Lectures 9

The Genetic Code

Acknowledgements:
Leninger Chap 27
Scitable
Internet Resources

Objectives of the lecture

1. Learn about the background research that led to the deciphering of the Genetic Code
2. The Experiment of Nirenberg
3. The establishment of the genetic code

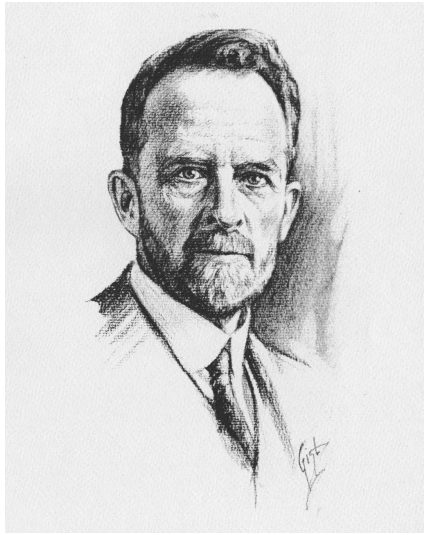
The protein synthesis process is complex

1. Eukaryotic protein synthesis involves more than
   a) 70 different ribosomal proteins
   b) 20 or more enzymes to activate the amino acid precursors
   c) a dozen or more auxiliary enzymes and other protein factors for the initiation, elongation, and termination of polypeptides
   d) 100 additional enzymes for the final processing of different protein
   e) 40 or more kinds of transfer and ribosomal RNAs
2. Overall, almost 300 different macromolecules cooperate to synthesize polypeptides
3. Every procaryote or eucaryote cell has thousands of copies of different RNAs and proteins which constitutes about 35% of the cell dry weight
4. Protein synthesis utilizes about 90% of the chemical energy of the cell

Important Contributions that Lead to the Deciphering of the Genetic Code

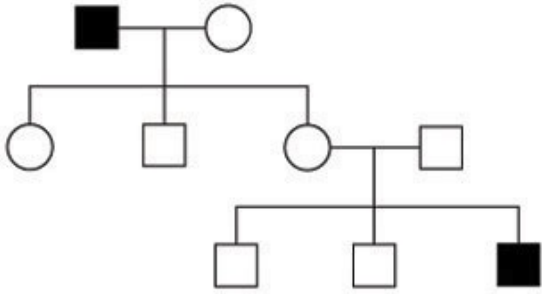
Gregor Mendel

Walter Sutton

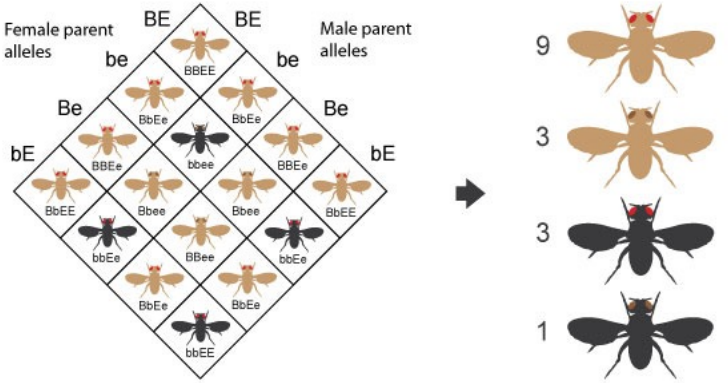


Thomas Hunt Morgan

- Elementen
- External resemblance
- Internal nature



Existence of Chromosomes in pairs



Deciphering the Genetic Code

- 1865 –Mendel defined the basic unit of inheritance as the gene
- 1900 –Mendel's forgotten work resurfaces; nature of gene is still unknown
- 1944 –it is established that a gene is made of DNA
- 1953 –Watson-Crick's double helix structure for DNA

DNA: L = {A, C, G, T}

RNA: L = {A, C, G, U}

Double Stranded DNA

5' A T T G C C C A T 3'

| | | | | | | | |

3` T A A C G G G T A 5'

One big question remained unanswered: how is the information in the DNA strand translated to protein?

