



Lecture 14:

The genetic code

Course 281

Lessons for life

Lack of direction,
not lack of time, is
the problem. We all
have twenty-four hour
days.

Zig Ziglar

meetville.com

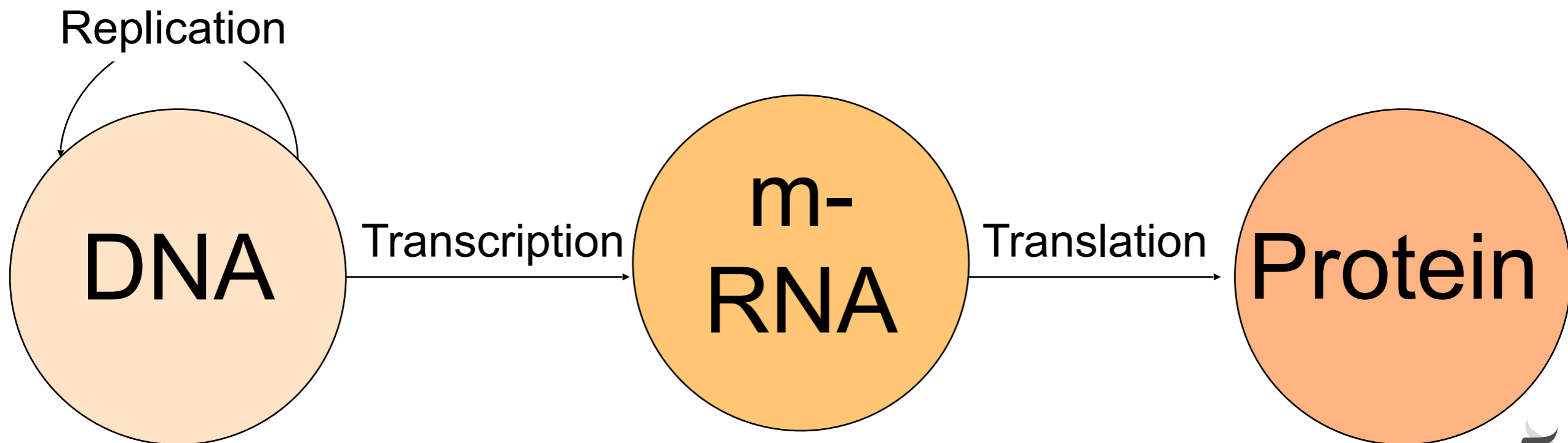
AIMS

- Understand the genetic code and how it was decoded.
- Understand the codons and what do they code for.
- Understand the general characteristics of the genetic code.

Gene expression

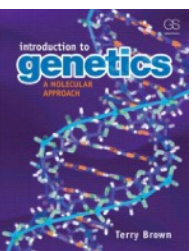
- Translating a protein coding gene is called **gene expression**.
- The path from genes to proteins go through an intermediate molecule called **m-RNA**.

What molecule gets translated into a protein?



The genetic code

- The rules that determine the order in which amino acids are joined together during translation of an mRNA.
- These rules are called the genetic code



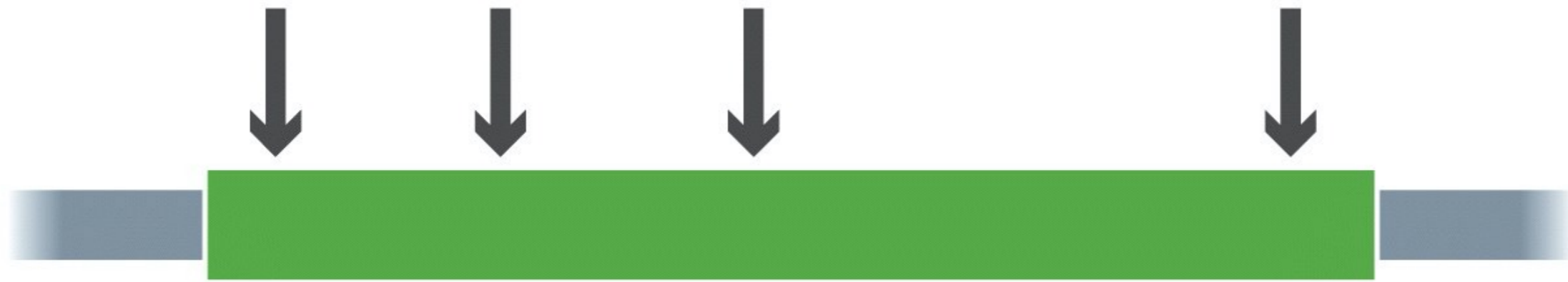
- There is a colinear relationship between a gene and its protein.
- The order of nucleotides in the gene correlates directly with the order of amino acids in the corresponding protein



Figure 7.13 Introduction to Genetics (© Garland Science 2012)

- In *E. coli* the gene coding for one of the two subunits—subunit A—of the enzyme tryptophan synthetase was changed
- A change in the subunit A nucleotide sequence gives rise to an amino acid alteration at the equivalent position in the subunit A protein.
- Subunit A gene is colinear with the subunit A protein
- With the amino terminus of the protein corresponding to the 5' end of the gene.

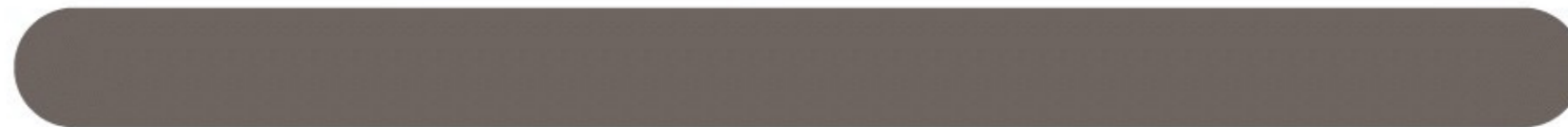
positions of nucleotide sequence alterations



gene

protein

? ? ? ?



where are the amino acid changes?

Figure. An experimental strategy for testing whether a gene is colinear with its protein

What about introns?

- The presence of introns means that a discontinuous eukaryotic gene is not colinear with its protein.
- The relationship is only linear, because two nucleotide changes on either side of an intron, will result in amino acid alterations that are much closer in the corresponding protein

positions of nucleotide sequence alterations

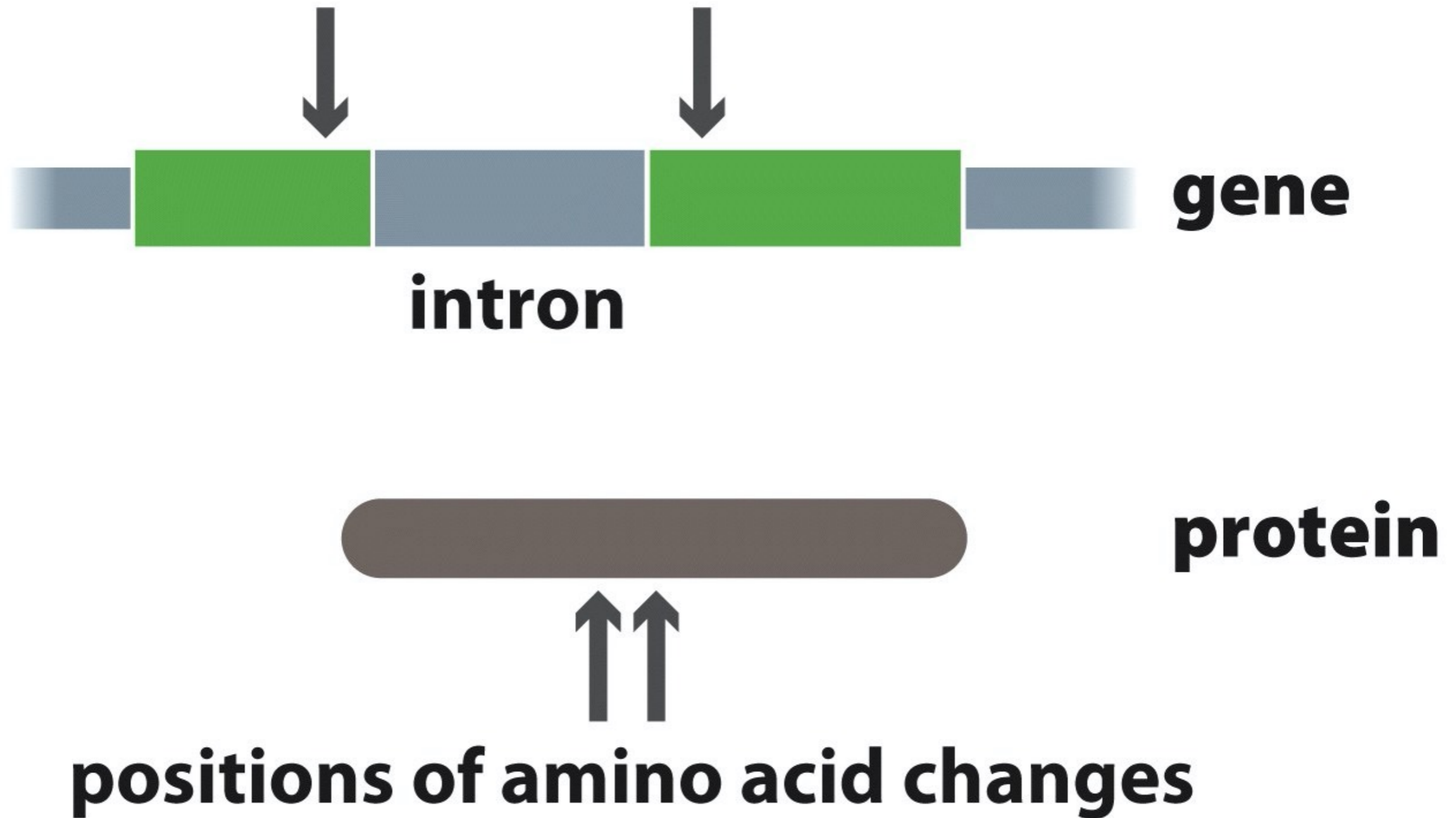
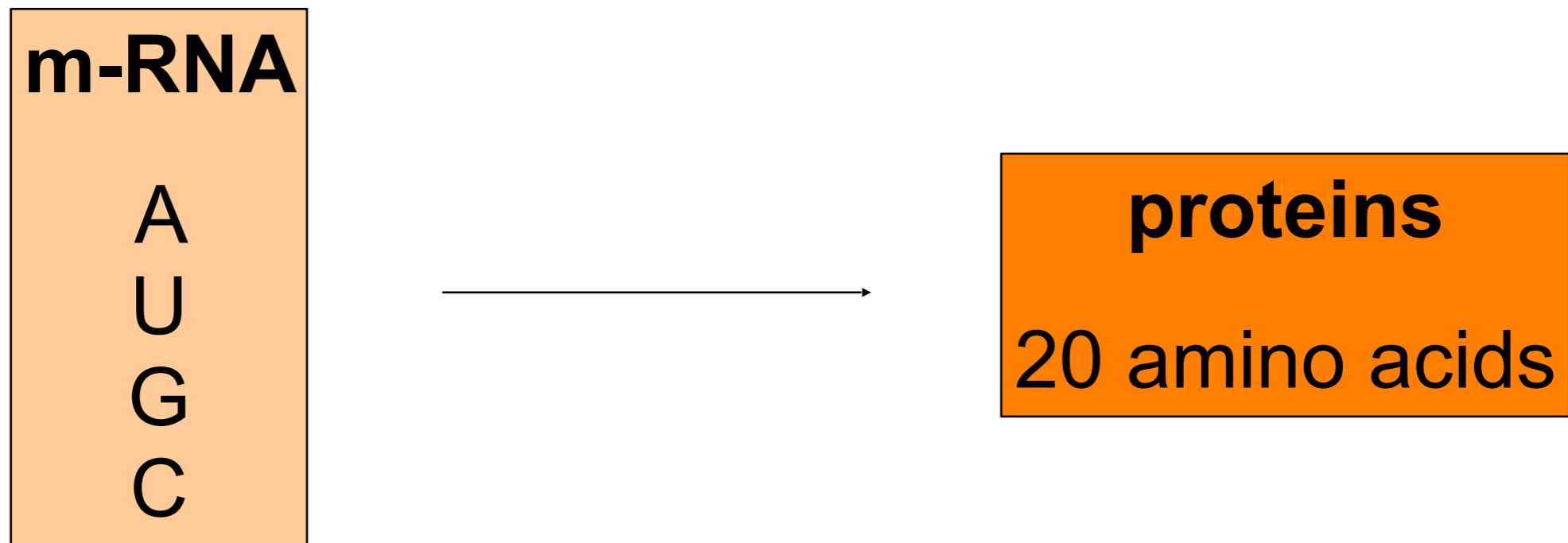


