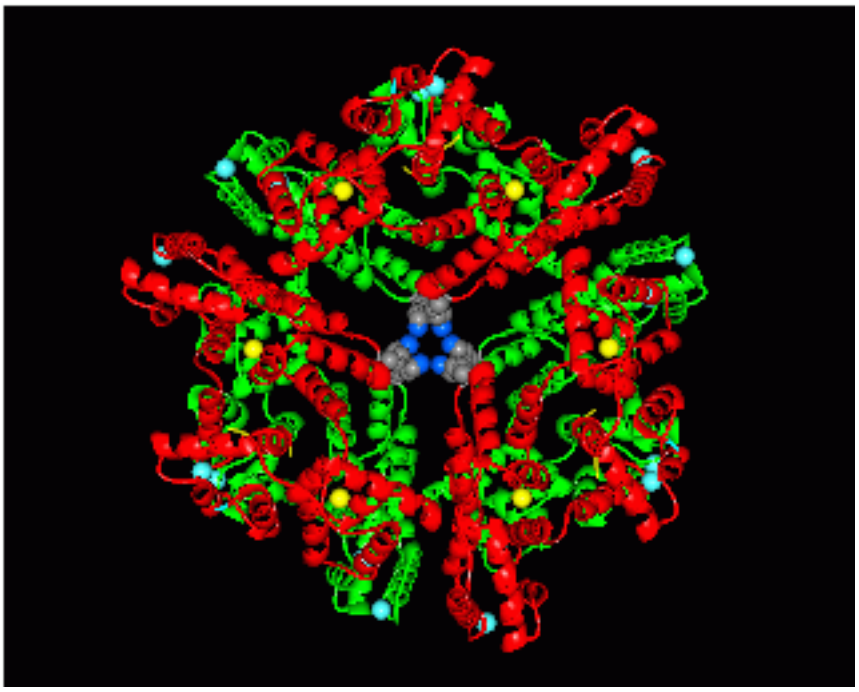
Lecture 1

Central Dogma and Deciphering the Genetic Code

Hartmut "Hudel" Luecke

Department of Molecular Biology and Biochemistry

Area of reseach: Structure & Function of Proteins



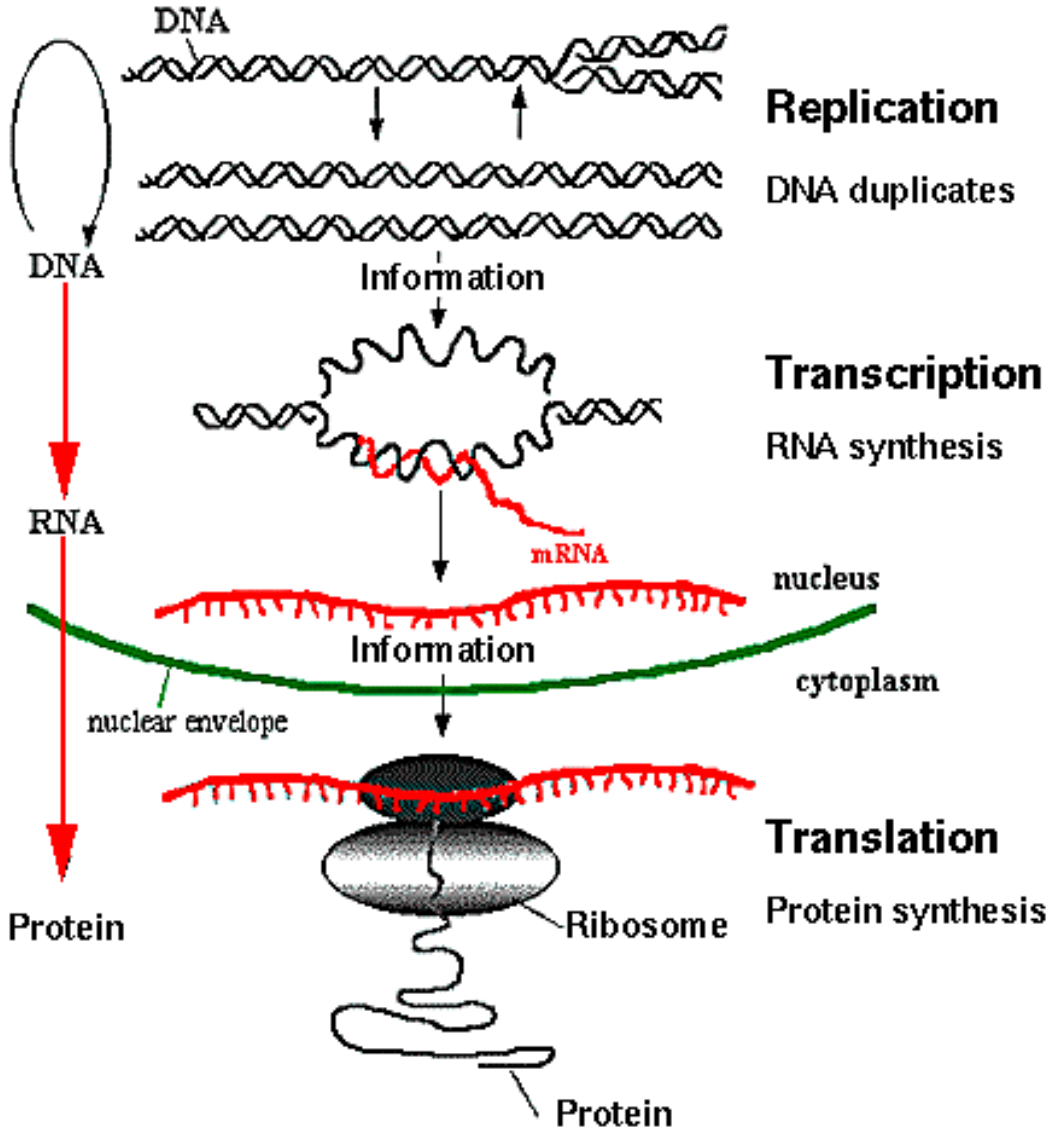
Email: HUDEL@UCI.EDU

http://bass.bio.uci.edu/~hudel/bs99a

http://bass.bio.uci.edu/~hudel/bs99b

Next: The Central Dogma

The Central Dogma



DNA	==>	RNA	==>	Protein
4 nucleotides (A,C,G, T)		4 nucleotides (A,C,G, U)		20 amino acids

TGGCGAACTGATGTG	***transcription*** by polymerase	**UGGCGAACUGAUGUG**	***translation*** by ribosome	**TrpArgThrAspVal**
phosphodiester bond		phosphodiester bond		peptide bond

Comparing proteins with nucleic acids:

Properties proteins have ***in common*** with nucleic acid:

- Linear heteropolymers with a defined sequence
- Individual building blocks (called amino acids or simply residues for proteins) are linked together through covalent (chemical) bonds

Properties ***different*** from nucleic acid:

- More diverse building blocks: 20 amino acids vs. 4 nucleic acids
- Large variety of functional groups: negatively charged, positively charged, hydrophobic, hydroxyl, sulfhydryl
- Vastly accelerate a multitude of chemical reactions (also: ribozymes)
- Assume a wealth of well-defined tertiary structures (shapes): helix bundles, beta sheets, alpha/beta barrels etc.

What do proteins do?

- Catalyze chemical reactions (enzymes): alcohol dehydrogenase
- Carry nutrients: hemoglobin is the oxygen carrier in your blood
- Signaling: peptide hormones bind to protein receptors, transcription factors
- Molecular recognition: antibodies bind to antigens
- Play structural roles: finger nails, hair, eye lens
- Function as motors & pumps: myosin-actin in your muscles, ion pumps

Proteins are the molecular workhorse of the cell

Proteins are of central importance in every cellular process

Proteins must be made by the cell with high fidelity

Examples of single amino-acid mutations that cause disease:

