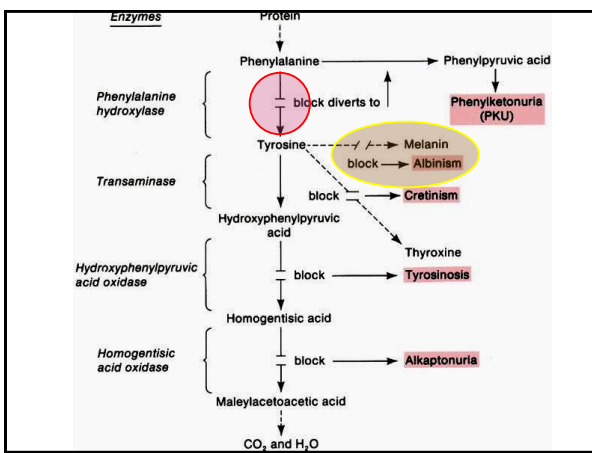


Metabolism Teaches Us About Genes

- Metabolic defects
 - studying metabolic diseases suggested that genes specified proteins
 - alkaptonuria (black urine from alkapton a.k.a. homogentisic acid)
 - PKU (phenylketonuria)
 - each disease is caused by non-functional enzyme

A → X → B → X → C → X → D → X → E



1 Gene – 1 Enzyme Hypothesis

- Beadle & Tatum
 - Compared mutants of bread mold, *Neurospora* fungus
 - created mutations by X-ray treatments
 - X-rays break DNA
 - inactivate a gene
 - wild type grows on “minimal” media
 - sugars + required precursor nutrient to synthesize essential amino acids
 - mutants require added amino acids
 - each type of mutant lacks a certain enzyme needed to produce a certain amino acid
 - non-functional enzyme = broken gene

Beadle & Tatum

1941 | 1958

George Beadle
Edward Tatum

Beadle & Tatum's *Neurospora* Experiment

| | Minimal medium (MM) | MM + Ornithine | MM + Citrulline | MM + Arginine |
|--|---------------------|----------------|-----------------|---------------|
| Wild type | ✓ | ✓ | ✓ | ✓ |
| Class I Mutants (mutation in gene A) | ✗ | ✓ | ✓ | ✓ |
| Class II Mutants (mutation in gene B) | ✓ | ✗ | ✓ | ✓ |
| Class III Mutants (mutation in gene C) | ✓ | ✓ | ✗ | ✓ |

(a) Experiment

| | Gene A | Gene B | Gene C |
|-------------------|--|--------|--------|
| Wild type | Precursor → Enzyme A → Ornithine → Enzyme B → Citrulline → Enzyme C → Arginine | | |
| Class I Mutants | Precursor → Enzyme A → Ornithine → Enzyme B → Citrulline → Enzyme C → Arginine | | |
| Class II Mutants | Precursor → Enzyme A → Ornithine → Enzyme B → Citrulline → Enzyme C → Arginine | | |
| Class III Mutants | Precursor → Enzyme A → Ornithine → Enzyme B → Citrulline → Enzyme C → Arginine | | |

(b) Interpretation

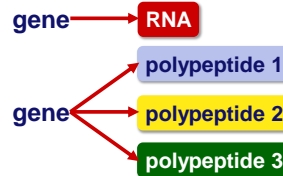
So... What is a Gene?

- One gene – one enzyme
 - ◆ all proteins are coded by genes
 - ◆ but not all proteins are enzymes
- One gene – one protein
 - ◆ each polypeptide has its own gene
 - ◆ but many proteins are composed of several polypeptides
- One gene – one polypeptide
 - ◆ but many genes only code for RNA
- One gene – one product
 - ◆ but many genes code for more than one product ...

Defining a Gene...

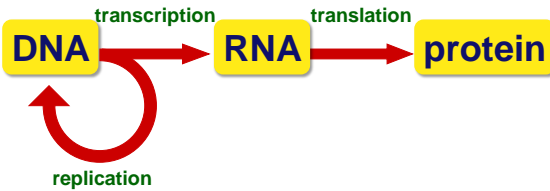
“Defining a gene is problematic because... one gene can code for several protein products, some genes code only for RNA, two genes can overlap, and there are many other complications.”