Figure 7.15 Introduction to Genetics (© Garland Science 2012)

The genetic code

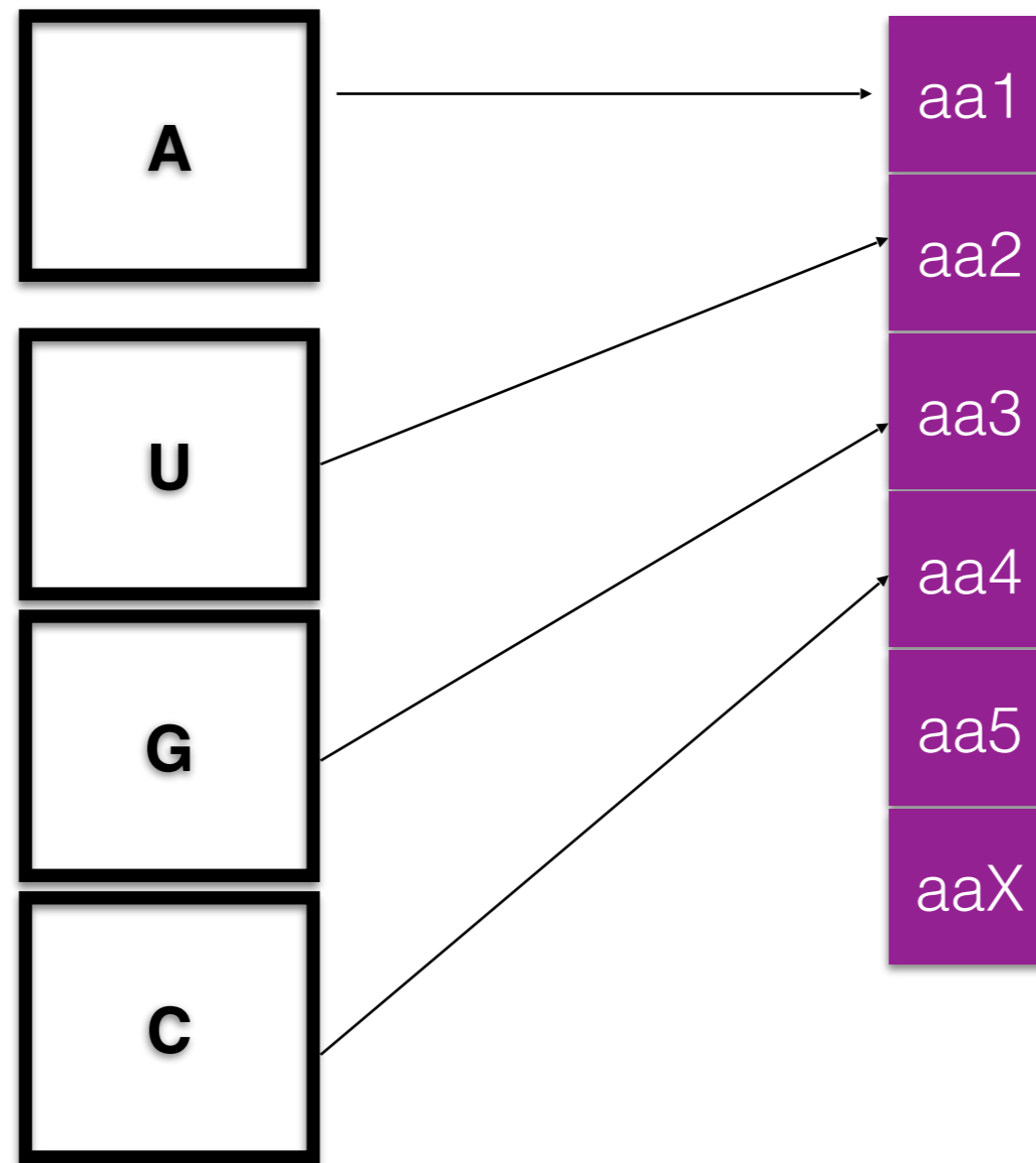
How do we get from mRNA → protein?



How do we get from 4 → 20?

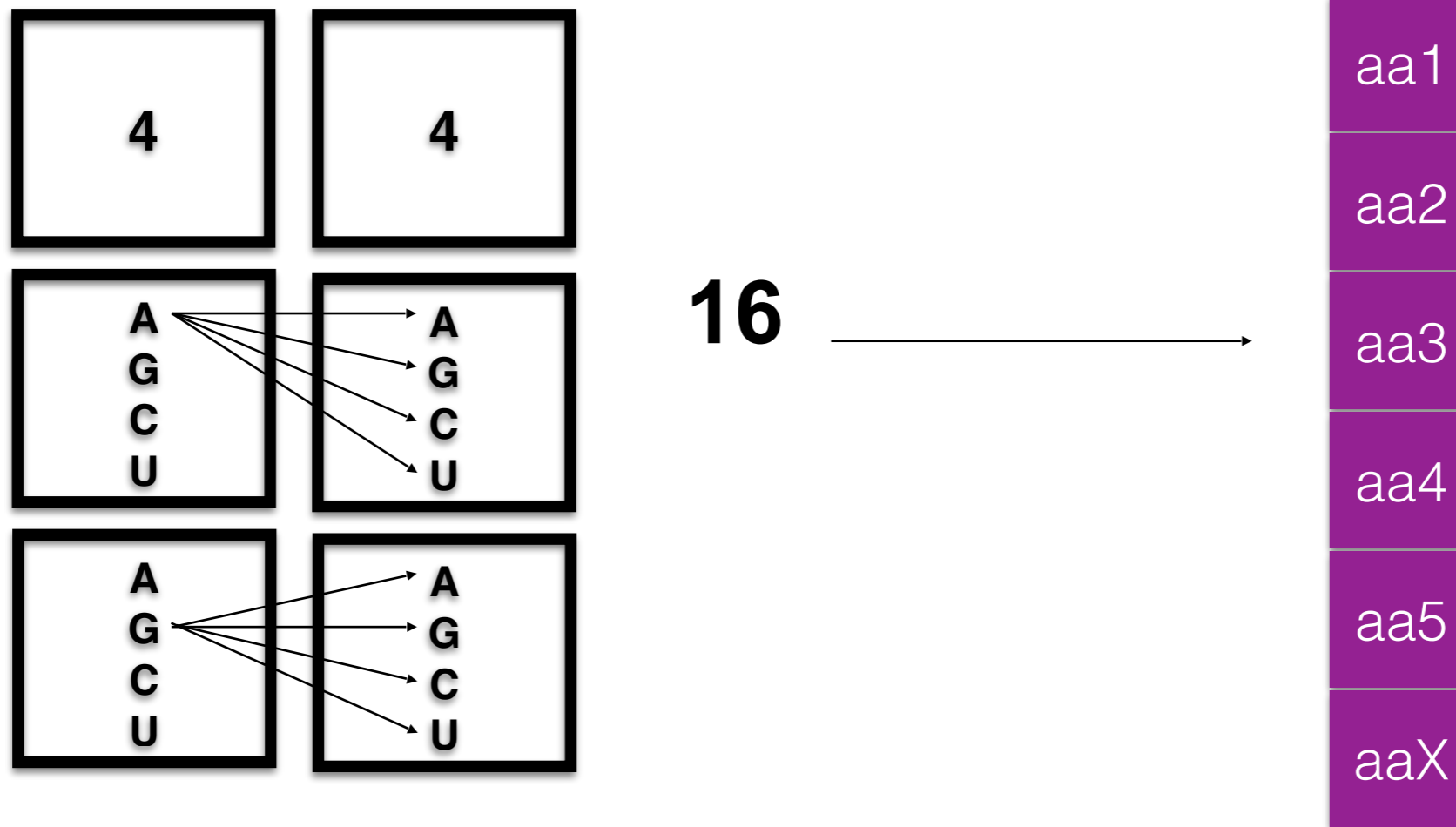
The genetic code

How do we get from mRNA → protein?



