Código genético

DNA		RNA		Proteína
4 Nucleótidos (ACGT)		4 Nucleótidos (ACGU)		20 aminoácidos
TGGCGAACTGATGTG	Transcripción por polimerasa	UGGCGAACUGAUGUG	Traducción por ribosomas	Trp-Arg-Thr-Asp-
Enlace fosfodiéster		Enlace fosfodiéster		Enlace peptídico

Comparing proteins with nuclei acid:

Properties proteins have in common with nucleic acid:

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

Properties different from nucleic acid:

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

What do proteins do?

Proteins are the molecular workhorse of the cell
Proteins are of central importance in every cellular process

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

Proteins must be made by the cell with high fidelity

Examples of single amino-acid mutations that cause disease:

a) Hemoglobin: glutamic acid (Glu) to valine (Val) mutation at position 6 of the β chain causes sickle-cell anemia
b) Fibroblast Growth Factor (FGF) receptor 3: glycine (Gly) to arginine (Arg) mutation results in achondroplasia
c) Enzyme uroporphyrinogen III cosynthase: causes congenital porphyria with symptoms such as skin photosensitivity & scarring, mutilating skin deformity, hyper trichosis, haemolytic anemia, red stained teeth.

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 mechanism 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 nucleotides) and translating it into the proper linear protein sequence of 20 different amino acids

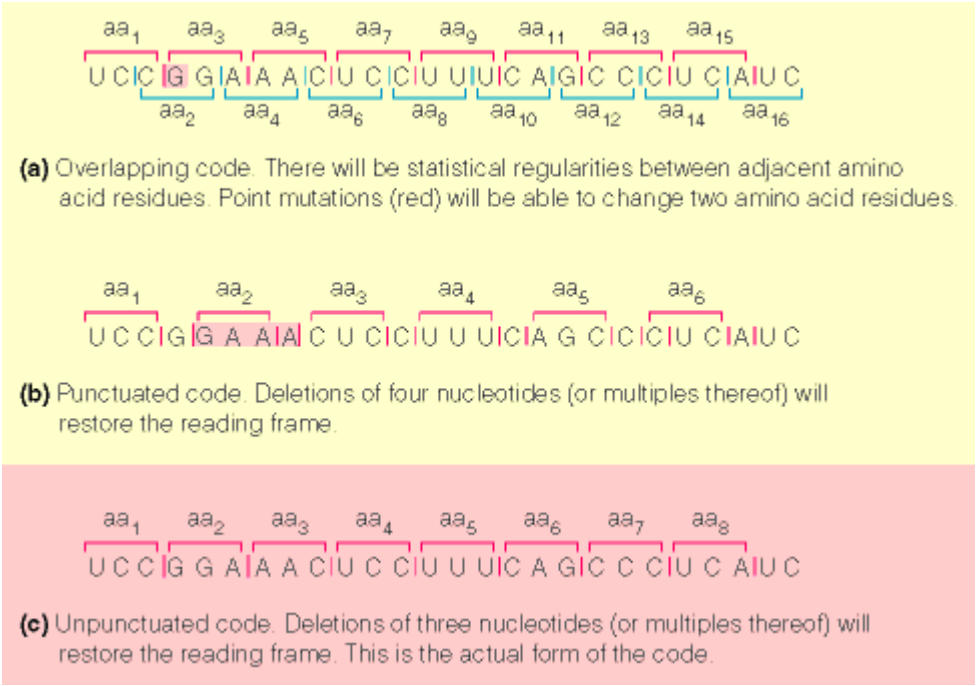
This process is performed in one of the most complex organelles of the cell, the ribosome. In the ribosome the mRNA sequence (information) is read and the corresponding polypeptide (protein) is assembled. The rules for translating the linear nucleic acid sequence (mRNA) into the linear amino acid sequence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

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 hended to code for all 20 amino acids. Thus Crick proposed that codon triplets code for individual amino acids

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



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)

