

Las bases de datos y su importancia en

La Dra. Georgina Hernández Montes es Química Farmacobióloga egresada de la Facultad de Química de la UNAM. Actualmente es parte de la Red de Apoyo a la Investigación (RAI) de la UNAM y se ha especializado en el área de bioinformática.

Esta publicación fue revisada por el comité editorial de la Academia de Ciencias de Morelos.

Desde que se desarrolló la escritura hace cuatro mil años en Sumeria, los seres humanos nos hemos dedicado a guardar registros de las actividades que consideramos importantes. En un inicio solo se registraban las actividades comerciales y administrativas, pero conforme se fueron desarrollando las civilizaciones la necesidad de registrar las observaciones e incluso el pensamiento de la época fue creciendo. Una vez que se tuvieron suficientes registros fue necesario organizarlos y catalogarlos dentro de lugares especializados, tales como las bibliotecas, para poder acceder a la información de manera rápida y dirigida, ya sea para tomar decisiones o generar nuevo conocimiento. Podría decirse que las bibliotecas fueron las primeras bases de datos.

La actividad científica de los últimos 50 años ha generado una gran cantidad de información que se ha organizado en bases de datos electrónicas y que son accesibles desde cualquier lugar del planeta donde se tenga una computadora y acceso a internet. Este hecho ha sido uno de los elementos clave que ha permitido que la respuesta a la pandemia de COVID-19, al menos desde el punto de vista científico, haya sido la más eficaz de toda la historia. Es por ello, estimado lector o lectora, que me gustaría poner a su consideración al menos cuatro de las bases de datos públicas que considero han sido fundamentales para generar nuevo conocimiento relacionado con la pandemia.

Literatura científica especializada

Uno de los sitios más populares para la búsqueda de literatura científica en el área de ciencias de la vida es el Centro Nacional de Información Biotecnológica (NCBI por sus siglas en inglés) de Estados Unidos. En este sitio se puede encontrar la información más relevante de cada publicación y un enlace al artículo original para una gran cantidad de temas y organismos. A continuación pongo la traducción de la misión del NCBI que aparece publicada: “[...] el NCBI se ha encargado de crear sistemas automatizados para almacenar y analizar conocimientos sobre biología molecular, bioquímica y genética; facilitar el uso de dichas bases de datos y software por parte de la comunidad médica y de investigación; coordinar esfuerzos para recopilar información biotecnológica tanto a nivel nacional como internacional.”

Sin embargo, desde que inició la emer-



FIGURA 1. IMAGEN del sitio del Consorcio mexicano de vigilancia genómica.

gencia sanitaria por el virus SARS-CoV-2 a principios del 2020, la actividad científica ha sido espectacular. Así, se han publicado un promedio de 300 artículos diarios que abordan diferentes aspectos de la pandemia, tales como la secuencia genómica del virus, los efectos de la enfermedad, los desarrollos de fármacos, los estudios clínicos de las vacunas, entre muchos otros.

Por esta razón el NCBI decidió crear LitCovid, una nueva base de datos exclusiva para COVID-19 en donde las publicaciones están organizadas por temas, localización geográfica y tipo de publicación, esta última estrechamente relacionada con el impacto y la calidad de la misma.

Este sitio no solo permite simplificar las búsquedas, sino hacer un más eficiente análisis de la información, ya que es posible descargarla en formatos de texto fácilmente manipulables para los científicos de datos. Si usted desea consultarla, puede dirigirse al siguiente sitio: <https://www.ncbi.nlm.nih.gov/research/coronavirus/>

Secuencias genómicas del virus SARS-CoV-2

En enero de 2020, en un hecho histórico por la rapidez de la respuesta entre la identificación de un agente infeccioso y disponibilidad de información sobre su estructura genómica, se hicieron públicas las primeras secuencias genómicas del virus del SARS-CoV-2 en la plataforma GISAID, que originalmente fue diseñada para compartir información genómica del virus de la influenza. Esta plataforma es una iniciativa internacional para promover que los diferentes

países compartan las secuencias genómicas de las variantes del virus detectadas localmente. A la fecha hay más de un millón de secuencias depositadas provenientes de al menos 72 países, lo que ha permitido hacer un seguimiento casi en tiempo real de la evolución del virus, de la aparición de variantes y de su dispersión.

Es importante mencionar que México y particularmente Morelos participan en la tarea de compartir datos a través del CoViGen-Mex, un consorcio de varias instituciones como el Instituto de Biotecnología de la Universidad Nacional Autónoma de México, el Instituto Mexicano del Seguro Social, el Instituto Nacional de Enfermedades Respiratorias, el Laboratorio Nacional de Genómica para la Biodiversidad, el Centro de Investigación en Alimentación y Desarrollo, la Universidad Autónoma de la Ciudad de México y la Universidad Autónoma de San Luis Potosí (Figura 1). En coordinación con el InDRE, estas instituciones se encargan de tomar las muestras de pacientes, secuenciarlas, hacer el análisis bioinformático y generar los reportes de la presencia de las diferentes variantes en el territorio nacional. Si usted desea conocer más acerca de esta iniciativa, este es el enlace: <http://mexcov2.ibt.unam.mx:8080/COVID-TRACKER/>

Estructura tridimensional del virus

Para poder dar respuesta a muchas de las preguntas que surgieron a partir de identificar la capacidad de infección del virus del SARS-CoV-2 y su letalidad, era necesario conocer no solo la estructura tridimensional del virus y de las proteínas que lo conforman, sino también a las proteínas de nuestro organismo.

