growing peptide chain.
- For example, the tRNA with the anticodon UAC corresponds with the codon AUG and attaches methionine amino acid onto the peptide chain.

Translation: tRNA

- mRNA is translated in 5' to 3' direction and the from N-terminal to C-terminus of the polypeptide.
- Elongation process (assuming polypeptide already began):
  - tRNA with the next amino acid in the chain binds to the A-site by forming base pairs with the codon from mRNA



  - Carboxyl end of the protein is released from the tRNA at the Psite and joined to the free amino group from the amino acid attached to the tRNA at the A-site; new peptide bond formed catalyzed by peptide transferase.
  - Conformational changes occur which shift the two tRNAs into the E-site and the P-site from the P-site and A-site respectively. The mRNA also shifts 3 nucleotides over to reveal the next codon.
  - The tRNA in the E-site is released
- GTP hydrolysis provides the energy to drive this reaction.

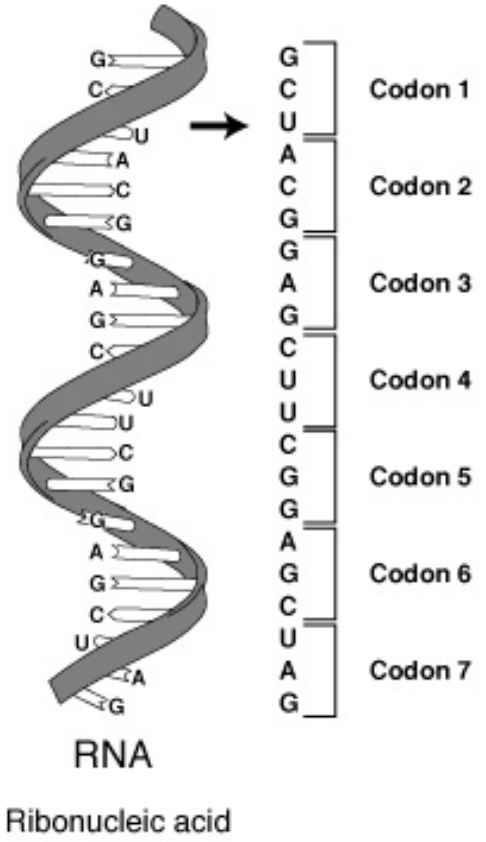
Terminology for Protein Folding

- Endoplasmic Reticulum: Membraneous organelle in eukaryotic cells where lipid synthesis and some posttranslational modification occurs.
- Mitochondria: Eukaryotic organelle where citric acid cycle, fatty acid oxidation, and oxidative phosphorylation occur.
- Molecular chaperone: Protein that binds to unfolded or misfolded proteins to refold the proteins in the quaternary structure.

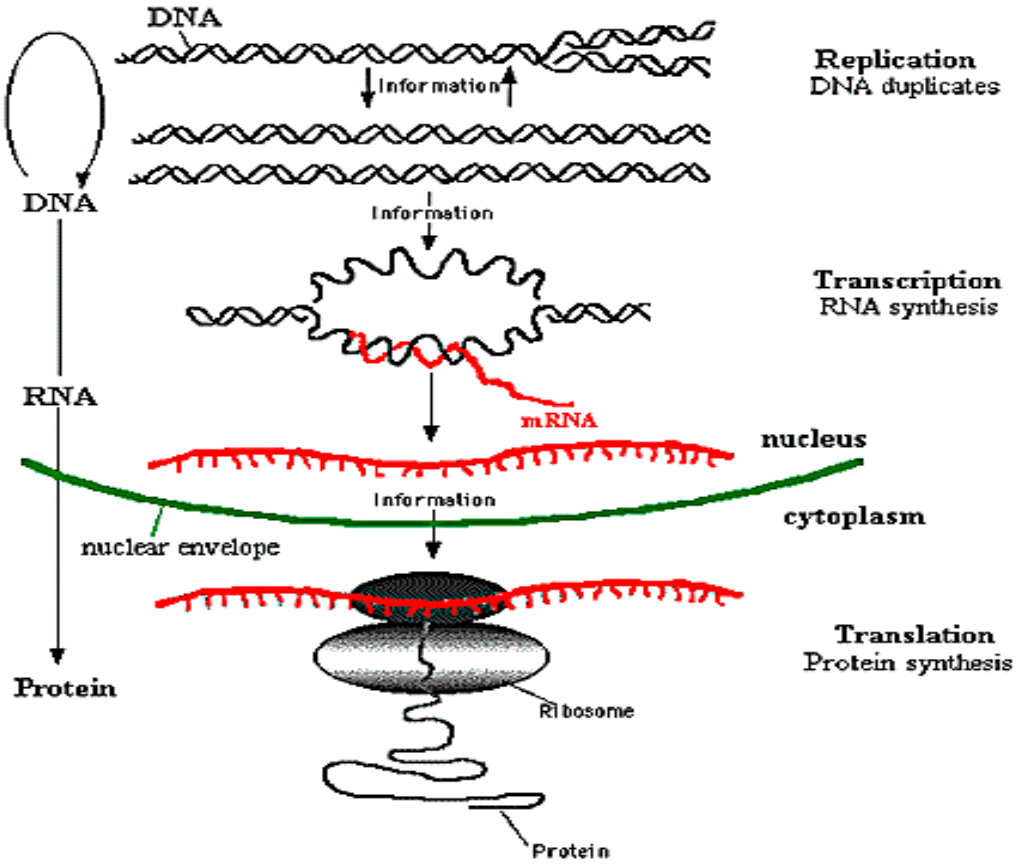
Revisiting the Central Dogma

- In going from DNA to proteins, there is an intermediate step where mRNA is made from DNA, which then makes protein
  - This known as **The Central Dogma**

- Why the intermediate step?
  - DNA is kept in the nucleus, while protein sythesis happens in the cytoplasm, with the help of ribosomes



The Central Dogma (cont'd)



The Central Dogma of Molecular Biology

RNA → Protein: Translation

- Ribosomes and *transfer-RNAs* (tRNA) run along the length of the newly synthesized mRNA, decoding one codon at a time to build a growing chain of amino acids ("peptide")
  - The tRNAs have anti-codons, which complimentarily match the codons of mRNA to know what protein gets added next
- But first, in eukaryotes, a phenomenon called splicing occurs
  - Introns are non-protein coding regions of the mRNA; exons are the coding regions
  - Introns are removed from the mRNA during splicing so that a functional, valid protein can form

Translation

- The process of going from RNA to polypeptide.
- Three base pairs of RNA (called a codon) correspond to one amino acid based on a fixed table.
- Always starts with Methionine and ends with a stop codon

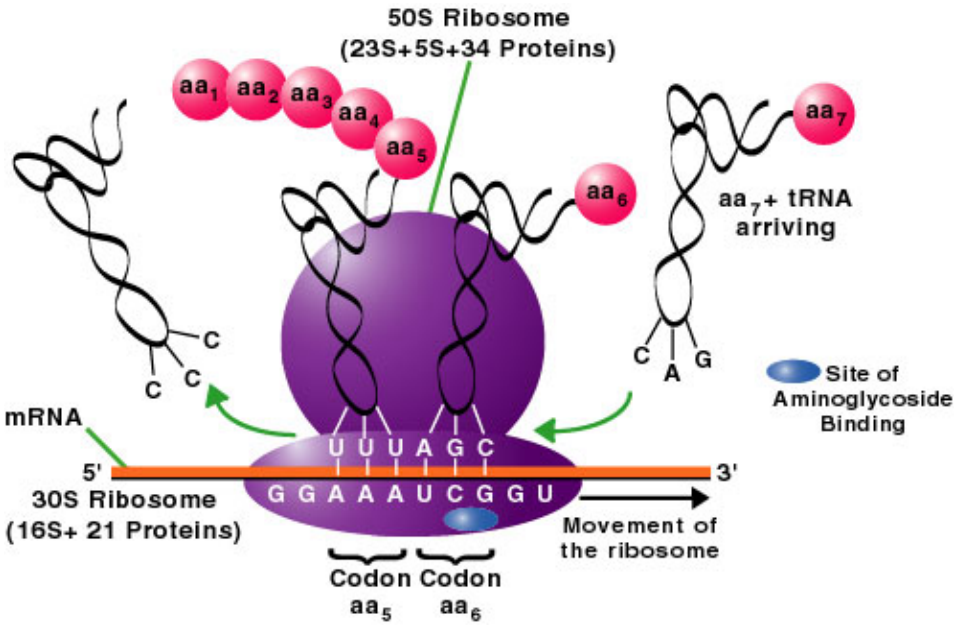
SECOND POSITION

FIRST POSITION	U	C	A	G	THIRD POSITION
U	phenyl-alanine	serine	tyrosine	cysteine	U
					C
	leucine		stop	stop	A
			stop	tryptophan	G
C	leucine	proline	histidine	arginine	U
					C
			glutamine		A
					G
A	isoleucine	threonine	asparagine	serine	U
					C
			lysine	arginine	A
	* methionine				G
G	valine	alanine	aspartic acid	glycine	U
					C
			glutamic acid		A
					G

