

**REPÚBLICA  
DA CIÊNCIA  
DIA ABERTO  
INSTITUTO  
GULBENKIAN  
DE CIÊNCIA**

**9.10.2010  
10-17h**

**Em  
busca  
de um  
gene  
mistério!**

Designways | www.designways.pt



**dia  
aberto 2010**

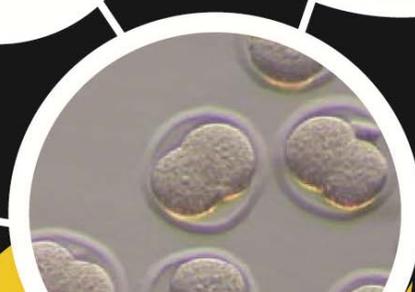
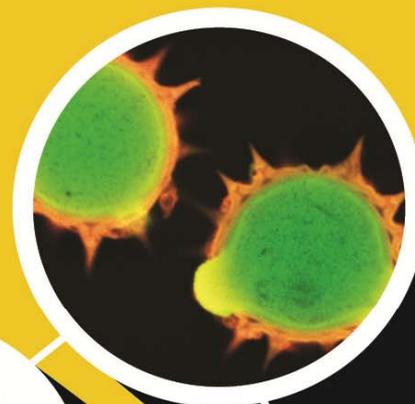
Instituto  
Gulbenkian  
de Ciência



FUNDAÇÃO CALOUSTE GULBENKIAN

Instituto Gulbenkian de Ciência

Bioinformatics and Computational Biology Unit





# I – Como é que o DNA codifica para proteínas?



## Sabia que:

- O **DNA** é constituído por fiadas de quatro nucleótidos diferentes, representados pelas letras A G T C, em que A é adenina, G guanina, T timina e C citosina?
- O DNA tem informação para fazer **proteínas**, a qual é ditada pela ordem das 4 letras ao longo do DNA?
- O DNA tem **sinais** que indicam onde começam e acabam as proteínas? O ATG (**codão de iniciação**) marca o início da proteína?
- As proteínas são fiadas de **aminoácidos**?
- Cada aminoácido é o resultado da descodificação de três nucleótidos (**codão**) do DNA? Existem 20 aminoácidos diferentes?
- Um **gene** é um segmento de DNA, que codifica *uma única proteína*?
- Os genes são herdados dos pais e determinam as **características genéticas** do indivíduo?

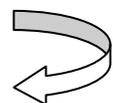
Agora que já tem estes conceitos  
siga as pistas para encontrar o  
gene mistério!

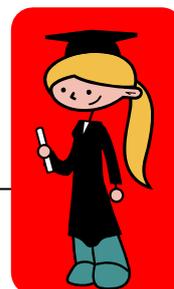
A primeira tarefa consiste em identificar o **ATG** no fragmento de uma sequência de DNA, que faz parte da sequência do gene mistério:

1 – Descubra o ATG na sequência mistério e faça um círculo à sua volta.

```
CCATGGTGGTGGTGATGATCTTTGCGTACTG  
CGTCTGCTGGGGACCCTACACCTTCTTCGCA
```

Quer saber mais sobre o código genético? Volte a folha!





## Quer saber mais sobre o código genético?

Os genes codificam instruções para fazer proteínas. Mas como é que o fazem? As proteínas são constituídas por sequências de aminoácidos, mais ou menos longas, tal como um colar feito de missangas.

Existem 20 aminoácidos diferentes e a cada um é dado uma letra de código. Por exemplo a metionina tem o código M, a Leucina L, a fenilalanina F, a prolina P, etc.

Um gene pode ser visto como uma lista de letras código para uma cadeia de aminoácidos: cada letra é constituída por três bases de DNA (codão). A maior parte das bases do ADN codifica para um aminoácido. Por exemplo, as letras TGC do DNA codificam para o aminoácido cisteína, enquanto que as letras TGG codificam para o aminoácido triptofano.