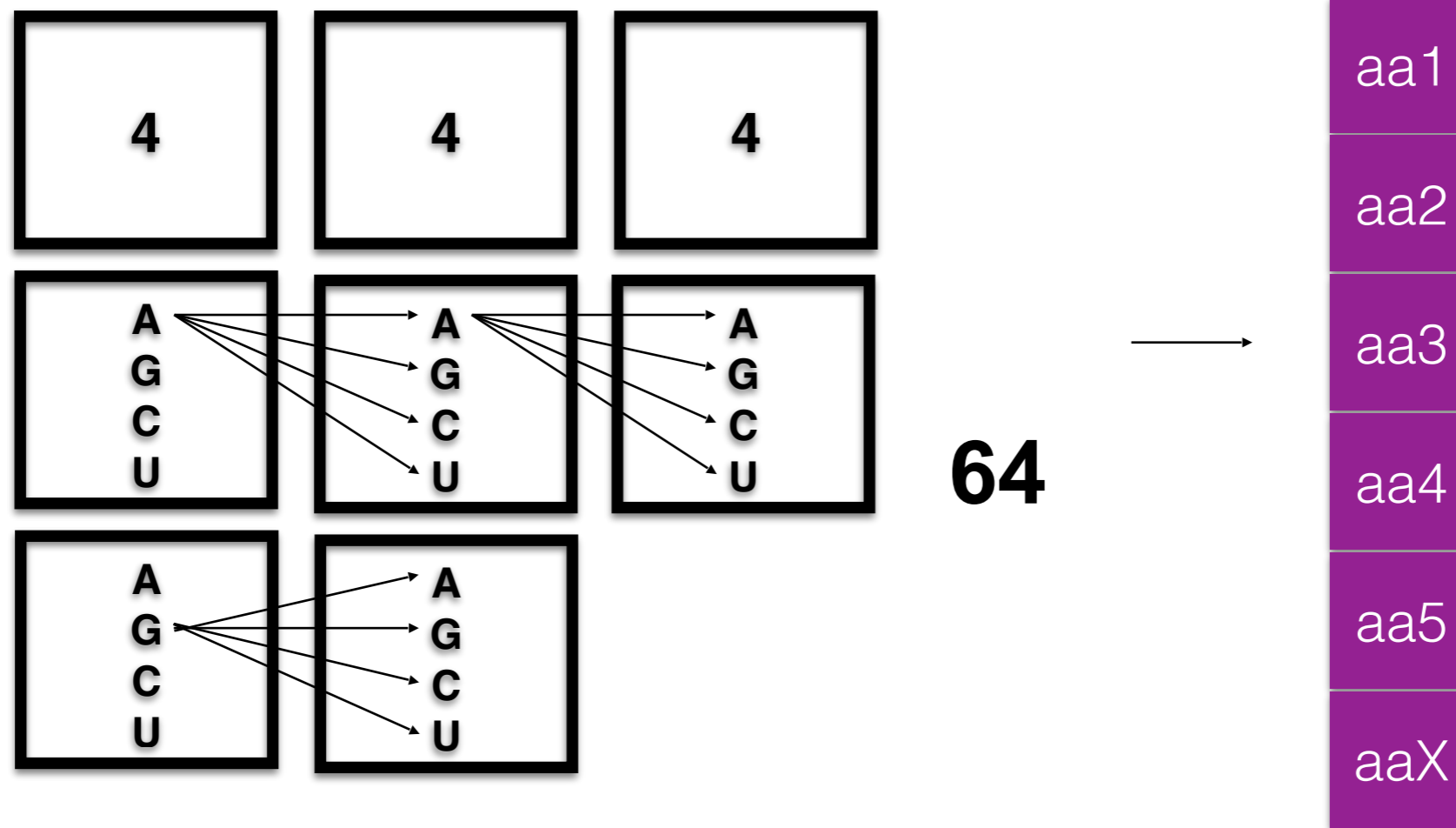
The genetic code

How do we get from mRNA → protein?



The genetic code

How do we get from mRNA → protein?



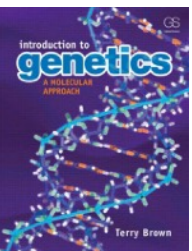
A mind experiment

- Each nucleotide codes for one amino acid.
Does not work ($4 \neq 20$)
- Each 2 nucleotide codes for one amino acid.
How many combinations of 2 nucleotides?
 $4 \times 4 = 16$ combinations
Does not work ($16 \neq 20$)
- Each three nucleotides codes for one amino acid.
How many combinations of 3 nucleotides?
 $4 \times 4 \times 4 = 64$ combinations
Can work ($64 > 20$)

Codon

- The group of nucleotides that code for a single amino acid
- Codons cannot be just single nucleotides (A, T, G, or C) because 20 amino acids found in proteins.
- A doublet code (codons such as AT, TA, TT, GC, etc.) seems unlikely as this would contain only $4^2 = 16$ different codons

- A triplet code (codons AAA, AAT, TAT, GCA, etc.), would be feasible as this would yield $4^3 = 64$ code words.
- Which would be more than enough.



A mind experiment

A code of three nucleotides coding for a single amino acid creates more than needed!

The genetic code

- The genetic code is made of triplets (3) nucleotides.
- **Codon:** three nucleotides in a m-RNA coding for a single specific amino acid.

How this was found?

The genetic code

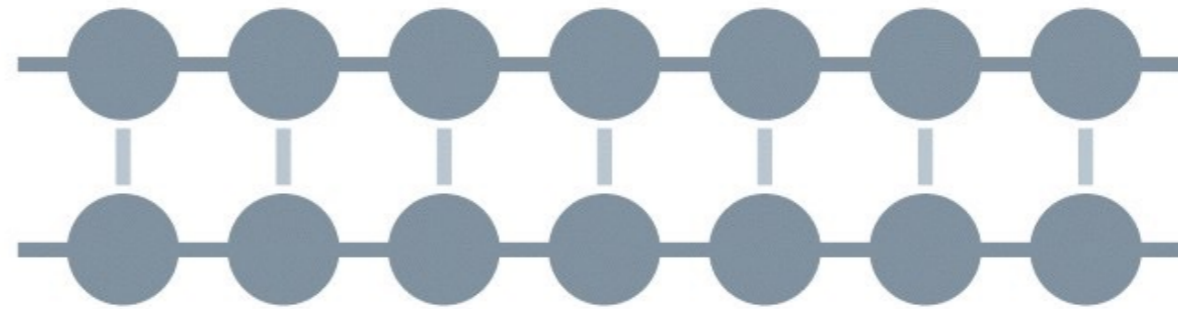
- Mutation experiments proved that only removal or addition of nucleotides by multiple of three can result in a functional protein.

	MUTATION	PHENOTYPE
Wild-type sequence ABCABCABCABCABCABCABCABC	NONE	rII ⁺
FC0 mutant ABCAABCABCABCABCABCABCABC	+	rII ⁻
Supression of FC0 ABCAABABCABCABCABCABCABC	+ -	rII ⁺
Two base additions ABCAABCABCBAABCABCABCABC	+ +	rII ⁻
Three base additions ABCAABCABCBAABCABCABCABCABC	+ + +	rII ⁺

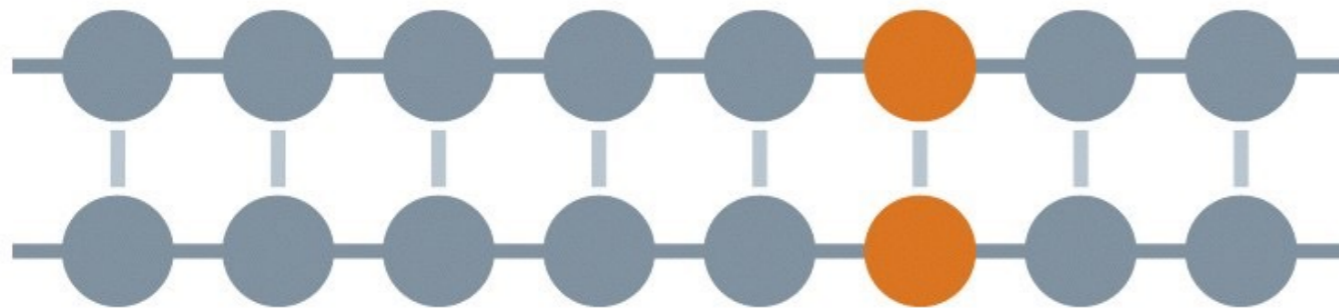
+ Base addition
- Base deletion

- The hypothesis that codons are triplets of nucleotides as first tested by experiments with *E. coli*
- These experiments made use of **proflavin**
 - one of the **acridine** dyes, a group of chemicals that cause base-pair deletions or additions in double-stranded DNA molecules

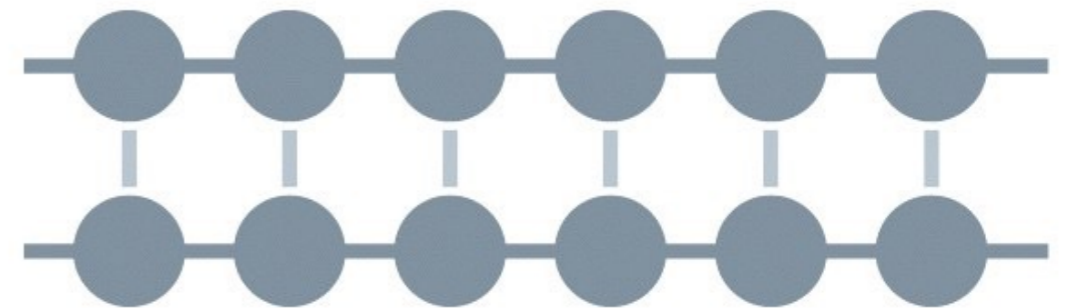
DNA molecule



proflavin



insertion



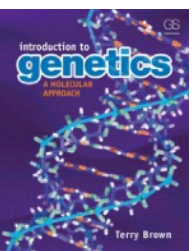
deletion

Figure. An acridine dye such as proflavin can cause an insertion or a deletion in a double-stranded DNA molecule.

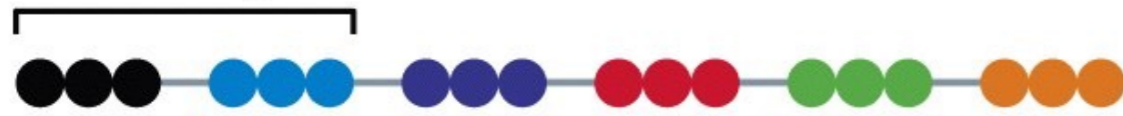
The rationale

- Some proteins contain segments where the amino acid sequence can be changed without altering the function of the protein.
- What if a series of insertions and/or deletions are introduced into the region of a gene coding for one of these tolerant segments of a protein?