* and start

Translation, continued

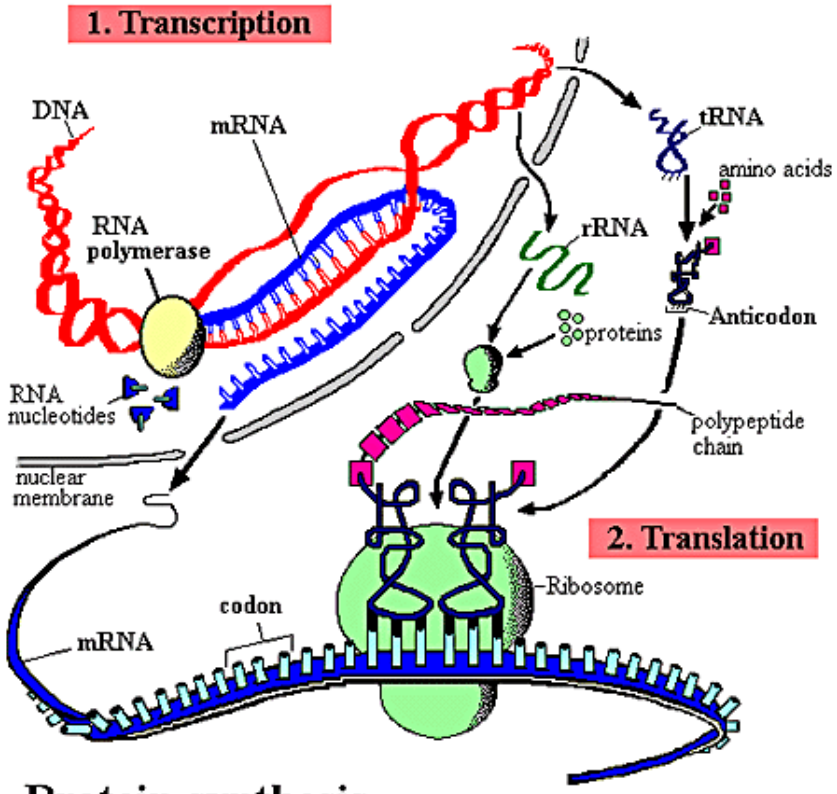
- Catalyzed by Ribosome
- Using two different sites, the Ribosome continually binds tRNA, joins the amino acids together and moves to the next location along the mRNA
- ~10 codons/second, but multiple translations can occur simultaneously



http://wong.scripps.edu/PIX/ribosome.jpg

Protein Synthesis: Summary

- There are twenty amino acids, each coded by three- base-sequences in DNA, called "codons"
  - This code is degenerate
- The **central dogma** describes how proteins derive from DNA
  - DNA → mRNA → (splicing?) → protein
- The protein adopts a 3D structure specific to it's amino acid arrangement and function



Protein synthesis

Proteins

- Complex organic molecules made up of amino acid subunits
- 20* different kinds of amino acids. Each has a 1 and 3 letter abbreviation.
- http://www.indstate.edu/thcme/mwking/amino-acids.html for complete list of chemical structures and abbreviations.
- Proteins are often enzymes that catalyze reactions.
- Also called "poly-peptides"

*Some other amino acids exist but not in humans.

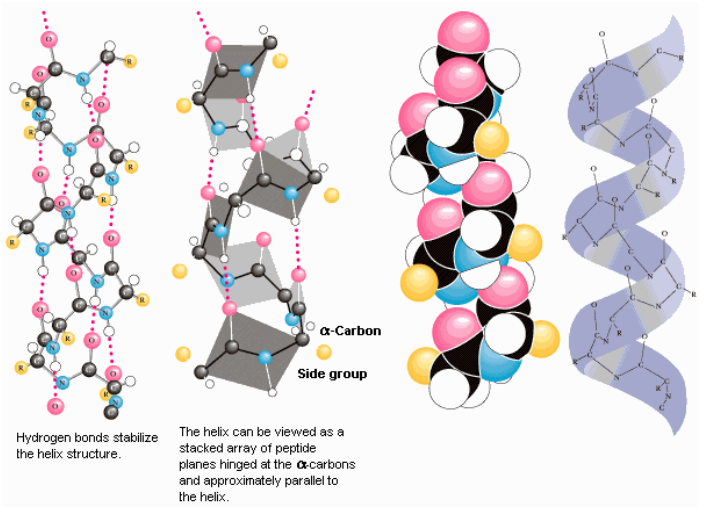
Polypeptide v. Protein

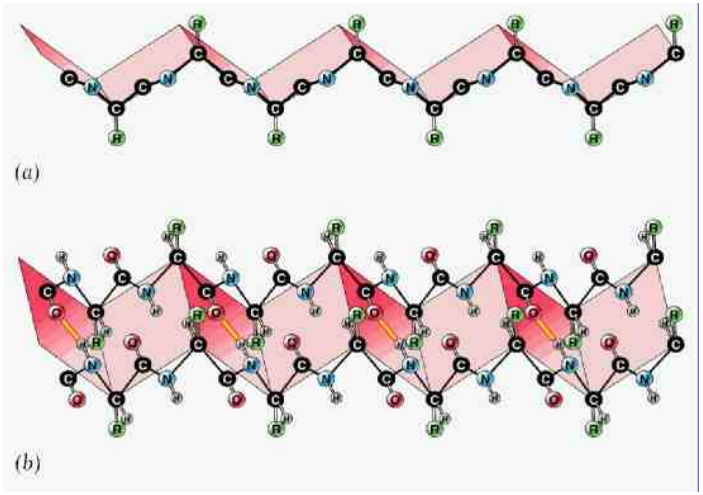
- A protein is a polypeptide, however to understand the function of a protein given only the polypeptide sequence is a very difficult problem.
- Protein folding an open problem. The 3D structure depends on many variables.
- Current approaches often work by looking at the structure of homologous (similar) proteins.
- Improper folding of a protein is believed to be the cause of mad cow disease.

http://www.sanger.ac.uk/Users/sgj/thesis/node2.html for more information on folding

Protein Folding

- Proteins tend to fold into the lowest free energy conformation.
- Proteins begin to fold while the peptide is still being translated.
- Proteins bury most of its hydrophobic residues in an interior core to form an α helix.
- Most proteins take the form of secondary structures α helices and β sheets.
- Molecular chaperones, hsp60 and hsp 70, work with other proteins to help fold newly synthesized proteins.
- Much of the protein modifications and folding occurs in the endoplasmic reticulum and mitochondria.



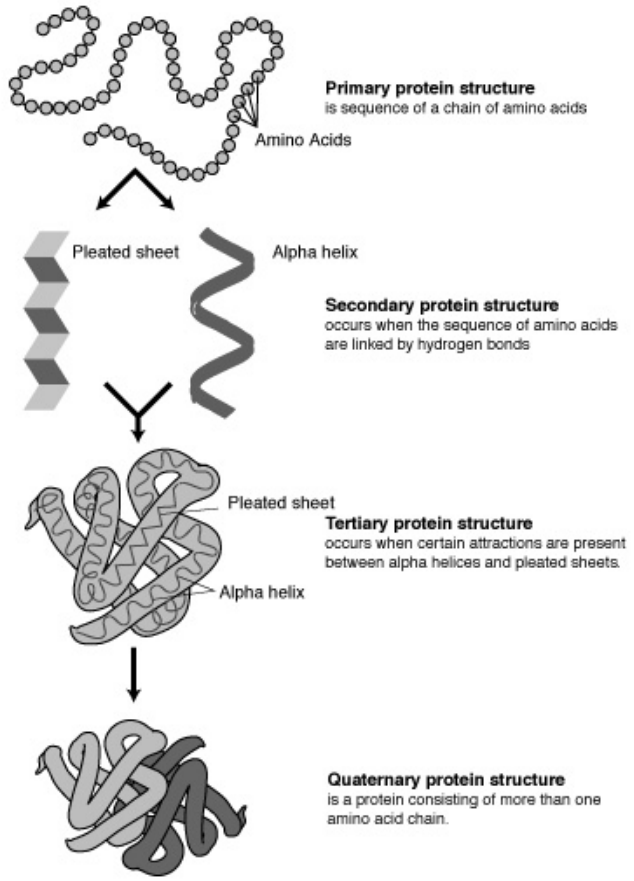


Protein Folding

- Proteins are not linear structures, though they are built that way
- The amino acids have very different chemical properties; they interact with each other after the protein is built
  - This causes the protein to start fold and adopting it's functional structure
  - Proteins may fold in reaction to some ions, and several separate chains of peptides may join together through their hydrophobic and hydrophilic amino acids to form a polymer

Protein Folding (cont'd)

- The structure that a protein adopts is vital to it's chemistry
- Its structure determines which of its amino acids are exposed carry out the protein's function
- Its structure also determines what substrates it can react with



Analyzing a Genome

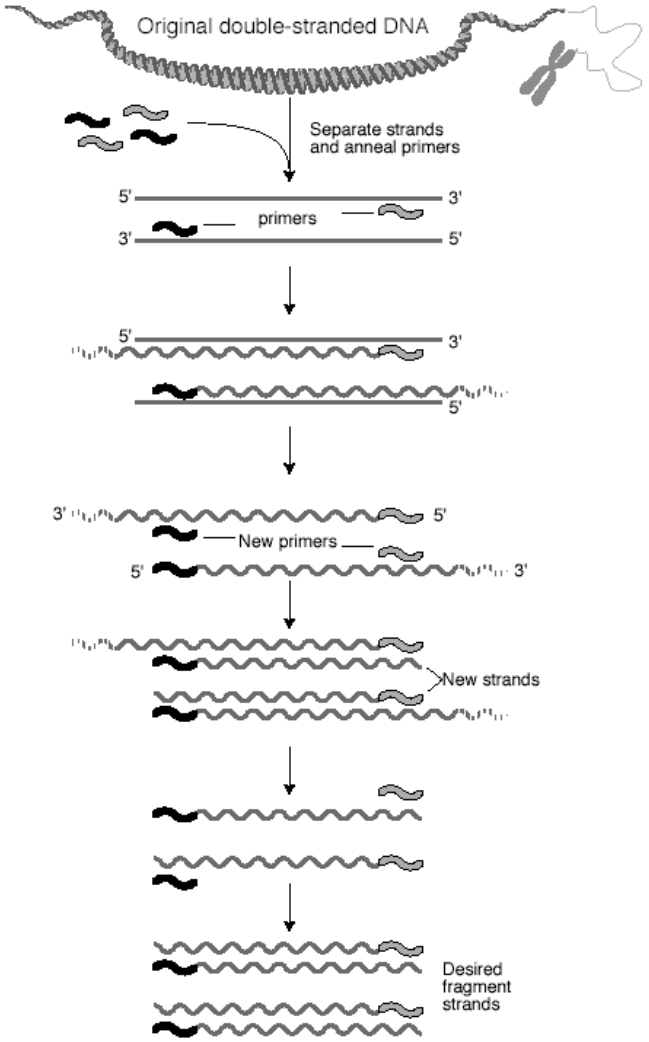
- How to analyze a genome in four easy steps.
  - Cut it
    - Use enzymes to cut the DNA in to small fragments.
  - Copy it
    - Copy it many times to make it easier to see and detect.
  - Read it
    - Use special chemical techniques to read the small fragments.
  - Assemble it
    - Take all the fragments and put them back together. This is hard!!!
- Bioinformatics takes over
  - What can we learn from the sequenced DNA.
  - Compare interspecies and intraspecies.