– Elizabeth Pennisi, Science 2003



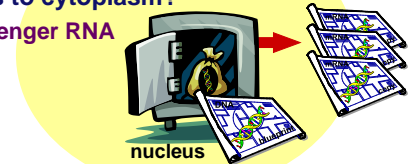
The “Central Dogma”

- How do we move information from DNA to proteins?



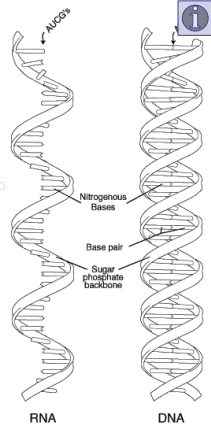
From nucleus to cytoplasm...

- Where are the genes?
 - ◆ genes are on chromosomes in nucleus
- Where are proteins synthesized?
 - ◆ proteins made in cytoplasm by **ribosomes**
- How does the information get from nucleus to cytoplasm?
 - ◆ messenger RNA



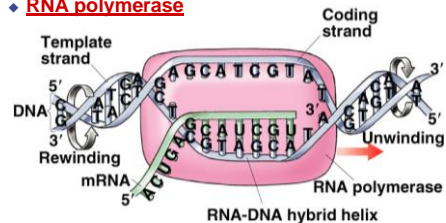
RNA

- ribose sugar
- N-bases
 - ◆ **uracil** instead of thymine
 - ◆ U : A
 - ◆ C : G
- single stranded
- mRNA, rRNA, tRNA, siRNA....



Transcription

- Transcribed DNA strand = **template strand**
 - ◆ untranscribed DNA strand = **coding strand**
- Synthesis of complementary RNA strand
 - ◆ transcription bubble
- Enzyme that facilitates the building of RNA:
 - ◆ **RNA polymerase**



Transcription in Prokaryotes

- **Initiation**
 - ◆ RNA polymerase binds to **promoter sequence** on DNA

Role of promoter

1. Where to start reading = starting point
2. Which strand to read = template strand
3. Direction on DNA = always reads DNA 3'→5'

Transcription in Prokaryotes

- **Promoter sequences upstream of gene**

RNA polymerase molecules bound to bacterial DNA

Transcription in Prokaryotes

- **Elongation**
 - ◆ RNA polymerase unwinds DNA ~20 base pairs at a time
 - ◆ reads DNA 3'→5'
 - ◆ builds RNA 5'→3' (the energy governs the synthesis!)

No proofreading

- 1 error/10⁵ bases
- many copies
- short life
- not worth it!

Transcription

Transcription in Prokaryotes

- **Termination**
 - ◆ RNA polymerase stops at **termination sequence**
 - ◆ mRNA leaves nucleus through pores

Completed RNA transcript

RNA GC hairpin turn

Prokaryote vs. Eukaryote Genetics

| | |
|--|---|
| <ul style="list-style-type: none"> ■ Prokaryotes <ul style="list-style-type: none"> ◆ DNA in cytoplasm ◆ circular chromosome ◆ naked DNA ◆ no introns | <ul style="list-style-type: none"> ■ Eukaryotes <ul style="list-style-type: none"> ◆ DNA in nucleus ◆ linear chromosomes ◆ DNA wound on histone proteins ◆ introns vs. exons |
|--|---|

intron = noncoding (inbetween) sequence

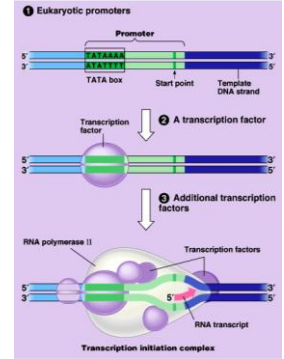
exon = coding (expressed) sequence

Transcription in Eukaryotes

- 3 RNA polymerase enzymes
 - ◆ RNA polymerase I
 - only transcribes rRNA genes
 - ◆ **RNA polymerase II**
 - transcribes genes into mRNA
 - ◆ RNA polymerase III
 - only transcribes rRNA genes
- ◆ each has a specific promoter sequence it recognizes

