

## Taller 7- Uniprot

### Grupo 1:

- Ana Sofia Villa Benavides 201923361
- Ronald Yesid Diaz Pardo 202111309
- Gabriela Moreno Prieto 201713956

1. Utilice UniProtKB para estudiar la proteína ACE2 de humanos, altamente relacionada con la pandemia COVID-19.

UniProtKB 48 results

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
Q9BYF1	ACE2_HUMAN	Angiotensin-converting enzyme 2[...]	ACE2, UNQ868/PRO1885	Homo sapiens (Human)	805 AA
P01019	ANGT_HUMAN	Angiotensinogen[...]	AGT, SERPINA8	Homo sapiens (Human)	476 AA

#### a. Acerca de su función:

##### Function<sup>i</sup>

Essential counter-regulatory carboxypeptidase of the renin-angiotensin hormone system that is a critical regulator of blood volume, systemic vascular resistance, and thus cardiovascular homeostasis (PubMed:27217402). Converts angiotensin I to angiotensin 1-9, a nine-amino acid peptide with anti-hypertrophic effects in cardiomyocytes, and angiotensin II to angiotensin 1-7, which then acts as a beneficial vasodilator and anti-proliferation agent, counterbalancing the actions of the vasoconstrictor angiotensin II (PubMed:10924499, PubMed:10969042, PubMed:11815627, PubMed:14504186, PubMed:19021774). Also removes the C-terminal residue from three other vasoactive peptides, neurotensin, kinetensin, and des-Arg bradykinin, but is not active on bradykinin (PubMed:10969042, PubMed:11815627). Also cleaves other biological peptides, such as apelin (apelin-13, [Pyr1]apelin-13, apelin-17, apelin-36), casomorphins (beta-casomorphin-7, neocasomorphin) and dynorphin A with high efficiency (PubMed:11815627, PubMed:27217402, PubMed:28293165). In addition, ACE2 C-terminus is homologous to collectrin and is responsible for the trafficking of the neutral amino acid transporter SL6A19 to the plasma membrane of gut epithelial cells via direct interaction, regulating its expression on the cell surface and its catalytic activity (PubMed:18424768, PubMed:19185582). <sup>6 Publications</sup>

(Microbial infection) Acts as a receptor for human coronaviruses SARS-CoV and SARS-CoV-2, as well as human coronavirus NL63/HCoV-NL63. <sup>13 Publications</sup>

#### i. ¿Cuál es su rol principal?

El rol principal de la proteína ACE2 es actuar como una carboxipeptidasa reguladora clave en el sistema renina-angiotensina, controlando el volumen sanguíneo, la resistencia vascular y la homeostasis cardiovascular. En relación al COVID-19, ACE2 actúa como el receptor para el virus SARS-CoV-2, permitiendo su entrada a las células humanas.

#### ii. ¿Posee actividad catalítica?

<b>Catalytic activity<sup>i</sup></b>	
<a href="#">Rhea 26554</a>	angiotensin II + H <sub>2</sub> O = angiotensin-(1-7) + L-phenylalanine This reaction proceeds in the forward direction. EC:3.4.17.23 ( <a href="#">UniProtKB</a>   <a href="#">ENZYME</a>   <a href="#">Rhea</a> ) <a href="#">View Rhea reaction</a>
<a href="#">Rhea 63532</a>	angiotensin I + H <sub>2</sub> O = angiotensin-(1-9) + L-leucine This reaction proceeds in the forward direction. <a href="#">View Rhea reaction</a>
<a href="#">Rhea 63536</a>	bradykinin(1-8) + H <sub>2</sub> O = bradykinin(1-7) + L-phenylalanine This reaction proceeds in the forward direction. <a href="#">View Rhea reaction</a>
<a href="#">Rhea 63540</a>	H <sub>2</sub> O + neurotensin = L-leucine + neurotensin-(1-12) This reaction proceeds in the forward direction. <a href="#">View Rhea reaction</a>
<a href="#">Rhea 63572</a>	H <sub>2</sub> O + neurotensin-(1-8) = L-arginine + neurotensin-(1-7) This reaction proceeds in the forward direction. <a href="#">View Rhea reaction</a>
<a href="#">Rhea 63544</a>	H <sub>2</sub> O + kinetensin = kinetensin-(1-8) + L-leucine This reaction proceeds in the forward direction. 

Si posee actividad catalítica en 12 reacciones diferentes.

iii. ¿En qué vías metabólicas interviene?

