

RESEARCH

Reposicionamiento de fármacos en COVID-19

Laura Núñez Jiménez^{*}, Inmaculada Jiménez Palomino and Talal Awija

^{*}Correspondence:

lauranunezj99@uma.es

ETSI Informática, Universidad de
Málaga, Málaga, España

Full list of author information is
available at the end of the article

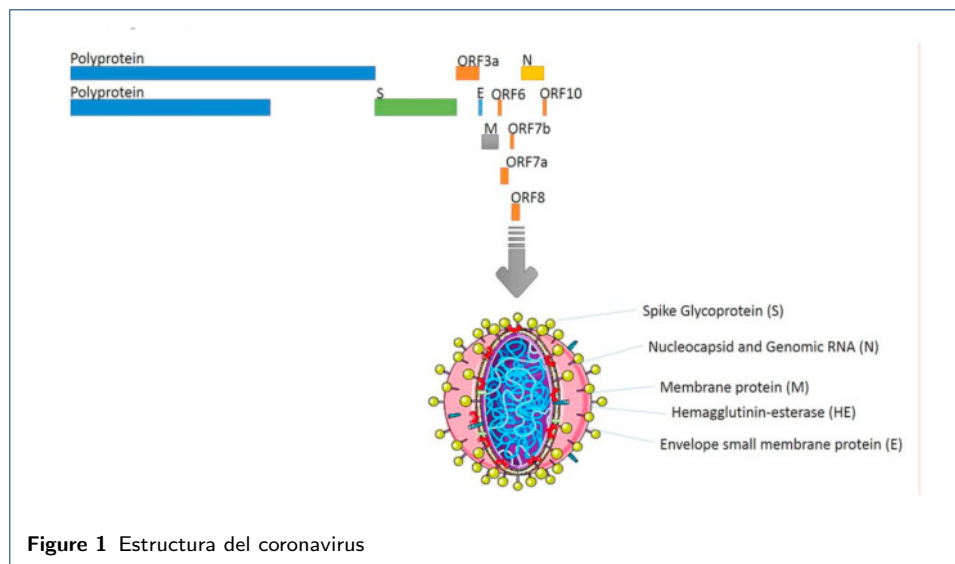
Abstract

A continuación, el informe trata sobre los fármacos que tienen efecto sobre las proteínas humanas que interaccionan con las 27 proteínas del Covid-19. Este caso surge debido a la pandemia vivida en nuestros días producida por el virus SARS-CoV-2, el cual está causando millones de contagios alrededor del mundo y miles de muertes. Una de las principales causas por lo que no se desarrolla de manera efectiva la creación de fármacos es por el poco conocimiento que se tiene sobre el virus. Por ello, se requiere de la búsqueda de fármacos ya existentes que puedan interaccionar con las conexiones entre proteína humana y el SARS-CoV-2, y así afectar a la reacción del virus. Nos basamos en implementar la reutilización de medicamentos basadas en la red que surge de esta relación, lo que permite clasificar todos los medicamentos aprobados en función de su posible eficacia. Usando las bases de datos STRING obtendremos dicha red, que nos ayudará a obtener las 332 proteínas humanas, las cuales, usaremos posteriormente en la base de datos de ChEMBL. Con dicha base de datos obtendremos en él, el experimento de fármacos potenciales, así como información sobre cada uno de ellos.

Keywords: Covid-19; SARS-CoV-2; ChEMBL; Fármacos experimentales

1 Introducción

En nuestros días estamos viviendo una pandemia que empezó en 2019 y sigue su crecimiento hasta día de hoy. El virus que originó este caos tanto económico como de salud pública mundial es el SARS-Cov-2, hallado en China a finales del año 2019. El SARS-Cov-2 tiene una alta tasa de contagios, lo que exige una actuación rápida para poder combatirlo. El virus afecta en mayor medida al sistema respiratorio, generando en cada persona diferentes grados de síntomas o en algunos casos hasta la muerte. Aunque, también se ha encontrado que el virus afecta a diferentes tejidos del cuerpo. Debido al gran impacto que ha generado en el mundo en todos los ámbitos, se necesita de manera urgente encontrar vacunas o medicamentos.



