The Genetic Code

How is the information for a polypeptide sequence stored within an mRNA molecule? There are twenty different common amino acids, but only four different bases in RNA (*A, C, G, and U*).

Base Arrangement	Possible Combinations
1	$4^1=4$
2	$4^2=16$
3	$4^3=64$
4	$4^4=256$

A triplet arrangement would seem to be the minimum possible combination necessary to code for the 20 different amino acids. Although, there are obviously going to be a lot of codons "left over". Most amino acids are coded for by more than a single unique triplet, and therefore the genetic code is said to be **degenerate**.

Experiments which led to the solution of the genetic code:

Nirenberg and Matthei (1961): Nirenberg and Matthei worked with bacterial extracts which contained everything needed for translation, with the exception of mRNA. To this they added either poly A, poly U or poly C RNA. The proteins produced by the translation of these RNA's was determined (*poly G did not work, probably due to conformational problems*):

Poly U	Poly A	Poly C
Phe	Lys	Pro

Genetic Code – Codon

Codon:

3-base RNA sequence

(5') . . . pNpNpN . . . (3') in mRNA

Base at 5' End of Codon ↓	Middle Base of Codon → U	C	A	G	Base at 3' End of Codon ↓
U	phe (UUU)	ser	tyr	cys	U
	phe	ser	tyr	cys	C
	leu	ser	termination	termination	A
	leu	ser	termination	trp	G
C	leu	pro	his	arg	U
	leu	pro	his	arg	C
	leu	pro	gln	arg	A
	leu	pro	gln	arg	G
A	ile	thr	asn	ser	U
	ile	thr	asn	ser	C
	ile	thr	lys	arg	A
	met (and initiation)	thr	lys	arg	G
G	val	ala	asp	gly	U
	val	ala	asp	gly	C
	val	ala		gly	A
	val	ala		gly	G

Stop codons

Start codon

List of Amino Acids

	Amino acid	Symbol	Codon
A	Alanine	Ala	GC*
C	Cysteine	Cys	UGU, UGC
D	Aspartic Acid	Asp	GAU, GAC
E	Glutamic Acid	Glu	GAA, GAG
F	Phenylalanine	Phe	UUU, UUC
G	Glycine	Gly	GG*
H	Histidine	His	CAU, CAC
I	Isoleucine	Ile	AUU, AUC, AUA
K	Lysine	Lys	AAA, AAG
L	Leucine	Leu	UUA, UUG, CU*

List of Amino Acids

	Amino acid	Symbol	Codon
M	Methionine	Met	AUG
N	Asparagine	Asn	AAU, AAC
P	Proline	Pro	CC*
Q	Glutamine	Gln	CAA, CAG
R	Arginine	Arg	CG*, AGA, AGG
S	Serine	Ser	UC*, AGU, AGC
T	Threonine	Thr	AC*
V	Valine	Val	GU*
W	Tryptophan	Trp	UGG
Y	Tyrosine	Tyr	UAU, UAC

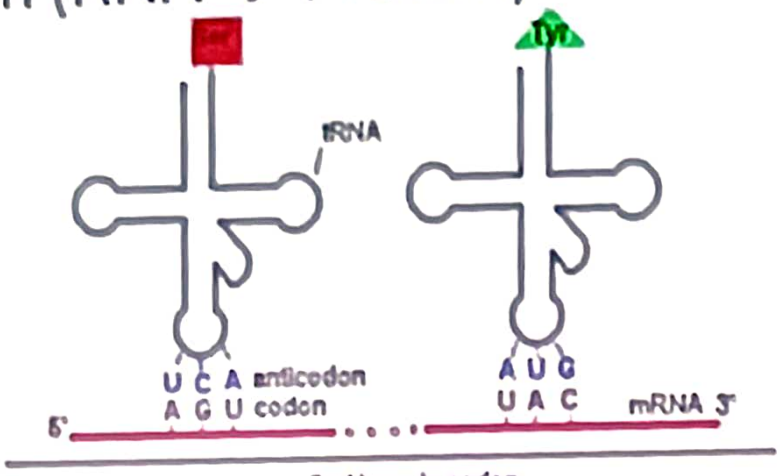
20 letters, no B J O U X Z

Codon and Reading Frame

- 4 letters → 4^3 = 64 triplet possibilities
- 20 (< 64) known amino acids
- *Wobbling* 3rd base
- Redundant → Resistant to mutation
- **Reading frame:** linear sequence of codons in a gene
- **Open Reading Frame (ORF),** definition varies:
  - a reading frame that begins with a start codon and end at a stop codon
  - a series of codons in a DNA sequence uninterrupted by the presence of a stop codon

  → a potential protein-coding region of DNA sequence

Translation (RNA → Protein)