### Enzyme and pathway databases

<a href="#">BRENDA</a>	<a href="#">3.4.15.1</a> 2681 <a href="#">3.4.17.23</a> 2681	<a href="#">SABIO-RK</a>	<a href="#">Q9BYF1</a>
<a href="#">BioCyc</a>	<a href="#">MetaCyc:ENSG00000130234-MONOMER</a>	<a href="#">SIGNOR</a>	<a href="#">Q9BYF1</a>
<a href="#">PathwayCommons</a>	<a href="#">Q9BYF1</a>	<a href="#">Signalink</a>	<a href="#">Q9BYF1</a>
<a href="#">Reactome</a>	<a href="#">R-HSA-2022377</a> Metabolism of Angiotensinogen to Angiotensins <a href="#">R-HSA-9678110</a> Attachment and Entry <a href="#">R-HSA-9679191</a> Potential therapeutics for SARS <a href="#">R-HSA-9694614</a> Attachment and Entry <a href="#">R-HSA-9733458</a> Induction of Cell-Cell Fusion	<a href="#">ENZYME</a>	<a href="#">Search...</a>

De acuerdo con la sección de enzimas y vías metabólicas, en específico en relación con reactome estas son las vías

Metabolismo de hormonas peptídicas
Sistema renina-angiotensina
Regulación de la presión arterial por señales químicas
Infección por SARS-CoV-2
Vías de señalización de neuropéptidos
Vías antihipertensivas
Vías de vasodilatación
Homeostasis cardiovascular

iv. ¿Con qué proteínas interacciona?

## Interaction<sup>i</sup>

### Subunit<sup>i</sup>

Homodimer (PubMed:32132184).

Interacts with the catalytically active form of **TMPPSS2** (PubMed:21068237).

Interacts with SLC6A19; this interaction is essential for expression and function of SLC6A19 in intestine (By similarity).

Interacts with ITGA5.ITGB1 (PubMed:15276642, PubMed:33102950).

Probably interacts (via endocytic sorting signal motif) with AP2M1; the interaction is inhibited by phosphorylation of Tyr-781 (PubMed:33436498).

Interacts (via PDZ-binding motif) with NHERF1 (via PDZ domains); the interaction may enhance ACE2 membrane residence (PubMed:34189428). By Similarity 6 Publications

(Microbial infection) Interacts with SARS coronavirus/SARS-CoV spike protein. 4 Publications

(Microbial infection) Interacts with SARS coronavirus-2/SARS-CoV-2 spike protein (via RBD domain). 10 Publications

(Microbial infection) Interacts with human coronavirus NL63 spike protein. 1 Publication

(Microbial infection) Interacts with human coronavirus NL63/HCoV-NL63 spike glycoprotein. 1 Publication

### Processed angiotensin-converting enzyme 2

(Microbial infection) Interacts with SARS coronavirus-2/SARS-CoV-2 spike protein; the interaction is increased by AVP/Arg-vasopressin with which they may form a complex. 1 Publication

### Binary interactions<sup>i</sup>

Q9BYF1 has binary interactions with 26 proteins

ACE2 interactúa directamente con 26 proteínas, esto lo pudimos observar en la sección de interacción, particularmente en *binary interaction*. De estas podemos destacar principalmente, que ACE2 interactúa con la proteína de spike NL63 del coronavirus SARS-2/SARS-CoV-2. A continuación, está un pantallazo de la tabla con las 26 proteínas reportadas, aquí se puede ver también información sobre el número de experimentos que respaldan esta interacción.