O DNA é constituído por fiadas de quatro nucleótidos diferentes, representados pelas letras A G T C, em que A é adenina, G guanina, T timina e C citosina. Com estas quatro letras podemos ter  $4 \times 4 \times 4 = 64$  combinações diferentes e cada uma pode ser usada para fazer um codão. Algumas destas 64 combinações codificam para o mesmo aminoácido. Por exemplo, GAA e GAG codificam ambos para o aminoácido glutamato E.

O codão ATG, que codifica para a metionina (M), é muito importante porque significa o início de uma proteína.

Os codões TAA, TGA e TAG, não codificam para qualquer aminoácido. São de extrema importância para a célula, pois estão responsáveis por indicar onde deve parar a leitura da proteína que é codificada a partir do DNA. Estes codões têm o nome de **codões stop**.

## II – Como ler a informação do DNA que vem a seguir ao ATG?

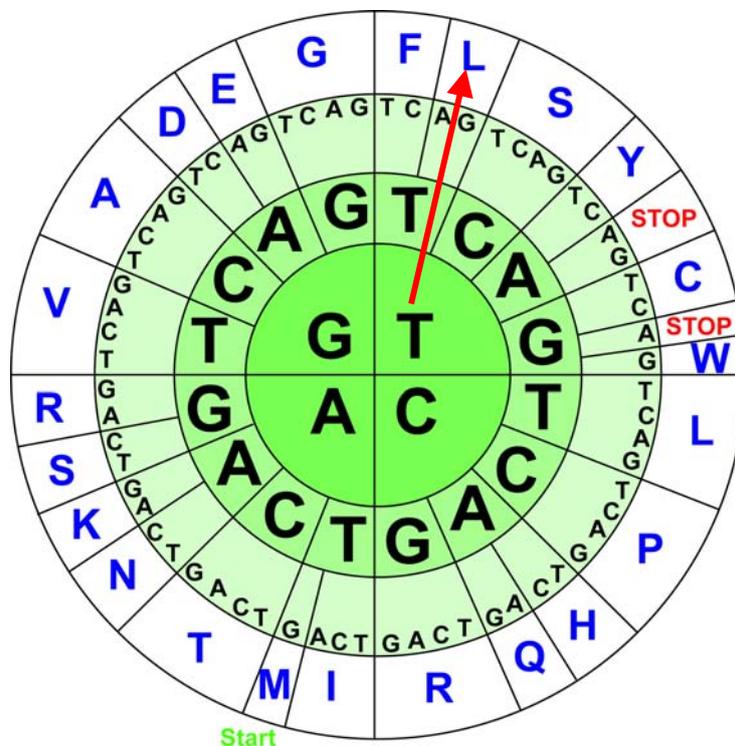
### 2 – Como descodificar o DNA?

Três nucleótidos codificam para um aminoácido.

Consegue preencher os quadrados vazios com o aminoácido correspondente, usando a roda do código genético?

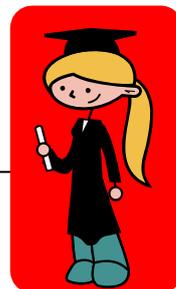
<b>Posição</b>	1	2	3	4	5	6	7	8	9	10
<b>Codão</b>	ATG	GTG	GTG	GTG	ATG	ATC	TTT	GCG	TAC	TGC
<b>Aminoácido</b>	M									
<b>Posição</b>	11	12	13	14	15	16	17	18	19	20
<b>Codão</b>	GTC	TGC	TGG	GGA	CCC	TAC	ACC	TTC	TTC	GCA
<b>Aminoácido</b>										

**Como usar a roda:** Se, por exemplo, quiser saber qual o aminoácido codificado pela sequência de DNA **TTG**, terá de começar no centro da roda, na posição **T**, e encontrar o **T** e o **G** adjacentes, (seta). O anel exterior indicar-lhe-á a letra L, e esta é o código correspondente ao aminoácido leucina.



Quer saber mais sobre a síntese das proteínas? Volte a folha!





## Quer saber mais sobre a síntese das proteínas?

Ficou a saber como uma sequência de ADN codifica para proteínas e como se lê o código genético.

