



Inteligencia artificial en el diseño de fármacos: hacia la inteligencia aumentada

Artificial intelligence in drug design: towards augmented intelligence

Fernanda I. Saldivar-González¹, Eli Fernández-de Gortari² y José L. Medina-Franco³

Recepción: 16/08/2022

Aceptación: 03/01/2023

Resumen

El desarrollo e investigación de fármacos es un proceso complejo que a menudo se percibe más como un arte que como una ciencia. Las muchas incógnitas de las enfermedades humanas, su complejidad, y el gran tamaño del espacio químico para encontrar fármacos potenciales, hacen que la toma de decisiones en Química Farmacéutica sea excepcionalmente exigente. Por tal motivo, durante décadas y especialmente en años recientes, se han integrado al proceso de diseño de fármacos modelos computacionales, y en particular, modelos de inteligencia artificial (IA) en combinación con su uso racional (inteligencia e intuición humana basada en la experiencia). Aquí se proporciona una descripción general y crítica de ejemplos recientes de aplicaciones de IA en el diseño y desarrollo de fármacos. Se hace una discusión de los conceptos y enfoques computacionales involucrados, y se analizan las posibilidades y limitaciones del diseño de fármacos utilizando IA en conjunto con su uso racional (o “el razonamiento humano”): inteligencia aumentada.

Palabras clave

Química Farmacéutica; diseño de fármacos asistido por computadora; inteligencia artificial; inteligencia aumentada; medicinal personalizada.

Abstract

Drug development and research is a complex process that is often perceived as more of an art than a science. The many unknowns of human diseases, their complexity, the patient-to-patient response, and the huge size of the chemical space for finding potential drugs make decision-making in Medicinal Chemistry exceptionally demanding. For this reason, for decades and especially in recent years, computational models, and in particular artificial intelligence (AI) models in combination with their rational use (human intelligence and intuition based on experience), have been integrated into the drug design process. An overview and critique of recent examples of AI applications in drug design and development are provided here. A discussion of the concepts and computational approaches involved is given, and the possibilities and limitations of drug design using AI coupled with human reasoning are analyzed: augmented intelligence.

Keywords

Medicinal Chemistry; Computer-Aided Drug Design; artificial intelligence; augmented intelligence; precision medicine.

¹ Grupo de investigación DIFACQUIM, Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México, México. <http://orcid.org/0000-0002-0435-8662>. Correo de contacto: fer.saldivarg@gmail.com

² Grupo NanoSafety, Laboratorio Internacional Ibérico de Nanotecnología, Braga, Portugal.

³ Grupo de investigación DIFACQUIM, Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México, México. <http://orcid.org/0000-0003-4940-1107>

Introducción

La inteligencia artificial (IA) es un término que quizás nos recuerda al terreno de la ciencia ficción, sin embargo, esta tecnología no es nueva y ya está integrada en nuestro día a día. Por ejemplo, los asistentes de voz y el reconocimiento de imágenes funcionan bajo IA. Los algoritmos, que son secuencias de instrucciones finitas y bien definidas, procesan y transforman imágenes, palabras o frases de un lenguaje natural a un formato legible para programas computacionales. Durante este proceso, los algoritmos identifican reglas para relacionar datos y extraer patrones e información que les permiten llegar a predicciones que facilitan la toma de decisiones y/o la resolución de problemas.

Al igual que otras disciplinas computacionales, como la quimioinformática, la IA es de gran utilidad para extraer información de un conjunto de datos y aportar conocimientos prácticos o útiles (Gasteiger, 2016). En la ciencia, cada vez es más común encontrar publicaciones que incluyen términos de IA como "*machine learning*" (ML), refiriéndose al aprendizaje automático o aprendizaje de máquinas, o como "*deep learning*" (DL), aludiendo al aprendizaje profundo. La relación entre los conceptos antes mencionados puede ser visualizada como círculos concéntricos, donde la IA representa el área general de estudio como el círculo más grande. Dentro se encuentra ML, que floreció después, y, finalmente, DL, que es parte de ML y está impulsando la explosión de la IA actual (figura 1).

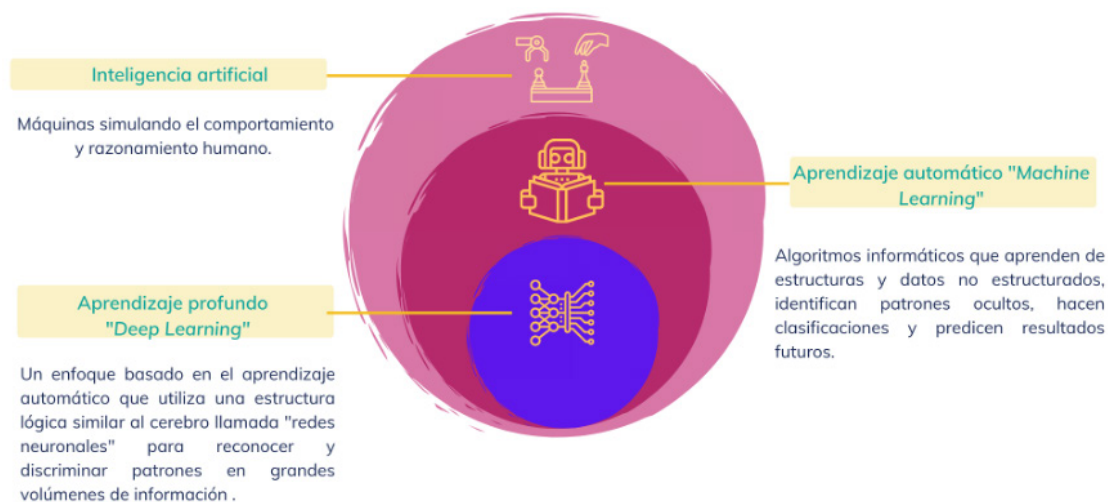


FIGURA 1. Esquema que muestra la relación entre la inteligencia artificial, el aprendizaje de máquinas y el aprendizaje profundo.

