

Como genes codificam proteínas

QBQ102

Prof. João Carlos Setubal



Universidade de São Paulo
Instituto de Química

Como DNA permite...

- A atividade da vida?
- A reprodução da vida?
- Hoje vamos ver a parte da **atividade da vida**

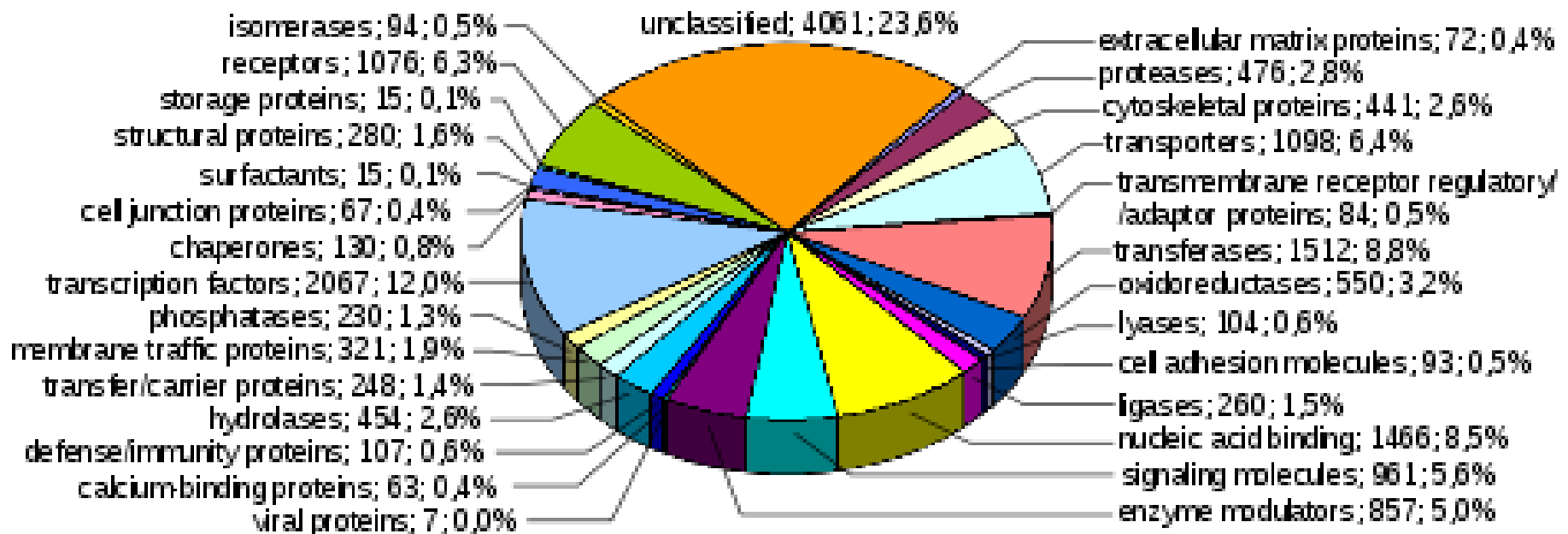
Atividade da vida significa...

- ...basicamente...
- Montar (ou sintetizar) proteínas

Proteínas são as moléculas trabalhadoras dos organismos

- “nós somos nossas proteínas”
- **Enzimas**: catalizam reações químicas, essenciais para o metabolismo celular
- Funções **estruturais** ou **mecânicas** (**actina** e **miosina** em músculos)
- **Sinalização celular**
- **Síntese de DNA**
 - A DNA polimerase é uma proteína

Variedade das proteínas humanas



Mas o que tem DNA a ver com proteínas?

- É no DNA que está a “receita” para a **fabricação das proteínas**
- Hélice dupla + complementaridade
 - ⇒ Replicação
- A cadeia de nucleotídeos
 - ⇒ **Informação**
 - ⇒ “receita” para sintetizar ou montar proteínas

Como assim?

- Da mesma forma como letras podem formar **palavras** em português que nós entendemos:
 - ANTICONSTITUCIONALÍSSIMAMENTE
- Em “celulês” ou “genomês” é possível formar palavras que a maquinaria da célula é capaz de entender:
 - ATGCCGGTCGTCGCGGACGACGACGG

Mas como são só 4 letras...

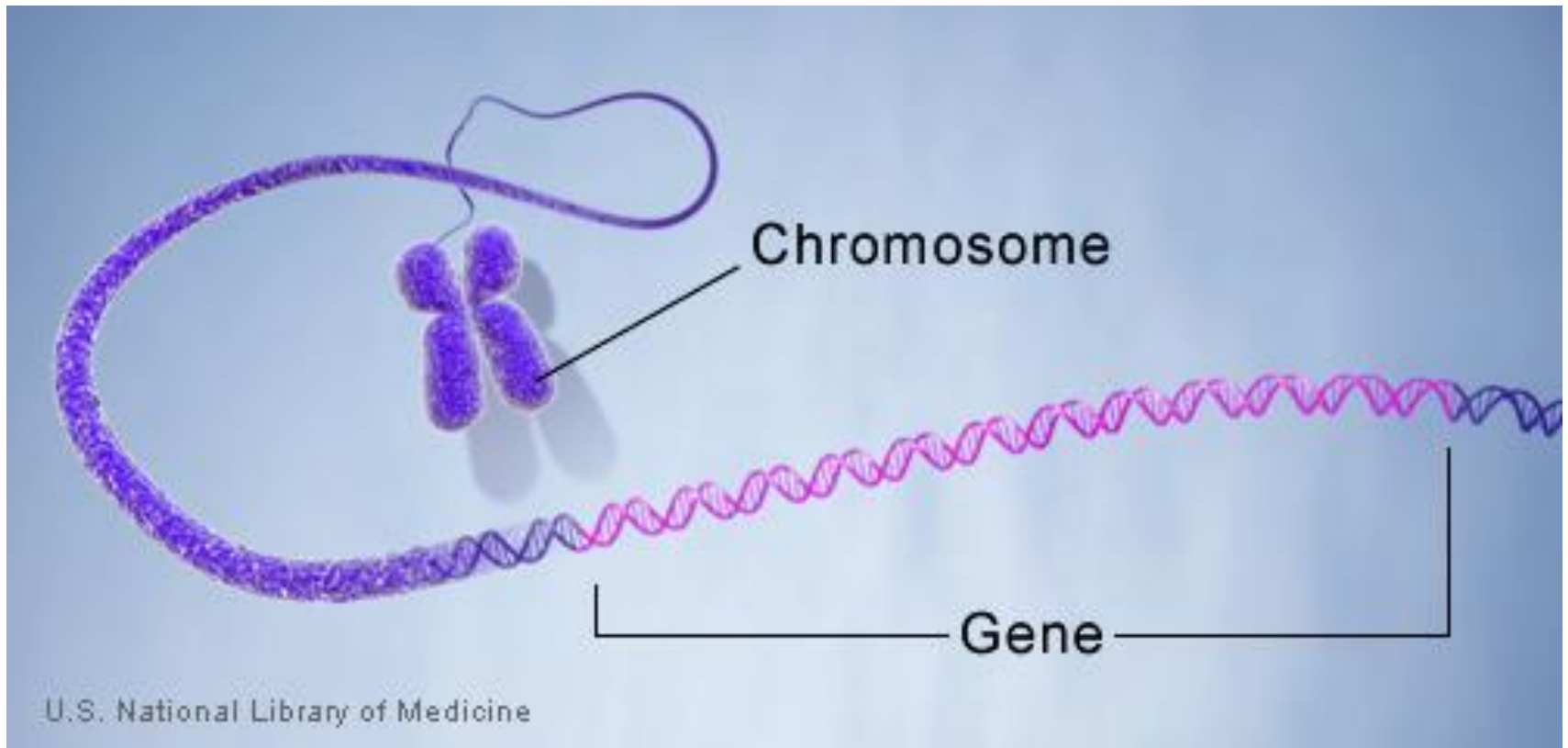
- Essas “palavras” são muito mais longas
- Tipicamente (numa bactéria) elas tem comprimento ~ 1000 pb
- Em outras palavras, um trecho de ~ 1000 pb é o tamanho da “receita” para montar uma proteína

Mas não é *qualquer* trecho de
~1000 pb...

- Assim como não é qualquer trecho de 10 letras que corresponde a uma palavra num texto em português
- Por exemplo:
 - quertre

Como se chamam os trechos com as receitas?

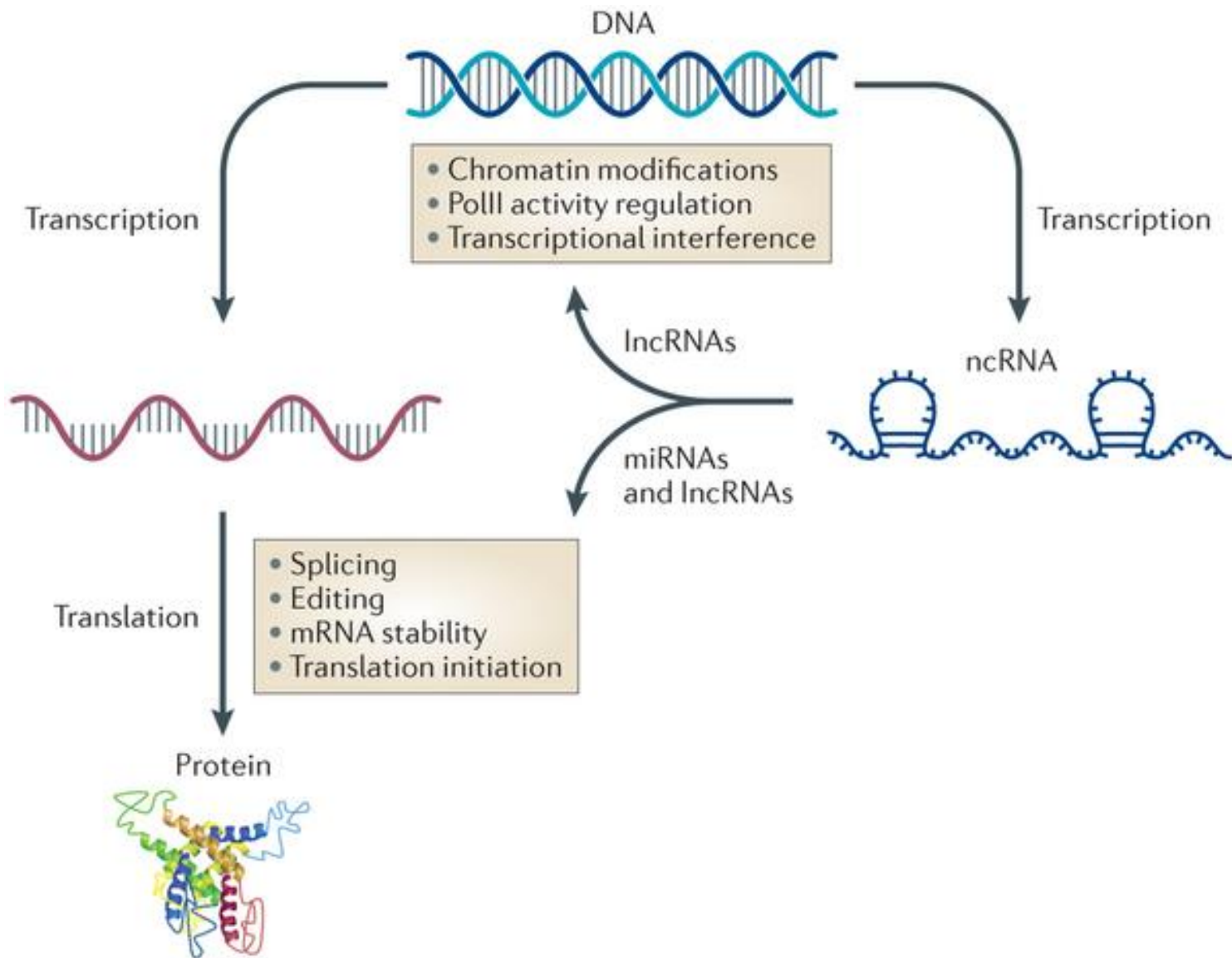
- Ou seja, as “palavras” em “genomês”?
- **Genes!**