George Gamow and the "RNA tie Club"

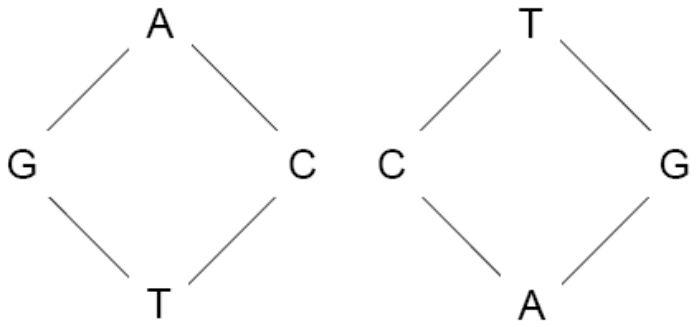
- Brotherhood consisted of 20 regular members, one for each amino acid
  - Watson was PRO (proline)
- Four honorary members, one for each nucleotide
- Eight of these members were or became Nobel Laureates

Georgiy Antonovich Gamov
March 4, 1904- August 19, 1968
Big Bang Theory
Formation of stars

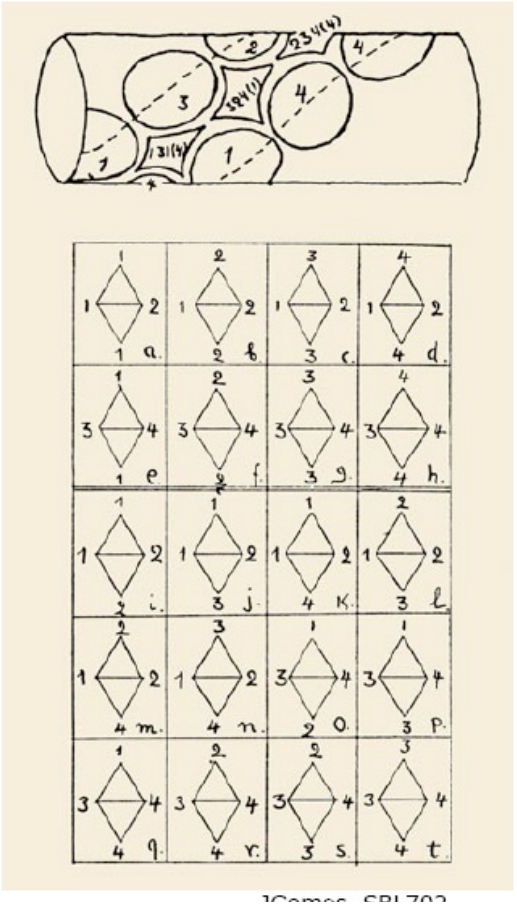
Some of the Ideas Proposed by the RNA Club

- The Adapter Hypothesis by Francis Crick – some unknown biological entity carried the amino acids and put them in the sequence order
- Gamow proposed that a three-letter code would be sufficient to define all 20 amino acids

Combinatorial Figures of Gamow's proposition



- Number of diamonds where top and bottom are identical
- $^4C_1 \times 2 = 8$
- Number of diamonds where top and bottom are different
- $^4C_2 \times 2 = 12$



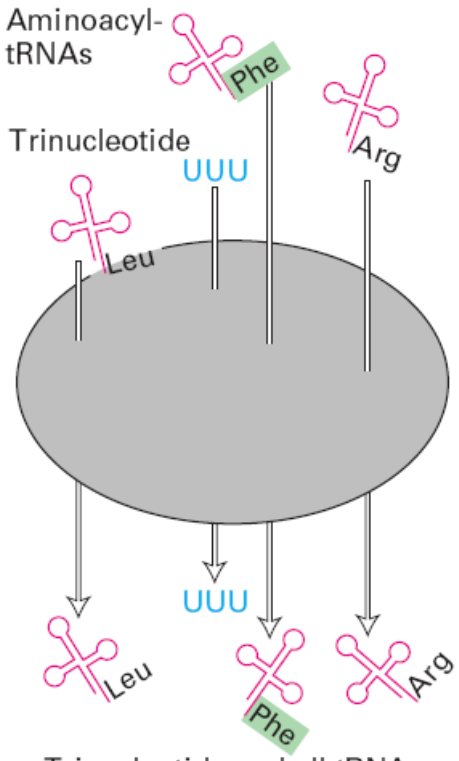
Analysis Presented by Crick

- Genetic code is in triplets (codon)
  - There are 20 amino acids
  - There are 4 alphabets A, T, G, C
  - 4 & $4^2 < 20$, $\therefore 4^3 = 64$ (but with redundancies ?)
- The genetic code should be comma free
- Only one valid reading frame
  - [abc][def][ghi][jkl]
  - NOT a [bcd][efg]hij]…
  - NOT ab[cde][fgh][ijk]…