Copying DNA

- Biologists needed to find a way to read DNA codes.
- How do you read base pairs that are angstroms in size?
  - It is not possible to directly look at it due to DNA's small size.
  - Need to use chemical techniques to detect what you are looking for.
  - To read something so small, you need a lot of it, so that you can actually detect the chemistry.
- Need a way to make many copies of the base pairs, and a method for reading the pairs.

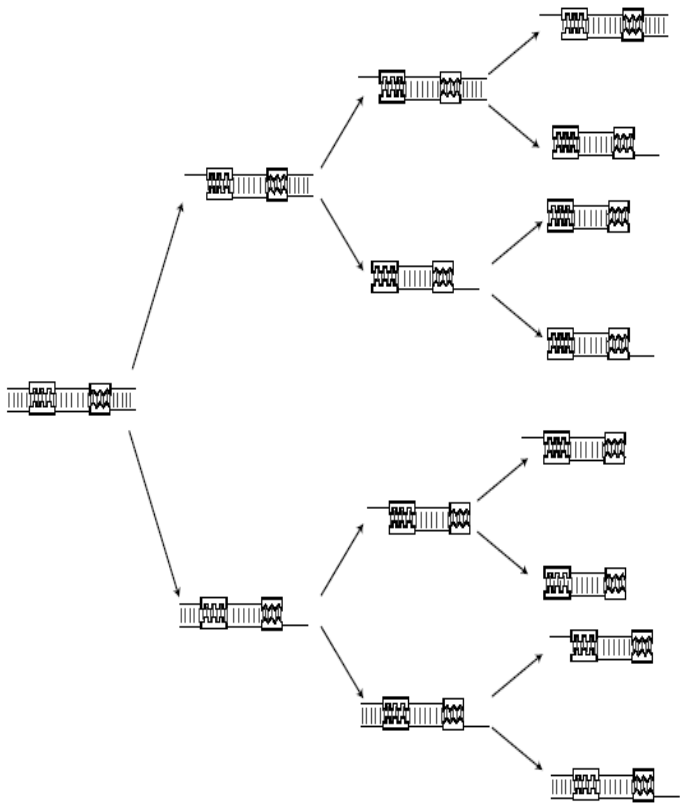
Polymerase Chain Reaction (PCR)

- Polymerase Chain Reaction (PCR)
  - Used to massively replicate DNA sequences.
- How it works:
  - Separate the two strands with low heat
  - Add some base pairs, primer sequences, and DNA Polymerase
    - Creates double stranded DNA from a single strand.
    - Primer sequences create a seed from which double stranded DNA grows.
  - Now you have two copies.
  - Repeat. Amount of DNA grows exponentially.
    - 1→2→4→8→16→32→64→128→256…

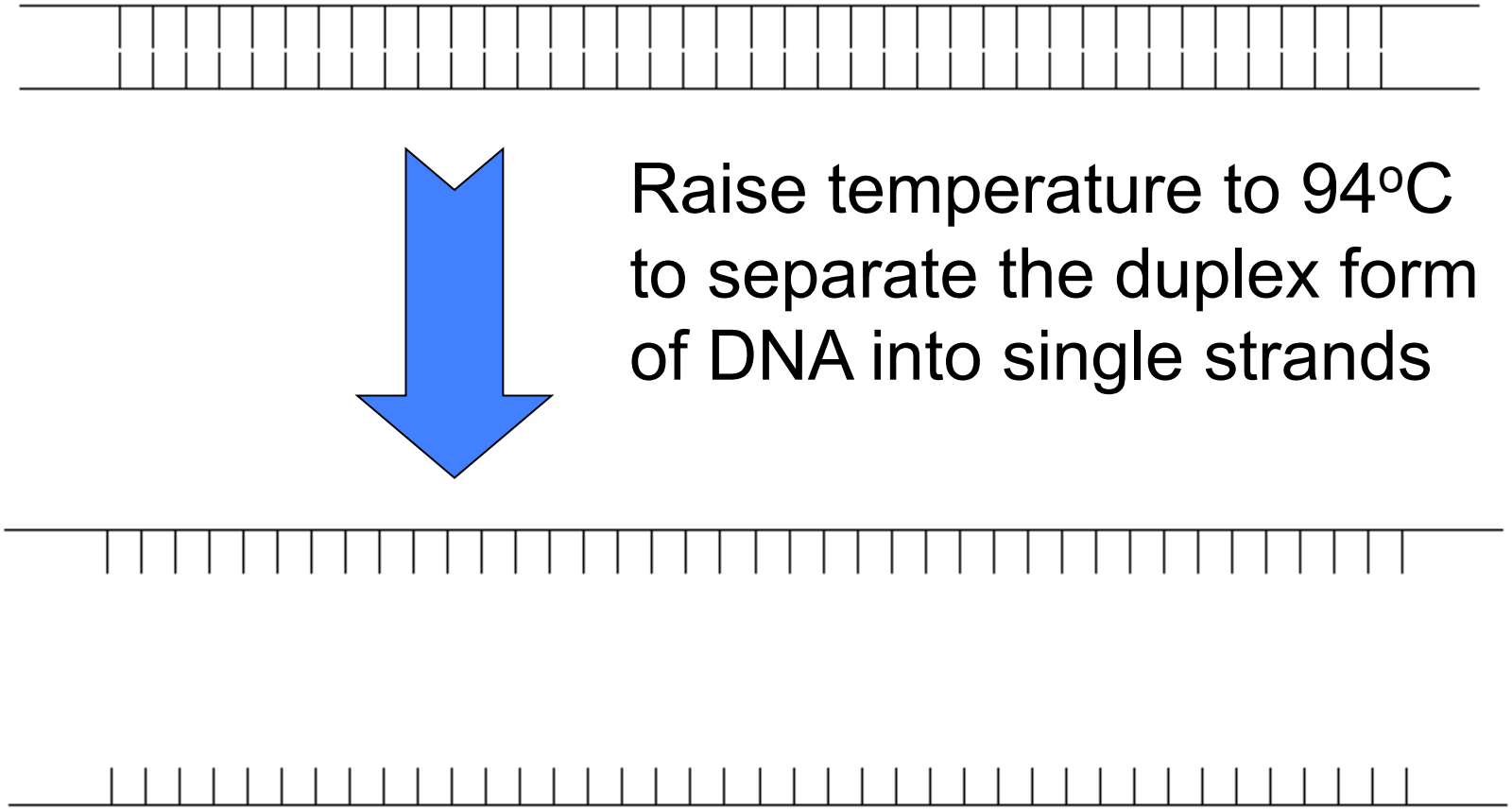


Polymerase Chain Reaction

- **Problem**: Modern instrumentation cannot easily detect single molecules of DNA, making amplification a prerequisite for further analysis
- **Solution**: PCR doubles the number of DNA fragments at every iteration

