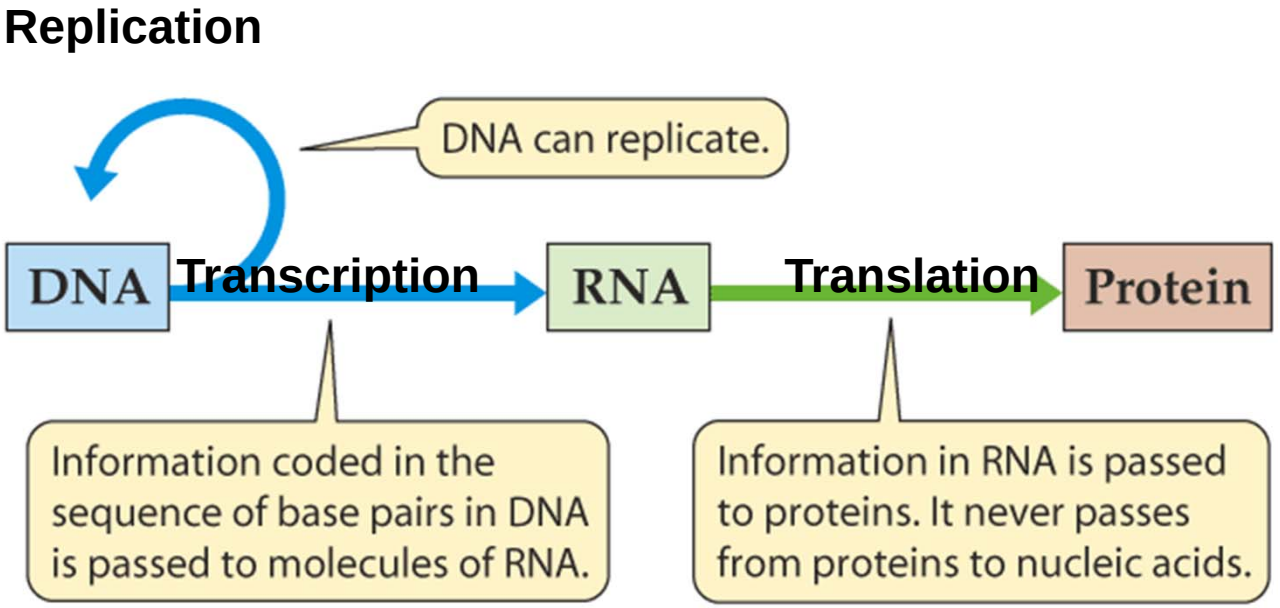
Introduction to Molecular Biology

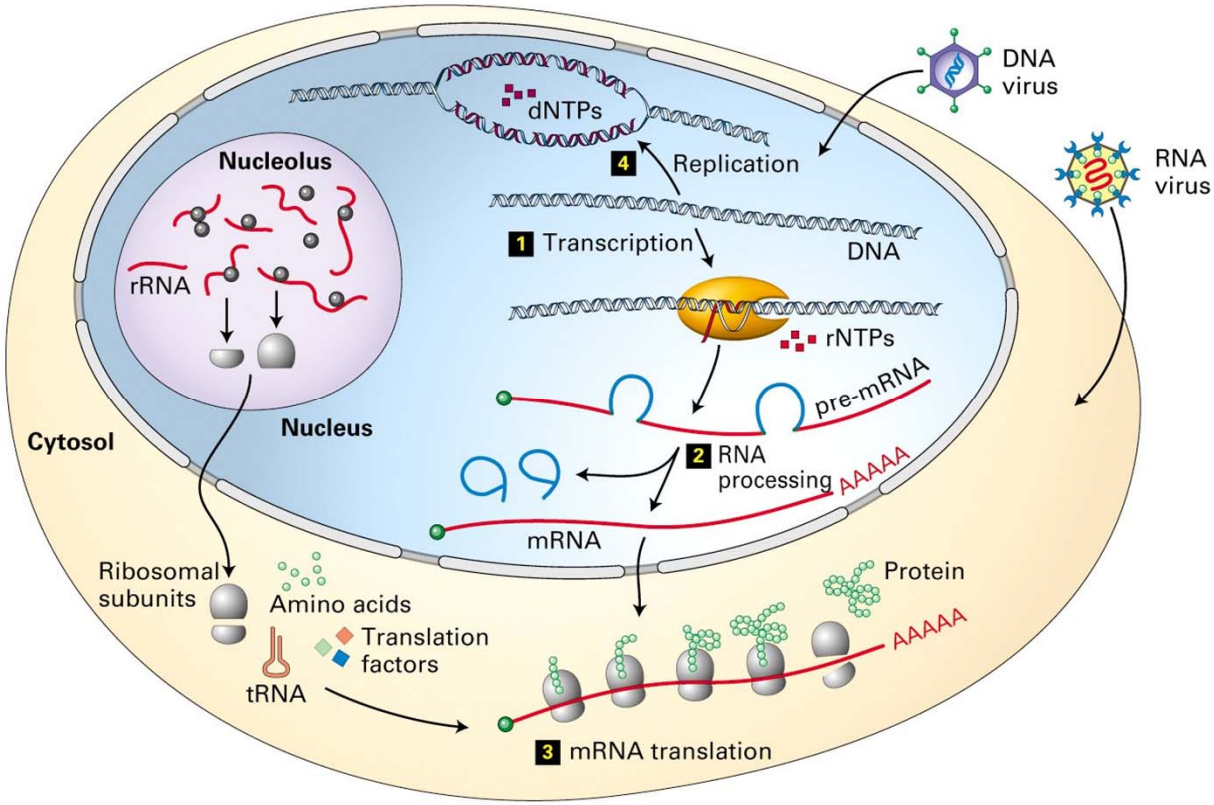
Part 2

DNA & RNA: Flow of Information



aka "The Central Dogma"!!

DNA to RNA to Protein

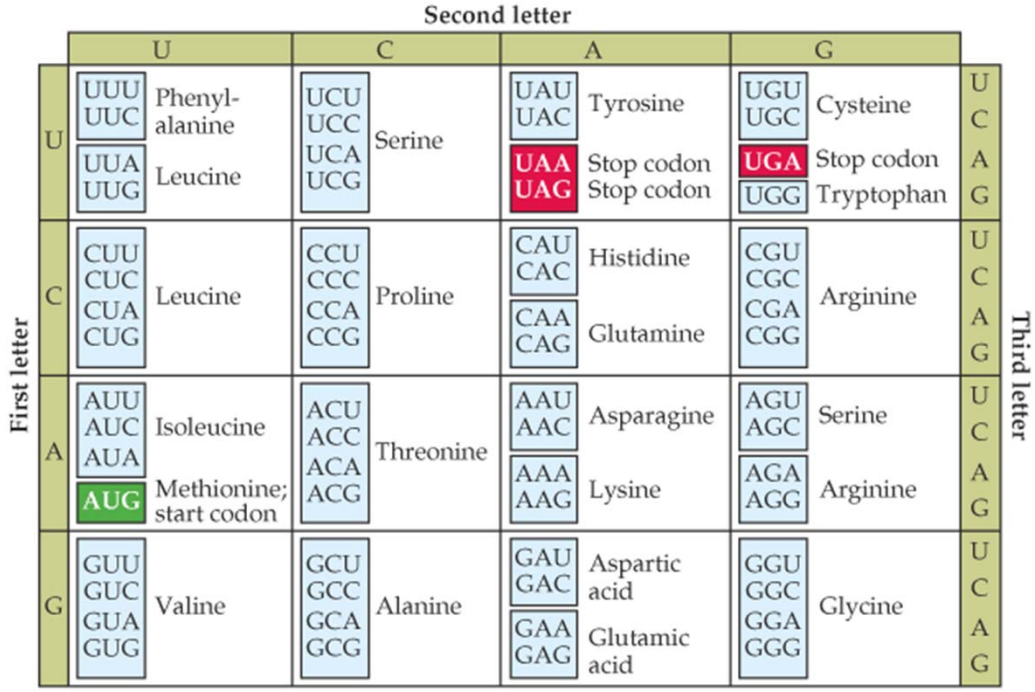


A gene is expressed in two steps

1. **Transcription:** RNA Synthesis
2. **Translation:** Protein Synthesis

The Code Book

- DNA, RNA, and Proteins are examples of strings written in either the four-letter nucleotide of DNA and RNA (A C G T/U)
- or the twenty-letter amino acid of proteins. Each amino acid is coded by 3 nucleotides called codons



First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC: Phenyl-alanine	UCU, UCC, UCA, UCG: Serine	UAU, UAC: Tyrosine	UGU, UGC: Cysteine	U C A G
	UUA, UUG: Leucine		UAA, UAG: Stop codon Stop codon	UGA: Stop codon; UGG: Tryptophan	
C	CUU, CUC, CUA, CUG: Leucine	CCU, CCC, CCA, CCG: Proline	CAU, CAC: Histidine	CGU, CGC, CGA, CGG: Arginine	U C A G
			CAA, CAG: Glutamine		
A	AUU, AUC, AUA: Isoleucine	ACU, ACC, ACA, ACG: Threonine	AAU, AAC: Asparagine	AGU, AGC: Serine	U C A G
	AUG: Methionine; start codon		AAA, AAG: Lysine	AGA, AGG: Arginine	
G	GUU, GUC, GUA, GUG: Valine	GCU, GCC, GCA, GCG: Alanine	GAU, GAC: Aspartic acid	GGU, GGC, GGA, GGG: Glycine	U C A G
			GAA, GAG: Glutamic acid		

DNA & RNA

- DNA = Deoxyribonucleic acid
- RNA = Ribonucleic acid
- They are almost the same…
- There is no T base in RNA
- A similar base U takes its place
- An oxygen atom is added to the sugar component of RNA

How Are Proteins Made?

```
DNA: TAC CGC GGC TAT TAC TGC CAG GAA GGA ACT
     ATG GCG CCG ATA ATG ACG GTC CTT CCT TGA
```

How Are Proteins Made?