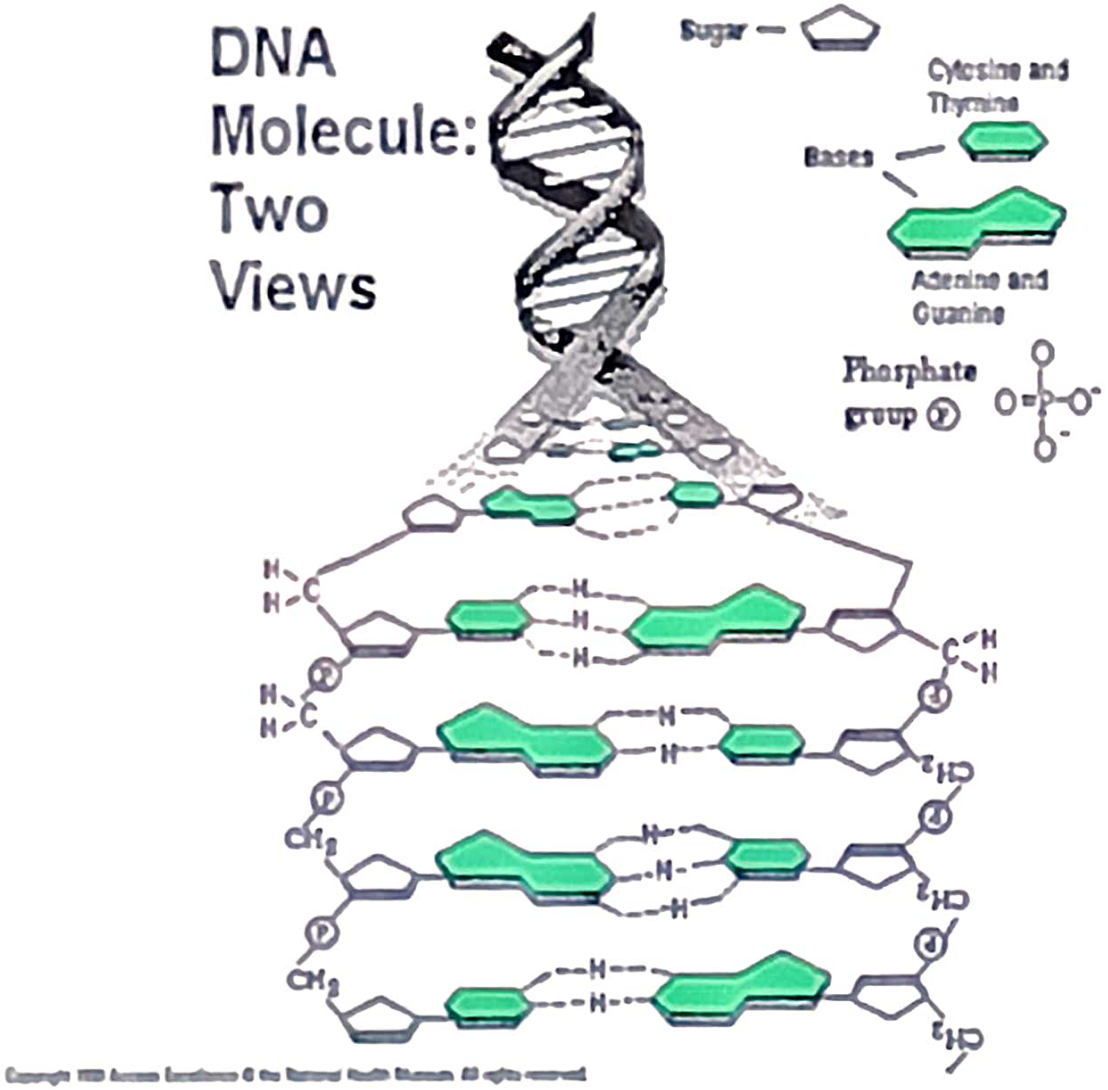


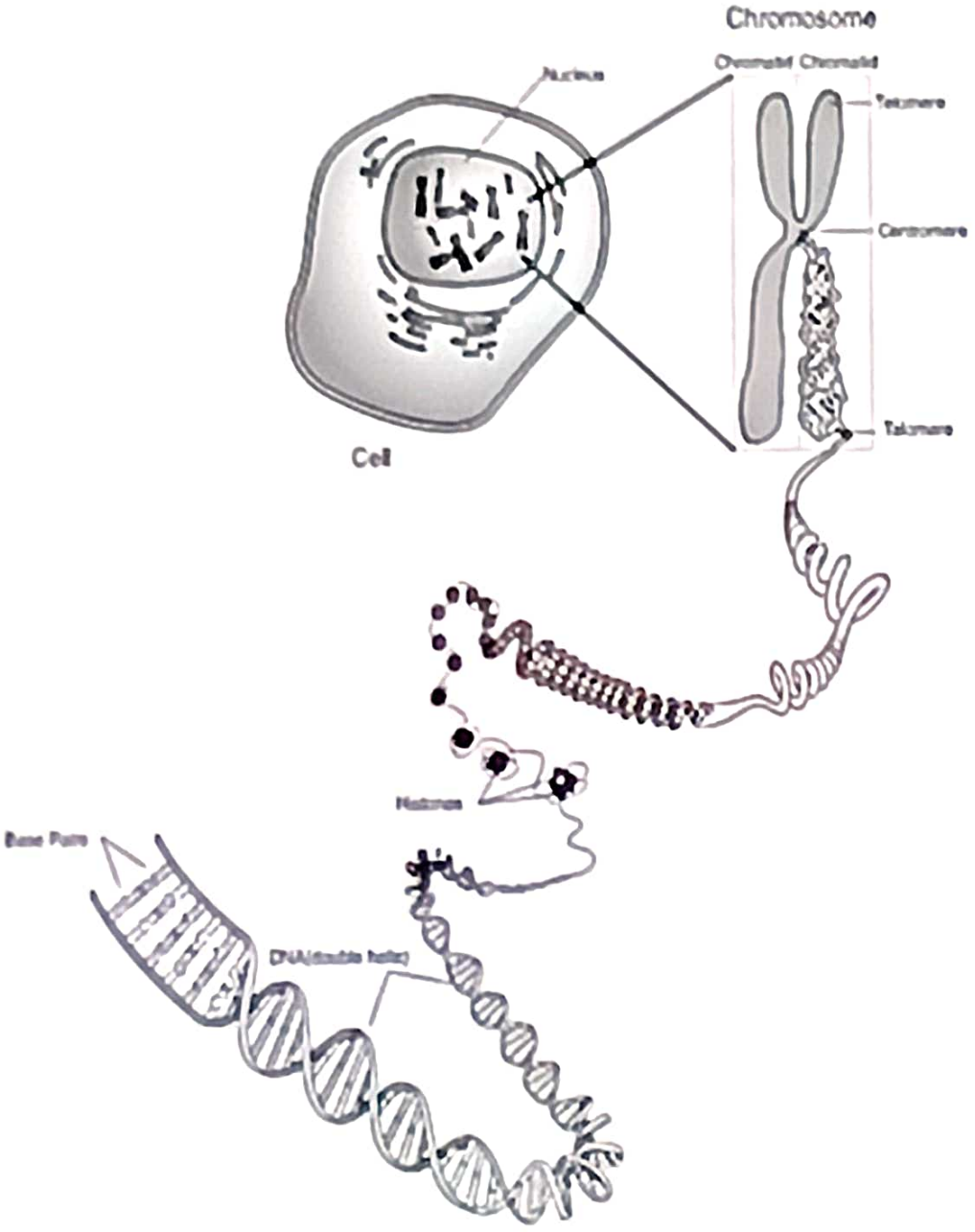
2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
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

The Genetic Code

DNA Structure





Genetic Material

- **DNA** (deoxyribonucleic acid) is the genetic material
- Information stored in DNA
  - the basis of inheritance
  - distinguishes living things from nonliving things
- **Genes**
  - various units that govern living thing's characteristics at the genetic level

Nucleotides

- Complicated genes can be many thousands of nucleotides long
- All of an organism's genetic instructions, its **genome**, can be maintained in millions or even billions of nucleotides

Orientation

- Strings of nucleotides can be attached to each other to make long **polynucleotide** chains
- 5' (5 prime) end
  - The end of a string of nucleotides with a 5' carbon not attached to another nucleotide
- 3' (3 prime) end
  - The other end of the molecule with an unattached 3' carbon.

Central dogma

Background

