

# Engenharia Genética

TP2 - Amplificação do Gene *pgmG* a partir do DNA cromossómico de  
*Sphingomonas elodea* ATCC 31461 por recurso à técnica de PCR

Joana Pinto | 78358

Rafaela Saraiva | 79613

Luís Rita | 78680

## Resultados / Discussão de Resultados / Conclusões

### a) Preparação, *in silico*, de uma Estratégia para Super-Produzir e Purificar um Gene Codificante para uma Proteína de Transporte

O objetivo desta primeira parte da atividade experimental consistiu em delinear uma estratégia para super-produzir e purificar o gene codificante da proteína em estudo. Para tal, recorreu-se à bactéria *E. coli* e utilizando ferramentas bioinformáticas, de modo a selecionar o vetor de clonagem, enzimas de restrição e desenhar os *primers* necessários ao sucesso da experiência.

Tal como previsto, o professor responsável por esta atividade laboratorial atribuiu um número de acesso a cada grupo, que tanto podia corresponder a uma proteína, como a um gene. Concluiu-se, a partir da base de dados do NCBI, com que tipo de estrutura biológica se estava a trabalhar: este código era de uma proteína, mais especificamente, uma proteína transportadora presente na espécie *Arabidopsis thaliana* [1].

Uma vez que se estava na presença de uma proteína e não de um gene, tornou-se essencial a determinação da sequência nucleotídica correspondente, obtida através da seguinte sequência de comandos: 1º - CDS; 2º - FASTA, foi possível revelar a composição nucleotídica associada à proteína em estudo.

transporter-like protein [Arabidopsis thaliana]

GenBank: BAB10596.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS	BAB10596	441 aa	linear	PLN 14-FEB-2004
DEFINITION	transporter-like protein [Arabidopsis thaliana].			
ACCESSION	BAB10596			
VERSION	BAB10596.1 GI:10177340			
DBSOURCE	accession <a href="#">AB011484.1</a>			
KEYWORDS	.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	<a href="#">Arabidopsis thaliana</a>			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis.			

Fig. 1 – Proteína em estudo

Grupo 2B5 - BAB10596.1

## Seleção do Vetor de Clonagem

O vetor de clonagem necessita de satisfazer determinadas características, nomeadamente:

- Ter a capacidade de se replicar na célula hospedeira (ter origem de replicação);
- Ter marcas de seleção (tais como resistência a antibióticos);
- Apresentar uma pequena dimensão (<10kb).

O PETBlueII é um vetor de clonagem que satisfaz todas estas características, na medida em que:

- É pequeno (3653 bp), facilitando a sua inserção em *E. coli*;
- Apresenta resistência à ampicilina (AMP<sup>R</sup>), possibilitando a seleção dos transformantes das colónias de *E. coli*. (As colónias que possuem o plasmídeo sobrevivem em meio complementado com o antibiótico ampicilina);
- Possui origem de replicação que é reconhecida pela *E. coli*;
- Cauda de histidinas (HIS tag), importante pelo facto de permitir separar a proteína em estudo das outras proteínas da *E. coli* por cromatografia.

### Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MXE10

GenBank: AB011484.1

[GenBank](#) [Graphics](#)

```
>(gi|2924734:2340-2409, 2514-2554, 2687-2747, 2832-2934, 3025-3127, 3214-3267,
3343-3405, 3502-3551, 3639-3714, 3804-3992, 4078-4165, 4241-4326, 4421-4528,
4615-4659, 4751-4839, 4981-5080) Arabidopsis thaliana genomic DNA, chromosome 5,
P1 clone:MXE10
ATGAAGAGGCTTCGCCGTGGATTTCCTTCTGGGAGCTTTTACCGTTTGGATCATCGTTCTTTGCACCG
CTTTGCCCATTTCTTCTCTATTCCCCTTTCTCTACTTTATGATTGATGATTTCACATTGCAAAGAAGGA
AGAGGACATTGGGTTCTACGCTGGATTGTGGTTGCTCTTTCATGCTCGGACGAGCTTTTACATCTGTG
GCATGGGGACTTGTGCTGATCGTTATGGTAGAAAACCTGTAATCCTCATAGGAACCGCTTCAGTGGTCG
TTTTAATACTCTGTTTGGCCTAAGTTTAAATTTCTGGATGGCTATTATCACAAAGATTTTGCCTCGGTAG
TTTCAACGGTTTACTTGGTCTTATCAAGGCTTACGCAATGGAATATTTCCGTGATGAGTATCAAGGTTTA
GCACTCTCAGCAGTTAGTACAGCTTGGGGAATTGGACTCATCTTGGCCCTGCTATAGGAGGTTTTCTTG
CTCAGCCTGCAAAGCAATATCCAAGTTTATTCTCACAGGACTCCATTTTGGCAAATTTCCCTTCTTTT
GCCATGCTTAGCAATATCCGTTTTTGCATTCTTGGTGACCATAGTTTCATCAAGGATTCGGAAACATTG
CACAAATCACAAAGTTTAAATGATGATGAGTCTTATGATGCTCTCAAAGATTTGTCTGATGACCTGAATCTA
ATAAAGTGGCAGAGAGAAATGGAAGTTCTCTCTTGAACAACTGGCCATTAATTTTCATCTATCATCGT
ATACTGCGTCTTTTCACTACATGATATGGCTTACACAGAGATCTTTTCATTGTGGGCAACAGTCCGAGG
AAATATGGAGGTTTGGGATACTCCACTGCAGATGTTGGTTCTGTTCTTGCCTTTTCAGGCTTTGGTCTCC
TTATCTTTCAGCTTTTCGCTCTACTCTTACGCAGAGAGGCTTTTAGGACCTATCATAGTTACACGTATATC
TGGGAGCCTAGCAATGGTCTGTTATCATGTTACCCACTAATAGCAAAGTTATCTGGTTTAGCCCTTACC
GTGACTGTAACCTTCTGCATCCGTAGCAAAGAGTGTTTTAGGTAATCTGCTATAACTGGATTATTCATCC
TTCAAACAAGGCTGTGAGACAAGACCAAGAGGAGCAGCTAATGGAATTGCCATGACAGCGATGTCTCT
TTTCAAAGCCATAGTCCAGCAGCAGCAGGAATCATTTTTCGTGGAGCGAGAAACGTCAGGGTGCTGCT
TTTCTCCCTGGTAATGTTAATCCTAATTCATGTTTCATGTCCATGTCCATGTCCATCTCTTAAGTAG
```