En diversas áreas de Química, incluyendo Química Farmacéutica, la IA se ha utilizado desde la década de los 50 (Gasteiger, 2020). Sin embargo, los avances tecnológicos y el desarrollo paralelo de teorías y modelos han acelerado su aplicación en años recientes. La idea de continuar adoptando la IA en el proceso de desarrollo de fármacos representa una manera de enfrentar desafíos como las bajas tasas de éxito, en donde cada vez se invierte más dinero en la obtención de nuevos medicamentos y solo pocas moléculas llegan a fases clínicas. En general, el tiempo medio requerido para el descubrimiento, estudios preclínicos y clínicos, aprobación y comercialización de un medicamento es, en total, de unos dieciséis años y cuesta una mediana de 985 millones de dólares, incluyendo los costos de los ensayos clínicos fallidos (Wouters et al., 2020). La implementación de IA, según los expertos de la

industria farmacéutica, puede reducir el tiempo de desarrollo de un nuevo medicamento a solo dos o tres años (Lamberti et al., 2019). Para ilustrar el alcance actual de los métodos computacionales (particularmente de la IA) en el diseño de fármacos, un caso destacado es el de la compañía estadounidense Insilico Medicine, que generó un modelo de IA llamado GENTRL (*GENerative Tensorial Reinforcement Learning*), el cual diseñó e identificó con éxito ¡en solo 21 días!, seis inhibidores de cinasas del receptor DPP1 (vinculado a la fibrosis pulmonar), tardando en total 46 días, incluyendo la síntesis y evaluación biológica (Zavoronkov et al., 2019). Esta gran velocidad de descubrimiento de compuestos candidatos también refleja una gran coordinación de los equipos de investigación multidisciplinarios (cómputo, síntesis química y evaluaciones biológicas). Desde luego, habrá que considerar también el tiempo y los costos de otras fases de desarrollo de los inhibidores de cinasas como fármacos.

Aunque los avances recientes en IA pueden verse como una amenaza a muchos empleos, especialmente los asociados al trabajo experimental (los referidos como *wet lab*), los avances en la IA se celebran porque se cree que, si son empleados con ética e interpretados en forma adecuada, contribuirán enormemente al mejoramiento de la sociedad. No obstante, en muchas ocasiones, el avance acelerado en la innovación y en la tecnología han generado expectativas exacerbadas, pero poco realistas. Esto se debe al uso irracional y únicamente guiado por la moda de usar nuevas tecnologías, sin llegar a estar conscientes de las limitaciones inherentes que tienen. En este sentido, la automatización y la innovación en el descubrimiento de fármacos deben ser rápidas, pero también sostenibles a largo plazo (Özdemir y Hekim, 2018).

Recientemente, se integró el término de “inteligencia aumentada” al modelo de asociación entre inteligencia humana (IH) e IA con el fin de mejorar el rendimiento cognitivo, el aprendizaje, la toma de decisiones y la generación de nuevas experiencias (Gartner Glossary, s.f.). Es decir, con este modelo no se busca que las máquinas reemplacen a los científicos, sino que los científicos comprendan las metodologías de IA, permitiéndoles pasar al ámbito de la biología de sistemas y aprovechar su propia experiencia, junto con las capacidades que ofrecen los modelos de IA (My Intelligent Machine, 2021; Bazoukis et al., 2022). Es así que se está dando un cambio en el paradigma del desarrollo de fármacos usando métodos de cómputo. En palabras del Dr. Artem Cherkasov (*University of British Columbia y Vancouver Prostate Center, Canada*) se pueden reconocer tres etapas:

- Pasado: Diseño de fármacos asistido por computadora.
- Presente (2023): Diseño de fármacos con inteligencia aumentada.
- Futuro: Diseño de fármacos asistido por el ser Humano.

En esta propuesta del Dr. Cherkasov se hace énfasis en el rol de la inteligencia aumentada y la tendencia hacia el futuro en el desarrollo de fármacos con la automatización *casí total* (“asistido por el ser Humano”).

Históricamente, los inicios de IA en el descubrimiento y desarrollo de fármacos están en los llamados estudios cuantitativos de relaciones estructura-actividad (QSAR, de las siglas en inglés *quantitative structure-activity relationships*) que comenzaron a emplearse en la década de los 60' (Hansch y Fujita, 1964). En estos estudios, que aún se siguen utilizando, el objetivo es encontrar un modelo matemático que sea capaz de aproximar

las asociaciones intrínsecas entre estructuras químicas —codificadas por una serie de descriptores o representaciones— que son una parte esencial de cualquier modelo de IA) con la actividad biológica. Aquí cabe mencionar que, como otras tecnologías, el QSAR en algunos casos es empleado de manera irracional y las malas prácticas incentivadas por la “popularidad” llevaron a crear falsas expectativas hasta el punto de crear la noción de la llamada “decepción del QSAR” (S. R. Johnson, 2008).