No final das tarefas propostas e se tiver tempo, pode consultar os seguintes endereços da Internet para ficar a saber mais sobre como são feitas as proteínas:

- <http://tinyurl.com/traducao-ribo>
- <http://tinyurl.com/s-ntese-proteica>

### III – Como posso saber qual é a proteína que contém estes aminoácidos?



#### Sabia que:

- Existem **bases de dados** onde as sequências de DNA e de proteínas estão guardadas?
- As bases de dados biológicos podem ser consultadas **livremente** através da Internet?
- É possível **comparar** as nossas sequências em estudo com as que já estão depositadas em bases de dados?
- Os cientistas que desenvolvem os programas que permitem recolher, armazenar e analisar os dados biológicos chamam-se **bioinformáticos**?

...continuando na pista do gene mistério...

#### Qual será a proteína que contém os aminoácidos que identificou?

Um programa chamado **Blast** vai ajudá-lo nesta tarefa:



Entre na página do programa Blast: <http://www.ebi.ac.uk/Tools/sss/ncbiblast/>

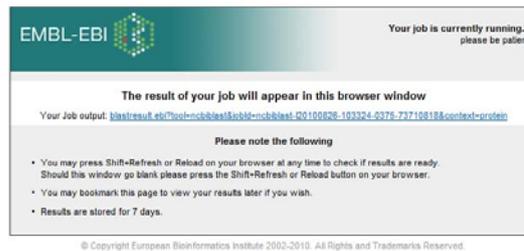
1) Escreva aqui a sequência de **aminoácidos** que determinou para a sequência de DNA. Verifique se escreveu correctamente a sequência!

2) Seleccione **UniProtKB/Swissprot** do menu DATABASE

3) A seguir clique no botão **Submit**.



.... **Seja paciente! O Blast leva algum tempo a dar o resultado da pesquisa...**  
Entretanto, no ecrã do seu computador, vai ter esta mensagem de espera:



1) Ao fim de alguns segundos o Blast terminou a pesquisa.

2) Com a ajuda do  faça descer a informação contida na janela. Irá aparecer uma tabela com o seguinte aspecto:

Align.	DB:ID	Source	Length	Score	Identities	Positives	E()
<input checked="" type="checkbox"/>	1 SP:OPSR_HUMAN	Red-sensitive opsin OS=Homo sapiens GN=OPN1LW PE=1 SV=1 Cross-references and related information in: ▶ Gene Expression ▶ Nucleotide Sequences ▶ Genomes ▶ Ontologies ▶ Protein Families ▶ Literature ▶ Macromolecular Structures ▶ Reactions & Pathways	384	117	100.0	100.0	4.0E-5
<input checked="" type="checkbox"/>	2 SP:OPSR_FELCA	Red-sensitive opsin OS=Felis catus GN=OPN1LW PE=2 SV=2 Cross-references and related information in: ▶ Nucleotide Sequences ▶ Genomes ▶ Ontologies ▶ Protein Families ▶ Literature	384	114	95.0	100.0	9.0E-5
<input checked="" type="checkbox"/>	3 SP:OPSG_SCICA	Green-sensitive opsin OS=Sciurus carolinensis GN=OPN1MW PE=2 SV=2 Cross-references and related information in: ▶ Nucleotide Sequences ▶ Ontologies ▶ Protein Families ▶ Literature	384	113	90.0	100.0	1.0E-5
<input checked="" type="checkbox"/>	4 SP:OPSG_RAT	Green-sensitive opsin OS=Rattus norvegicus GN=Opn1mw PE=1 SV=2 Cross-references and related information in: ▶ Nucleotide Sequences ▶ Ontologies ▶ Protein Families ▶ Literature	359	113	90.0	100.0	1.0E-5
<input checked="" type="checkbox"/>	5 SP:OPSG_RABIT	Green-sensitive opsin OS=Oryctolagus cuniculus GN=OPN1MW PE=1 SV=2 Cross-references and related information in: ▶ Nucleotide Sequences ▶ Genomes ▶ Ontologies ▶ Protein Families ▶ Literature	384	113	90.0	100.0	1.0E-5
<input checked="" type="checkbox"/>	6 SP:OPSG_MOUSE	Green-sensitive opsin OS=Mus musculus GN=Opn1mw PE=1 SV=1 Cross-references and related information in: ▶ Gene Expression ▶ Nucleotide Sequences ▶ Genomes ▶ Ontologies ▶ Protein Families ▶ Literature	359	113	90.0	100.0	1.0E-5
<input checked="" type="checkbox"/>	7 SP:OPSR1_DANRE	Red-sensitive opsin-1 OS=Danio rerio GN=opn1lw1 PE=1 SV=1 Cross-references and related information in: ▶ Gene Expression ▶ Nucleotide Sequences ▶ Genomes ▶ Ontologies ▶ Protein Families ▶ Literature	357	112	95.0	95.0	2.0E-5
<input checked="" type="checkbox"/>	8 SP:OPSR_CAPHI	Red-sensitive opsin OS=Capra hircus GN=OPN1LW PE=1 SV=1 Cross-references and related information in: ▶ Nucleotide Sequences ▶ Ontologies ▶ Protein Families ▶ Literature	384	111	90.0	100.0	2.0E-5
<input checked="" type="checkbox"/>	9 SP:OPSR_BOVIN	Q9BG17 Red-sensitive opsin OS=Bos taurus GN=OPN1LW... Cross-references and related information in: ▶ Nucleotide Sequences ▶ Genomes ▶ Ontologies ▶ Protein Families	384	111	90.0	100.0	2.0E-5