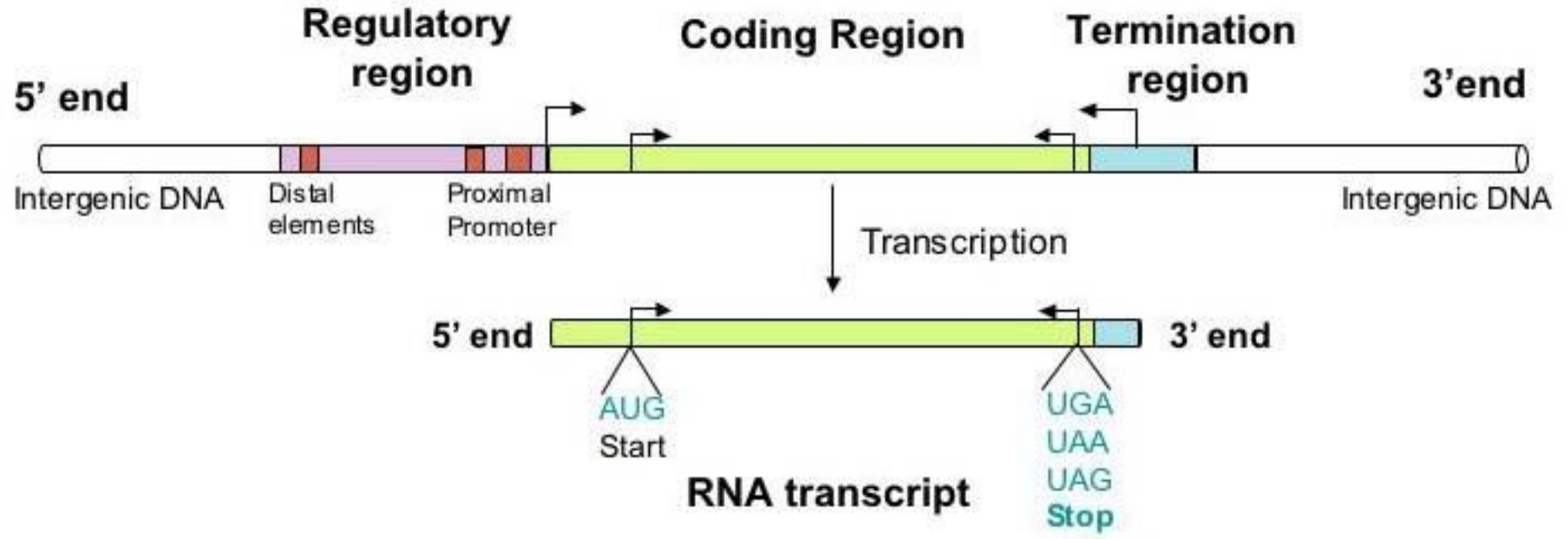
U.S. National Library of Medicine

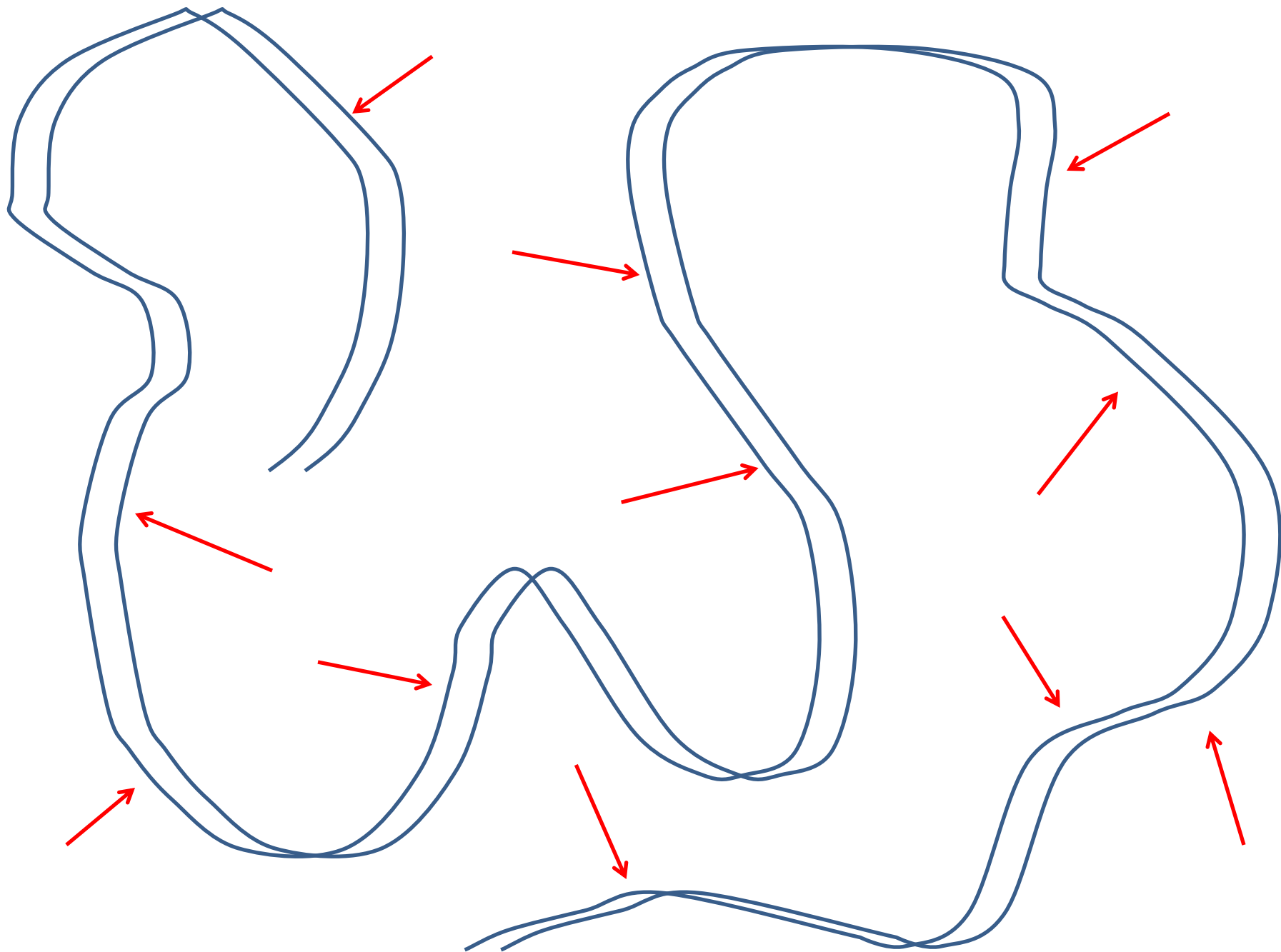
Informações armazenadas num genoma

- Genes **codificadores de proteína**
- Genes de **RNA**
 - tRNA
 - RNA ribossomal
 - Outros pequenos RNAs, a maioria **reguladores**



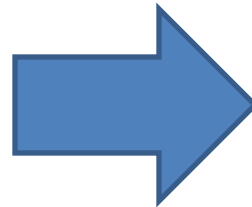
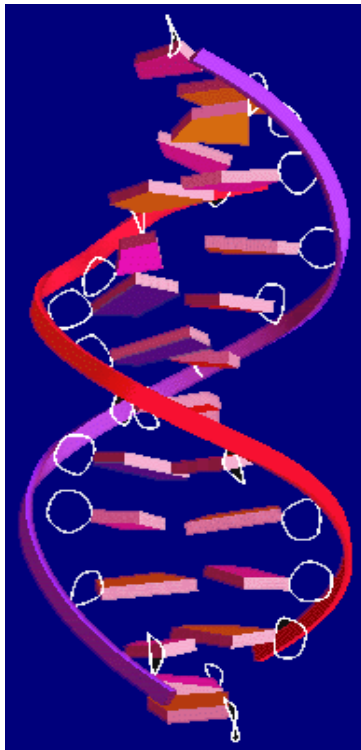
Prokaryotic Gene Structure



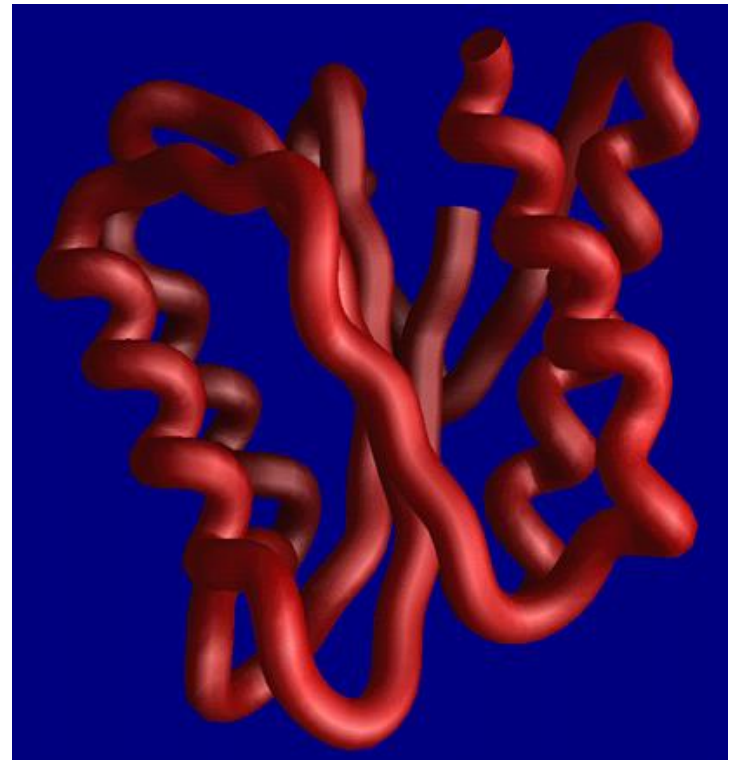


Genes e proteínas

DNA



Proteína



Para continuar...

- Precisamos falar de proteínas

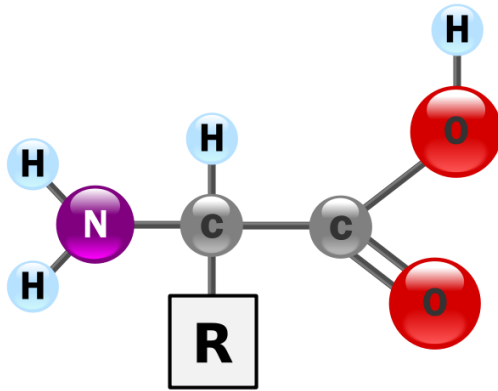
Proteínas são macromoléculas

São cadeias de **aminoácidos**

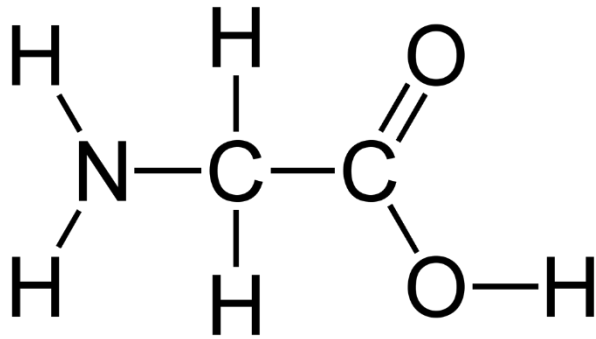
> Protein sequence

```
MKIVYWSGTGNTTEKMAELIAKGIIESGKDVNTINVSDVNI  
DELLNEDILILGCSAMGDEVLEESEFEPFIEEISTKISGK  
KVALFGSYGWGDGKWMRDFEERMNGYGCVVETPLIVQNE  
PDEAEQDCIEFGKKIANI
```

