



Análise dos genes MC4R e SEC16B e a sua relação com a diabetes tipo 2

Bioinformática - Opção A

Mestrado em Engenharia Biomédica

Grupo 17: PG50255 Beatriz Silva Matos

PG50291 Catarina Rodrigues Pereira

PG50443 Isabel Maria Ferreira da Silva

PG50626 Mariana Fernandes Gonçalves

Docentes: *Miguel Rocha, Luís Melo*

Data de entrega: 3 fevereiro, 2023

Sumário

Este trabalho promove a utilização avançada de ferramentas de Bioinformática, para análise de um conjunto de genes e sequências relacionadas. Os genes em estudo são MC4R e SEC16B que, segundo estudos de GWAS, indicam relação com diabetes tipo 2.

Índice

1	Introdução	1
2	Análise da Literatura	2
2.1	MC4R - Recetor de Melanocortina 4	2
2.2	SEC16B - Proteína transportadora da SEC16B	2
3	Análise da sequência e das <i>features</i> presentes no NCBI	3
3.1	Gene MC4R	3
3.2	Gene SEC16B	5
4	Análise de homologias por BLAST	7
4.1	MC4R - Recetor de Melanocortina 4	7
4.2	SEC16B - Proteína transportadora da SEC16B	8
5	Ferramentas de análise das propriedades da proteína	10
5.1	Recetor de Melanocortina 4 (MC4R)	10
5.1.1	Proteína codificada pelo gene MC4R	10
5.1.2	Estrutura da Proteína	10
5.1.3	Localização sub-celular	11
5.1.4	Funções gerais	12
5.1.5	Locais de expressão do gene	13
5.1.6	Família e Domínios	14
5.1.7	Domínios Transmembranares	15
5.1.8	Alterações pós-tradução	16
5.2	Proteína Transportadora da Proteína SEC16B	17
6	Alinhamento múltiplo e filogenia	18
7	Regulação	19
8	Comparação de variantes do gene e o seu impacto biológico	20
9	Conclusões	21

1 Introdução

A diabetes é uma doença crónica que se caracteriza pelo aumento da glicose no sangue, dando origem à hiperglicemia. Estes níveis elevados devem-se à produção insuficiente de insulina ou a uma atuação ineficaz da insulina e, em alguns casos, à combinação dos dois fatores.

A diabetes tipo 2 é uma das doenças que mais afeta a humanidade atualmente e é responsável por mais de 90% dos pacientes com diabetes. Apresenta um impacto social e económico enorme, nomeadamente na diminuição da esperança e da qualidade de vida e também apresenta efeitos nos sistemas de saúde. A incidência e a prevalência desta doença continuam a aumentar globalmente, sendo esta desenvolvida essencialmente devido a hábitos de vida pouco saudáveis e um estilo de vida sedentário, levando a complicações microvasculares e macrovasculares. Alguns dos sintomas clássicos da doença são a sensação de boca seca, aumento da vontade de urinar e de beber água e, até, perda de peso sem causa aparente. Quanto a medidas preventivas destaca-se um estilo de vida saudável com uma alimentação equilibrada e atividade física regular, além de manter os níveis de colesterol e de pressão arterial controlados. Relativamente ao tratamento da diabetes este passa por mudanças no estilo de vida e administração de insulina, em casos mais avançados.

A obesidade é um distúrbio do peso corporal caracterizado pelo excesso de adiposidade que aumenta o risco de desenvolvimento de diabetes tipo 2.

Ambos os genes em estudo, MC4R e SEC16B, estão relacionados com a obesidade e, consequentemente, com a diabetes tipo 2, uma vez que a obesidade é um dos fatores de risco da doença.

Com a realização deste trabalho, pretende-se analisar estes dois genes escolhidos, focando-se no estudo das suas sequências e da sua estrutura. Assim, o principal objetivo será utilizar diversas ferramentas bioinformáticas e interpretar os resultados obtidos, de forma a ser possível perceber a função dos genes em estudo, as possíveis interações e relações com a diabetes e com os possíveis fatores ambientais relacionados.

2 Análise da Literatura

2.1 MC4R - Recetor de Melanocortina 4

MC4R é um gene humano constituído por 332 aminoácidos e que codifica uma proteína chamada *recetor de melanocortina 4*. Este é um recetor acoplado à proteína G (GPCR), que é a maior família de recetores transmembranares em humanos, abrangendo quase 800 genes distintos e os seus produtos gênicos correspondentes. A estrutura dos GPCR consiste em sete domínios transmembranares hidrofóbicos, que se organizam na membrana plasmática da célula, bem como três domínios extracelulares e três intracelulares que são relativamente hidrofílicos. Estes elementos extracelulares e intracelulares servem como transdutores ideais de estímulos através da membrana celular. Os GPCRs reconhecem uma elevada variedade de estímulos extracelulares, nomeadamente pequenas moléculas, péptidos e proteínas, e comunicam os estímulos através de uma barreira de membrana impermeável ao domínio intracelular para efetuar alterações na função celular. A sinalização é transmitida através de uma cascata de interações de proteínas e eventos de ativação. [1]

MC4R é altamente expresso no hipotálamo, uma região do cérebro intimamente envolvida na regulação do apetite, desempenhando um papel central na homeostase energética, no crescimento somático e na regulação da formação de tecido adiposo. Assim, é essencial para a regulação do apetite e do gasto energético. [2]

O corpo humano é constituído por inúmeros sensores para os níveis de energia. Quando detetam elevados níveis de energia, são enviados sinais de saciedade de forma a ativar os neurónios através do *recetor de melanocortina 4*, alertando o corpo de saciedade. Quando detetam baixos níveis de energia, são enviados sinais de insaciedade que vão ativar os neurónios também através do *recetor de melanocortina 4*, produzindo a sensação de fome.

Uma disfunção deste gene pode causar hiperfagia, saciedade prejudicada e obesidade. Mutações inativadoras do MC4R são a causa mais comum de obesidade monogénica, ou seja, obesidade determinada por uma mutação génica específica. Mais de 160 mutações e variantes do MC4R foram descritas, sendo que algumas reduzem a função do gene e outras eliminam completamente a sua função. A perda da função devido a mutações de MC4R leva a um aumento do apetite do indivíduo, apresentando uma característica típica designada de hiperfagia, que consiste numa condição caracterizada por um apetite insaciável. [2]

Vários estudos revelam que o gene MC4R está relacionado à obesidade monogénica, como já referido, mas também à obesidade comum. Interações de variantes no gene MC4R com massa gorda e com o gene associado à obesidade aumenta significativamente o risco de obesidade, principalmente na adolescência. Por outro lado, cerca de 15% das variantes do gene MC4R resultam em um ganho de função que protege contra a obesidade e está associado a um perfil metabólico favorável. [3]

Além disso, estudos também indicam que o gene MC4R é um possível candidato a causar diabetes tipo 2 e está associado a doenças cardiovasculares. [4]

2.2 SEC16B - Proteína transportadora da SEC16B

3 Análise da sequência e das *features* presentes no NCBI

Com o intuito de perceber melhor a influência destes genes na patologia da diabetes, foi necessário proceder a uma análise da sua sequência e *features*. Esta análise foi conseguida com o auxílio de scripts em BioPython.

