

RESEARCH

Reposicionamiento de fármacos en COVID-19

Irene Sánchez Jiménez , Clara Jiménez Valverde^{*}, Patricia Trujillo Rodríguez and Lucía Valverde Martínez

^{*}Correspondence:

clarajimenez@uma.es

ETSI Informática, Universidad de
Málaga, Málaga, España

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

Abstract

En esta investigación, se busca encontrar fármacos ya existentes que puedan combatir la acción del virus SARS_Cov-2. Debido a la alta tasa de contagio que tiene, nos urge encontrar tratamientos eficaces en poco tiempo. Aunque se han probado medicamentos útiles en casos de virus similares, para encontrar un tratamiento adecuado contra el SARS_Cov-2 es esencial cambiar la aproximación reduccionista por una más centrada en el sistema. Por ello, en este trabajo se propone usar las proteínas humanas descubiertas del interactoma funcional SARS-humano, junto a otras relacionadas, para encontrar fármacos ya existentes que afecten la acción del virus. Usando las bases de datos STRING y ChEMBL, obtenemos en el experimento fármacos potenciales, así como información sobre ellos, que se podrá usar en estudios posteriores.

Keywords: COVID-19; fármacos; proteínas; interactoma

1 Introducción

A día de hoy, puede que la enfermedad más mencionada en todo el mundo sea el SARS-CoV-2 (coronavirus 2 del síndrome respiratorio agudo severo). Este virus, que se originó en la ciudad China de Wuhan en diciembre de 2019, se ha extendido a todos los países del mundo rápidamente. Aunque la tasa de mortalidad del 2% no es especialmente alta, el SARS-CoV-2 es altamente contagioso y transmisible de persona en persona con un período de incubación de hasta 24 días [1]. Este brote de SARS-CoV2 ha provocado graves impactos en la salud social y la economía en todos los niveles, convirtiéndolo en una prioridad encontrar tratamientos eficientes para la enfermedad.

Como podemos ver en la revisión bibliográfica [1], ya en los primeros 75 días del brote de COVID-19 se propusieron algunos tratamientos médicos para combatir el SARS-CoV-2. Los más usados fueron antivirales (oseltamivir, ganciclovir y KALETRA), agentes antibióticos (cefalosporina, quinolona y carbapenem) para prevenir infecciones secundarias, y en los casos más graves terapia de oxígeno y terapia de reemplazo renal. El uso de estos tratamientos se debe en su mayoría a la eficacia que han tenido en virus similares como es la influenza.

Para maximizar la eficacia de los tratamientos empleados, es vital diseñar fármacos que ataquen al virus de forma directa. Sin embargo, esto supone dos inconvenientes. El primero es conocer una forma eficaz de combatir la enfermedad y el segundo es el largo proceso que conlleva diseñar un fármaco.

Encontramos numerosos artículos que tratan de desentrañar el interactoma SARS-humano ([2], [3], [4], [5], [6]). Conociendo las proteínas humanas que son afectadas por el virus, podemos encontrar fármacos que perjudiquen la acción de éste en los seres humanos.

Debido a la urgencia de encontrar tratamientos efectivos, en nuestra investigación se propone encontrar fármacos ya existentes que actúen sobre las proteínas del interactoma entre SARS y humanos.

Para hacer esta búsqueda, usaremos la base de datos ChEMBL. ChEMBL es un conjunto de datos 'quimogenómicos' que une información sintética, de bioactividad y genómica de medicamentos. Surgió ante la necesidad de indexar productos genéticos y conectarlos con medicamentos para su posterior análisis y divulgación por parte de farmacéuticas, pequeñas organizaciones y establecimientos académicos.

En nuestro caso, ChEMBL nos permitirá obtener fármacos cuyas dianas sean proteínas humanas del interactoma SARS-Humano.

2 Materiales y métodos

Partimos nuestra investigación buscando el interactoma funcional SARS-humano. Aunque el SARS-CoV-2 es un virus reciente, encontramos en la literatura científica numerosos papers que investigan este tema ([2], [3], [4], [5], [6]). Como podemos ver en [6], el genoma viral del SARS-CoV-2 es capaz de crear 29 proteínas diferentes. En el trabajo [4], Gordon et al. consiguieron aislar 27 de estas proteínas para estudiar aquellas proteínas humanas que interaccionan con las del SARS-CoV-2. Esta investigación encontró interacciones de proteínas humanas con 26 de las proteínas del SARS-CoV-2.

Las proteínas humanas que pertenecen al interactoma funcional SARS-humano pueden ser la clave para frenar al virus, impidiendo que éste infecte a las células, se replique y se expanda. Sin embargo, diseñar fármacos específicos para el SARS-CoV-2 puede llevar mucho tiempo, por lo que vamos a estudiar qué fármacos existen actualmente cuya diana sean estas proteínas humanas.

2.1 STRING, base de datos de interacción de proteínas

Como se habla en [5], cuando usamos fármacos con dianas específicas, podemos encontrar efectos en proteínas relacionadas. Esta idea se enmarca dentro de la corriente holística, donde debemos estudiar el sistema como un conjunto, en vez de centrarnos en las partes.

Siguiendo esta reflexión, buscar proteínas que se relacionen directamente con nuestras proteínas diana, que llamaremos proteínas secundarias, y hallar fármacos que actúen sobre estas proteínas secundarias, puede llevarnos a tener tratamientos que indirectamente combatan la acción del virus.