TYPE	ENTRY 1	ENTRY 2	NUMBER OF EXPERIMENTS	INTACT
-- Select --	Q9BYF1			
BINARY	Q9BYF1	ACE2 Q9BYF1	8	EBI-7730807, EBI-7730807 <a href="#">↗</a>
BINARY	Q9BYF1	AGT PRO_0000032457 P01019	2	EBI-7730807, EBI-25493366 <a href="#">↗</a>
BINARY	Q9BYF1	AGT PRO_0000032458 P01019	5	EBI-7730807, EBI-6622938 <a href="#">↗</a>
BINARY	Q9BYF1	AP2M1 Q96CW1	2	EBI-7730807, EBI-297683 <a href="#">↗</a>
BINARY	Q9BYF1	APP PRO_0000000092 P05067	3	EBI-7730807, EBI-821758 <a href="#">↗</a>
BINARY	Q9BYF1	CLEC4M Q9H2X3	3	EBI-7730807, EBI-1391211 <a href="#">↗</a>
BINARY	Q9BYF1	DEFA5 PRO_0000417390 Q01523	2	EBI-7730807, EBI-25634589 <a href="#">↗</a>
BINARY	Q9BYF1	GRM2 Q14416	2	EBI-7730807, EBI-10232876 <a href="#">↗</a>
BINARY	Q9BYF1	HSPA5 P11021	6	EBI-7730807, EBI-354921 <a href="#">↗</a>
BINARY	Q9BYF1	ITGB1 P05556	4	EBI-7730807, EBI-703066 <a href="#">↗</a>
BINARY	Q9BYF1	KDM1A O60341	19	EBI-7730807, EBI-710124 <a href="#">↗</a>
BINARY	Q9BYF1	KNG1 PRO_0000006688 P01042	5	EBI-7730807, EBI-6623273 <a href="#">↗</a>
XENO	Q9BYF1	Kpna2 P52293	4	EBI-7730807, EBI-3043908 <a href="#">↗</a>
BINARY	Q9BYF1	NHERF1 O14745	7	EBI-7730807, EBI-349787 <a href="#">↗</a>
BINARY	Q9BYF1	NTS PRO_0000019524 P30890	2	EBI-7730807, EBI-6655817 <a href="#">↗</a>
BINARY	Q9BYF1	PDZK1 Q5T2W1	3	EBI-7730807, EBI-349819 <a href="#">↗</a>
XENO	Q9BYF1	S A0A6B9WHD3	2	EBI-7730807, EBI-25496413 <a href="#">↗</a>
XENO	Q9BYF1	S A0A6G6A1M4	4	EBI-7730807, EBI-26997256 <a href="#">↗</a>
XENO	Q9BYF1	S A0A6M3G9R1	4	EBI-7730807, EBI-26997195 <a href="#">↗</a>
XENO	Q9BYF1	S P0DTC2	252	EBI-7730807, EBI-25474821 <a href="#">↗</a>
XENO	Q9BYF1	S PRO_0000449647 P0DTC2	2	EBI-7730807, EBI-25490323 <a href="#">↗</a>
XENO	Q9BYF1	S P59594	57	EBI-7730807, EBI-15582614 <a href="#">↗</a>
XENO	Q9BYF1	S PRO_0000037209 P59594	3	EBI-7730807, EBI-25475261 <a href="#">↗</a>
XENO	Q9BYF1	S Q5GDB5	2	EBI-7730807, EBI-25566398 <a href="#">↗</a>
XENO	Q9BYF1	S Q6Q1S2	9	EBI-7730807, EBI-15814420 <a href="#">↗</a>
BINARY	Q9BYF1	SFTPD P35247	2	EBI-7730807, EBI-11316157 <a href="#">↗</a>
BINARY	Q9BYF1	SHANK1 Q9Y566	2	EBI-7730807, EBI-3442234 <a href="#">↗</a>
BINARY	Q9BYF1	SLC6A19 Q695T7	4	EBI-7730807, EBI-25475705 <a href="#">↗</a>
BINARY	Q9BYF1	SNX27 Q96L92	2	EBI-7730807, EBI-2514865 <a href="#">↗</a>
BINARY	Q9BYF1	TMPPSS2 O15393	3	EBI-7730807, EBI-12549863 <a href="#">↗</a>
BINARY	Q9BYF1	VIM P08670	4	EBI-7730807, EBI-353844 <a href="#">↗</a>

b. Acerca de su secuencia y estructura:

i. ¿Cuántas isoformas se han identificado?

## Sequence & Isoform<sup>i</sup>

Align isoforms (2)  Add isoforms

Sequence status<sup>i</sup> | Complete



This entry describes 2 isoforms<sup>i</sup> produced by **Alternative splicing**.






















Se han identificado 2 isoformas.

ii. ¿Cuántos sitios activos posee?

Features

Showing features for binding site: **Active site**

  Download

TYPE	ID	POSITION(S)	DESCRIPTION
Binding site	169		chloride (UniProtKB   CHEBI C <sup>1</sup> )   
Binding site	273		substrate 
Binding site	345-346		substrate 
Binding site	374		Zn <sup>2+</sup> (UniProtKB   CHEBI C <sup>1</sup> ); catalytic  
<b>Active site</b>	375		Proton acceptor  
Binding site	378		Zn <sup>2+</sup> (UniProtKB   CHEBI C <sup>1</sup> ); catalytic  
Binding site	402		Zn <sup>2+</sup> (UniProtKB   CHEBI C <sup>1</sup> ); catalytic  
Binding site	477		chloride (UniProtKB   CHEBI C <sup>1</sup> )   
Binding site	481		chloride (UniProtKB   CHEBI C <sup>1</sup> )   
<b>Active site</b>	505		Proton donor  



Posee 2 sitios activos, uno en la posición 375 y otro en la 505. Ambos están descritos como aceptores de protones.