Previamente a esta análise, foi realizado o *download* dos ficheiros, em formato *genbank*, através do NCBI, nos quais compilavam a informação acerca da sequência dos mesmos - "*sequence-MC4R.gb*" e "*sequence-SEC16B.gb*". Estes ficheiros podem ser acedidos no repositório associado a este trabalho.

As sequências foram lidas recorrendo à interface *Sequence Input/Output* (SeqIO) do BioPython de acordo com o código abaixo.

```
#Gene MC4R
geneM= SeqIO.read("sequence-MC4R.gb", "genbank")
geneM
#Gene SEC16B
geneS= SeqIO.read("sequence-SEC16B.gb", "genbank")
geneS
```

3.1 Gene MC4R

Relativamente ao gene MC4R, através da descrição dada pelo código abaixo, verifica-se que este gene é o gene do recetor de melanocortina 4 do organismo *Homo sapiens* e está localizado no cromossoma 18. Para além disto, através deste, também foi possível obter a sequência completa formada por 1714 nucleótidos.

```
print ("Descricao:", geneM.description)
print (geneM.seq)
print ("Numero de nucleotidos:", len(geneM.seq))
```

De seguida, foi também analisada a informação presente nas anotações do ficheiro (`geneM.annotations`), apresentadas na Figura 1, as quais revelaram informações como o tipo de molécula - molécula de DNA-, a topologia - linear-, a taxonomia e ainda um comentário.

```
{'molecule_type': 'DNA',
  'topology': 'linear',
  'data_file_division': 'PRI',
  'date': '19-FEB-2021',
  'accessions': ['NG_016441', 'REGION:', '4994..6707'],
  'sequence_version': 1,
  'keywords': ['RefSeq', 'RefSeqGene'],
  'source': 'Homo sapiens (human)',
  'organism': 'Homo sapiens',
  'taxonomy': ['Eukaryota',
    'Metazoa',
    'Chordata',
    'Craniata',
    'Vertebrata',
    'Euteleostomi',
    'Mammalia',
    'Eutheria',
    'Euarchontoglires',
    'Primates',
    'Haplorrhini',
    'Catarrhini',
    'Hominidae',
    'Homo'],
  'comment': 'REVIEWED REFSEQ: This record has been curated by NCBI staff in collaboration with LRG Consortium. The reference sequence was derived from AC091576.11. This sequence is a reference standard in the RefSeqGene project. Summary: The protein encoded by this gene is a membrane-bound receptor and member of the melanocortin receptor family. The encoded protein interacts with adrenocorticotrophic and MSH hormones and is mediated by G proteins. This is an intronless gene. Defects in this gene are a cause of autosomal dominant obesity. [provided by RefSeq, Jan 2010].'}
```

Figura 1: Anotações do ficheiro "*sequence-MC4R.gb*"

O comentário contém informação de interesse acerca da proteína codificada por este gene, como a sua interação com hormonas e outras proteínas. Estas interações serão discutidas mais à frente neste trabalho.

Para além disto, o comentário também revela que a sequência deste gene não é interrompida por intrões e que defeitos no gene MC4R são uma das causas da obesidade, daí a sua relação com a diabetes tipo 2.

Um dos objetivos desta parte era a análise das features. Esta foi feita com a ajuda das funções abaixo.

```

1 def typelocal(features): #Dá-nos todos os tipos e localizações das features
2     print ("Tipo e Local das Features")
3     for f in features:
4         print(f.type, f.location)
5
6 def qualifiersCDS(features): #Significado Biológico das proteínas codificadas pelo gene
7     print ("Qualifiers de CDS")
8     for f in range (len(features)):
9         if features[f].type=="CDS":
10             print("Localização:",features[f].location)
11             print (features[f].qualifiers)
12             print()
13
14 def misc(features): #Retorna a localização e descrição de locais de interesse da sequência do gene
15     print("Features de Interesse")
16     for f in range (len(features)):
17         if features[f].type=="misc_feature":
18             print("Localização:",features[f].location)
19             print(features[f].qualifiers)
20             print()

```

Figura 2: Código relativo a funções utilizadas para a análise das *features* dos genes

A função `typelocal` é uma função inicial que sumariza, de certa forma, o conteúdo das *features*, dando-nos vários tipos existentes e a sua localização. Esta foi utilizada com o intuito de ver quais as *features* com maior interesse para analisar. Daqui retirou-se que os tipos de maior interesse seriam, então, a CDS, isto é, sequência codificante de proteína, e *misc_features*, ou seja, *features* que não podem ser descritas por mais nenhuma *feature key* mas que têm interesse biológico.

Desta forma, desenvolveu-se as funções `qualifiersCDS` e `misc` para que fosse retornado o local e o significado biológico das proteínas codificadas e locais de interesse no gene, respetivamente.

Relativamente à CDS do gene MC4R, verificou-se que a sequência entre os nucleotídeos 426 e 1425 codifica a proteína com o mesmo nome do gene, isto é, o recetor de melanocortina 4, e que esta é a única proteína a ser codificada por este gene. É esta proteína vai utilizada no BLAST no ponto 4.1 e na análise no ponto 5.1

Relativamente às *misc_features*, observa-se que a sequência do genoma apresenta 3 locais de N-glicosilação e 7 domínios transmembranares, todos localizados dentro do intervalo codificante da proteína MC4R, confirmando a literatura anterior.

3.2 Gene SEC16B

No que toca ao gene SEC16B, o processo de análise foi em tudo semelhante ao anterior, utilizando linhas de código semelhantes para obter a sequência nucleotídica completa, o seu comprimento e a sua descrição.

Desta forma, obteve-se que o gene SEC16B é composto por 55497 nucleotídeos no total e pertence ao cromossoma 1 no organismo *Homo sapiens*. Relativamente às anotações do SEC16B, estas demonstraram que tinha uma topologia e taxonomia igual à do gene anterior, uma vez que são sequência do mesmo tipo e do mesmo organismo.

No que diz respeito à análise das *features*, através da função `typelocal`, viu-se que este não apresentava *misc_features*, no entanto, continha várias CDS.

Após correr a função `qualifiersCDS` nas *features* do gene SEC16B, observou-se que este codificava os isómeros 1,2 e 4 da proteína transportadora da proteína SEC16B. Para além disto,

focando na localização destas CDS, verifica-se que estas são compostas por várias junções, significando que a sequência codificante é interrompida por vários intrões, ao contrario do gene MC4R. Isomeros de proteínas são proteínas semelhantes umas às outras que atuam, também, de forma semelhante.