Não se preocupe se à primeira vista esta tabela lhe parecer difícil:

- No lado esquerdo da tabela, a coluna **DB:ID** mostra os nomes de cada uma das proteínas encontradas na base de dados.
- A seguir, a coluna **Source** mostra uma pequena descrição de cada uma das proteínas.
- No lado direito e em frente de cada nome/descrição das proteínas, aparecem uns números que indicam o grau de semelhança encontrado entre cada proteína da base de dados e a sua proteína mistério.
- O primeiro resultado da lista é a **proteína** mistério.

3 – Qual o nome da proteína mistério? \_\_\_\_\_



Siga o **link** associado a esta proteína (SW:OPSR\_HUMAN).

## IV – Onde podemos obter informação sobre esta proteína?

O *link* leva-o para uma base de dados de proteínas chamada **Swiss-Prot**:

- A Swiss-Prot é como se fosse uma **enciclopédia** de proteínas.
- Toda a informação sobre cada uma das proteínas, cuja sequência de aminoácidos é já conhecida, é guardada aqui.

Uma entrada na **Swiss-Prot** tem o seguinte aspecto:

<b>P04000</b> (OPSR_HUMAN) ★ Reviewed, UniProtKB/Swiss-Prot Last modified August 10, 2010. Version 113. <a href="#">History...</a>	
<a href="#">Clusters with 100%, 90%, 50% identity</a>   <a href="#">Documents (7)</a>   <a href="#">Third-party data</a>	
<a href="#">Customize display</a>   <a href="#">Names</a>   <a href="#">Attributes</a>   <a href="#">General annotation</a>   <a href="#">Ontologies</a>   <a href="#">Sequence annotation</a>   <a href="#">Sequences</a>   <a href="#">References</a>   <a href="#">Web links</a>   <a href="#">Cross-refs</a>   <a href="#">Entry info</a>   <a href="#">Documents</a>	
<b>Names and origin</b>	
Protein names	<i>Recommended name:</i> <b>Red-sensitive opsin</b> <i>Short name:</i> RCP <i>Alternative name(s):</i> Red cone photoreceptor pigment
Gene names	Name: <b>OPN1LW</b> Synonyms:RCP
Organism	<b>Homo sapiens (Human)</b> [Complete proteome]
Taxonomic identifier	<b>9606</b> [NCBI]
Taxonomic lineage	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominoidea > Homo
<b>Protein attributes</b>	
Sequence length	364 AA.
Sequence status	Complete.
Protein existence	Evidence at protein level.
<b>General annotation (Comments)</b>	
Function	Visual pigments are the light-absorbing molecules that mediate vision. They consist of an apoprotein, opsin, covalently linked to cis-retinal.
Subcellular location	<a href="#">Membrane</a> ; <a href="#">Multi-pass membrane protein</a> .
Tissue specificity	The three color pigments are found in the cone photoreceptor cells.
Post-translational modification	Phosphorylated on some or all of the serine and threonine residues present in the C-terminal region.
Involvement in disease	Defects in OPN1LW are the cause of partial colorblindness protan series (CBP) [MIM:303900]; also known as protanopia. <a href="#">Ref 5</a>
Sequence similarities	Belongs to the <a href="#">G-protein coupled receptor 1 family</a> . <a href="#">Opsin subfamily</a> .
Biophysicochemical properties	<a href="#">Absorption</a> : Abs(max)=560 nm