Por ejemplo, la estructura de los receptores dentro de nuestras células con las que interacciona la proteína “Spike” del virus nos pueden explicar mucho de su mecanismo de invasión. La base de datos que contiene esta información es el Protein Data Bank o PDB por sus siglas en inglés. En general, este sitio contiene la estructura de todas las proteínas de estructura conocida, hasta la fecha cerca de unas 180,000, la mayor parte obtenidas por difracción de rayos X.

Como es de suponerse, contiene también la estructura de las 27 proteínas del virus, y en particular, la estructura de decenas de proteínas con las que puede interaccionar, como receptores, inhibidores, anticuerpos, etc.

De hecho, en la sección “la molécula del mes” del PDB se publicó en el mes de junio de 2020, la estructura de la ahora famosa proteína espiga (Spike en inglés) del SARS-CoV-2 (<https://pdb101.rcsb.org/motm/246>). En la Figura 2 se muestra un modelo de la interacción entre la proteína Spike del SARS-CoV-2 y un anticuerpo que se está probando como potencial tratamiento. Gracias a esta información se puede ver con mucho detalle cómo están funcionando los anticuerpos. Con este tipo de información también se pueden generar modelos teóricos de interacción entre probables fármacos y proteínas de interés.

Ensayos clínicos

Uno de los procesos determinantes para aprobar la administración y comercialización de un fármaco o una vacuna son los ensayos clínicos, que constan de diferentes etapas, para probar cómo administrar la vacuna o fármaco; cuál es la mejor dosis; si son seguras y si no causan reacciones adversas graves. Cada ensayo cuenta con una estrategia o plan de acción en el que se establece el objeti-



la pandemia de COVID-19.



FIGURA 2. ESTRUCTURA tridimensional de la proteína spike del virus y el anticuerpo neutralizante COVA2-04. Imagen tomada de <https://www.rcsb.org/structure/7JMO>

vo de la prueba, qué se hará, cómo se hará y la justificación de cada etapa del ensayo. Este protocolo debe ser evaluado por un comité que incluye a varios miembros de la comunidad, tales como médicos, estadistas, científicos, entre otros, para asegurar que el estudio cumpla con las reglas de la ética, que protege los derechos y el bienestar de los participantes y que asegura que los riesgos del ensayo clínico sean menores que los beneficios potenciales que el resultado de la prueba aportará a la humanidad.

Uno de los sitios de referencia donde se dan a conocer los estudios clínicos para diferentes fármacos y de diferentes países es la base de datos "clinical trials", que pertenece también al NCBI. Esta base de datos contiene más de trescientos cincuenta mil estudios de 220 países, en los que se reporta información acerca de su protocolo y eventualmente de los resultados. Gracias a esta información, la sociedad puede informarse sobre cuáles fueron las condiciones que se usaron para los estudios clínicos, tanto de las vacunas hasta ahora desarrolladas como de algunos tratamientos que están en curso contra el COVID-19. Si usted desea más información acerca de este sitio o de algún estudio clínico puede visitar el siguiente enlace: <https://clinicaltrials.gov/>

Buscando fármacos

Una de las estrategias que cobró mayor relevancia para encontrar una molécula efectiva para tratar el COVID-19 es el "reposicionamiento de fármacos". Esta estrategia consiste en buscar moléculas ya evaluadas y autorizadas para tratar otros padecimientos, con el fin de probar su eficacia contra el SARS-CoV-2, ya sea inhibiendo su capacidad de infección, inhibiendo su replicación o al menos disminuyendo los efectos de la infección en el organismo. Esta estrategia no solo tiene el potencial de disminuir los tiempos de desarrollo de antivirales sino también disminuir los costos de producción, evaluación y distribución. De esta estrategia surgió el único antiviral disponible avalado por la FDA: el remdesivir. Este fármaco ya se había utilizado previamente para tratar el ébola, una enfermedad mortal que produce hemorragias agudas.

Una de las bases de datos más útiles ha sido el DrugBank, plataforma que contiene la información de la estructura de las moléculas, de su toxicidad, de la interacción con otras moléculas, de sus efectos secundarios, de la farmacocinética, de la farmacodinámica o bien de lo que el fármaco le hace al cuerpo y de cuáles son las formas de

administración disponibles. Esta información permite seleccionar candidatos a probar y puede complementarse con la información sobre la estructura tridimensional como la del PDB. En el caso de que el antiviral se trate de una proteína o de que existan estructura de las proteínas del virus en complejo con estas. Si usted desea conocer o consultar esta base de datos puede dirigirse al siguiente enlace: <https://go.drugbank.com/>

Finalmente me gustaría mencionar un sitio que, aunque no es una base de datos propiamente, es un lugar que va recopilando y actualizando la información sobre el avance de la vacunación en el mundo y que muestra gráficas comparativas entre los diferentes países. Es la sección de vacunas del sitio "Our World in Data". Contar con esta información nos permite evaluar el desempeño de las políticas que se han implementado en cada país para el manejo de la pandemia. En la Figura 3 se muestra una imagen del avance de la vacunación al día 22 de julio de 2021.