iii. Entre las posibles variantes, ¿Hay alguna/s que inhiban la interacción con la proteína Spike de SARS-CoV-2? Si lo hay, cite la bibliografía correspondiente

Disease & Variants<sup>i</sup>

Features

Showing features for mutagenesis<sup>1</sup>, natural variant<sup>1</sup>:


► Mutagenesis	24-26	Slightly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	41	Strongly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	68	Slightly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	82-84	<b>Inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	355	Strongly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	357	Strongly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	383	Slightly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	389	Slightly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	393	Slightly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	425-427	Slightly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 
► Mutagenesis	559	Slightly <b>inhibits</b> interaction with SARS-CoV spike glycoprotein. 

Manual assertion based on experiment<sup>1</sup>

Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2.

Li W, Zhang C, Sul J, Kuhn JH, Moore MJ, Luo S, Wong S-K, Huang L-C, Xu K, [...] , Farzan M.

[View abstract](#)

 PubMed C<sup>1</sup>  
 Europe PMC C<sup>1</sup>  
 EMBO J 24:1634-1643 (2005) C<sup>1</sup>

 Cited in  Mapped to

De todas las posibles variantes recopilamos las que mencionaban inhibición de la interacción con la proteína spike. Encontramos que hay algunas variantes que ligeramente inhiben, otras que lo hacen fuertemente y una sola que indica inhibir sin indicar la fuerza de la inhibición. Todas las variantes de este tipo están soportadas por un mismo artículo de Li et al., (2005). Este se llama "Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2". Sin embargo, ninguna de estas variantes hace referencia al SARS-CoV-2 en particular si no al SARS-CoV.

Hay otras variantes que si mencionan puntualmente al SARS-CoV-2, sin embargo, estas no inhiben y en cambio pueden aumentar ligeramente la interacción. Estas variables todas están reportadas por otro artículo: "Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2" (Chan et al., 2020).

► Mutagenesis	386	Increases slightly the interaction with RBD domain of SARS-CoV-2 spike protein. In sACE2.v2.2; increases interaction with of SARS-CoV-2 spike protein; when associated with Y-27 and Y-330.	1 Publication
► Mutagenesis	389	Slightly inhibits interaction with SARS-CoV spike glycoprotein.	1 Publication
► Mutagenesis	389	Increases very slightly the interaction with RBD domain of SARS-CoV-2 spike protein.	1 Publication
► Mutagenesis	393	Slightly inhibits interaction with SARS-CoV spike glycoprotein.	1 Publication
► Mutagenesis	393	Increases very slightly the interaction with RBD domain of SARS-CoV-2 spike protein.	1 Publication

2. Dada la siguiente lista de identificadores Entrez ID utilice la herramienta de mapeo de UniProt: 258469 5008 13986 7078 10236 8906 58638

- a. Identifique los Entrez ID que no fueron mapeados a proteínas en UniProt. ¿Cuántos son?

Para realizar la búsqueda de estos IDs se ha usado la herramienta de ID Mapping de Uniprot provista en el siguiente [link](#). Allí se ingresaron los IDs y se realizó inicialmente la búsqueda con los valores por defecto de UniProt, en este caso por defecto la base de datos escogida es UniProtKB AC/ID y UniProtKB, no obstante, dentro de esta base de datos no se hallaron resultados que coincidan con lo que se busca, así se cambió la base de datos 1 a la de Genome Annotation Databases /GeneID, esto debido a que se identifica que estos IDs hacen referencias a IDs provistos en NCBI, específicamente Gene ID, por lo cual hacer la búsqueda en esta base de datos nos dan resultados adecuados, dejando la base de datos #2 como UniProtKB. De acuerdo con lo mencionado la query de búsqueda luce de esta manera:

## Retrieve/ID mapping

Enter one of more IDs (100,000 max). You may also [load from a text file](#). Separate IDs by whitespace (space, tab, newline) or commas. This service can also be used [programmatically](#). Alternatively, the [underlying data](#) can be downloaded.

258469 5008 13986 7078 10236 8906 58638

📘 Your input contains 7 unique IDs

From database

GeneID

To database

UniProtKB

Name your ID Mapping job

258469 +6 GeneID → UniProtKB

Finalmente, como resultado de esta búsqueda se encontró que en total 2 IDs no fueron mapeados: 58638 y 13986, y de los que si fueron mapeados se obtuvieron 13 resultados en total, el acceso a este trabajo puede ser obtenido en este [link](#).

- b. Genere un archivo (.csv o .tab) que contenga una tabla de todas las proteínas mapeadas con las siguientes columnas:
  - i. UniProt ID, UniProt name, Protein name, Gene name, Organism, Entrez ID.