DNA: TAC CGC GGC TAT TAC TGC CAG GAA GGA ACT

How Are Proteins Made?

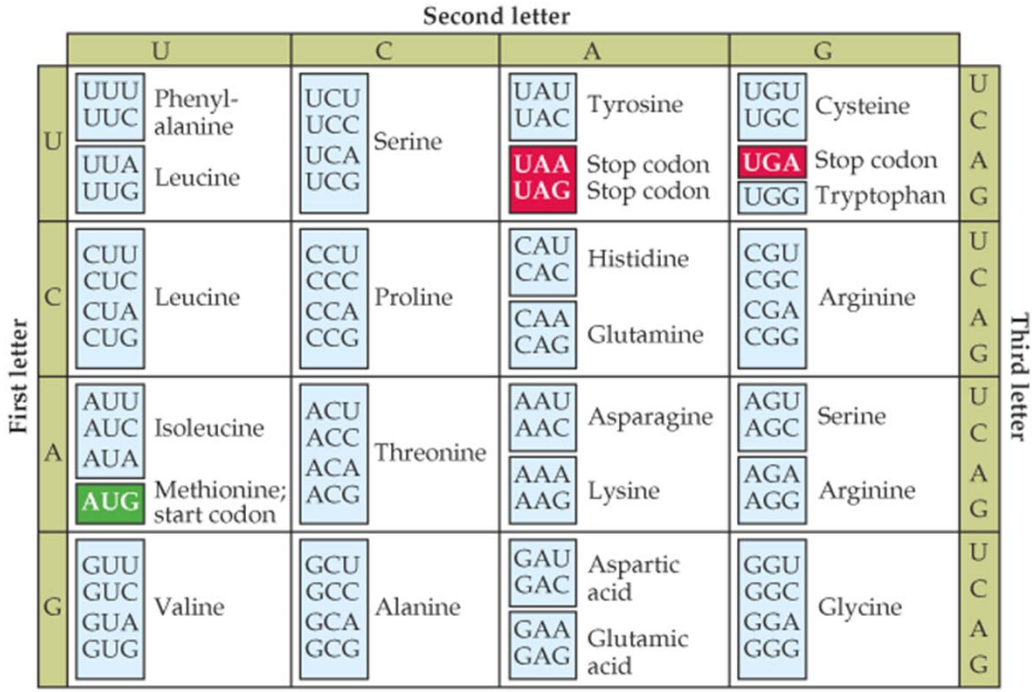
DNA: TAC CGC GGC TAT TAC TGC CAG GAA GGA ACT
RNA: AUG GCG CCG AUA AUG ACG GUC CUU CCU UGA

How Are Proteins Made?