- If each code word is a triplet of nucleotides.
- Then a single insertion or deletion would give rise to a nonfunctional protein.
- Because all the codewords downstream of the insertion or deletion would be altered.
- Including those in the nontolerant segment following the tolerant region.



tolerant segment



normal sequence of codons



insert one nucleotide –
all the following codons are changed



insert two nucleotides –
all the following codons are changed



insert three nucleotides –
reading frame is maintained

Figure. The rationale behind an experiment to test whether codons are triplets of nucleotides.

- Two insertions or deletions (although not one of each) would have the same effect.
- But three insertions or deletions in the tolerant region would maintain the correct **reading frame** in the nontolerant region.
- Would be predicted to have no effect on the function of the protein.

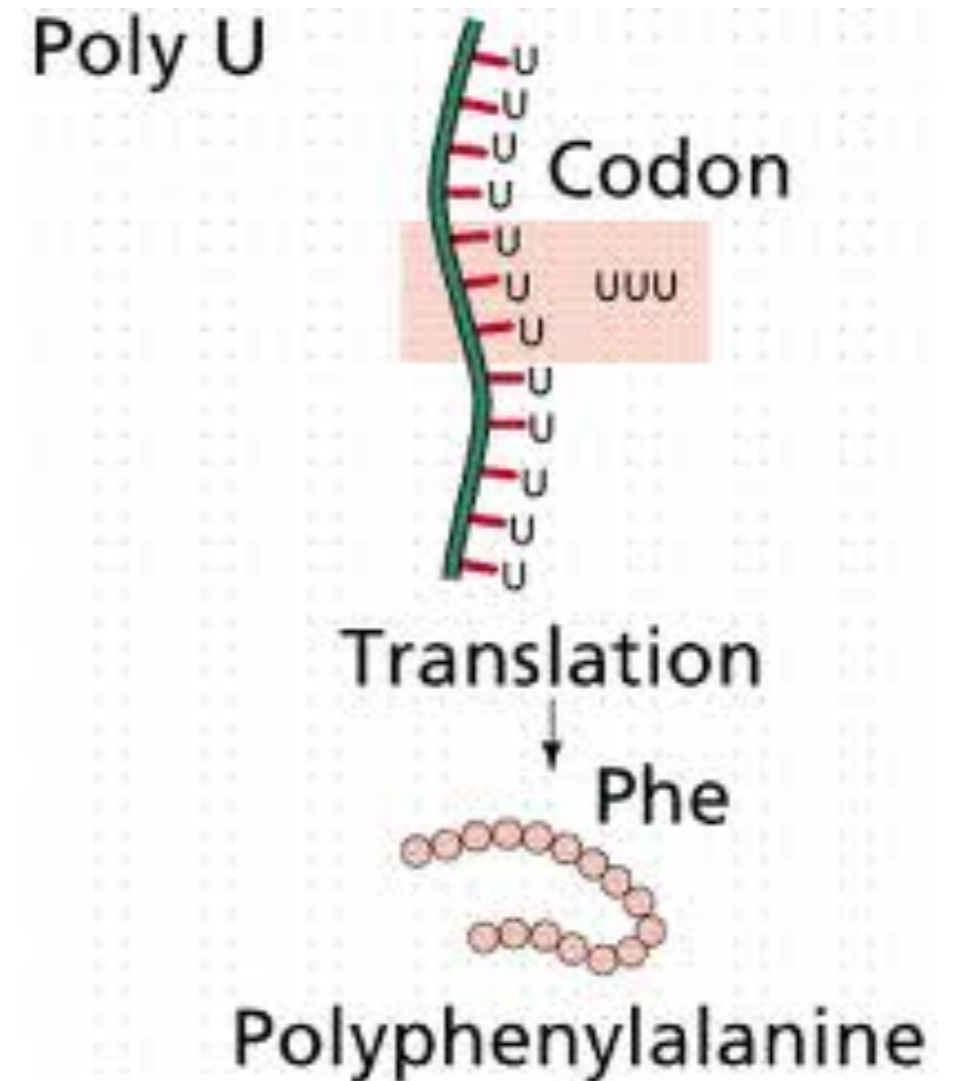
- An elegant experiment of this kind was first carried out successfully with a gene from the *E. coli* bacteriophage called T4.
- This work established the triplet nature of the code.

What codons code for what amino acid?

First experiment: using mononucleotide polymers as the mRNA.

Poly(U) mRNA gives poly phenylalanine amino acids. Thus **UUU** codes for **phenylalanine**.

Can we do the same for the other three nucleotides?



What codons code for what amino acid?

Poly(A) mRNA gives poly lysine amino acids. Thus **AAA** codes for **lysine**.

Poly(C) mRNA gives poly proline amino acids. Thus **CCC** codes for **proline**.

Poly(G) could not be done for structural difficulties.

What codons code for what amino acid?

Second experiment: using random copolymer mRNA of two different nucleotides.

Make a copolymer of (A and C).

What are the outcomes?

What codons code for what amino acid?

- AAA (we already know)
- CCC (we already know)

- CCA
- CAC
- AAC
- CAA
- ACC
- ACA

**What do they
code for?**

Asparagine
Glutamine
Histidine
Threonine

How Do we know?

What codons code for what amino acid?

Second experiment: using random copolymer mRNA of two different nucleotides.

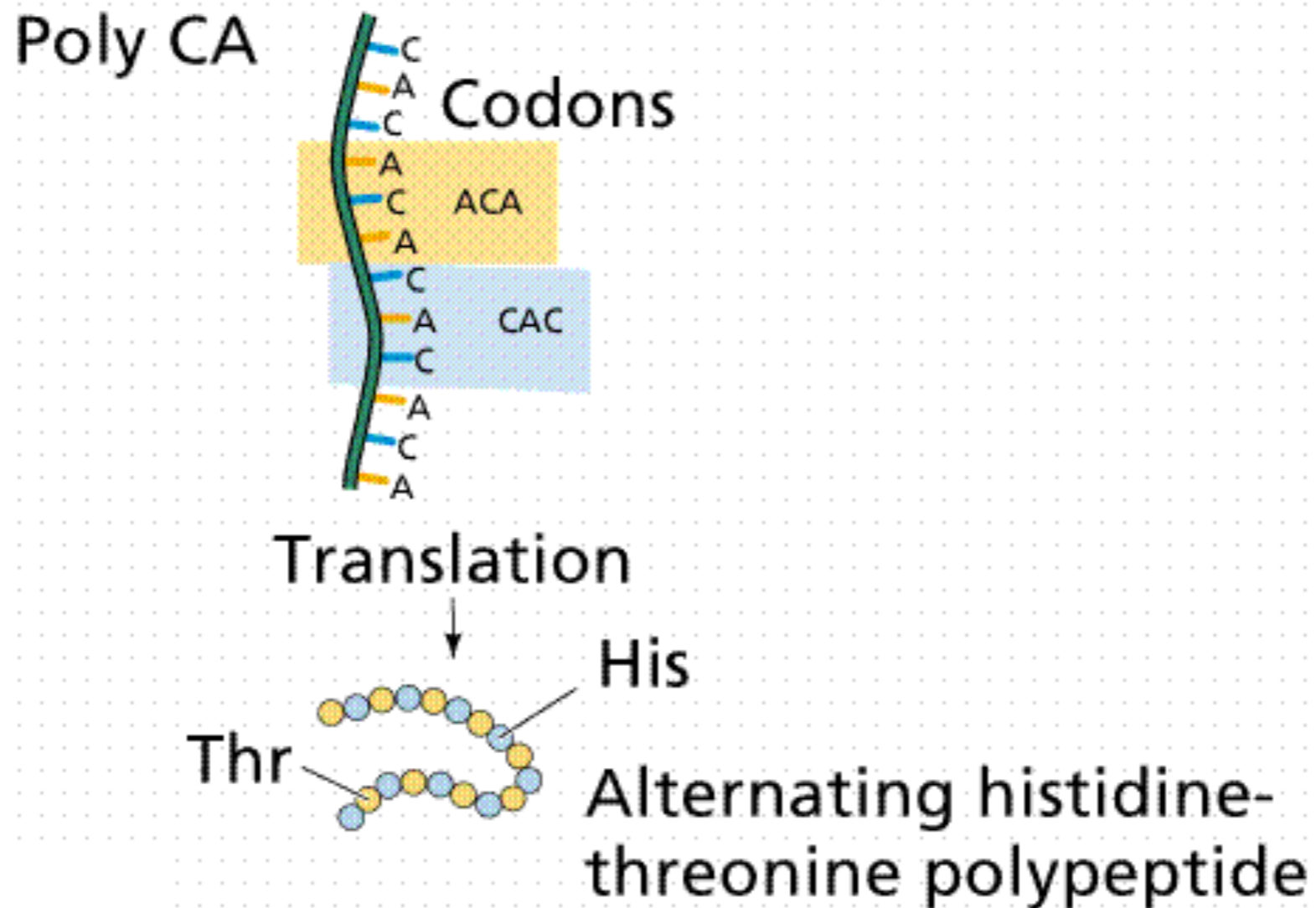
- (1) Play with the ratio (add more A than C)
- (2) Get more Asparagine than histidine
- (3) Thus Asparagine must be coded by 2As and histidine by 2Cs

This experiment tells us about the composition of the codon rather than the sequence of the codon!

What codons code for what amino acid?

Result: UCU and CUC code for leucine and serine

But can not tell which is which!



What codons code for what amino acid?

Fourth experiment: using the translation process to determine the code.

The approach used was called “**ribosome binding assay**”

The experiment determined the specific sequence of the codons.

