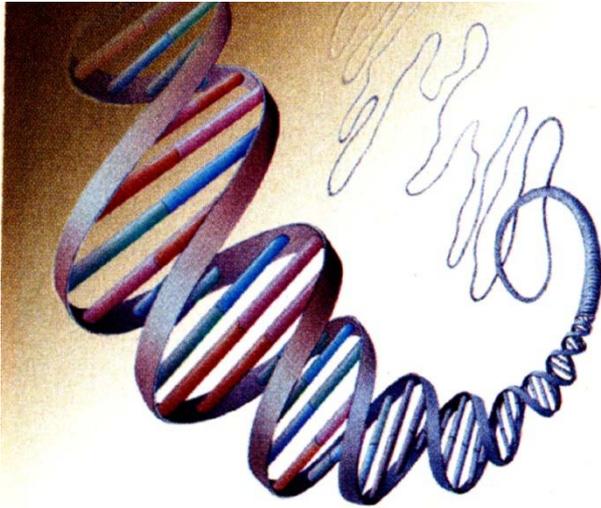


# Código Genético

**Prof. Eduardo Moraes Rego Reis**  
**Instituto de Química – USP**

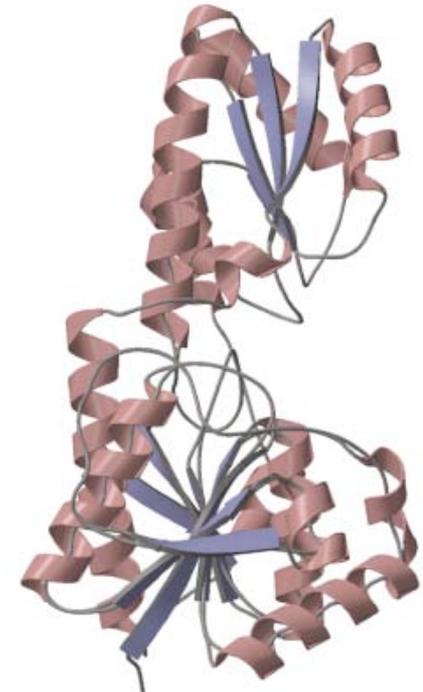
**QBQ1354 – Biologia Molecular**  
**Química**

# O CÓDIGO GENÉTICO



código 4 letras  
(AGCT)

LEITURA

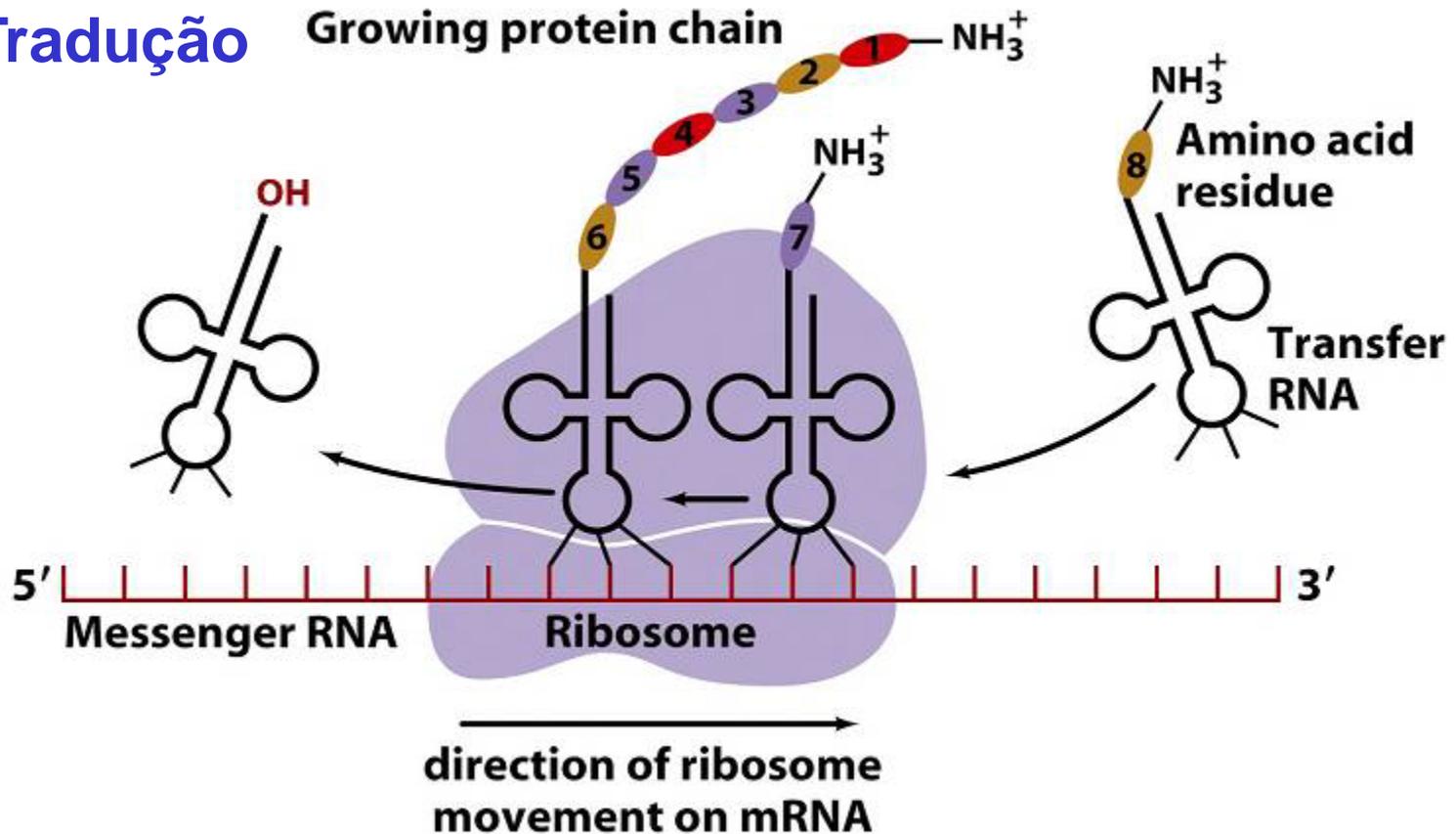


código 20 letras  
(aminoácidos)

**Relação entre a seqüência de bases no DNA  
(mRNA) e a seqüência de aminoácidos na  
proteína**

# Como a seqüência de bases do mRNA é interpretada para compor a seqüência de aminoácidos na proteína?

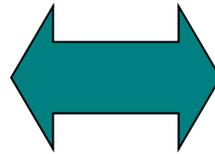
## Tradução



# Decifrando o Código Genético

Quantos nucleotídeos seriam necessários para compor a unidade mínima do código?

4 nucleotídeos  
diferentes no  
DNA



20 aminoácidos  
diferentes na  
proteína

Nº máximo de combinações

Código de um nucleotídeo ( $4^1$ ) = 4 combinações



Código de dois nucleotídeos ( $4^2$ ) = 16 combinações

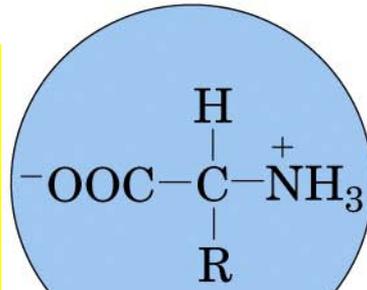


Código de três nucleotídeos ( $4^3$ ) = 64 combinações



•Francis Crick sugeriu (1955) a existência de **moléculas adaptadoras** para interagir com os aminoácidos e o mRNA.