Fig. 2 – Sequência nucleotídica correspondente à proteína em estudo

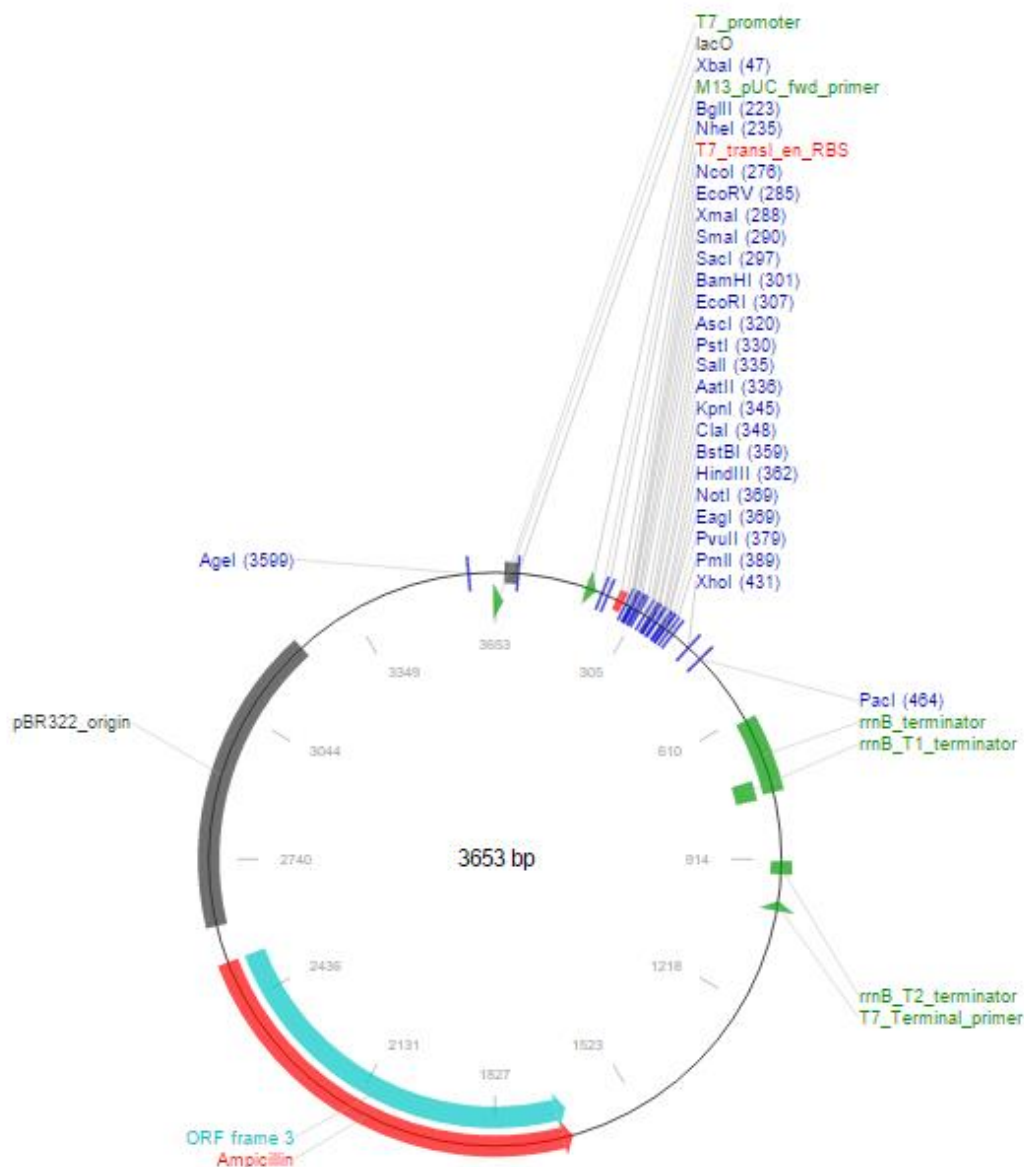


Fig. 3 - Vetor PETBlue II

## Seleção das enzimas de restrição

Após a seleção do vetor a usar, é importante proceder à seleção das enzimas de restrição. As enzimas de restrição devem ser escolhidas de forma a que cortem o vetor de clonagem na zona “multiple cloning site” (MCS), mas que não cortem a sequência do gene. A região MCS pode ser visualizada na figura 3 (região a azul escuro).

Introduzindo, no NEBcutter, a sequência de nucleótidos a que aquela proteína corresponde, visualizou-se o seu mapa de restrição, inferindo sobre as enzimas de restrição que cortam o gene, comprometendo, assim, a sua utilização.

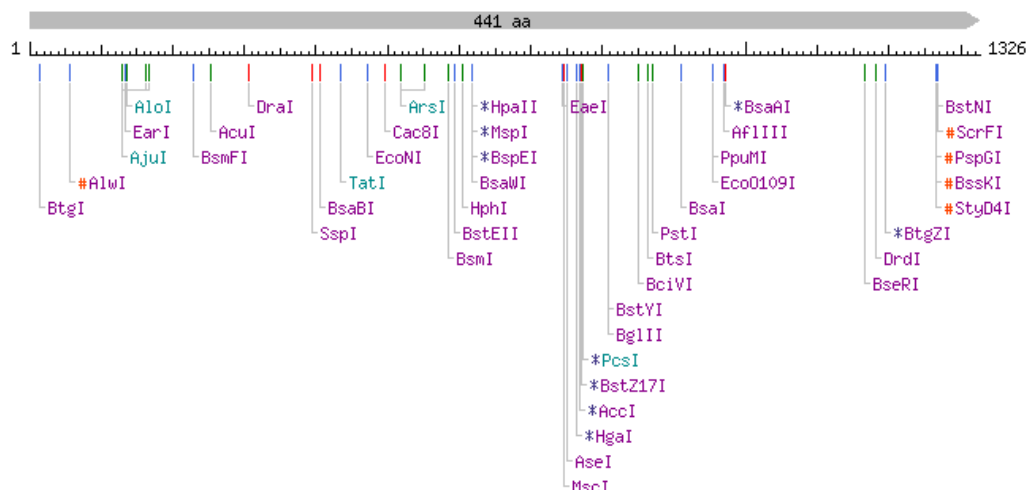


Fig. 4 - Mapa de restrição do fragmento de DNA que contém a sequência de nucleótidos da proteína.

