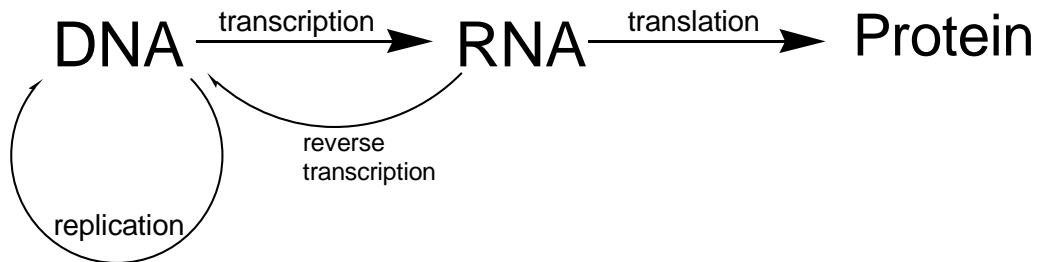


# *Introduction to Molecular Genetics*

## *I. INTRODUCTION*

- A. Have you thought much about why you are like your parents, and your kids are (will be) like you? (Identical twins?)
1. Inherited traits result from the transfer of specific molecules to offspring.
  2. All living things use DNA to store inherited information and transfer it to offspring.
  3. Some viruses use RNA for information storage.
- B. Organisms organize their genetic information.
1. Chromosomes (bacteria may have just one)
  2. Operons (grouped, related genes: prokaryotes)
  3. Genes
- C. “The Central Dogma” of molecular biology:  
DNA makes RNA makes Protein.

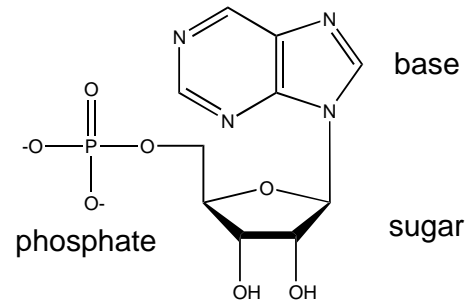


## *NUCLEIC ACID COMPONENTS*

- A. General thoughts
1. Two main categories
    - a) RNA: ribonucleic acid
    - b) DNA: 2'-deoxyribonucleic acid

2. A nucleotide has 3 components:

- a) base
- b) sugar
- c) phosphate(s)



## B. Bases

1. RNA has mostly G, C, A, & U:

- a) Guanine (G)
- b) Cytosine (C)
- c) Adenine (A)
- d) Uracil (U)
- e) Also, many (tRNA) other bases derived from a-d, above.

2. DNA has

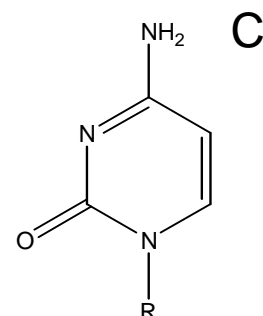
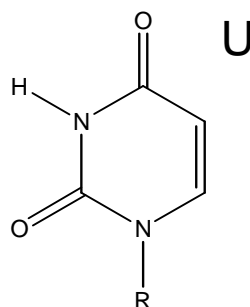
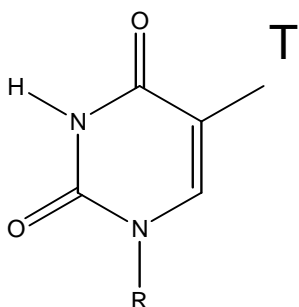
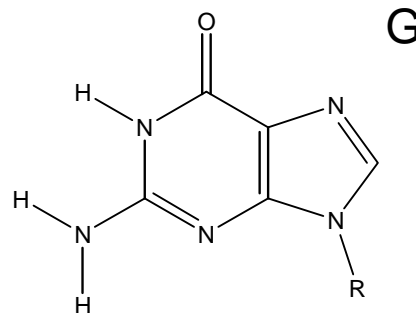
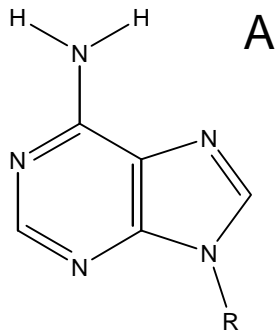
- a-c) the same as RNA
- d) Thymine (T) in place of U. Why? (later!)
- e) Also, other bases derived from a-d, above.