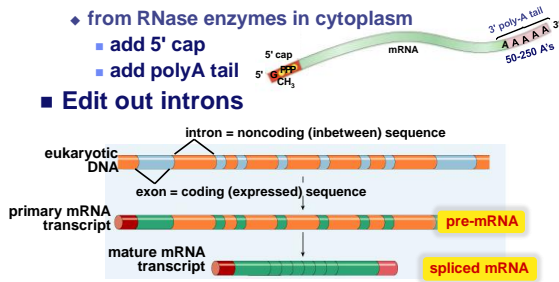
Transcription in Eukaryotes

- **Initiation complex**
 - ◆ **transcription factors** bind to **promoter region** upstream of gene
 - proteins which bind to DNA & turn on or off transcription
 - **TATA** box binding site
 - ◆ only then does RNA polymerase bind to DNA



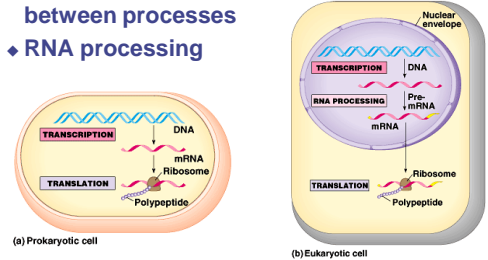
Eukaryotic Post-transcriptional Processing

- **Primary transcript**
 - ◆ eukaryotic mRNA needs work after transcription
- **Protect mRNA**
 - ◆ from RNase enzymes in cytoplasm
 - add 5' cap
 - add polyA tail
- **Edit out introns**

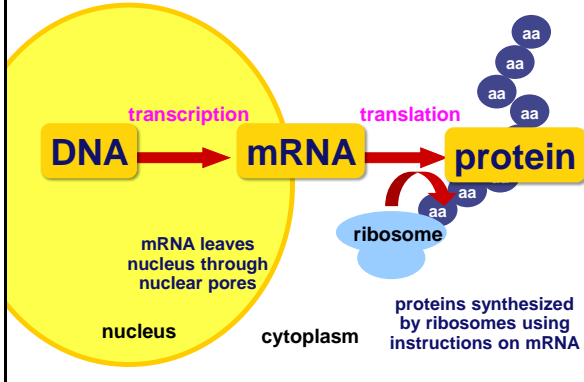


Prokaryote vs. Eukaryote Genetics

- **Differences between prokaryotes & eukaryotes**
 - ◆ time & physical separation between processes
 - ◆ RNA processing

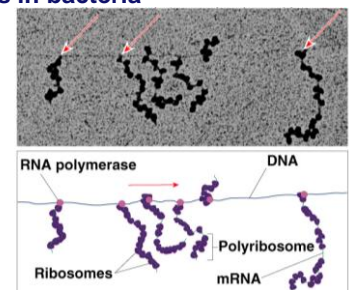


From Gene to Protein



Translation in Prokaryotes

- **Transcription & translation are simultaneous in bacteria**
 - ◆ DNA is in cytoplasm
 - ◆ no mRNA editing needed