Recientemente, dentro de la investigación de fármacos, los algoritmos de IA han ganado atención de manera constante y se están viendo numerosas aplicaciones dentro de las diferentes etapas del proceso de descubrimiento y desarrollo de fármacos (figura 2). Por ejemplo, en las primeras etapas, los métodos de IA han tenido un impacto sustancial en la predicción de la estructura tridimensional (3D) de proteínas (Jumper et al., 2021). También, han ayudado a sugerir un gran número de compuestos similares a fármacos para ensayar de manera efectiva su actividad biológica/bioquímica en estudios de cribado de alto rendimiento (HTS, por sus siglas en inglés, *high throughput screening*). Los datos de compuestos químicos, proteínas y ensayos biológicos generados hasta el momento han sido utilizados y transformados por los métodos de IA para identificar de manera efectiva objetivos biológicos de compuestos bioactivos y para predecir el potencial terapéutico de pequeñas moléculas orgánicas. El papel de la IA en la predicción de propiedades relacionadas con la absorción, distribución, metabolismo, eliminación y toxicidad (ADME-Tox) ha sido parte importante en la toma de decisiones, ya que facilita la selección de compuestos para su síntesis y evaluación biológica. Un área de investigación que crece en paralelo con las tecnologías disponibles es la relacionada con la síntesis química. En esta área, los modelos de IA se utilizan para planear y optimizar las síntesis de compuestos, para estimar la accesibilidad sintética de compuestos presentes en bibliotecas químicas (o quimiotecas) que pueden llegar a tener miles o millones de compuestos, y para automatizar el proceso de síntesis (Schneider, 2018).



FIGURA 2. Ejemplos de aplicaciones actuales de la inteligencia artificial en el descubrimiento y diseño de fármacos.
Abreviaciones: ADME-Tox, absorción, distribución, metabolismo, excreción y toxicidad.

Los métodos de IA también se emplean en etapas preclínicas para identificar biomarcadores específicos de tejidos (ensayos *in vitro*) y en etapas clínicas para la selección de pacientes en ensayos clínicos, para identificar nuevos usos de medicamentos ya existentes (reposicionamiento de fármacos) (Olascoaga-del Angel, 2022; Saldivar-González y Medina-Franco, 2021) y en farmacovigilancia, para comprender y prevenir efectos adversos de los

fármacos (Mak y Pichika, 2019). Sin embargo, estas últimas aplicaciones quedan fuera de la revisión de este artículo, que tiene como objetivo revisar y discutir las aplicaciones de IA en la identificación de compuestos bioactivos como potenciales candidatos a fármacos.

Métodos de IA en el diseño de fármacos

La ciencia del siglo XXI está sufriendo grandes transformaciones debido al creciente acceso de información disponible en bases de datos, muchas de ellas de acceso público. En cierto modo, estamos ante una nueva manera de percibir la Química y la Biología, centradas en la obtención y procesamiento cuantitativo de datos biológicos (bioinformática), datos de moléculas “pequeñas” con peso molecular menor a 1000 Da (quimioinformática), e incluso datos clínicos (informática biomédica) que propician el análisis inductivo y la simulación de sistemas vivos (López-López et al., 2021). Para obtener sentido de la información contenida en los datos se requiere de nuevas formas de pensar y aprender e, incluso, nuevas metodologías de trabajo (Hadley, 2004). La IA, y en particular las metodologías de ML y DL, han mostrado ser eficientes tanto en términos de mejora del rendimiento como de reducción de gastos dentro del diseño de fármacos (Kolluri et al., 2022). Cabe enfatizar que en todos los casos, ML y DL dependen de la existencia de datos y, como en cualquier proceso de aprendizaje, la calidad de los modelos y el eventual conocimiento que se genere dependerá en gran medida de la calidad y confiabilidad de los datos que se emplearon para el entrenamiento.

Aprendizaje automático en el diseño y descubrimiento de fármacos

Los enfoques de ML proporcionan un conjunto de herramientas y algoritmos bien establecidos que permiten crear modelos que se entrenan usando una gran cantidad de datos para realizar predicciones y ofrecer soluciones a problemas específicos. Según la tarea que se quiera efectuar y los datos disponibles, será más adecuado trabajar con un algoritmo u otro, o eventualmente combinaciones de ellos. No obstante, es importante tener en mente que, el aprendizaje automático puede usar los datos de varias maneras diferentes para obtener información significativa.

En la literatura científica, se pueden encontrar aplicaciones exitosas de ML en el área de diseño de fármacos que usan diversos algoritmos (Priya et al., 2022; Vamathevan et al., 2019), algunos de estos algoritmos y aplicaciones se encuentran descritos en la figura 3. Como se observa, los algoritmos de ML tienen tres vertientes principales: el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje por refuerzo.

En el aprendizaje supervisado, los algoritmos “aprenden” de los datos introducidos y etiquetados (clasificados) con ayuda de intervención humana. Existen dos tipos principales de algoritmos de aprendizaje supervisado:

- Clasificación: como su nombre lo indica, estos algoritmos clasifican un objeto dentro de dos o más clases (por ejemplo: activo e inactivo; tóxico o no tóxico). Un ejemplo que emplea algoritmos de clasificación es el programa de predicción de actividades biológicas PASS (*Prediction of Activity Spectra for Biologically Active Substances*) (Lagunin et al., 2000). Las predicciones de actividad basada en la estructura de nuevos compuestos que realiza PASS, se basan en un clasificador (Naive Bayes) entrenado con datos de más de 30000 compuestos biológicamente activos.