Para la reutilización de medicamentos o de la creación de nuevos medicamentos, se necesita saber como actúa el SARS-CoV-2 durante la infección y los cambios moleculares en las proteínas humanas. Hoy en día, no hay medicamentos antivirales contra el coronavirus. Dicho anteriormente, el desconocimiento de detalles de la molécula del SARS-Cov-2, impide que se obtenga una evaluación precisa para la creación de fármacos. Esto supone un largo tiempo de investigación.

Después de ver unos cuantos artículos sobre como resuelven este problema de la reutilización del fármaco, concluimos en que la mejor estrategia es la basada en la elección de fármacos a través de la creación de una red interactoma humano perturbado por el SARS-Cov-2. A partir de ella, se verá los tejidos afectados por el virus y se priorizará los medicamentos existentes en función de las proteínas humanas y dichas interacciones. Para interrumpir el interactoma del SARS-CoV-2, buscamos ligandos de proteínas humanas que interactúen con proteínas virales.

2 Materiales y métodos

En este apartado se explicará los diferentes métodos que se han seguido a lo largo de un proceso para la reutilización de fármacos.

El inicio de nuestra investigación se ha basado leyendo un número de artículos para ponernos en situación del virus SARS-CoV-2. Encontramos en unos de los paper [1], que el virus forma 29 proteínas, pero que tan solo han sido clonadas y expresadas 26 proteínas de las 29. Estas 26, las señaladas en la tabla 1, son las usada en la investigación y con las que será generado el interactoma proteínas humanas-SARS-Cov-2.

2.1 STRING, Protein-Protein Interaction Networks

Visto anteriormente, la principal idea es obtener que relaciones tienen las proteínas del virus con nuestro cuerpo. De esta manera, sabremos con que proteínas humanas interactúan y así encontrar fármacos que afecten directamente a ellas. El pensamiento es que estos fármacos eviten en cierta medida, la clonación y la infección del virus de manera indirecta.

Para encontrar las proteínas secundarias, llamadas de esta forma en los artículos leídos, se hará uso de la base de datos STRING. STRING se usa para obtener las relaciones de las proteínas del SARS-Cov-2, con todas las humanas. Las proteínas humanas mostradas en STRING son 332 que serán clave en la investigación de los fármacos.

2.2 redSring.R

Una vez hecho la búsqueda de las 332 proteínas humanas, las guardamos en un fichero de entrada. Posteriormente este archivo, creará un mapeo de estas mismas proteínas, pero queriendo reducir el número de ellas.

El archivo ejecutará un mapeo sobre las 332 proteínas humanas encontradas en STRING. Utilizará el fichero donde vienen las proteínas y creará un dataframe donde vendrán el nombre de las proteínas y además se añadirá una columna con STRINGID de cada una de ellas. Después, la intención es obtener una subred de las interacciones entre esas 332, para generar una red secundaria de las interacciones más fuertes. Para ello hemos utilizado un combined score mayor a 900.

Este proceso se utiliza, porque nos ayudará a tener unos resultados más fiables a la hora de encontrar fármacos. Debido a que encontraremos fármacos que afecten a relaciones muy estrechas.

2.3 convertirUniprot.R

Este fichero es auxiliar. Nos va a ayudar a convertir los Gene Name de las proteínas obtenidas en el archivo redString.R, en su nombre de UniProt. Por lo que, devolverá un fichero .txt con los respectivos nombres.

Se realiza esta conversión, para tener los datos preparados en la búsqueda de fármacos en ChEMBL.

2.4 Bases de datos ChEMBL y fichero consultarChembl.py

2.4.1 Bases de datos ChEMBL y consultas en MySQL

En el informe de la explicación de la práctica, nos dice que: Las proteínas humanas con las que interactúan las proteínas virales de SARSCoV2 pueden ser targets farmacológicos de interés, por lo que en este proyecto se pretende encontrar fármacos de la base de datos ChEMBL que tengan como diana las proteínas o módulos del interactoma funcional SARS-humano. Por lo tanto, todo lo que se ha hecho hasta ahora es facilitar el trabajo para que una vez hayamos obtenidos esos denominados targets, se busquen en esta bases de datos.