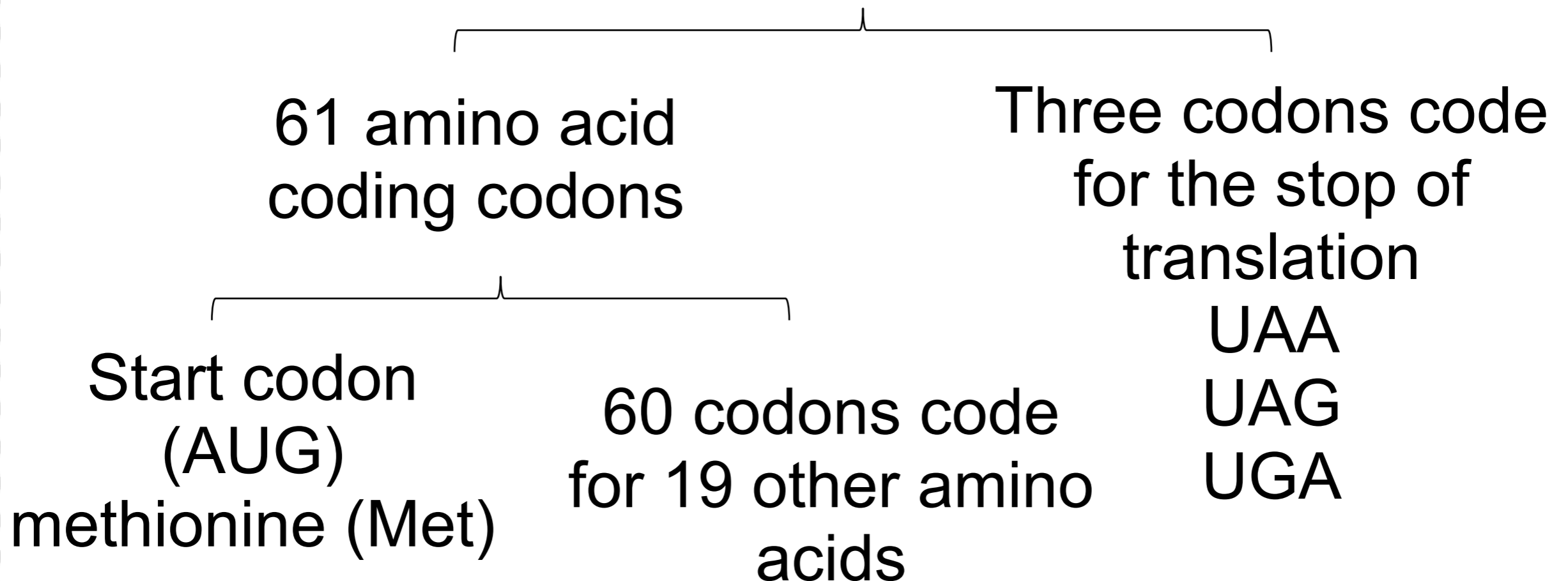
What codons code for what amino acid?

When decoding the mRNA codons, 1 amino acid goes to the ribosome and binds (tRNA).

This approach determined the sequence of the majority of the codons.

The genetic code

The genetic code is composed of 64 codons



The genetic code

**The codons are more than what we need to
translate the 20 amino acids**

We will learn how and why later!

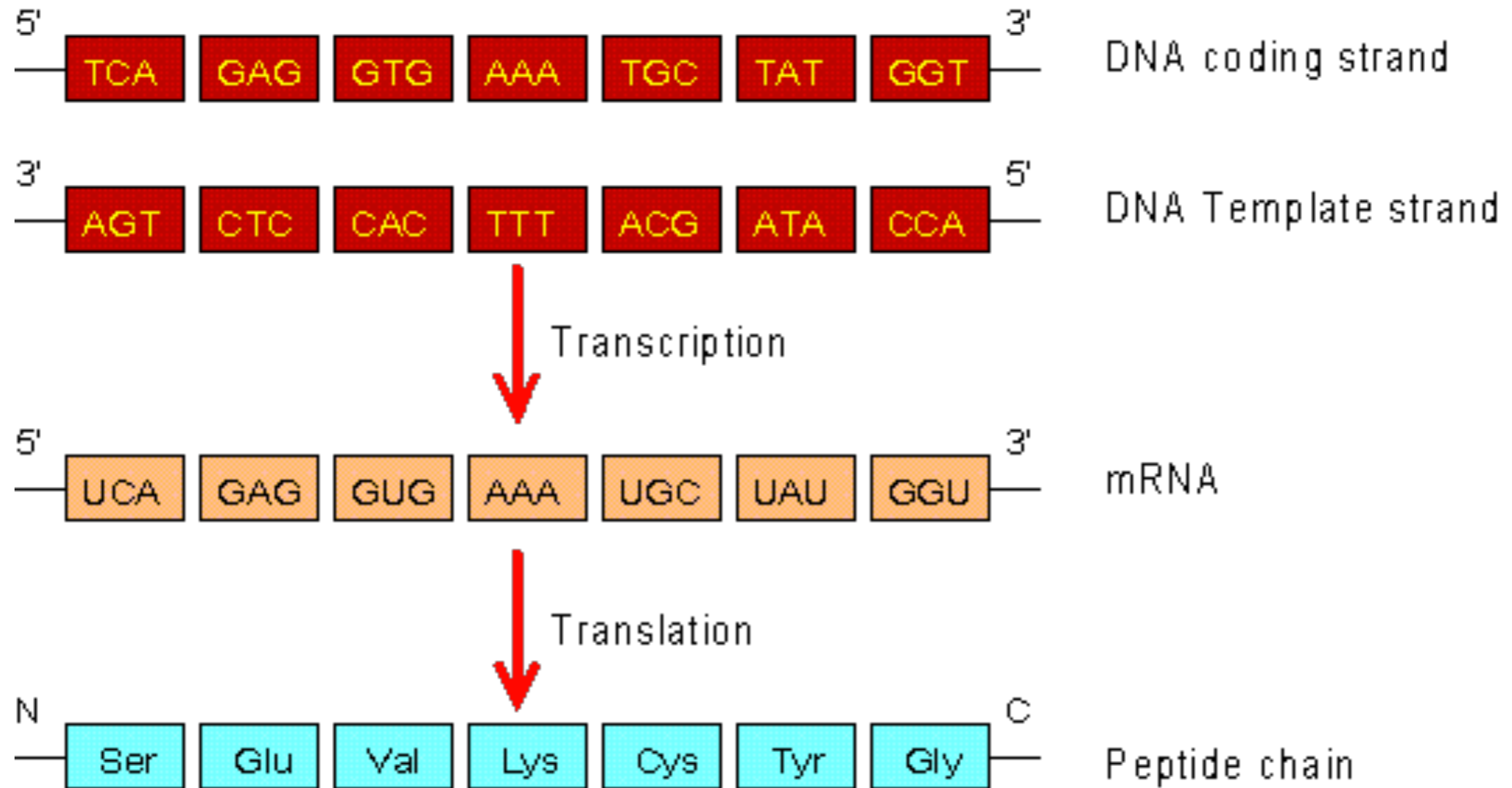
Characteristics of the genetic code

- The genetic code is made of triplets of nucleotides (3nts) called codons.