- Regresión: generalmente asociados a la predicción de un valor numérico. En química farmacéutica, los modelos QSAR son el ejemplo más representativo. Estos modelos permiten relacionar las propiedades fisicoquímicas y moleculares (en general, descriptores) de compuestos con una actividad biológica determinada *in silico* o experimentalmente, para posteriormente predecir valores como la concentración inhibitoria media (IC_{50}), constante de inhibición (K_i) y dosis efectiva media (DE_{50}) de nuevos compuestos.

A diferencia del caso anterior, los métodos no supervisados se emplean con fines exploratorios para desarrollar modelos que permitan agrupar los datos de una forma no especificada por el usuario (aquí no existe la intervención humana para etiquetar los datos). Existen tres tipos de algoritmos para ML no supervisado:

- Agrupamiento (*clustering*): clasifican en grupos los datos de salida. Estos algoritmos suelen usarse para identificar características comunes de una clase particular de compuestos. Una aplicación sencilla es la selección de compuestos en estudios de cribado virtual, en donde los compuestos se agrupan de acuerdo con propiedades o resultados del acoplamiento molecular similares a un compuesto de consulta (generalmente un compuesto activo y validado experimentalmente). Otra aplicación se encuentra en el perfilado basado en imágenes, en donde la basta cantidad de datos presentes en las imágenes biológicas se representa como un perfil multidimensional, del cual pueden extraerse patrones relevantes asociados a enfermedades, a la comprensión de los mecanismos de la enfermedad y a la predicción de la actividad, la toxicidad o el mecanismo de acción de un fármaco (Chandrasekaran et al., 2021).
- Asociación: determinan reglas de asociación dentro de conjuntos de datos. Los algoritmos de asociación son capaces de reconocer patrones ocultos en moléculas que escapan de modelos tradicionales y la intuición de la química farmacéutica debido al gran conjunto de variables y parámetros que deben considerarse en paralelo para el diseño de fármacos. Con estos algoritmos es posible identificar asociaciones objetivo-objetivo, ligando-objetivo y ligando-efecto biológico, lo cual puede ayudar en la búsqueda de eventos adversos (Noguchi et al., 2018) y para el reposicionamiento de fármacos (Masoudi-Sobhanzadeh y Masoudi-Nejad, 2020).
- Reducción de dimensiones: usados para la visualización de datos. Los algoritmos de reducción de dimensiones mapean conjuntos de datos a subespacios de menor dimensión derivados del espacio original, generalmente a espacios de 2 o 3 dimensiones, que son los que puede interpretar el ojo humano. La posibilidad de integrar más elementos visuales a estas representaciones facilita las interpretaciones y teorías sobre los datos que obtenemos y analizamos. Por ejemplo, es posible mapear propiedades de interés como la actividad biológica mediante una escala de color continua, o bien, variando el tamaño de los puntos o usando formas para variables categóricas. En el diseño de fármacos, se utilizan para la visualización del espacio químico, que a su vez permite evaluar la diversidad de diferentes conjuntos de datos, explorar las relaciones estructura-propiedad (actividad biológica) y en el diseño de bibliotecas químicas (Saldivar-González y Medina-Franco, 2022).

En el aprendizaje por refuerzo, los algoritmos aprenden por sí solos el comportamiento a seguir mediante la observación del estado de su entorno y mediante su interacción con una serie de acciones por las cuales recibe recompensas y penalizaciones. Como resultado, una secuencia de decisiones exitosas hará que el proceso se "refuerce" porque resuelve mejor el problema en cuestión. Algunos enfoques de diseño *de novo* están utilizando algoritmos de aprendizaje por refuerzo para obtener compuestos dirigidos que sean eficaces y seguros (ver Sección Diseño de novo de nuevos compuestos) (Popova et al., 2018). Otra aplicación se encuentra en la farmacología de sistemas, enfoque potencialmente efectivo para diseñar terapias personalizadas para enfermedades complejas e intratables a la fecha (Tan et al., 2022).