•Sugeriu ainda que o código **seria lido em trincas**.



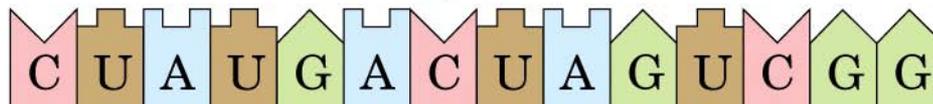
**Aminoácido**

**Sítio de ligação do Aminoácido**

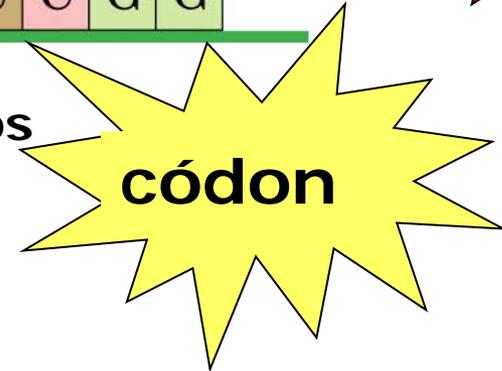
**Molécula adaptadora**



mRNA

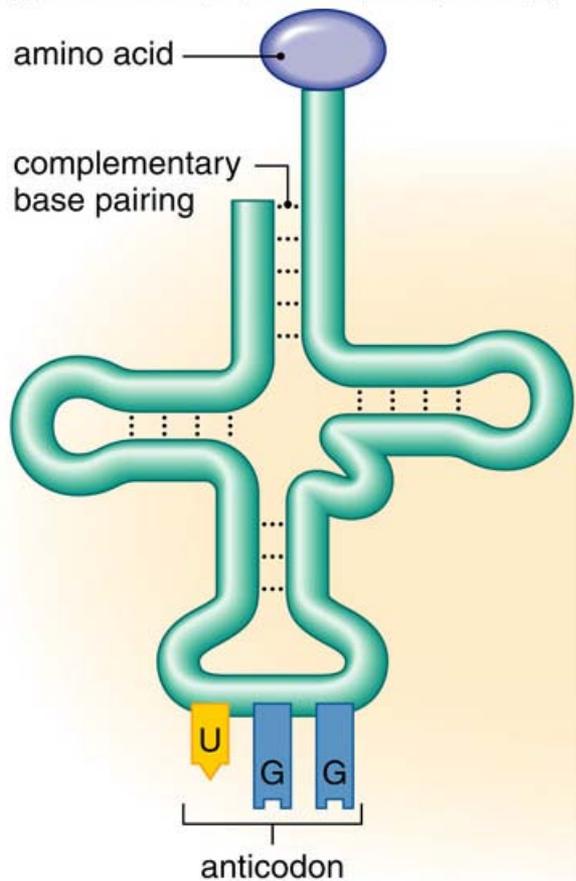


**Triplete de nucleotídeos codificando para um aminoácido**

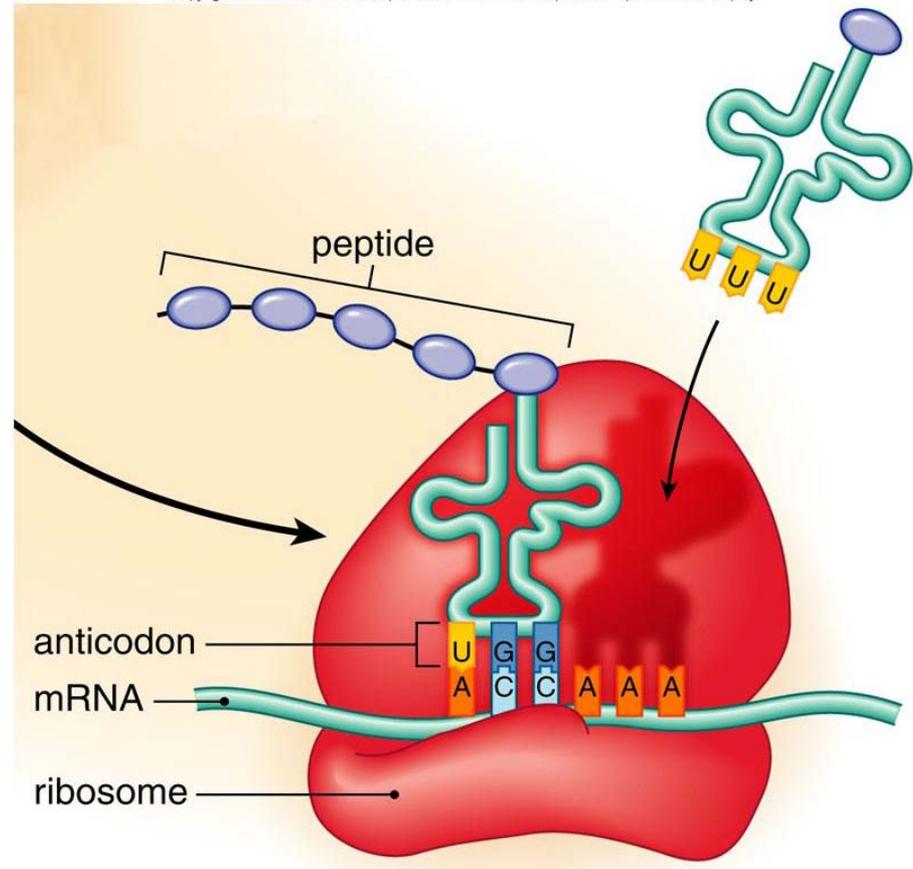


# RNAs transportadores (tRNAs):

moléculas adaptadoras que integram o código do DNA com a maquinaria de síntese protéica