- hemoglobin: **Glutamic acid** (Glu) to **valine** (Val) mutation at position 6 of the beta chain causes ***sickle-cell anemia***.
- Fibroblast Growth Factor (FGF) receptor 3: **Glycine** (Gly) to **arginine** (Arg) mutation results in ***achondroplasia***, a form of dwarfism for which the gene was discovered here at UCI by Drs. Thompson and Wasmuth.
- Enzyme uroporphyrinogen III cosynthase: causes ***congenital porphyria*** with symptoms such as skin photosensitivity & scarring, mutilating skin deformity, hypertrichosis, hemolytic anemia, red stained teeth.

Next: Translation

Please report typos, errors etc. by EMAIL (mention the title of this page).

Translation

The first step in following the blueprint of DNA to make a protein is *transcription* which generates an mRNA copy from the DNA template.

Translation* is the 2nd step: It uses the mRNA template to make the protein polymer. This process is also called *protein synthesis.

The reasons for having two steps instead of one are:

- Amplification: a **single copy** gene on DNA can be transcribed into **many copies** of mRNA
- Increased levels of control: regulation of ***transcription*** as well as ***translation***
- Ability to separate the mechanisms for DNA ***replication & transcription*** from ***protein synthesis***
- In eukaryotes: Ability to spatially separate ***replication & transcription*** (nucleus) from ***protein synthesis*** (cytoplasm)

Translation **is the process of reading the copy of genetic information on the mRNA (linear sequence of 4 different nucleic acids) and translating it into the proper linear protein sequence of 20 different amino acids.**

This process is performed in one of the most complex organelles of the cell, the *ribosome*. In the ribosome the mRNA sequence (*information*) is read and the corresponding polypeptide (protein) is assembled. The rules for translating the linear nucleic acid sequence (mRNA) into the linear amino acid sequence (protein) are called the *Genetic Code.*

Next: The Genetic Code

Please report typos, errors etc. by EMAIL (mention the title of this page).

The Genetic Code

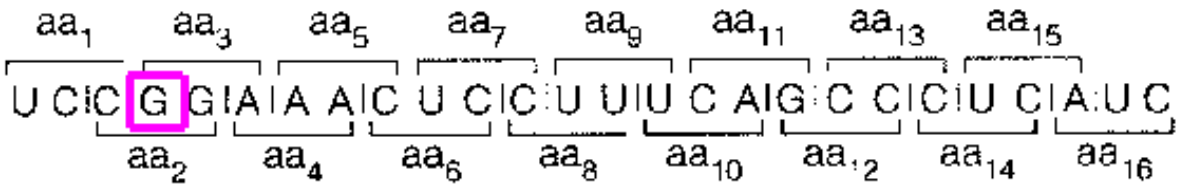
How to encode a 20-letter alphabet (protein) with a 4-letter alphabet (DNA)?

1 nucleotide	**4**	**A, C, G, U**	**4 amino acids**
2 nucleotides	**4 x 4 = 16**	**AU, AG, CA, UU etc.**	**16 amino acids**
3 nucleotides	**4 x 4 x 4 = 64**	**AUG, UGC, CGA etc.**	**64 amino acids**

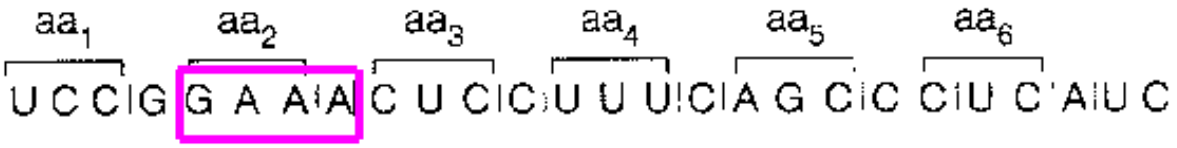
Since two nucleotides are not enough (16), three nucleotides are needed to code for all 20 amino acids

Thus Watson & Crick proposed that *codon triplets* code for individual amino acids.