Para encontrar estas proteínas secundarias, vamos a hacer uso de STRING. STRING es una base de datos de interacciones de proteínas. Introducimos en STRING las proteínas humanas del interactoma SARS-humano que conocemos (un total de 332 proteínas) y hacemos un filtrado de la búsqueda.

Primero, seleccionamos que la puntuación de la interacción (interaction score) sea de al menos 0.700, lo que nos asegura interacciones de alta confianza. A continuación, pedimos que se nos muestren hasta 100 proteínas que interaccionan con cada una de las introducidas. De esta forma, encontramos las proteínas secundarias para nuestro estudio.

2.1.1 Mapeo de códigos

Tras hacer nuestra búsqueda en STRING, tenemos un listado de genes que codifican proteínas del interactoma SARS-Humano (332 genes) y otro listado que corresponde a los genes que codifica a las proteínas secundarias (115 genes).

A continuación, debemos preparar estos datos para nuestro flujo, que requiere que los nombres de los genes se mapeen con el ID de ChEMBL. Este proceso lo hemos hecho usando la herramienta de mapeo Retrieve/ID que proporciona Uniprot. Sin embargo, durante el mapeo, el número de genes disminuye mucho, teniendo finalmente 88 proteínas del interactoma y 30 proteínas secundarias.

2.2 Código desarrollado

Para poder obtener los datos de ChEMBL hemos encontrado una serie de paquetes tanto en R como en Python que contienen las funciones deseadas.

2.2.1 Trabajo en Python

El principal paquete que existe y tiene acceso a ChEMBL es un cliente web que trabaja a través del lenguaje Python. Su nombre es *chembl_webresource_client*, y ChEMBL tiene una sección web dedicada a explicar su funcionamiento.

Al usar este paquete se nos presentan dos problemas. El primero es que los datos que obtenemos desde este servicio web no están conectados entre sí de la misma manera en que lo están los datos dentro de ChEMBL. Por tanto, acceder a un medicamento a partir de los targets (proteínas) disponibles no es algo fácil de hacer. El segundo problema se presenta una vez averiguamos cómo hacer esto, y es que los datos descargados no son siempre correctos, y en algunos casos una misma petición devuelve respuestas diferentes en cada ejecución. Esto nos hizo llegar a la conclusión de que no podíamos hacer un programa fiable con este paquete, y por tanto pasamos a otras opciones.

2.2.2 Trabajo en R

El paquete que tenemos disponible en R es muy básico, se llama *chemblr* y fue modificado por última vez hace siete años, por lo que no está demasiado actualizado.

En vez de basarnos en este paquete, vamos a obtener nuestra información a partir de peticiones GET a la web de ChEMBL. Usaremos de nuevo la página en la que se explica el funcionamiento del servicio web de python, solo que aplicaremos sus ejemplos sobre R. Para hacer esto, nos hemos ayudado de una página de la web de EMBL.

El paquete *chemblr* lo usaremos simplemente para obtener los json en un formato cómodo. Todos los paquetes requeridos se instalarán en la carpeta *software/deps* dentro de *code*.

2.2.3 setup.sh

Para la carga de los distintos paquetes que vamos a usar durante el desarrollo de la actividad, utilizaremos la ventana de comandos.

Hemos diseñado un script de bash llamado *setup.sh*. En este script, creamos en primer lugar un repositorio local **deps** que deberá crearse dentro de la carpeta **software**.

Antes de proceder a la carga de todas las librerías dentro de este repositorio, concedemos permisos de administrador a la misma.

Hemos definido dos listas de paquetes, una para aquellos que se obtienen desde CRAN (jsonlite, httr, devtools, sjmisc, tidyverse, networkD3, magrittr) y otra para aquellos que se obtienen a través del paquete previamente instalado devtools, que en nuestro caso es solo una (chemblr).

Para cada una de estas listas, ejecutamos un bucle for dentro del cual se instalarán las librerías una por una.

2.2.4 launch.sh

Para ejecutar todo el flujo de trabajo que hemos generado, utilizamos el archivo *setup.sh*. A este archivo le pasamos como argumento un número, que puede ser **1**, **2**, **3** o **4**. Si le pasamos **1**, trabajaremos con fármacos en fase 3 y proteínas de primer grado de interacción; en el caso de pasar un **2**, utilizaremos fármacos en fase

3 y proteínas secundarias; de la misma manera ocurre con el **3** y el **4**, a diferencia de que en este caso los fármacos serán de fase 4.

Este archivo primero cargará todas las librerías que hemos descargado previamente con **setup.sh** y estarán en la carpeta **deps** dentro de **software**.

Posteriormente, se establecen las variables locales que vamos a usar durante el estudio. Seguidamente se cargan las funciones realizadas y se pasa a la ejecución, donde se utilizan esas funciones.

Finalmente, se procede a guardar los archivos **CSV** y las gráficas que hayamos generado durante la actividad en la carpeta **results**.