Para obtener los datos en un formato de .csv o .tab se identificó que UniProt no ofrece la posibilidad de descargar los datos en estos formatos, es por ello por lo que se ha desarrollado una herramienta de acceso programático que accede al API de UniProt con la URL generada de una búsqueda con las tablas requeridas:

URL de request: [https://rest.uniprot.org/idmapping/uniprotkb/results/stream/560328d6804c8449f2e60266cf18bf468be93f39?compressed=true&fields=accession%2Creviewed%2Cid%2Cprotein\\_name%2Cgene\\_names%2Corganism\\_name%2Clength%2Cxref\\_geneid&format=tsv](https://rest.uniprot.org/idmapping/uniprotkb/results/stream/560328d6804c8449f2e60266cf18bf468be93f39?compressed=true&fields=accession%2Creviewed%2Cid%2Cprotein_name%2Cgene_names%2Corganism_name%2Clength%2Cxref_geneid&format=tsv)

Tener en cuenta que este acceso programático se está haciendo con el valor de parámetro de compressed = True, por lo que los datos están comprimidos y requieren ser descomprimidos al obtenerlos para poder visualizarlos correctamente y no en lenguaje binario.

Código de acceso programático con archivo resultante mapping.csv visualizado por medio de Pandas:

```
request_url = "https://rest.uniprot.org/idmapping/uniprotkb/results/stream/560328d6804c8449f2e60266cf18bf468be93f39?compressed=true&fields=accession%2Creviewed%2Cid%2Cprotein_name%2Cgene_names%2Corganism_name%2Clength%2Cxref_geneid&format=tsv"

response = requests.get(request_url)

# the results are compressed to make it easier to retrieve data of the API, so we need to decompress it
data_filename = "mapping"

decompressed_file = gzip.decompress(response.content)

# save the data to a file of format csv
with open(data_filename + ".csv", "wb") as file:
    file.write(decompressed_file)

# visualize info of the data

data = pandas.read_csv(data_filename + ".csv", sep="\t")

data
```

[15] ✓ 2.2s Open 'data' in Data Wrangler

	From	Entry	Reviewed	Entry Name	Protein names	Gene Names	Organism	Length	GeneID
0	258469	Q0VEL5	unreviewed	Q0VEL5_MOUSE	Olfactory receptor	Or2h2 Olfr90	Mus musculus (Mouse)	310	258469;
1	5008	P13725	reviewed	ONCM_HUMAN	Oncostatin-M (OSM)	OSM	Homo sapiens (Human)	252	5008;
2	5008	B5MCX1	unreviewed	B5MCX1_HUMAN	Oncostatin M	OSM	Homo sapiens (Human)	231	5008;
3	7078	P35625	reviewed	TIMP3_HUMAN	Metalloproteinase inhibitor 3 (Protein MIG-5) ...	TIMP3	Homo sapiens (Human)	211	7078;
4	10236	O43390	reviewed	HNRNPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R (hnR...	HNRNPR HNRPR	Homo sapiens (Human)	633	10236;
5	10236	B4DMB1	unreviewed	B4DMB1_HUMAN	cDNA FLJ53358, highly similar to Heterogeneous...	NaN	Homo sapiens (Human)	595	10236;
6	10236	B4DMD1	unreviewed	B4DMD1_HUMAN	cDNA FLJ53360, highly similar to Heterogeneous...	NaN	Homo sapiens (Human)	473	10236;
7	10236	B4DT28	unreviewed	B4DT28_HUMAN	Heterogeneous nuclear ribonucleoprotein R (cDN...	HNRNPR HNRPR hCG_38907	Homo sapiens (Human)	494	10236;
8	10236	Q0VGD6	unreviewed	Q0VGD6_HUMAN	HNRNPR protein	HNRNPR	Homo sapiens (Human)	607	10236;
9	10236	Q6MZS5	unreviewed	Q6MZS5_HUMAN	Uncharacterized protein DKFZp686A13234	DKFZp686A13234	Homo sapiens (Human)	613	10236;
10	8906	O75843	reviewed	AP1G2_HUMAN	AP-1 complex subunit gamma-like 2 (Gamma2-adap...	AP1G2	Homo sapiens (Human)	785	8906;
11	8906	Q86U03	unreviewed	Q86U03_HUMAN	Full-length cDNA clone CS0DI003YF22 of Placent...	NaN	Homo sapiens (Human)	404	8906;
12	8906	Q86V28	unreviewed	Q86V28_HUMAN	AP-1 complex subunit gamma	NaN	Homo sapiens (Human)	640	8906;