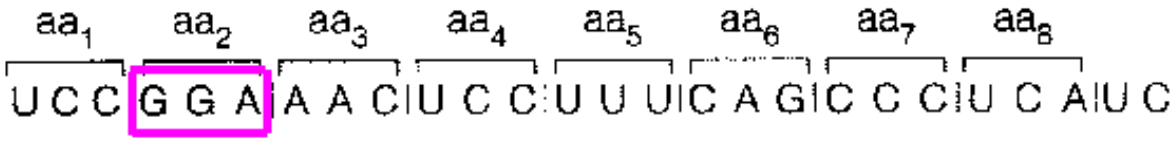
There are several possibilities how triplets might code for amino acids:



(a) Overlapping code. There will be statistical regularities between adjacent amino acid residues. Point mutations (red) will be able to change two amino acid residues.



(b) Punctuated code. Deletions of four nucleotides (or multiples thereof) will restore the reading frame.



(c) Unpunctuated code. Deletions of three nucleotides (or multiples thereof) will restore the reading frame. This is the actual form of the code.

To verify that the Code uses triplets and to determine:

- overlapping vs. non-overlapping code,
- punctuated vs. unpunctuated code,
- the redundancy of the code (64 triplets for 20 amino acids)

the following experiments were performed in the early 60s:

Next: Experimental Evidence for the Genetic Code

Please report typos, errors etc. by EMAIL (mention the title of this page).

Experimental Evidence

Crick & Brenner (1961) showed the effect of successive deletions of nucleotides in bacteriophage T4 DNA:

ATG CT*G* CTC TGT GCC GCC Met Leu Leu Cys Ala Ala	Original sequence
ATG CT*C* TCT GTG CCG CC. Met Leu Ser Val Pro Pro	1 nucleotide deleted
ATG CT*T* CTG TGC CGC C.. Met Pro Leu Cys Arg ...	2 nucleotides deleted
ATG CTC TGT GCC GCC ... Met Leu Cys Ala Ala ...	3 nucleotides deleted

Deletion of 1 or 2 nucleotides (*frame shift mutation*) results in non-functional protein

Deletion of 3 nucleotides results only in deletion of 1 amino acid (but protein could still be dysfunctional)

Insertion of 3 nucleotides results in insertion of 1 amino acid (not shown)

Change of 1 nucleotide results in either a *sense or silent mutantion* or in a *missense mutation*

Nucleotides are read as triplets without overlap or punctuation

Next: Deciphering the Genetic Code

Please report typos, errors etc. by EMAIL (mention the title of this page).

Deciphering the Genetic Code

Which triplet codon corresponds to which amino acid?

Nirenberg (1961) added synthetic homo-polynucleotides to bacterial lysate:

poly U	**UUUUUUUUUUUU**	**Phe-Phe-Phe-Phe**
poly A	**AAAAAAAAAAAA**	**Lys-Lys-Lys-Lys**
poly C	**CCCCCCCCCCCC**	**Pro-Pro-Pro-Pro**

Thus the first codon was determined: UUU codes for phenylalanine

Alternatively, AAA code for lysine, and CCC for proline.

Khorana (mid 60s) was able to synthesize triplet repeats such as $(AAG)_n$:

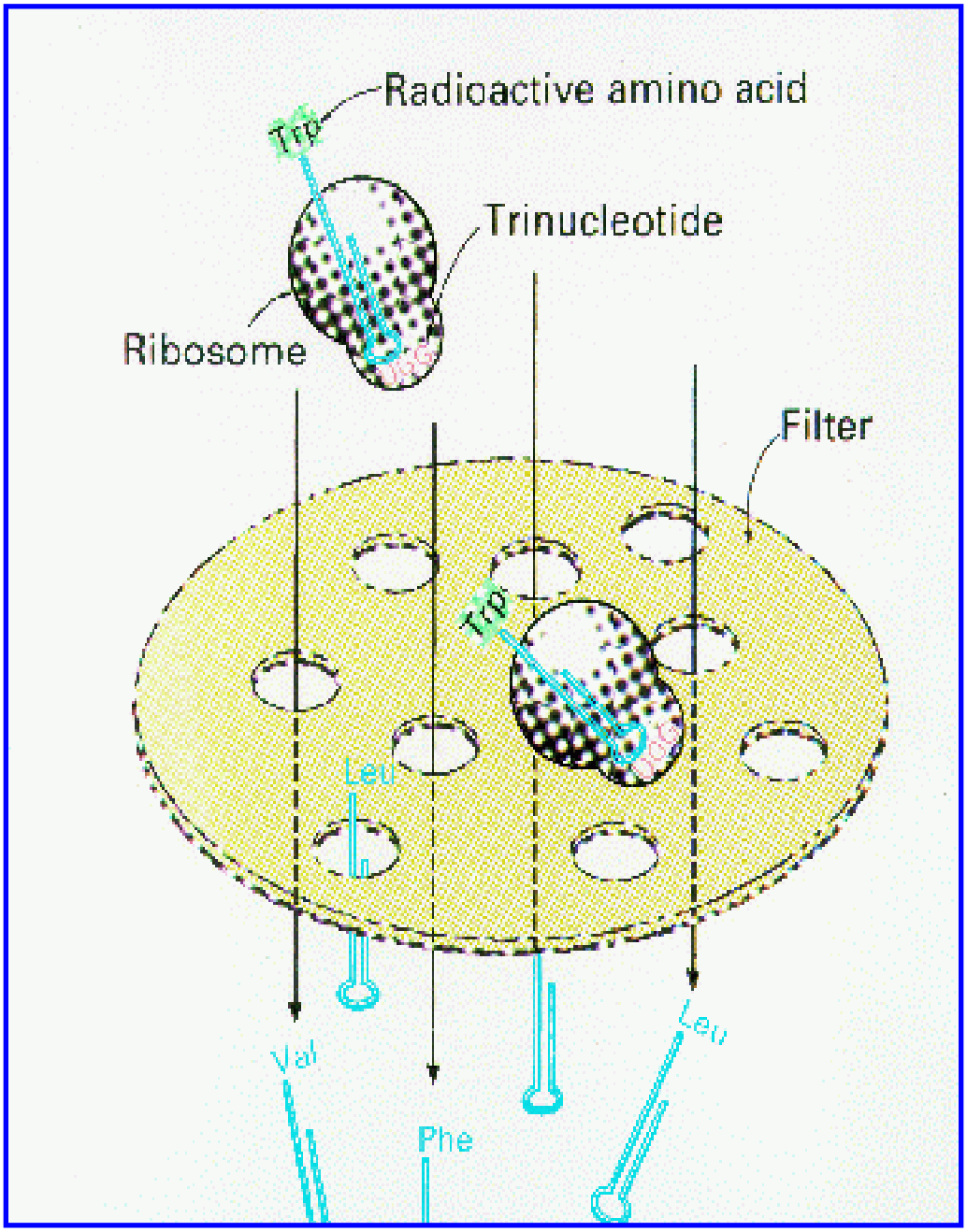
When **AAGAAGAAGAAG** was incubated with bacterial lysate containing ribosomes, tRNAs etc., the following results were obtained:

Reading frame 1	Reading frame 2	Reading frame 3
-AAG-AAG-AAG-AAG- -Lys-Lys-Lys-Lys-	A-AGA-AGA-AGA-AG -Arg-Arg-Arg-	AA-GAA-GAA-GAA-G -Glu-Glu-Glu-

Upon translation, a mixture of homo-polypeptides (poly-lysine, poly-arginine and poly-glutamic acid) was obtained, according to the 3 possible reading frames. No hetero-polypeptides were produced, confirming the absence of overlap and punctuation in the Genetic Code.