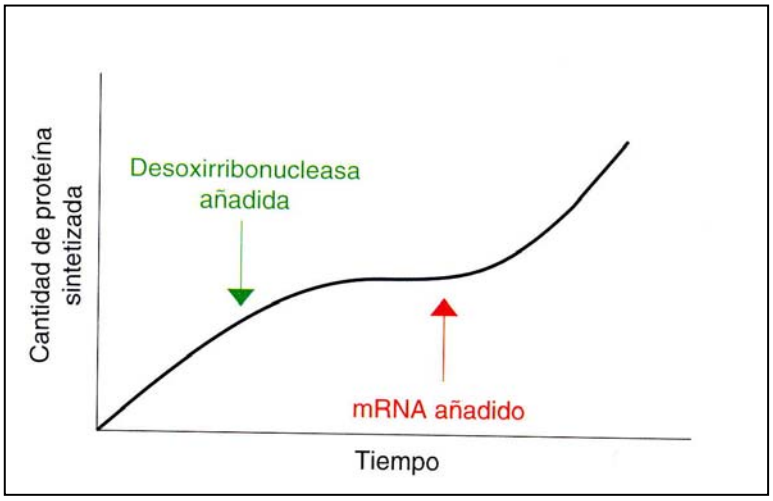
The following experiments were performed in the early 60s. Crick & Brenner (1961) showed the effect of successive deletions of nucleotides in bacteriophage T4 DNA.

ATG CTG CTC TGT GCC GCC **Met Leu Leu Cys Ala Ala**	**Original sequence**
ATG CTC TCT GTG CCG CC . **Met Leu Ser Val Pro Pro**	**1 nucleotide deleted**
ATG CTT 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**

1-Deletion of 1 or 2 nucleotides (frame shift mutation) results in non-functional protein
2-Deletion of 3 nucleotides results only in deletion of 1 amino acid
3-Insertion of 3 nucleotides results in insertion of 1 amino acid
4-Change of 1 nucleotide results in either a sense or silent mutation or in a missense mutation

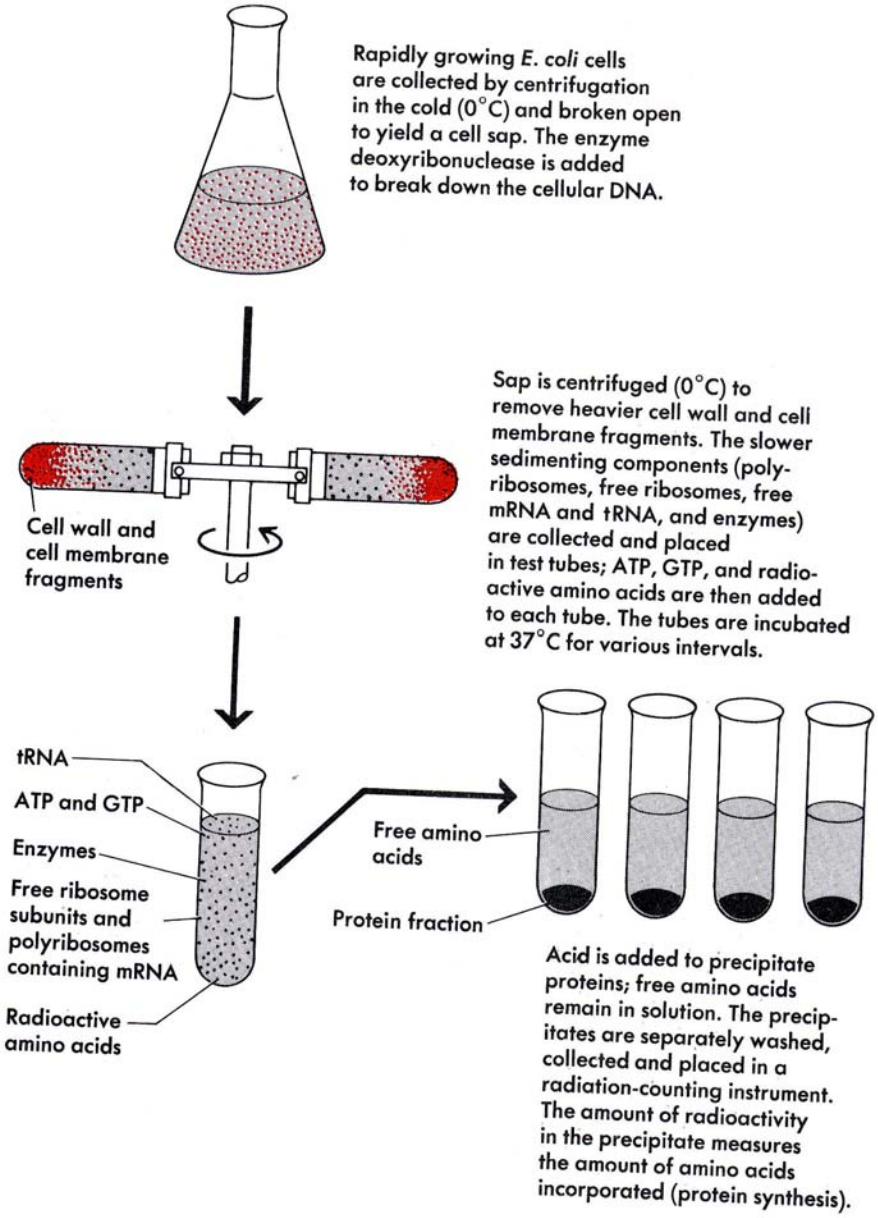
Nucleotides are read as triplets without overlap or punctuation