3. Purines: G & A. Pyrimidines: C, T, & U.

- a) Hydrogen Bonding sites are extremely important
- b) Identify all sites in & on 6-membered rings below

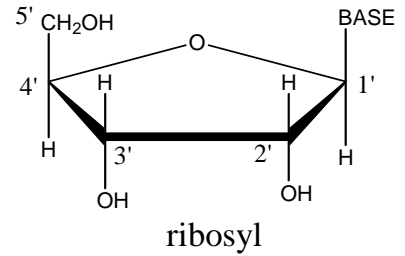
Use "d" for the donor sites.

Draw non-bonding pairs for acceptor sites:



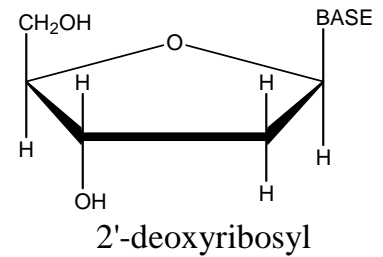
### C. Sugar

1. Ribose in RNA
2. 2'-Deoxyribose in DNA
3. *Nucleoside*: base linked to sugar



### D. Phosphate(s)

1. Usually a 5'-linkage
2. *Nucleotides*: mono-, di-, triphosphates
3. Other linkages (3',5'-cAMP)
4. Comment: ATP in metabolism.