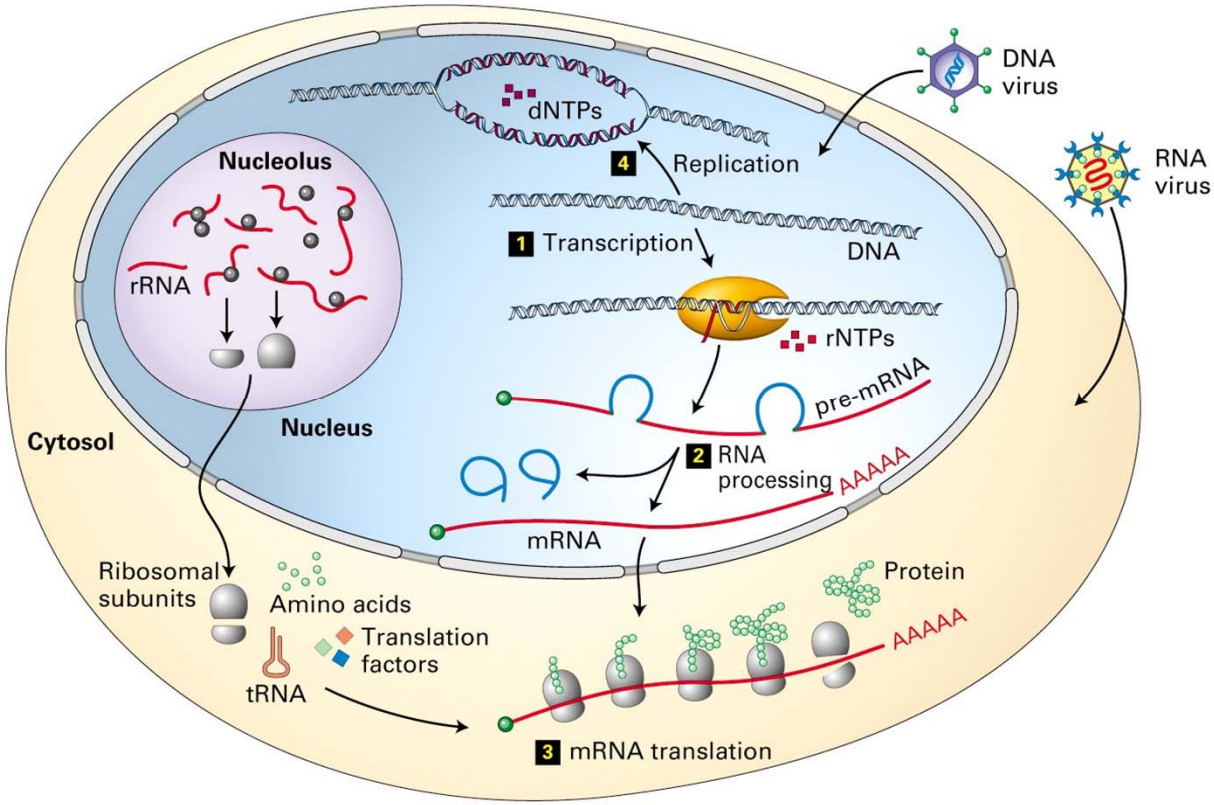
```
    DNA: TAC CGC GGC TAT TAC TGC CAG GAA GGA ACT
    RNA: AUG GCG CCG AUA AUG ACG GUC CUU CCU UGA
Protein: Met Ala Pro Ile Met Thr Val Leu Pro Stop
```

How Are Proteins Made?

DNA: TAC CGC GGC TAT TAC TGC CAG GAA GGA ACT
RNA: AUG GCG CCG AUA AUG ACG GUC CUU CCU UGA
Protein: Met Ala Pro Ile Met Thr Val Leu Pro Stop