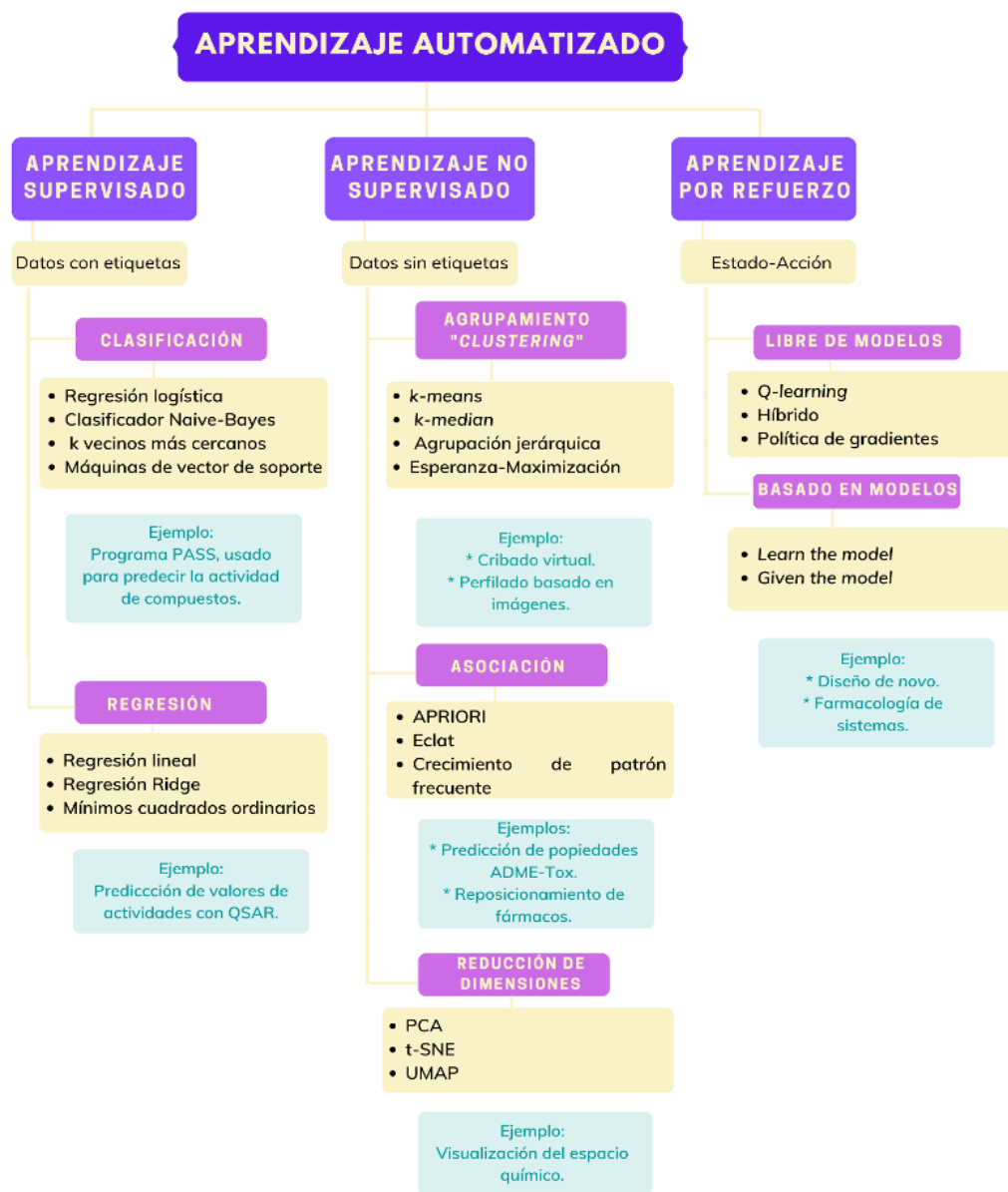


FIGURA 3. Algoritmos utilizados en aprendizaje de máquinas con aplicaciones en diseño de fármacos. Abreviaciones: ADME-Tox, absorción, distribución, metabolismo, excreción y toxicidad; Eclat, del inglés *equivalence class clustering and bottom-up lattice traversal*; PASS, del inglés *Prediction of Activity Spectra for Biologically Active Substances*; PCA, análisis de componentes principales; QSAR: estudios cuantitativos de relaciones estructura-actividad; t-SNE, incrustación de vecinos estocásticos distribuidos en t; UMAP, aproximación y proyección múltiple uniforme.

Redes neuronales y aprendizaje profundo en el diseño de fármacos

El DL es un método específico de ML que incorpora redes neuronales para aprender de los datos de manera iterativa a partir de una combinación de algoritmos supervisados y no supervisados. Si bien, existe cierta controversia sobre si los modelos de DL son superiores en cuanto a rendimiento a enfoques de ML, DL ofrece varias ventajas (Chen et al., 2018). Por ejemplo, los modelos de DL permiten la extracción automática de características durante el procedimiento de entrenamiento. En este proceso se generan representaciones comprimidas de estructuras moleculares (que únicamente son interpretables por la computadora). Esta compresión de las estructuras es prometedora para trabajar con compuestos de mayor tamaño y complejidad como péptidos y microciclos (moléculas que contienen un anillo con nueve o más átomos)(David et al., 2020), y para el aprendizaje multitarea (Ramsundar et al., 2015). En principio, estas representaciones comprimidas, que no podemos interpretar directamente, pero sí interconvertir después a formatos legibles, podrían codificar información relacionada con varios parámetros en paralelo, como: selectividad, absorción, estabilidad metabólica, toxicidad y síntesis, para optimizar el diseño de nuevos compuestos.

En diseño de fármacos, los algoritmos de DL también se han utilizado en tareas como el diseño *de novo* (sección 3.2), para obtener información contenida en estudios de relaciones estructura-actividad, en cribado virtual y en la predicción de propiedades farmacocinéticas, por mencionar algunos ejemplos (Gawehn et al., 2016). Dentro de las limitaciones de DL, que se discuten en la sección 5, se encuentran su baja interpretabilidad y bajo rendimiento en escenarios de pocos datos (Lenselink et al., 2017).

Aplicaciones de IA en diseño de fármacos

En esta sección abordaremos ejemplos sobresalientes del uso de métodos de IA en el proceso de descubrimiento y desarrollo de nuevos compuestos como candidatos a fármacos. Mencionaremos aplicaciones en la predicción de estructuras 3D de proteínas, en el diseño de compuestos, en la evaluación sistemática de objetivos o blancos moleculares, en la predicción de propiedades ADME-Tox y en la síntesis de nuevos compuestos.