4 Análise de homologias por BLAST

4.1 MC4R - Recetor de Melanocortina 4

De forma a ser possível fazer uma análise de homologias utilizou-se a ferramenta BLAST, tendo-se encontrado várias sequências homólogas à sequência da proteína codificada pelo gene em estudo, MCR4. De forma geral, todas as sequências encontradas apresentam um elevado grau de compatibilidade e semelhança.

Na Figura 3 estão apresentadas algumas das proteínas com sequências homólogas. A primeira linha corresponde à proteína codificada pelo gene em estudo, MCR4, o qual possui o nome de proteína *Melanocortin receptor 4*, com o organismo correspondente *Homo sapiens (Human)* e com comprimento de sequência de 332 aminoácidos.

A sequência mais homóloga à proteína em estudo possui como número de acesso Q8HXX3.1, apresentando o nome de proteína *Melanocortin receptor 4*, com comprimento de 332 aminoácidos e associado ao organismo *Macaca fascicularis*. Analisando a sequência, verificamos que apresenta um **Query Cover** de 100%, **Max Score** e **Total Score** de 664, **E-Value** de 0, **Percent Identities** de 98,49% e um comprimento de sequência de 332 aminoácidos.

Sequences producing significant alignments		Download	Select columns	Show	100				
<input checked="" type="checkbox"/> select all 100 sequences selected		GenPept	Graphics	Distance tree of results	Multiple alignment	MSA Viewer			
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Homo sapiens]	Homo sapiens	673	673	100%	0.0	100.00%	332	P32245.2
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Macaca fascicularis]	Macaca fascicularis	664	664	100%	0.0	98.49%	332	Q8HXX3.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Sus scrofa]	Sus scrofa	647	647	99%	0.0	96.39%	332	Q97504.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Vulpes vulpes]	Vulpes vulpes	645	645	99%	0.0	96.08%	332	Q02819.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Mus musculus]	Mus musculus	634	634	99%	0.0	93.98%	332	P56450.3
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Rattus norvegicus]	Rattus norvegicus	634	634	99%	0.0	93.98%	332	P70596.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Bos taurus]	Bos taurus	625	625	99%	0.0	93.37%	332	Q9GLJ8.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 4; Short=MC4-R [Danio rerio]	Danio rerio	466	466	95%	4e-165	71.69%	326	B0V1P1.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 5; Short=MC5-R [Rattus norvegicus]	Rattus norvegicus	404	404	92%	5e-141	65.06%	325	P35345.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 5; Short=MC5-R [Mus musculus]	Mus musculus	398	398	92%	1e-138	64.42%	325	P41149.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 5; Short=MC5-R [Pan troglodytes]	Pan troglodytes	394	394	94%	5e-137	62.73%	325	Q9IT23.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 5; Short=MC5-R; AltName: Full=MC-2 [Homo sapiens]	Homo sapiens	393	393	94%	2e-136	62.73%	325	P33032.3
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 3; Short=MC3-R [Mus musculus]	Mus musculus	382	382	91%	2e-132	59.93%	323	P33033.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 3; Short=MC3-R [Homo sapiens]	Homo sapiens	378	378	88%	9e-131	61.33%	323	P41968.3
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 3; Short=MC3-R [Rattus norvegicus]	Rattus norvegicus	378	378	91%	1e-130	59.28%	323	P32244.2
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 5; Short=MC5-R [Bos taurus]	Bos taurus	374	374	87%	8e-129	64.38%	325	P56451.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 5; Short=MC5-R [Ovis aries]	Ovis aries	370	370	94%	2e-127	60.19%	325	P41983.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocortin receptor 5; Short=MC5-R [Sus scrofa]	Sus scrofa	362	362	86%	1e-124	62.93%	294	Q9MZV8.1
<input checked="" type="checkbox"/>	RecName: Full=Melanocyte-stimulating hormone receptor; Short=MSH-R; AltName: Full=BDF3; AltName: Full=Mel...	Bos taurus	296	296	87%	1e-98	51.71%	317	P47798.2

Figura 3: Sequências homólogas à sequência da proteína em estudo MC4R.

O **Query Cover** corresponde à percentagem de cobertura entre as duas sequências nucleotídicas. Sendo o valor obtido de 100% isto significa que tem uma cobertura da sequência completa, ou seja, toda a sequência alvo é abrangida pela sequência em estudo.

O **Total Score** corresponde à soma das pontuações do alinhamento de todos os segmentos da mesma sequência. O **Max Score** é a pontuação de alinhamento máxima calculada a partir da soma das recompensas relativas aos aminoácidos coincidentes e penalidades por mismatches e gaps. A pontuação do Total Score é diferente do Max Score se várias partes das sequências da base de dados corresponderem a diferentes partes da sequência alvo. Neste caso, os valores

destes dois parâmetros são iguais e apresentam um valor bastante próximo do valor apresentado pela proteína codificada pelo gene em estudo.

Quanto ao **E-Value**, o seu valor indica o número de alinhamentos com o mesmo score (ou maior) se todas as sequências na base de dados fossem aleatórias. Quanto menor o seu valor, ou mais próximo de zero, mais "significativa" é a correspondência. Neste caso, o valor obtido foi zero o que corresponde a uma probabilidade arredondada e significa que não há praticamente nenhuma probabilidade de o alinhamento ocorrer por acaso. A correspondência é, por isso, muito significativa e apresenta uma compatibilidade e homologia entre as sequências total.

O valor de **Percent Identities** demonstra quão semelhantes são as sequências a serem comparadas e corresponde à extensão em que as duas sequências têm os mesmos resíduos na mesma posição, sendo que quanto maior, mais significativa é a correspondência. Neste caso o valor é de 98.49%, e por isso, a correspondência é bastante significativa e as sequências são muito semelhantes.

Entre os organismos que apresentam as proteínas homólogas estão a *macaca fascicularis*, *sus scrofa* (javali), *vulpes vulpes* (raposa-vermelha), *mus musculus* (rato-doméstico) entre outros.

Relativamente às funções desempenhadas, a proteína em estudo é um recetor específico para o núcleo heptapeptídeo comum ao hormônio adrenocorticotrófico e alfa, beta e gama-MSH e desempenha um papel central na homeostase energética e no crescimento somático. Este recetor é mediado por proteínas G que estimulam a adenilato ciclase (cAMP). É importante referir que as sequências homólogas desempenham a mesma função, o que é verificado pelo elevado grau de compatibilidade com o gene em estudo.

4.2 SEC16B - Proteína transportadora da SEC16B

Para realizar a análise de homologia para o segundo gene em estudo, SEC16B, utilizou-se novamente a ferramenta BLAST tal como referido para o primeiro gene. Assim, foi possível encontrar as suas várias sequências homólogas.