How Does DNA Code for Proteins

DNA TACGCACATTACGTACGCGG

↓

mRNA AUGCGUGUAAUGCAUGCGCC

↓

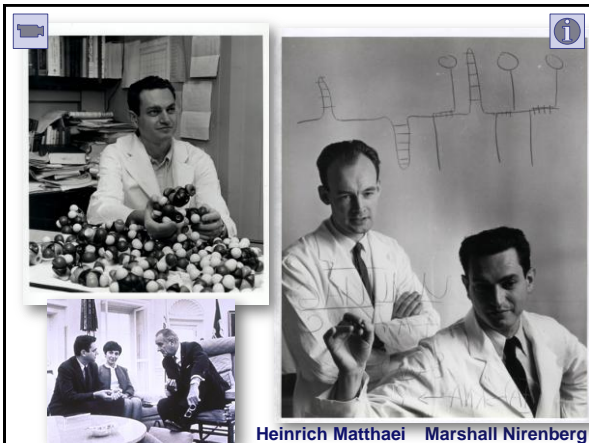
protein Met Arg Val Asn Ala Cys Ala

Cracking the Code

1960 | 1968

- Nirenberg & Matthaei
 - ◆ determined 1st codon–amino acid match
 - UUU coded for phenylalanine
 - ◆ created artificial poly(U) mRNA
 - ◆ added mRNA to test tube of ribosomes, tRNA & amino acids
 - mRNA synthesized single amino acid polypeptide chain

phe-phe-phe-phe-phe



Translation

- Codons
 - ◆ blocks of 3 nucleotides decoded into the sequence of amino acids

mRNA Codes for Proteins in Triplets

DNA TACGCACATTACGTACGCGG

↓

mRNA AUGCGUGUAAUGCAU ribosome

↓

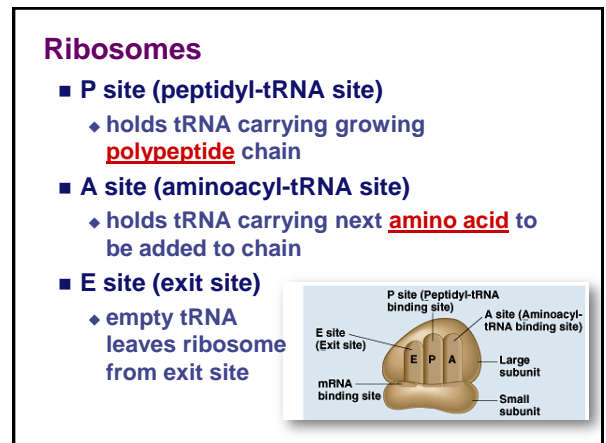
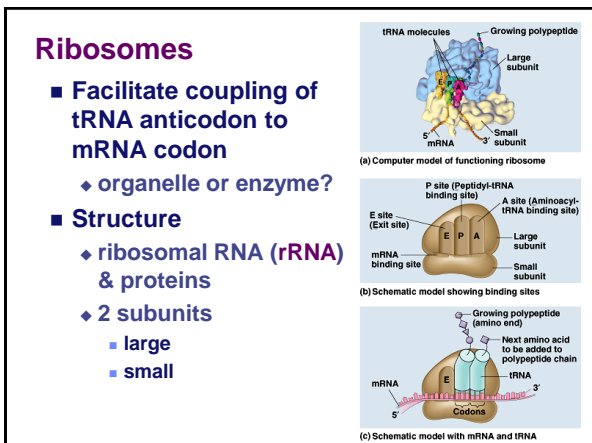
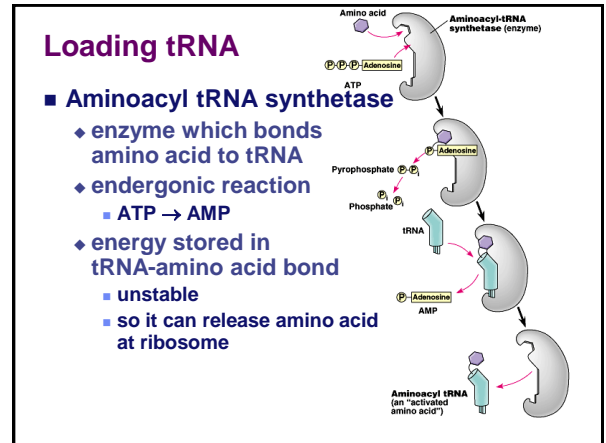
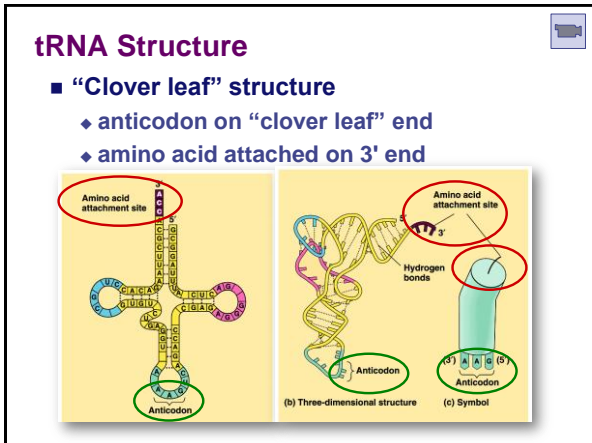
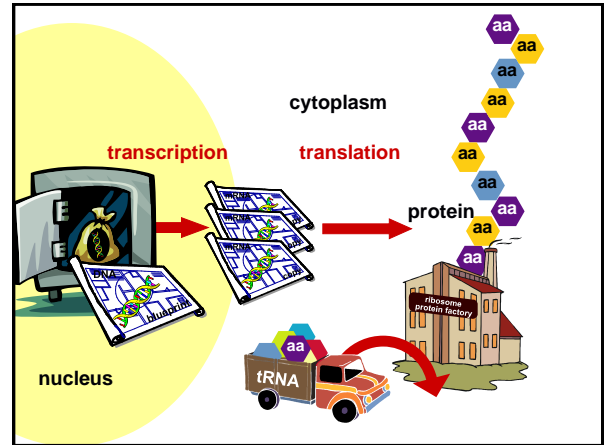
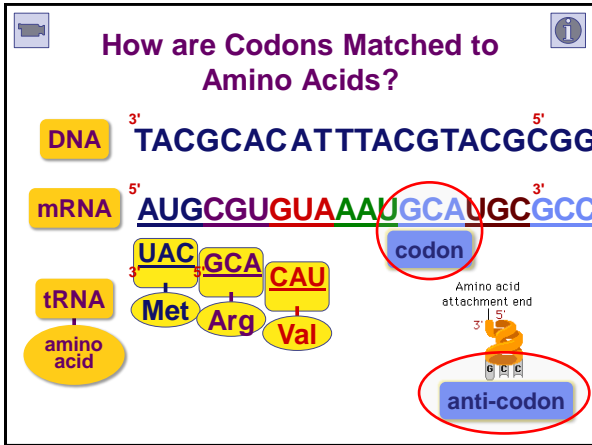
protein Met Arg Val Asn Ala Cys Ala