Esta página contém muita informação sobre a proteína mistério.  
Veja se descobre:

**4 - O número de entrada (*Accession Number*) da proteína na Swiss-Prot**

\_\_\_\_\_

**5 - O organismo a que pertence esta a proteína**

\_\_\_\_\_

**6 - O nome do gene mistério**

\_\_\_\_\_



## V- Onde se localiza o gene desta proteína?

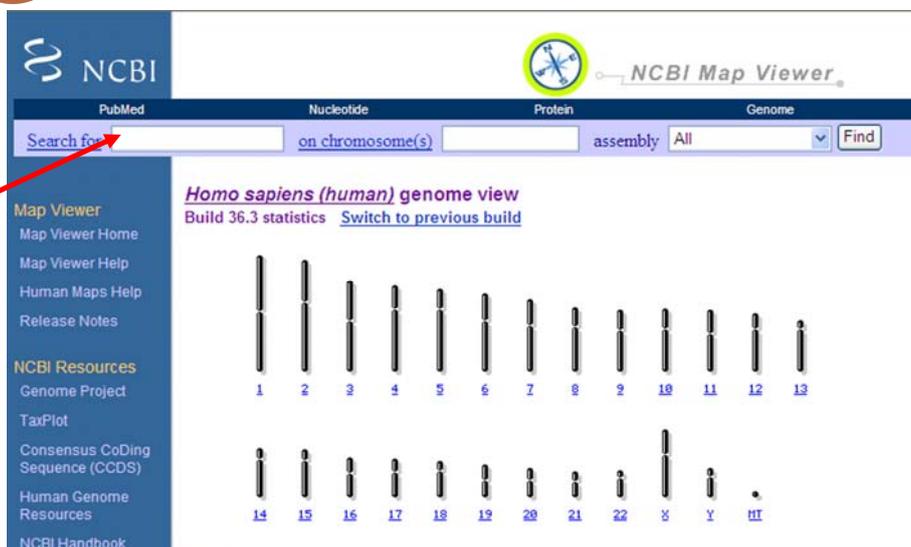


**Sabia que:**

- A sequência do **Genoma Humano** foi conhecida em Abril de **2003**?
- O DNA humano tem **3 mil milhões de nucleótidos**?
- O DNA humano tem informação para cerca de **25 mil genes**?



Entre na página: <http://tinyurl.com/viewhuman>



- Escreva aqui o nome do gene mistério e clique no botão **Find** que se encontra à direita.



Veja se descobre:

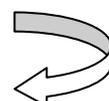
**7 – O cromossoma em que se localiza o gene mistério:**

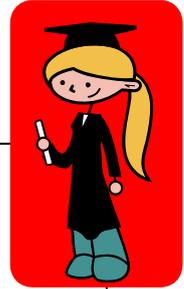
\_\_\_\_\_

**8 – Quantos pares de cromossomas existem numa célula humana?**

\_\_\_\_\_

Quer saber mais sobre cromossomas? Volte a página!





## Quer saber mais sobre cromossomas?

Um cromossoma não é mais do que uma longa molécula de DNA associada a proteínas.

Cada espécie tem um número específico de cromossomas.

Herdámos um conjunto de cromossomas do nosso pai e outro da nossa mãe e portanto temos dois conjuntos.

Os cromossomas foram numerados de acordo como seu tamanho, sendo o cromossoma 1 o maior de todos. Isto só não se verifica em dois cromossomas especiais, chamados X e Y, que determinam o nosso sexo.

Os rapazes herdaram o cromossoma X da mãe e o Y do pai. As raparigas herdaram um cromossoma X da mãe e outro do pai.