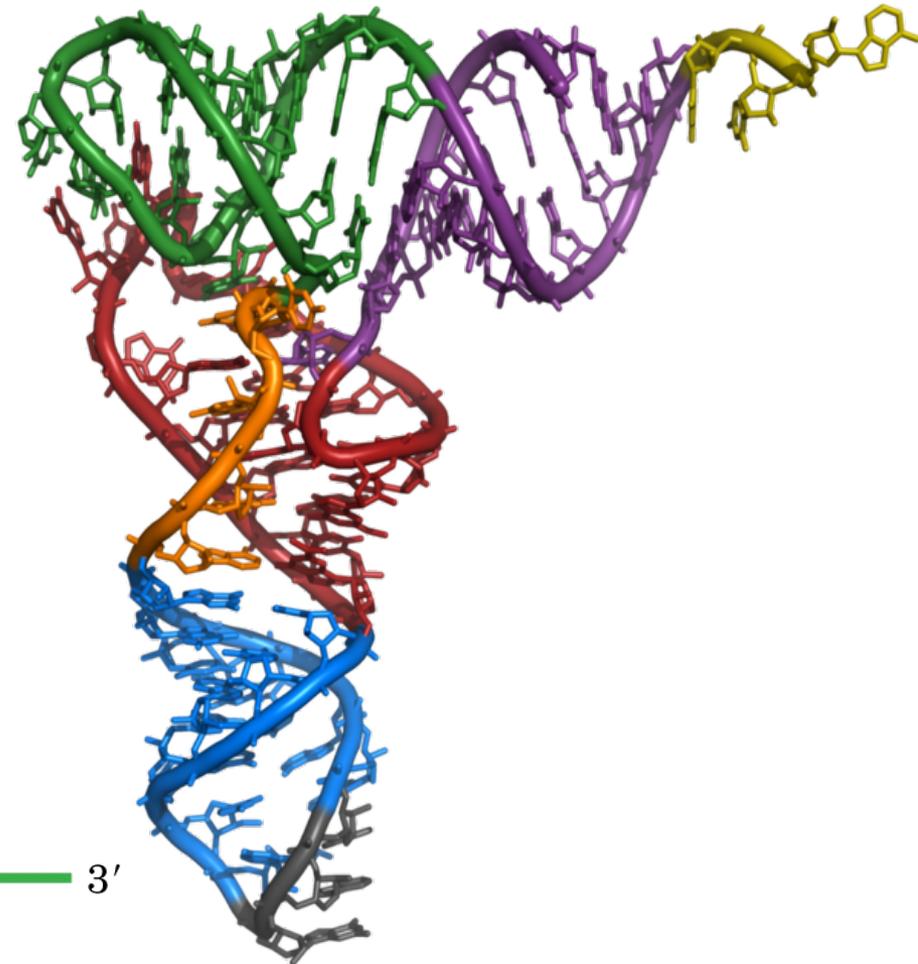
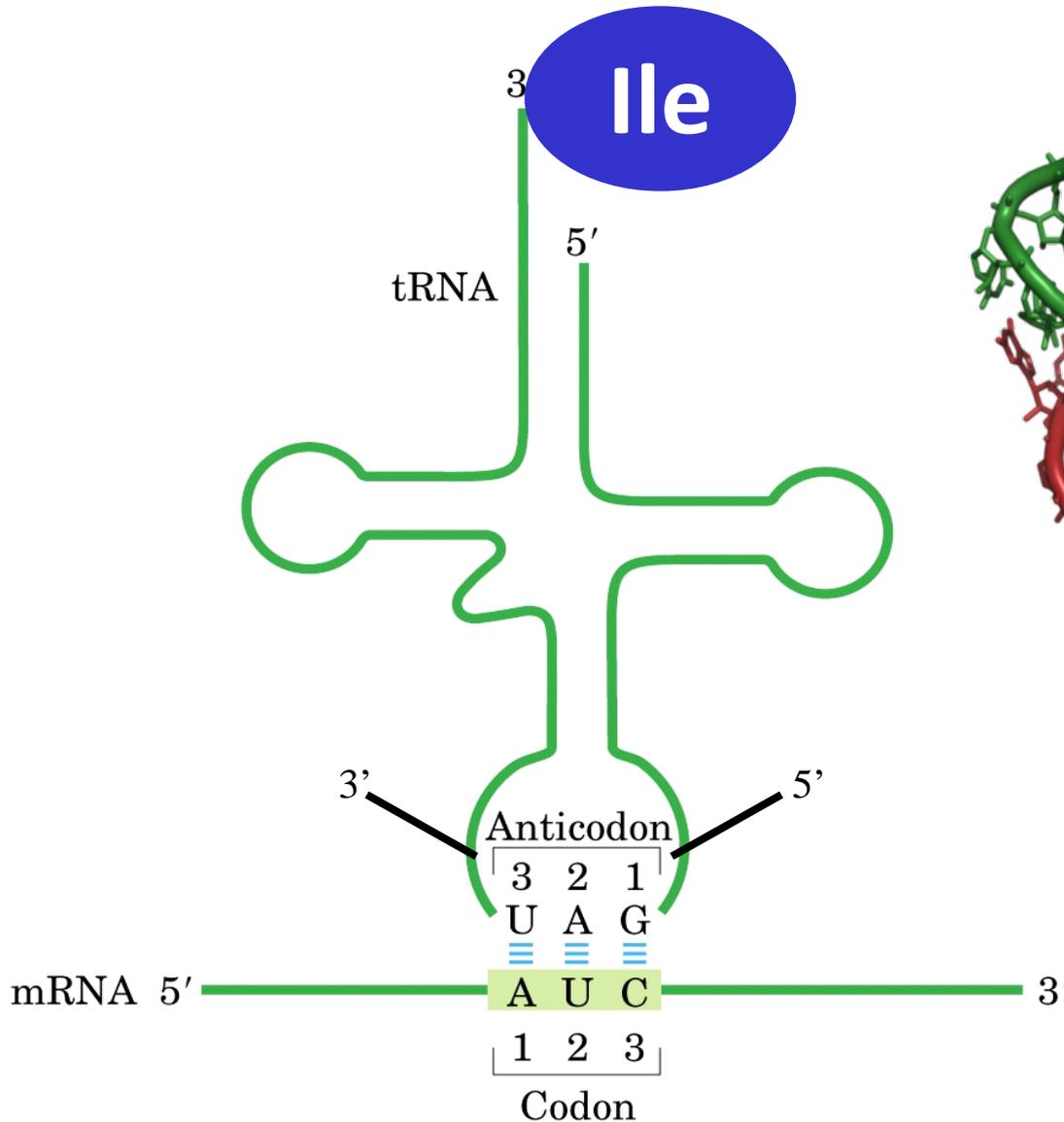


**tRNA carregado com aminoácido  
(aminoacil-tRNA)**



**aminoacil-tRNA no ribosomo**

# Códon do mRNA pareia com o anticódon do tRNA



# Decifrando o Código Genético

Como o código foi decifrado?

Tanto as sequências de bases do DNA como a de aminoácidos das proteínas eram desconhecidas!!!

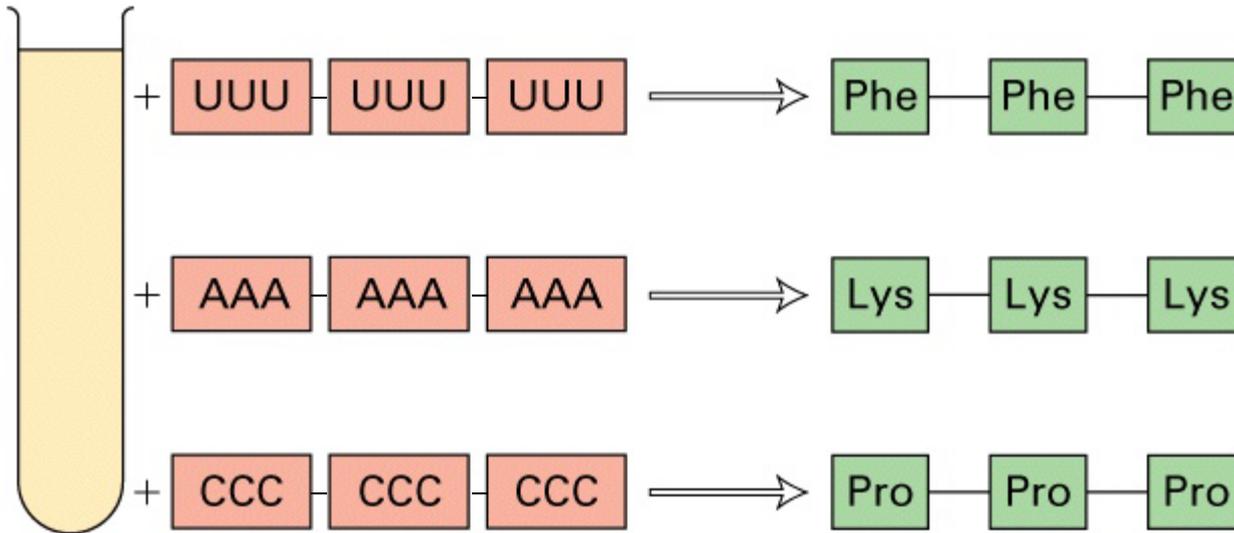
Nirenberg em 1961 mostrou que a adição do RNA sintético poliuridilato (poli U \*) em um sistema de síntese proteica livre de células (extrato de *E.coli*) levava a síntese de polifenilalanina:

**UUU = Phe**

\* Poli U foi sintetizado *in vitro* pela polinucleotídeo fosforilase.

# Decifrando o código genético

Extrato bacteriano capaz de síntese protéica in vitro



RNA sintético

Polipeptídeos obtidos

Contem todos os ingredientes celulares necessários para a síntese de proteínas:

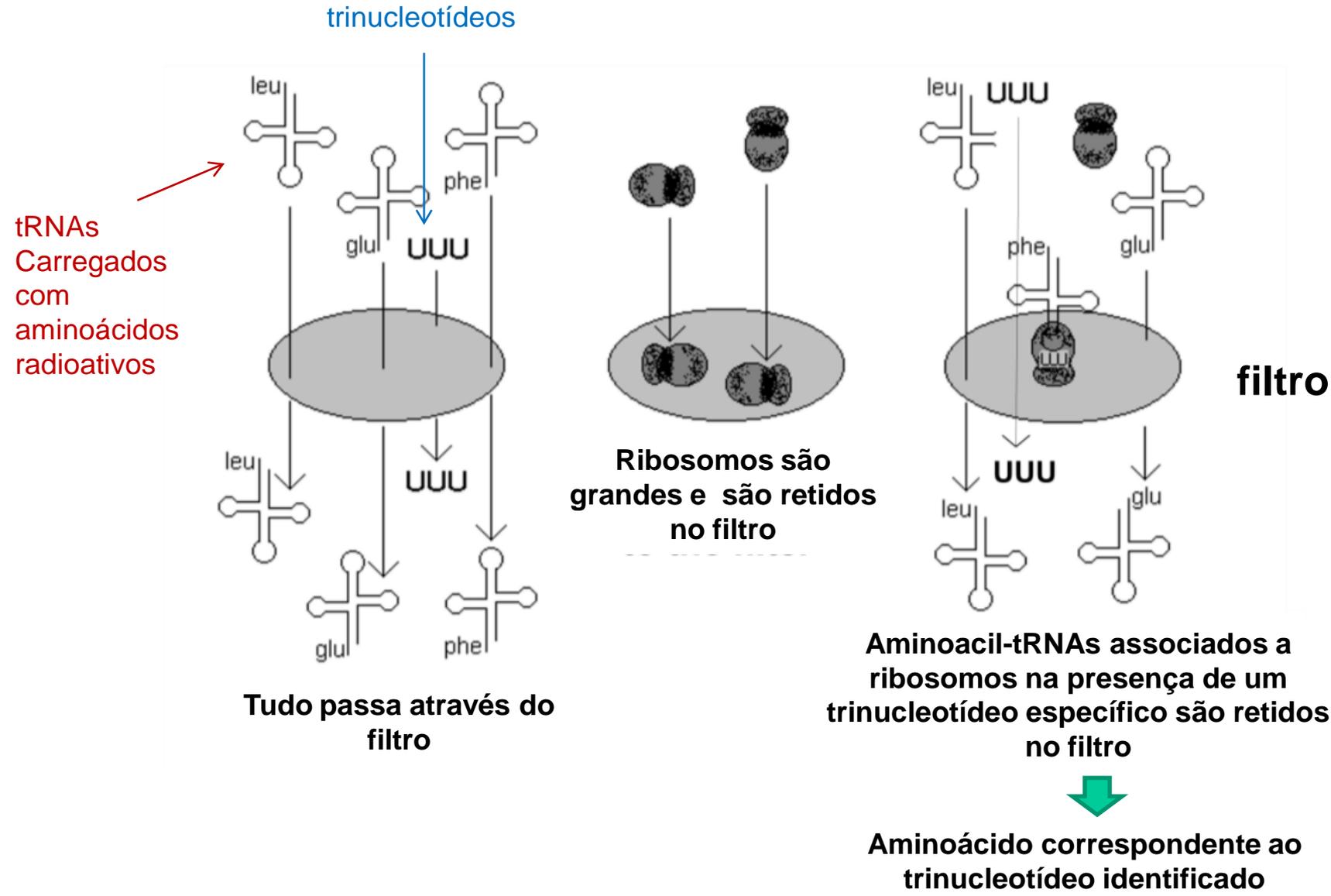
- Ribosomos
- Aminoacil tRNAs
- Co-fatores