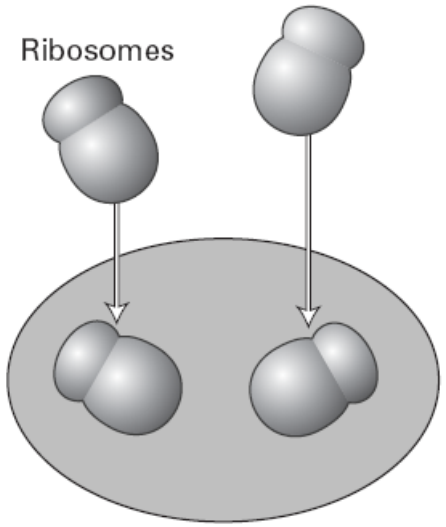
(AND THE FOLLOWING THOUGH NOT QUITE CORRECT)

- AAA, TTT, GGG, CCC are not possible because for example AAAAAA the reading frame is ambiguous
  - That leaveS 64-4 = 60
- ATGATG must be read unambiguously, So, whenever ATG is a codon, TGA or GAT is not
- That gives (1/3)*60 = 20

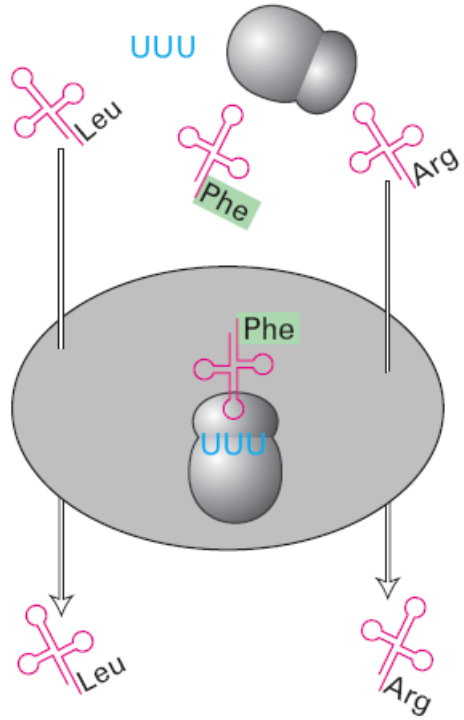
Marshall Nirenberg Deciphered the Genetic Code in 1961



Trinucleotide and all tRNAs pass through filter



Ribosomes stick to filter



Complex of ribosome, UUU, and Phe-tRNA sticks to filter

Marshall Warren Nirenberg
April 10, 1927 – January 15, 2010, ; Jewish American biochemist and geneticist