A few years after he and James Watson had proposed the double helical structure for DNA, Francis Crick (with other collaborators) proposed that a less stable nucleic acid, RNA, served as a messenger RNA that provided a transient copy of the genetic material that could be translated into the protein product encoded by the gene. Such mRNAs were indeed found. These and other studies led Francis Crick to formulate this **"central dogma"** of molecular biology.

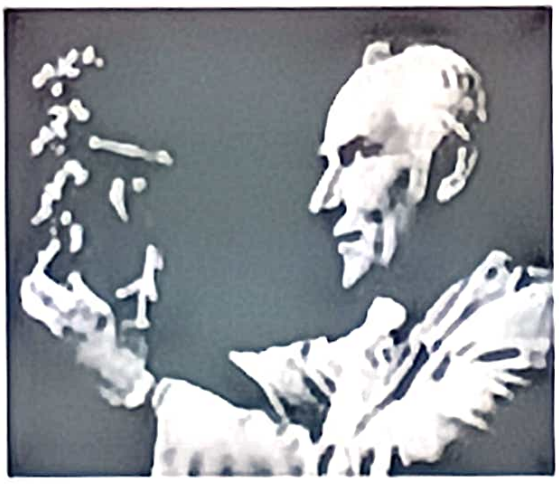
This model states that **DNA serves as the repository of genetic information**. It can be **replicated** accurately and indefinitely. The **genetic information is expressed** by the DNA first serving as a template for the **synthesis of (messenger) RNA**; this occurs in a process called **transcription**. The mRNA then serves as a template, which is read by ribosomes and **translated** into **protein**. The protein products can be enzymes that catalyze the many metabolic transformations in the cell, or they can be structural proteins.