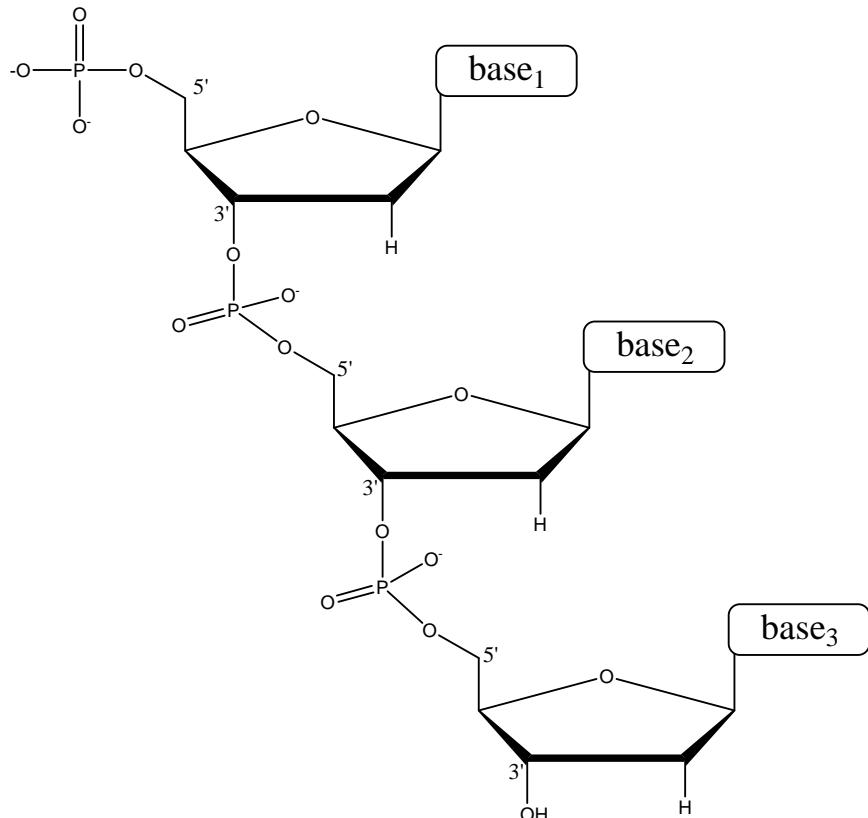


## STRUCTURE OF DNA AND RNA

### A. The first level is 1° structure (sequence).

1. Like proteins, nucleic acid polymers have distinct ends: the 5'-end & 3'-end.

### DNA Sugar-phosphate Backbone

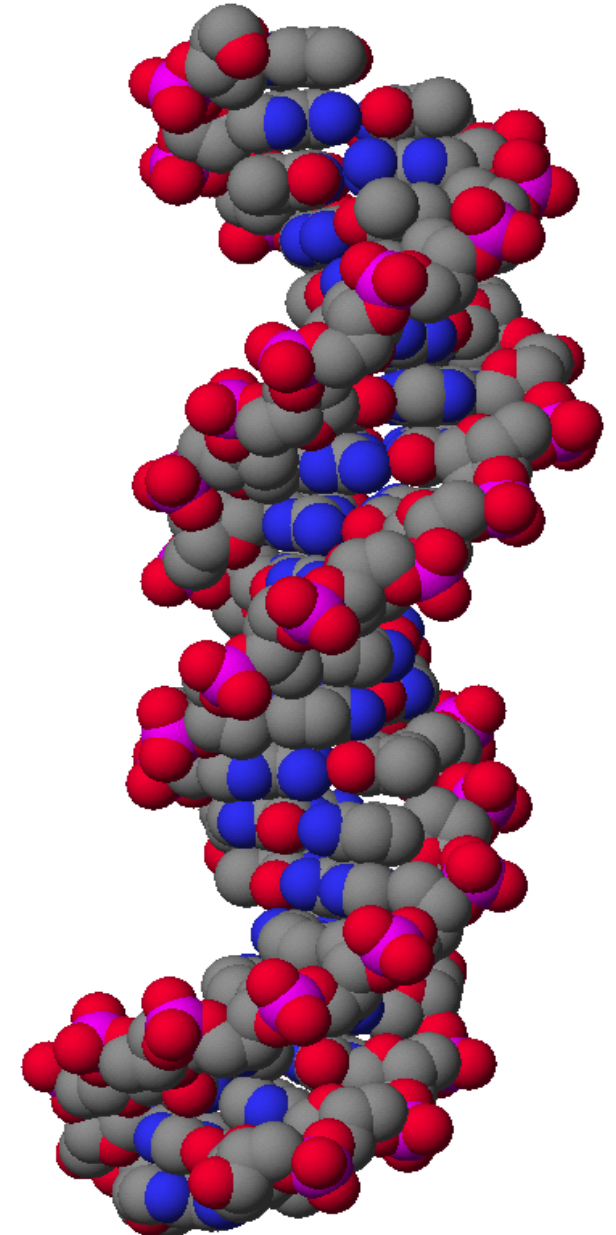