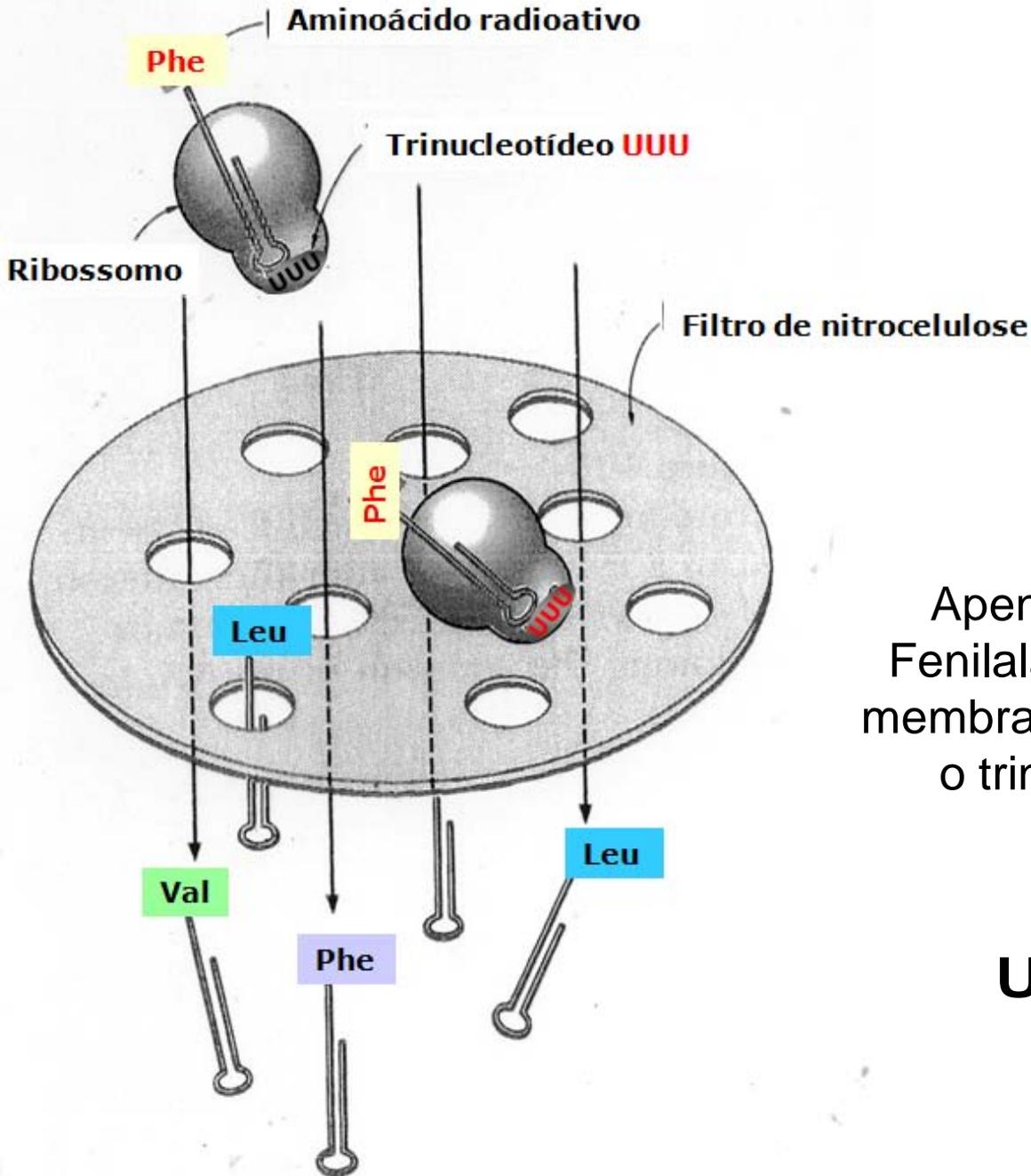
**Poli U**  
UUU = Phe

**Poli A**  
AAA = Lys

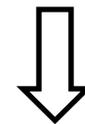
**Poli C**  
CCC = Pro

Niremberg e Leder utilizaram a propriedade de **tRNAs carregados com aminoácidos se associarem a ribonucleotídeos específicos** no ribosomo para confirmar todos os 64 códons.





Apenas o aminoácido  
Fenilalanina fica retido na  
membrana quando se utiliza  
o trinucleotídeo UUU



**UUU = Phe**

# Decifrando o Código Genético

## Trinucleotides Induce Specific Binding of Aminoacyl-tRNAs to Ribosomes

Trinucleotide	<sup>14</sup> C-Labeled aminoacyl-tRNA bound to ribosome*		
	Phe-tRNA <sup>Phe</sup>	Lys-tRNA <sup>Lys</sup>	Pro-tRNA <sup>Pro</sup>
UUU	4.6	0	0
AAA	0	7.7	0
CCC	0	0	3.1

**Source:** Modified from Nirenberg, M. & Leder, P. (1964) RNA code words and protein synthesis. *Science* **145**, 1399.

\*Each number represents the factor by which the amount of bound <sup>14</sup>C increased when the indicated trinucleotide was present, relative to a control in which no trinucleotide was added.

# Decifrando o Código Genético

O código é degenerado

64 códons para codificar 20 aminoácidos



61 códons codificam 20 aminoácidos proteínogênicos

Para a maioria dos aminoácidos há mais de um códon

3 códons não codificam aminoácidos (UAA, UAG, UGA)

São códons de terminação da síntese de proteínas

## 2a. Letra do códon

U C A G

## Degeneração do código genético

### 1a. Letra do códon

U

C

A

G

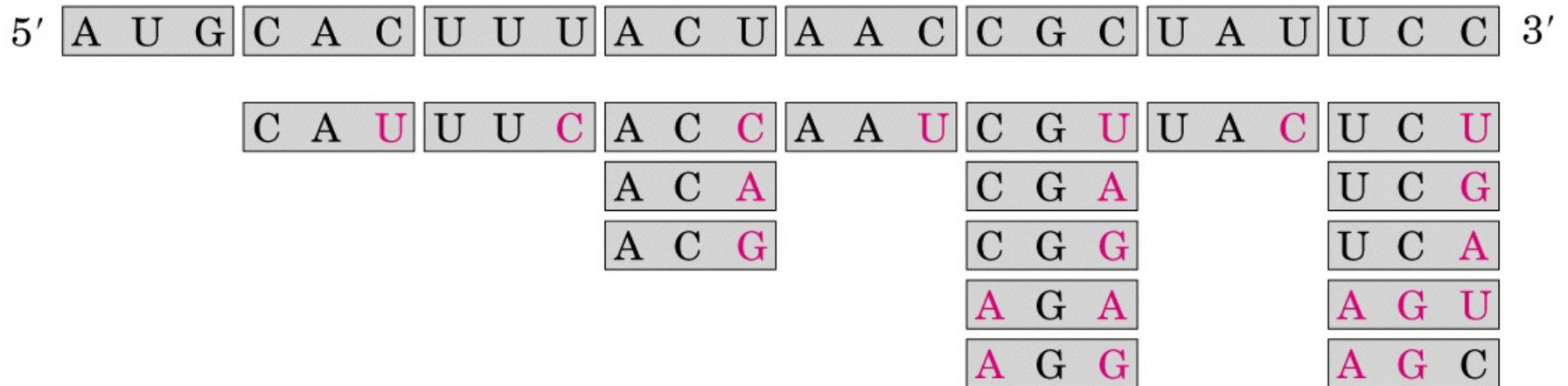
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
U	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser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

Amino acid	Number of codons
------------	------------------

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

# Consequência da degeneração do código genético

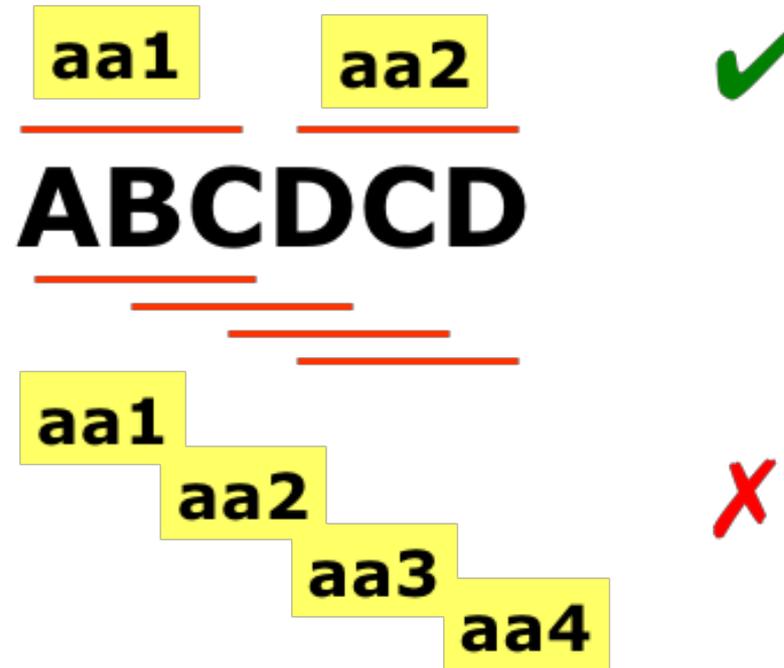
Met — His — Phe — Thr — Asn — Arg — Tyr — Ser



seqüências diferentes de mRNA podem especificar a mesma seqüência de aminoácidos

# Características do Código Genético

## O código não é superposto



•Análise da seqüência de aminoácidos de mutantes da proteína da capa do vírus mosaico do tabaco mostraram que o código não era superposto.