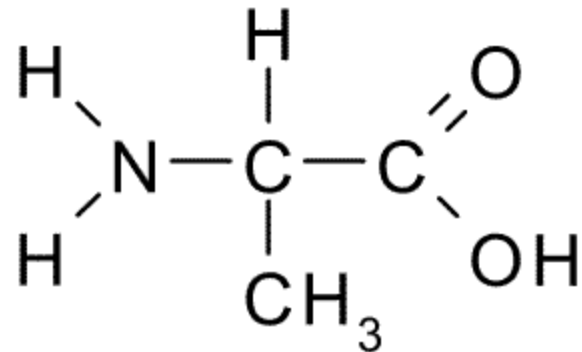
Aminoácidos



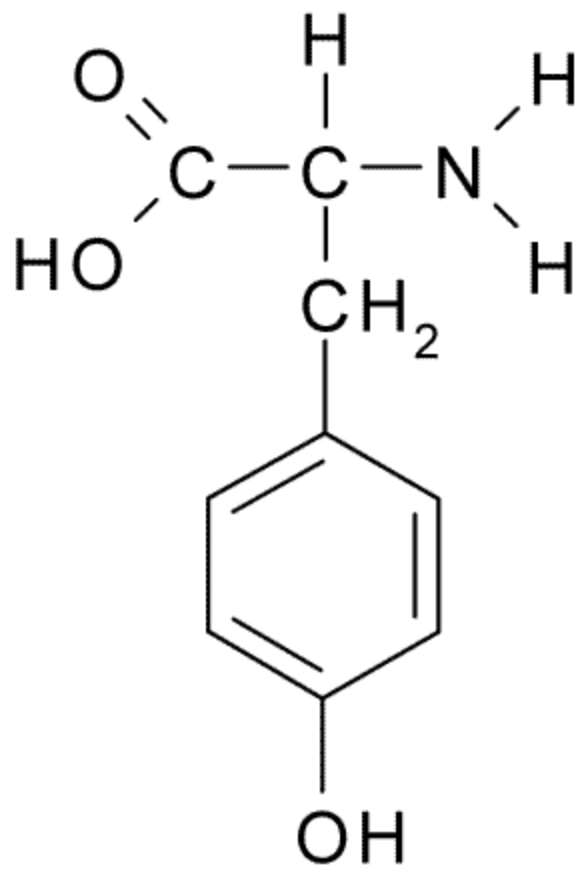
Estrutura genérica de um aminoácido



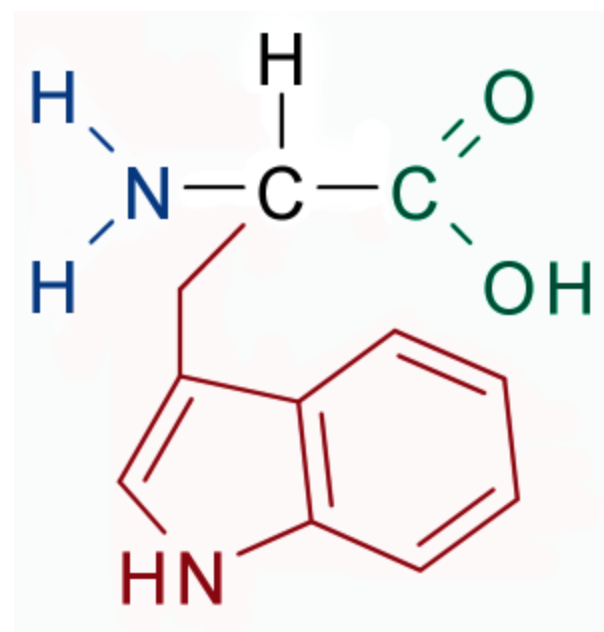
glicina



alanina



tirosina

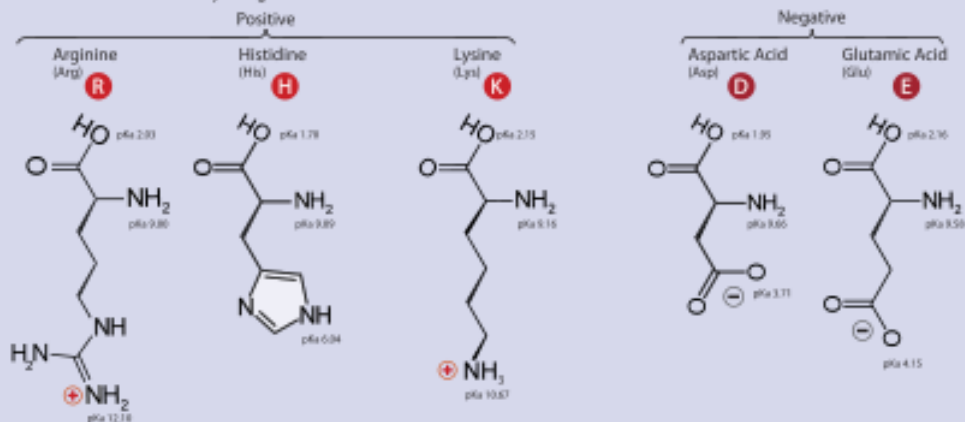


triptofano

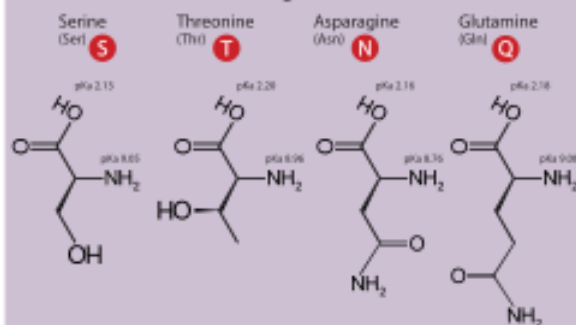
Twenty-One Amino Acids

⊕ Positive ⊖ Negative
 • Side chain charge at physiological pH 7.4

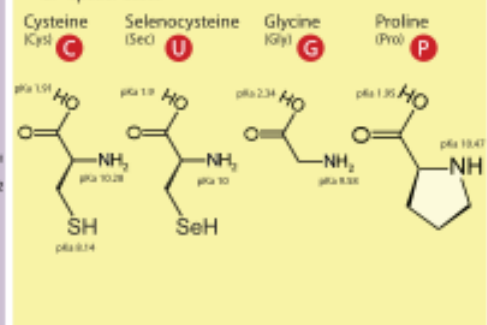
A. Amino Acids with Electrically Charged Side Chains



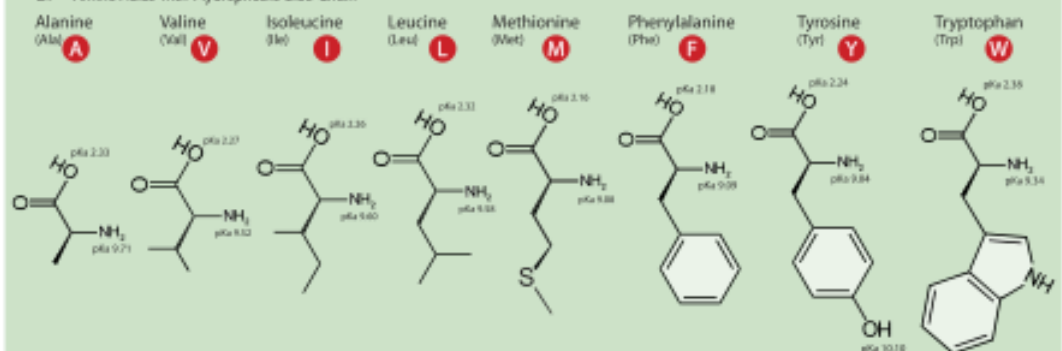
B. Amino Acids with Polar Uncharged Side Chains