A Figura 4 apresenta algumas das sequências homólogas, sendo que a primeira linha corresponde à proteína codificada pelo gene em estudo o qual possui o nome de proteína *Protein transport protein Sec16B*, com o organismo correspondente *Homo sapiens (Human)* e com comprimento de sequência 1060 aminoácidos.

A sequência mais homóloga possui número de acesso Q75NY9.1, que corresponde ao gene *SEC16B* apresentando o nome *Protein transport protein sec16* e com comprimento de 1052 aminoácidos. O organismo associado a este gene corresponde ao *Bos Taurus*. Analisando a sequência e os valores apresentados, verificamos que apresenta um **Query Cover** de 99%, **Total Score** e **Max Score** de 1501, **Percent Identities** de 76,22%, um **E-Value** de 0 e um comprimento de sequência de 1052 aminoácidos.

Sequences producing significant alignments				Download	Select columns	Show	100	
<input checked="" type="checkbox"/> select all 21 sequences selected				GenPept	Graphics	Distance tree of results	Multiple alignment	MSA Viewer
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein Sec16B; AltName: Full=Leucine zipper transcription regulator 2; AltName: ... Homo sapiens	Homo sapiens	2180	2180	100%	0.0	100.00%	1060	Q96JF7.2
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein Sec16B; AltName: Full=Regucalcin gene promoter region-related protein... Bos taurus	Bos taurus	1501	1501	99%	0.0	76.22%	1052	Q75NY9.1
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein Sec16B; AltName: Full=Leucine zipper transcription regulator 2; AltName: ... Mus musculus	Mus musculus	1454	1454	100%	0.0	72.75%	1051	Q91XT4.3
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein Sec16B; AltName: Full=Leucine zipper transcription regulator 2; AltName: ... Rattus norvegicus	Rattus norvegicus	1446	1446	100%	0.0	72.50%	1057	Q75N33.1
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein Sec16B; AltName: Full=Regucalcin gene promoter region-related protein... Oryctolagus cuni...	Oryctolagus cuni...	1362	1362	100%	0.0	69.20%	1045	Q6BCB4.1
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein Sec16B; AltName: Full=Regucalcin gene promoter region-related protein... Gallus gallus	Gallus gallus	690	690	92%	0.0	45.32%	929	Q6AW68.1
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein SEC16A; AltName: Full=SEC16 homolog A; Short=p250 [Homo sapiens] Homo sapiens	Homo sapiens	480	480	43%	7e-144	50.21%	2357	Q15027.4
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein SEC16A; AltName: Full=SEC16 homolog A; Short=p250 [Mus musculus] Mus musculus	Mus musculus	478	478	43%	3e-143	49.89%	2357	E9QAT4.1
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein sec16; AltName: Full=Protein transport protein sec16 [Neurospora cr... Neurospora cras...	Neurospora cras...	139	139	37%	8e-32	28.63%	2084	Q9HEC9.1
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein SEC16; AltName: Full=Protein transport protein SEC16 [Chaetomium glo... Chaetomium glo...	Chaetomium glo...	135	135	38%	7e-31	27.54%	1865	Q2GXM1.1
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein SEC16B homolog; AltName: Full=Protein MAIGO 5 homolog [Arabidopsis thali... Arabidopsis thali...	Arabidopsis thali...	127	127	44%	3e-28	26.40%	1361	Q8FGK8.1
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein SEC16A homolog; AltName: Full=Protein MAIGO 5 [Arabidopsis thaliana] Arabidopsis thali...	Arabidopsis thali...	126	126	21%	5e-28	34.36%	1350	Q8FGK9.1
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein sec16; AltName: Full=Protein transport protein sec16 [Aspergillus ni... Aspergillus niger...	Aspergillus niger...	110	110	30%	3e-23	27.27%	1914	A2R4T4.2
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein sec16; AltName: Full=Protein transport protein sec16 [Aspergillus fu... Aspergillus fumig...	Aspergillus fumig...	106	106	27%	7e-22	25.22%	1832	Q4WD95.1
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein sec16; AltName: Full=Protein transport protein sec16 [Aspergillus fis... Aspergillus fisch...	Aspergillus fisch...	105	105	27%	1e-21	26.25%	1835	A1DLN3.1
<input checked="" type="checkbox"/> RecName: Full=Protein transport protein sec-16A.1; AltName: Full=Secretory cargo traffic protein sec-16A.1 [Caenorhabditis e... Caenorhabditis e...	Caenorhabditis e...	103	103	37%	6e-21	23.02%	1232	P34643.1
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein sec16; AltName: Full=Protein transport protein sec16 [Aspergillus or... Aspergillus oryza...	Aspergillus oryza...	101	101	27%	2e-20	26.77%	1843	Q2U968.1
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein sec16; AltName: Full=Protein transport protein sec16 [Aspergillus ter... Aspergillus terre...	Aspergillus terre...	98.6	98.6	37%	2e-19	24.41%	1806	Q0CGY7.1
<input checked="" type="checkbox"/> RecName: Full=COPII coat assembly protein sec16; AltName: Full=Protein transport protein sec16 [Aspergillus cl... Aspergillus claya...	Aspergillus claya...	96.3	96.3	34%	9e-19	24.53%	1875	A1CME3.1

Figura 4: Sequências homólogas à sequência da proteína em estudo SEC16B.

O valor de **Query Cover** indica que 99% da sequência é abrangida pela sequência em estudo. O valor de **Total Score** e **Max Score** de 1501 apresenta diferenças significativas comparativamente ao valor que apresenta a proteína em estudo, sendo este 2180. O valor de **Percent Identities** de 76,22%, indica a extensão em que as duas sequências têm os mesmos resíduos na mesma posição. Quanto ao **E-Value**, com o valor zero, isto significa uma elevada compatibilidade e homologia entre as sequências.

Analisando ainda os dados obtidos e apresentados na imagem acima, entre os organismos associados às sequências homólogas encontram-se o *bos taurus*, *mus musculus*, *rattus norvegicus*, entre outros.

No que diz respeito às funções desempenhadas, a proteína em estudo desempenha um papel na organização dos locais de saída do retículo endoplasmático (ERES), também conhecido como retículo endoplasmático transicional (tER), é necessário para o tráfego de carga secretora do retículo endoplasmático para o aparelho de Golgi, está envolvido na biogênese do peroxissomo e regula o transporte dos fatores de biogênese peroxissomal PEX3 e PEX16 do RE para os peroxissomos. É importante referir que as sequências homólogas desempenham funções semelhantes ou mesmo idênticas.

5 Ferramentas de análise das propriedades da proteína

Recorrendo a várias ferramentas de bioinformática, foi possível reunir diversas informações dos genes em estudo o que permitiu realizar uma análise das propriedades das proteínas codificadas por esses genes.