Predicción de estructura 3D de blancos moleculares terapéuticos

Una de las estrategias del diseño de fármacos que ha sido exitosa, es el diseño basado en la estructura del blanco terapéutico (como receptores de membrana, incluyendo los receptores acoplados a proteínas G, enzimas, ácidos nucleicos, etc.). Ejemplos notables de diseño de fármacos basados en el conocimiento de la estructura —o modelos de ella— son antivirales como el relenza y fármacos empleados para el tratamiento del síndrome de inmunodeficiencia adquirida (SIDA) (Kumar et al., 2014). Técnicas como la cristalografía de rayos X, resonancia magnética nuclear y microscopía electrónica, están haciendo contribuciones fundamentales para elucidar la estructura 3D de blancos moleculares. Esto, a su vez, aporta información invaluable para dirigir el diseño de compuestos que interactúan de manera favorable con estas estructuras. La hipótesis de trabajo es que compuestos con afinidad hacia sitios funcionales (clave) de estas estructuras, potencialmente pueden inducir o traducirse en un efecto clínico con utilidad terapéutica. A su vez, las estructuras

3D se pueden emplear en el diseño asistido por computadora, utilizando técnicas como acoplamiento molecular, diseño basado en fragmentos y diseño *de novo* (Batoool et al., 2019).

A pesar de que a la fecha de escritura de este manuscrito (julio, 2022) están disponibles en *Protein Data Bank* (Berman et al., 2000) (el Banco de Datos de Proteínas más grande de dominio público) las coordenadas 3D de más de 192,000 estructuras, incluyendo más de 78,000 proteínas, aún resulta tardado y costoso elucidar experimentalmente las estructuras de todos los blancos terapéuticos. En este sentido, la IA está teniendo avances sorprendentemente útiles para predecir, con gran precisión, el plegamiento de las proteínas a partir del conocimiento de su secuencia. En 2021 se publicó el primer método computacional capaz de predecir estructuras de proteínas con precisión atómica, incluso en casos en los que no se conoce una estructura similar. La red neuronal, llamada **AlphaFold**, analiza la distancia entre los aminoácidos adyacentes y los ángulos correspondientes de los enlaces peptídicos para predecir con precisión la estructura 3D de proteínas (Jumper et al., 2021). AlphaFold ha generado la imagen más completa y precisa del proteoma humano (Tunyasuvunakool et al., 2021) y ya ha demostrado su utilidad a la comunidad experimental. Algunos investigadores han usado AlphaFold para dar sentido a los datos experimentales que se obtienen mediante la cristalografía de rayos X y la microscopía crioelectrónica (M. Gupta et al., 2021). Se espera que su uso acelere la investigación de fármacos basada en la estructura y genere nuevas vías de conocimiento científico.

Diseño de novo de nuevos compuestos

El diseño *de novo*, es decir, el diseño “desde cero”, no es un concepto que únicamente se aplique en la química. En 2018, un retrato hecho por un algoritmo de IA abrió nuevas interrogantes respecto a la creación y concepción de lo que consideramos arte. La obra Edmond de Belamy generó numerosas críticas que cuestionan la capacidad creativa (o imitativa) de los modelos de IA, que incluso pueden trasladarse a la Química Farmacéutica (Epstein et al., 2020). Otro ejemplo notable de los modelos generativos se observa en la creación de obras literarias, en donde se han generado desde poemas hasta libros de campos expertos del área de investigación (Petsko, 2019). En el diseño de fármacos, los modelos generativos se utilizan principalmente para diseñar nuevos compuestos químicos (Bajorath, 2022). La figura 4 resume algunas arquitecturas de DL empleadas en el diseño *de novo*, que son: Autocodificadores Variacionales (VAE, del inglés, *Variational Autoencoders*) (4a), Redes Generativas Antagónicas (GAN, *Generative Adversarial Networks*) (4b) y Redes Neuronales Recurrentes (RNN, *Recurrent Neural Network*) (4c). Adicionalmente, pueden diseñarse enfoques híbridos que usan una de las arquitecturas mencionadas anteriormente en conjunto con algoritmos de aprendizaje por refuerzo. Un ejemplo de enfoque híbrido es GENTRL (Zhavoronkov et al., 2019), el modelo desarrollado por Insilico Medicine para generar inhibidores de cinasas (discutido en la Introducción) (4d). GENTRL usa VAE que están compuestas de tres elementos: un codificador, un decodificador y una función de pérdida. El codificador toma el dato de entrada y aprende a representarlos de manera compacta, mientras que el decodificador toma esa versión comprimida y reconstruye el dato original a la salida. La función de pérdida se aplica al momento de obtener la representación compacta, también conocida como “espacio latente”, y consiste en mantener este espacio dentro de una distribución normal para así minimizar la diferencia entre la imagen original y la imagen reconstruida. Al momento de generar el espacio latente

también se agrega un componente de ruido que permite agregar aleatoriedad a las VAE con el fin de que generen moléculas novedosas (diferentes a las usadas en el entrenamiento).