Utilizou-se, ainda, a ferramenta designada por “O cutters”, que apresentou uma lista com as enzimas que não cortavam o gene. Assim, escolheu-se duas enzimas que estivessem presentes em ambos os locais: Na figura 3 e na tabela fornecida pela página da internet [2].

Considerou-se relevante a escolha de duas enzimas para diminuir o risco do vetor recircularizar e também para que a sequência de nucleótidos ficasse com uma correta orientação. Para além dos fatores referidos anteriormente a ter em conta na seleção das enzimas, outro ponto relevante está relacionado com a solução tampão de restrição, pois as enzimas devem ser usadas preferencialmente na mesma solução tampão, de forma a serem compatíveis e terem eficiência máxima. Este facto também leva a que se possa realizar a digestão de DNA por estas enzimas em simultâneo. A informação sobre os meios de solução tampão de cada uma das enzimas encontra-se na página de internet *clontech* [3].

As enzimas que selecionadas são: **HindIII** e **NheI**. Estas enzimas podem ser utilizadas na solução tampão tipo M.

## Seleção dos *primers*

Os *primers* a desenhar devem ser sintetizados tendo em conta a amplificação de todo o gene, para que toda a proteína seja expressa. São necessários 2 *primers*, um para cada cadeia, sendo que um deles recebe o nome de *forward* (*Fw*) e o outro de *reverse* (*Rv*). O *Primer forward* coincide com os primeiros 20 nucleótidos da sequência de DNA da proteína e o *Primer reverse* é o inverso do complementar dos últimos 20 nucleótidos da sequência de DNA.

O *primer forward* vai ligar-se à cadeia de baixo (cadeia complementar à sequência de DNA do gene) e a replicação é feita de 5'→3'. No caso do *primer reverse*, este vai ligar-se à cadeia de DNA de cima, e a replicação também é feita no mesmo sentido.

*Primer forward1* → ATG AAG AGG CTT CGC CGT GG

*Primer reverse1* → CTA GTT AAG AGA TGG ACA TG

Para além disto, tem de se ter em conta que o gene que codifica a proteína deverá ter extremidades coesivas correspondentes às do plasmídeo PETBlue-II para que possa ser facilmente inserido e fique bem orientado. Assim, adicionou-se, aos *primers*, locais de restrição para as enzimas que foram selecionadas no passo anterior (HindIII e NheI). Ao *Primer forward* foi adicionado o local de restrição da enzima HindIII e ao *primer reverse* o da enzima NheI (sequências introduzidas nas extremidades 5' dos *primers*). Os locais de restrição das enzimas puderam ser consultados num documento disponibilizado na página da cadeia.

*Primer forward2* → AAG CTT ATG AAG AGG CTT CGC CGT GG

*Primer reverse2* → GCT AGC CTA GTT AAG AGA TGG ACA TG

Foram ainda adicionados três nucleótidos à esquerda dos *primers* que permitem um melhor funcionamento da enzima e aumentam o conteúdo GC da sequência de nucleótidos. (Estes 3 nucleótidos foram retirados do mesmo documento referido anteriormente).

*Primer forward3* → CCC AAG CTT ATG AAG AGG CTT CGC CGT GG

*Primer reverse3* → CTA GCT AGC CTA GTT AAG AGA TGG ACA TG

Para se poder verificar a qualidade dos *primers*, utilizou-se o software *OligoAnalyzer* [4], onde se obteve informações acerca do conteúdo GC e da temperatura de fusão dos *primers*.

Primers	% GC	T <sup>m</sup> (°C)	Tamanho (b)
Primer forward1	60	62.1	20
Primer reverse1	40	47.1	20

Tabela I - Características dos *primers* desenhados, nomeadamente conteúdo GC e temperatura de fusão.

Foram analisados os *Primers1*, obtidos sem ter em conta os locais de restrição das enzimas. Isto porque os nucleótidos acrescentados aos *Primers3* não hibridam no processo de

PCR pois não têm homologia na sequência de DNA onde se vão ligar. Consequentemente, não vão ter influência na temperatura de fusão nem no conteúdo GC relevante para a ligação.

Antes de se proceder à análise dos resultados, é importante referir quais as características que um *primer* deve possuir para ser considerado “de qualidade”:

- Conteúdo GC entre 40 a 60% [5];
- $T^m$  entre 55 e 80°C;
- Comprimento entre 18-27 bases;
- Terminação 3' com C, G, CG, GC e as extremidades não devem ser complementares;

Passando agora à análise dos resultados, podemos concluir que o *Primer forward1* possui bastante qualidade pois tem um conteúdo GC de 60%, uma temperatura de fusão de 62.1 °C e um tamanho de 20 bases, tudo compreendido nos intervalos referidos anteriormente.

Relativamente ao *Primer reverse1*, apesar do seu conteúdo GC ser baixo (40%), o que resulta também numa temperatura mais baixa que a ideal (47.1°C), considera-se que a qualidade ainda é razoável, pois o conteúdo GC encontra-se no limite de conteúdo GC aconselhável para a realização da técnica. Ainda assim, pode acontecer que as ligações do *primer* ao fragmento sejam fracas e possam ocorrer hibridações não perfeitas, indesejáveis - o que poderia ser corrigido acrescentando nucleótidos ao *primer* que lhe conferissem um aumento do seu conteúdo GC.