M. W. Nirenberg and P. Leder, 1964, *Science* **145**:1399

The Genetic Code

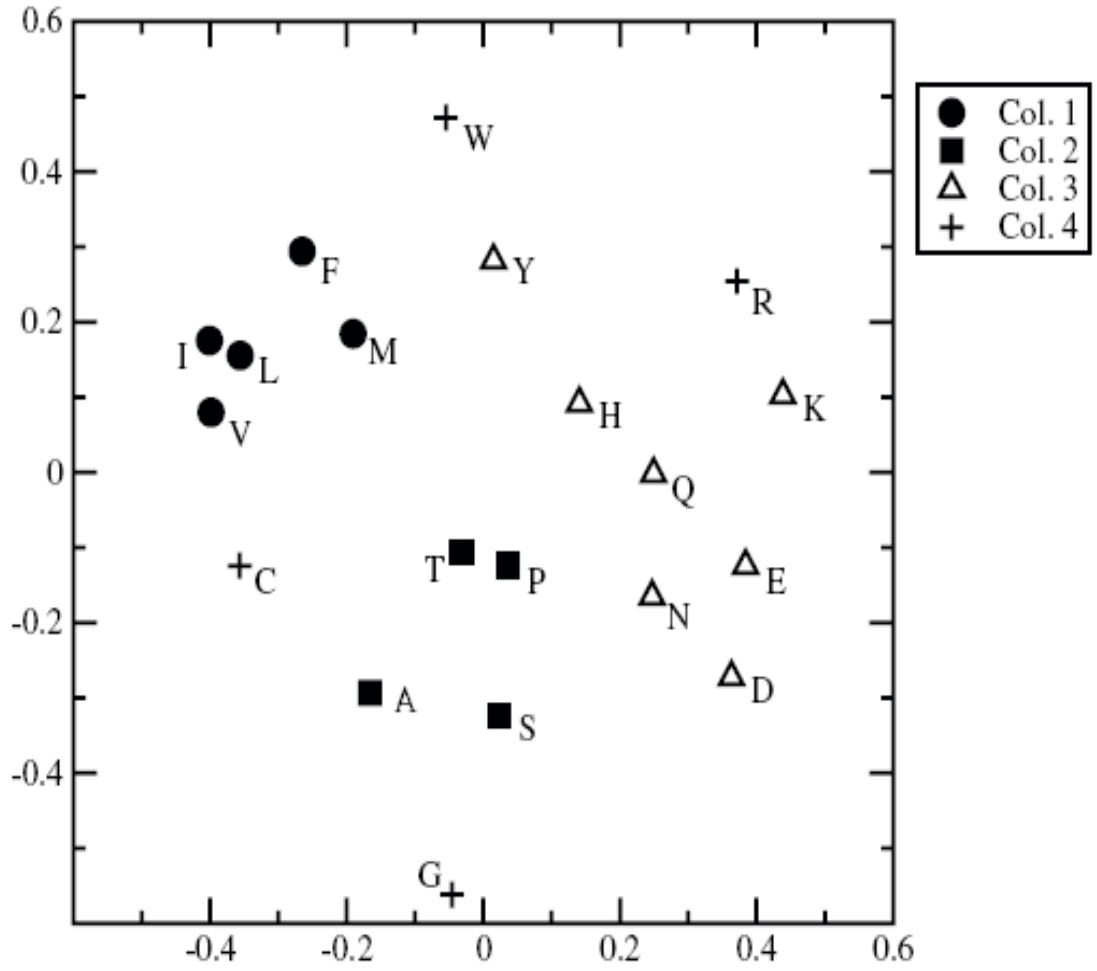
The genetic code is a map of Codons “C” to Amino Acids “A”

g: C → A

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

Grouping by Physical Properties of Amino Acids Best Explains the Genetic Code Table

1st base	2nd base: U	2nd base: C	2nd base: A	2nd base: G	3rd base
U	UUU *Phe*	UCU *Ser*	UAU *Tyr*	UGU *Cys*	U
	UUC *Phe*	UCC *Ser*	UAC *Tyr*	UGC *Cys*	C
	UUA *Leu*	UCA *Ser*	UAA Stop	UGA Stop	A
	UUG *Leu*	UCG *Ser*	UAG Stop	UGG *Trp*	G
C	CUU *Leu*	CCU *Pro*	CAU *His*	CGU *Arg*	U
	CUC *Leu*	CCC *Pro*	CAC *His*	CGC *Arg*	C
	CUA *Leu*	CCA *Pro*	CAA *Gln*	CGA *Arg*	A
	CUG *Leu*	CCG *Pro*	CAG *Gln*	CGG *Arg*	G
A	AUU *Ile*	ACU *Thr*	AAU *Asn*	AGU *Ser*	U
	AUC *Ile*	ACC *Thr*	AAC *Asn*	AGC *Ser*	C
	AUA *Ile*	ACA *Thr*	AAA *Lys*	AGA *Arg*	A
	AUG *Met*	ACG *Thr*	AAG *Lys*	AGG *Arg*	G
G	GUU *Val*	GCU *Ala*	GAU *Asp*	GGU *Gly*	U
	GUC *Val*	GCC *Ala*	GAC *Asp*	GGC *Gly*	C
	GUA *Val*	GCA *Ala*	GAA *Glu*	GGA *Gly*	A
	GUG *Val*	GCG *Ala*	GAG *Glu*	GGG *Gly*	G