ChEMBL tiene la posibilidad de ser descargada en diferentes formatos para ser trabajada desde Oracle, MySQL, XML, etc. La base de datos ChEMBL ha sido descargada en formato MySQL para realización del informe. Mediante un esquema de las tablas proporcionadas y una guía donde explica cada tabla y cada elemento, hemos podido hacer la consulta sobre los fármacos que queremos. Hemos realizado una ruta entre las tablas que necesitamos acordes con los datos preparados.

La consulta en cuestión es:

```

SELECT DISTINCT compound_name, max_phase, ACTION_T.action_type, M.pref_name, CS.accession

FROM compound_records C, assays A, target_dictionary T, activities ACT, molecule_dictionary M, drug_mechanism D, action_type
ACTION_T, component_sequences CS, target_components TC

WHERE C.record_id = ACT.record_id AND ACT.assay_id = A.assay_id AND A.tid = T.tid AND pchembl_value >=6 AND M.molregno =
C.molregno AND D.molregno = M.molregno AND D.action_type = ACTION_T.action_type AND max_phase = "+ sys.argv[2]+" AND
CS.component_id = TC.component_id AND TC.tid = T.tid AND CS.accession= " + sys.argv[2] + " + sys.argv[3] + "

```

Figure 2 Consulta MySQL

Tal y como podemos ver en esta consulta, accedemos a varias tablas de la base de datos. Para explicar esto mejor, vamos a mostrar el esquema de la base de datos de ChEMBL, pero únicamente las tablas que hemos utilizado:

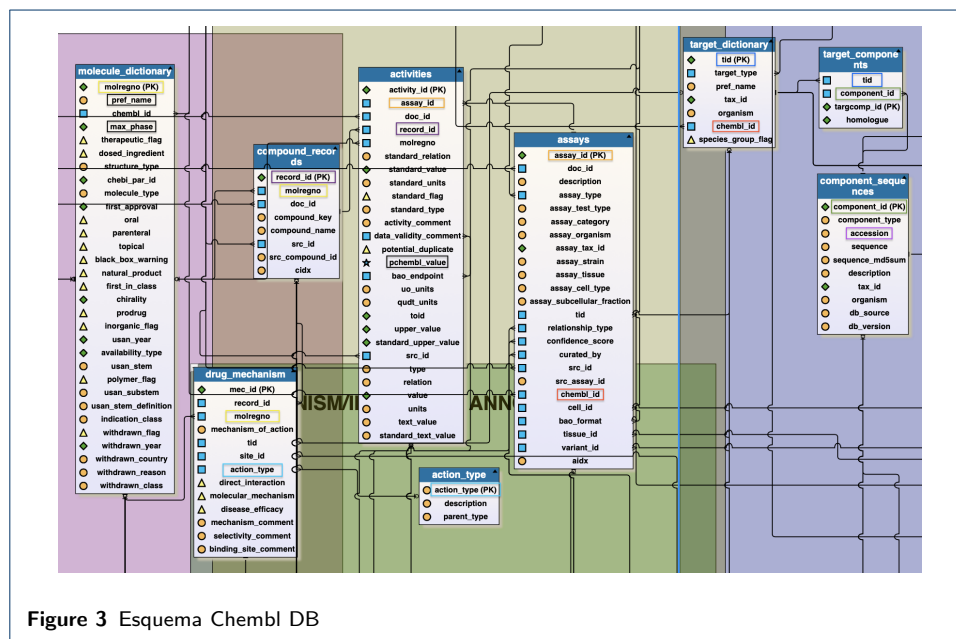


Figure 3 Esquema ChEMBL DB

Accedemos a este esquema por el atributo **accession** de la tabla **component sequences**, en las que introducimos el UniprotID de las 178 proteínas humanas que hemos extraído anteriormente.