Vacunación

Como hemos visto, en esta era de la información, las bases de datos son recursos valiosísimos que nos permiten utilizar la información acumulada a lo largo del tiempo para generar nuevo conocimiento, para ahorrar tiempo en el desarrollo de estrategias terapéuticas en el caso de una emergencia sanitaria e incluso para tomar decisiones estratégicas de acuerdo con la evolución del virus y su dispersión por el mundo. De alguna manera son también un elemento que permite validar información que circula en redes sociales. Es por ello que es necesario invertir en su construcción y mantenimiento tanto recursos económicos como recursos humanos, pues estas bases de datos necesitan estar hospedadas ya sea en la computadora local de algún

centro de investigación o bien en alguna nube, deben tener interfaces amigables, que solo se logran con buenos programadores, la información debe ser verificada para evitar errores, sesgos y fraudes y también debe ser actualizada constantemente para que no pierdan valor y credibilidad.

El estadístico estadounidense William Edwards Deming decía que "sin datos no eres más que una persona con una opinión". Es justo en estos momentos donde necesitamos más datos que opiniones. Es fundamental que en el país se invierta en buenas bases de datos que nos permitan poner a disposición no solo de la comunidad científica sino de la población en general información confiable y actualizada para ser consultada y utilizada, no solo porque eso nos va a permitir tomar mejores decisiones sino porque una buena parte de esa información ha sido generada a partir de recursos públicos.

Esta columna se prepara y edita semana con semana, en conjunto con investigadores morelenses convencidos del valor del conocimiento científico para el desarrollo social y económico de Morelos. Desde la Academia de Ciencias de Morelos externamos nuestra preocupación por el vacío que genera la extinción de la Secretaría de Innovación, Ciencia y Tecnología dentro del ecosistema de innovación estatal que se debilita sin la participación del Gobierno del Estado.

Share of people vaccinated against COVID-19, Jul 22, 2021

This data is only available for countries which report the breakdown of doses administered by first and second doses.

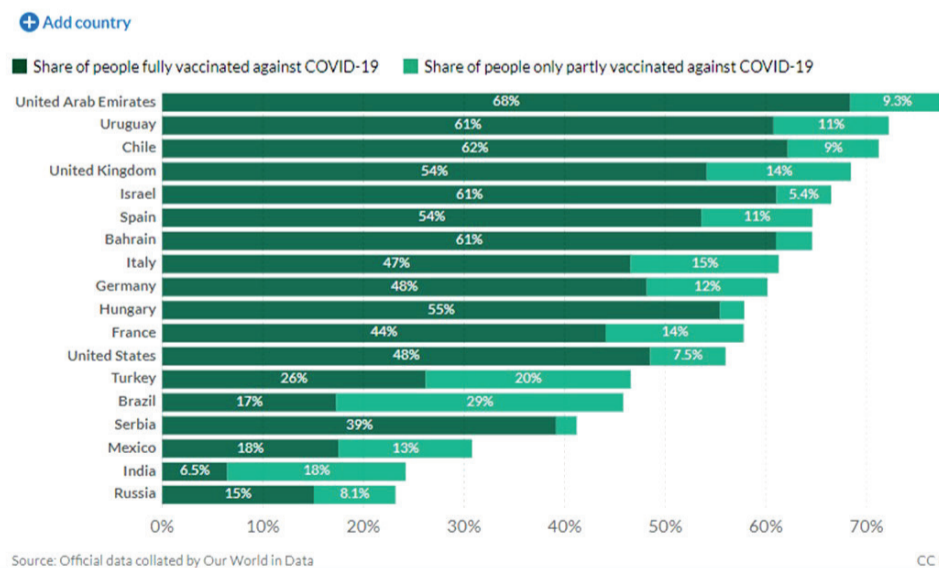


FIGURA 3. PORCENTAJE de la población vacunada contra COVID-19 en diferentes países incluido México al día 22 de julio de 2021. Imagen tomada de <https://ourworldindata.org/covid-vaccinations>

Sitios de interés

- > <https://medlineplus.gov/spanish/clinicaltrials.html>
- > <https://go.drugbank.com/>
- > <https://clinicaltrials.gov/>
- > <https://www.ncbi.nlm.nih.gov/research/coronavirus/>
- > <https://ourworldindata.org/covid-vaccinations>
- > <http://mexcov2.ibt.unam.mx:8080/COVID-TRACKER/>
- > <https://www.rcsb.org/>