Aquí podemos ver cómo quedaría el flujo:

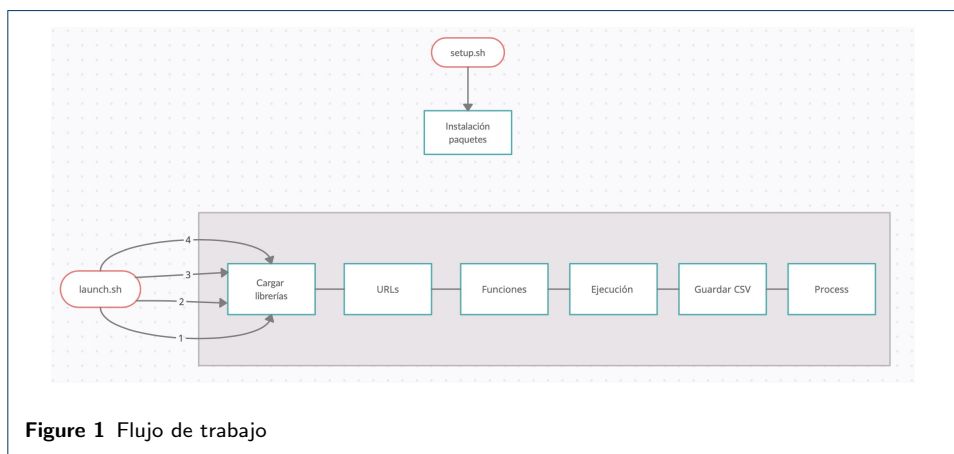


Figure 1 Flujo de trabajo

3 Resultados

3.1 Información de los fármacos

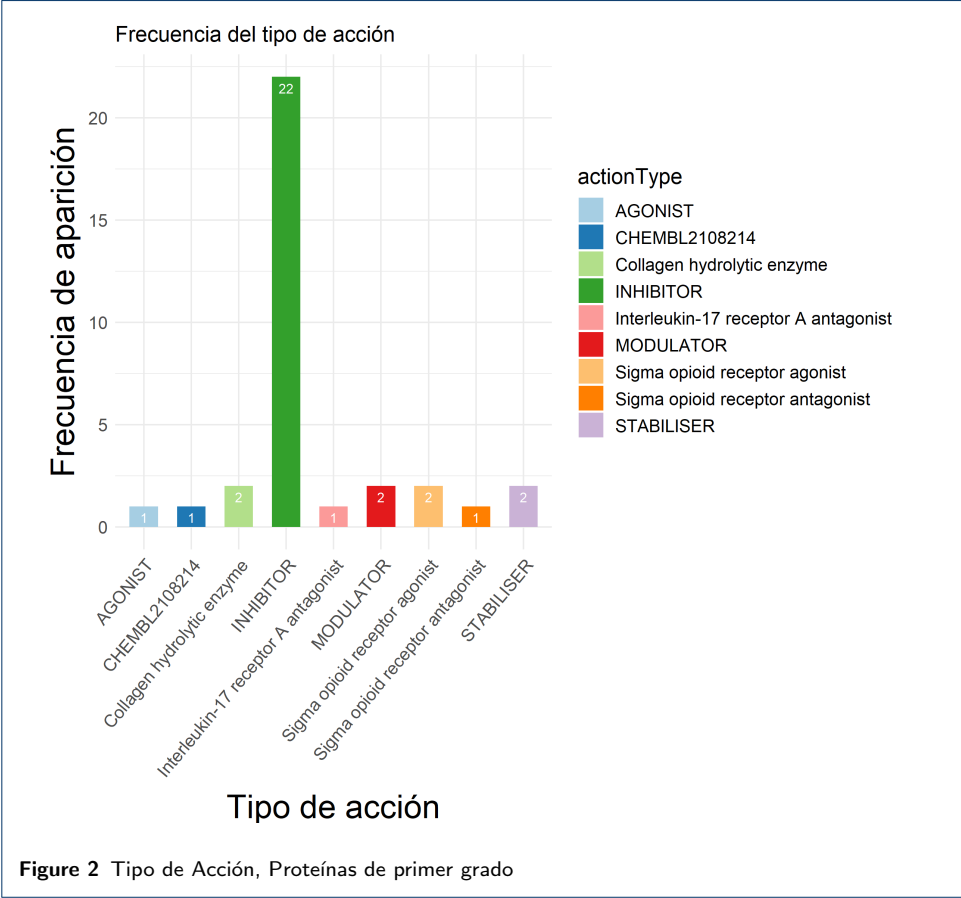
Tras la ejecución de nuestro flujo de bash con sus cuatro opciones, en la carpeta de resultados vamos a tener un directorio para cada combinación de fase del ensayo clínico y del conjunto de proteínas diana buscadas. En la carpeta encontramos dos archivos csv, uno contiene información genérica del fármaco (id de ChEMBL del fármaco y de la proteína objetivo, fecha de aprobación, etc.) y el otro corresponde a la información química de cada fármaco encontrado.