mRNA codes for proteins in triplets...
CODONS!

The Code!

- For ALL life!
 - ◆ strongest support for a common origin for all life
- Code is redundant
 - ◆ several codons for each amino acid
- Start codon
 - ◆ AUG
 - ◆ methionine
- Stop codons
 - ◆ UGA, UAA, UAG

| | | Second base | | | | |
|---------------------|------------------|-------------|---------|----------|----------|---|
| | | U | C | A | G | |
| First base (5' end) | U | UUU Phe | UCU | UAU Tyr | UGU Cys | U |
| | U | UUC | UCC Ser | UAC Tyr | UGC Cys | C |
| | U | UUA Leu | UCA | UAA Stop | UGA Stop | A |
| | U | UUG | UCG | UAG Stop | UGG Trp | G |
| C | CUU | CCU | CAU His | CGU | | U |
| | CUC | CCC | CAC His | CGC | | C |
| | CUA | CCA | CAA Gln | CGA | | A |
| | CUG | CCG | CAG Gln | CGG | | G |
| A | AUU | ACU | AAU Asn | AGU Ser | | U |
| | AUC | ACC | AAC Asn | AGC Ser | | C |
| | AUA | ACA | AAA Lys | AGA Arg | | A |
| | AUG Met or start | ACG | AAG Lys | AGG Arg | | G |
| G | GUU | GCU | GAU Asp | GGU | | U |
| | GUC | GCC | GAC Asp | GGC Gly | | C |
| | GUA | GCA | GAA Glu | GGG Gly | | A |
| | GUG | GCG | GAG Glu | GGG Gly | | G |