No obstante, se requiere de un refinamiento de las columnas para adecuar los nombres de estas las columnas que se piden, por lo cual se obtiene la siguiente tabla luego del refinamiento de acuerdo con las columnas de datos que se solicitan:

	EntrezID	UniProtID	Reviewed	UniProtName	Protein names	Gene Names	Organism	Length
0	258469	Q0VEL5	unreviewed	Q0VEL5_MOUSE	Olfactory receptor	Or2h2 Olf90	Mus musculus (Mouse)	310
1	5008	P13725	reviewed	ONCM_HUMAN	Oncostatin-M (OSM)	OSM	Homo sapiens (Human)	252
2	5008	B5MCX1	unreviewed	B5MCX1_HUMAN	Oncostatin M	OSM	Homo sapiens (Human)	231
3	7078	P35625	reviewed	TIMP3_HUMAN	Metalloproteinase inhibitor 3 (Protein MIG-5) ...	TIMP3	Homo sapiens (Human)	211
4	10236	O43390	reviewed	HNRPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R (hnR...	HNRNPR HNRPR	Homo sapiens (Human)	633
5	10236	B4DMB1	unreviewed	B4DMB1_HUMAN	cDNA FLJ53358, highly similar to Heterogeneous...	NaN	Homo sapiens (Human)	595
6	10236	B4DMD1	unreviewed	B4DMD1_HUMAN	cDNA FLJ53360, highly similar to Heterogeneous...	NaN	Homo sapiens (Human)	473
7	10236	B4DT28	unreviewed	B4DT28_HUMAN	Heterogeneous nuclear ribonucleoprotein R (cDN...	HNRNPR HNRPR hCG_38907	Homo sapiens (Human)	494
8	10236	Q0VGD6	unreviewed	Q0VGD6_HUMAN	HNRPR protein	HNRPR	Homo sapiens (Human)	607
9	10236	Q6MZS5	unreviewed	Q6MZS5_HUMAN	Uncharacterized protein DKFZp686A13234	DKFZp686A13234	Homo sapiens (Human)	613
10	8906	O75843	reviewed	AP1G2_HUMAN	AP-1 complex subunit gamma-like 2 (Gamma2-adap...	AP1G2	Homo sapiens (Human)	785
11	8906	Q86U03	unreviewed	Q86U03_HUMAN	Full-length cDNA clone CSOD1003YF22 of Placent...	NaN	Homo sapiens (Human)	404
12	8906	Q86V28	unreviewed	Q86V28_HUMAN	AP-1 complex subunit gamma	NaN	Homo sapiens (Human)	640

- c. Genere un archivo (.csv o .tab) que contenga una tabla sólo de las proteínas curadas con la siguiente información:
  - i. UniProt ID, UniProt name, Protein name, Gene name, Organism, Entrez ID, Sequence, Length, Alternative products, Anotaciones en GO-BP, GO-CC, GO-MF.

Para descargar los datos que corresponden a proteínas curadas se ha seleccionado la sección de Reviewed dentro de la página de búsqueda:

Status

Reviewed (Swiss-Prot) (4) X

Popular organisms

Human (4)

Taxonomy

Filter by taxonomy

Proteins with

3D structure (4)

Alternative products (isoforms) (1)

Alternative splicing (1)

Beta strand (3)

Binary interaction (4)

More items

ID mapping 4 results found for GeneID → UniProtKB

OverviewInput ParametersAPI Request

ToolsDownload (4)AddView: CardsTableCustomize columnsResubmitShare

5 IDs were mapped to 4 results  
2 ID were not mapped:  
[Show IDs](#)

From	Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> 10236	O43390	HNRPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R[...]	HNRNPR, HNRPR	Homo sapiens (Human)	633 AA
<input type="checkbox"/> 8906	O75843	AP1G2_HUMAN	AP-1 complex subunit gamma-like 2[...]	AP1G2	Homo sapiens (Human)	785 AA
<input type="checkbox"/> 5008	P13725	ONCM_HUMAN	Oncostatin-M[...]	OSM	Homo sapiens (Human)	252 AA
<input type="checkbox"/> 7078	P35625	TIMP3_HUMAN	Metalloproteinase inhibitor 3[...]	TIMP3	Homo sapiens (Human)	211 AA

Aquí dentro, se extrajo la URL de API para realizar la búsqueda con las columnas seleccionadas, por lo cual se obtuvo el siguiente resultado luego de filtrar:



EntrezID	UniProtID	UniProtName	Protein names	Gene Names	Organism	Length	Sequence	Alternative products (isoforms)	Gene Ontology (biological process)	Gene Ontology (cellular component)	Gene Ontology (molecular function)
10236	O43390	HNRPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R (hnR...	HNRNPR HNRPR	Homo sapiens (Human)	633	MANQVNGNAVQLKEEPMDSVTHTETHTYKTLIEAGLPQKVAERL...	ALTERNATIVE PRODUCTS: Event=Alternative splic...	mRNA processing [GO:0006397]; mRNA splicing, v...	spliceosome [GO:0071013]; end...	mRNA binding [GO:0003729]; RNA binding [GO:000...
8906	O75843	AP1G2_HUMAN	AP-1 complex subunit gamma-like 2 (Gamma2-adap...	AP1G2	Homo sapiens (Human)	785	MVVPSLKLQDLIEBIRGAQTQAQEREVIQKECAHIRASFRDGDVPH...	ALTERNATIVE PRODUCTS:	Golgi to vacuole transport [GO:0006896]; intra...	AP-1 adaptor complex [GO:0030121]; endosome me...	clathrin adaptor activity [GO:0035615]
5008	P13725	ONCM_HUMAN	Oncostatin-M (OSM)	OSM	Homo sapiens (Human)	252	MGVLITQRTLLSLVALLFSPMASMAAIGSCSKEYRVLLGQLQKQT...	ALTERNATIVE PRODUCTS:	immune response [GO:0006955]; negative regulat...	extracellular region [GO:0005576]; extracellul...	cytokine activity [GO:0005125]; growth factor ...
7078	P35625	TIMP3_HUMAN	Metalloproteinase inhibitor 3 (Protein MIG-5) ...	TIMP3	Homo sapiens (Human)	211	MTPWLGLVLGLGWSLGDWGAEEACTSPSPHQDAFCNSDIVIRAKV...	ALTERNATIVE PRODUCTS:	cellular response to organic substance [GO:007...	basement membrane [GO:0005604]; collagen-conta...	metal ion binding [GO:0046872]; metalloendopep...

El acceso a las herramientas programáticas usadas en el código de cuaderno Jupyter y a los archivos resultantes .csv puede ser encontrado en nuestro repositorio GitHub en este link.

3. Análisis a partir de la secuencia de AA provista,
  - a. Realice un BLAST para determinar las proteínas similares a la secuencia objetivo.

Las proteínas similares son las denominadas: Glicoproteína de espiga, glicoproteína S, E2, proteína peplómero, escindida en proteína de espiga S1, proteína de espiga S2, proteína de espiga S2'

Tools Download Add Customize columns Resubmit						
Entry	Entry Name	Protein Names	Gene Names	Organism	Length	
<input type="checkbox"/> P0DTC2	SPIKE_SARS2	Spike glycoprotein[...]	S.2	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	1,273 AA	98.9% 6477
<input type="checkbox"/> USWLK5	USWLK5_SARS	Spike glycoprotein[...]	S	Bat SARS-like coronavirus RsSHC014	1,256 AA	76.7% 6050
<input type="checkbox"/> P59594	SPIKE_SARS	Spike glycoprotein[...]	S.2	Severe acute respiratory syndrome coronavirus (SARS-CoV)	1,255 AA	75.4% 6000
<input type="checkbox"/> Q3LZX1	SPIKE_BCH3	Spike glycoprotein[...]	S.2	Bat coronavirus HKU3 (BtCoV) (SARS-like coronavirus HKU3)	1,342 AA	75.2% 6899
<input type="checkbox"/> Q3ISJ5	SPIKE_BCRP3	Spike glycoprotein[...]	S.2	Bat coronavirus Rp3/2004 (BtCoV/Rp3/2004) (SARS-like coronavirus Rp3)	1,241 AA	74.7% 6850

- b. Identifique a qué proteína es más probable que pertenezca la secuencia provista.

La proteína con la que tiene mayor identidad es la denominada P0DTC2 con una identidad del 98.9%

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	
<input type="checkbox"/> P0DTC2	SPIKE_SARS2	Spike glycoprotein[...]	S.2	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	1,273 AA	98.9% 6477

- c. Determine si las diferencias entre la secuencia provista y la proteína más similar ocurren en features de interés

Al resaltar el BLAST de las dos secuencias por dominio se observa una zona donde se cambian 3 aminoácidos. Un aminoácido cerca a los 296 y dos aminoácidos seguidos cerca de los 518. Cuando las secuencias de proteínas tienen diferencias en la zona de un dominio, esto puede tener implicaciones importantes, ya que los dominios proteicos son las regiones estructurales y funcionales clave de una proteína.