Deciphering the Genetic Code
Which triplet codon corresponds to which amino acid?



Marshall Nirenberg

Síntesis proteica en un sistema libre de células detenida a los pocos minutos después de la adición de desoxirribonucleasa y continuada por la adición de mRNA.



La polinucleótido fosforilasa forma polímeros del tipo RNA aleatorios

Marianne
Grunberg-Manago

Severo Ochoa

En 1955 Marianne Grunberg-Manago y Severo Ochoa descubrieron el enzima bacteriano polinucleótido fosforilasa, que cataliza in vitro la reacción

$$(RNA)_n + \text{ribonucleósido difosfato} \underset{\text{In vivo}}{\overset{\text{In vitro}}{\rightleftharpoons}} (RNA)_{n+1} + Pi$$

El enzima requiere los 5'difosfatos de los ribonucleótidos y no esta dirigída por un molde. El polimero de RNA formado no tiene una secuencia de baser específica. La reacción transcurre tanto con un sólo nucleósido difosfato como con los cuatro. **"La composición de bases del polímero sintetizado por el enzima refleja las concentraciones relativas de los sustratos 5'-difosfato en el medio".**

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

Poly U	UUUUUUUUUU.	Phe-Phe-Phe-Phe-.
Poly A	AAAAAAAAAA.	Lys-Lys-Lys-Lys-.
Poly C	CCCCCCCCCC.	Pro-Pro-Pro-Pro-.

Thus the first codon was determined: **UUU** codes for phenylalanine; alternatively, **AAA** code for lysine and **CCC** for proline.

Mixed Copolymers Allowed Additional Codon Assgnments

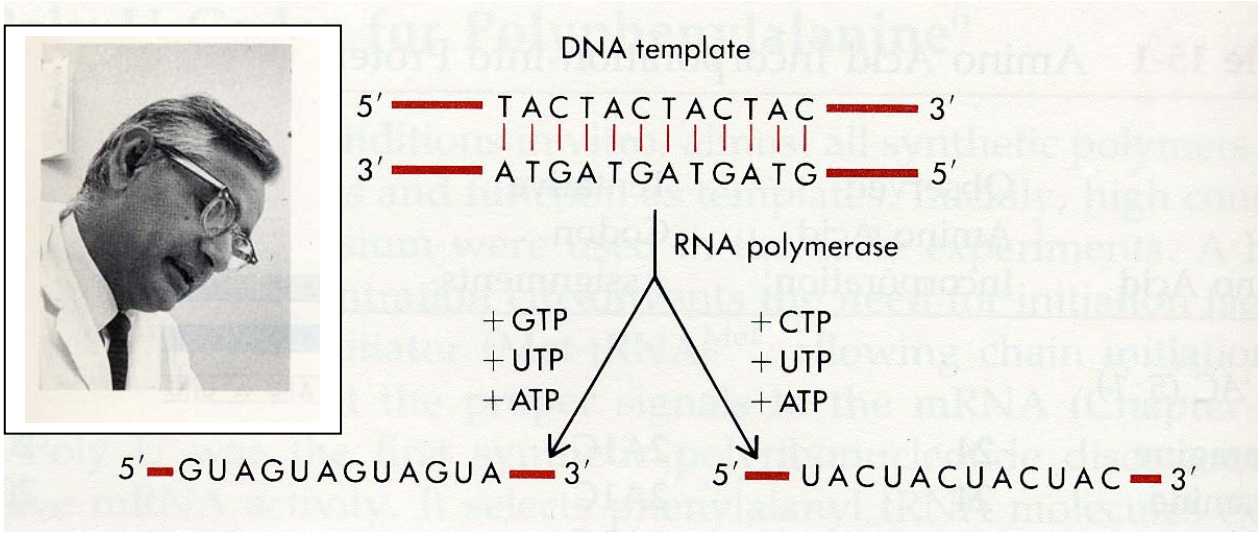
Poly AC molecules can contain eight different codons, CCC, CCA, CAC, ACC, CAA, ACA, AAC, and AAA, whose proportions vary with the copolymer A/C ratio. When AC copolymers attach to ribosomes, they cause the incorporation of asparagine, glutamine, histidine, and threonine, in addition to the proline previously assigned to CCC codons and the lysine previously assigned to AAA codons. The proportions of these amino acids incorporated into polypeptide products depend on the A/C ratio. Thus, since an AC copolymer containing much more A than C promotes the incorporation of many more asparagine than histidine residues, we conclude that asparagine is coded by two As and one C and that histidine is coded by two Cs and one A (Table 15-1). Similar experiments with other copolymers allowed a number of additional assignments. Such experiments, however, did not reveal the order of the different nucleotides within a codon. There is no way of knowing from random copolymers whether the histidine codon containing two Cs and one A is ordered CCA, CAC, or ACC.