Building a Polypeptide

- **Initiation**
 - ◆ brings together mRNA, ribosome subunits, proteins & initiator tRNA
- **Elongation**
- **Termination**

Elongation: Growing a Polypeptide

Termination: Release Polypeptide

- **Release factor**
 - ◆ “release protein” bonds to A site
 - ◆ bonds water molecule to polypeptide chain

Protein Targeting

- **Signal peptide**
 - ◆ address label

Destinations:

- secretion
- nucleus
- mitochondria
- chloroplasts
- cell membrane
- cytoplasm

ex. start of a secretory pathway

Can you tell the eukaryotic story?

Putting it all together...

The loss of a single nucleotide from a 1,000-nucleotide gene can completely destroy the gene's function.

Mutations

DNA → Transcription → mRNA → Translation → Protein

10 μm

Val | His | Leu | Thr | Pro | Val | Glu | ...
1 2 3 4 5 6 7

(b) Sickled red blood cells and the primary structure of sickle-cell hemoglobin

Universal Code

- Code is redundant
 - several codons for each amino acid
 - “wobble” in the tRNA
 - “wobble” in the aminoacyl-tRNA synthetase enzyme that loads the tRNA

Activating enzyme
tRNA^{Trp}
Anticodon

| | Second base | | | |
|---|------------------|-----|----------|----------|
| | U | C | A | |
| U | UUU Phe | UCU | UAU Tyr | UGU Cys |
| | UUC | UCC | UAC | UGC |
| | UUA Leu | UCA | UAA Stop | UGA Stop |
| C | CUU | CCU | CAU His | CGU |
| | CUC | CCC | CAC | CGC |
| | CUA Leu | CCA | CAA | CGA |
| A | AUU | ACU | AAU Asn | AGU Ser |
| | AUC | ACC | AAC | AGC |
| | AUA | ACA | AAA | AGA |
| G | AUG Met or start | ACG | AAG | AGG |
| | GUU | GCU | GAU Asp | GGU |
| | GUC | GCC | GAC | GGC |
| | GUA Val | GCA | GAA | GGA |
| | GUG | GCG | GAG | GGG |

Mutations

- Point mutations
 - single base change
 - base-pair substitution
 - silent mutation
 - no amino acid change
 - redundancy in code
 - missense
 - change amino acid
 - nonsense
 - change to stop codon

A Mutation Leads to Sickle Cell Anemia

Wild-type hemoglobin DNA: 3' C T T 5'

Mutant hemoglobin DNA: 3' C A T 5'

mRNA: 5' G A A 3' (Wild-type) vs 5' G U A 3' (Mutant)

Normal hemoglobin: Glu

Sickle-cell hemoglobin: Val

Sickle Cell Anemia

10 μm

Val | His | Leu | Thr | Pro | Glu | Glu | ...
1 2 3 4 5 6 7

(a) Normal red blood cells and the primary structure of normal hemoglobin

Val | His | Leu | Thr | Pro | Val | Glu | ...
1 2 3 4 5 6 7

(b) Sickled red blood cells and the primary structure of sickle-cell hemoglobin

Mutations

- Frameshift
 - shift in the reading frame
 - changes everything “downstream”
 - insertions
 - adding base(s)
 - deletions
 - losing base(s)

Wild type mRNA: 5' A U G A A G U U U G G C U U A A 3'

Protein: Met | Lys | Phe | Gly | Stop

Base-pair insertion or deletion

Frameshift causing extensive missense: Extra U → Met | Lys | Leu | Ala | ...

Frameshift causing immediate nonsense: Extra U → Met | Stop

Insertion or deletion of 3 nucleotides: no frameshift; extra or missing amino acid → Met | Phe | Gly | Stop