5.1 Recetor de Melanocortina 4 (MC4R)

5.1.1 Proteína codificada pelo gene MC4R

O gene MC4R codifica uma proteína com o mesmo nome, ou seja, *recetor de melanocortina 4*, tal como já referido anteriormente. A proteína pertence ao organismo *Homo Sapiens (Human)*. Esta informação foi confirmada recorrendo à base de dados UniProt, como se pode verificar na Figura 5

P32245 · MC4R_HUMAN			
Protein ¹	Melanocortin receptor 4	Amino acids	332
Gene ²	MC4R	Protein existence ³	Evidence at protein level
Status ¹	UniProtKB reviewed (Swiss-Prot)	Annotation score ¹	5.3
Organism ¹	Homo sapiens (Human)		

Figura 5: Proteína codificada pelo gene MC4R. Retirado de UniProt.

5.1.2 Estrutura da Proteína

O recetor de melanocortina 4 é um recetor acoplado à proteína G (GPCR) que tem uma estrutura de sete domínios transmembranares. A estrutura da proteína em estudo, apresentada na Figura 6 consiste em um N-terminal extracelular, sete hélices transmembranares e um C-terminal intracelular.

O domínio N-terminal extracelular é composto por uma combinação de hélices alfa e folhas beta e liga-se aos ligantes peptídicos de melanocortina. Relativamente às hélices transmembranares, estas formam uma bolsa no espaço extracelular, onde o ligante se liga, e abrangem a membrana celular. O domínio C-terminal intracelular é composto por uma combinação de hélices alfa e folhas beta e está envolvido na transdução do sinal, na interação com as proteínas G e na ativação de vias de sinalização intracelular.

As informações estruturais de MC4R são utilizadas para o desenvolvimento de novos medicamentos direcionados a este recetor, isto porque o design de medicamentos com base na estrutura é uma ferramenta poderosa para a descoberta de novos medicamentos.

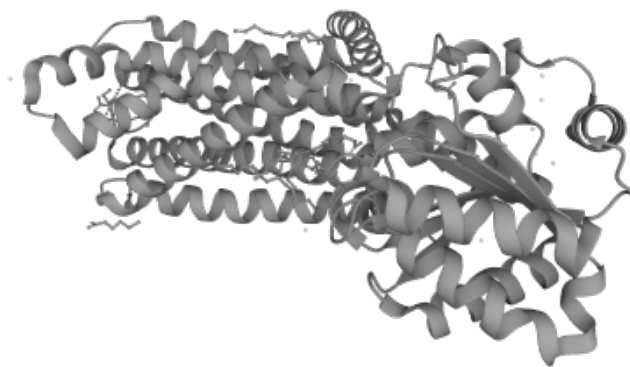


Figura 6: Estrutura da proteína MC4R. Retirado de UniProt.

5.1.3 Localização sub-celular

Através da base de dados UniProt e da ferramenta Hum-mPLoc 2.0 e LOCtree, foi possível verificar que a proteína localiza-se, principalmente, na **membrana plasmática**. Além disso, pode também ser encontrada no citoplasma e no núcleo.

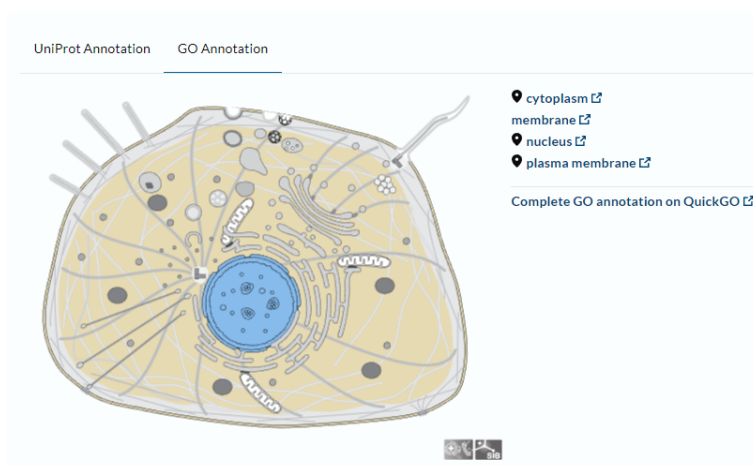


Figura 7: Localização sub-celular. Retirado de UniProt.

Hum-mPLoc 2.0: Predicting subcellular localization of human proteins including those with multiple sites

[Read Me](#) | [Data](#) | [Citation](#) |

Your input sequence (332aa) is:

```
>sp|P32245|MC4R_HUMAN OS=Homo sapiens OX=9606 GN=MC4R PE=1 SV=2
MVNSTHRGMHTSLHLWNRSSYRLHSNASESLGKGYSDGGCYEQLFVSPEVFTLVGLVISLL
ENILVIVIAKKNKLNLSHPMYFFICSLAVADMLVSVSNGSETIVITLLNSTDTDAQSFTVN
IDNVDSVICSSLLASICSLLSIAVDRYFTIFALQYHNIMTVKRVGHIISCIWAACVTS
GILFIYSDSSAVIICLITMFFFTMLALMASLYVHMFLMARLHIKRIAVLPGTGAIKQGAN
MKGAILTLILIGVFVVCWAPFFLHLIFVISCQNPYCVCFMSHFNLYLILMCNSIIDPL
IYALRSQELRKTKEIICCYPLGGLCDLSSRY
```


----- Hum-mPLoc 2.0 Computation Result -----

Query protein	Predicted location(s)
sp P32245 MC4R_HUMAN OS=Homo sapiens OX=9606 GN=MC4R PE=1 SV=2	Plasma membrane.

Figura 8: Localização sub-celular. Retirado de Hum-mPLoc 2.0.

Domain: Eukaryota

Details	Protein ID	Score	Expected Accuracy	Localization Class	Gene Ontology Terms	Annotation Type
Details	sp P32245 MC4R_HUMAN	100	99%	plasma membrane	integral to membrane GO:0016021(TAS); nucleus GO:0005634(IEA); plasma membrane GO:0005886(TAS);	PSI-BLAST

 Mouse click on Details/Protein ID leads to the detailed description of a prediction.

Please cite:

Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Nielsen H, Rost B et al. LocTree3 prediction of localization. *Nucleic Acids Research* 2014. PMID: 24848019

Goldberg T, Hamp T and Rost B. LocTree2 predicts localization for all domains of life. *Bioinformatics* 2012, 28:1458-1465. PMID: 22962467

Figura 9: Localização sub-celular. Retirado de LOCTree.

5.1.4 Funções gerais

MC4R é um recetor específico do núcleo de heptapéptido comum à hormona adrenocorticotrófica e à hormona alfa, beta, e gama-MSH. Desempenha um papel central na homeostase energética e no crescimento somático. Este recetor é mediado por proteínas G que estimulam a adenilato ciclase (cAMP).

Além disso, através da base de dados UniProt encontraram-se mais algumas funções gerais, apresentadas na Figura 10. As funções moleculares incluem atividade de recetor acoplado a proteína G, de recetor de melanocortina e de recetor da hormona estimulante de melanócitos. Pode também ligar-se a neuropeptídeos, a proteínas ubiquitina ligase e a hormonas peptídicas.