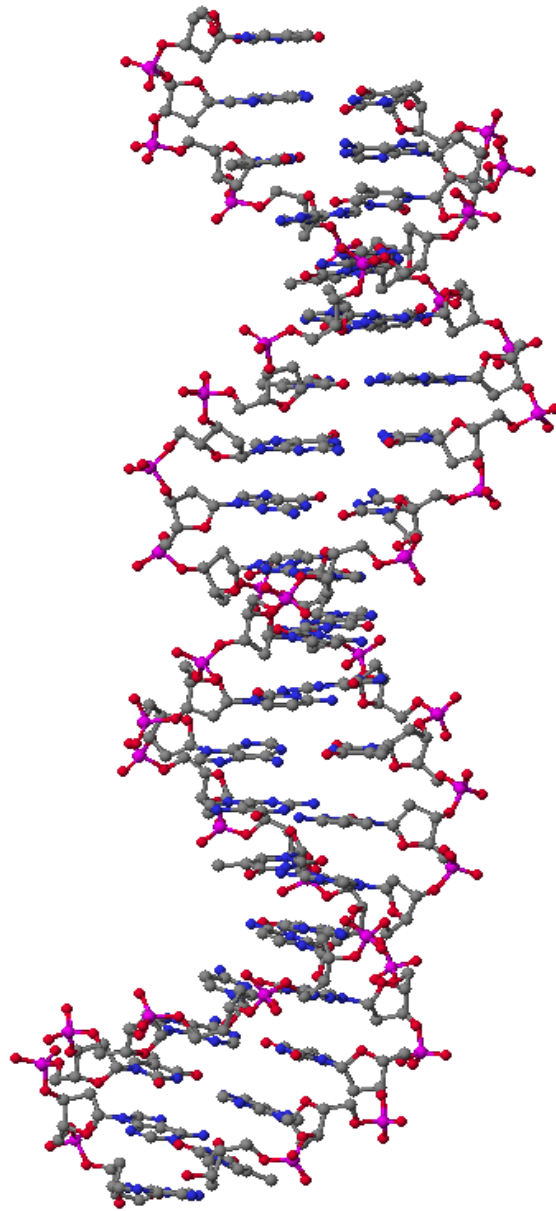
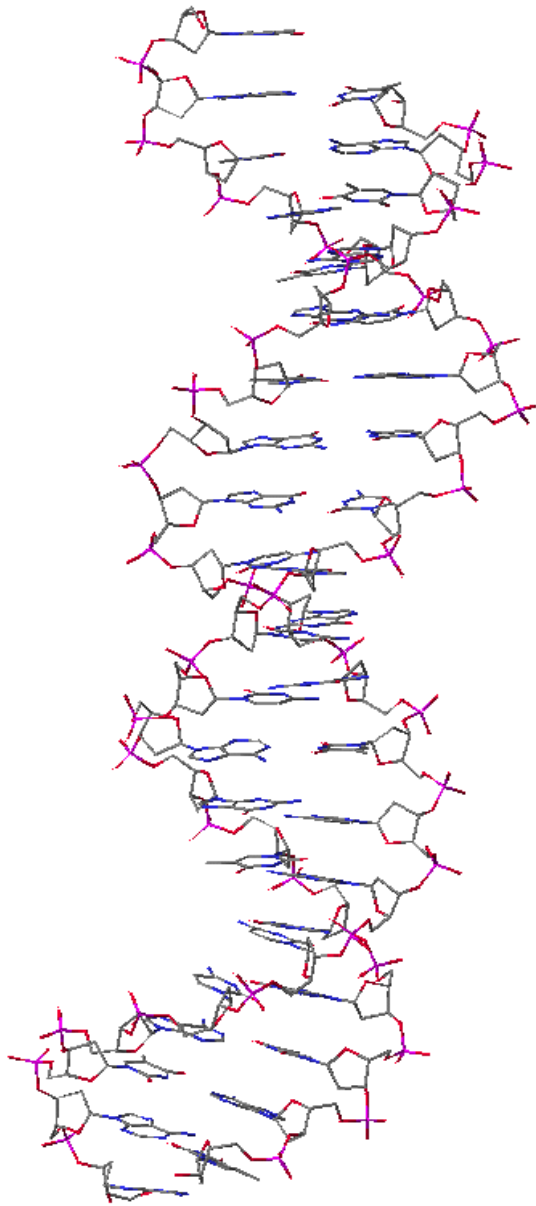


2. There *are* cyclic DNA molecules. Their chains still have 5'- and 3'- directionality, even though they don't formally have ends.
3. To find direction of chain, pick a sugar. Find 5'- & 3'- sites, determine which end is which.