C. Special Cases

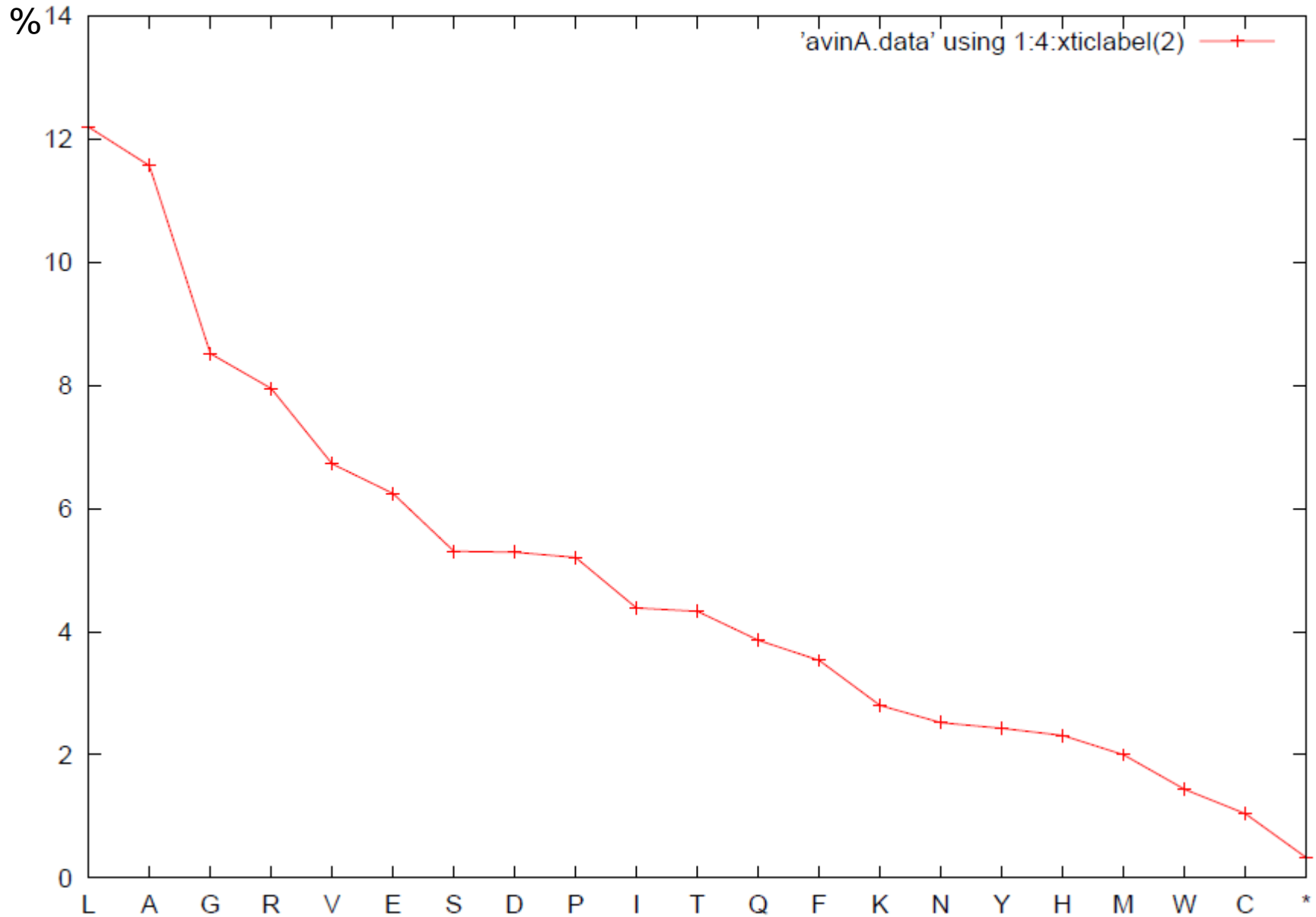


D. Amino Acids with Hydrophobic Side Chain

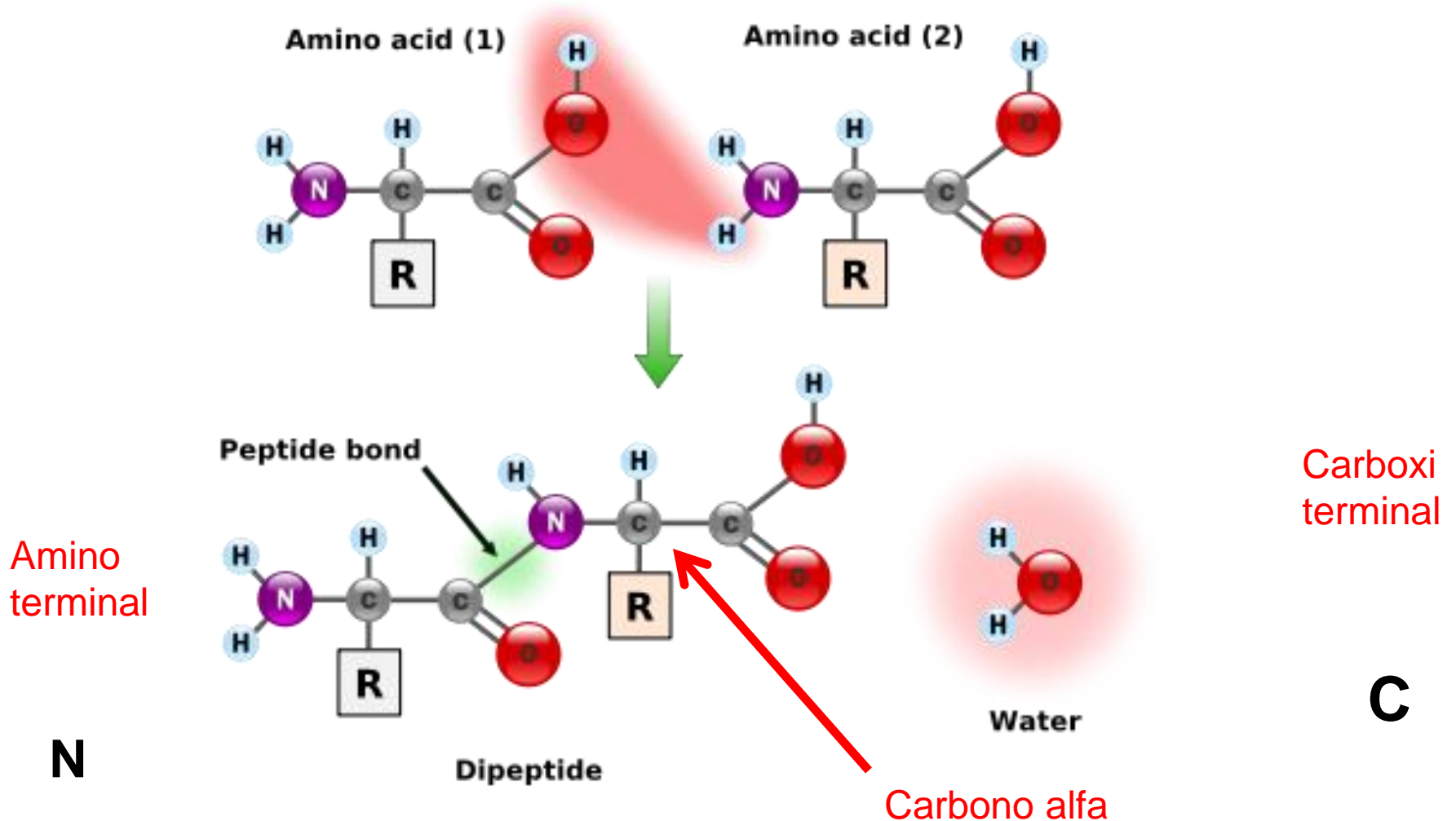


| Amino acid | 3-letter code | 1-letter code | MW (Da) | Structure |
|---------------|---------------|---------------|---------|--|
| Alanine | Ala | A | 89.1 | $\text{CH}_3\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Arginine | Arg | R | 174.2 | $\text{HN}=\text{C}(\text{NH}_2)\text{-NH-}(\text{CH}_2)_3\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Asparagine | Asn | N | 132.1 | $\text{H}_2\text{N-CO-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Aspartic Acid | Asp | D | 133.1 | $\text{HOOC-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Cysteine | Cys | C | 121.2 | $\text{HS-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Glutamic Acid | Glu | E | 147.1 | $\text{HOOC-}(\text{CH}_2)_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Glutamine | Gln | Q | 146.1 | $\text{H}_2\text{N-CO-}(\text{CH}_2)_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Glycine | Gly | G | 75.1 | $\text{NH}_2\text{-CH}_2\text{-COOH}$ |
| Histidine | His | H | 155.2 | $\text{NH-CH}=\text{N-CH}=\text{C-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ _____ |
| Isoleucine | Ile | I | 131.2 | $\text{CH}_3\text{-CH}_2\text{-CH}(\text{CH}_3)\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Leucine | Leu | L | 131.2 | $(\text{CH}_3)_2\text{-CH-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Lysine | Lys | K | 146.2 | $\text{H}_2\text{N-}(\text{CH}_2)_4\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Methionine | Met | M | 149.2 | $\text{CH}_3\text{-S-}(\text{CH}_2)_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Phenylalanine | Phe | F | 165.2 | $\text{Ph-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Proline | Pro | P | 115.1 | $\text{NH-}(\text{CH}_2)_3\text{-CH-COOH}$ _____ |
| Serine | Ser | S | 105.1 | $\text{HO-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Threonine | Thr | T | 119.1 | $\text{CH}_3\text{-CH}(\text{OH})\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Tryptophan | Trp | W | 204.2 | $\text{Ph-NH-CH}=\text{C-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ _____ |
| Tyrosine | Tyr | Y | 181.2 | $\text{HO-p-Ph-CH}_2\text{-CH}(\text{NH}_2)\text{-COOH}$ |
| Valine | Val | V | 117.1 | $(\text{CH}_3)_2\text{-CH-CH}(\text{NH}_2)\text{-COOH}$ |

Frequência de aminoácidos em proteínas da bactéria *Azotobacter vinelandii*

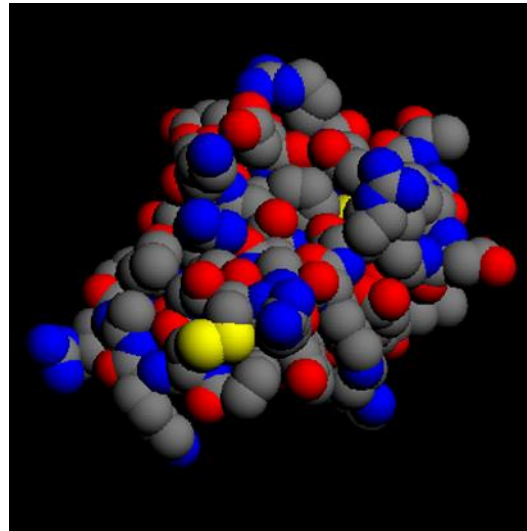


Os aminoácidos se ligam **entre si** por **ligações peptídicas**



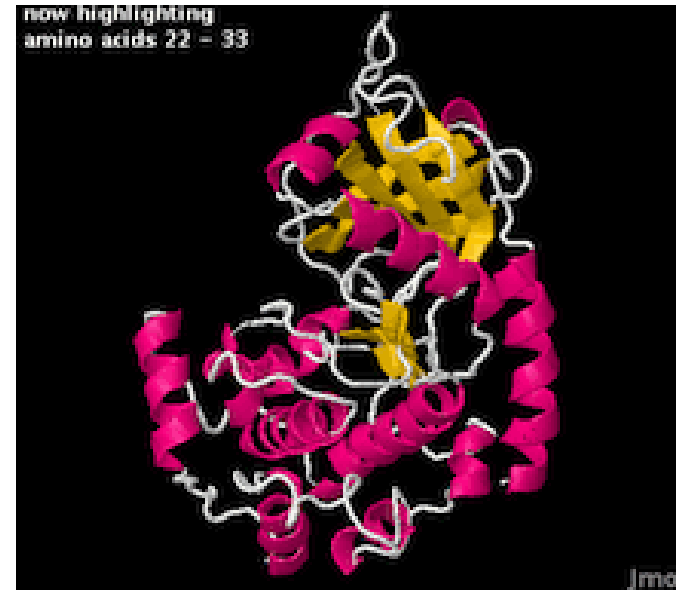
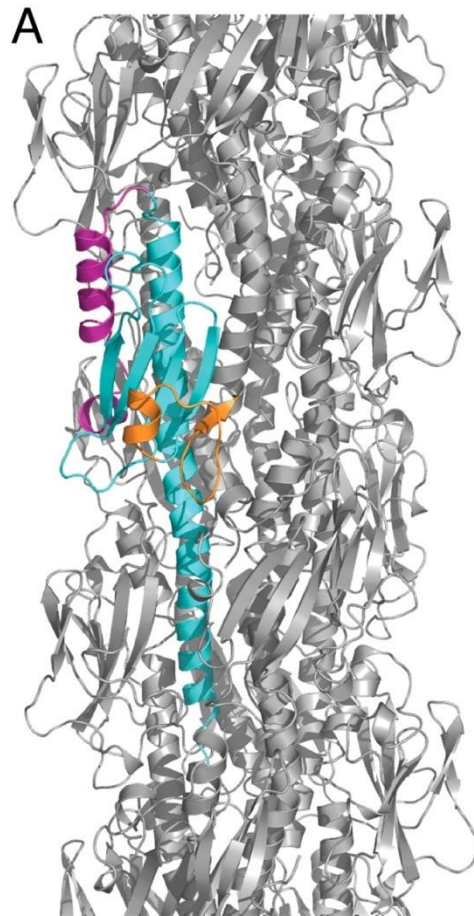
Proteínas formam estruturas tridimensionais

- São complexas e variadas
 - Diferentes proteínas tem diferentes estruturas



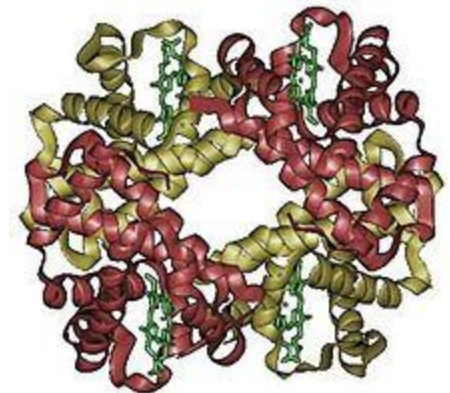
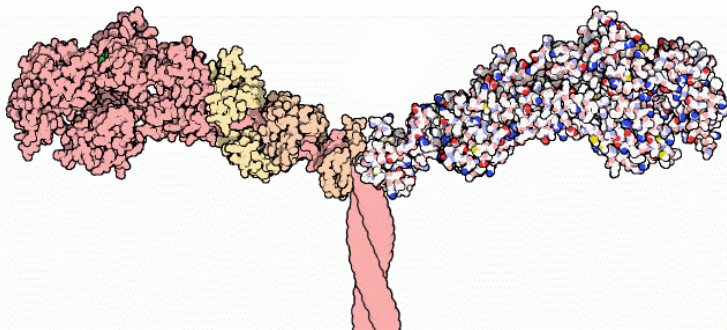
- (DNA é sempre hélice dupla)

Exemplos de proteínas



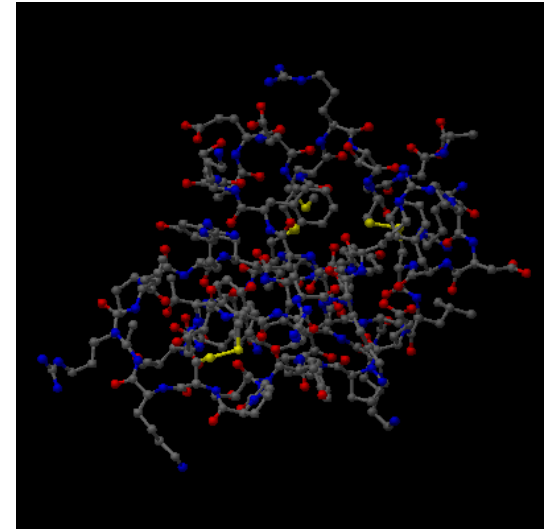
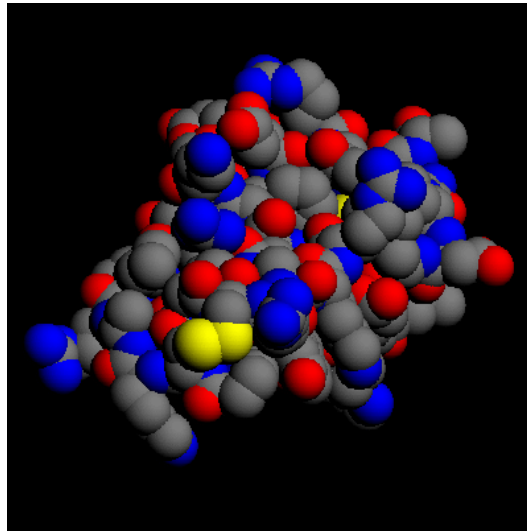
Diferentes estruturas conferem diferentes **funções**