Desde ahí, recorreremos un camino, marcado en la figura con recuadros en los atributos de cada tabla, en el que pasamos por la tabla **target components** a través del component id de la proteínas, cogemos el tid y llegamos a la tabla **target dictionary**, de donde sacaremos el ChEMBL ID de la proteína para llegar a la tabla **assays**.

Con el assay id accedemos a la tabla **activities**, que además de formar parte de la ruta desde la proteína hasta sus fármacos asociados, contiene el valor de P ChEMBL, que vamos a utilizar para filtrar los fármacos: con un P ChEMBL value mayor a 6 nos aseguramos que los compuestos estén estrechamente relacionados con las proteínas. En esta misma tabla, nos encontramos el record id, que nos servirá para enlazar con la tabla **compound records**, nuestro objetivo.

Una vez llegados aquí, ya tenemos la lista de fármacos asociados, pero hemos indagado más, recurriendo a las tablas **molecule dictionary**, **drug mechanism** y **action type** para incluir en nuestra consulta, además, el tipo de acción que ejerce

el fármaco, la fase de ensayo en la que se encuentra (que hemos dividido en fármacos experimentales, fase 3, y fármacos aprobados, fase 4) y el nombre común que tienen los compuestos, ya que el científico resultaba muy poco evidente.

2.4.2 Código en python. *consultarChembl.py*

Debido al gran número de proteínas obtenidas, sería muy costoso y lento ir buscando en la consulta de MySQL una a una. Por lo que, se ha creado un programa Python, que lee el fichero donde se encuentran las 178 proteínas con sus respectivos nombres y UniProt ID.

El programa se conecta a las bases de datos descargada en el ordenador y procede a hacer un bucle leyendo cada línea del fichero. De esta manera, va leyendo cada UniProt ID y la va introduciendo en la consulta para ir obteniendo nuestro resultado final que sería los fármacos que estábamos buscando al inicio de la investigación.

Se ha decidido que los fármacos buscados sean del tipo 3 y 4. Donde el tipo 3 son fármacos experimentales y de tipo 4 fármacos aprobados. Esta decisión es debida a que así tendremos una certeza adecuada de funcionalidad del fármaco hallado.

2.5 analisisResultados.R

La ejecución de este fichero sirve para obtener graficas sobre los fármacos encontrados y poder evaluar los resultados obtenidos. El archivo que recoge el fichero es el obtenido en el programa de Python.

2.6 launch.sh

Es un flujo del trabajo completo, el fichero que hay que ejecutar para que se lleve a cabo todo el proceso. Es el fichero principal, un flujo bash

3 Resultados

Tras ejecutar todos los pasos, todos los resultados obtenidos se irán guardando en la carpeta results. Como se han seguido una serie de pasos durante la investigación, en este apartado se irán mostrando los resultados de cada sección.

Antes de hacer la investigación, leímos varios artículos para encontrar el número de proteínas del SARS-CoV-2. Las proteínas encontradas fueron las mostradas en la tabla.

Table 1 Proteínas del SARS-CoV-2 usadas en la red

Proteínas del covid			
"NSP6"	"NSP1"	"NSP2"	"NSP5"
"NSP7"	"NSP5"	"NSP8"	"NSP9"
"NSP11"	"NSP12"	"NSP13"	"NSP14"
"NSP15"	"S"	"E"	"N"
"M"	"ORF3a"	"ORF3b"	"ORF8"
"ORF6"	"ORF7a"	"ORF9b"	"ORF9c"
"ORF10"			

En los artículos venían que había 29 proteínas, pero que solo se utilizaron 26 de ellas. Por eso en los resultados, sólo hay 26. El fichero que contienen las proteínas las introducimos en STRING para encontrar el interactoma y tuvimos de resultado la siguiente red.