Assim, conclui-se que a técnica de PCR pode ser utilizada para amplificar a sequência de nucleótidos da proteína em estudo.

## b) Previsão da Função de uma Proteína Não Caraterizada

Mesmo conhecendo *a priori* a função e a localização celular da proteína em estudo, realizou-se um conjunto de passos padrão que deve ser seguido sempre que se pretende averiguar a função e a localização de um polipéptido desconhecido.

Uma vez determinada a sequência nucleotídica correspondente à proteína, o próximo passo era a **identificação de ORF's** (*Open Reading Frames*) na sequência de nucleótidos gerada anteriormente. Por outras palavras, procurou-se locais que se iniciassem por codões de iniciação (ATG), terminassem com codões de finalização (TAA, TAG ou TGA) e que apresentassem um tamanho mínimo, de forma a ser possível obter uma proteína a partir deles. Ou seja, tratam-se de locais no DNA que, hipoteticamente, poderão codificar para uma proteína. De forma a verificar a existência deste tipo de regiões, recorreu-se a um *site* denominado *ORF Finder*, onde se inseriu a sequência de nucleótidos a analisar [6].

transporter-like protein [Arabidopsis thaliana]				
GenBank: BAB10596.1				
<a href="#">Identical Proteins</a> <a href="#">FASTA</a> <a href="#">Graphics</a>				
<a href="#">Go to:</a> <input type="checkbox"/>				
LOCUS	BAB10596	441 aa	linear	PLN 14-FEB-2004
DEFINITION	transporter-like protein [Arabidopsis thaliana].			
ACCESSION	BAB10596			
VERSION	BAB10596.1 GI:10177340			
DBSOURCE	accession <a href="#">AB011484.1</a>			
KEYWORDS	.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	<a href="#">Arabidopsis thaliana</a> Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis.			

Fig. 5 – Proteína em estudo



Fig. 6 – Arabidopsis thaliana



Obtiveram-se 10 regiões ORF na sequência de nucleótidos exposta abaixo (Figura 8). Metade destas obtiveram-se a partir da cadeia introduzida no site e a outra metade a partir da cadeia de DNA complementar. A presença de 6 barras na figura (3 em cima e outras 3 em baixo) explica-se facilmente, uma vez que, dependendo do local onde

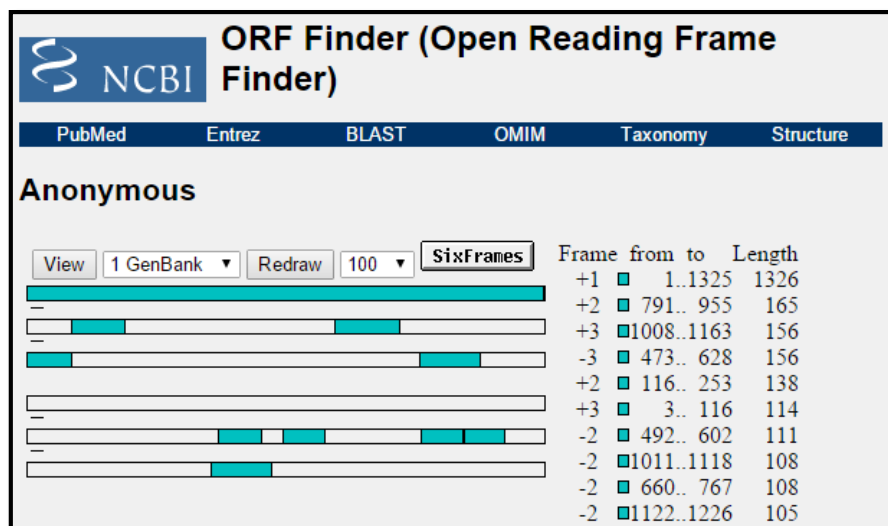


Fig. 7 – 10 Regiões ORF previstas pelo ORF Finder

se começa a leitura, o mesmo nucleótido poderá integrar 3 codões distintos.

```
ATGAAGAGGCTTCGCCGTGGATTCCCTTCTGGGAGCTTTTACCCTTGGATCATCGTCTTCTTGACCG
CTTTGCCATTTCTTCTCTATTCCCTTTCTCTACTTTATGATTGATGATTCAACATTGCAAGAAGGA
AGAGGACATTGGGTTCTACGCTGGATTGTTGGTGTCTTTTCATGCTCGGACGAGCTTTTACATCTGTG
GCATGGGACTTGTGTGATGCTTATGGTAGAAAACCTGTAATCCTCATAGGAACCGCTTCAGTGGTCG
TTTTAATACTCTGTTGGCTAAGTTTAAATTTCTGGATGGCTATTATCACAGATTGTTGCCTCGGTAG
TTTCAACGGTTTACTTGGTCTATCAAGGCTTACGCAATGGAAATATTCGATGATGATCAAGGTTTA
GCACTCTCAGCAGTTAGTACAGCTTGGGGAATTGGACTCATCTTGGCCCTGCTATAGGAGGTTTCTTG
CTCAGCCTGCAAGCAATATCCAAGTTTATCTCACAGGACTCCATTTTGGCAAAATTTCCCTCTCTTTT
GCCATGCTTAGCAATATCCGTTTTCGATTCTTGGTGACCATAGTTTCATCAAGGATTCGGAAACATTG
CACAAATCACAAAGTTTATGATGATGAGTCTTATGATGCTCTCAAGATTGTTGCTGATGACCTGAATCTA
ATAAAGTGGCAGAGAGAAATGGAAAAAGTTCTCTTGAACAACTGGCCATTATTTATCATCTATCATCGT
ATCTGCGTCTTTTCACTACATGATATGGCTTACACAGAGATCTTTTCATTGTGGGCAACAGTCCGAGG
AAATATGAGGTTTGGGACTCTCACTGAGATGTTGGTCTGTTCTTGGCTTTTCAGGCTTGGTCTCC
TTATCTTTAGCTTTGCTCTACTCTTACGCAAGAGGCTTTTAGGACCTATCATAGTTACACGTATATC
TGGGAGCCTAGCAATGCTGCTTATCATGTTTACCACTAATAGCAAGTTATCTGTTTATGCGCTTACC
GTGACTGTAACCTTCTGCATCCGTAGCAAGAGTGTGTTAGGTAATCTGCTATAACTGGATTATTCATCC
TTCAAAACAGGCTGTGAGACAAAGCAAGAGGAGCAGCTAATGGAATTGCCATGACAGCGATGCTCT
TTTCAAAGCCATAGGTCCAGCAGCAGAGGAATCAATTTTCTGAGGAGCAGAAACGTCAGGGTGTCTCT
TTTCTCCTGGAATGTAATCTAATCATGTTCCATGTCATGTCATCTCTTAAGT
```