Hemoglobina: transporta oxigênio no sangue



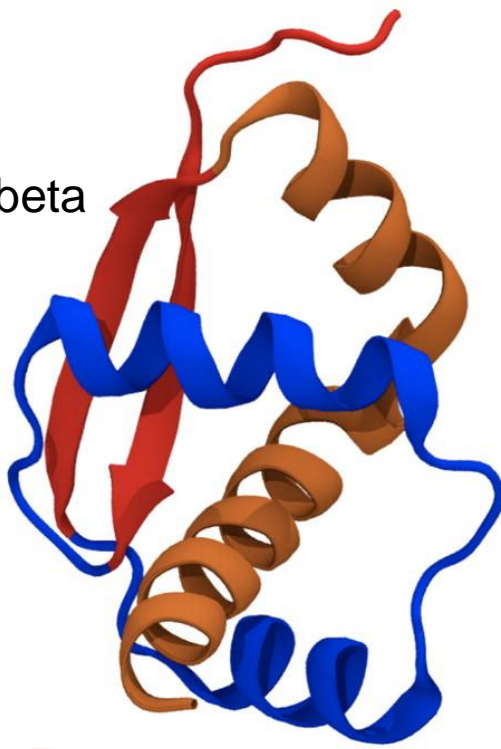
Miosina: contração do músculo

Proteínas podem ser visualizadas de diferentes formas

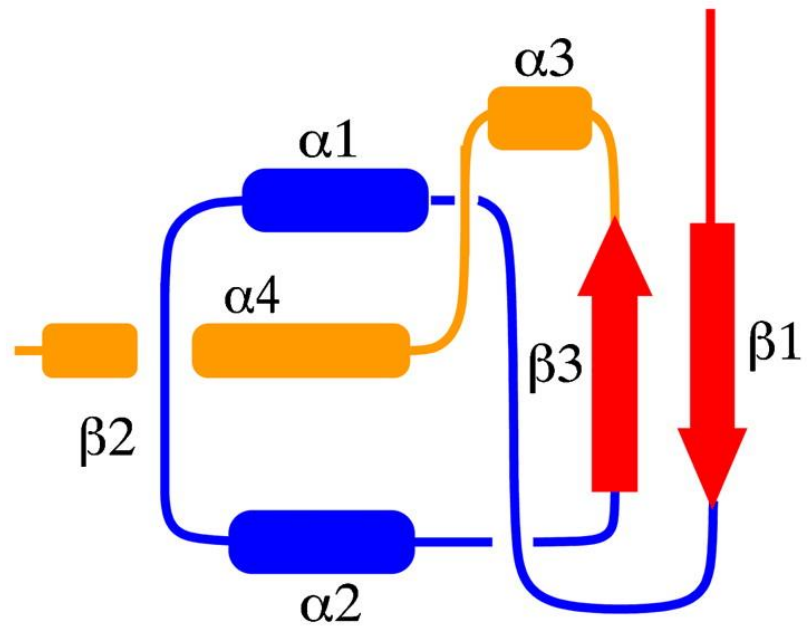
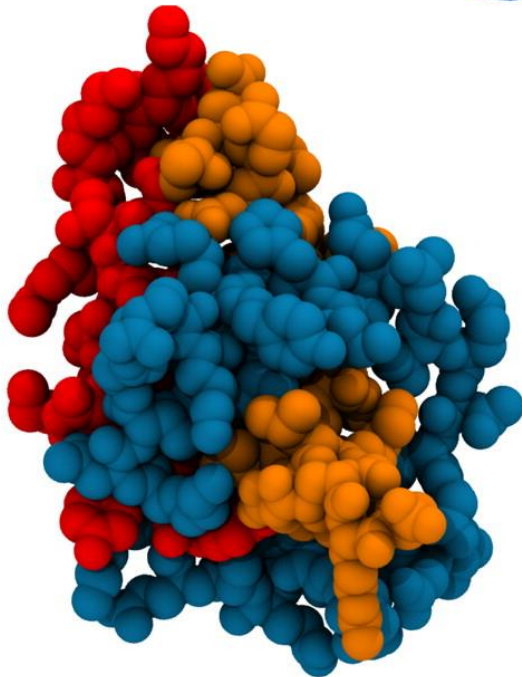
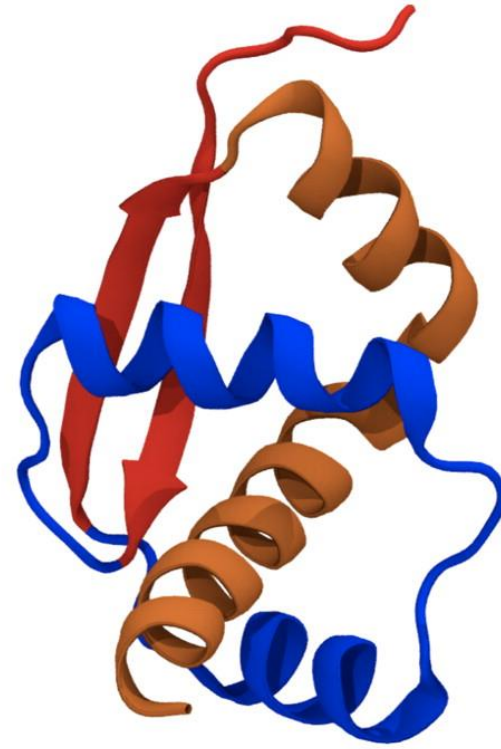


> Protein sequence
MKIVYWSGTGNTTEKMAELIAKGIIESGKDVNTINVSDVNI
DELLNEDILILGCSAMGDEVLEESEFEPFIEEISTKISGK
KVALFGSYGWGDGKWMRDFEERMNGYGCVVETPLIVQNE
PDEAEQDCIEFGKKIANI

Folha beta



Hélice alfa



Os aminoácidos estão para as proteínas assim como os nucleotídeos estão para o DNA

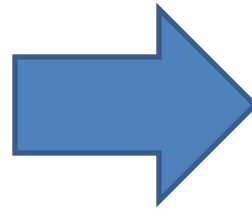
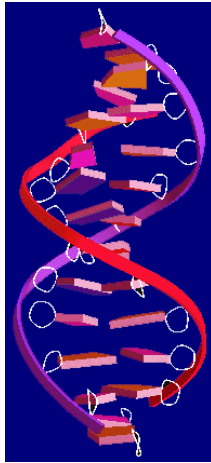
- Proteínas: 20 aminoácidos
- DNA: 4 nucleotídeos
- Ambas as cadeias têm **direcionalidade**
- DNA: 5' → 3'
- Proteína: Amino terminal (N) → Carboxi terminal (C)
- Proteínas são bem mais **curtas** do que DNA
- Protéina típica tem **300 aa**

Se 1 angstrom = 1 mm

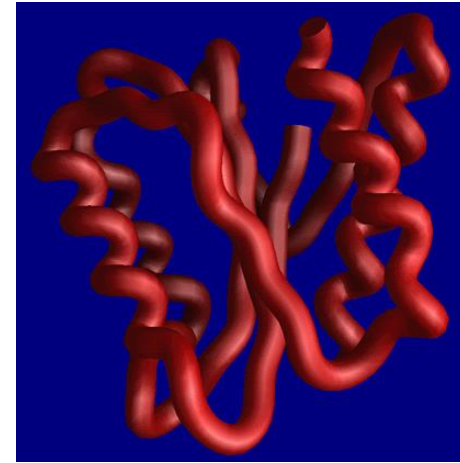
- Um cromossomo de 100.000.000 bp teria **340 km de comprimento**
- Uma proteína de 300 aa teria **3 metros**

Genes e proteínas

DNA



Proteína



> DNA sequence

```
ATGTCATGAAAATCGTATACTGGTCTGGTACCGGCAACAC
TGAGAAAATGGCAGAGCTCATCGCTAAAGGTATCATCGAA
TCTGGTAAAGACGTCAACACCATCAACGTGTCTGACGTTA
ACATCGATGAACTGCTGAACGAAGATATCCTGATCCTGGG
TTGCTCTGCCATGGGCGATGAAGTTCTCGAGGAAAGCGAA
TTTGAACCGTTCATCGAAGAGATCTCTACCAAATCTCTG
GTAAGAAGGTTGCGCTGTTTCGGTTCCTTACGGTTGGGGCGA
CGGTAAGTGGATGCGTGACTTCGAAGAACGTATGAACGGC
TACGGTTGCGTTGTTGTTGAGACCCCGCTGATCGTTCAGA
ACGAGCCGGACGAAGCTGAGCAGGACTGCATCGAATTTGG
TAAGAAGATCGCGAACATCTAGTAGA
```

> Protein sequence

```
MKIVYWSGTGNTTEKMAELIAKGIIESGKDVNTINVSDVNI
DELLNEDILILGCSAMGDEVLEESEFEPFIEEISTKISGK
KVALFGSYGWGDGKWMRDFEERMNGYGCVVVETPLIVQNE
PDEAEQDCIEFGKIANI
```


Como um pedaço de DNA (gene) pode gerar uma proteína?

- **Informacionalmente** por meio de um **código** (o famoso **código genético**)
- **Mecanicamente** por meio de processos celulares chamados de
 - **Transcrição**
 - **Tradução**
- Esta aula: o processo informacional

“Dogma Central” da Biologia Molecular

Replicação

DNA

Transcrição

RNA mensageiro

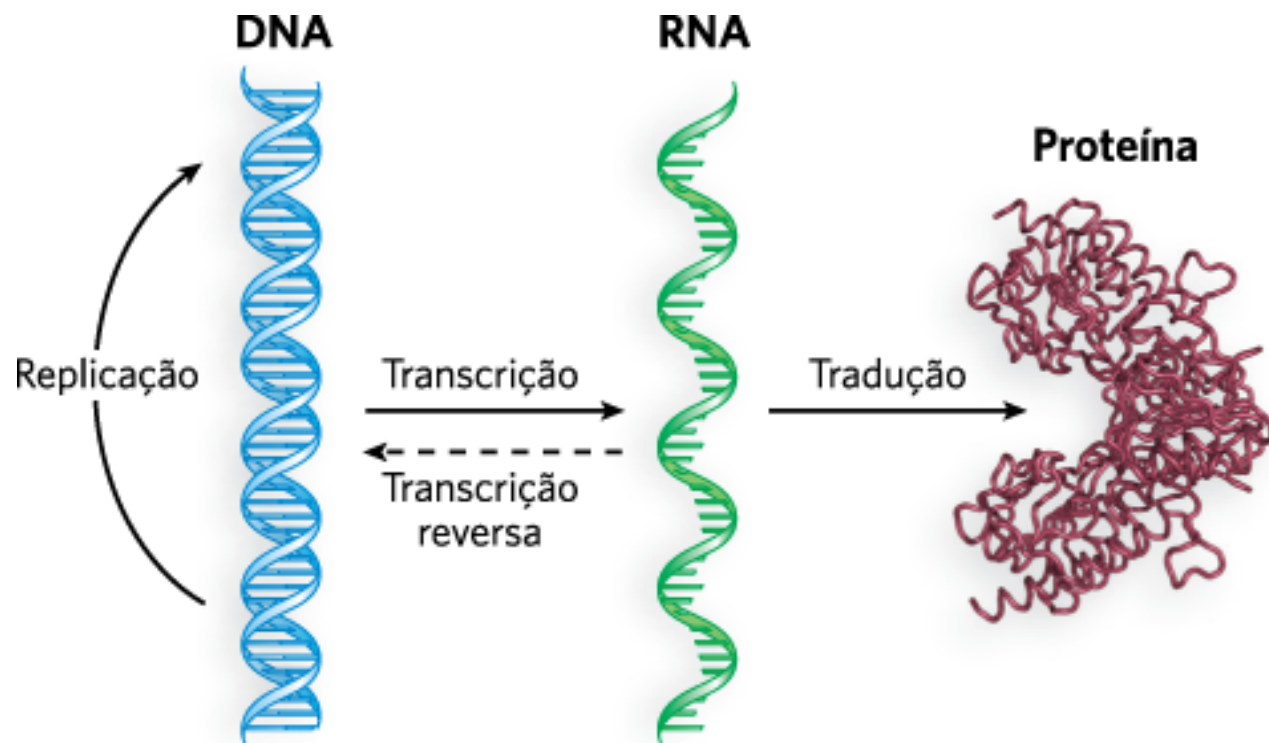
RNA

Tradução de mRNAs

Proteína

Usa **Uracila** ao invés de Timina

Ocorre no **ribossomo**



O dogma central do fluxo da informação:

DNA→**RNA**→**proteína**. A informação para replicar o DNA está inerente na sua estrutura (seta curvada). A informação flui do DNA para o RNA por transcrição. A informação flui do RNA para a proteína por tradução. Em alguns casos, a informação também pode fluir de volta, do RNA para o DNA (transcrição reversa). Não existe nenhuma evidência de informação que flui da proteína para o ácido nucleico.