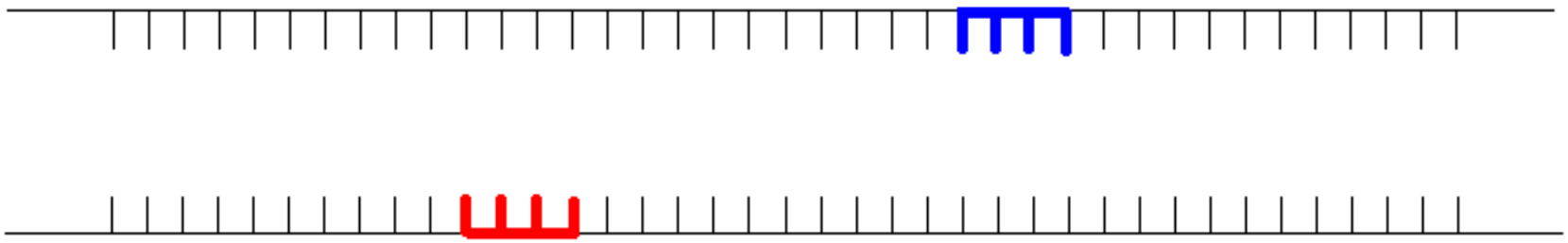


Denaturation



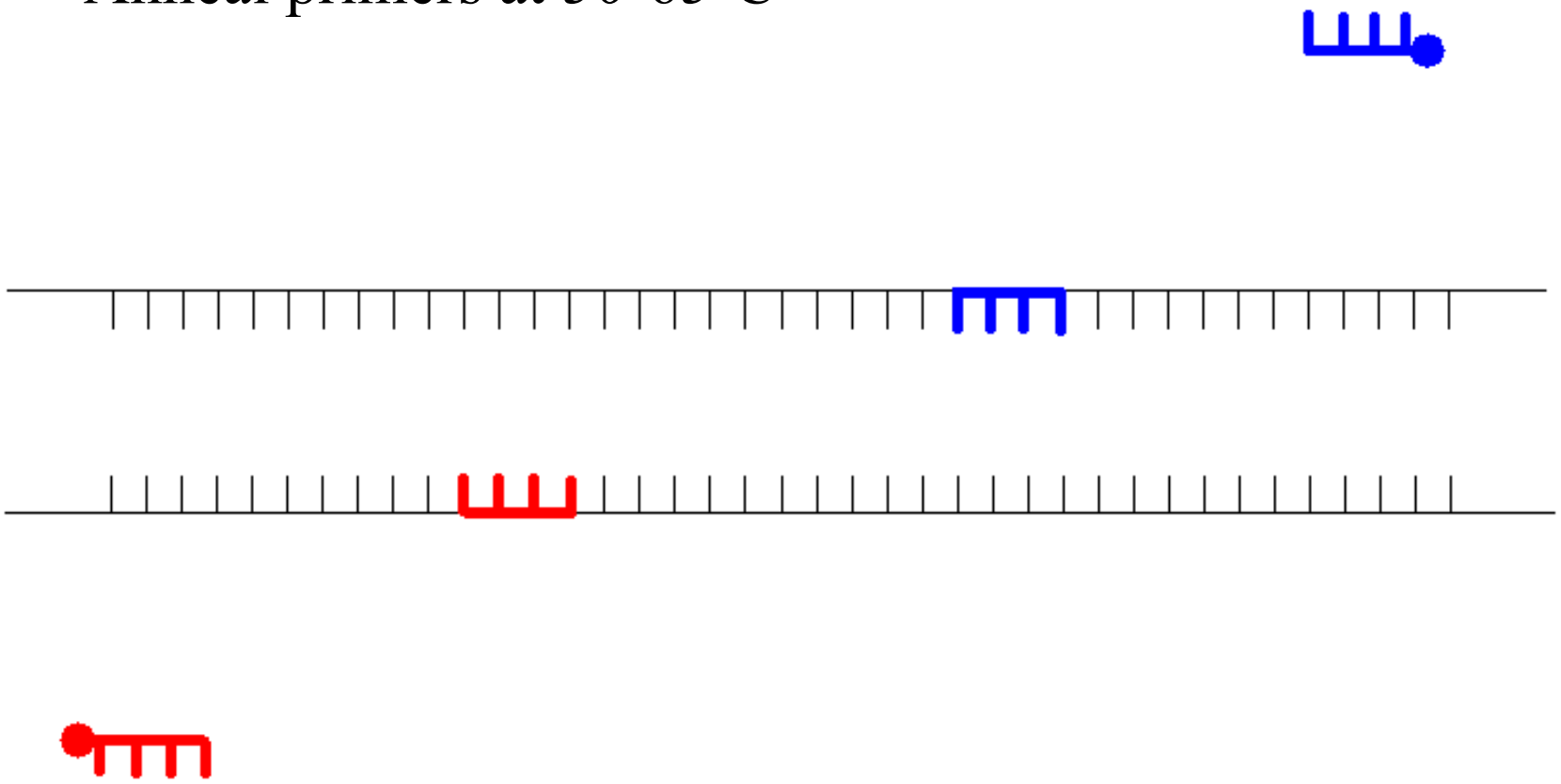
Design primers

- To perform PCR, a 10-20bp sequence on either side of the sequence to be amplified must be known because DNA polymerase requires a primer to synthesize a new strand of DNA



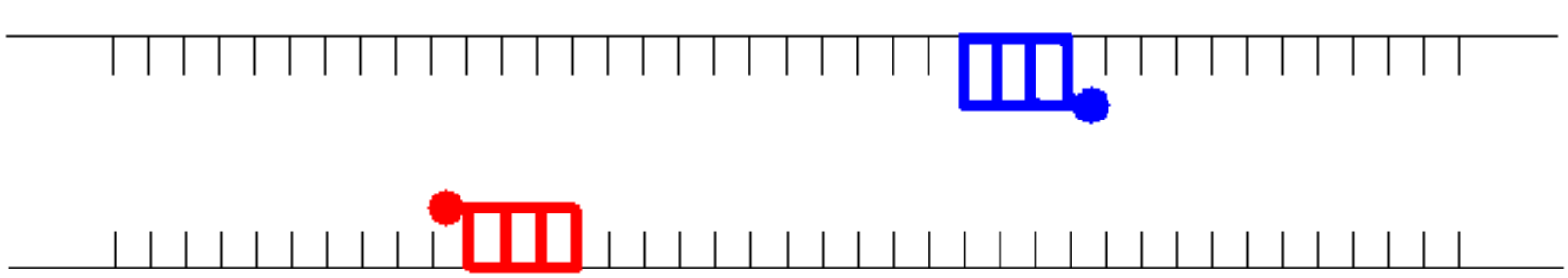
Annealing

- Anneal primers at 50-65^{o}C



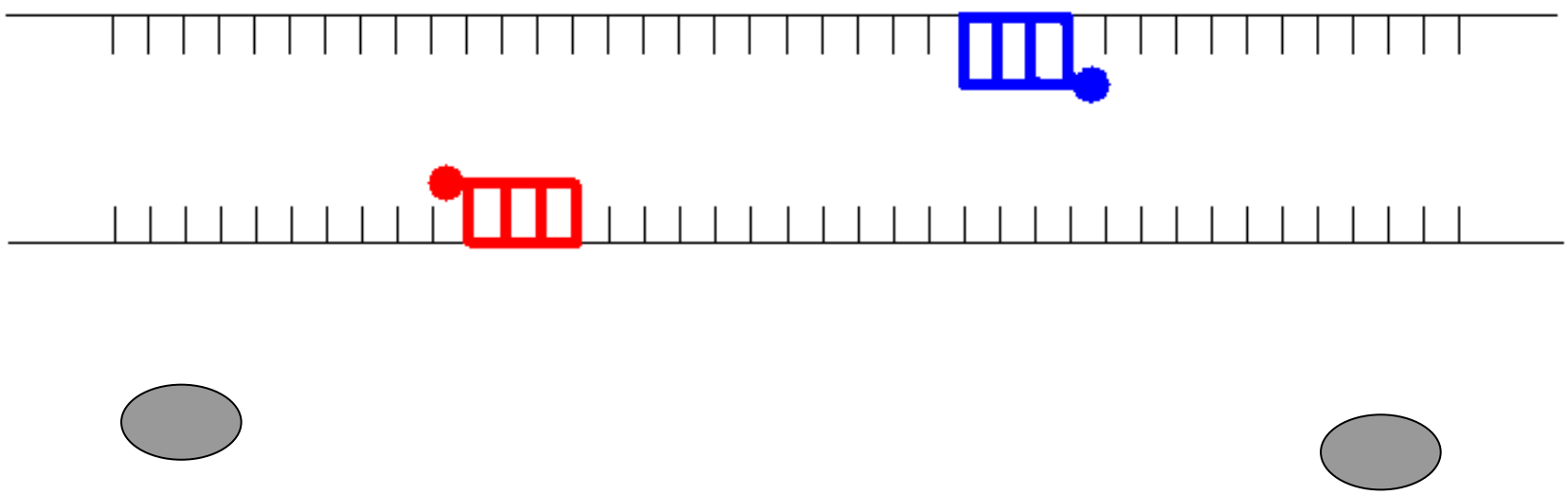
Annealing

- Anneal primers at 50-65^{o}C



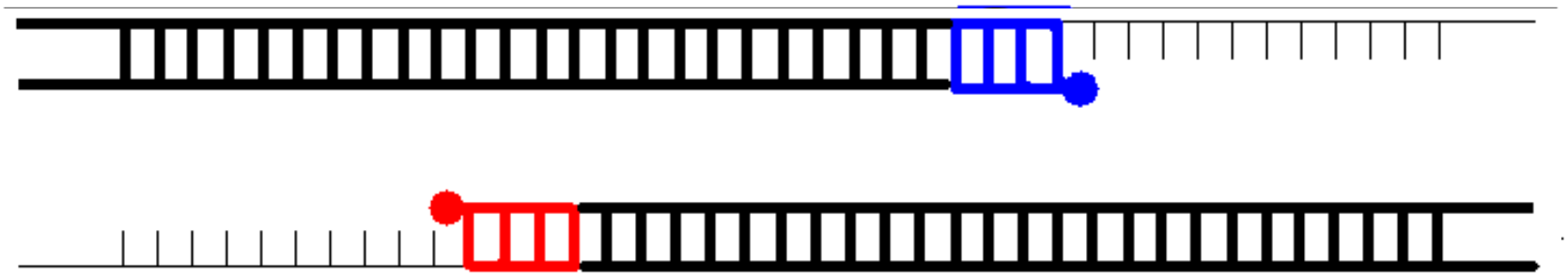
Extension

- Extend primers: raise temp to 72^{o}C, allowing Taq DNA polymerase to attach at each priming site and extend a new DNA strand