•A mutação em um nucleotídeo leva a mudança de um aminoácido e não de dois ou três aminoácidos.

# Características o Código Genético

3 fases de leitura são teoricamente possíveis em cada uma fitas.

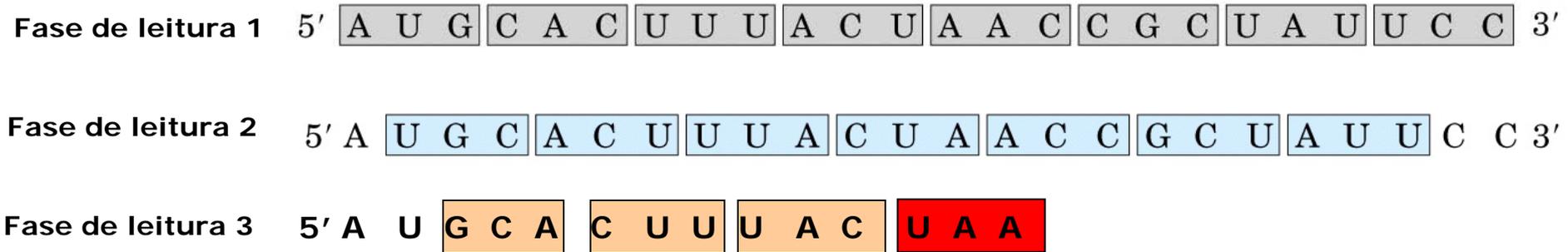
3 sequencias proteicas possíveis na fita + (+1, +2, +3)

DNA

```
R S R A F W S P M S A A D S S * K A A P F T N R A S N R Q P R T A K
D L G R S G R R C R R P T H L E R L H R S R T G R R T G N R G R R
I S G V L V A D V G G R L I L K G C T V H E P G V E P A T A D G E
CGATCTCGGGCGTTCTGGTCGCCGATGTCGGCGGCCGACTCATCTTGAAAGGCTGCACCGTTACGAACCGGGCGTCGAACCGGCAACCGCGGACGGCGA
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
GCTAGAGCCCGCAAGACCAGCGGCTACAGCCCGCGGCTGAGTAGAACTTCCGACGTGGCAAGTGCTTGGCCCGCAGCTTGGCCGTGGCGCCTGCCGCT
R D R A N Q D G I D A A S E D Q F A A G N V F R A D F R C G R V A
S R P R E P R R H R R G V * R S L S C R E R V P R R V P L R P R R
A I E P T R T A S T P P R S M K F P Q V T * S G P T S G A V A S P S
```

3 sequencias proteicas possíveis na fita - (-1, -2, -3)

# Características o Código Genético



- 3 fases de leitura são teoricamente possíveis em cada uma fitas.
- código é lido sequencialmente sem pausas a partir do início determinado, até um códon de terminação

# Diferentes fases de leitura da sequencia de DNA/RNA resultam em diferentes sequencias de aminoácidos

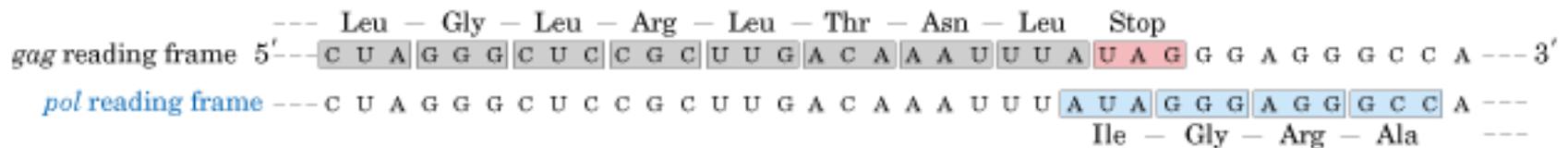
Amino acid sequence Met — His — Phe — Thr — Asn — Arg — Tyr — Ser

Fase de leitura 1 5' A U G C A C U U U A C U A A C C G C U A U U C C 3'

Fase de leitura 2 5' A U G C A C U U U A C U A A C C G C U A U U C C 3'

Amino acid sequence Cys — Thr — Leu — Leu — Thr — Ala — Ile

# Uso alternativo de fase de leitura pode dar origem a proteínas diferentes a partir do mesmo RNA



Ex.:

Genes gag-pol no genoma de RNA de retrovírus (ex. HIV)

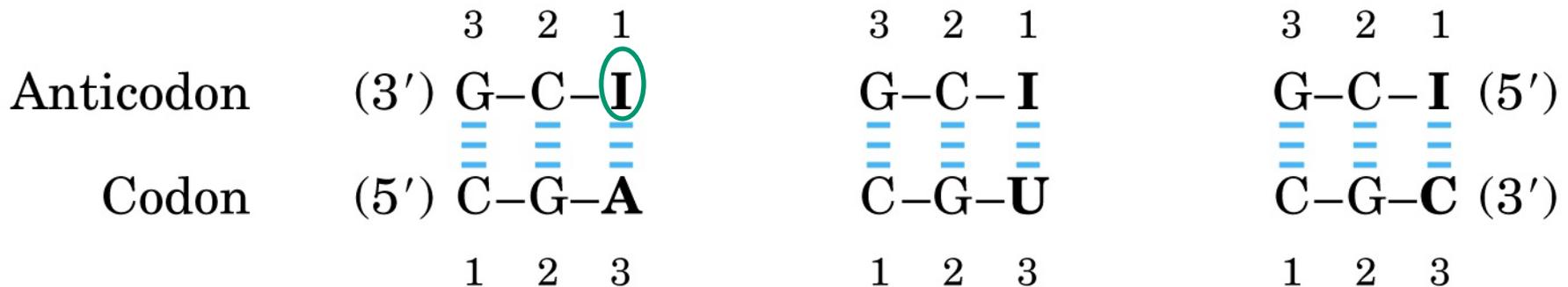
- Codificam para a proteína gag (capsídeo) e pol (transcriptase reversa) em fases diferentes de leitura
- Tradução do RNA da proteína gag ocorre na proporção 20:1 em relação a proteína pol