Código Genético

- Funciona como uma tabela
- Nucleotídeos → Aminoácidos
- Semelhante ao código Morse

International Morse Code

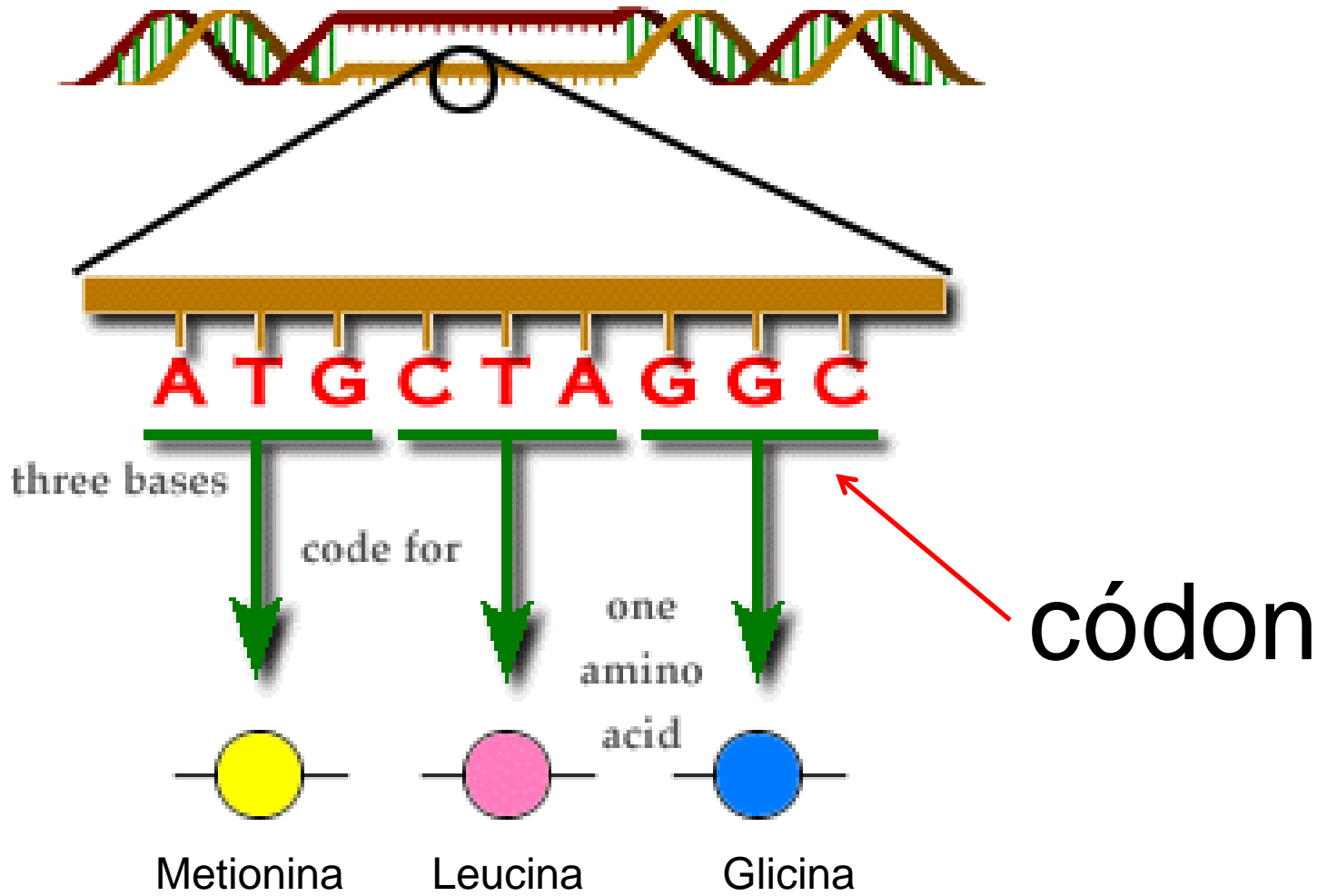
1. The length of a dot is one unit.
2. A dash is three units.
3. The space between parts of the same letter is one unit.
4. The space between letters is three units.
5. The space between words is seven units.

| | | | |
|---|---------|---|-----------|
| A | • — | U | • • — |
| B | — • • • | V | • • • — |
| C | — • — • | W | • — — |
| D | — • • | X | — • • — |
| E | • | Y | — • — — |
| F | • • — • | Z | — — • • |
| G | — — • | | |
| H | • • • • | | |
| I | • • | | |
| J | • — — — | | |
| K | — • — — | | |
| L | • — • • | | |
| M | — — | | |
| N | — • | | |
| O | — — — | | |
| P | • — — • | | |
| Q | — — • — | | |
| R | • — • | | |
| S | • • • | | |
| T | — | | |
| | | 1 | • — — — — |
| | | 2 | • • — — — |
| | | 3 | • • • — — |
| | | 4 | • • • • — |
| | | 5 | • • • • • |
| | | 6 | — • • • • |
| | | 7 | — — • • • |
| | | 8 | — — — • • |
| | | 9 | — — — — • |
| | | 0 | — — — — — |

Código genético

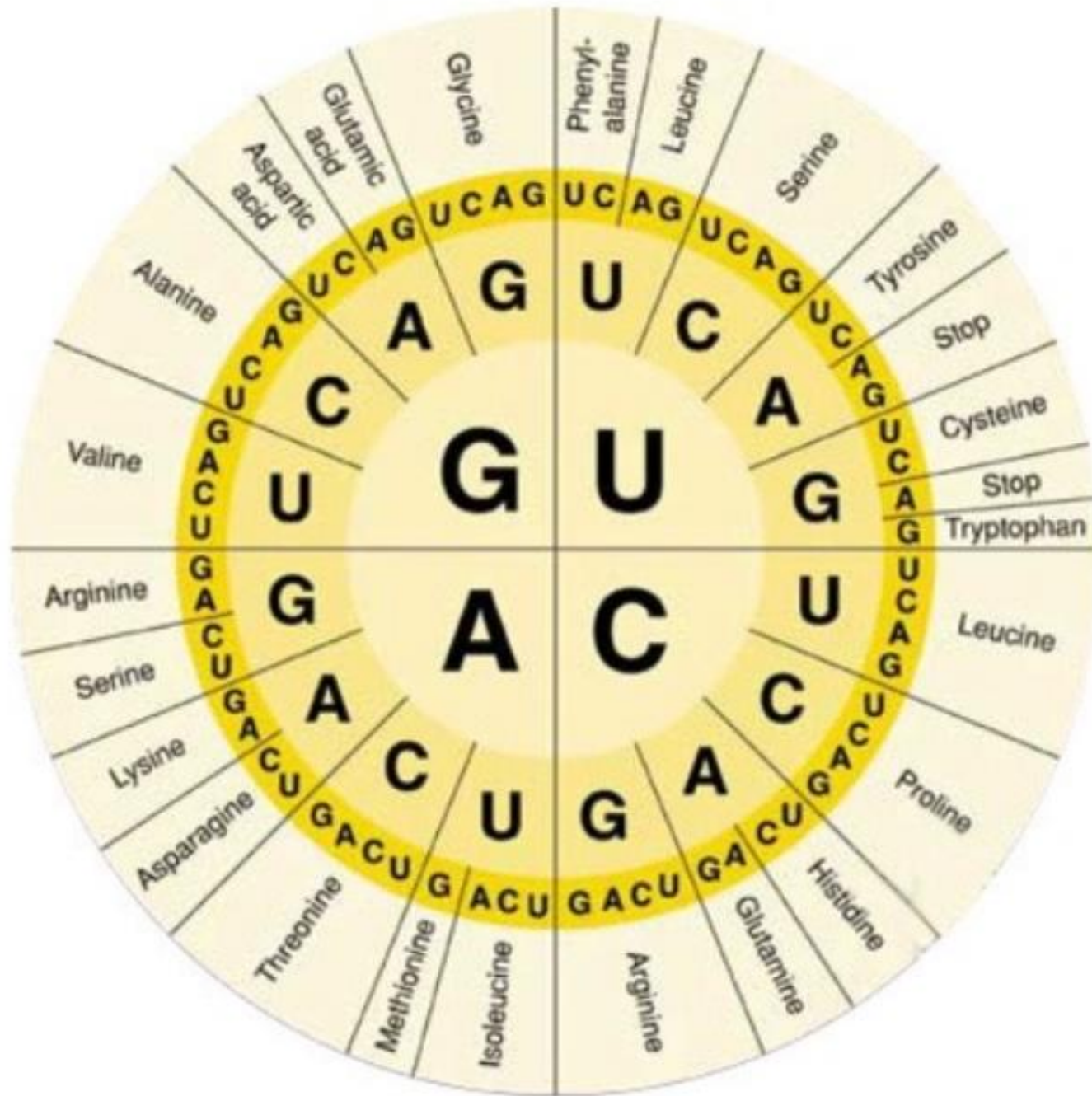
- 4 bases e 20 aminoácidos
- Um código 1:1 não dá
- Um código 2:1 também não dá
 - Apenas 16 possíveis pares
- Um código 3:1 dá (e sobra)
 - 64 possíveis **trincas**
- Lembrar que é preciso um sistema de pontuação (início e fim da região codificadora)

The Genetic Code



A tabela com o código genético

| 1st base | 2nd base | | | | | | | | 3rd base |
|----------|--------------------|-----------------------|-----|-------------------|-----|-----------------------|-----|--------------------|----------|
| | U | | C | | A | | G | | |
| U | UUU | (Phe/F) Phenylalanine | UCU | (Ser/S) Serine | UAU | (Tyr/Y) Tyrosine | UGU | (Cys/C) Cysteine | U |
| | UUC | | UCC | | UAC | | UGC | | C |
| | UUA | (Leu/L) Leucine | UCA | | UAA | Stop (Ochre) | UGA | Stop (Opal) | A |
| | UUG | | UCG | | UAG | Stop (Amber) | UGG | (Trp/W) Tryptophan | G |
| C | CUU | (Leu/L) Leucine | CCU | (Pro/P) Proline | CAU | (His/H) Histidine | CGU | (Arg/R) Arginine | U |
| | CUC | | CCC | | CAC | | CGC | | C |
| | CUA | | CCA | | CAA | (Gln/Q) Glutamine | CGA | | A |
| | CUG | | CCG | | CAG | | CGG | | G |
| A | AUU | (Ile/I) Isoleucine | ACU | (Thr/T) Threonine | AAU | (Asn/N) Asparagine | AGU | (Ser/S) Serine | U |
| | AUC | | ACC | | AAC | | AGC | | C |
| | AUA | | ACA | | AAA | (Lys/K) Lysine | AGA | (Arg/R) Arginine | A |
| | AUG ^[A] | ACG | AAG | | AGG | | G | | |
| G | GUU | (Val/V) Valine | GCU | (Ala/A) Alanine | GAU | (Asp/D) Aspartic acid | GGU | (Gly/G) Glycine | U |
| | GUC | | GCC | | GAC | | GGC | | C |
| | GUA | | GCA | | GAA | (Glu/E) Glutamic acid | GGA | | A |
| | GUG | | GCG | | GAG | | GGG | | G |



Exercício de tradução

- Dada uma sequência em DNA, mostrar sua tradução em aminoácidos
- **Informacionalmente**, podemos “pular” o passo de mRNA; ou seja, do DNA ir direto para proteína
- Na célula, nunca ocorre esse “pulo”

Exercícios

- Que aminoácidos são codificados por
TCGTCTGATATTCTA?
- E por
CGGCCCTGGCCTCCGACATCGGCGCC?