Important points related to translation

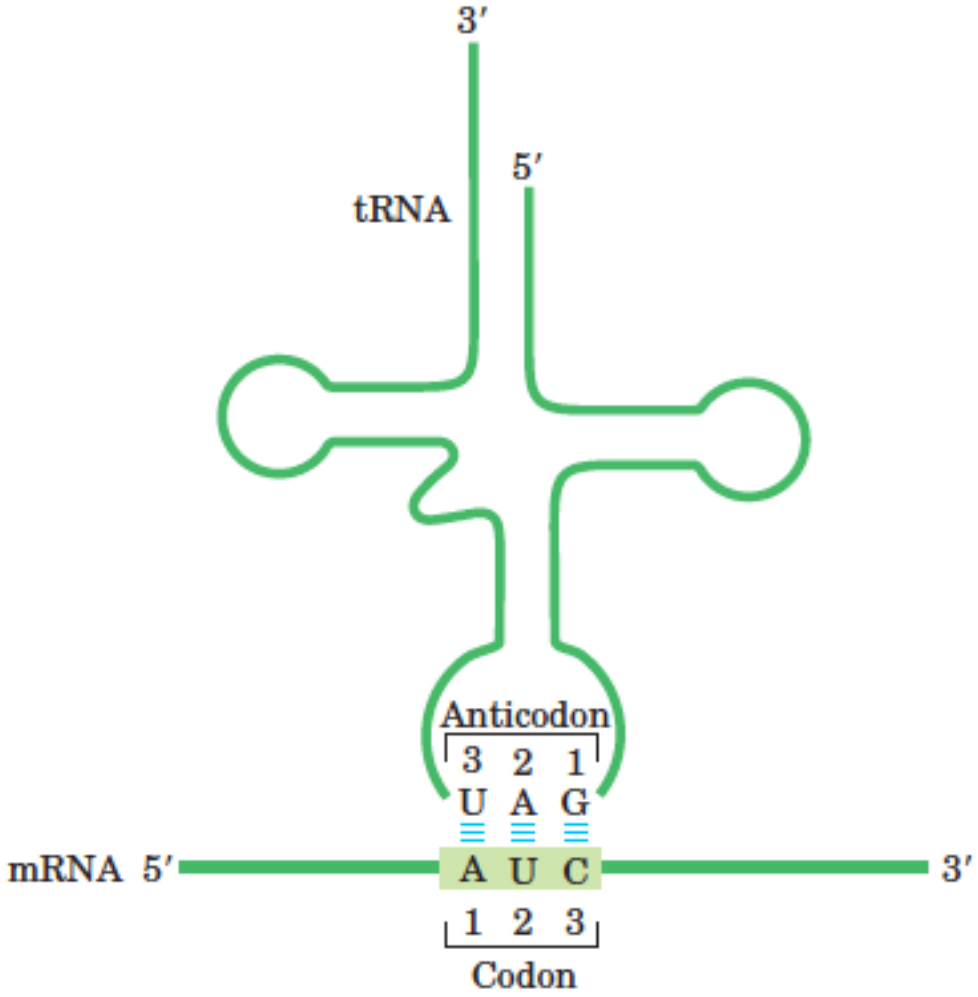
- The particular amino acid sequence of a protein is constructed through the translation of information encoded in mRNA. This process is carried out by ribosomes.
- Amino acids are specified by mRNA codons consisting of nucleotide triplets. Translation requires adaptor molecules, the tRNAs, that recognize codons and insert amino acids into their appropriate sequential positions in the polypeptide.
- The base sequences of the codons were deduced from experiments using synthetic mRNAs of known composition and sequence.
- The codon AUG signals initiation of translation. The triplets UAA, UAG, and UGA are signals for termination.

Degeneracy of the Genetic Code

Amino acid	*Number of codons*	*Amino acid*	*Number of codons*
Met	1	Tyr	2
Trp	1	Ile	3
Asn	2	Ala	4
Asp	2	Gly	4
Cys	2	Pro	4
Gln	2	Thr	4
Glu	2	Val	4
His	2	Arg	6
Lys	2	Leu	6
Phe	2	Ser	6

The Wobble Hypothesis

- Alignment of the two RNAs is antiparallel. The tRNA is shown in the traditional cloverleaf configuration
- Three different codon pairing relationships are possible when the tRNA anticodon contains inosinate.