# 32 tRNAs são suficientes para decodificar os 61 códons

## Como explicar?

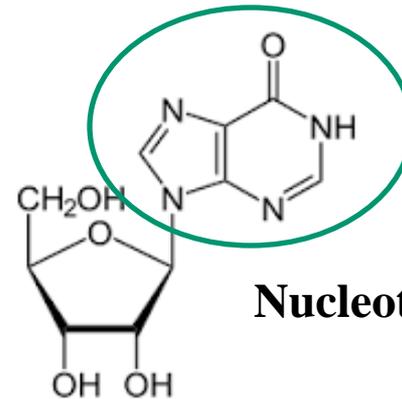
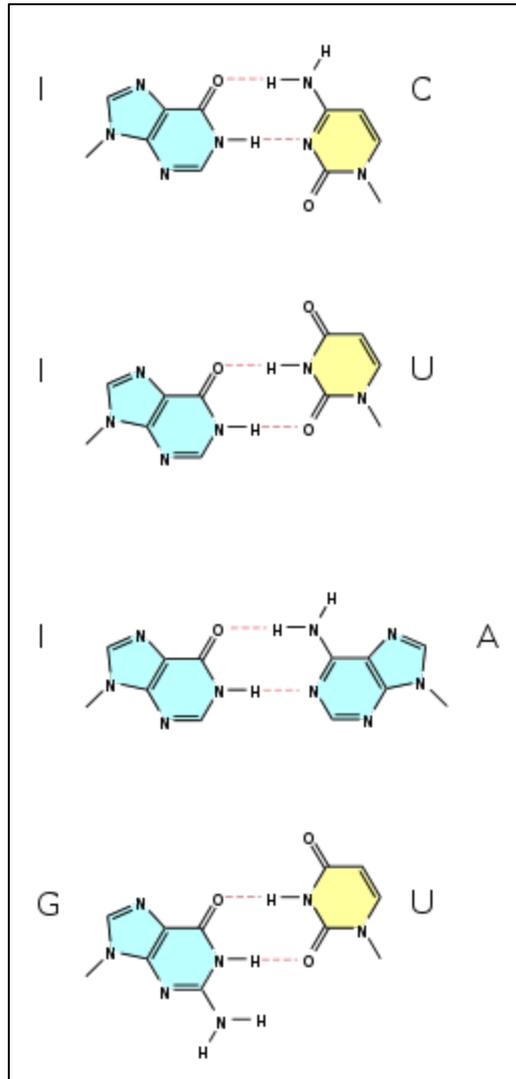
**Pareamento oscilante** (em inglês, "wobble") da 3ª base do códon com a 1ª base do anticodon permite que um mesmo tRNA se ligue a diferentes codons que codificam o mesmo aminoácido

I = Inosina (base modificada)



- Os 3 códons acima codificam para o aminoácido **arginina**
- **Um mesmo arg-tRNA liga-se em 3 códons distintos**

# Pareamento oscilante de bases inosina e uracila



Nucleotídeo de inosina

Pareamentos estáveis entre I = C

I = U

I = A

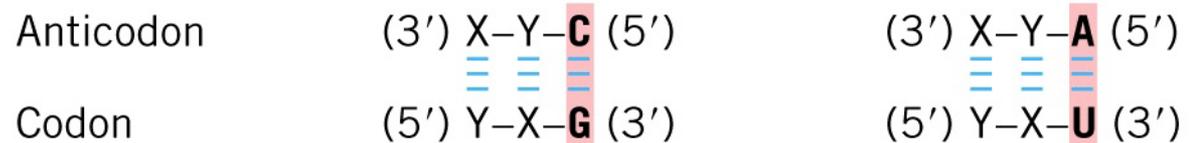
G = U

- Intermoleculares (códon do mRNA e anticódon do tRNA)

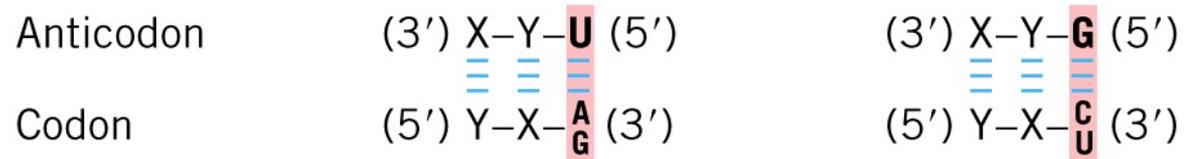
- Intramoleculares (estruturas secundárias de RNAs)

# Pareamento oscilante da terceira base do códon ("wobble base-pairing")

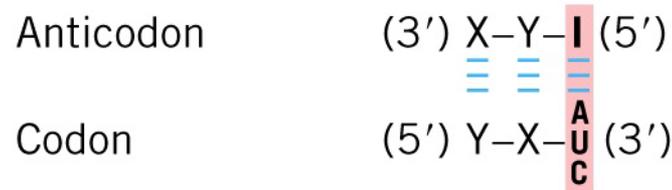
1. One codon recognized:



2. Two codons recognized:



3. Three codons recognized:



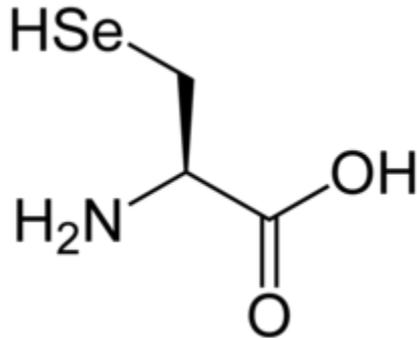
\*X and Y denote complementary bases capable of strong Watson-Crick base pairing with each other. The bases in the wobble positions—the 3' position of codons and 5' position of anticodons—are shaded in red.

**Nucleotídeo presente na posição 1 do anticódon  
determina as propriedades de pareamento oscilante**

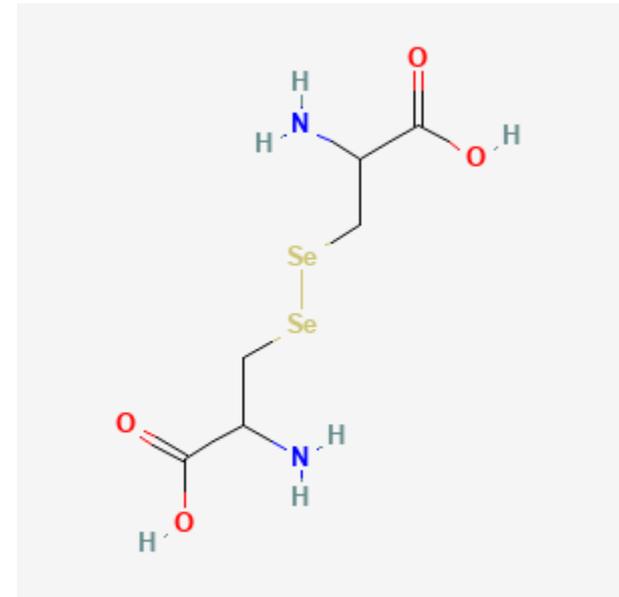
# Algumas proteínas contêm um ou mais resíduos de selenocisteína, um análogo de cisteína