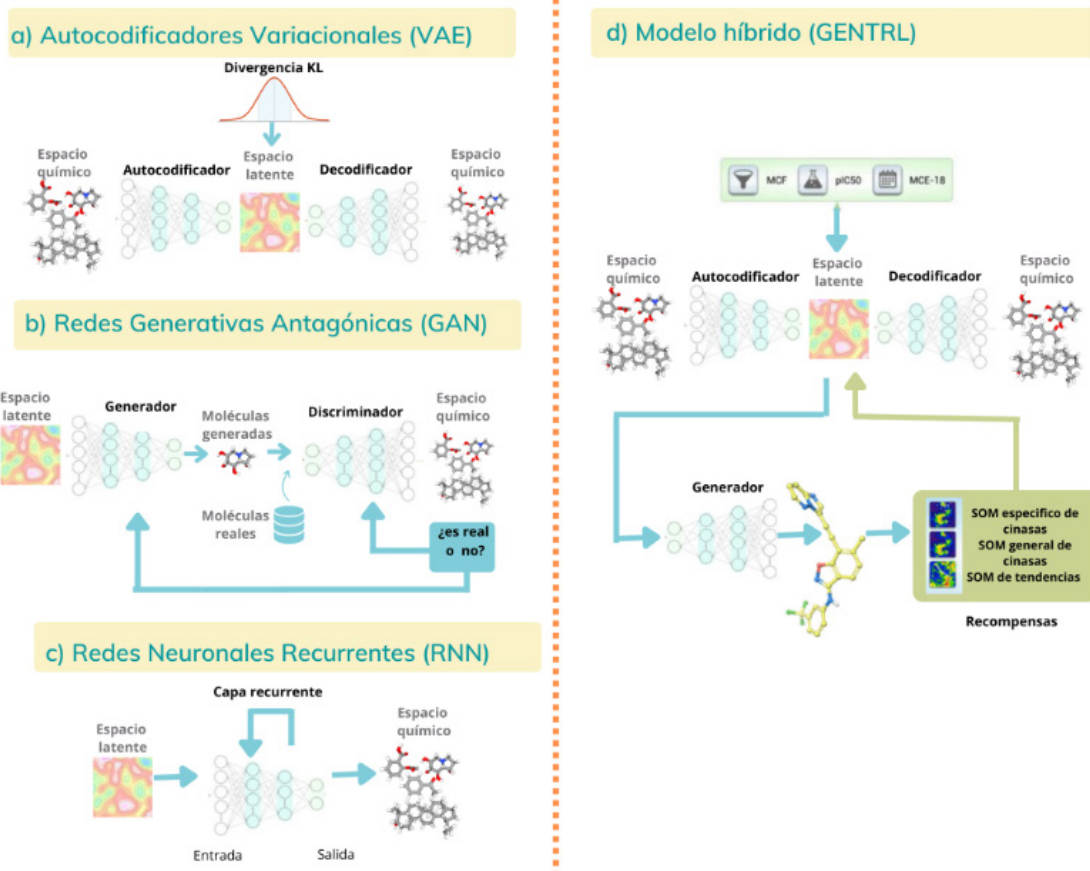


FIGURA 4. Arquitecturas de aprendizaje profundo para el diseño *de novo*.

a) Autocodificadores variacionales (VAE),
b) Redes Generativas Antagónicas (GAN);
c) Redes Neuronales Recurrentes (RNN);
d) Modelo híbrido GENTRL, usado para descubrir inhibidores del receptor DDR1, adaptado de (Zhavoronkov et al., 2019).

Adicional a esto, se puede agregar otra red neuronal, que se encarga de optimizar el proceso, garantizando que las moléculas generadas no solo sean viables estructuralmente, sino que además tengan el potencial de inhibir o modular las proteínas de interés. Para el caso de GENTRL, se utilizó aprendizaje por refuerzo. De esta forma, el espacio latente generado por las VAE tiene características similares a las de moléculas que ya se usan para el tratamiento de la fibrosis pulmonar. La red neuronal genera una señal de recompensa que indica a las VAE que el código generado resulta adecuado. De lo contrario, genera una señal de castigo que les enseña a no generar este tipo de moléculas.

Las GAN, por su parte, están integradas por dos redes neuronales que compiten entre sí: un generador y un discriminador. El objetivo del generador es fabricar artificialmente salidas que sean indistinguibles para el discriminador de los datos reales. Mientras que el objetivo del discriminador es diferenciar ejemplos reales de aquellos que fueron producidos por el generador adversario. Este tipo de arquitectura ha sido muy utilizada en la generación de imágenes y obras de arte e incluso se ha alertado de aplicaciones de software para la manipulación mal intencionada de imágenes (Wang et al., 2022). En el diseño de fármacos, las GAN han permitido generar moléculas a partir de la expresión génica inducida por compuestos (Mendez-Lucio et al., 2021) y se explora su aplicación para

encontrar compuestos potencialmente activos en campañas de alto rendimiento, en donde equipos automatizados analizan rápidamente miles o millones de muestras en busca de compuestos activos en modelos con tejidos, órganos o células aisladas.

Las RNN, se diferencian por analizar datos secuenciales como texto, voz, audio, video, entre otros. En el diseño de nuevos compuestos, las RNN extraen la información automáticamente a nivel gramatical y semántico a partir de representaciones moleculares en formato de cadenas de texto, como son los SMILES (*Simplified Molecular Input Line Entry System*) (David et al., 2020). Cuando se utilizan RNN para generar moléculas, se seleccionan las mejores moléculas y se utilizan para volver a entrenar la red a través del aprendizaje por transferencia. El aprendizaje por transferencia permite que el conocimiento se transfiera entre tareas y ha demostrado ser una forma eficiente de mejorar la precisión de los modelos y en este caso, de generar una mayor proporción de moléculas químicamente válidas (A. Gupta et al., 2018).