Table 15-1 Amino Acid Incorporation into Proteins*

Amino Acid	Observed Amino Acid Incorporation	Tentative Codon Assignments	Calculated Triplet Frequency				Sum of Calculated Triplet Frequencies
			3A	2A1C	1A2C	3C	
Poly AC (5:1)							
Asparagine	24	2A1C		20			20
Glutamine	24	2A1C		20			20
Histidine	6	1A2C			4.0		4
Lysine	100	3A	100				100
Proline	7	1A2C, 3C			4.0	0.8	4.8
Threonine	26	2A1C, 1A2C		20	4.0		24
Poly AC (1:5)							
Asparagine	5	2A1C		3.3			3.3
Glutamine	5	2A1C		3.3			3.3
Histidine	23	1A2C			16.7		16.7
Lysine	1	3A	0.7				0.7
Proline	100	1A2C, 3C			16.7	83.3	100
Threonine	21	2A1C, 1A2C		3.3	16.7		20

*The amino acid incorporation into proteins was observed after adding random copolymers of A and C to a cell-free extract similar to that described in Figure 15-1. The incorporation is given as a percentage of the maximal incorporation of a single amino acid. The copolymer ratio was then used to calculate the frequency with which a given codon would appear in the polynucleotide product. The relative frequencies of the codons are a function of the probability that a particular nucleotide will occur in a given position of a codon. For example, when the A/C ratio is 5:1, the ratio of AAA/AAC = 5 × 5 × 5:5 × 5 × 1 = 125:25. If we thus assign to the 3A codon a frequency of 100, then the 2A and 1C codon is assigned a frequency of 20. By correlating the relative frequencies of amino acid incorporation with the calculated frequencies with which given codons appear, tentative codon assignments can be made.

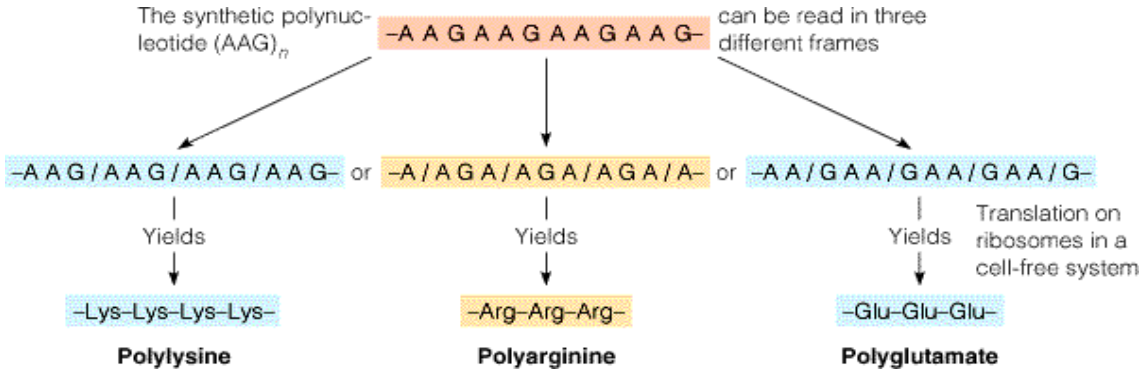
Codon Assignments from repeating copolimers (Khorana)



Using a combination of organic synthesis and copying by DNA polymerase I, double-stranded DNA with simple repeating sequences can be generated. RNA polymerase will then synthesize long polyribonucleotides corresponding to one or the other DNA strand, depending on the choice of ribonucleoside triphosphates added to the reaction mixture.