Fig. 8 – Sequência de nucleótidos da proteína em estudo

Em suma, as barras azuis presentes na figura 7 correspondem aos locais no ácido desoxirribonucleico que satisfazem os critérios especificados anteriormente. Na figura 9, pode verificar-se o que foi dito atrás, isto é, para a primeira sequência nucleotídica apresentada na figura 8, os codões

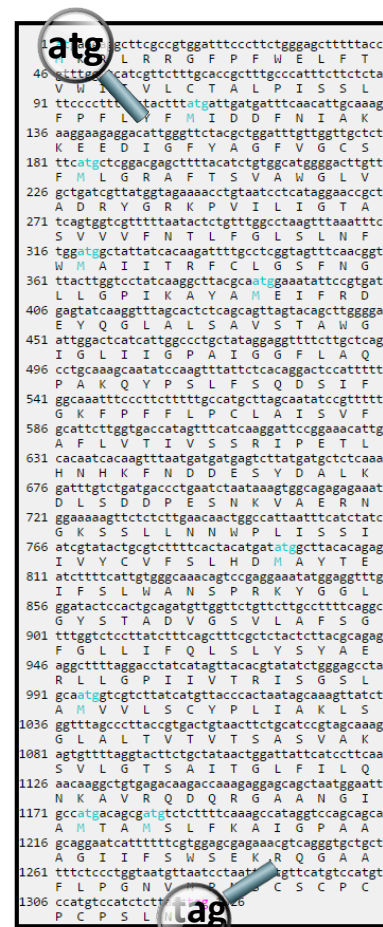


Fig. 9 – ORF: 1→1325

de iniciação e terminação são, respetivamente: ATG e TAG.

Antes de se avançar no trabalho computacional, pretendeu-se compreender que tipo de proteína poderia estar a ser codificada pelo 1º ORF. Tal como era expectável, o BLAST indicou que a sequência de nucleótidos inserida codificava para a proteína em estudo a 100%.

Feito isto, o próximo passo foi tentar encontrar **regiões homólogas noutras sequências conhecidas e presentes na base de dados NCBI BLAST** [7]. Para tal, inseriu-se a sequência de nucleótidos previamente obtida e aguardou-se. O resultado final é o que se encontra presente na figura 10 e tabela II. Verifica-se, assim, a presença de 6 sequências nucleotídicas presentes na espécie *Arabidopsis thaliana* que se assemelham em 95% à obtida anteriormente. Previsivelmente, uma delas corresponde à proteína inicial, da qual se procurou a sequência de nucleótidos.

À partida, poder-se-ia esperar obter uma homologia de 100%, no entanto há que relembrar que o código genético é redundante e que é possível obter-se a mesma proteína a partir de sequências de nucleótidos distintas. Facilmente, conclui-se que algumas mutações ao nível do código genético da proteína em estudo levaram a esta divergência de resultados relativamente à sequência nucleotídica presente na base de dados BLAST.

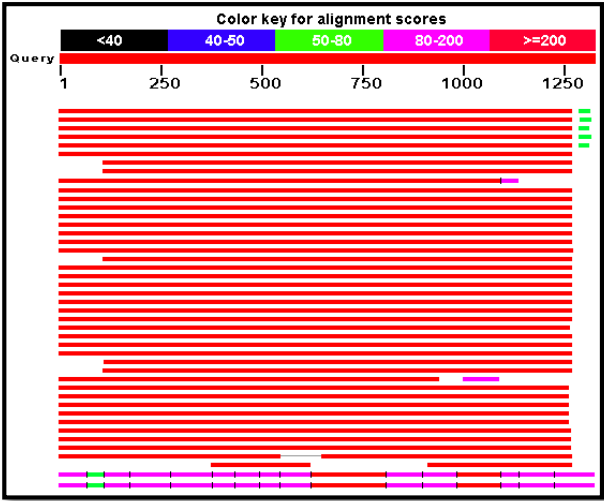


Fig. 10 – Grau de semelhança da sequência de nucleótidos da proteína em estudo com diferentes sequências codificantes da base de dados.

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana zinc induced facilitator-like 1 protein mRNA, complete cds</a>	2348	2348	95%	0.0	100%	<a href="#">NM_180705.2</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana AT5g13750/MXE10_2 mRNA, complete cds</a>	2348	2348	95%	0.0	100%	<a href="#">AY102150.1</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana transporter-like protein (AT5g13750: MXE10.2) mRNA, complete cds</a>	2348	2348	95%	0.0	100%	<a href="#">AY062867.1</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana AT5g13750/MXE10_2 mRNA, complete cds</a>	2348	2348	95%	0.0	100%	<a href="#">AY039542.1</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSLTPGH38ZE06 of Hormone Treated Callus of strain col-0 of Arabidopsis thaliana (thale cress)</a>	2338	2338	95%	0.0	99%	<a href="#">BX831902.1</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana zinc induced facilitator-like 1 protein mRNA, complete cds</a>	2335	2335	95%	0.0	99%	<a href="#">NM_001036800.2</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana zinc induced facilitator-like 1 protein mRNA, complete cds</a>	2145	2145	87%	0.0	100%	<a href="#">NM_121378.4</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana clone 27439 mRNA, complete sequence</a>	2135	2135	87%	0.0	99%	<a href="#">AY086758.1</a>
<input type="checkbox"/>	<a href="#">Arabidopsis thaliana AT5G13750 mRNA, complete cds, clone: RAFL08-17-F01</a>	2012	2096	85%	0.0	99%	<a href="#">AK318697.1</a>
<input type="checkbox"/>	<a href="#">Arabidopsis lyrata subsp. lyrata hypothetical protein, mRNA</a>	1873	1873	95%	0.0	93%	<a href="#">XM_002871537.1</a>

Tabela II - Lista de Resultados BLAST Relativa à Análise da Sequência Nucleotídica

No entanto, é importante reforçar o facto de que, mesmo com estas bases azotadas distintas das que poderíamos estar à espera, a proteína mantém-se funcional, porque diferentes codões podem codificar um mesmo aminoácido (propriedade de degenerescência do código genético). Estas regiões, com conteúdo nucleotídico distinto aparecem representadas a verde na figura 10.