O código genético é quase **universal**
(o mesmo para todas as formas de vida)

- Em *Mycoplasma* (bactéria), UGA é **Trp**
- Em *Candida* (fungo), CUG é **Ser**
- ...e outras pequenas variações

O código genético é degenerado

| # codons | Aminoácidos | # aa | Total codons |
|----------|--|------|--------------|
| 6 | Leu, Ser, Arg | 3 | 18 |
| 4 | Ala, Thr, Pro, Gly, Val | 5 | 20 |
| 3 | Ile, Stop | 2 | 6 |
| 2 | Phe, Tyr, His, Gln, Asn, Lys, Asp, Glu, Cys | 9 | 18 |
| 1 | Met, Trp | 2 | 2 |
| Totais | | 20+1 | 64 |

A degeneração ocorre principalmente por meio da **terceira base**

| codon | AA |
|-------|----------|
| GG | Gly |
| CC | Pro |
| GC | Ala |
| CG | Arg (+2) |
| GU | Val |
| CU | Leu (+2) |
| UC | Ser (+2) |
| AC | Thr |

AU: 3 possibilidades correspondem a Isoleucina; a outra é Met
UA: 2 possibilidades correspondem a STOP e 2 a Tirosina

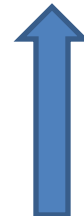
Degeneração significa **redundância**

- Robustez em relação a erros
- Mutações **sinônimas** (ou **silenciosas**)
- Mutações **não-sinônimas**

Mutação sinônima ou silenciosa

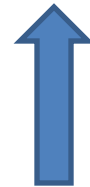
Wild Type DNA TAC GGG AAA GTC CGT GGC
Wild Type mRNA AUG CCC UUU CAG GCA CCG
Amino acids Met -Pro- Phe- Gln- Ala- Pro

Mutated DNA TAC GGG AAG GTC CGT GGC
Mutated mRNA AUG CCC UUC CAG GCA CCG
Amino acids Met -Pro- Phe- Gln- Ala- Pro



Mutação não-sinônima

| | | | | | |
|----------------|-----|-------------|------|------|------|
| Wild Type DNA | TAC | CAC | CCC | GCC | ATC |
| Wild Type mRNA | AUG | GUG | GGG | CGG | UAG |
| Amino acids | Met | -Val- | Gly- | Arg- | Stop |
| Mutated DNA | TAC | <u>G</u> AC | CCC | GCC | ATC |
| Mutated mRNA | AUG | <u>C</u> UG | GGG | CGG | UAG |
| Amino acids | Met | -Leu- | Gly- | Arg- | Stop |



Mutações são boas ou más?

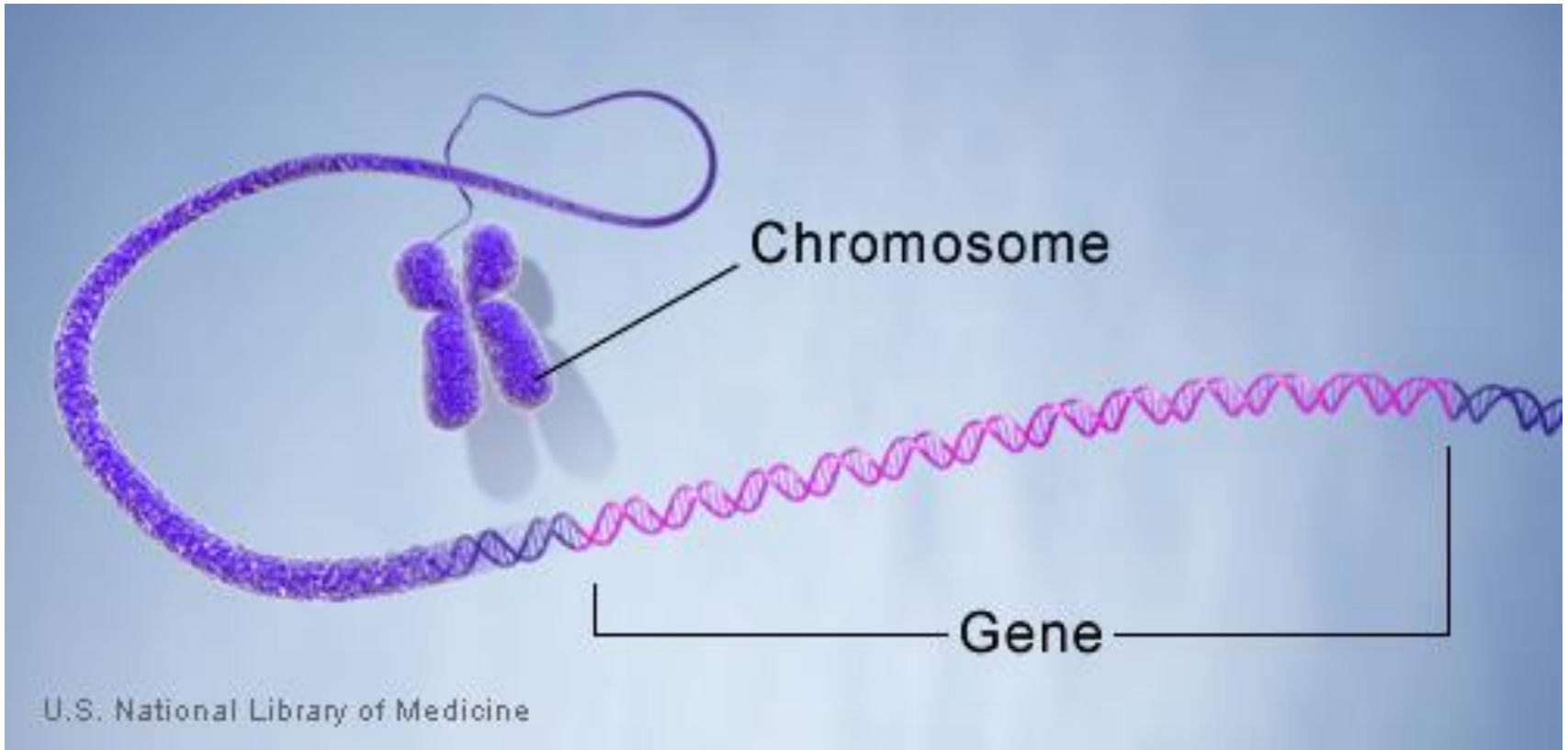
- Neutras
 - exemplo: mutação sinônima
- Más
 - Quando causam doença
- Boas
 - Quando dão uma vantagem competitiva ao indivíduo
 - Combustível da evolução!
- Algumas mutações são boas e más ao mesmo tempo!

Anemia falciforme

- Mutação não sinônima **numa única posição** de hemoglobina (amino ácido num. 7)
- GAA (glu) → GUA (val)
- GAG (glu) → GUG (val)
- Valina é **hidrofóbica** e Ácido glutâmico **não é**

A mutação de Af é boa e má ao mesmo tempo

- Má: Anemia falciforme é uma doença
- Boa: Indivíduos com essa mutação tem **proteção** contra **malária!**
- Anemia falciforme é prevalente nas regiões da África que historicamente foram (e são) afetadas pela malária

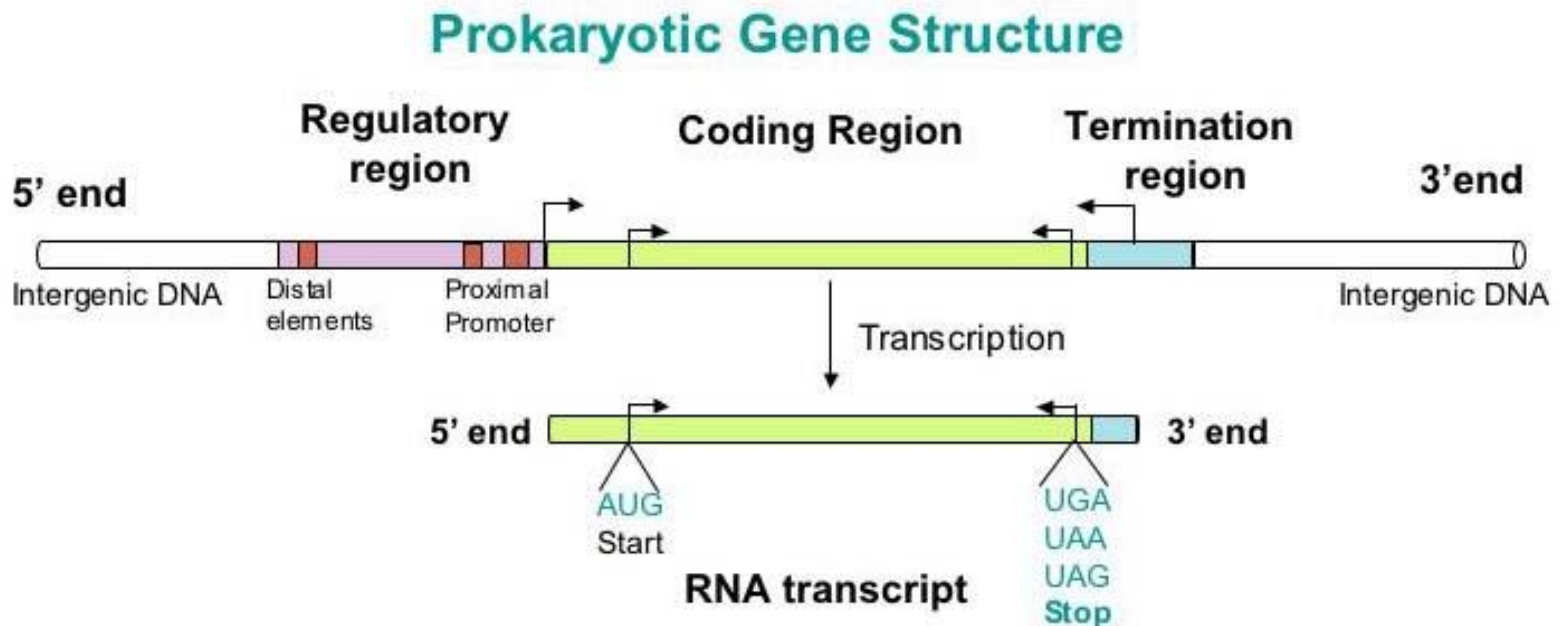


Um gene numa sequência

AGCTCGCGCTCCGCATCCATCCAGTAGGGTTCGGTGTGACGAGCGTGCC
GTCCATATCCCAGAAGACGGCGGCCGGCATCGCGTGCGGAGTCAGTTCGG
TCACGGCTGACAAGTCTATCCCGCGGCCCGGGCCTATTCTTGAGGGAC
GGCGTCCTGACCGGTGCGCCGATGAAAGGACCAGAACGCCCCGTGACTGA
CGCGAACAGCATCCTC**GGAGGGCGCATCCTCATGGTGGCCTTCGAAGGGT**
GGAACGACGCTGGCGAGGCCGCCAGCGGGGCCGTCAAGACGCTCAAGGAC
CAGCTGGATGTCGTCCCGTTCGCCGAGGTTCGATCCCGAGCTGTA
CTTCCAGTTCAACCGGCCGGTTCGTTCGCCGACGACGACGGCCGCCGGCGCC
TCATCTGGCCGTCCGCCGAGATCCTGGGCCAGCTCGCCCCGGCGACACC
GGCGATGCGCGCCTGGACGCCACCGGCCCAACGCGGGCAATATCTTCCT
TCTCCTCGGCACCGAGCCGTTCGCCGAGCTGGCGCAGCTTACCGCGGAGA
TCATGGATGCGGCCCTGGCCTCCGACATCGGCGCCATCGTCTTCCTCGGT
GCGATGCTGGCGGACGTACCGCACACCCGCCCATCTCCATCTTCGCTTC
GAGCGAGAACGCGGCCGTCCGTGCGGAGCTCGGCATCGAACGCTCTTCGT
ACGAGGGGCCGGTTCGGTATCCTGAGCGCGCTCGCCGAAGGGGGCGGAGGAC
GTGGGCATTCCGACCATCTCCATCTGGGCGTTCGGTTCCGCACTATGTCCA
CAATGCGCCCAGCCCCGAAGGCGGTGCTCGCACTGATCGACAAGCTCGAAG
AGCTGGTGAATGTCACCATCCCGCGTGGCTCGCTGGTGGAGGAGGCCACG
GCCTGGGAAGCCGGGATCGACGCGCTGGCTCTGGACGACGACGAGATGGC
TACGTACATCCAGCAGCTGGAGCAGGCACGCGACACCGTGGACTCCCCTG
AGGCCAGCGGCGAGGCGATCGCCAGGAGTTCGAGCGCTACCTCCGCCGC
CGCGACGGCCGCGCCGGCGATGACCCCCGCCGTGGCTGACGTCACCCCT
CTCTGCGTCCGCCGTCTCTGTTCCCCCGCTCGGCCTCCCCTGAGGCCG
AGGAGTCGCGCCACATGCCGAACTCCTCCTTTCTGACTTTCTGGAG