## VI – A estrutura 3D da proteína mistério



**Sabia que:**

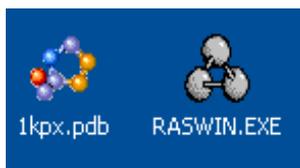
- A função de uma proteína está associada à sua **configuração** (estrutura terciária)?
- A **mutação num gene** pode alterar a configuração e a proteína **perder a função**?
- A configuração é determinada pela **ordem dos aminoácidos** ao longo da proteína?

### Como posso ver a estrutura 3D da proteína?



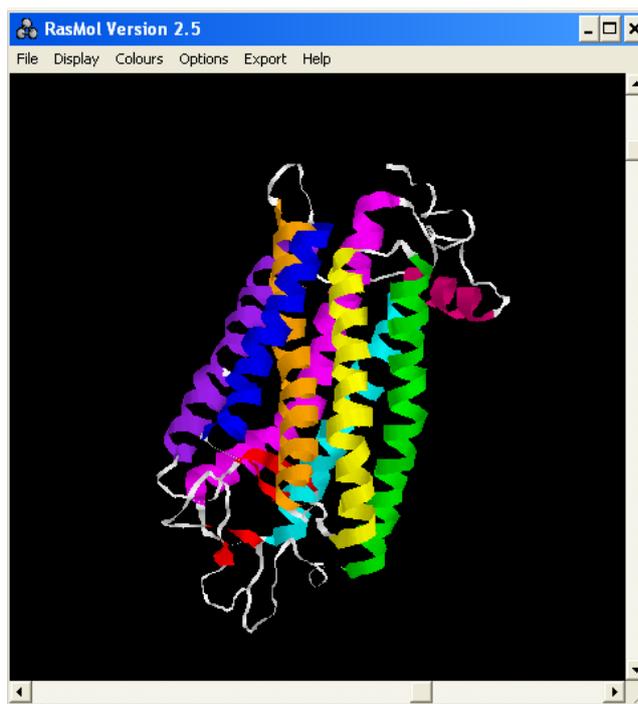
Vá à página da PDB (Protein Data Bank) - <http://www.pdb.org>

Escreva o código **1KPX** no campo **PDB ID or keyword**. Terá acesso a um ficheiro que contém informação sobre a posição de todos os átomos que fazem parte da proteína (OPSR). Nós já fizemos isso por si, se quiser, tente fazê-lo também em casa.



No ecrã do seu computador tem estes 2 ficheiros:

1. Clique no **RASWIN.EXE**, programa que lhe permite visualizar a estrutura 3D.
2. Com ajuda do  arraste o ficheiro **1kpx.pdb** para dentro da janela preta do RASWIN.EXE



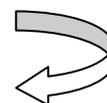
### Agora experimente:

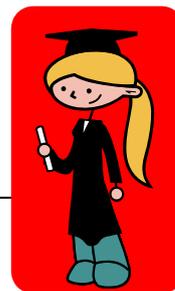
1. Rodar a estrutura 3D movendo simplesmente o rato em cima dela
2. Observar diferentes aspectos da configuração usando o menu do Raswin (barra de topo):

Por exemplo:

- Clique em **Display** e escolha **Cartoons**
- Clique em **Colours** e escolha **Structure**

Quer saber mais sobre a estrutura das proteínas? Volte a página!





## Quer saber mais sobre a estrutura das proteínas?

As longas cadeias de aminoácidos codificadas pelos genes podem dobrar-se em diferentes tipos de configurações como, por exemplo, cadeias em hélices (*alpha helices*) ou cadeias pregueadas (*beta sheets*).

Isto é possível porque cada aminoácido tem propriedades fisico-químicas diferentes.

Alguns têm carga eléctrica e, por exemplo, o aminoácido glutamato que tem carga negativa, é atraído pelo aminoácido de carga positiva, como a lisina. Outros, como a leucina e valina, são repelidos pela água (hidrofóbicos) e preferem ficar no interior das proteínas.

Resumindo, proteínas com seqüências de aminoácidos diferentes originam configurações diferentes.