Por tudo o que foi dito no parágrafo anterior, facilmente se deduz que uma análise ao nível dos aminoácidos da proteína será muito mais rigorosa para

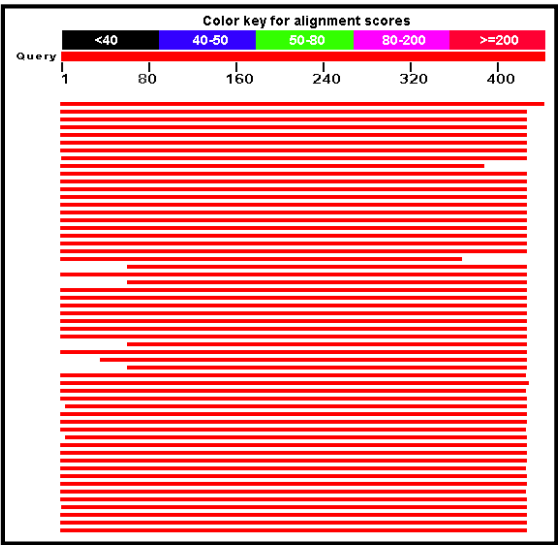


Fig. 11 – Grau de semelhança da sequência de aminoácidos da proteína em estudo com diferentes sequências da base de dados

verificar se existe alguma estrutura biológica com composição semelhante à proteína em estudo e, posteriormente, retirar conclusões acerca da função da mesma.

Assim, optou-se pela **comparação entre a sequência de aminoácidos da proteína em estudo, com as disponíveis na ferramenta BLAST**. Para tal, utilizou-se a ferramenta Blastp para obter os resultados que a seguir se apresentam [8].

#### Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

<div>Alignments</div> <div>Download</div> <div>GenPept</div> <div>Graphics</div> <div>Distance tree of results</div> <div>Multiple alignment</div>							
		Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">transporter-like protein [Arabidopsis thaliana]</a>	892	892	100%	0.0	100%	<a href="#">BAB10596.1</a>
<input type="checkbox"/>	<a href="#">zinc induced facilitator-like 1 protein [Arabidopsis thaliana]</a>	854	854	96%	0.0	99%	<a href="#">NP_851036.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein ARALYDRAFT_488196 [Arabidopsis lyrata subsp. lyrata]</a>	797	797	96%	0.0	93%	<a href="#">XP_002871583.1</a>
<input type="checkbox"/>	<a href="#">PREDICTED: protein ZINC INDUCED FACILITATOR-LIKE 1 [Camelina sativa]</a>	783	783	96%	0.0	91%	<a href="#">XP_010492149.1</a>
<input type="checkbox"/>	<a href="#">PREDICTED: protein ZINC INDUCED FACILITATOR-LIKE 1-like isoform X1 [Camelina sativa]</a>	779	779	96%	0.0	91%	<a href="#">XP_010453472.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein CARUB_v10002619mq [Capsella rubella]</a>	778	778	96%	0.0	89%	<a href="#">XP_006286669.1</a>
<input type="checkbox"/>	<a href="#">PREDICTED: protein ZINC INDUCED FACILITATOR-LIKE 1-like [Camelina sativa]</a>	777	777	96%	0.0	90%	<a href="#">XP_010419980.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein AALP_AA8G135600 [Arabis alpina]</a>	773	773	96%	0.0	88%	<a href="#">KFK25600.1</a>
<input type="checkbox"/>	<a href="#">zinc induced facilitator-like 1 protein [Arabidopsis thaliana]</a>	765	765	87%	0.0	98%	<a href="#">NP_001031877.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein EUTSA_v10013417mq [Eutrema salsugineum]</a>	762	762	96%	0.0	88%	<a href="#">XP_006399873.1</a>

Tabela III – Lista de Resultados BLAST Relativa à Análise da Sequência de Aminoácidos

```

MKRLRRGFPFWELFTVWIIVLCTALPISSLPFLYFMIDDFNIAKKEEDIGFYAGFVGCSFMLGRAFTSV
AWGLVADRYGRKPVILIGTASVVVNTLFGLSLNFWMAITRFLGSGFNLLGPIKAYAMEIFRDEYQGL
ALSAVSTAWGIGLIIGPAIGGFLAQPAKQYPSLFSQDSIFGKFPFLPCLAISVFAFLVTIVSSRIPETL
HNHKNDDDESVDKDLSDDPESNKVAERNKGSSLLNNWPLISSIIVYCVFSLHDMAYTEIFSLWANSR
KYGGLGYSTADVGSLAFSGFGLLIFQLSLSYAERLLGPIIVTRISGSLAMVVLSCYPLIAKLSGLALT
VTVTASVAKSVLGTSAITGLFILQNKAVRQDQGAANGIAMTAMSLFKAIGPAAAGIIFSWSEKRQGAA
FLPGNVNPNNSCSCPCPCPSLN

```

Fig. 12 – Sequência de aminoácidos da proteína em estudo

Tal como expectável, encontrou-se uma sequência de aminoácidos rigorosamente igual à com que se começou a atividade computacional. O que permitiria concluir, que a atividade da proteína em estudo é igual à de uma já conhecida. Isto seria particularmente importante, caso se desconhecasse por completo a função na célula da proteína em estudo.