Assignment of Codons Using Repeating Copolymers Built from Two or Three Nucleotides

Copolymer	Codons Recognized	Amino Acids Incorporated or Polypeptide Made	Codon Assignment
$(CU)_n$	CUC|UCU|CUC . . .	Leucine Serine	5′-CUC-3′ UCU
$(UG)_n$	UGU|GUG|UGU . . .	Cysteine Valine	UGU GUG
$(AC)_n$	ACA|CAC|ACA . . .	Threonine Histidine	ACA CAC
$(AG)_n$	AGA|GAG|AGA . . .	Arginine Glutamine	AGA GAG
$(AUC)_n$	AUC|AUC|AUC . . . UCA|UCA|UCA . . . CAU|CAU|CAU . . .	Polyisoleucine Polyserine Polyhistidine	5′-AUC-3′ UCA CAU



Polipéptidos producidos en respuesta a polímeros sintéticos de RNA con secuencias repetitivas de tres y cuatro bases

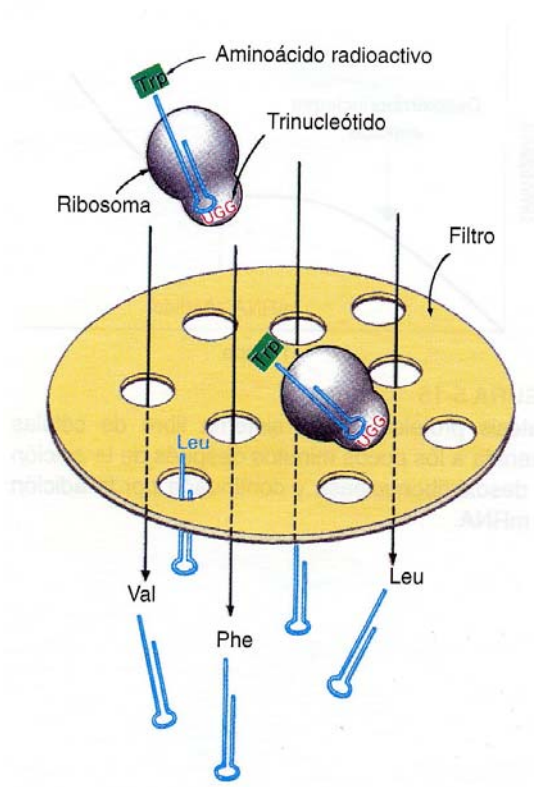
Polinucleótido	Productos polipeptídicos
Repeticiones de trinucleótidos	
$(UUC)_n$	$(Phe)_n$, $(Ser)_n$, $(Leu)_n$
$(AAG)_n$	$(Lys)_n$, $(Arg)_n$, $(Glu)_n$
$(UUG)_n$	$(Leu)_n$, $(Cys)_n$, $(Val)_n$
$(CCA)_n$	$(Pro)_n$, $(His)_n$, $(Thr)_n$
$(GUA)_n$	$(Val)_n$, $(Ser)_n$, (terminador de cadena)*
$(UAC)_n$	$(Tyr)_n$, $(Thr)_n$, $(Leu)_n$
$(AUC)_n$	$(Ile)_n$, $(Ser)_n$, $(His)_n$
$(GAU)_n$	$(Asp)_n$, $(Met)_n$, (terminador de cadena)*
Repeticiones de tetranucleótidos	
$(UAUC)_n$	$(Tyr–Leu–Ser–Ile)_n$
$(UUAC)_n$	$(Leu–Leu–Thr–Tyr)_n$
$(GUAA)_n$	Di- y tripéptidos*
$(AUAG)_n$	Di- y tripéptidos*

*** Con estos polinucleótidos los patrones de incorporación de aminoácidos en polipéptidos están afectados por la presencia de codones que son señales de terminación para la biosíntesis de proteínas . En las secuencias repetitivas de tres nucleótidos, uno de los tres marcos de lectura sólo incluye codones de terminación y, por tanto, sólo se observan dos homopolipéptidos. En algunas de las secuencias repetitivas de cuatro nucleótidos, cada cuarto codón es una señal de parada en cada marco de lectura; por tanto, sólo se producen péptidos cortos.**

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

1) Midiendo la unión codón –dependiente de moléculas específicas de tRNA a los ribosomas.
2) Utilizando como moldes poirribonucleótidos sintéticos con una secuencia ordenada.