In 1964 Nirenberg & Leder developed a filter-binding assay:

Ribosomes were incubated with one radioactively labeled aminoacyl tRNA* (Trp) and unlabeled (non-Trp) tRNAs. Subsequently, one type of synthetic RNA trinucleotide (**AAG** or **CGA** or **UGG**) was added. When this solution was washed over Millipore filters, **only** ribosome/tRNA/ mRNA complexes where tRNA* (Trp) **AND** mRNA (**UGG**) were ***complementary*** remained on the filter. Free, non-complementary tRNAs and mRNA were washed off.



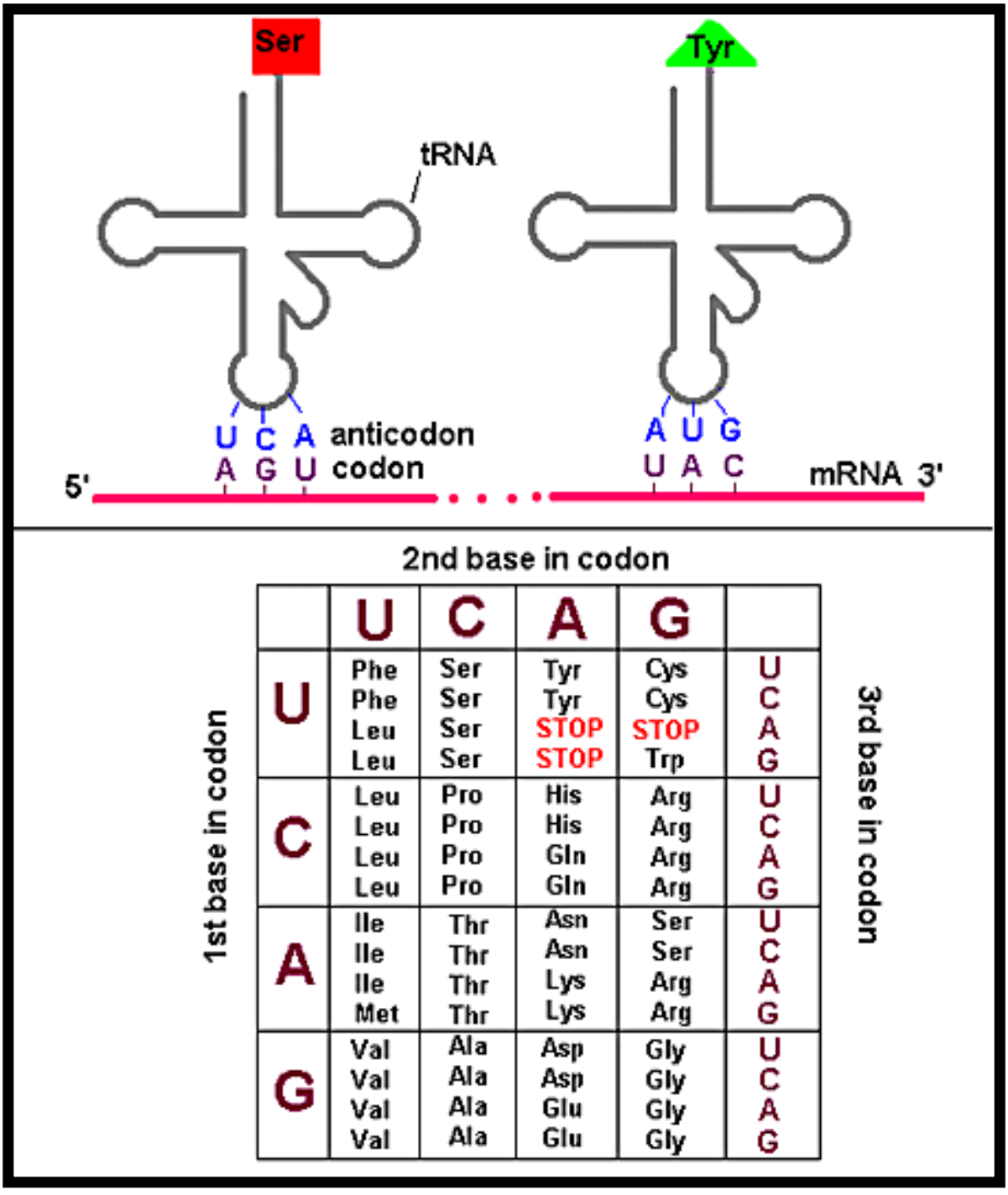
UGG	Trp* tRNA
UUU **UUC**	Phe* tRNA

In this fashion, the remaining code pairs were determined.