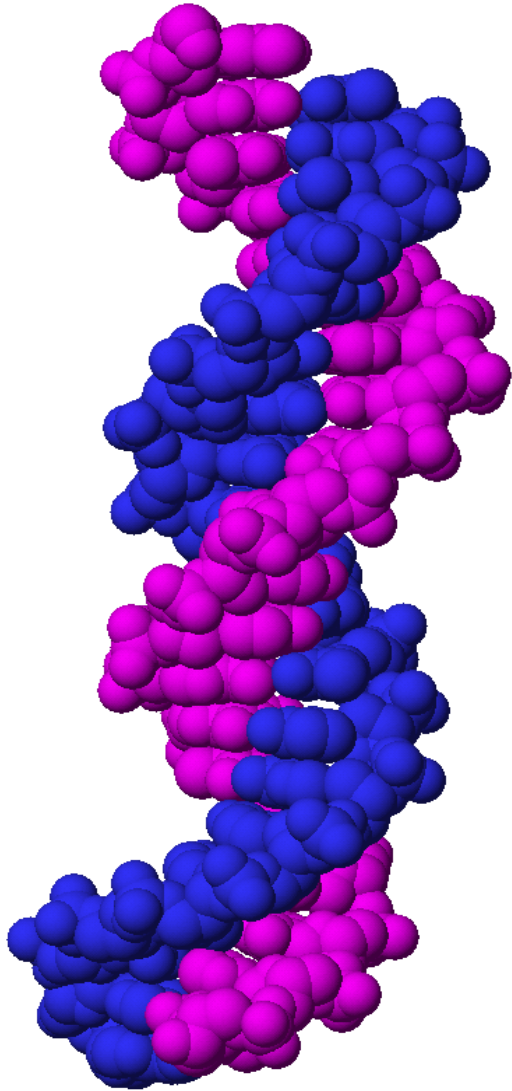
## B. Secondary structure: *The Double Helix!!!*

1. Aside:
  - a) In RNA this can be viewed as 2° structure.
  - b) In DNA it is usually more accurately viewed as 2° & 4° structure combined. (Tertiary structure? \_\_\_\_\_  
\_\_\_\_\_)
2. Determination of the double helical nature of DNA was among the most important biology-chemistry achievements in the 20<sup>th</sup> century.
  - a) *Circa* 1950, people knew:
    - i) DNA had regular, repeating structures
    - ii) A & T occur in = %, as do C & G
  - b) Watson & Crick (W-C) used this info to propose a structural model for DNA.
  - c) The structure of this model immediately clarified many aspects of DNA function.
3. Key to determining DNA structure was seeing the importance of **complimentary Hydrogen Bonding** between A & T and G & C.
4. Look at at DNA double helix. (See next page.) This figure was prepared from pdb file 1lmb, which is the lambda repressor protein bound to its operator DNA. The repressor functions to regulate transcription (see below) from nearby genes. The file was transferred to CACHE and the protein deleted to aid in viewing.
5. Comment on:
  - a) strands are anti-parallel (5'→3' vs. 3'→5')
  - b) non-W-C Hydrogen Bonding areas of the bases “see” solvent, re. DNA function.

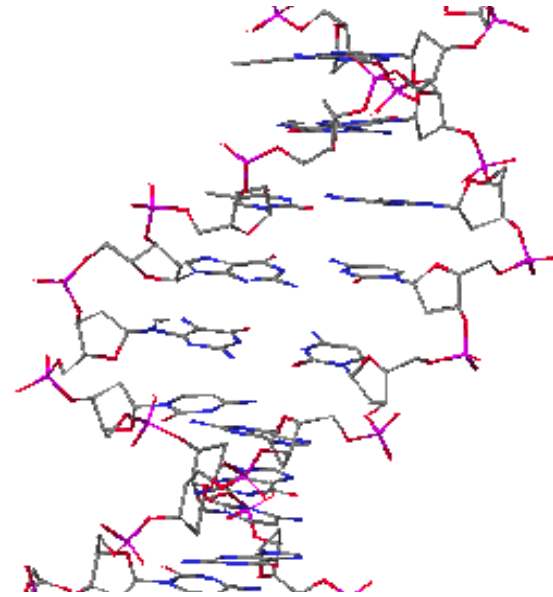
# Bacteriophage lambda operator DNA (largley in B form)