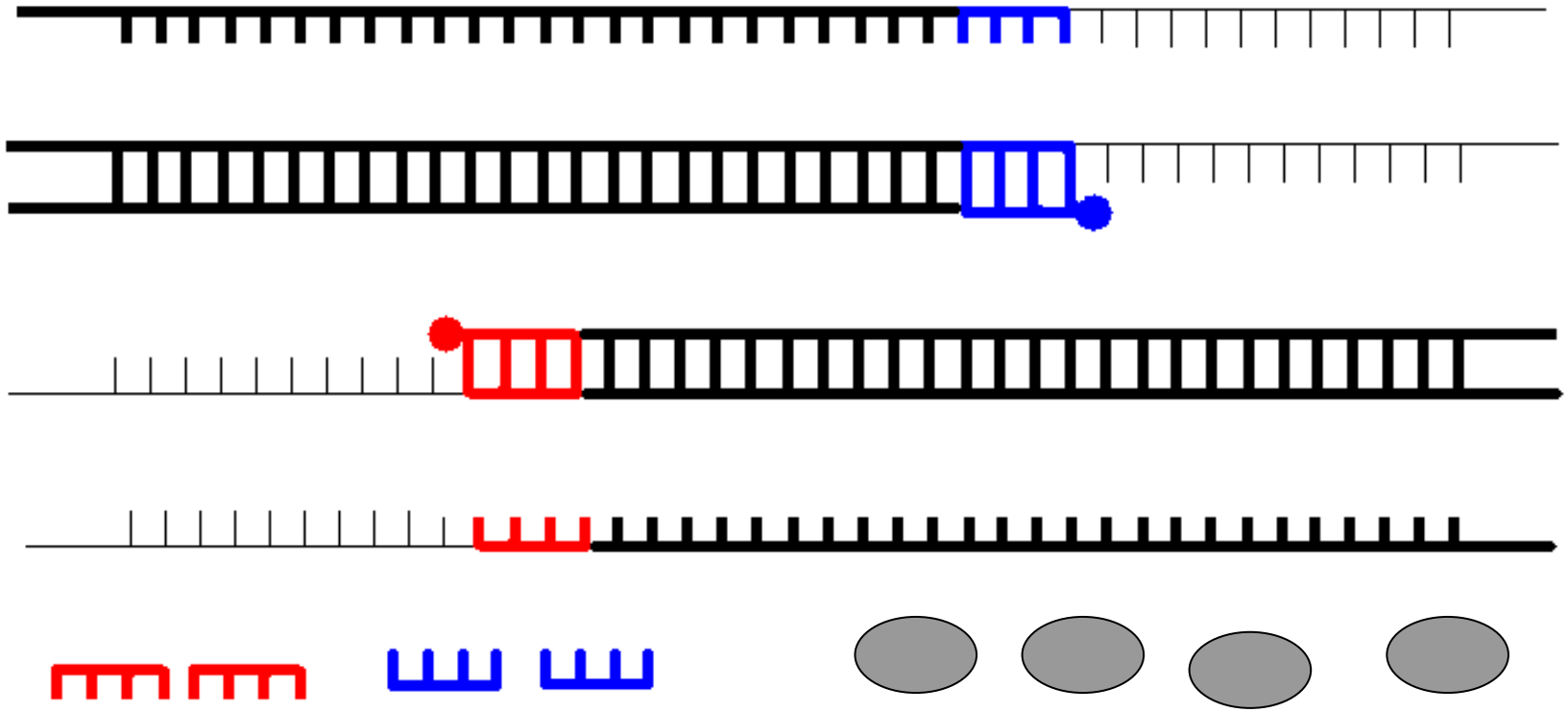
Extension

- Extend primers: raise temp to 72°C, allowing Taq pol to attach at each priming site and extend a new DNA strand

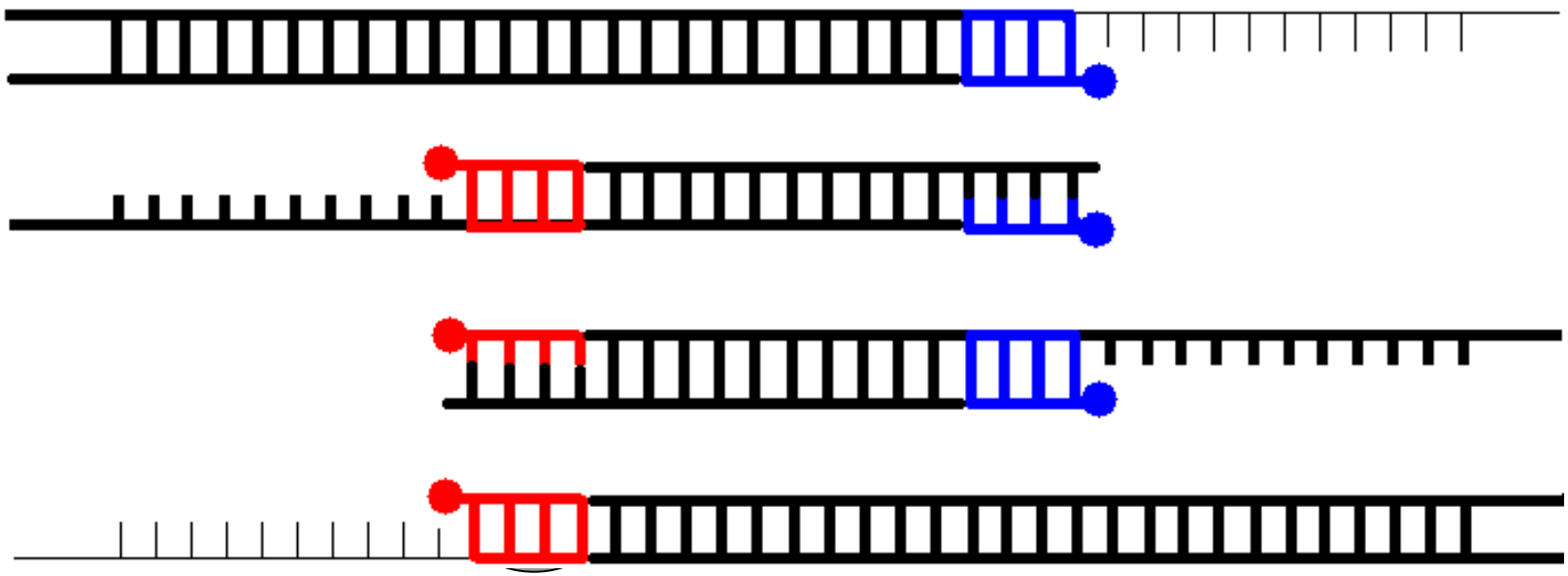


Repeat

- Repeat the Denature, Anneal, Extension steps at their respective temperatures…



Polymerase Chain Reaction



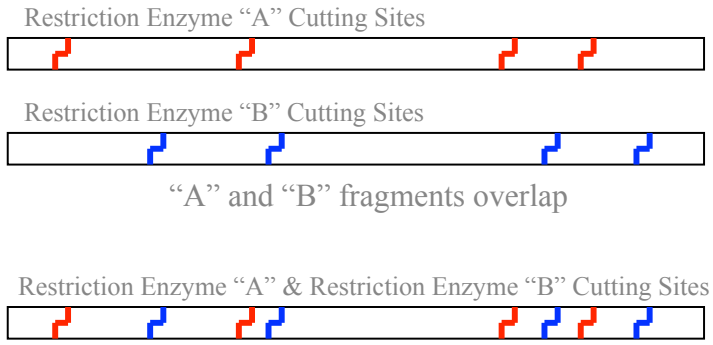
Cutting and Pasting DNA

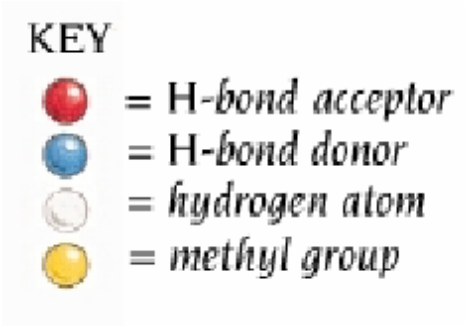
Restriction Enzymes

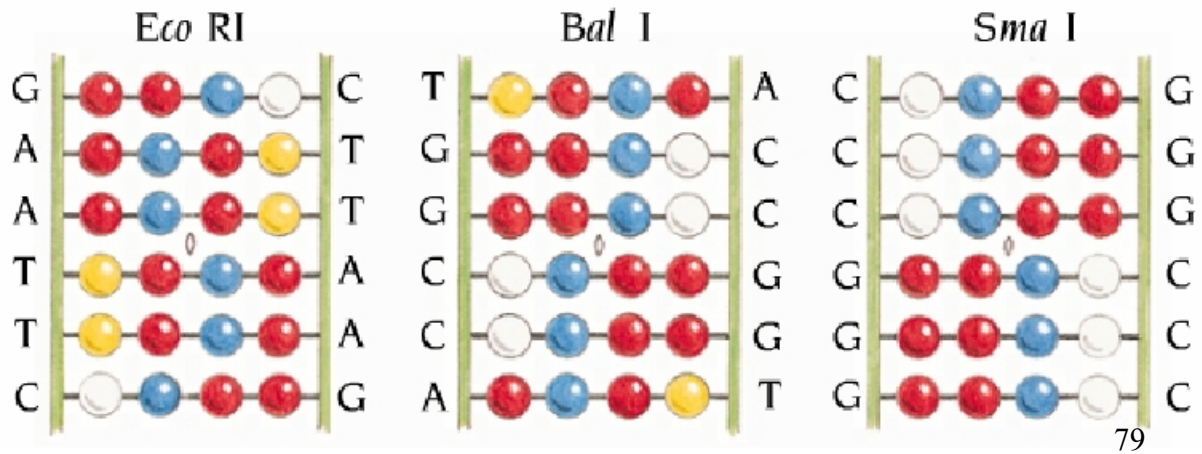
- Discovered in the early 1970's
  - Used as a defense mechanism by bacteria to break down the DNA of attacking viruses.
  - They cut the DNA into small fragments.
- Can also be used to cut the DNA of organisms.
  - This allows the DNA sequence to be in a more manageable bite-size pieces.
- It is then possible using standard purification techniques to single out certain fragments and duplicate them to macroscopic quantities.

Cutting DNA

- Restriction Enzymes cut DNA
  - Only cut at special sequences
- DNA contains thousands of these sites.
- Applying different Restriction Enzymes creates fragments of varying size.