Next: The Genetic Code

Please report typos, errors etc. by EMAIL (mention the title of this page).

Table of the Genetic Code



	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

In 1968 Nirenberg & Khorana jointly were awarded the Nobel Prize for the elucidation of the Genetic Code.

Noteworthy observations:

- Most codons for a given amino acid differ only in the last (third) base of the triplet (exceptions: Leu, Arg, Ser)

- One codon (**AUG** or Met) also signals the **START** of a polypeptide chain.
- Three codons (**UAA**, **UAG** and **UGA**) are used to signal the **END** of a polypeptide chain (STOP codons)

For a historical account of the cracking of the Genetic Code click here.

Next: Summary

Please report typos, errors etc. by EMAIL (mention the title of this page).

Summary

Features of the Genetic Code

- The Code transfers ***information*** from mRNA to proteins with high fidelity
- It is ***redundant*** or degenerate: 61 mRNA triplets code for 20 amino acids
- Contains START (1) and STOP (3) codons
- The Genetic Code is nearly ***universal***: correspondence between a nucleotide triplet and an amino acid is identical from viruses to mammals. The rare exceptions are mitochondria and unicellular protozoa.
- The ***universality*** of the Genetic Code is a result of strong evolutionary pressure: a change in a single codon would alter nearly every protein made by an organism.
- The universality is the basis for recombinant protein technology: mammalian mRNA sequences inserted into bacteria will be correctly expressed (translated). More about this in the last lecture: *Expression Systems for Recombinant Proteins*.

Next lecture: tRNA: Structure & Function

Information Transfer and The Central Dogma

DNA -> DNA	DNA -> RNA	RNA -> Protein
Replication	***Transcription***	***Translation***
Substrates: **dNTPs** (A,T,C,G)	Substrates: **rNTPs** (A,U,C,G)	Substrates : **Amino Acids** (20)
Chain growth: **5' to 3'**	Chain growth: **5' to 3'**	Chain growth: **N to C**
by **DNA Polymerase**	by **RNA Polymerase**	Step 1: **tRNA Synthetases** (different for each AA) use energy from ATP to couple amino acids to cognate tRNAs.
Requires template and **primer.**	Requires only template.	Charging specificity determined by 3D structural features unique to each tRNA amino acid pair.
Bases added to 3' OH of primer according to Watson-Crick pairing with template.	Bases added to 3' OH of growing chain according to Watson-Crick pairing with template.	Step 2 by **Ribosomes:**
Initiated at replication **Origins.**	Initiated at **Promoters.**	Initiated at **Start Codon** (AUG). In prokaryotes, preceded by ribosome binding site.
In general double stranded and stable.	Two classses of RNAs:	Ribosomes provide a platform for binding of tRNA anticodon to individual triplet codons in mRNA according to the **Genetic Code**.
Many mechanisms to assure fidelity during replication (proofreading) and maintenance between replicative rounds (recombination and repair).	1. Messenger (mRNA): single stranded, rapid turnover.	Amino acids from charged tRNAs are joined to the carboxyl end of the growing chain. Elongation requires GTP hydrolysis.

	2. Stable RNA (eg. tRNA, ribosomal RNA, snRNAs: folds into compact structures or ribonucleoprotein complexes (RNPs).	
Processing:	Processing:	Processing:
Methylation of bases, ligation of chains, chain cleavage by nucleases. Topological.	Base modification, ligation, cleavage, splicing & editing, polyadenylation, 5' capping.	Phosphorylation, acetylation, chain cleavage by proteases, disulfide crosslinking, lipidation, glycosylation etc.

Please report typos, errors etc. by EMAIL (mention the title of this page).