Los trinucleótidos promueven el acoplamiento de moléculas específicas de tRNA a los ribosomas.



Ensayo de unión a filtro para la detección de la unión de un trinucleótido a una molécula específica de tRNA. Este complejo se une a un ribosoma que se adhiere al filtro. Por el contrario, el tRNA solo pasa a través del filtro. Las moléculas de tRNA correspondientes a un determinado aminoácido se marcan específicamente con un aminoácido radioactivo por esterificación enzimática. En el ejemplo representado aquí, pUGG se une al tRNA específico del triptófano.

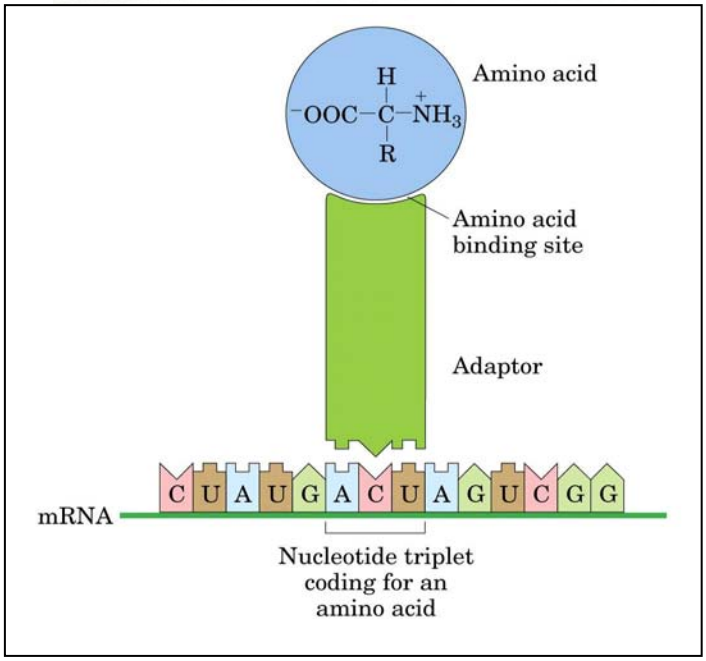


Table 15-2 Binding of Aminoacyl tRNA Molecules to Trinucleotide-Ribosome Complexes

Trinucleotide						AA~tRNA Bound
5′-UUU-3′	UUC					Phenylalanine
UUA	UUG	CUU	CUC	CUA	CUG	Leucine
AAU	AUC	AUA				Isoleucine
AUG						Methionine
GUU	GUC	GUA	GUG	UCU*		Valine
UCU	UCC	UCA	UCG			Serine
CCU	CCC	CCA	CCG			Proline
AAA	AAG					Lysine
UGU	UGC					Cysteine
GAA	GAG					Glutamic acid

*Note that this codon was misassigned by this method.

Trinucleotides Bound to Ribosomes Promote the Binding of Specific Aminoacyl-tRNAs