		3	2	1		3	2	1		3	2	1	
Anticodon	(3′)	G	C	I		G	C	I		G	C	I	(5′)
Codon	(5′)	C	G	A		C	G	U		C	G	C	(3′)
		1	2	3		1	2	3		1	2	3	

1. One codon recognized:

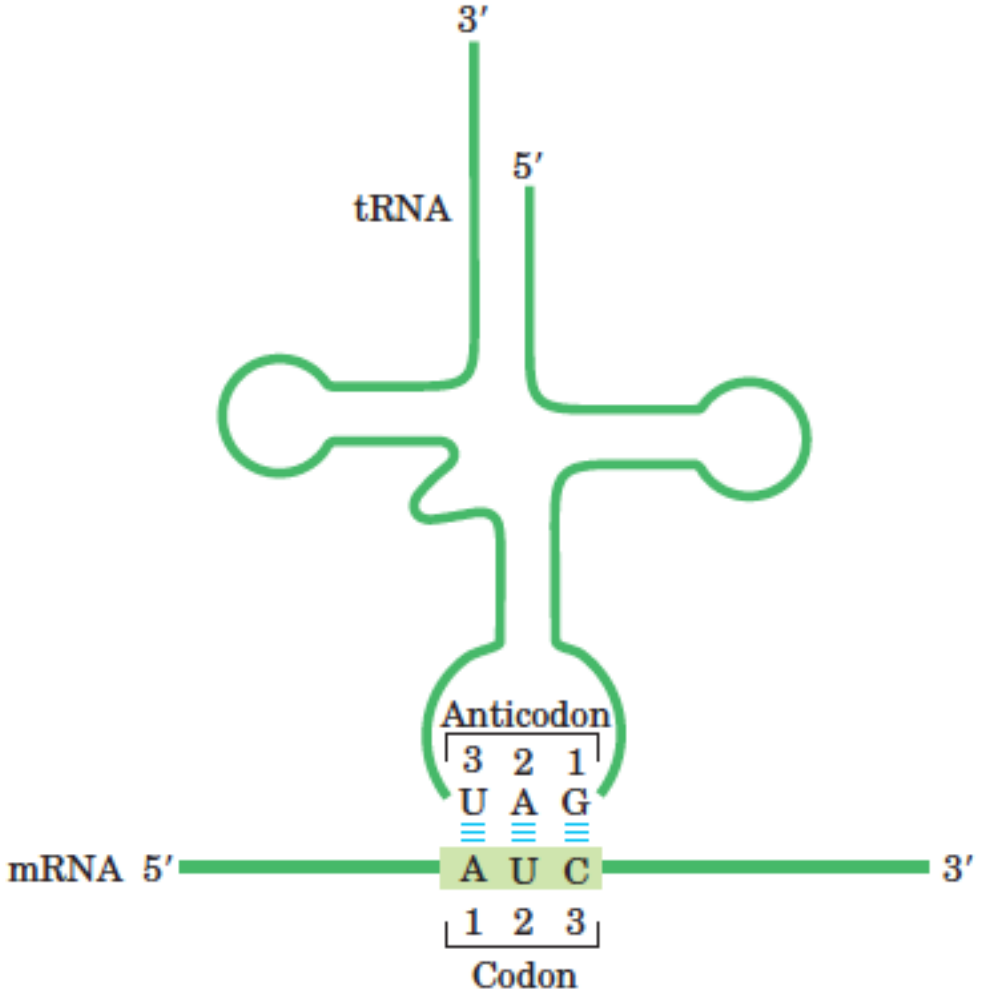
1. Anticodon	(3′) X–Y–**C** (5′)	(3′) X–Y–**A** (5′)
Codon	(5′) Y–X–**G** (3′)	(5′) Y–X–**U** (3′)

2. Two codons recognized:

1. Anticodon	(3′) X–Y–**U** (5′)	(3′) X–Y–**G** (5′)
Codon	(5′) Y–X–**A G** (3′)	(5′) Y–X–**C U** (3′)

3. Three codons recognized:

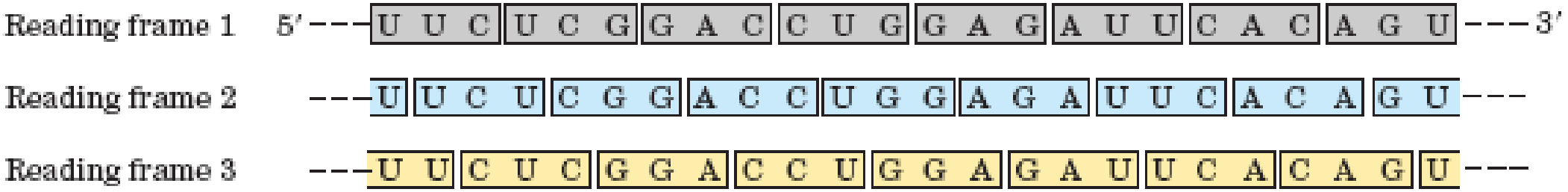
1. Anticodon	(3′) X–Y–**I** (5′)	
Codon	(5′) Y–X–**A U C** (3′)	