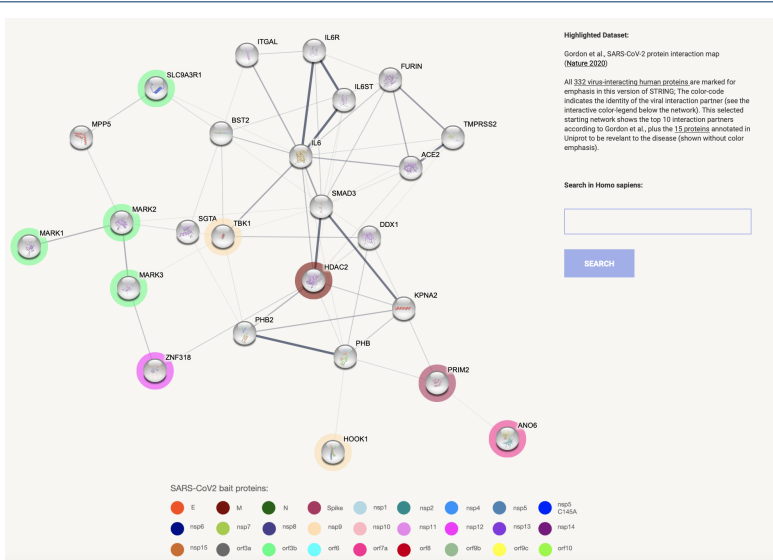


Figure 4 Red proteínas humanas-proteínas SARS-CoV-2

En STRING obtuvimos 332 proteínas que interactuaban con las proteínas del SARS-CoV-2. Pero la red se divide en varios módulos, porque cada proteína del SARS-CoV-2 interactúa con un número X de proteínas humanas. Por lo que, descargamos las 332 proteínas humanas para crear posteriormente una red entre ellas y elegir las que más fuerza de interacción tengan.

Ejecutamos el fichero redString y nos da el resultado de 178 proteínas secundarias, que son las que usaremos para la búsqueda de los fármacos.

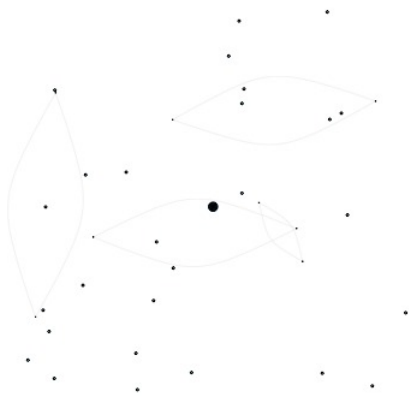


Figure 5 Red de proteínas humanas

Todas esos genenames de proteínas las pasamos a Uniprot ID.

P48556	P14735	Q9UHD2	Q96JC1
Q9NQT4	Q14975	O00469	Q9BVL2
Q9NW13	O00116	P38606	B4DP20
O43818	E9PH64	P55085	A0A024R7H8
P62877	Q7Z3B4	Q9HAV7	A0A140VKB1
A0A024R6Y2	A0A0S2Z5A8	P82675	A6NMQ1
Q8TD19	Q96CW1	Q9NQT5	A0A0S2Z6B4
P55789	Q9Y673	O95613	Q9Y3U8
B2R4A9	A0A158RFU6	Q9BS26	Q8NEJ9
Q96DZ1	Q9Y2I6	P19784	P51148
P06280	A0A024R7I3	Q08117	O43823
Q66GS9	A0A0S2Z5U1	O43292	Q2T9J0
Q92820	Q15056	P49642	Q9BVK2
P11233	Q12907	A0A024RDU9	Q9H845
P61019	O00566	Q13206	Q13438
Q8TEM1	A0A024R006	B2RP65	A4D0Z4
B3KM34	Q9BZQ6	Q10713	A0A024R6G3
P62072	B3KR21	Q13868	A7MCY6
P61026	O94826	B3KPY8	Q5VU43
A0A0S2Z4V6	Q08431	Q9BSF4	Q15904
Q14745	P62330	P35658	Q5JQI0
A0A024R2W3	F8VSL7	P35555	A0A1U9X7J2
O76061	Q15370	P12109	Q8TEQ8
A0A024R712	P17612	Q9NXX8	Q6UX04
A0A024R9L0	B7ZAJ4	A0A024RB33	Q8N4C6
P82663	Q8NBL9	O94973	Q8IWF2
Q9BQ95	B9EG74	Q86YT6	P35241
B4DYP7	Q14181	Q9BSC4	Q7KZ17
Q8NBJS	O95391	O60573	Q6Y7W6
Q7L2J0	P52948	A0A024R845	B3KX12
O76094	A0A140VJP2	Q14789	A0A0A0MT23
Q9Y399	Q5GIA7	Q9NQC3	Q08379
I6L9H2	J3KNF8	P52306	Q4G0J3
Q14157	A6NHH0	P63218	P35556
A0A024R467	P00387	O14578	Q6PKB3
B4DK95	P09132	P05556	Q92769
Q9ULT8	A0A0S2Z366	A0A088QCU6	Q96HW7
Q9BV73	Q04724	A0A0S2Z5J4	G3XAN8
P61011	P78406	Q96DA6	Q9BSJ2
Q96HR9	A0A024R9C1	A0A024R648	A0A024R6C0
Q00765	Q96B26	P62820	A0A0S2Z3L0
A0A024R324	B7Z879	A0A024R7L8	Q04726
A0A024R7Y5	A0A024QZF1	Q9NY61	B3KVK2
		Q9NP72	P61962
		Q92615	P49643
			B3KXHO