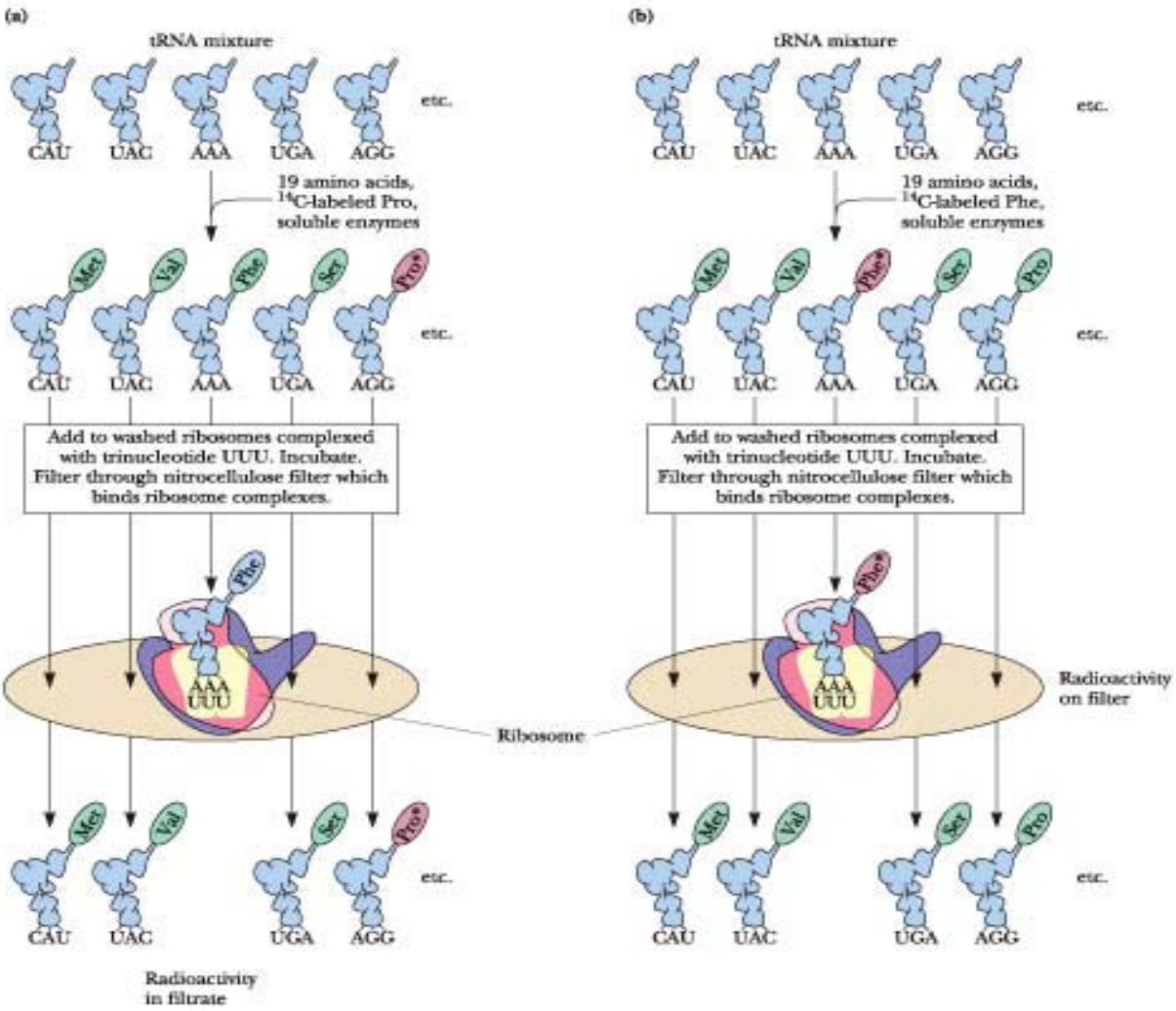


Figure 32.3 · The filter-binding assay for elucidation of the genetic code. Reaction mixture includes washed ribosomes, Mg2 1, a particular trinucleotide (pUpUpU in this example), and all 20 aminoacyl-tRNAs, one of which is radioactively (14C) labeled. (a) 14C-labeled prolyl-tRNA. (b) 14C-labeled phenylalanyl-tRNA. Only the aminoacyl-tRNA whose binding is directed by the trinucleotide codon will become bound to the ribosomes and retained on the nitrocellulose filter. The amount of radioactivity retained by the filter is a measure of trinucleotide-directed binding of a particular labeled aminoacyl-tRNA by ribosomes. Use of this binding assay to test the 64 possible codon trinucleotides against the 20 different amino acids quickly enabled researchers to assign triplet code words to the individual amino acids. The genetic code was broken. (Adapted from Nirenberg, M. W., and Leder, P., 1964. RNA codewords and protein synthesis. Science 145:1399 - 1407)