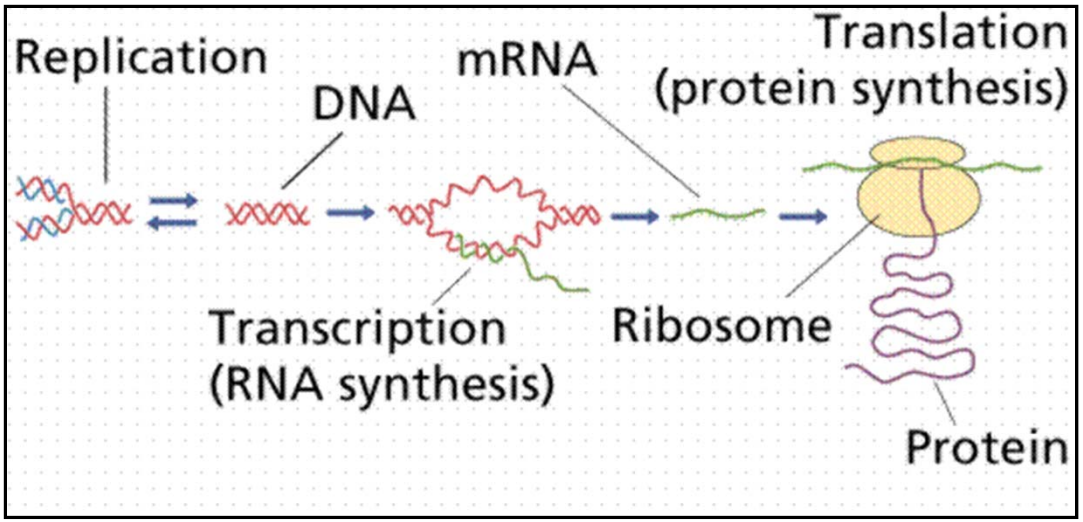
DNA to RNA to Protein

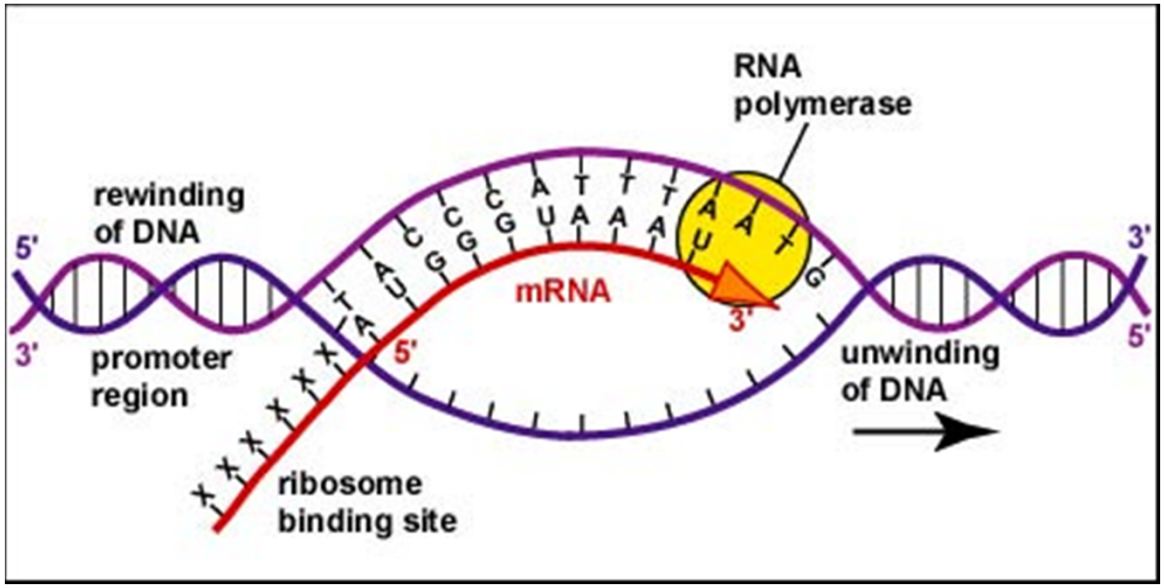


A gene is expressed in two steps

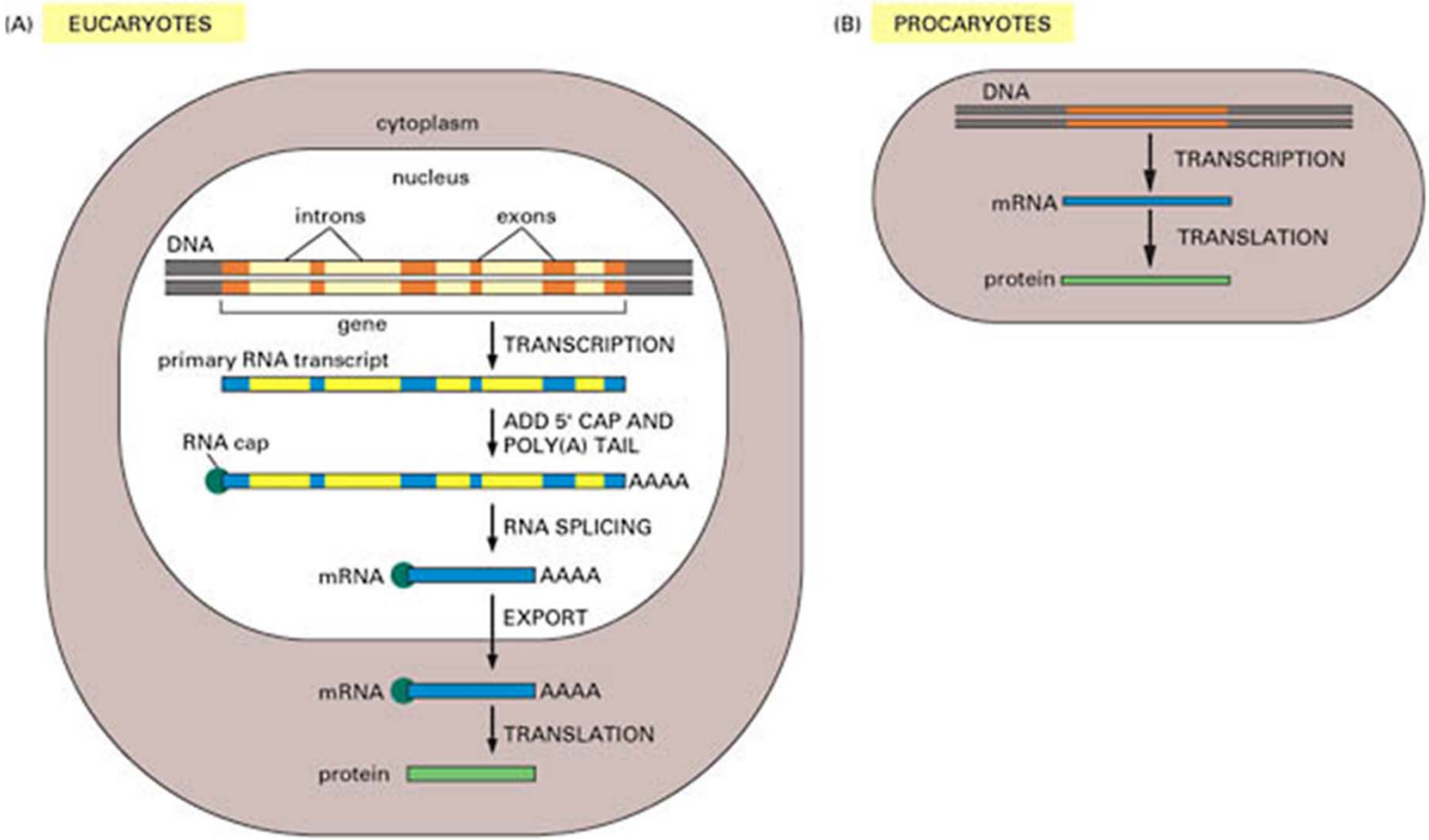
1. **Transcription:** RNA Synthesis
2. **Translation:** Protein Synthesis