Figure 6 Uniprot ID de las proteínas obtenidas en la red

Una vez obtenidos los UnitProt ID, buscamos los fármacos en la base de datos ChEMBL. Tras obtener los datos de los fármacos tanto para la fase de aprobación como para la experimental, se realiza una serie de gráficos. A continuación, vamos a mostrar dos gráficas, una donde serán los tipos de fármacos que se utilizan en ensayos ya aprobados y la otra grafica los tipos de los ensayos experimentales.

Se obtienen alrededor de 43 fármacos de ensayos experimentales y unos 66 de fármacos en los ensayos aprobados.

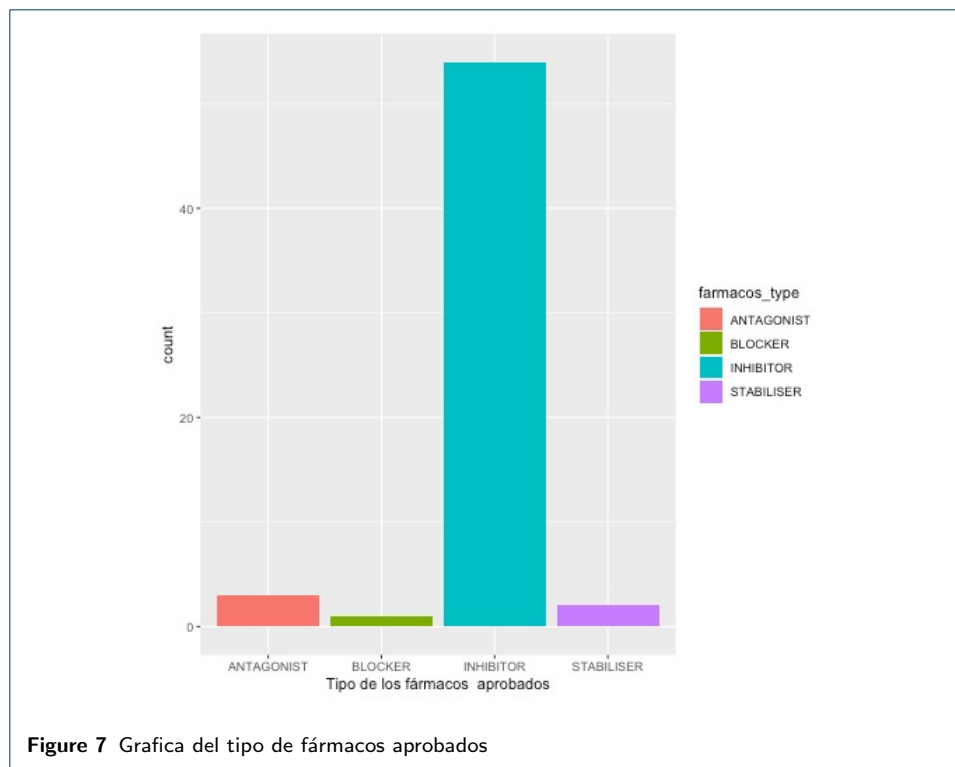


Figure 7 Grafica del tipo de fármacos aprobados

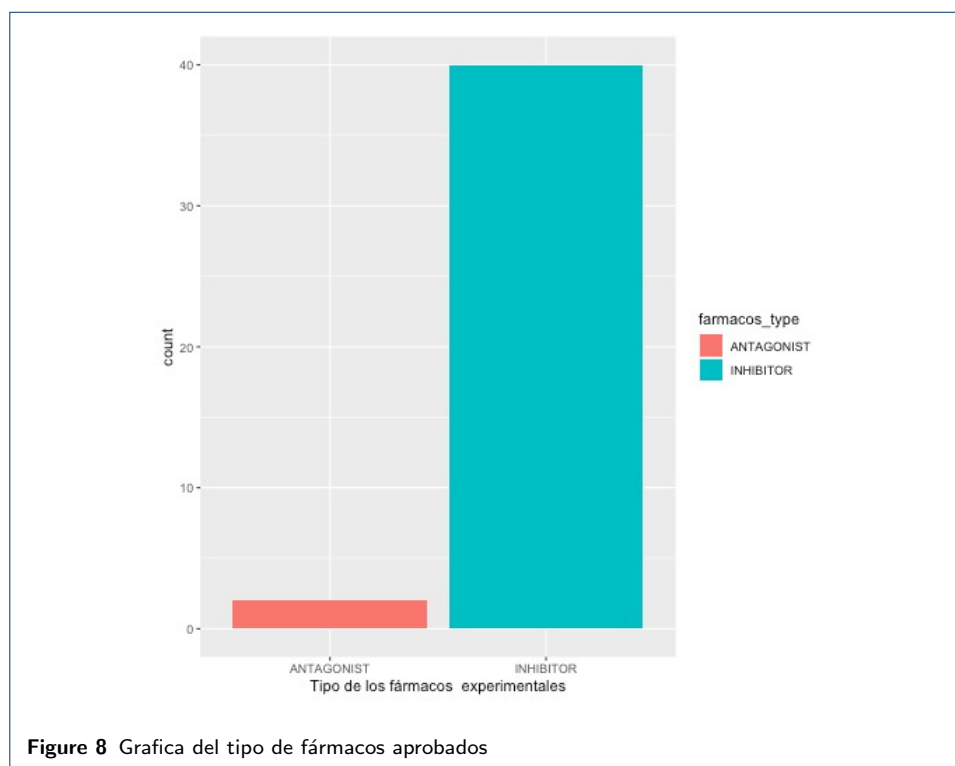


Figure 8 Grafica del tipo de fármacos aprobados

Como observamos en esta primera grafica y en la segunda, la mayor parte de fármacos son de tipo inhibidor.

4 Discusión

Lo más destacable de los resultados es que, el tipo de fármaco que predomina sobre todos son los inhibidores, que disminuyen o cesan la acción o creación de otra sustancia.

Nos podemos imaginar y deduciendo de los artículos leídos [1] [2] [3] que los fármacos pueden bloquear, en cierto modo, que las células del SARS-CoV-2 invadan las células humanas o limitar entre proceso que causa la infección. De esta manera, los síntomas y el tiempo de la enfermedad se podrían ver limitados.

Todo lo desarrollado es en general, permite que se puedan buscar fármacos en otras fases de ensayos y especificar más que realiza cada fármaco. Por lo que, no solo existen los fármacos de nuestra lista.