Início e fim da porção codificadora de um gene de procarionto

- O início é quase sempre um **ATG = metionina**
- O final é sempre um **codon de parada**



Sequenciamento de DNA

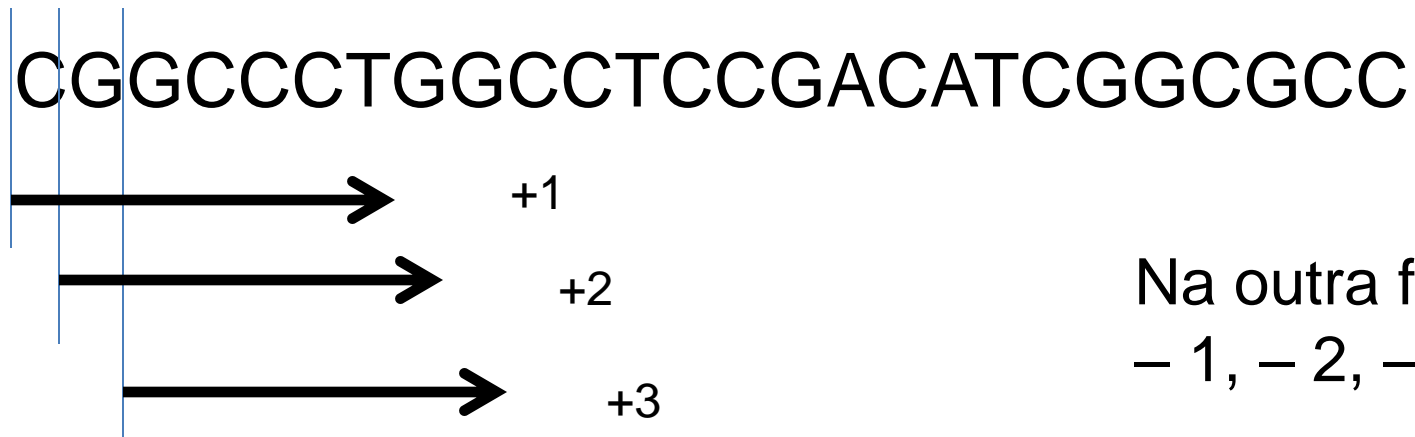
- Será visto em detalhe em outra aula
- Gera imensas quantidades de sequências de DNA, armazenadas em arquivos de computadores
- Em geral são fragmentos onde a informação da fita e da localização no cromossomo foi perdida
- Para achar onde estão os genes de forma computacional, precisamos do conceito de **Quadro de Leitura**

Quadros de leitura

Uma fita dupla de DNA admite

6 quadros de leitura

(3 em cada fita)



Cada quadro tem sua própria tradução

V L L A * S K K S N L T Y Y L V L F Y S I # N Y H N N L H L S N
A C S W H D Q R R V I # L I I + C C F I P Y K I I I I T Y T # V
R A L G M I K E E # F N L L F S A V L F H I K L S # # L T L K #
CGTGCTCTGGCATGATCAAAGAAGAGTAAITTAACCTTATTATTAGTGCTGTTTTATTCCATATAAAAATTATCATAATAACTTACACTTAAGTAA
41240 41260 41280 41300 41320
GCACGAGAACCGTACTAGTTTCTTCTCATTAAATTGAATAATAAATCACGACAAAATAAGGTATATTTAATAGTATTATTGAATGTGAATTCATT
L H E Q C S * L L T I # S I I # H Q K I G Y L I I M I V # V # T I
R A R P M I L S S Y N L K N N L A T K N W I F N D Y Y S V S L Y
T S K A H D F F L L K V # # K T S N # E M Y F # * L L K C K L L

stop

stop

Fita de cima

Coordenada dentro deste segmento

Fita de baixo

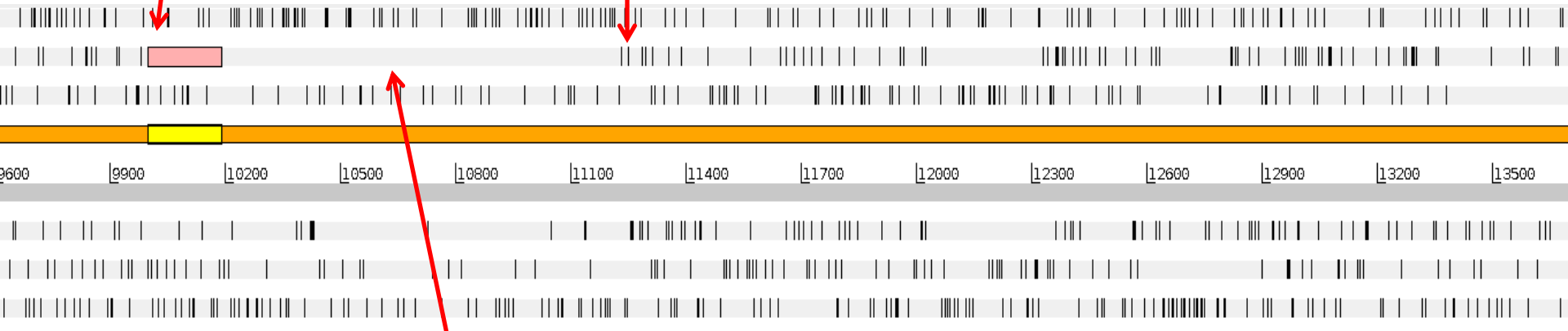
Quadro aberto de leitura

- *Open reading frame (ORF)*
- É um quadro de leitura
 - Com número de bases **múltiplo de 3**
 - Terminando em STOP
 - Sem outros STOPS no meio
- A **porção codificadora** (de proteína) de um **gene bacteriano** é um quadro aberto de leitura iniciado por **ATG** (muito mais raramente por GTG ou CTG)

início

fim

Tracinhos verticais são stops



L F N A F I + D I L I # D K R N Y F I * # S K F A N H F L L F Y N S I I D N H I F Y Y I F G * F I S # K N S N * Y Y D Q F C H I I
F L M R L F R I F L S K I K G I I L S D K A S L L I I F C Y S I I P L L I I I F F I I S L V D L Y H K K I A I D T T I N S V I L
T F # C V Y L G Y S Y L R # K E L F Y I K Q V C # S F F V I L # F H Y * # S Y F L L Y L W L I Y I I K K + Q L I L R S I L S Y Y
C T T T A A T G C G T T A T T A G G A T A T C T T A T C T A A G A T A A A G G A A T T A T T T A T C T G A T A A G C A A G T T G C T A A T C A T T T T T G T T A T T C T A T A A T C C A T T A T T G A T A A T C A T A T T T T A T T A T A T C T T T G G T T G A T T A T A T C A T A A A A A A A T A G C A A T T G A T A C T A C G A T C A A T T C T G T C A T A T T A
10000 10020 10040 10060 10080 10100 10120 10140 10160 10180
G A A A A A T T A C G C A A A T A A A T C C T A T A A G A A T A G A T T C A T T T T C C T T A A T A A A A T A G A C T A T T T C G T T C A A A C G A T T A G T A A A A A A C A A T A A G A T A T T A A G G T A A T A A C T A T T A G T A T A A A A A A T A A T A T A G A A A C C A A C T A A A T A T A G T A T T T T T A T C G T T A A C T A T G A T G C T A G T T A A G A C A G T A T A A T A
K K I R K N L I N K D L I F P I I K D S L A L K S I M K Q # E I I G N N I I M N K I I D K T S K Y * L F I A I S V V I L E T M N D
K # H T # K P Y E # R L Y F S N N # R I F C T Q + D N K T I R Y N W # Q Y D Y K K N Y R Q N I # I M F F Y C N I S R D I R D Y # I
S K L A N I # S I R I I + S L L F # K I Q Y L L N A L * K K N N + L E M I S L * I K # # I K P Q N I D Y F F L L Q Y + S * N Q * I I

Exemplo de uma ORF bacteriana

Estas imagens vieram do navegador de genomas ARTEMIS

Disponível (de graça) em

<https://www.sanger.ac.uk/resources/software/artemis/>

Exercício

- Dada uma sequência de DNA, achar uma ORF fazendo a tradução nos 6 quadros de leitura
- Verifique o resultado usando

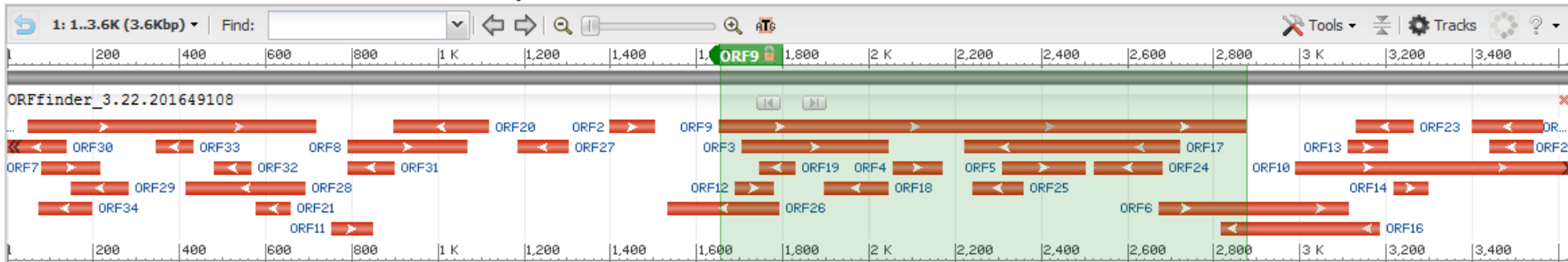
<https://www.ncbi.nlm.nih.gov/orffinder/>

ORFfinder PubMed Search

Open Reading Frame Viewer

Sequence

ORFs found: 34 Genetic code: 1 Start codon: 'ATG' only



Add six-frame translation track

ORF9 (407 aa) Display ORF as...

```
>1c1|ORF9
MLSPACFPVIPGGHVCTVSCRSRILRTDRHAGLQRAHRSM
HVVRLSIHRLRRFQTVLHPSALNLLTGDNGAKTSVLE
ALHLMAYGRSFRGRVVDGLIQQGANDLEVFVEWKEGGAA
VERTRRAGLRHSGQEWIGRLDGEDVAQLGSLCAALAVTF
EPGSHVLISGGGEPRRRFLDWGLFHVEPDFLTLWRRYARA
LKQRNALLKQGAQPRMLDAWDNELAESGETLTSRRMRYLE
RLQDRLVPVADAIAPALGLSALTFAPGWKRHEVSLADALL
LARERDRQNGYTSQGPHRADWMPSPHALPGKDALSQGAQ
LTALACLLAQAEFAFERGEWPFVIALDDLGSELDRHHQGR
VLQRLASAPAQVLITATETPPGLADAAALLQQFHVEHGQI
ARQATVN
```

BLAST Database:

Mark subset... Marked: 0 Download marked set as Protein FASTA

| Label | Strand | Frame | Start | Stop | Length (nt aa) |
|-------|--------|-------|-------|-------|------------------|
| ORF9 | + | 2 | 1652 | 2875 | 1224 407 |
| ORF1 | + | 1 | 49 | 717 | 669 222 |
| ORF10 | + | 2 | 2990 | >3640 | 651 216 |
| ORF17 | - | 1 | 2722 | 2222 | 501 166 |
| ORF6 | + | 1 | 2671 | 3114 | 444 147 |
| ORF16 | - | 1 | 3184 | 2816 | 369 122 |
| ORF3 | + | 1 | 1705 | 2046 | 342 113 |
| ORF8 | + | 2 | 791 | 1069 | 279 92 |
| ORF28 | - | 2 | 693 | 415 | 279 92 |
| ORF26 | - | 2 | 1701 | 1524 | 258 85 |