## selenocisteína

análogo de cisteína com selenio ao invés de enxofre



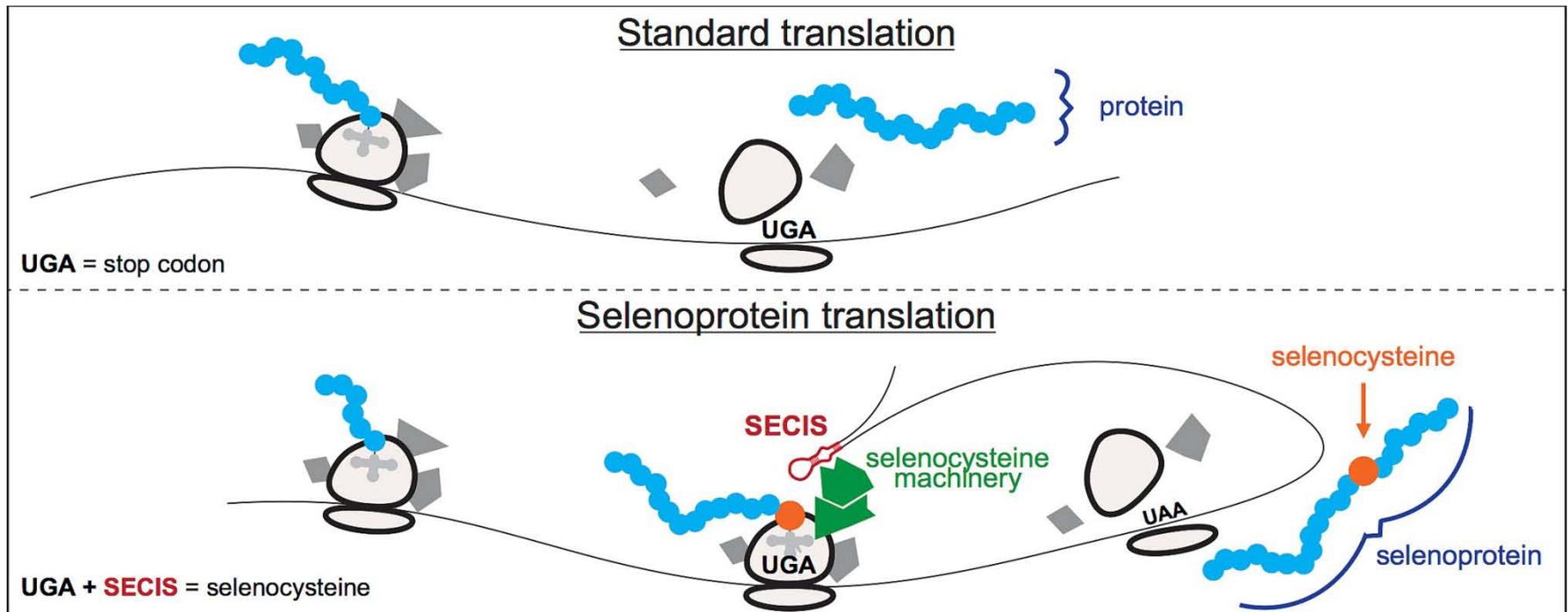
## selenocistina (forma oxidada)



- Selenocisteína é o 21<sup>o</sup> aminoácido proteinogênico
- Potencial de redução mais baixo que cisteína
- Presente em selenoproteínas, várias com ação anti-oxidante (ex. glutathiona peroxidase, tioredoxina redutase)

# A selenocisteína (SeCys) é codificada pelo codon UAG (codon de parada)

- Normalmente o codon UAG é um codon de parada da tradução (codon "nonsense").
- RNAs que codificam para selenoproteínas possuem trechos com ~60 nt na vizinhança dos UAGs onde ocorre a inserção de selenocisteínas ("selenocystein insertion sequence, SECIS") durante a tradução.
- O reconhecimento do codon UAG é feito por um tRNA específico (tRNA-SeCys)



# O conhecimento do código genético tem levado ao desenvolvimento de sistemas sintéticos para síntese de proteínas contendo aminoácidos não-canônicos

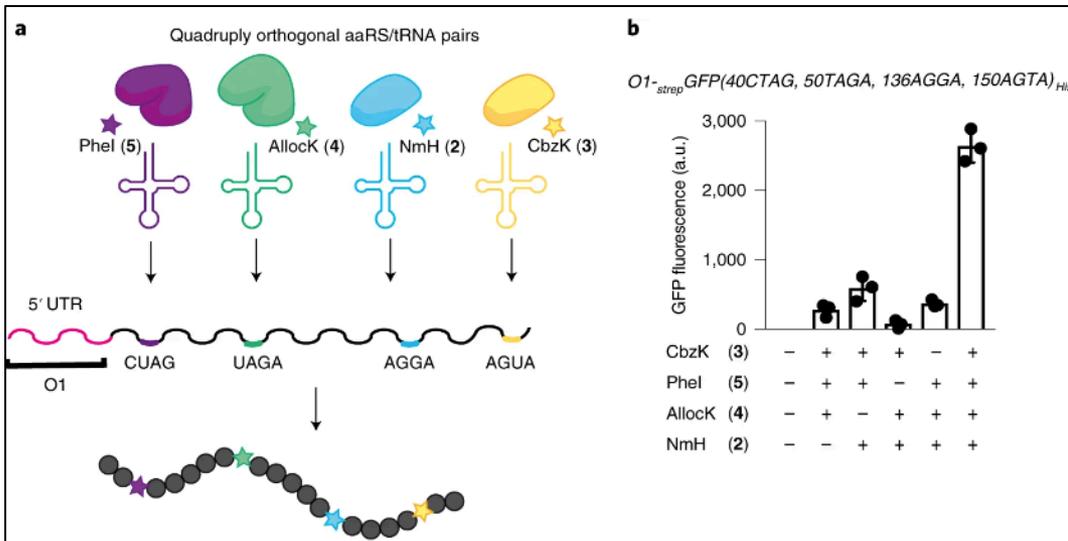
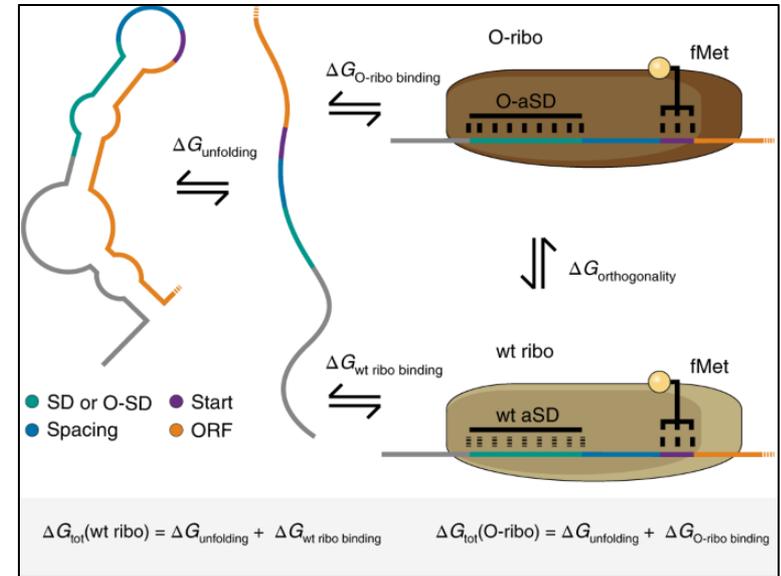
Article | [Published: 23 August 2021](#)

## A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design

[Daniel L. Dunkelmann](#), [Sebastian B. Oehm](#), [Adam T. Beattie](#) & [Jason W. Chin](#) ✉

*Nature Chemistry* **13**, 1110–1117 (2021) | [Cite this article](#)

7568 Accesses | 13 Citations | 86 Altmetric | [Metrics](#)



Células manipuladas geneticamente para produzir ribosomos não naturais (O-ribosomos) capazes de se ligar a mRNAs sintéticos (O-mRNAs) que contêm codons artificiais compostos por 4 bases, reconhecidos por tRNAs não-naturais acoplados a aminoácidos não canônicos