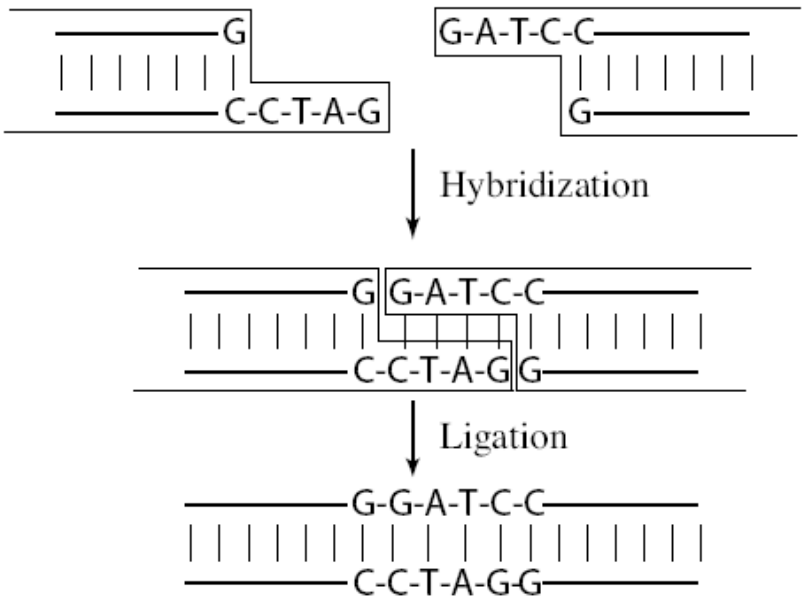






Pasting DNA

- Two pieces of DNA can be fused together by adding chemical bonds
  - Hybridization – complementary base-pairing
  - Ligation – fixing bonds with single strands

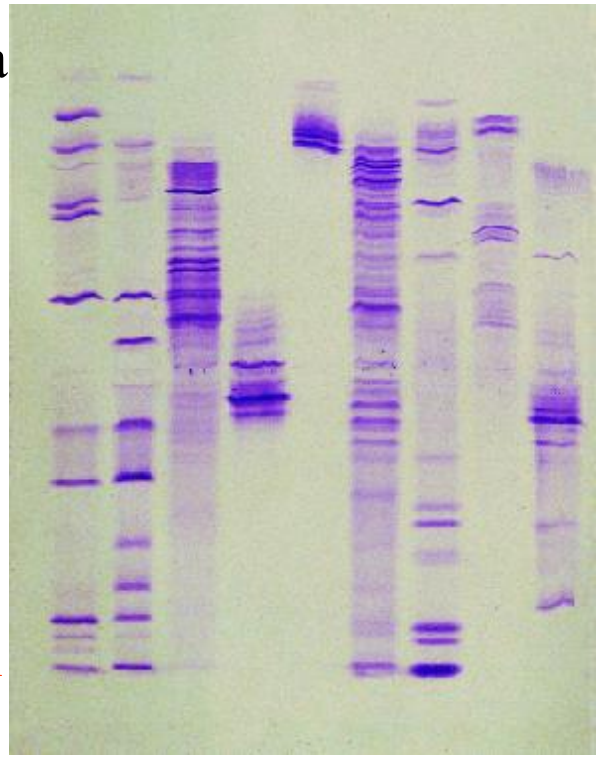


Measuring DNA Length

Electrophoresis

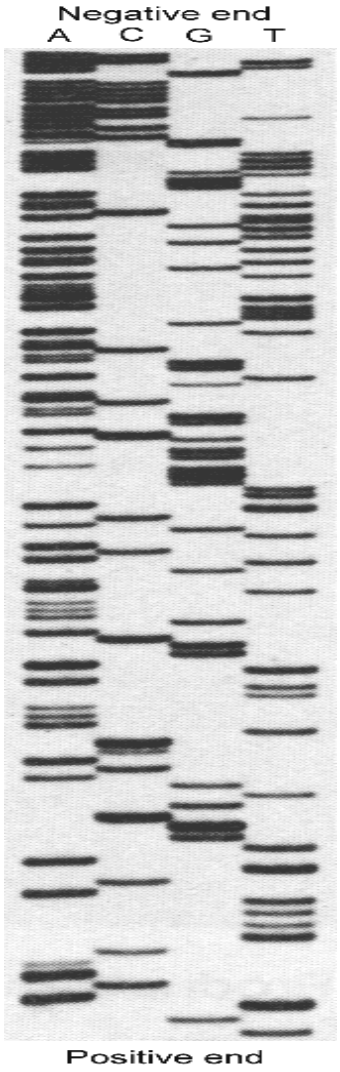
- A copolymer of mannose and galactose, agaraose, when melted and recooled, forms a gel with pores sizes dependent upon the concentration of agarose

- The phosphate backbone of DNA is highly negatively charged, therefore DNA will migrate in an electric field
  - The size of DNA fragments can then be determined by comparing their migration in the gel to known size standards.



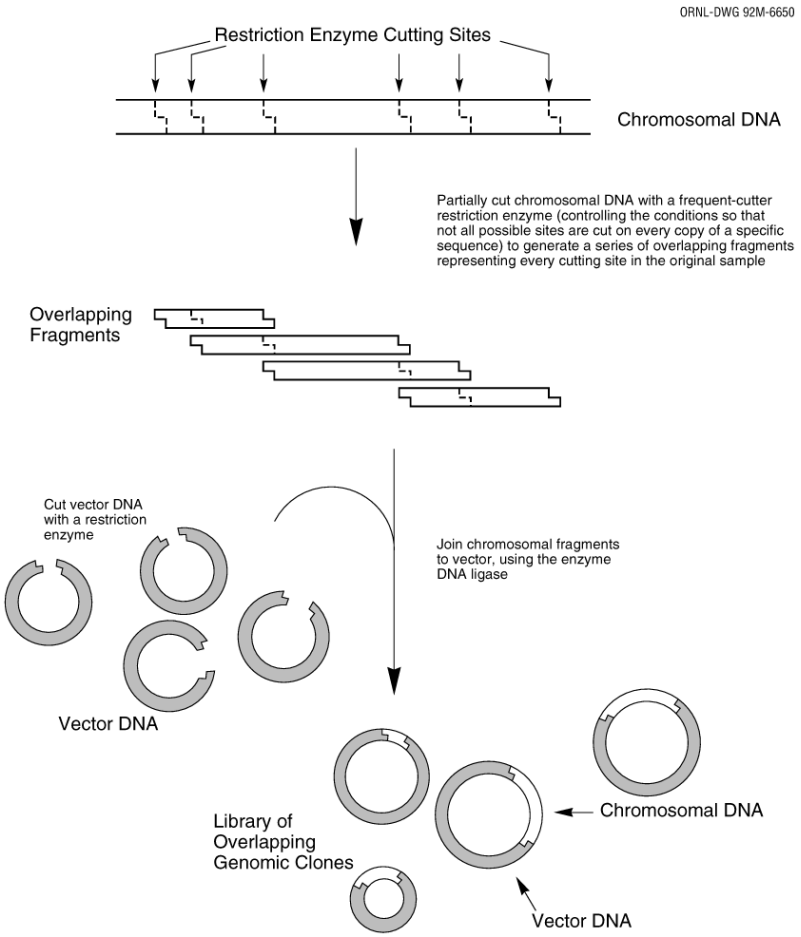
Reading DNA

- Electrophoresis
  - Reading is done mostly by using this technique. This is based on separation of molecules by their size (and in 2D gel by size and charge).
  - DNA or RNA molecules are charged in aqueous solution and move to a definite direction by the action of an electric field.
  - The DNA molecules are either labeled with radioisotopes or tagged with fluorescent dyes. In the latter, a laser beam can trace the dyes and send information to a computer.
  - Given a DNA molecule it is then possible to obtain all fragments from it that end in either A, or T, or G, or C and these can be sorted in a gel experiment.
- Another route to sequencing is direct sequencing using gene chips.



Assembling Genomes

- Must take the fragments and put them back together
  - Not as easy as it sounds.
- SCS Problem (Shortest Common Superstring)
  - Some of the fragments will overlap
    - Fit overlapping sequences together to get the shortest possible sequence that includes all fragment sequences



Assembling Genomes

- DNA fragments contain sequencing errors
- Two complements of DNA
  - Need to take into account both directions of DNA
- Repeat problem
  - 50% of human DNA is just repeats
  - If you have repeating DNA, how do you know where it goes?

Probing DNA

DNA probes

- Oligonucleotides: single-stranded DNA 20-30 nucleotides long
- Oligonucleotides used to find complementary DNA segments.
- Made by working backwards---amino acid sequence----mRNA---cDNA.
- Made with automated DNA synthesizers and tagged with a radioactive isotope.

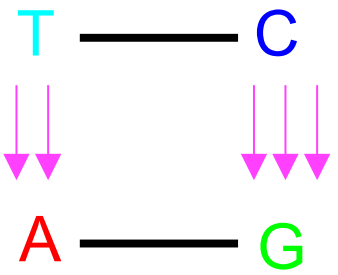
87

DNA Hybridization

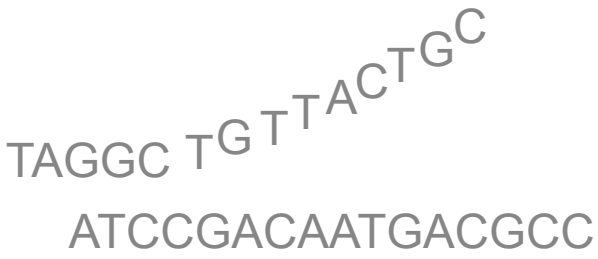
- Single-stranded DNA will naturally bind to complementary strands.
- Hybridization is used to locate genes, regulate gene expression, and determine the degree of similarity between DNA from different sources.
- Hybridization is also referred to as annealing or renaturation.

88

1. Hybridization is binding two genetic sequences. The binding occurs because of the hydrogen bonds [pink] between base pairs.

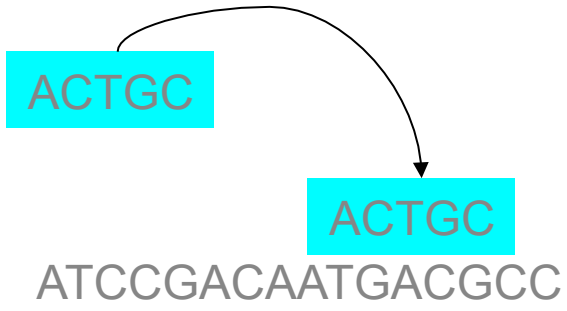


2. When using hybridization, DNA must first be denatured, usually by using heat or chemicals.



Create a Hybridization Reaction Cont.