Quanto a atividades biológicas, estas incluem regulação do comportamento alimentar, processo metabólico de reserva de energia, ativação da via de sinalização do recetor acoplado à proteína G por ativação da atividade da adenilato ciclase, entre outros.

Molecular Function	G protein-coupled receptor activity	Source:GO_Central	1 Publication
Molecular Function	melanocortin receptor activity	Source:Protinc	1 Publication
Molecular Function	melanocyte-stimulating hormone receptor activity	Source:UniProtKB	1 Publication
Molecular Function	neuropeptide binding	Source:UniProtKB	1 Publication
Molecular Function	peptide hormone binding	Source:Ensembl	
Molecular Function	ubiquitin protein ligase binding	Source:UniProtKB	1 Publication
Biological Process	adenylate cyclase-activating G protein-coupled receptor signaling pathway	Source:BioRx-UCL	2 Publications
Biological Process	adenylate cyclase-modulating G protein-coupled receptor signaling pathway	Source:Protinc	1 Publication
Biological Process	diet induced thermogenesis	Source:Ensembl	
Biological Process	energy reserve metabolic process	Source:Ensembl	
Biological Process	feeding behavior	Source:Protinc	1 Publication
Biological Process	insulin secretion	Source:Ensembl	
Biological Process	negative regulation of feeding behavior	Source:Ensembl	
Biological Process	positive regulation of bone resorption	Source:HGNC-UCL	1 Publication
Biological Process	regulation of eating behavior	Source:ARUK-UCL	
Biological Process	regulation of grooming behavior	Source:Ensembl	
Biological Process	regulation of metabolic process	Source:GO_Central	1 Publication
Biological Process	response to insulin	Source:Ensembl	
Biological Process	response to melanocyte-stimulating hormone	Source:ARUK-UCL	

Figura 10: Funções gerais associadas à proteína. Retirado de UniProt.

5.1.5 Locais de expressão do gene

Através das ferramentas *Genevisible* e *The Human Protein Atlas* é possível identificar os locais onde o gene é expresso e, conseqüentemente, onde a proteína codificada pelo gene se encontra.

Analisando os dados obtidos, verificamos que os órgãos que mais se destacam e onde o gene é mais expresso, são o **cérebro**, destacando-se o hipotálamo, basal ganglia, tálamo, córtex cerebral, entre outros; os **olhos**, em específico, a retina; e o **sistema reprodutor feminino**, nomeadamente na trompa de Falópio. Sendo o cérebro o órgão onde MC4R é mais expresso, pois a sua expressão ocorre em diversas zonas deste órgão. Além desses órgãos, o gene também é expresso, embora de uma forma mais reduzida, no trato gastrointestinal, como por exemplo, no estômago e no duodeno, entre vários outros representados na Figura 11 e Figura 12.

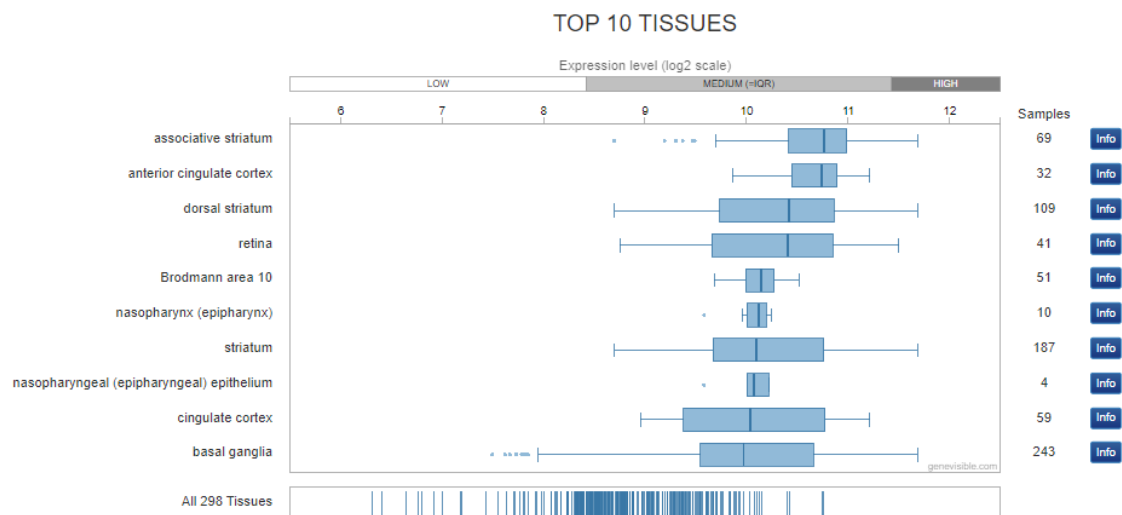


Figura 11: Tecidos onde MC4R é expresso. Retirado de Genevisible.

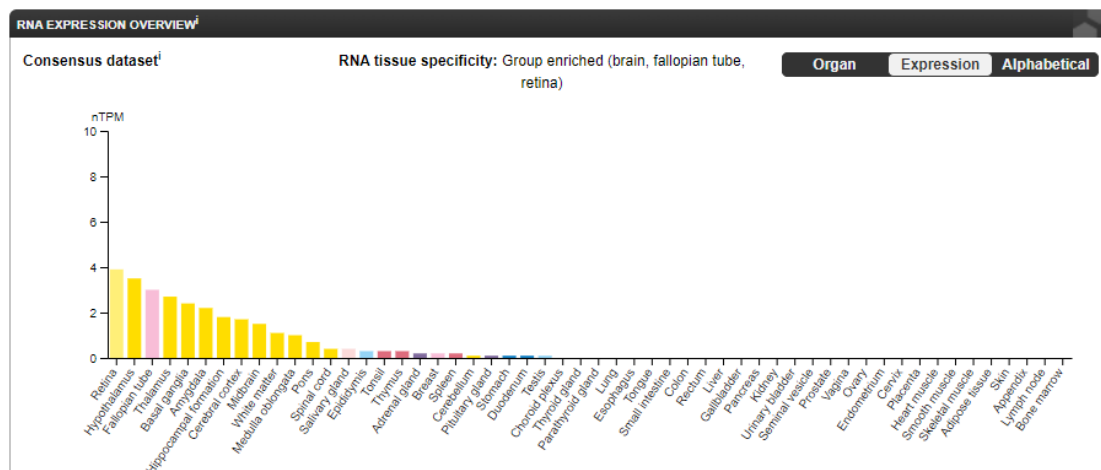


Figura 12: Locais de expressão de MC4R. Retirado de The Human Protein Atlas.