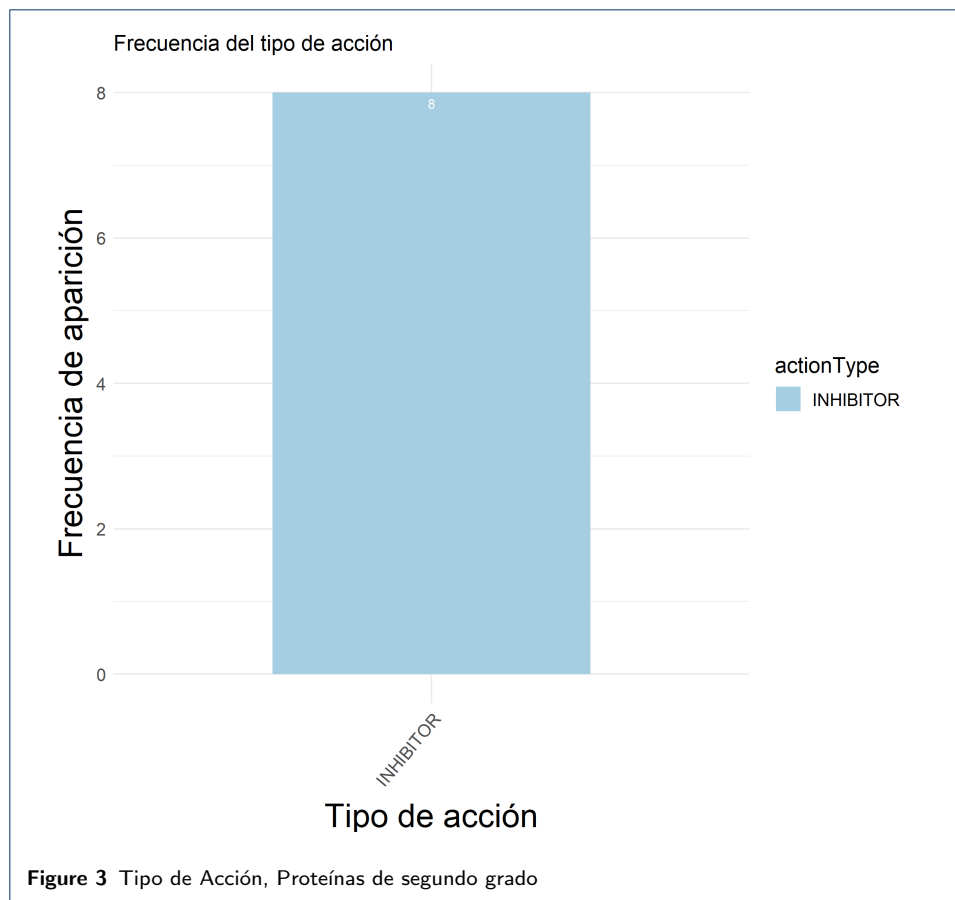
En nuestro trabajo, no usamos la información química obtenida para analizar los fármacos resultantes de nuestra búsqueda. Sin embargo, como vemos en [7] datos químicos como el número de anillos aromáticos pueden ser de gran ayuda a la hora de estudiar los efectos de un fármaco sobre el organismo. Por ello, guardamos esta información y definimos en el anexo el significado de cada variable de los archivos csv.

3.2 Gráficas en R

Tras obtener los datos de los fármacos para fase del ensayo clínico 3 y 4 y para proteínas del interactoma SARS-Humano y las proteínas secundarias, vamos a realizar una serie de gráficos que nos ayuden a analizarlos. A continuación, mostraremos solo aquellos correspondientes a los fármacos de fase cuatro debido a que los de fase tres nos proporcionan muy pocos resultados.

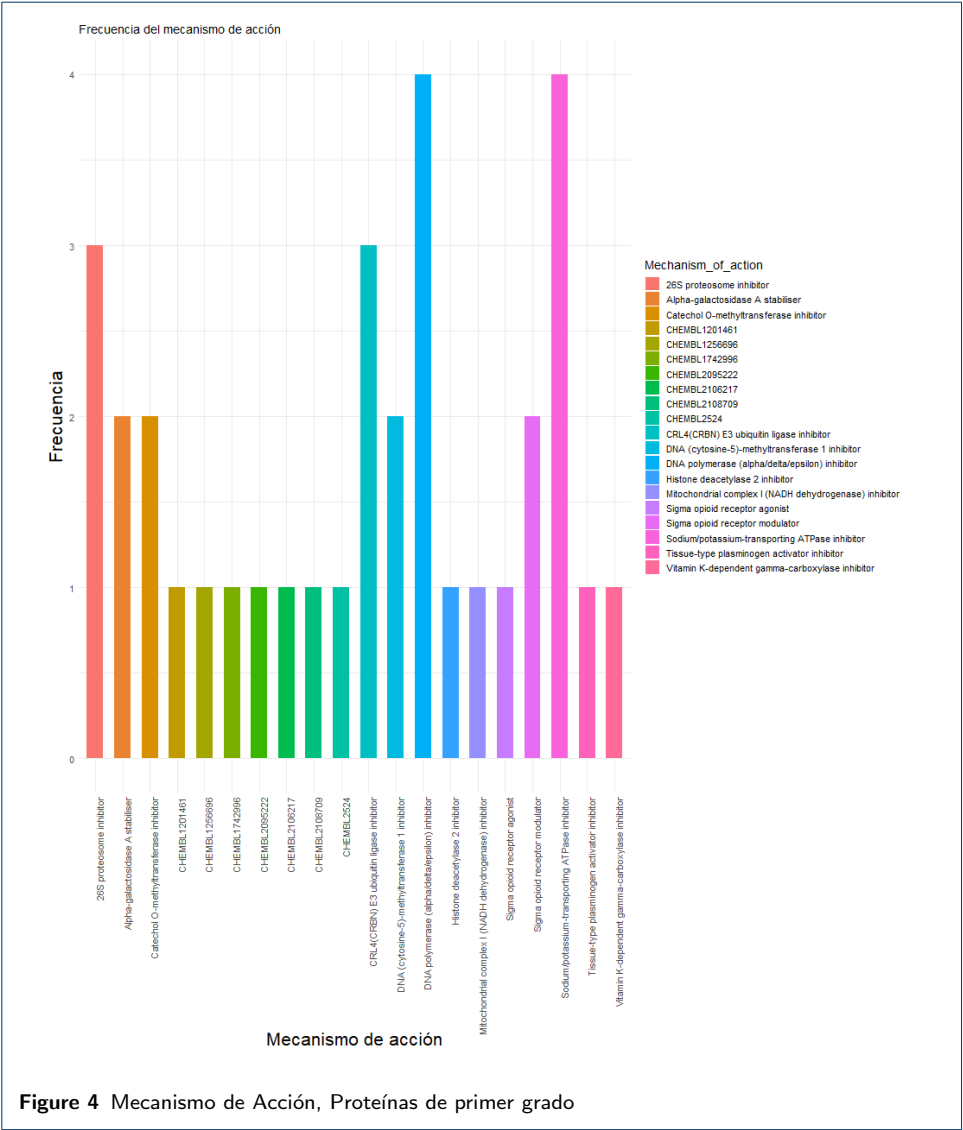
Lo primero que creamos es una gráfica de barras en la que representamos la frecuencia de cada tipo de acción tanto para las proteínas de primer grado como para las de segundo. Con esto buscamos obtener una idea de cuál es más viable para ser utilizado contra el COVID-19. Estos tipos de acción se refieren a la manera en la que el fármaco interactúa con la proteína o con su entorno.