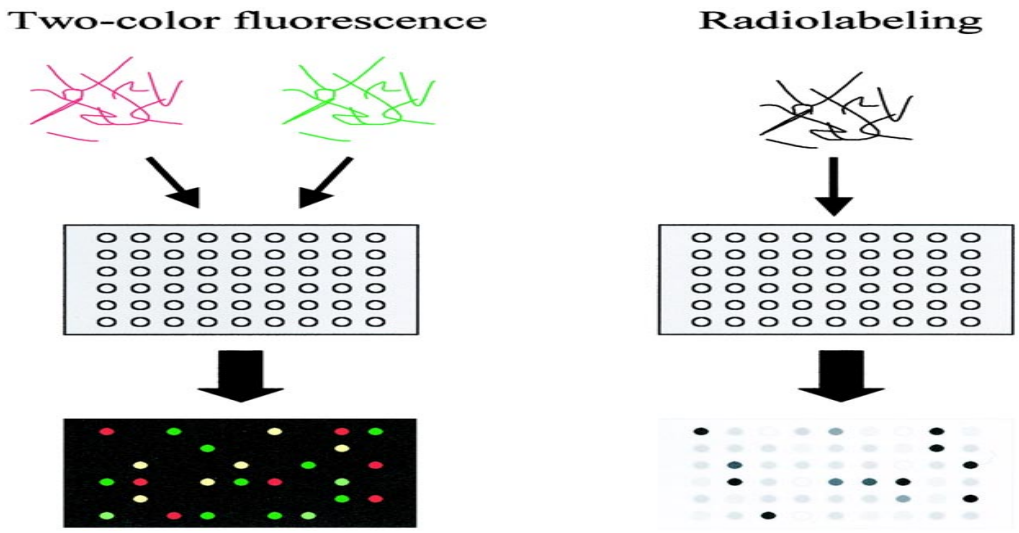
3. Once DNA has been denatured, a single-stranded radioactive probe [light blue] can be used to see if the denatured DNA contains a sequence complementary to probe.



4. Sequences of varying homology stick to the DNA even if the fit is poor.

90

Labeling technique for DNA/RNA arrays



RNA samples are labeled using fluorescent nucleotides (*left*) or radioactive nucleotides (*right*), and hybridized to arrays. For fluorescent labeling, two or more samples labeled with differently colored fluorescent markers are hybridized to an array. Level of RNA for each gene in the sample is measured as intensity of fluorescence or radioactivity binding to the specific spot. With fluorescence labeling, relative levels of expressed genes in two samples can be directly compared with a single array.

DNA Arrays--Technical Foundations

- An array works by exploiting the ability of a given mRNA molecule to hybridize to the DNA template.

- Using an array containing many DNA samples in an experiment, the expression levels of hundreds or thousands genes within a cell by measuring the amount of mRNA bound to each site on the array.

- With the aid of a computer, the amount of mRNA bound to the spots on the microarray is precisely measured, generating a profile of gene expression in the cell.

92

An experiment on a microarray

In this schematic:

GREEN represents Control DNA

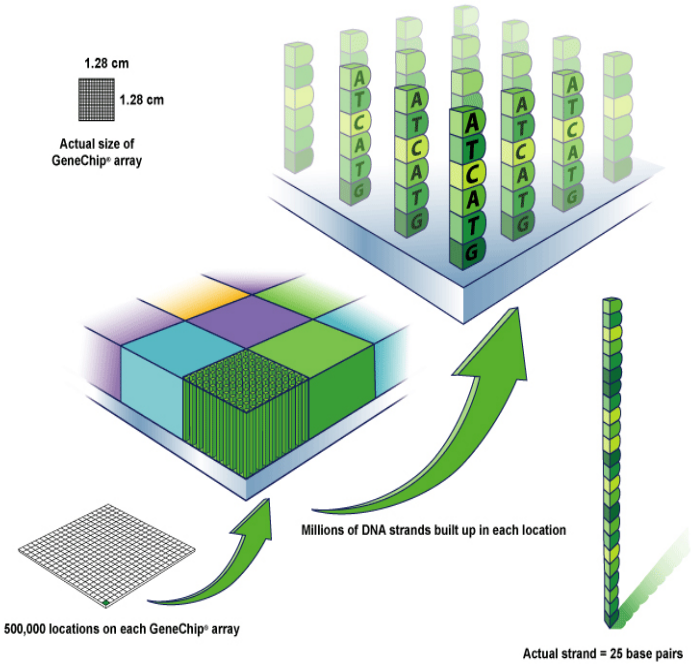
RED represents Sample DNA

YELLOW represents a combination of Control and Sample DNA

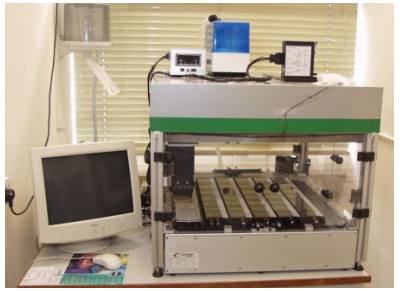
BLACK represents areas where neither the Control nor Sample DNA

Each color in an array represents either healthy (control) or diseased (sample) tissue. The location and intensity of a color tell us whether the gene, or mutation, is present in the control and/or sample DNA.

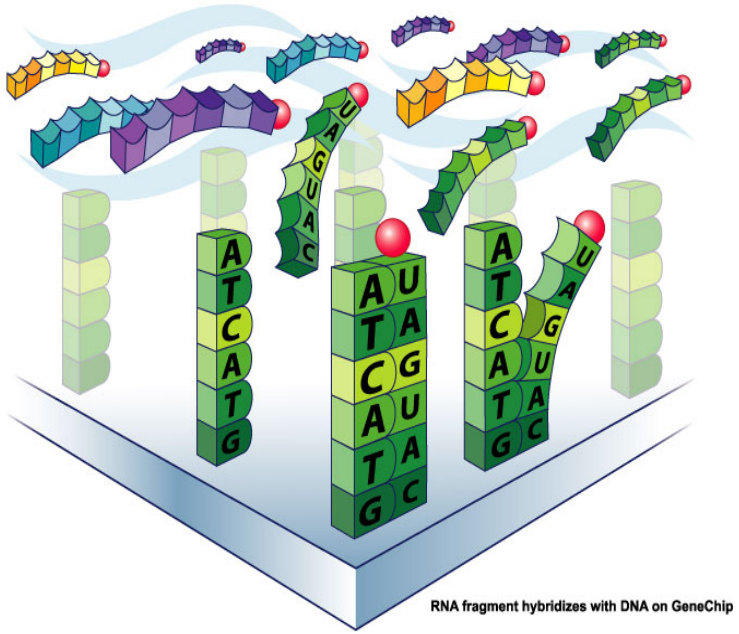
DNA Microarray



Millions of DNA strands build up on each location.



RNA fragments with fluorescent tags from sample to be tested



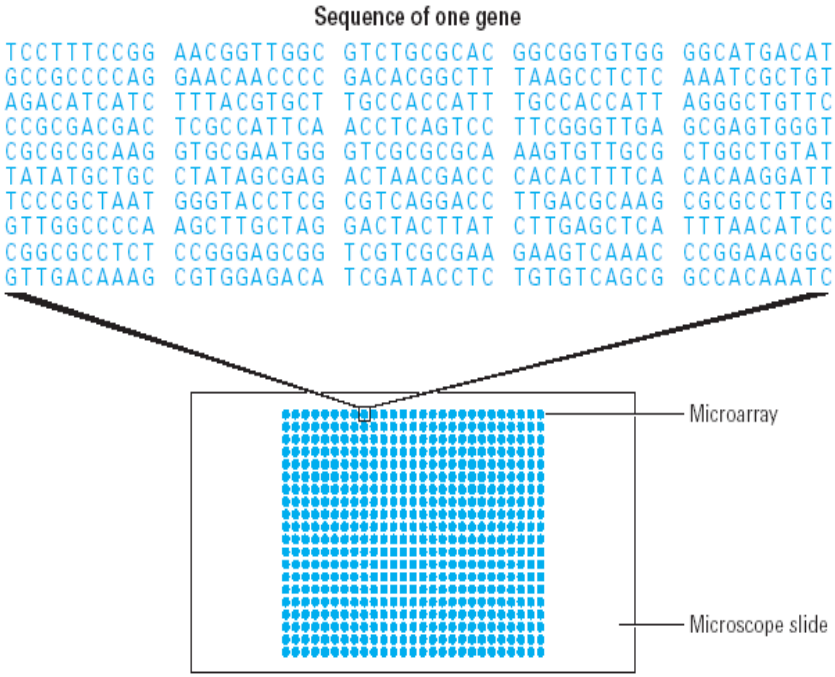
Tagged probes become hybridized to the DNA chip's microarray.

94

DNA Microarray

Affymetrix

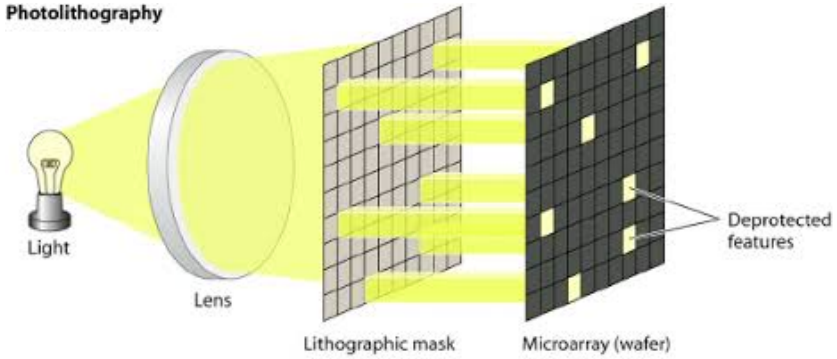
Microarray is a tool for analyzing gene expression that consists of a glass slide.