Transcription

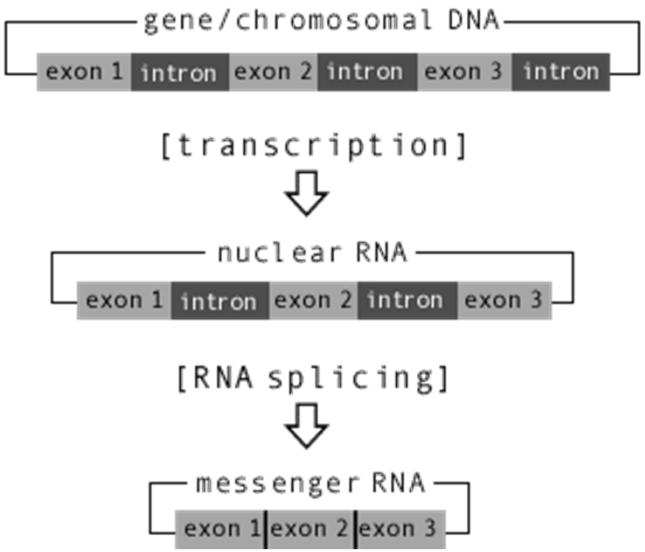




DNA to RNA to Protein



Splicing


gene/chromosomal DNA
exon 1
intron
exon 2
intron
exon 3
intron
[transcription]
nuclear RNA
exon 1
intron
exon 2
intron
exon 3
[RNA splicing]
messenger RNA
exon 1
exon 2
exon 3

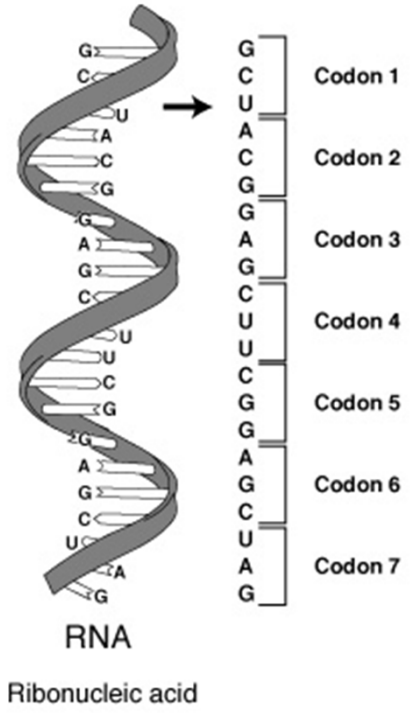
Terminology

- **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.

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

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



Proteins

- 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 are often enzymes that catalyze reactions.
- Also called “poly-peptides”

Polypeptide *vs* 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 is 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.