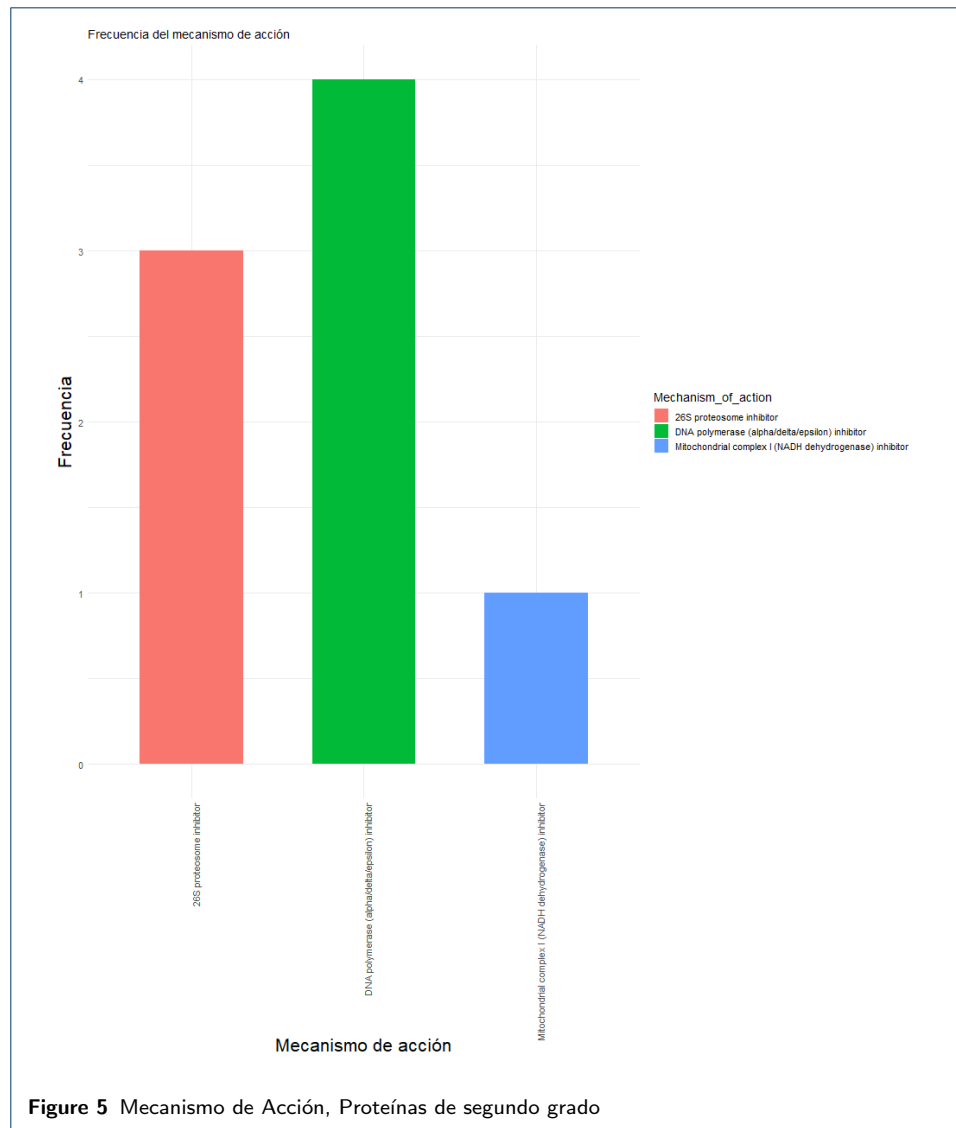




Como se puede comprobar, los inhibidores prevalecen. En el caso de las proteínas de primer grado, veintidós de los treinta y cuatro fármacos encontrados son inhibidores. En el caso de las de segundo grado, la totalidad de los fármacos (ocho) tiene acción inhibidora, significando esto que la mayoría de los fármacos encontrados para estas proteínas se dedican a reducir la actividad de enzimas de manera reversible, acoplándose a ellas.