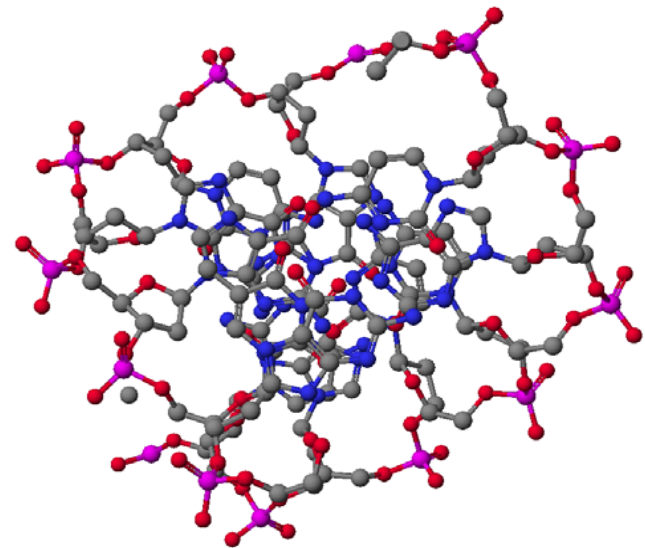
Doubled stranded color emphasis



Detail showing W-C base pairing



View along helical axis for 1/2 turn



C. Chromosomal structure see: <http://www.youtube.com/watch?v=IUESmHDrN40>  
It runs from full folded, wrapped, etc. chromosome down to double helix.

## *DNA REPLICATION (initiate-elongate-terminate)*

Go to dnai.org & click on CODE. Go to *Finding the structure*, then to problem, pieces of the puzzle, and putting it together. **(Do the interactives!!!)**

A. DNA replication is *semi-conservative*.

1. Parent DNA strands separate (at specific sites).
2. Each parent strand serves as a template to make a new daughter strand.
3. This gives 2 “half-new” complementary strands. Emphasize:
  - a) A pairs only with T                      See pdb structure 137d
  - b) G pairs only with C

B. **Many** enzymes & proteins are involved in DNA replication (3 different types of polymerases!).

1. The main protein involved in making the new DNA copies is called *DNA polymerase*.
2. Some proteins help polymerase get started.
3. Other proteins help the DNA unwind and keep short stretches of DNA single-stranded.
4. Still other proteins help rewind & terminate.

C. DNA polymerases synthesize the new strand in the 5' to 3' direction.

Therefore: at dnai.org go to *Copying the Code*

<http://www.dnai.org/a/index.html>

1. Leading strand: topologically 5' to 3' direction.
2. Lagging strand: looks 3'→5' direction, but is actually 5'→3' in short (100 base pair) chunks.

D. Many bacterial DNA's are circular.

E. Is your nuclear DNA circular? \_\_\_\_\_

1. This can lead to problems with shortened ends.
2. *Telomerases* help cope with this.
3. Some immortal (including cancer) cells have enhanced telomerase activity.

# INFORMATION FLOW IN BIOLOGICAL SYSTEMS

DNA → RNA → protein

## A. Main types of RNA:

1. Messenger RNA (mRNA)
2. Transfer RNA (tRNA)
3. Ribosomal RNA (rRNA)
4. Other RNAs

## B. mRNA codes for the synthesis of proteins. The largest part of your DNA that codes for RNA codes for this type (20,000+ different mRNAs!)

1. Bacterial mRNAs correlate directly with the genes that code for them.
2. Most of your mRNAs are made from much larger precursors that you trim down to the right size. Draw picture on the board.
  - a) Eucaryotic genes often contain *introns*. Introns (or intervening sequences) are spliced out of the initial RNA molecule shortly after it is made.
  - b) We make additional chemical modifications to the 5' and 3' ends of the mRNAs before they are used in protein synthesis.