5 Conclusiones

Finalmente, el objetivo que teníamos al inicio de la investigación de encontrar fármacos que puedan actuar sobre el virus del SARS-CoV-2 usando la base de datos ChEMBL ha sido completada. Al inicio del trabajo creamos el interactoma de las proteínas, pero debido a que hay muchos módulos dentro de ese mismo interactoma, es difícil encontrar un interactoma funcional completo. Otro dato, es que al hacer el mapeo solo se pierden la mitad casi de las proteínas, es decir, tenemos un alto nivel de interacción entre bastantes proteínas secundarias, por lo que será de mayor ayuda en la búsqueda de los fármacos, ya que saldrán resultados más certeros. Aun así, nos sale de resultados un gran número de fármacos ya aprobados. Además, dentro de estos fármacos podemos observar, que se encuentran un mecanismo de acción mayoritario, es el caso de fármacos inhibidoras. Como consecuencia de todo podemos concluir que hay que estudiar qué proteínas del interactoma funcional son las mejores candidatas que desarrollar un fármaco.

Disponibilidad de datos y materiales

https://github.com/talal314/project_template.git

Contribución de los autores

T.A: Uso y descarga de la base de datos ChEMBL, consultas y recorrido de tablas sobre la bases de datos, Connectar a la base de datos desde python para guardar los resultados en una tabla y filtración de datos con R.

I.J.P: Obtención de las proteínas humanas, creación de igraph y mapeo de dichas proteínas, filtración de las interacciones con combined score y transformación del id a uniprot id.

L.N.J: Redacción del informe de la investigación. Consultar y buscar información en artículos y la obtención de las 29 proteínas del SARS-CoV-2.

Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

References

- Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., Tummino, T.A., Hüttenhain, R., Kaake, R.M., Richards, A.L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B.J., Braberg, H., Fabius, J.M., Eckhardt, M., Southeray, M., Bennett, M.J., Cakir, M., McGregor, M.J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Kain, A.M., Miorin, L., Moreno, E., Naing, Z.Z.C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I.T., Melnyk, J.E., Chiorba, J.S., Lou, K., Dai, S.A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C.J.P., Perica, T., Pilla, K.B., Ganesan, S.J., Saltzberg, D.J., Rakesh, R., Liu, X., Rosenthal, S.B., Calviello, L., Venkataraman, S., Liboy-Lugo, J., Lin, Y., Huang, X.P., Liu, Y.F., Wankowicz, S.A., Bohn, M., Safari, M., Ugur, F.S., Koh, C., Savar, N.S., Tran, Q.D., Shengjuler, D., Fletcher, S.J., O'Neal, M.C., Cai, Y., Chang, J.C.J., Broadhurst, D.J., Klippsten, S., Sharp, P.P., Wenzell, N.A., Kuzuoglu-Ozturk, D., Wang, H.Y., Trenker, R., Young, J.M., Caverio, D.A., Hiatt, J., Roth, T.L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R.M., Frankel, A.D., Rosenberg, O.S., Verba, K.A., Agard, D.A., Ott, M., Emerman, M., Jura, N., von Zastrow, M., Verdini, E., Ashworth, A., Schwartz, O., d'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H.S., Fujimori, D.G., Ideker, T., Craik, C.S., Floor, S.N., Fraser, J.S., Gross, J.D., Sali, A., Roth, B.L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., García-Sastre, A., Shokat, K.M., Shoichet,

- B.K., Krogan, N.J.: A sars-cov-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020). doi:10.1038/s41586-020-2286-9
2. Gysi, D.M., Ítalo Do Valle, Zitnik, M., Ameli, A., Gan, X., Varol, O., Ghiassian, S.D., Patten, J., Davey, R., Loscalzo, J., Barabási, A.-L.: Network medicine framework for identifying drug repurposing opportunities for covid-19 (2020)
 3. Hosseini, E.S., Kashani, N.R., Nikzad, H., Azadbakht, J., Bafrani, H.H., Kashani, H.H.: The novel coronavirus disease-2019 (covid-19): Mechanism of action, detection and recent therapeutic strategies. *Virology* **551**, 1–9 (2020). doi:10.1016/j.virol.2020.08.011