Las siguientes dos gráficas muestran la frecuencia de cada mecanismo de acción existente. Conocer el mecanismo de acción de un fármaco aumenta la información conocida sobre su tipo de acción, pues especifica con mayor profundidad la actividad del fármaco sobre la proteína y la reacción del paciente a él. Con esto podremos ver si algún mecanismo de acción es muy común.





Observando estas gráficas podemos ver que para las proteínas de primer grado existen cuatro grupos principales de inhibidores, que actúan sobre el proteasoma 26S, la ubiquitina ligasa, la DNA polimerasa y la ATPasa transportadora de sodio y potasio. Otros mecanismos de los que encontramos repeticiones son los estabilizadores de alfa-galactosidasa, inhibidores de metiltransferasa y moduladores de receptor opiode.

En el caso de las proteínas de segundo grado, volvemos a encontrarnos con cuatro fármacos que inhiben a la DNA polimerasa, y tres al proteasoma 26S, además de una que inhibe al complejo mitocondrial. Que se repitan dos mecanismos en ambos tipos de proteínas podría indicar la relevancia de ellos.

Por último, representaremos redes que reflejan la relación entre proteínas y los medicamentos que las afectan. Por tanto podremos encontrar proteínas con mayor cantidad de medicamentos que las puedan afectar.

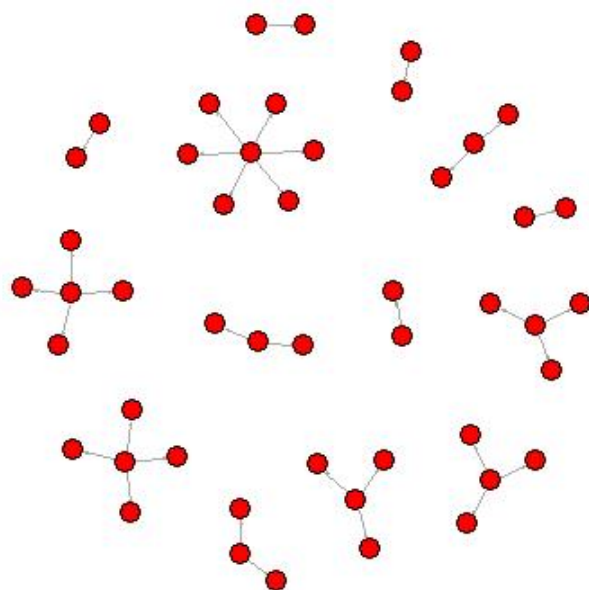
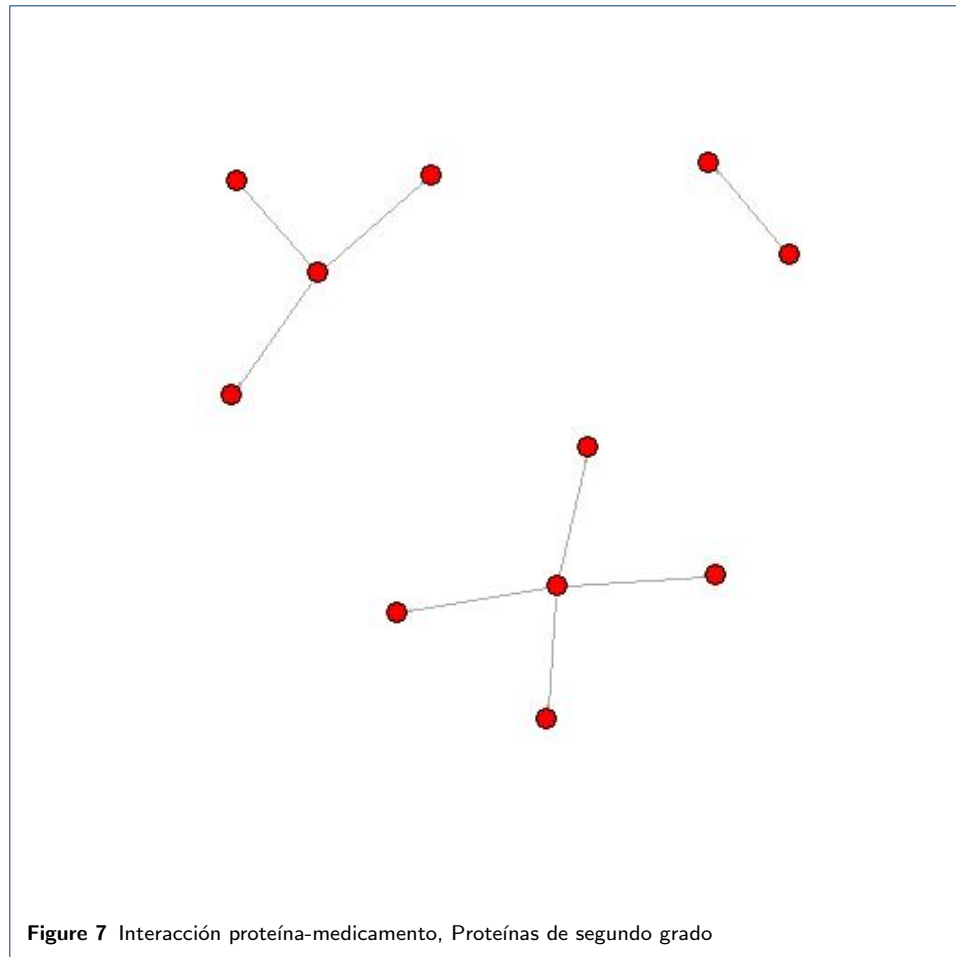