5.1.6 Família e Domínios

Com a utilização da ferramenta *InterPro* foi possível determinar a família a que a proteína MC4R pertence e também o domínio a si associado, tendo-se obtido os resultados apresentados na Figura 13. Adicionalmente, recorreu-se também ao banco de dados *NCBI Conserved Domain* para confirmar a existência do domínio conservado.

A proteína em estudo pertence à família de recetores acoplados à proteína G (GPCR). GPCRs são uma grande família de proteínas transmembranares que estão envolvidas numa ampla gama de processos fisiológicos e são alvos de muitos fármacos. São designados por recetores acoplados à proteína G uma vez que interagem com proteínas G para mediar as respostas celulares aos sinais extracelulares. A família GPCR é a maior família de recetores no genoma humano e é dividida em várias subfamílias com base na similaridade de sequência e especificidade do ligante. O MC4R pertence à subfamília dos GPCRs semelhantes à rodopsina, que também inclui outros recetores, como os recetores opioides, adrenérgicos e serotoninérgicos.

Através do banco de dados *NCBI Conserved Domain*, Figura 14, verificou-se a existência de domínios conservados, nomeadamente os sete domínios transmembranares, 7tmA_MC4R, que são característicos dos recetores acoplados à proteína G. Assim, a superfamília principal é a Superfamília 7tm_GPCRs.

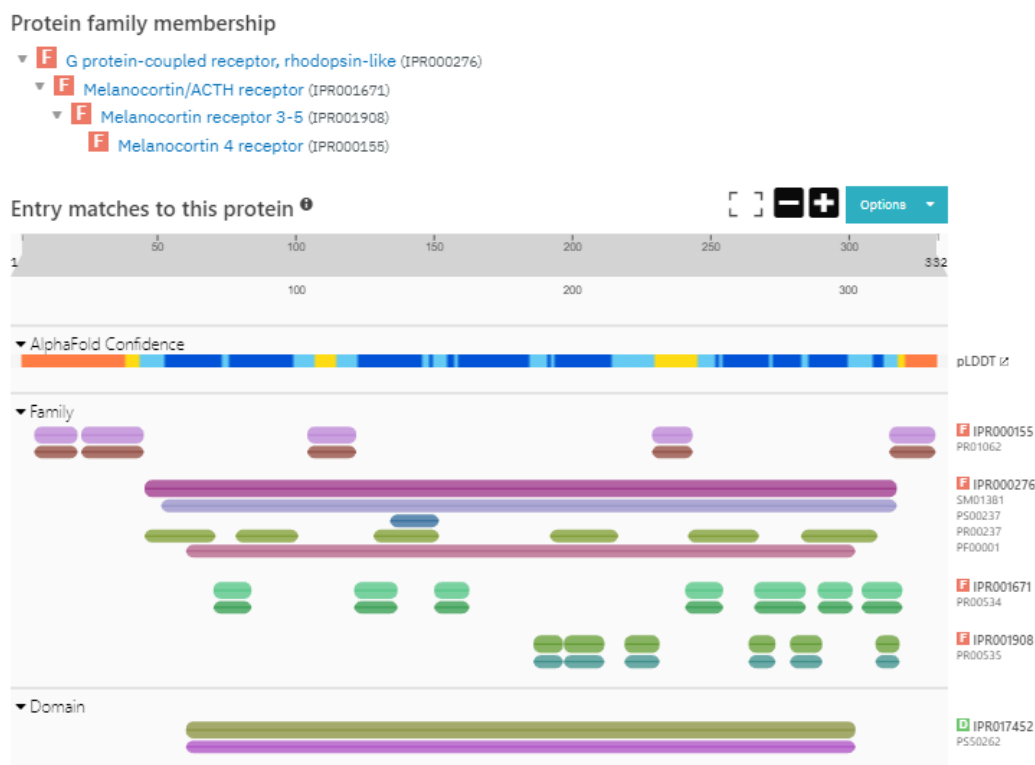


Figura 13: Família e Domínios da proteína. Retirado de InterPro.

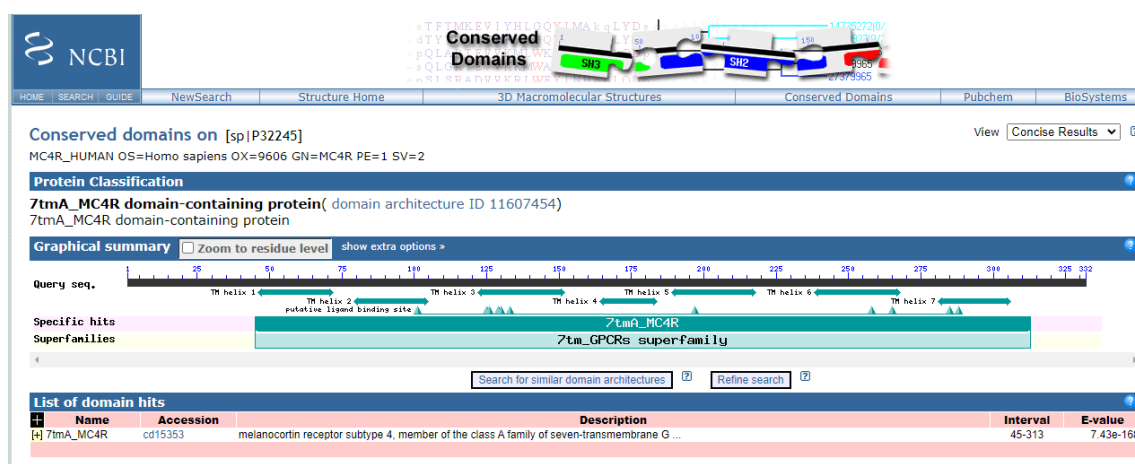


Figura 14: Domínios conservados proteína. Retirado de NCBI Conserved Domain.

5.1.7 Domínios Transmembranares

Para verificar a existência de domínios transmembranares, recorreu-se às ferramentas *DeepTMHMM* e *Phobius*. A informação obtida está presente na Figura 16.

Com a análise dos gráficos, podemos verificar que o recetor de melanocortina 4 possui sete domínios transmembranares (7TM), tal como já referido anteriormente, sendo estes domínios compostos por alfa-hélices.