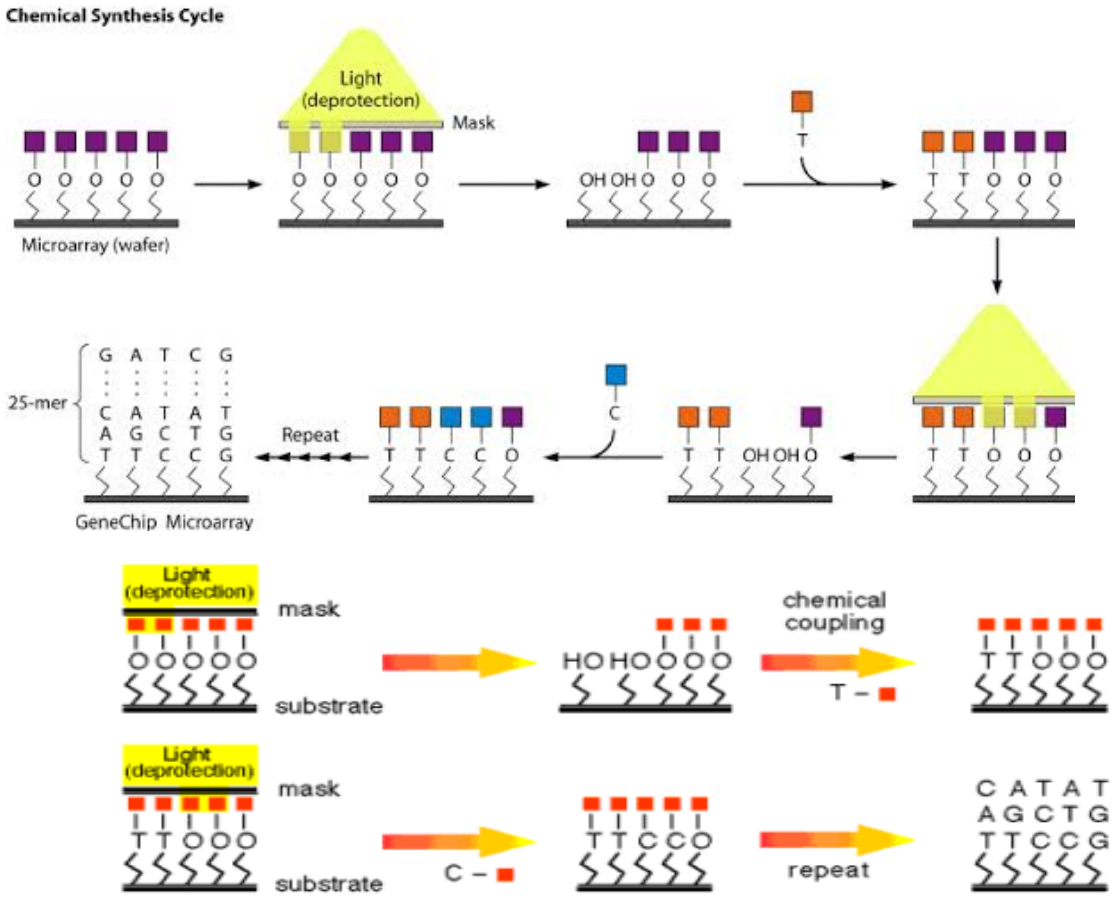


Each blue spot indicates the location of a PCR product. On a real microarray, each spot is about 100um in diameter.

95

Photolithography

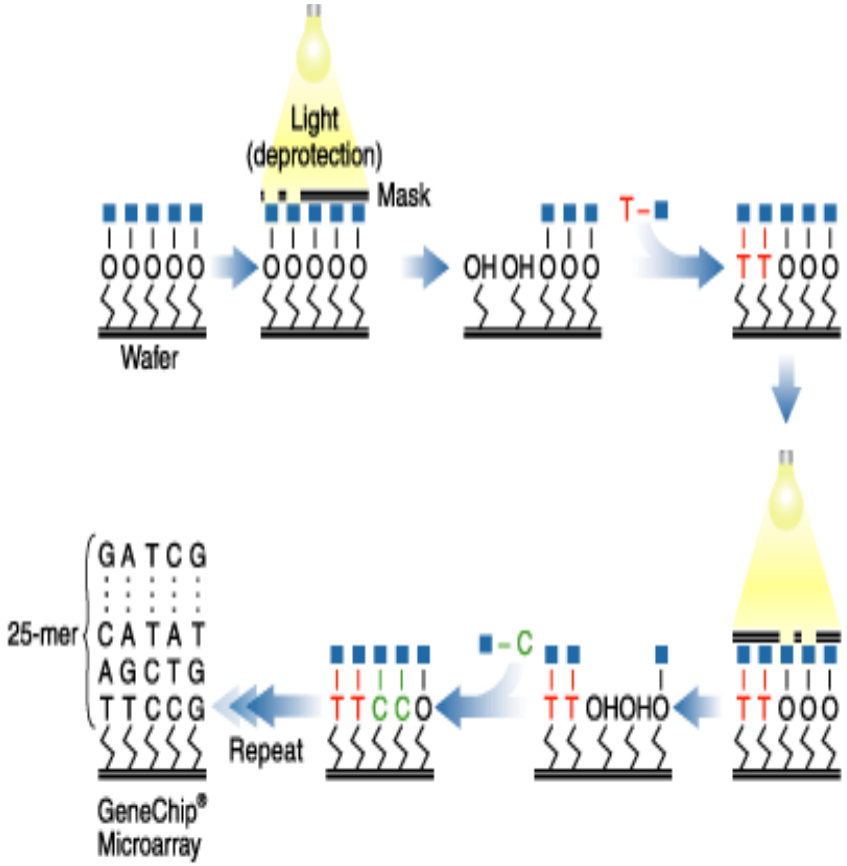


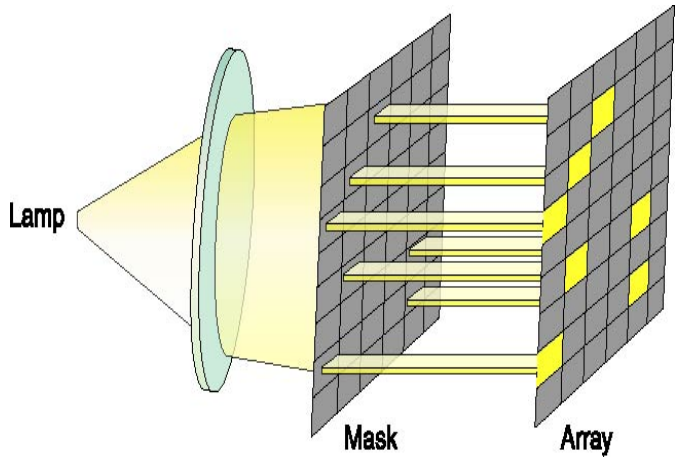


- Light directed oligonucleotide synthesis.
- A solid support is derivatized with a covalent linker molecule terminated with a photolabile protecting group.
- Light is directed through a mask to deprotect and activate selected sites, and protected nucleotides couple to the activated sites.
- The process is repeated, activating different set of sites and coupling of different bases allowing arbitrary DNA probes to be constructed at each site.

96

Affymetrix GeneChip® Arrays

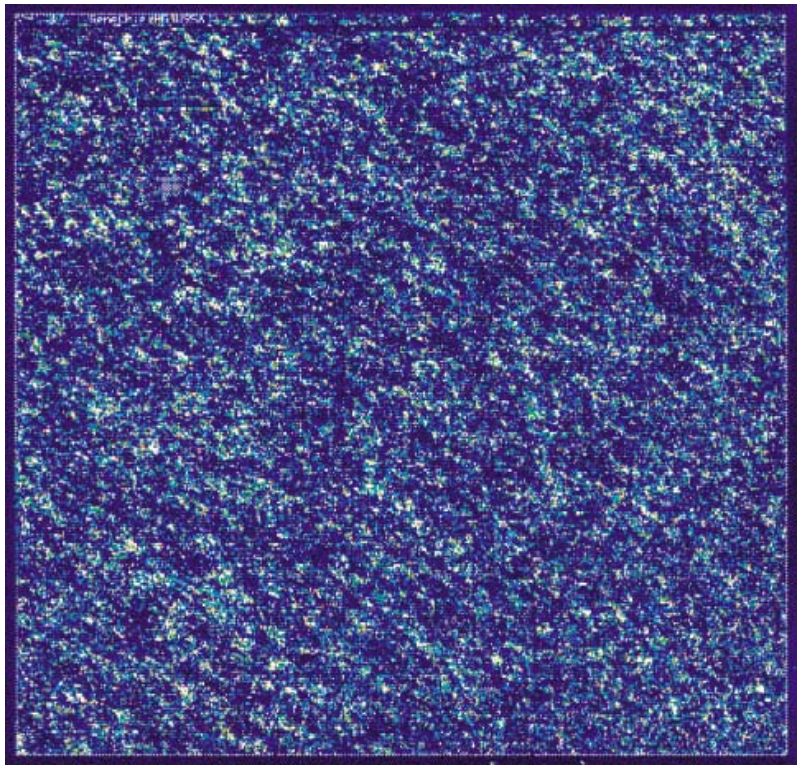




- A combination of photolithography and combinatorial chemistry manufactures GeneChip® Arrays.
- With a minimum number of steps, Affymetrix produces arrays with thousands of different probes packed at extremely high density.
- Enables to obtain high quality, genome-wide data using small sample volumes.

Affymetrix GeneChip® Arrays

Data from an experiment showing the expression of thousands of genes on a single GeneChip® probe array.



http://www.affymetrix.com/corporate/media/image_library/image_library_1.affx