## C. tRNA serves an important translational function in protein synthesis at the ribosome.

1. At the ribosome one end (anti-codon loop) of tRNAs binds (by W-C base pairing) to complimentary RNA triplets in the mRNA.
2. The other end covalently links to an amino acid.
3. The middle provides recognition sites so the tRNA charging enzymes will link the correct amino acid to the correct tRNA.

## D. rRNA is an important *catalytic* and structural component of the ribosomes. Ribosomes contain more than one rRNA.

## E. Other RNAs. There are many, but they are slightly beyond our scope.

## F. Transcription is catalyzed by RNA polymerase. See [dnai.org](http://dnai.org) *Copying the Code*. Again, the pattern is:

1. Initiation (at specific sites identified by specific DNA sequences)
2. Elongation (70- a few thousand bases are added)
3. Termination (at specific sites identified by specific DNA sequences)

G. Many eukaryotic mRNAs are processed before being translated into the amino acid sequences of proteins.

1. Noncoding sequences, called *introns*, are removed.
2. Coding regions (*exons*) are spliced together to form a continuous strand, the 5' end is capped, & the 3' end has poly A added.
3. In some mRNAs: intron RNA is 10-30 times longer than exon RNA.
4.  $\approx 3\%$  of our DNA appears to code for proteins.

## *THE GENETIC CODE* You don't need to memorize it.

A. How do you get from the language of nucleotides to the language of amino acids?

1. How many aa's could you code for using a 4 base alphabet (A, U, G, & C) and one letter long words? \_\_\_\_\_
2. How many aa's could you code for using a 4 base alphabet and two letter long words? \_\_\_\_\_
3. How many aa's could you code for using a 4 base alphabet and three letter long words? \_\_\_\_\_
4. How many "common amino acids" are there? \_\_\_\_\_

The words are called *triplet codons*.

B. Comments on genetic code: See for example:

<http://images.google.com/images?hl=en&q=genetic+code&btnG=Search+Images&gbv=2>

1. It is redundant, multiple codons for most aa.
2. Three of the codons are *stop signals*.
3. The code is nearly universal. (re. evolution)
  - a) Bacteria use the same code we do.
  - b) Exceptions to pattern fit nicely into established evolutionary patterns.
    - i) mostly mitochondrial ( $\approx 10-20$  proteins)
    - ii) derivation from an "original" code

C. *Codons* are within the mRNA. An *anti-codon* is located in the anticodon loop of each tRNA. See pdb structure 4tna.

# PROTEIN SYNTHESIS

A. Energetics of protein synthesis.  $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$

1. Is this rxn. favored:  $\text{protein} + \text{H}_2\text{O} \rightarrow \text{amino acids}$   
What does meat tenderizer do, and how does it work?
2. If rxn.:  $\text{protein} + \text{H}_2\text{O} \rightarrow \text{amino acids}$  has negative  $\Delta G^\circ$ , can the reverse rxn. have negative  $\Delta G^\circ$  (be favored)? \_\_\_\_\_
3. Since  $G^\circ$  of free amino acids is too low for their use as reactants in protein synthesis, we need to make a higher  $G^\circ$  reactant. (Graph)
4. If proteins are less stable than their monomers, why don't we (our proteins) just fall apart?  
Comment on kinetic vs. equilibrium "stability."  
(Graph?)
5. Amino acids are activated (converted to higher  $G^\circ$  compounds):
  - a) by reaction with ATP followed by
  - b) ester linkage to 3' ribose of proper tRNA
  - c) rxns. a) & b) catalyzed by *aminoacyl tRNA synthetases*.  
Essentially a different charging enzyme for each different tRNA.

See [http://www.biostudio.com/demo\\_freeman\\_protein\\_synthesis.htm](http://www.biostudio.com/demo_freeman_protein_synthesis.htm) animation.

See also at dnai.org *Reading the Code*.