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

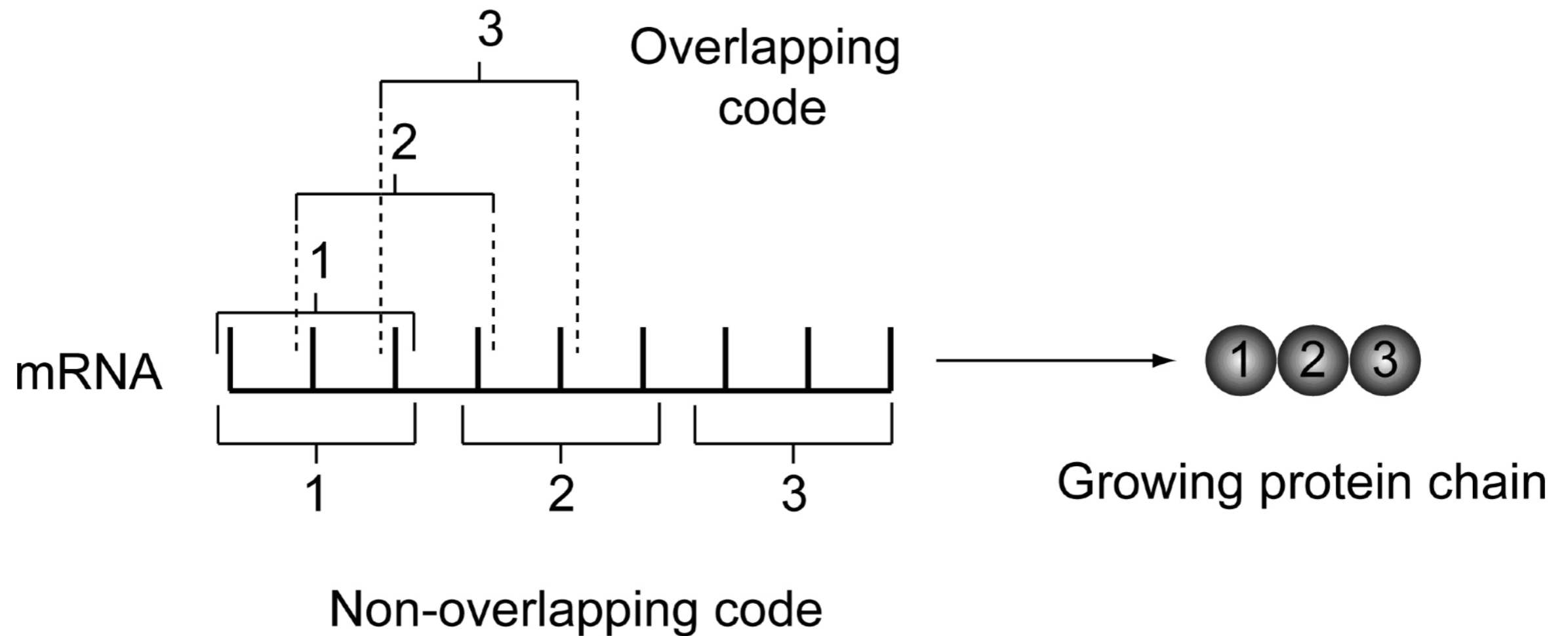
Characteristics of the genetic code

2. The genetic code is continuous (no skipping)



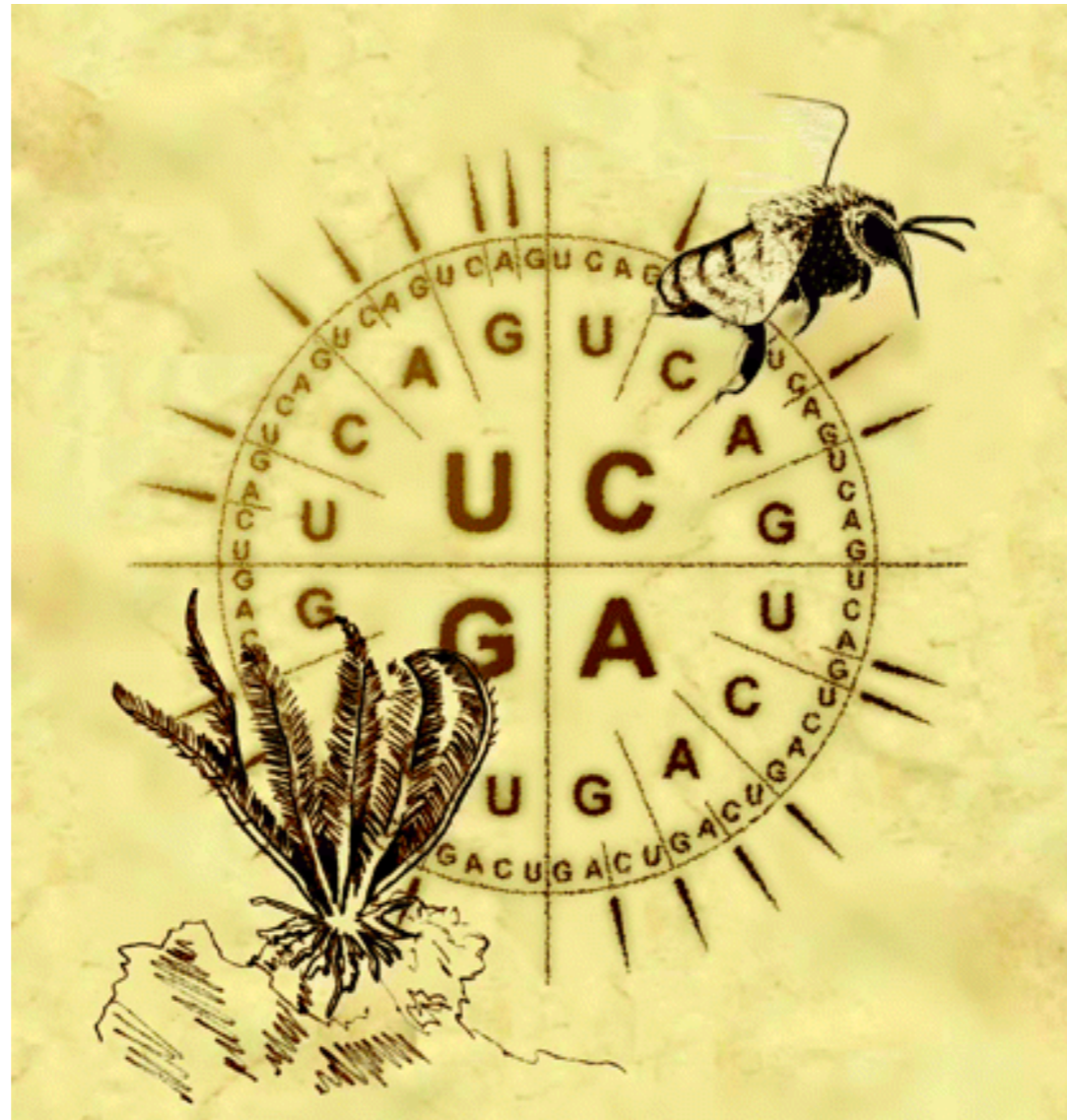
Characteristics of the genetic code

- The code is not overlapping. Every three nucleotides in a sequence code for one codon.



Characteristics of the genetic code

4. The genetic code is universal (almost).
All living organisms have the same code and the system of the code.



The genetic code is not universal

- The code holds for the vast majority of genes in the vast majority of organisms.
- But deviations are widespread.
- Genes present in human and plant mitochondrial DNA often use a nonstandard code.

TABLE 7.2 NONSTANDARD CODONS IN HUMAN MITOCHONDRIAL DNA

Codon	Should code for	Actually codes for
UGA	Stop	Tryptophan
AGA, AGG	Arginine	Stop
AUA	Isoleucine	Methionine

- Nonstandard codes are also known in the nuclear genes of lower eukaryotes.
- Often a modification is restricted to just a small group of organisms.
- Frequently it involves reassignment of the termination codons.
- Modifications are less common among bacteria.

TABLE 7.3 EXAMPLES OF NONSTANDARD CODONS IN NUCLEAR GENOMES

Organism	Codon	Should code for	Actually codes for
Several protozoa	UAA, UAG	Stop	Glutamine
<i>Candida rugosa</i> (yeast)	CUG	Leucine	Serine
<i>Euplotes</i> spp. (ciliated protozoan)	UGA	Stop	Cysteine

Table 7.3 Introduction to Genetics (© Garland Science 2012)

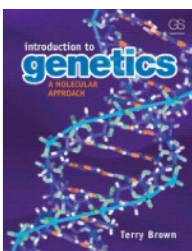
context-dependent codon reassignment

- A second type of code variation.
- Occurs when the protein to be synthesized contains either selenocysteine or pyrrolysine.
- Proteins containing pyrrolysine are rare and are probably present only in some archaea and a very small number of bacteria.
- But proteins containing selenocysteine are widespread in many organisms.

TABLE 7.4 EXAMPLES OF PROTEINS THAT CONTAIN SELENOCYSTEINE

Organisms	Protein
Mammals	Glutathione peroxidase
	Thioredoxin reductase
	Iodothyronine deiodinase
Bacteria	Formate dehydrogenase
	Glycine reductase
	Proline reductase
Archaea	Hydrogenase
	Formyl-methanofuran dehydrogenase

Table 7.4 Introduction to Genetics (© Garland Science 2012)



The enzyme glutathione peroxidase

- Helps protect the cells of humans and other mammals against oxidative damage.
- Selenocysteine is coded by 5'-UGA-3', which therefore has a dual meaning because it is still used as a termination codon in the organisms concerned.

- A 5'-UGA-3' codon that specifies selenocysteine is distinguished from true termination codons by the presence of a hairpin loop structure in the mRNA.
- Positioned just downstream of the selenocysteine codon in prokaryotes and in the 3' untranslated region in eukaryotes.

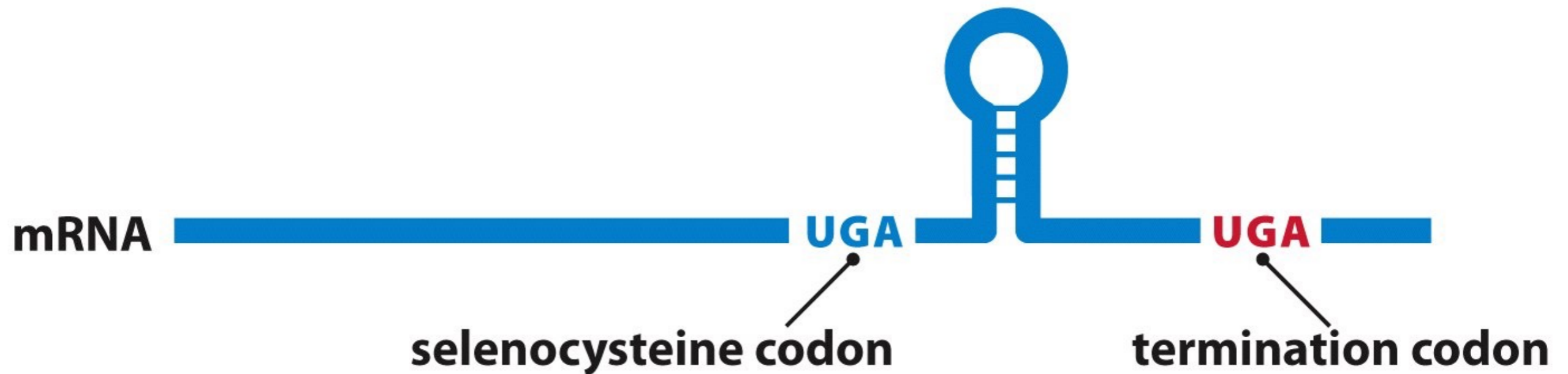


Figure. A 5'-UGA-3' codon that specifies selenocysteine is distinguished from a termination codon by the presence of a hairpin loop in the mRNA.

- Recognition of the selenocysteine codon requires interaction between the hairpin and a special protein that is involved in translation of these mRNAs.
- A similar system probably operates with pyrrolysine
 - which is specified by a second termination codon, 5'-UAG-3'.

Characteristics of the genetic code

5. The code has specific signals for start of translation and stop of translation.

The start codon (AUG) codes for a methionine amino acid.

Three stop codons (UAA, UAG, UGA) code for a stop **WITHOUT** amino acid.

The stop codons are also called **nonsense codons**, or **chain termination codons**.

Characteristics of the genetic code

6. The genetic code is “degenerate”.

Degenerate means redundant.

Remember 61 codons code for 20 amino acids

More than one codon for the same one amino acid

Characteristics of the genetic code



أَسَدٌ

حَيْدَرٌ

أَسَامَةٌ

ثَبِيبٌ

لَيْثٌ

غَضَنَفَرٌ

بَاسِلٌ

سَبِيعٌ

حَمْرَةٌ

ضَرْغَامٌ



Characteristics of the genetic code



أَسَد

حَيْدَر

أَسَامَة

ثَبِيبُ

لَيْث

غَضَنَفَر

بَاسِل

سَبْع

حَمْرَة

ضَرْغَام



هل من الممكن الإشارة للئيس مثلاً بأي من هذه الأسماء؟

Characteristics of the genetic code



أسد

حيدر

أسامة

ثبيل

ليث

غضنفر

باسيل

سبع

حمرّة

ضرغام



لا تستخدم الاسماء هذه إلا للإشارة لل
(lion) بينما ال (lion) يشار له بعدة أسماء
يختص بها عن غيره

Characteristics of the genetic code

Remember:

Each codon codes for one amino acid

BUT

**An amino acid can be coded by more than one
codon**

Characteristics of the genetic code

- The genetic code is degenerate
 - some amino acids are specified by more than one codon
 - All amino acids except methionine and tryptophan are specified by more than one codon
- The code will include **punctuation codons**
 - Special codons that indicate the **start** and **end** of the nucleotide sequence that must be translated into protein

DNA

gene



mRNA



initiation codon

termination codon



Figure. The positions of the punctuation codons in an mRNA.

- The work that led to elucidation of the genetic code was carried out during 1961–1966
- The first codon to be assigned was 5'-UUU-3', which specifies phenylalanine
- This was worked out by showing that poly(U), an RNA molecule that contains just U nucleotides, directs synthesis of polyphenylalanine
- Equivalent experiments enabled 5'-AAA-3' to be assigned to lysine and 5'-CCC-3' to proline

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

Figure. The genetic code.

- Most synonymous codons are grouped into families
- This similarity between synonymous codons is relevant to the way the code is deciphered during protein synthesis

The code include **punctuation codons**

- Special codons that indicate the **start** and **end** of the nucleotide sequence that must be translated into protein

The initiation codon: 5'-AUG-3'

- occurs at the start of most genes and marks the position where translation should begin
- Most newly synthesized polypeptides have this amino acid at the amino terminus
- Methionine may subsequently be removed after the protein has been made
- AUGs that are not initiation codons may be found in the internal region of a gene

With a few genes

- A different triplet such as 5'-GUG-3' or 5'-UUG-3' is used as the initiation codon.