De seguida, partiu-se para a **análise físico-química da proteína**, recorrendo-se ao *site* da *ExPASy*, onde se introduziu a sequência de aminoácidos em estudo e se obteve os resultados apresentados de seguida [9]. Esta avaliação é de extrema importância, uma vez que

permitirá avaliar os cuidados que se deverão ter ao manusear esta proteína em laboratório. Exemplificando, uma das características que poderá ser importante conhecer é o seu ponto isoelétrico. Esta medida permite ao utilizador saber para que valores de pH a proteína apresenta uma carga global neutra: neste caso, o valor teórico é de 8.72.

Nº de Aminoácidos			441		
Peso Molecular (Dalton)			47905.9		
pI Teórico			8.72		
Composição de aminoácidos					
Ala (A)	42	9.5 %	Lys (K)	17	3.9 %
Arg (R)	16	3.6 %	Met (M)	9	2.0 %
Asn (N)	17	3.9 %	Phe (F)	35	7.9 %
Asp (D)	15	3.4 %	Pro (P)	22	5.0 %
Cys (C)	10	2.3 %	Ser (S)	44	10.0 %
Gln (Q)	9	2.0 %	Thr (T)	18	4.1 %
Glu (E)	12	2.7 %	Trp (W)	8	1.8 %
Gly (G)	37	8.4 %	Tyr (Y)	14	3.2 %
His (H)	3	0.7 %	Val (V)	27	6.1 %
Ile (I)	35	7.9 %	Pyl (O)	0	0.0 %
Leu (L)	51	11.6 %	Sec (U)	0	0.0 %

Tabela IV – Propriedades Físico-Químicas da Proteína em Estudo

<b>Número Total de Resíduos Carregados Negativamente ( Asp + Glu )</b>		27
<b>Número Total de Resíduos Carregados Positivamente ( Arg + Lys )</b>		33
<b>Composição Atómica</b>	Carbono	2218
	Hidrogénio	3414
	Azoto	546
	Oxigénio	598
	Enxofre	19
<b>Fórmula</b>		C <sub>2218</sub> H <sub>3414</sub> N <sub>546</sub> O <sub>598</sub> S <sub>19</sub>
<b>Número total de Átomos</b>		6795

Tabela V - Propriedades Físico-Químicas da Proteína em Estudo

<b>Índice de Instabilidade</b>	36.54 (Proteína Estável)
<b>Índice Alifático</b>	103.33
<b>GRAVY (Grand Average of Hydropathicity)</b>	0.507

Tabela VI - Propriedades Físico-Químicas da Proteína em Estudo

Em seguida, **procurou-se regiões transmembranares** - os dados obtidos neste ponto permitirão retirar conclusões importantíssimas acerca da localização da proteína na célula [10].

Tal como se pode verificar, esta ferramenta prevê vários locais transmembranares, interiores e exteriores à membrana plasmática. Estes resultados foram obtidos com base na polaridade da proteína nos vários locais, permitindo, assim, determinar quais os locais mais hidrofóbicos/hidrofílicos. Isto vai de encontro ao que se estava à espera, uma vez que se sabia desde início que esta se tratava de uma proteína transmembranar, pelo que não é de espantar a existência desta alternância entre zonas com afinidade à água e hidrofóbicas.

Finalmente, analisou-se a proteína quanto à **existência de locais com modificações pós- traducionais**. Tal como se sabe, um enorme número de polipéptidos, após a tradução do mRNA correspondente, são enviados para o complexo de *Golgi* para sofrerem algumas alterações à sua estrutura e composição química. Está-se a falar não só de fosforilações, mas também de desfosforilações, introdução de regiões glicídicas e/ou lipídicas, entre outras...

Por forma a prever a existência deste género de alterações utilizou-se a base de dados *InterProScan* [11], que contém uma lista de padrões (sequência de aminoácidos dispostos de determinada forma) que se confirmou experimentalmente estarem associados a alguma propriedade biológica. Confirmou-se que à proteína em estudo não estão associados nenhum destes padrões, ou seja, nenhuma modificação pós-traducional.

Após a análise dos resultados obtidos, verifica-se que a **proteína foi incluída na categoria das MFS (Major Facilitator Superfamily)**, uma superfamília que inclui uma enorme variedade de proteínas membranares presentes em eucariotas, que apenas conseguem transportar solutos pequenos, a favor de um gradiente quimiosmótico. Foram também detetadas regiões de aminoácidos associadas à resistência ao antibiótico tetraciclina. Por fim, esta pode, também, ser integrada na seguinte família: *Tetracycline resistance protein, TetA/multidrug resistance protein MdtG*.

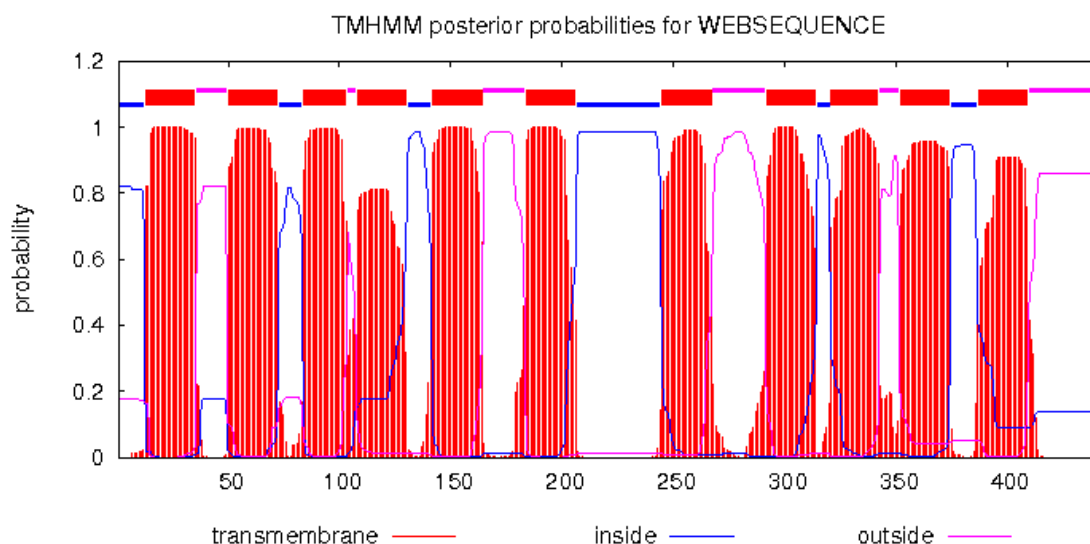
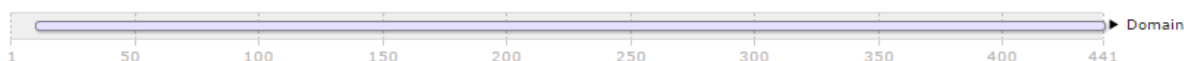