Figure 6 Interacción proteína-medicamento, Proteínas de primer grado



A partir de estas redes podemos ver que del primer grupo de proteínas extraemos catorce proteínas relacionadas con nuestros treinta y cuatro fármacos, de las que destaca una con seis medicamentos que la tienen como objetivo.

En el segundo grupo solo se localizan fármacos para tres de las proteínas, pues a dos de ellas las afecta más de un fármaco.

4 Discusión

Una de las principales apreciaciones de los resultados es que, en ambos caso, el tipo de acción predominante son los **inhibidores**, que disminuyen o cesan la acción o creación de otra sustancia.

Por otro lado, podemos ver los mecanimos de acción más presentes:

- 1 **Inhibidor de DNA polimerasa**, fundamentales para la replicación del ADN.
- 2 **Inhibidor 26s proteosoma**, un complejo protéico que se encarga de degradar proteínas no necesarias o dañadas.
- 3 **Inhibidor de la bomba de Sodio/Potasio**, una enzima que realiza varias funciones en la celula, como mantener el gradiente Sodio/Potasio, el potencial de membrana o de transporte.
- 4 **Inhibidor E3 ubiquitina ligasa**, proteína que recluta una enzima de ubiquitina E2, reconoce un sustrato de proteína y ayuda o cataliza directamente la transferencia de ubiquitina desde el E2 al sustrato de proteína.

Tiene sentido que muchos de nuestros fármacos compartan mecanismo de acción ya que están destinados a grupos de proteínas relacionadas, que comparten función biológica.

5 Conclusiones

El objetivo de esta investigación ha sido encontrar fármacos que puedan combatir el virus SARS-Cov-2 usando la base de datos ChEMBL. Para ello, hemos comenzado nuestro trabajo buscando el interactoma funcional SARS-humano en la literatura científica. Aunque se han realizado numerosos artículos en poco tiempo, aún no se ha descubierto el interactoma funcional completo. Debido a ello, decidimos ampliar la búsqueda de fármacos usando proteínas humanas que interaccionan directamente con las proteínas humanas del interactoma funcional.

Aunque intentamos partir de un conjunto numeroso de proteínas de segundo grado, no ha sido posible debido a que se pierden muchas de ellas al mapear su código de gen al código ChEMBL. Esto se debe a que no existen fármacos en ChEMBL cuyo objetivo sean estas proteínas. Aún así, gracias a las proteínas de segundo grado se consiguen otros ocho fármacos potenciales de fase cuatro.

Dentro de los fármacos obtenidos, encontramos algunos mecanismos de acción que son más frecuentes que los demás. Como se ha dicho previamente, esto puede deberse a que algunas de las proteínas comparten función biológica. También, pueden tratarse de mecanismos muy estudiados.

Para futuros trabajos, sería interesante utilizar la información química obtenida de los fármacos potenciales para estudiar el efecto de todos ellos sobre la acción del virus, así como las rutas que siguen en el organismo humano. Además, otra línea de investigación sería estudiar cuál de las proteínas del interactoma funcional es una candidata óptima a desarrollar un fármaco, simulando la repercusión que tendría en el sistema.

6 Anexos

En el siguiente anexo vamos a explicar el significado de cada variable guardada en los archivos csv resultados de la ejecución de nuestro flujo de bash.

6.0.1 Archivo con información general

A continuación, se muestran las variables guardadas y su significado:

- idFarmaco: es el identificador de ChEMBL que tiene el fármaco.
- idTarget: es el identificador de ChEMBL que tiene la proteína diana del fármaco.
- fechaAprobación: fecha en la que el fármaco fue aprobado.
- canonicalSmile: (Simplified Molecular Input Line Specification o SMILES) es una especificación para describir sin ambigüedades la estructura de una molécula usando cadenas ASCII cortas.
- acciónType: indica que tipo acción realiza el fármaco.
- mechanismOfAction: atributo la forma en la que actúa el fármaco.

6.0.2 Archivo con información química

A continuación, se muestran las variables guardadas y su significado:

- AlogP: valor calculado de la lipofiliidad de una molécula expresado como log (coeficiente de partición octanol / agua).
- Aromatics_rings: número de anillos aromáticos.
- cx_logd: se define como la relación de concentraciones de todas las especies moleculares (neutras e ionizadas) en octanol dividida por la concentración de todas las especies en medios acuosos al pH especificado.
- cx_logp: es el coeficiente de Partituton octanol/agua calculado utilizando un método basado en fragmentos desarrollado por ACDlabs.
- cx_most_apka: es el pKa para el grupo más ácido de la molécula.
- cx_most_bpka: es el pKa para el grupo más básico de la molécula.
- hba: número de aceptores de enlaces de hidrógeno.
- hba_lipsinki: número de nitrógenos y oxígenos en la molécula.
- hbd: número de aceptores de enlaces de hidrógeno.
- hbd_lipsinki: número de hidrógenos unidos a átomos de nitrógeno u oxígeno.
- heavy_atomos: número de moléculas que no sean hidrógeno en la molécula.
- molecular_species: descripción de la especie predominante con pH 7.4. Puede ser ACID, BASE, NEUTRAL o ZWITTERION.
- mw_freebase: peso molecular de la forma padre de la molécula.
- mw_monoisotopic: a masa monoistópica del compuesto calculada como la suma de las masas de los isótopos más abundantes en el compuesto.
- psa: el área de la superficie polar se calcula mediante el método de P Erti. El cálculo rápido del área de superficie polar molecular es la suma de contribuciones basadas en fragmentos y su aplicación a la predicción de propiedades de transporte de fármacos, Ertl, P., Rohde, B., Selzer, P., J. Medicina. Chem. 2000, 43, 3714-3717.
- qed_weight: esta es la estimación cuantitativa de la similitud con las drogas como se describe en: "Cuantificando la belleza química de las drogas" G. Richard Bickerton, Gaia V. Paolini, Jeremy Besnard, Sorel Muresana y Andrew L. Hopkins, Nature Chemistry, 2012, 4, 90-98 Los valores oscilan entre

0-1, donde 1 es el más parecido a una droga y 0 el menos parecido a una droga.

- ro3_pass: Regla de 3 pases. Se sugiere que los compuestos que pasan estos criterios tienen más probabilidades de ser aciertos en la selección de fragmentos.
- rtb: número de enlaces rotativos en la molécula.

Disponibilidad de datos y materiales

El repositorio de Github usado es: https://github.com/clarajvv/repo_farm_COVID

Contribución de los autores

L.V.M : Búsqueda de las proteínas del interactoma SARS-Humano, búsqueda de proteínas secundarias en STRING, mapeo de ICHEMBL, búsqueda inicial de métodos para la extracción de datos de ChEMBL, desarrollo de funciones en R para obtener las relaciones entre proteínas y fármacos y los datos de estos en ChEMBL, creación de gráficas de tipos de acción y mecanismos de acción y guardado de las mismas, escritura anexo, escritura de metodología.

C.J.V : Búsqueda inicial de métodos para la extracción de datos de ChEMBL, desarrollo de código en Python para extracción de información desde ChEMBL (fallido), desarrollo en R de pruebas de extracción de relaciones entre proteínas y fármacos y los datos de estos desde ChEMBL, búsqueda de métodos de traducción de IDs de proteínas, ayuda en prueba y error de archivos bash, inclusión de la bibliografía en el informe, comunicación con los profesores, labores de coordinación, maquetación y comentario de las gráficas en el informe, correcciones finales del informe, escritura de metodología.

I.S.J : Documentación sobre el problema [6], búsqueda de métodos de extracción de información de STRING y GO desde R, documentación sobre uso de bash, creación, correcciones y actualizaciones de archivo bash para la instalación de paquetes, creación, correcciones y actualizaciones de archivo de bash para la ejecución del código R, escritura de metodología.

P.T.R : Búsqueda de métodos de extracción de información de STRING y GO desde R, recopilación de la bibliografía y creación del archivo para mostrarla, creación de gráficas de interacción y guardado de las mismas, escritura de discusiones y anexo.

Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

References

1. Yan, Y., Shin, W.I., Pang, Y.X., Meng, Y., Lai, J., You, C., Zhao, H., Lester, E., Wu, T., Pang, C.H.: The first 75 days of novel coronavirus (SARS-CoV-2) outbreak: Recent advances, prevention, and treatment. MDPI AG (2020). doi:[10.3390/ijerph17072323](https://doi.org/10.3390/ijerph17072323). <https://pubmed.ncbi.nlm.nih.gov/32235575/>
2. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., Müller, M.A., Drosten, C., Pöhlmann, S.: SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**(2), 271–2808 (2020). doi:[10.1016/j.cell.2020.02.052](https://doi.org/10.1016/j.cell.2020.02.052)
3. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., Zhou, Q.: Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* (80-.). **367**(6485), 1444–1448 (2020). doi:[10.1126/science.abb2762](https://doi.org/10.1126/science.abb2762)
4. Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., Gallagher, T.: A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. *J. Virol.* **85**(2), 873–882 (2011). doi:[10.1128/jvi.02062-10](https://doi.org/10.1128/jvi.02062-10)
5. Chan, J.F.W., Kok, K.H., Zhu, Z., Chu, H., To, K.K.W., Yuan, S., Yuen, K.Y.: Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **9**(1), 221–236 (2020). doi:[10.1080/22221751.2020.1719902](https://doi.org/10.1080/22221751.2020.1719902)
6. Gysi, D.M., Valle, Í.D., 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. *arXiv* (2020). [2004.07229](https://arxiv.org/abs/2004.07229)
7. Polêto, M.D., Rusu, V.H., Grisci, B.I., Dorn, M., Lins, R.D., Verli, H.: Aromatic rings commonly used in medicinal chemistry: Force fields comparison and interactions with water toward the design of New Chemical Entities. *Frontiers in Pharmacology* **9**(APR), 1–20 (2018). doi:[10.3389/fphar.2018.00395](https://doi.org/10.3389/fphar.2018.00395)