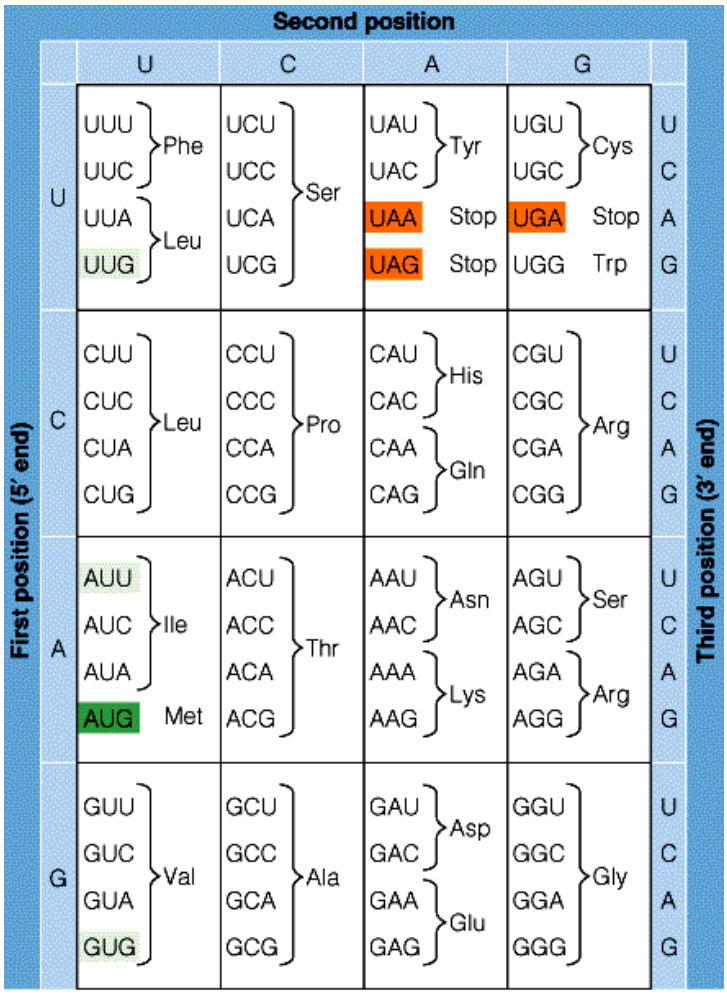
In 1964, Marshall Nirenberg and Philip Leder reported that trinucleotides bound to ribosomes directed the binding of specific aminoacyl-tRNAs. That is, ternary ribosome:trinucleotide:aminoacyl-tRNA complexes could be formed, provided the right trinucleotide and aminoacyl-tRNA combination was present. Aminoacyl-tRNAs were prepared by adding all 20 amino acids to a purified tRNA mixture in the presence of a soluble E. coli fraction containing the necessary aminoacyl-tRNA synthetases. Only one of the amino acids was 14C-labeled in any one binding assay. Trinucleotides are the equivalent of codons, so if a specific trinucleotide promoted the binding of a particular 14C-labeled amino-acyl-tRNA, the base sequence of the trinucleotide must be the code word for that amino acid. Binding was detected because the ribosomes were retained on a nitrocellulose filter while free aminoacyl-tRNAs passed through; only aminoacyl-tRNAs bound by ribosomes were retained (Figure 32.3).

This system was quickly exploited to elucidate the genetic code. Elucidation of the genetic code was probably the greatest scientific achievement of the 1960s. For their roles in it, Marshall Nirenberg and H. Gobind Khorana shared in the 1968 Nobel Prize for physiology or medicine.

1Because polyguanylic acid (poly[G]) has a very strong tendency to form multistranded helices, it was a poor template for protein synthesis. The fact that GGG codes for Gly was not learned until later.

Código genético

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



Codon	Usual Use	Alternate Use	Where Alternate Use Occurs
AGA AGG	Arg	Stop, Ser	Some animal mitochondria, some protozoans
AUA	Ile	Met	Mitochondria
CGG	Arg	Trp	Plant mitochondria
CUU CUC CUA CUG	Leu	Thr	Yeast mitochondria
AUU GUG UUG	Ile Val Leu	Start (*N*-fMet)	Some prokaryotes[a]
UAA UAG	Stop	Glu	Some protozoans
UGA	Stop	Trp Selenocysteine	Mitochondria, mycoplasmas *E. coli*[a]

[a]Depends on context of message, other factors

Features of the Genetic Code

a) The Code transfers **information** from mRNA to proteins with high fidelity

b) It is **redundant or degenerate**: 61 mRNA triplets code for 20 amino acids

c) Most codons for **a given amino acid differ only in the last (third) base** of the triplet (exeptions. Leu, Arg y Ser)

d) One codon (**AUG** or GUG, UUG and AUU) also signals the **START** of a plypeptide chain

e) Three codons (**UAA, UAG and UGA**) are used to signal the **END** of a polypeptide chain (**STOP** codons)

f) 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

g) The universality is the basis for **recombinant protein technology**: mammalian mRNA sequences inserted into bacteria will be correctly expressed (translated)

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.