Note the **static role of DNA** in this process. Implicit in this model is the idea that DNA does not provide an active cellular function, but rather it encodes macromolecules that are functional.

Central Dogma of Molecular Biology

"The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid."

Francis Crick, 1958



Central Dogma of Molecular Biology

- DNA: information storage
- Protein: function unit, such as *enzyme*
- Gene: instructions needed to make protein
- Central dogma

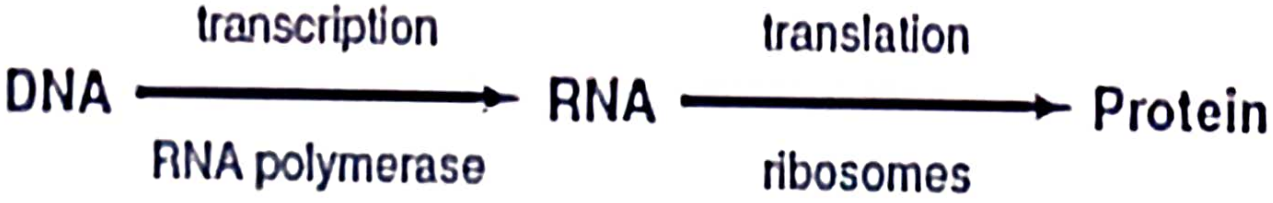
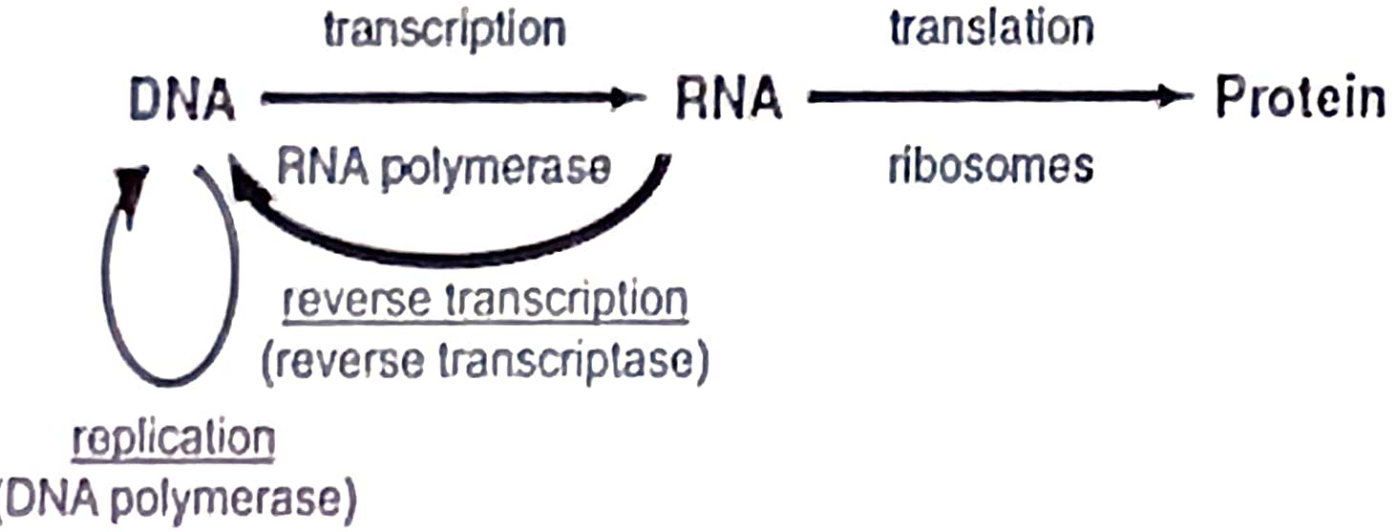


FIGURE 1.4 *The central dogma of molecular biology. Information in cells passes from DNA to RNA to proteins. RNA is made from DNA molecules during transcription by RNA polymerases. Proteins are made from the information content of RNA molecules as they are translated by ribosomes. DNA polymerases also make copies of DNA molecules during the replication process of cell*

Central Dogma of Molecular Biology



- DNA obtained from reverse transcription is called complementary DNA (cDNA)