B. Three major stages:

1. initiation
2. elongation
3. termination

C. Initiation

1. mRNA, small ribosomal subunit, & charged initiator  $\text{tRNA}^{\text{met}}$  form *ternary complex*. A specific region (base sequence) of the mRNA binds the ribosomal subunit.

What minimal 3 base sequence of the mRNA must be involved?\_\_\_\_\_

2. A number of initiation factors (protein) also act.
3. Large ribosomal subunit now binds.
4. Aside on ribosomes (from rat liver unless noted):
  - a) They are *giant* enzymes. (MW =  $4.22 \times 10^6$ )

What rxn. do they catalyze?\_\_\_\_\_

- b) 2 subunits 40S/60S (30S/50S in prokaryotes)
- c) 82 different proteins (40% of weight)
- d) 4 different rRNAs (60% of weight)  
See pdb structures 1ffk and 1gix.
5. GTP hydrolysis is required. (Why do we need to take in energy?)

#### D. Elongation

1. Next charged tRNA<sup>xxx</sup> binds to the complex at the aminoacyl (A) site of the ribosome. XXX determined by anti-codon:codon base pairing.
2. C-terminal end of growing peptide chain forms peptide bond with amino group of aa XXX bound to tRNA at A site.
3. “Empty” tRNA dissociates from P site.
4. Peptidyl-tRNA at A site is translocated to P site. Requires GTP.
5. Elongation factors (protein) are involved.

#### E. Termination

1. When a stop codon is encountered, termination occurs.
2. Protein *release factors* (not tRNA) recognize stop codons.
3. Specific sequences often occur near stop codon that aid in termination.

#### F. General protein synthesis comments.

1. mRNA read 5'→3'
2. Protein synthesized from -N to C-terminal end.
3. More than one ribosome can read an mRNA at a time. (Amplification!)
4. Very few errors occur.

### *Aside on GENE REGULATION*

#### A. General considerations

1. Do all cells in your body need to make equal amounts of all proteins all of the time? \_\_\_\_\_
2. Do you need hemoglobin in your nerve cells? \_\_\_\_
3. Do you need to make antibodies against measles when you have anthrax?

#### B. Operons (prokaryotes)

1. Related genes, clustered on the chromosome.
2. A single long mRNA is made that contains the coding information of all of these genes.
3. The ribosomes make proteins for each of the genes located on the mRNA. (start efficiency?)

C. Response elements: These regions of the DNA contain binding sites (specific base sequences) for factors that influence efficiency of transcription. (Steroid hormone-receptor complexes bind response elements.)

D. Transcription factors (usually protein) bind DNA at specific sites (response elements) and influence transcription rates from (usually) nearby genes.

## *MUTATIONS, DNA REPAIR & GENETIC DISEASES*

#### A. Mutations

1. A mutation is a change (fixed) in DNA.
2. Mutation types
  - a) silent (ex.: DNA coded mRNA of UUU → UUG)
  - b) conservative [mRNA AUU (Ile) → GUU (val)]
  - c) Other. These cause substantial change in protein function. (Sickle-cell anemia. Can you figure out the DNA change from Glu→Val?)

#### B. Mutagens

1. Mutagens cause mutations
2. We think of mutations as random, but specific mutagens do not necessarily occur or act randomly.

3. UV light can cause mutations (Thymine dimers).
4. Why does DNA have T, not U?

You correct  $\approx$  1 million of these/day.

Could you correct these if your DNA normally contained U?

5. Generally, mutagens are carcinogens and teratogens. (What are these?)

C. Genetic Diseases occur when a mutation has a deleterious effect.

1. Sickle-cell anemia (always deleterious?)
2. Xeroderma pigmentosum (thymine dimer repair)

D. Molecular aspects of evolution:

1. Mutations can lead to genetic advancement as well as disease.

Example: Sickle-cell anemia.

2. Comment on odds for “good” vs. “bad” mutations, related to randomness. More later?