PODTC2:Domain		
Query	LEPLVDLP IGINITRFQTLHALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPL	296
Match:PODTC2	LEPLVDLP IGINITRFQTLALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPL	296
Query	VGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLYRYGFGPTNGVGYQPYRVVLSFELL	518
Match:PODTC2	VGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFGPTNGVGYQPYRVVLSFELL	518
PODTC2:Domain		

También, al resaltar el BLAST por mutagénesis se observa una zona cerca a los 720 aminoácidos donde hay una delección de 4 aminoácidos. Al revisar las anotaciones de esta zona se observa que está asociado a: la perdida completa de la O-glicosilación por parte del huésped, pérdida completa de la escisión por la furina del huésped, mejora parcial de la escisión por la furina del huésped y aproximadamente un 50% más de eficiencia de entrada en células Vero E6, mientras que un 30% menos de eficiencia de entrada en células BHK que expresan hACE2.

Query	CLIGAEHVNSYECDIPIGAGICASYQTQTNSTNSRVSASQSIAYTMSLGAENSVAYSNNIAIPTNFTI	720
Match:PODTC2	CLIGAEHVNSYECDIPIGAGICASYQTQTNSTNSRVSASQSIAYTMSLGAENSVAYSNNIAIPTNFTI	720
PODTC2:Mutagenesis		

Por último, al resaltar el BLAST por topological domain está asociado a la parte extracelular y se observa que va desde el aminoácido 12 hasta el 1211. Donde están incluidas las diferencias ya mencionadas y otras adicionales como la delección de 4 aminoácidos cerca de los 897 y el cambio de dos aminoácidos por los 1104 y 1173 aminoácidos. Estos cambios pueden tener cambios significativos en la proteína como: la interacción con otras moléculas o señales y cambiar la orientación de la proteína dentro de la membrana para su correcta localización.

Query	AD---KQYGDCLGDIAARDLICAKFNGLTVLPPLLTDEMI AQYTSALLAGTITSGWTFGAGAALQIP	897
Match:PODTC2	ADAGFIKQYGDCLGDIAARDLICAKFNGLTVLPPLLTDEMI AQYTSALLAGTITSGWTFGAGAALQIP	897
PODTC2:Topological domain		
Query	QSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNETTAPAI CHDGAHFPRREGVVFVSNGTHWFV	1104
Match:PODTC2	QSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNETTAPAI CHDGAHFPRREGVVFVSNGTHWFV	1104
PODTC2:Topological domain		
Query	TQRNFYEPQIITTDNTFVRGNCDDVIGIVNNTVYDPLQPELDSFKEELDQYFKNHTSPDVLGDISGIN	1173
Match:PODTC2	TQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDQYFKNHTSPDVLGDISGIN	1173
PODTC2:Topological domain		

MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLP  
FFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTK  
QSLIVNNATNVVIKVFCEFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEY  
SQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYKHTPINLVRDLPQGFSALEPLVD  
LPIGINITRFQTLHALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTI  
TDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVF  
NATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADS  
FVIRGDEVQRQIAPGQTGKIADYNYKLDDFTGCVIAWNSNLDKSVGGNYNYLYRLF  
RKS NLKPFERDISTEIQAGSTPCNGVEGFNCYFPLYRYGFGPTNGVGYQPYRVVVL  
SFELLHAPATVCGPKKSTNLVKNKCLNFNFNGLTGTGVLTESNKKFLPFQFGRDIA  
DTTDAVRDPQTLILDITPCSGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHAD  
QLTPTWRVYSTGSNVFQTRAGCLIGAEHVNSYECDIPIGAGICASYQTQTNRSVA

SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGD  
STECNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNF  
SQILPDPSKPSKRSFIEDLLFNKVTLADKQYGDCLGDIAARDLICAQKFNGLTVLPP  
LLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQ  
KLIANQFNSAIGKIQDLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVL  
NDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL  
GQSKRVDFCGKFYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAHFPR  
EGVFVSNGTHWFVTQRNFYEPQIITDNTFVRGNCDVVIGIVNNTVYDPLQPELDSF  
KEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK  
YEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCCLKGCCS

## Referencias

Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S. K., Huang, I. C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., & Farzan, M. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *The EMBO journal*, 24(8), 1634–1643. <https://doi.org/10.1038/sj.emboj.7600640>

Chan, K. K., Dorosky, D., Sharma, P., Abbasi, S. A., Dye, J. M., Kranz, D. M., Herbert, A. S., & Procko, E. (2020). Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science (New York, N.Y.)*, 369(6508), 1261–1265. <https://doi.org/10.1126/science.abc0870>