Termination codons

- Three triplets:
 - 5'-UAA-3'
 - 5'-UAG-3'
 - 5'-UGA-3'
- They do not code for amino acids
- One of these three always occurs at the end of a gene at the point where translation must stop

Characteristics of the genetic code

7. The Wobble effect of the third base in the codon

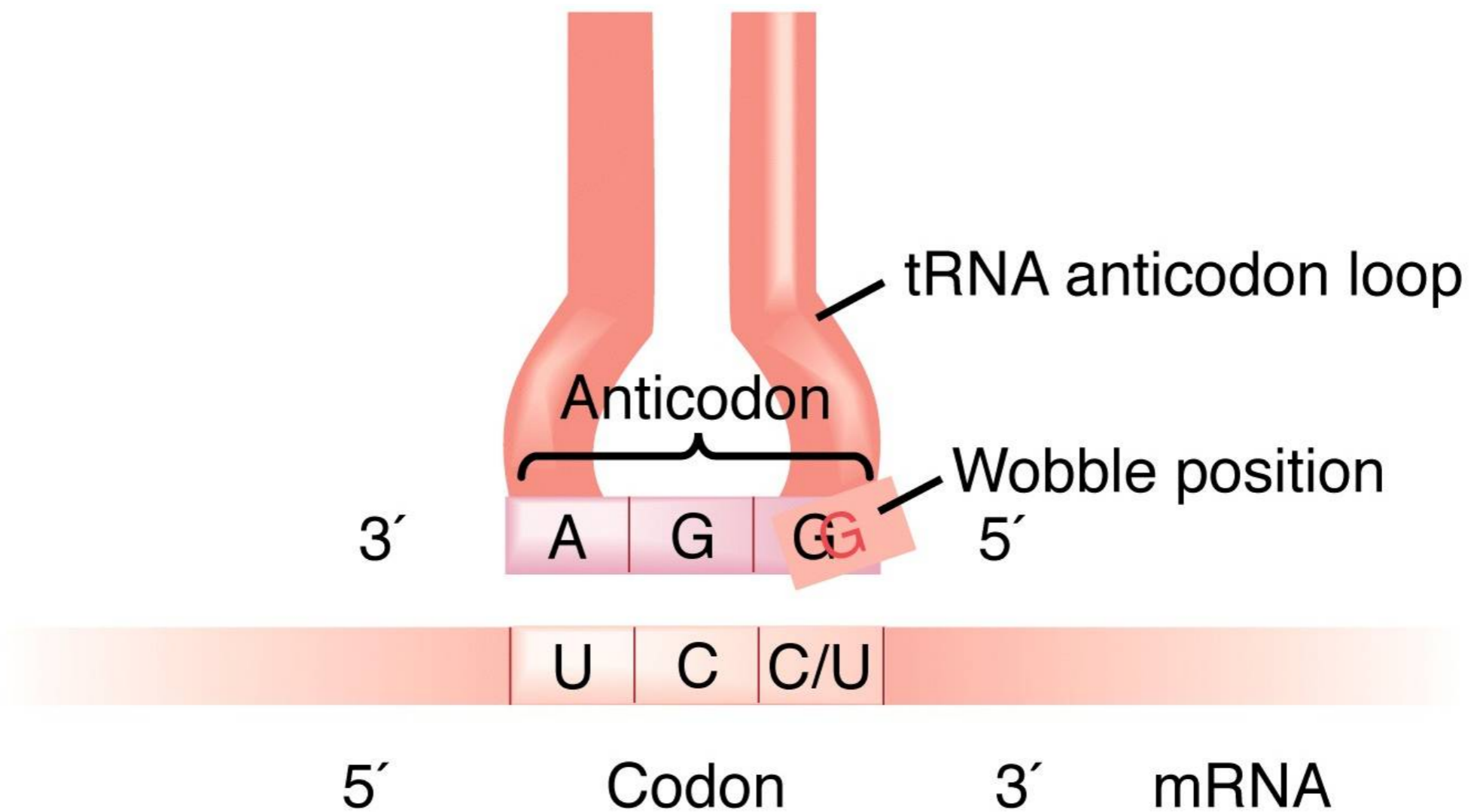
The third nucleotides in some codons are not essential for determining the identity of the amino acid.

Characteristics of the genetic code

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

Characteristics of the genetic code

This is due to the base pairing between the codon in the mRNA and the anti-codon in the tRNA during the translation process.



The wobble hypothesis

- 61 different types of tRNA molecule in each cell
- One for each of the codons that specify an amino acid
- However, there are substantially fewer than 61 different tRNA molecules
- Usually between 30 and 50 depending on the organism

The wobble hypothesis

- The anticodon loop and the anticodon itself is not a perfectly linear trinucleotide
- The short double helix formed by base pairing between the codon and anticodon do not have the precise configuration of a standard RNA helix
- Its dimensions are slightly altered
- Nonstandard base pairs can form at the “wobble position”
- Between the third nucleotide of the codon and the first nucleotide of the anticodon

- Because of the wobble pairing a single anticodon may be able to base-pair with more than one codon
- This means that a single tRNA might decode more than one member of a codon family
- However, the base-pairing rules do not become totally flexible at the wobble position
- Only a few types of unusual base pairs are allowed
- G–U base pairs are a common example

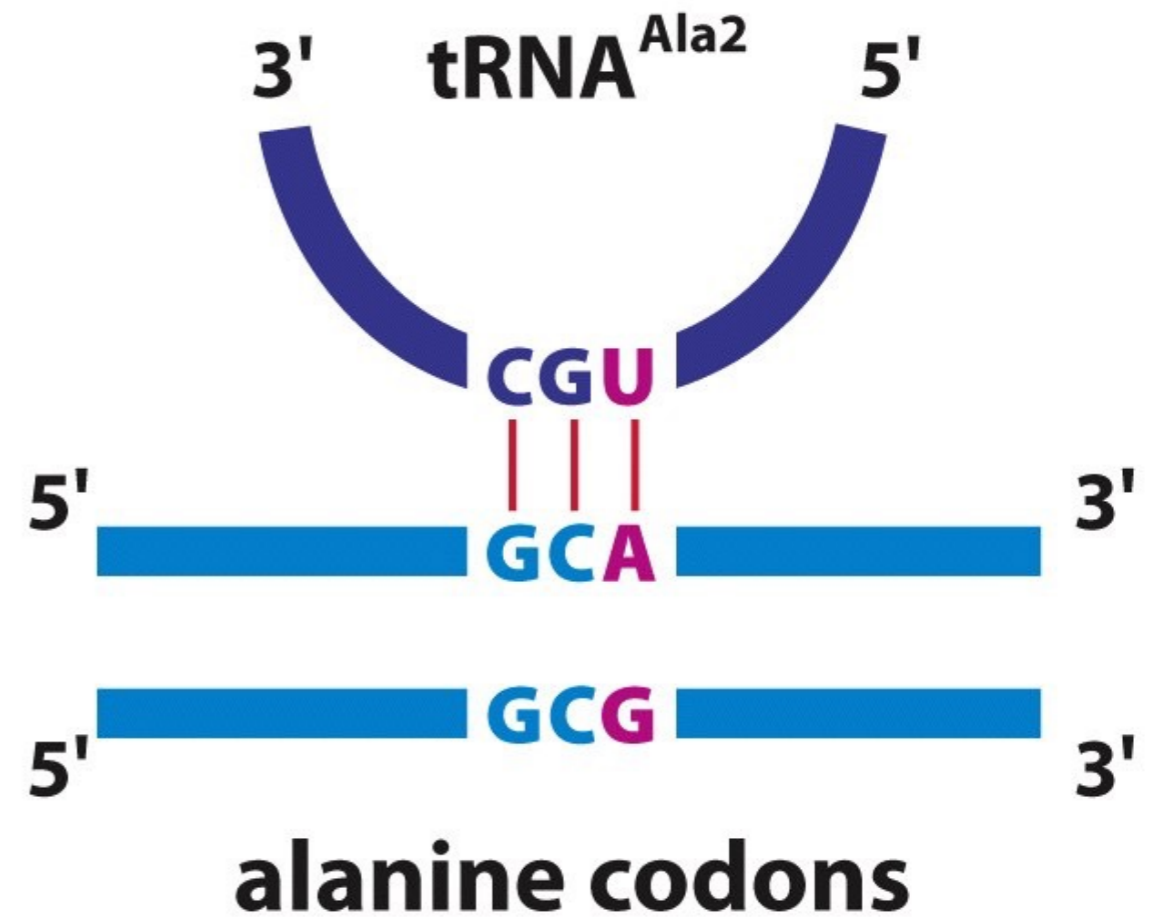
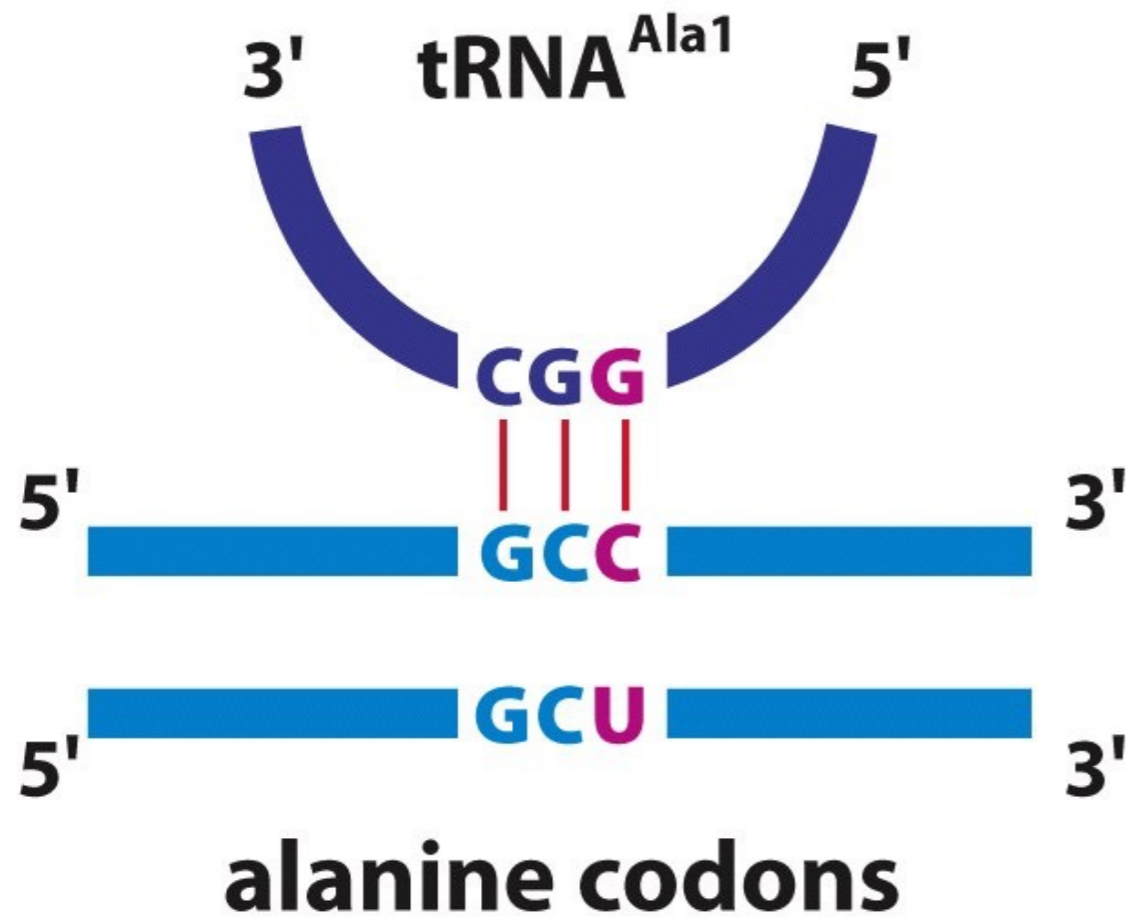
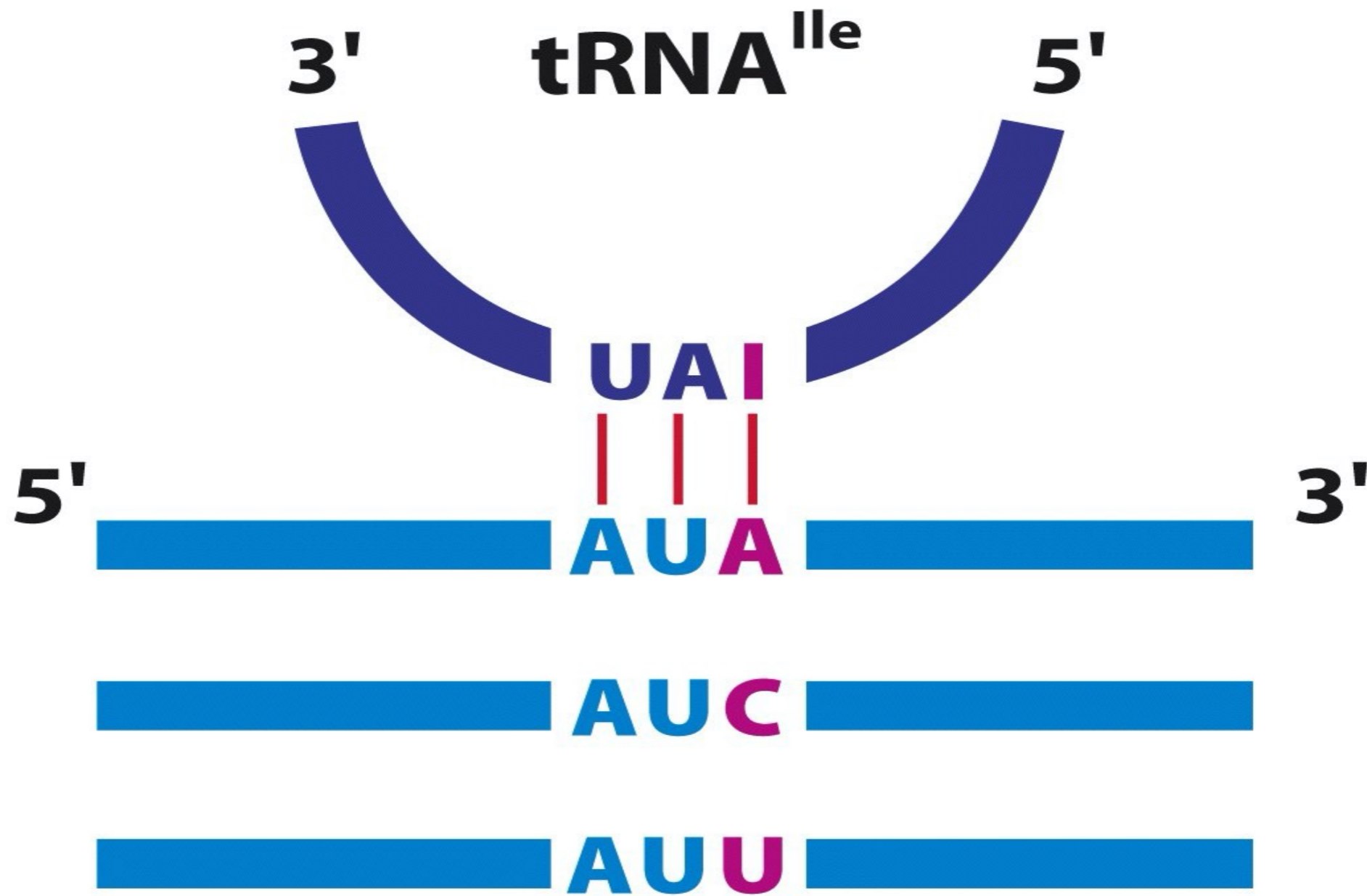