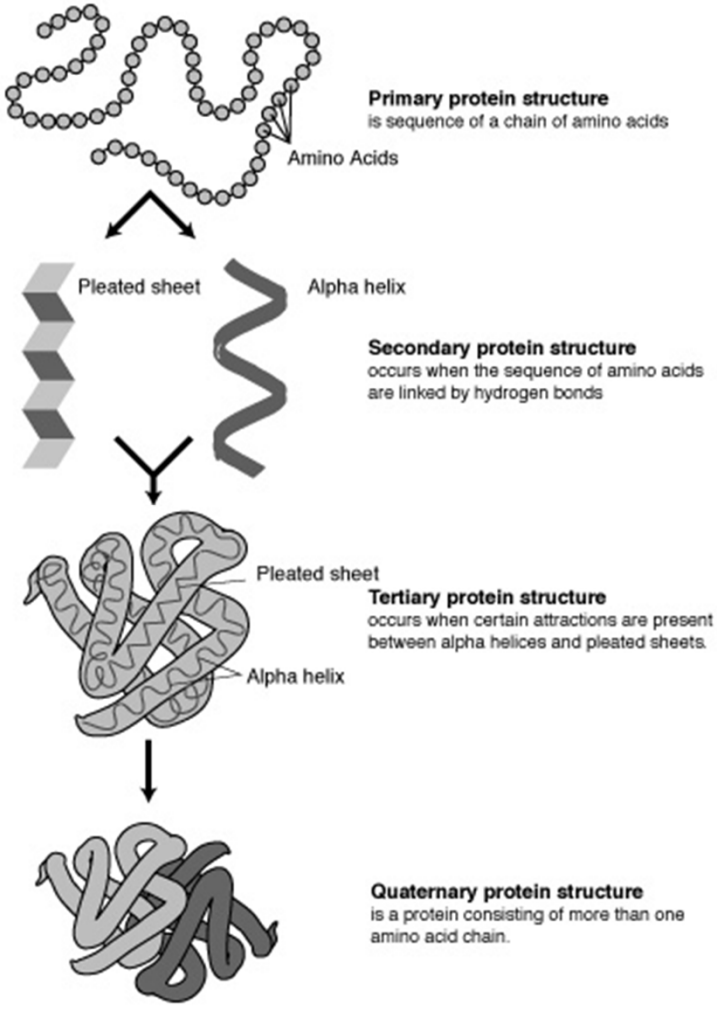
Protein Folding

- 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 fold and adopt 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