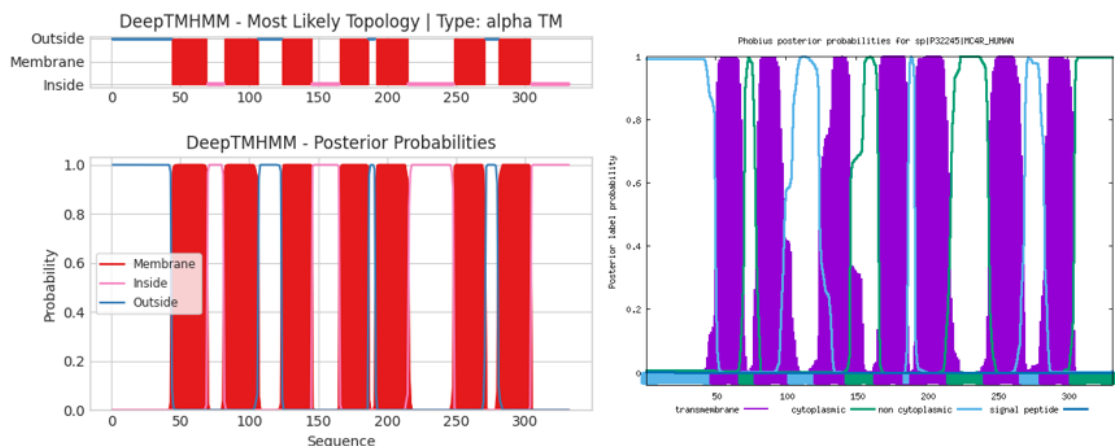


Figura 15: Domínios Transmembranares da proteína. Retirado de DeepTMHMM e Phobius, respectivamente.

5.1.8 Alterações pós-tradução

Alterações pós-tradução são modificações que ocorrem na proteína após a tradução da sequência de nucleótidos em aminoácidos. Estas modificações podem determinar a atividade, a localização, e interações com outras proteínas. Além de alterarem as propriedades da proteína, tornando-se relevantes para aumentar a diversidade funcional do proteoma.

Glicosilação é um processo que consiste na ligação de moléculas de açúcar a proteínas, sendo importante para a localização sub celular, interações célula-célula e interações proteína-ligando. Através da ferramenta NetNGlyc foram identificados 5 locais de N-glicosilação na proteína em estudo, uma vez que existem 5 aminoácidos que ultrapassam o valor de threshold.

Fosforilação é o mecanismo mais comum de regulação da função da proteína e transmissão de sinais na célula. Além disso, é um mecanismo importante de regulação celular, pois permite que as células respondam rapidamente a estímulos externos e internos. Este ocorre pela adição de um grupo fosfato a uma proteínas ou outra molécula, e é importante para a ligação de proteínas, ciclos celulares, vias de sinalização e regulação de enzimas. Através da ferramenta NetPhos são identificados 30 locais de fosforilação de aminoácidos, nomeadamente da serina, treonina e tirosina. Estes locais estão representados pelas linhas que ultrapassam o valor de threshold.

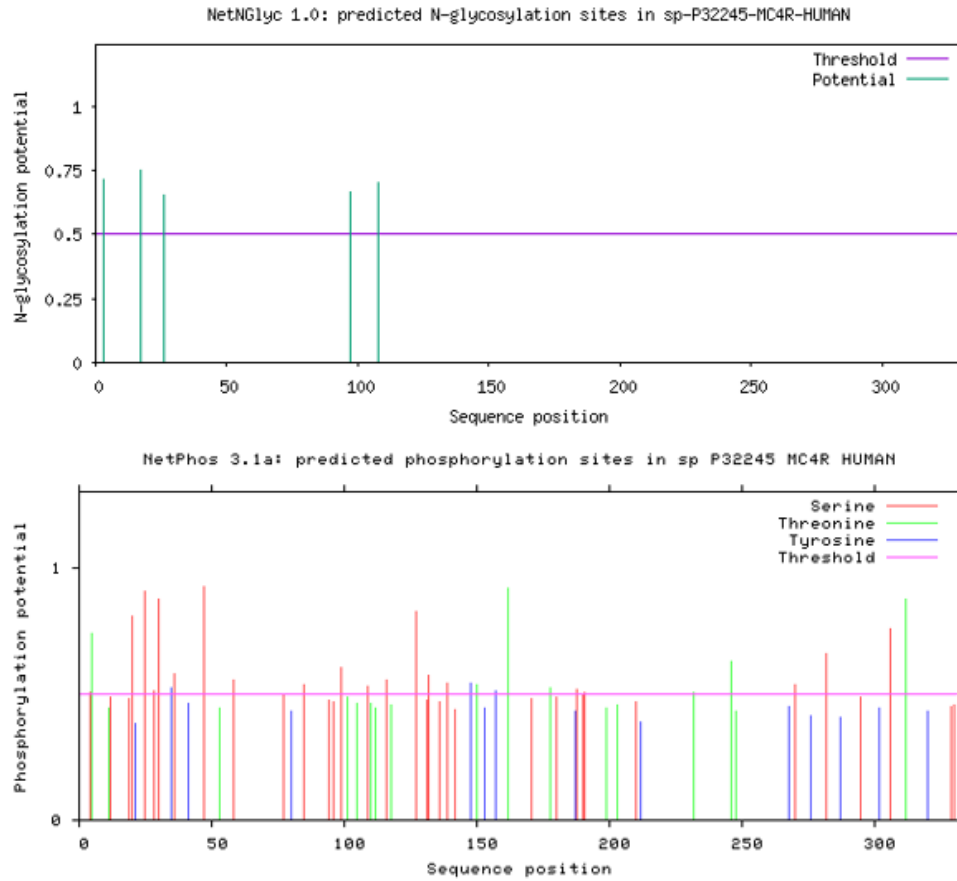


Figura 16: Alterações pós-tradução: glicosilação e fosforilação. Retirado de NetNGlyc e NetPhos, respectivamente.

5.2 Proteína Transportadora da Proteína SEC16B

6 Alinhamento múltiplo e filogenia

7 Regulação

8 Comparação de variantes do gene e o seu impacto biológico

9 Conclusões

Bibliografia

- [1] J. P. L. Gonçalves, D. Palmer, and M. Meldal, “Mc4r agonists: Structural overview on antiobesity therapeutics,” *Trends in Pharmacological Sciences*, 2018. DOI:<https://doi.org/10.1016/j.tips.2018.01.004>.
- [2] P. J. D. Delhanty, E. Bouw, M. Huisman, R. M. L. Vervenne, A. P. N. Themmen, A. J. van der Lely, and E. L. T. van den Akker, “Functional characterization of a new human melanocortin-4 receptor homozygous mutation (n72k) that is associated with early-onset obesity,” *Molecular Biology Reports*, vol. 41, p. 7967–7972, 2014. <https://doi.org/10.1007/s11033-014-3691-7>.
- [3] V. HAINER, I. A. HAINEROVÁ, M. KUNEŠOVÁ, R. T. BRAUNEROVÁ, H. ZAMRAZILOVÁ, and B. BENDLOVÁ, “Melanocortin pathways: suppressed and stimulated melanocortin-4 receptor (mc4r),” *Physiological Research*, vol. 69, pp. 245–254, 2020. DOI:10.33549/physiolres.934512.
- [4] J. W. Sull, G. Kim, and S. H. Jee, “Association of mc4r (rs17782313) with diabetes and cardiovascular disease in korean men and women,” *BMC Medical Genetics*, 2020. <https://doi.org/10.1186/s12881-020-01100-3>.

ANEXOS