Figure. Wobble involving a G–U base pair

A second type of wobble involves inosine

- One of the modified nucleotides present in tRNA
- Inosine can base-pair with A, C, and U
- The triplet 3'-UAI-5' is sometimes used as the anticodon in a tRNA^{Ile} molecule
- Because it pairs with 5'-AUA-3', 5'-AUC-3', and 5'-AUU-3'
- These triplets form the three-codon family for isoleucine in the standard genetic code



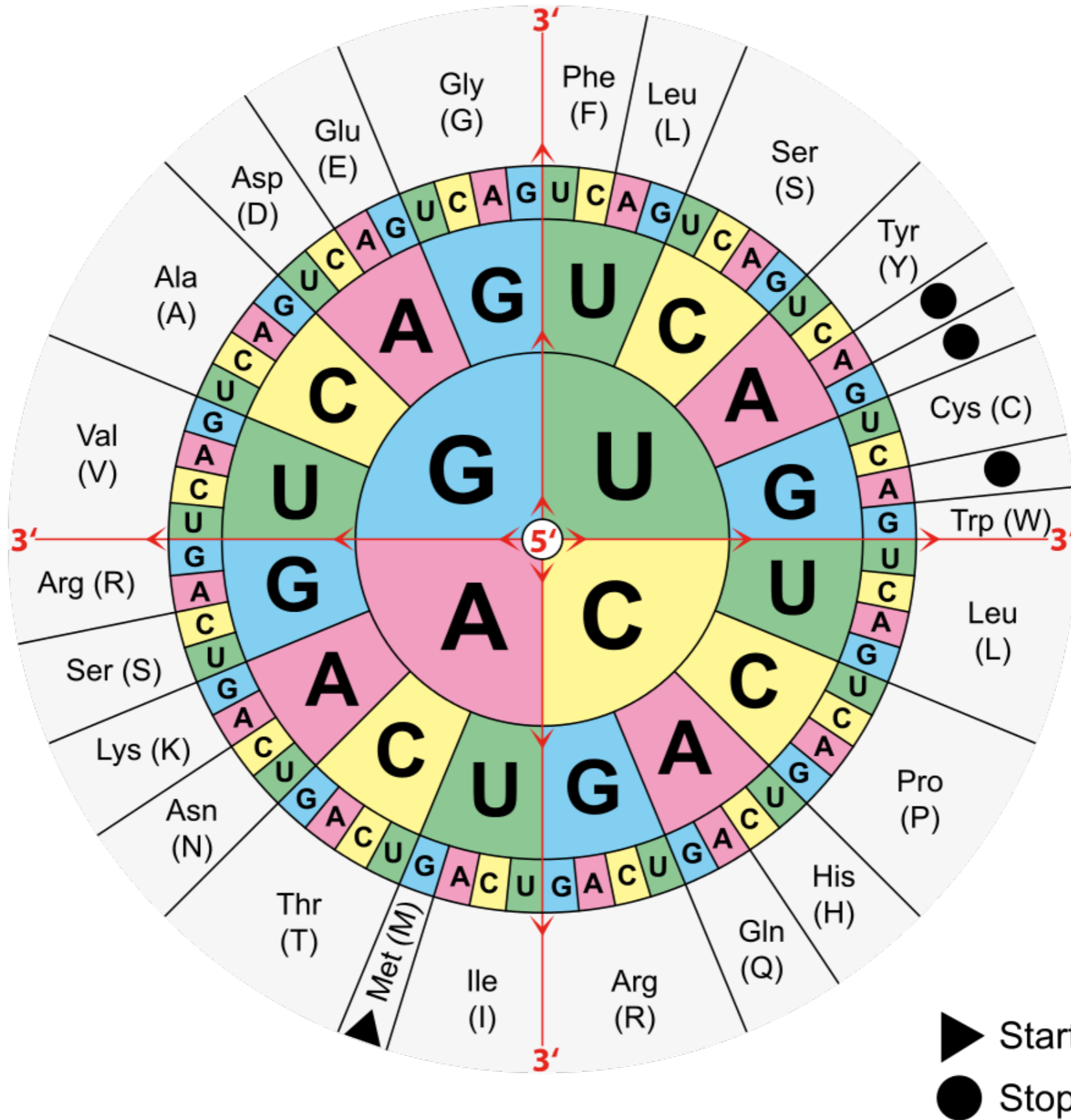
isoleucine codons

Figure. Wobble involving inosine

- Wobble reduces the number of tRNAs needed in a cell
- by enabling one tRNA to read two or possibly three codons
- Hence bacteria can decode their mRNAs with as few as 30 tRNAs.
- Eukaryotes also make use of wobble but in a restricted way

- The human genome has 48 tRNAs
- Of these, 16 are predicted to use wobble to decode two codons each
- The remaining 32 are specific for just a single triplet
- Wobble does not violate the rules of the genetic code
- The protein that is made during translation is always synthesized strictly in accordance with the nucleotide sequence of the relevant mRNA

The genetic code



Stuff to know



Chain termination codons

Ribosome binding assay

Start codon

UGA

copolymer

AUG

codon

Polar amino acids

Nonsense codons

Mononucleotide polymer

triplets

Stop codon

UAG

Random copolymer

UUA

Expectations

- You know how the mRNA carries the genetic code and how the sequence is mean to be read.
- You understand the experiments that lead to the discovery of the genetic code.
- You know the characteristics of the genetic code.

We have to