Fig. 13 – Probabilidade da proteína conter regiões intracelulares/extracelulares/transmembranares

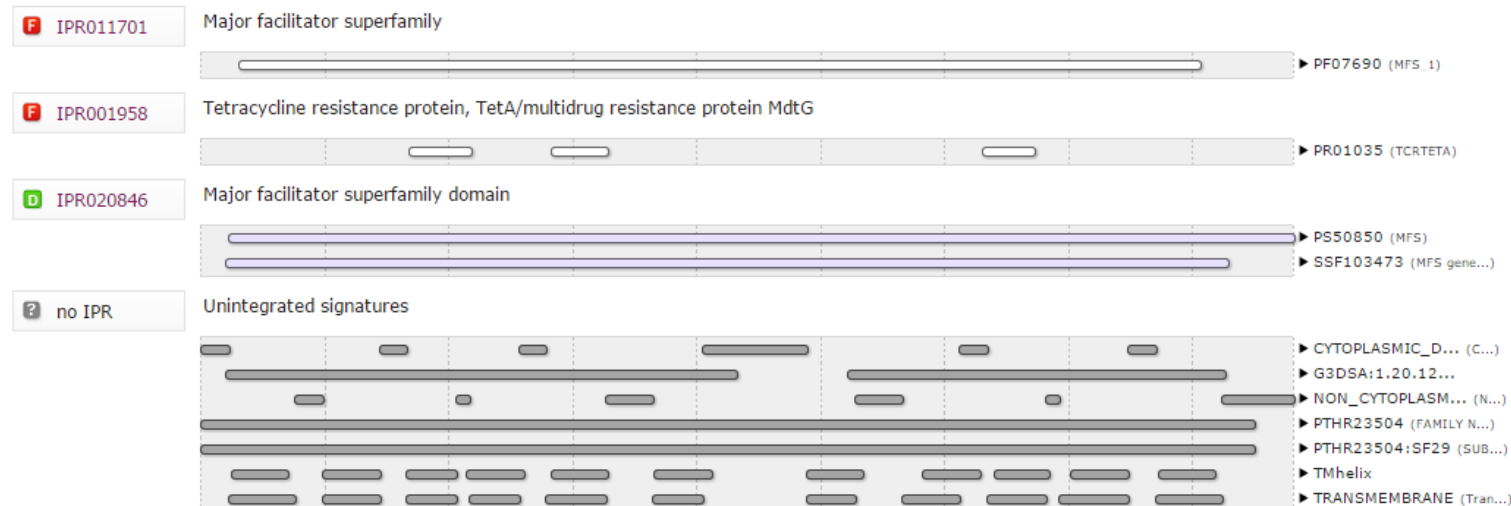
## Protein family membership

- Major facilitator superfamily (IPR011701)
- Tetracycline resistance protein, TetA/multidrug resistance protein MdtG (IPR001958)

## Domains and repeats



## Detailed signature matches



## GO term prediction

### Biological Process

GO:0055085 transmembrane transport

### Molecular Function

GO:0005215 transporter activity

### Cellular Component

GO:0016021 integral component of membrane

Fig. 14 – Domínios típicos da família MFS

De acordo com o já previsto, o processo biológico em que a proteína em estudo se encontra envolvida é o transporte de substâncias através da membrana plasmática e a função molecular corresponde à de transportadora.

## Bibliografia

- [1] <http://www.ncbi.nlm.nih.gov/protein/BAB10596.1>
- [2] <http://nc2.neb.com/NEBcutter2/listbycuts.php?name=78bbca2b-&numcuts=0>
- [3] [http://www.clontech.com/takara/US/Support/Applications/Restriction Enzymes/Buffer Activity Chart](http://www.clontech.com/takara/US/Support/Applications/Restriction_Enzymes/Buffer_Activity_Chart)
- [4] <http://eu.idtdna.com/calc/analyzer>
- [5] [http://www.premierbiosoft.com/tech\\_notes/PCR Primer Design.html](http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html)
- [6] [www.ncbi.nlm.nih.gov/gorf/gorf.html](http://www.ncbi.nlm.nih.gov/gorf/gorf.html)
- [7] [http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\\_TYPE=BlastSearch&LINK\\_LO C=blasthome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)
- [8] [http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LO C=blasthome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)
- [9] <http://web.expasy.org/protparam/>
- [10] <http://www.cbs.dtu.dk/services/TMHMM-2.0/>
- [11] <http://www.ebi.ac.uk/interpro/search/sequence-search>
- [12] Guia de trabalhos laboratoriais de Engenharia Genética, Mestrado em Engenharia Biológica, Mestrado em Engenharia Biomédica, TP2-AMPLIFICAÇÃO DO GENE *pgmG* A PARTIR DO DNA CROMOSSÓMICO DE *Sphingomonas elodea* ATCC 31461 POR RECURSO À TÉCNICA DE PCR, Moreira, L. M., Viegas, C. A., Fialho, A., Leitão, J. H., Sá-Correia, I., Área de Ciências Biológicas, IST, 2015/2016
- [13] Powerpoints das aulas teóricas de Engenharia Genética, Prof. Leonilde Moreira, 2015/2016