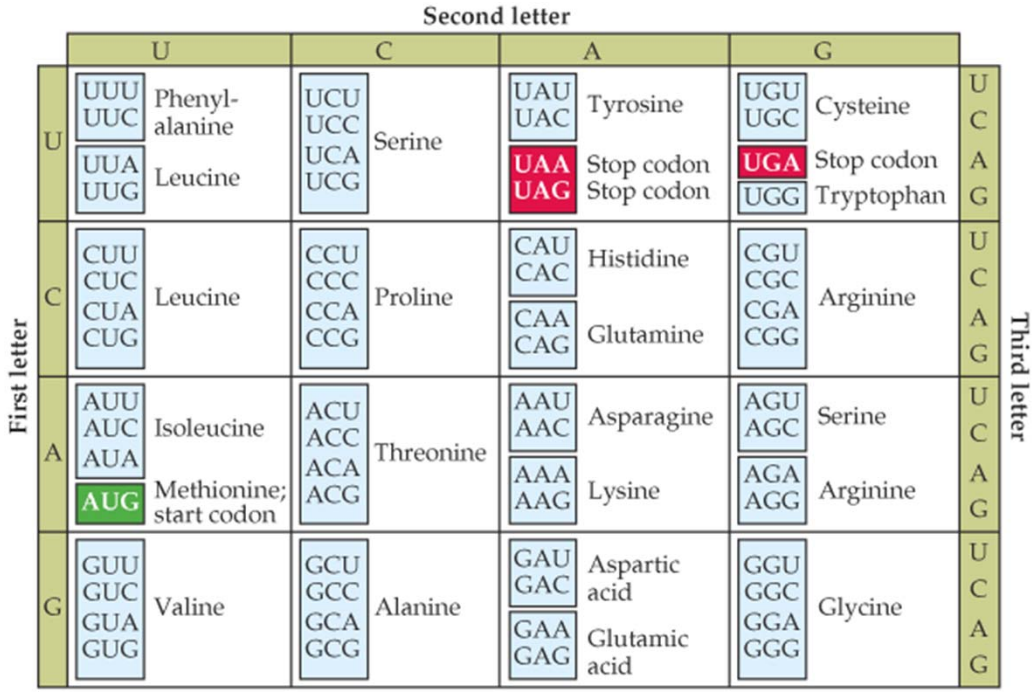
Protein Folding

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

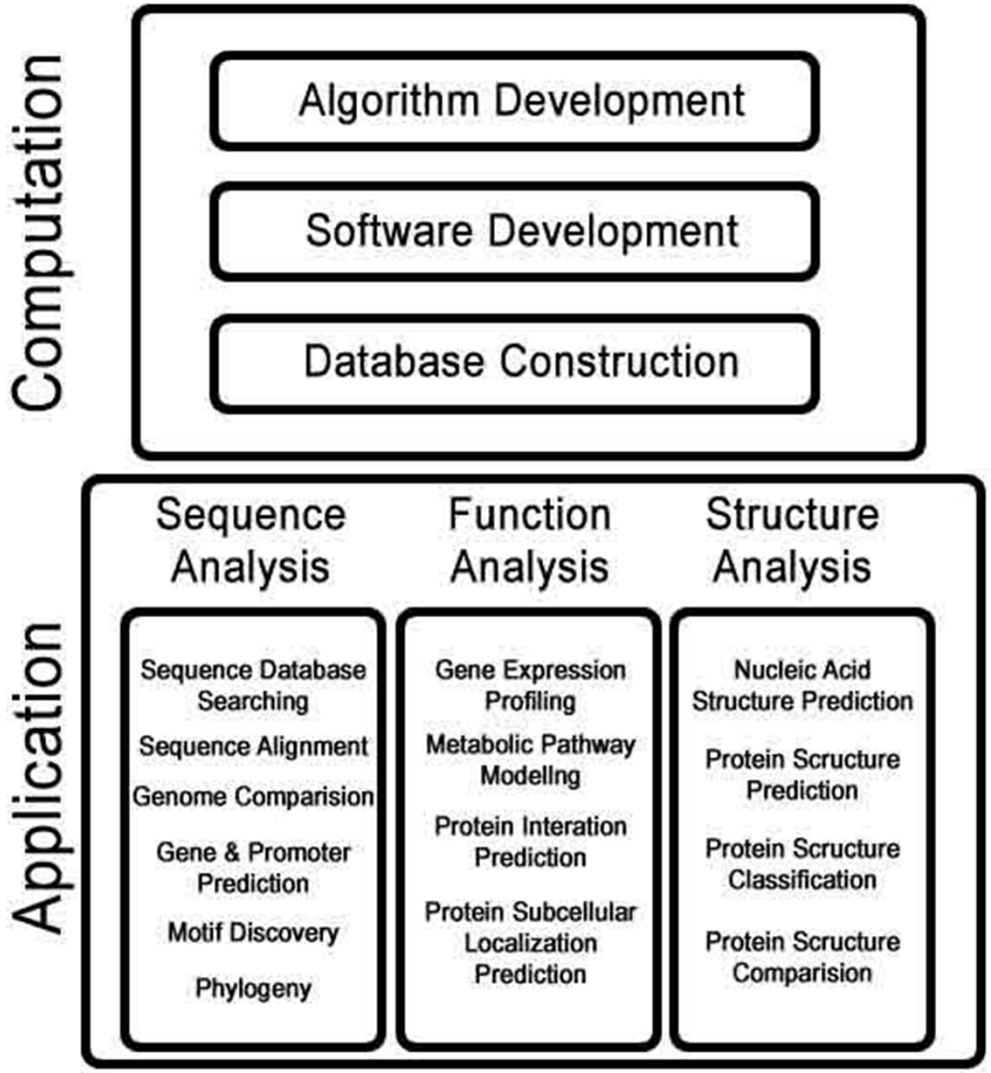


How Are Proteins Made?

```
    DNA: TAC CGC GGC TAT TAC TGC CAG GAA GGA ACT
    RNA: AUG GCG CCG AUA AUG ACG GUC CUU CCU UGA
Protein: Met Ala Pro Ile Met Thr Val Leu Pro Stop
```



Bioinformatics



Sequence Analysis

- **Sequence Databases** (e.g. GenBank)
  Primary (raw sequence data), secondary (biological knowledge)

- **Sequence Alignment** (global, local, multiple)
  Needed for structural, functional, and evolutionary inferences. Motifs, domains…

- **Gene & Promoter Prediction**
  open reading frames, exons, introns, …

- **Molecular Phylogenetics**
  Evolutionary history of living organisms, phylogenic tree construction, …

Structural Bioinformatics

- **Protein Structure**

  Protein functions are determined by their structure
  Databases, Visualization, Classification

- **Protein Structure Prediction**

- **Protein Structure Comparison**

- **RNA Structure Prediction**

Genomics & Proteomics

- **Structural Genomics**

  Genome mapping, sequence, assembly, annotation, comparison

- **Functional Genomics**

  Gene expression

- **Proteomics**

  Entire set of expressed proteins in a cell

Computation

- **Exhaustive Search**
  regulatory motifs in DNA, profiles
- **Greedy Algorithms**
  genome rearrangements, motif search
- **Dynamic Programming Algorithms**
  DNA sequence comparison, alignment, gene prediction
- **Divide-and-Conquer Algorithms**
  sequence alignment
- **Graph Algorithms**
  DNA sequencing, fragment assembly, peptide sequencing
- **Combinatorial Pattern Matching**
  similarity search, database searches
- **Clustering and Trees**
  gene expression analysis, tree construction
- **Hidden Markov Models**
  profile alignment
- **Randomized Algorithms**
- Machine Learning

Molecular Biology: Challenges

How does the structure and function at the molecular level account for the hierarchy?

- Molecular
- Intracellular
- Intercellular
- Tissue
- Organism
- Communities


DNA WAY
FORBES BLVD

References

- Neil C. Jones and Pavel A. Pevzner, *An Introduction to Bioinformatics Algorithms*, MIT Press 2004.
- Adapted from slides posted at the web site of the above book.
- Francis Crick, *Central Dogma of Molecular Biology*, Nature, Volume 227, August 1970.
- Luciano Floridi, *Information: A Very Short Introduction*, Oxford 2010.