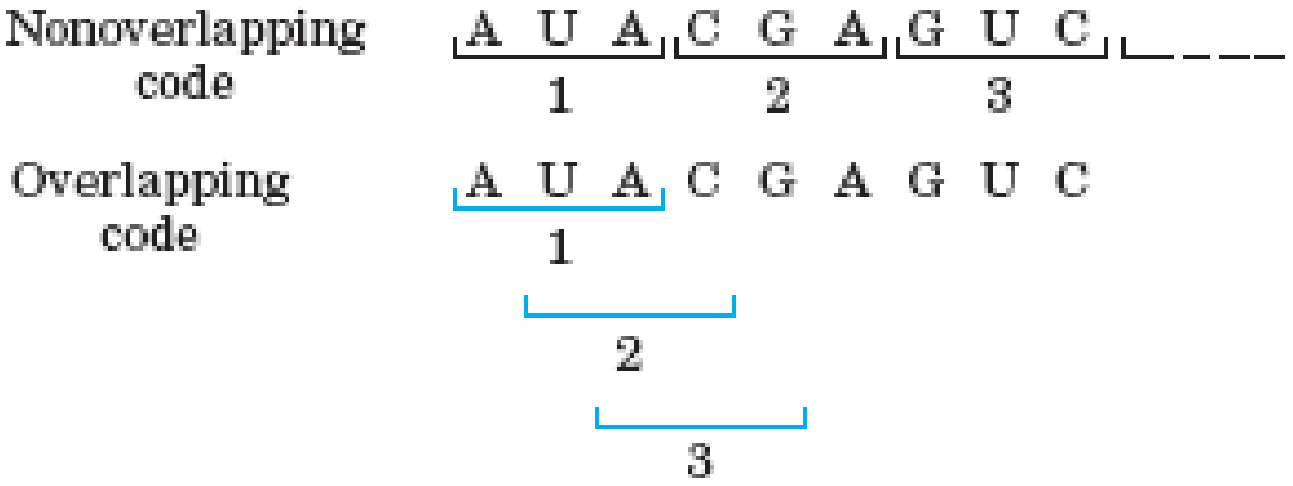


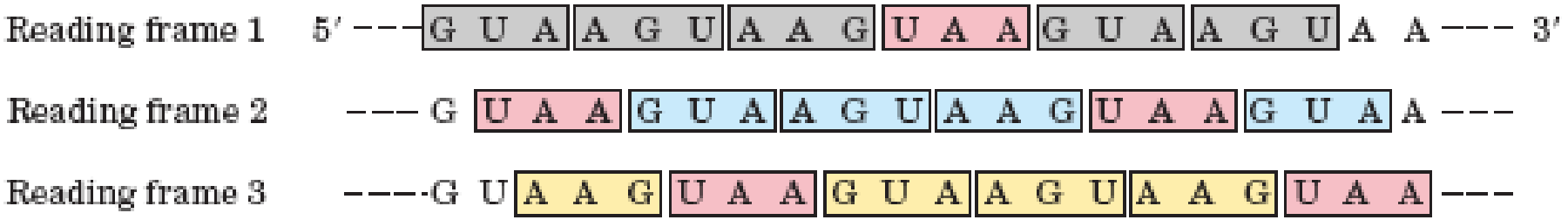
G	GUU, GUC, GUA, GUG } Val	GCU, GCC, GCA, GCG } Ala	GAU, GAC } Asp; GAA, GAG } Glu	GGU, GGC, GGA, GGG } Gly	U C A G

C	CUU, CUC, CUA, CUG } Leu	CCU, CCC, CCA, CCG } Pro	CAU, CAC } His; CAA, CAG } Gln	CGU, CGC, CGA, CGG } Arg	U C A G

Reading Frames







The Triplet Non-overlapping Code