El diseño generativo ha logrado avances considerables en un período de tiempo relativamente corto, sin embargo, aún se requiere generar más evidencia de su eficiencia con respecto a casos de prueba validados experimentalmente (Schneider y Clark, 2019). Los desafíos actuales de los enfoques generativos incluyen la síntesis (la mayoría de los modelos generativos no brindan información sobre cómo se puede sintetizar una molécula) y la existencia de múltiples soluciones para cualquier propiedad dada (Jiménez-Luna et al., 2021).

Evaluación sistemática de objetivos o blancos moleculares

Comprender el mecanismo por el cual un compuesto ejerce efecto terapéutico es esencial para el diseño y optimización de fármacos. Actualmente, la identificación de objetivos biológicos para compuestos bioactivos se considera un cuello de botella para el diseño informado de “sondas químicas” (moléculas que ayudan a elucidar la función biológica de proteínas) y fármacos líderes (Rodrigues y Bernardes, 2020).

Afortunadamente, la experimentación de alto rendimiento ha generado una gran cantidad de datos químicos y biológicos que permiten producir modelos para vincular objetivos biológicos con compuestos bioactivos. Esta información también está generando evidencia creciente de que los compuestos rara vez son selectivos para un solo objetivo biológico dado. La interacción de un compuesto que involucra a varias macromoléculas relacionadas o no relacionadas es conocida como polifarmacología, y está siendo aprovechada en enfoques racionales para diseñar fármacos multi-objetivo y en el reposicionamiento de fármacos (Reddy y Zhang, 2013).

Los crecientes esfuerzos computacionales para la identificación de objetivos incluyen el desarrollo de modelos de ML y servidores web para analizar el potencial medicinal de compuestos anotados en las bases de datos químicas públicas (Rodrigues y Bernardes, 2020). La predicción que hacen los modelos se basan en el principio de similitud química, el cual sugiere que moléculas con estructuras químicas semejantes tienen propiedades semejantes y por ende se unen a objetivos biológicos similares (Maggiora et al., 2014); y/o en los principios de la biofísica de las interacciones moleculares, cuanto mayor sea la complementariedad entre ligando y receptor, mayor será la afinidad y mayor será su efecto en la regulación de la actividad biológica.

Pocas de estas herramientas incluyen estructuras complejas como productos naturales y péptidos en sus conjuntos de entrenamiento, por lo que podrían no predecir correctamente los objetivos biológicos de estos compuestos (Saldivar-González et al., 2022), ni podrían descubrir una nueva biología como si lo hacen experimentos de laboratorio. No obstante, representan una oportunidad única para analizar el potencial medicinal de la gran cantidad de compuestos presentes en repositorios públicos.

Predicción de propiedades ADME-Tox de moléculas pequeñas

Como vimos en secciones anteriores, la cantidad de moléculas orgánicas que potencialmente podrían sintetizarse en un programa de descubrimiento de fármacos es muy grande, aproximadamente de 10^{63} (Walters, 2019). Esta es una cifra incluso mayor que el número de estrellas en el cielo que se conocen a la fecha. Para limitar el número de moléculas a considerar en el mundo real, han surgido aplicaciones de IA para predecir las propiedades ADME-Tox de una molécula antes de su síntesis. Propiedades como la solubilidad acuosa, la permeabilidad intestinal y la unión a proteínas plasmáticas son parámetros importantes que considerar para conseguir la exposición sistémica de un fármaco oral. De igual forma, conocer con qué rapidez se metaboliza y se elimina un fármaco, cuál es el mecanismo de acción y qué metabolitos forma, determinan el éxito de un medicamento, ya que estas propiedades tienen un impacto significativo sobre la biodisponibilidad y seguridad de un fármaco. La toxicidad también es un parámetro crítico que evaluar en el diseño de fármacos y en industrias como la agroquímica y la cosmética (Roncaglioni et al., 2022). La cantidad de causas posibles para que un compuesto falle y tenga efectos adversos es grande y, como tal, la cantidad de propiedades para verificar en un laboratorio puede volverse muy costosa y consumir mucho tiempo. Por este motivo, se han hecho esfuerzos considerables para desarrollar modelos computacionales que permitan predecir conjuntos heterogéneos de propiedades ADME-Tox (Butina et al., 2002). En este contexto, los modelos de IA han ayudado a aumentar la precisión de las predicciones sobre la eficacia y seguridad de los medicamentos (Maltarollo et al., 2015).

Actualmente, existen repositorios públicos que están haciendo más accesible el uso y la comparación de modelos de ML para la predicción de propiedades ADME-Tox (Bhatarai et al., 2019; Roncaglioni et al., 2022).

Síntesis química

Las técnicas de IA también se están utilizando para analizar información de síntesis química disponible en la literatura y bases de datos. Esto, con la finalidad de establecer diversas mejoras en el proceso de diseño de fármacos, que se van desde las primeras etapas, centrándose en los ciclos de diseño, síntesis, prueba y análisis, hasta los procesos de fabricación a gran escala (Miljković et al., 2021). Por ejemplo, se pueden comparar y procesar diferentes reacciones para producir un compuesto determinado. Es importante que, además de los reactivos involucrados, se consideren otras variables como solventes (si es el caso), temperatura, tiempo, presión y, en general, las condiciones de reacción. Esta información puede derivar en: a) análisis retrosintéticos que dividen el problema en subproblemas y generan sugerencias de síntesis paso a paso, b) en la recomendación de las condiciones de reacción que conducirán a una reacción directa exitosa y c) en la

predicción de productos a partir de un conjunto de materiales y condiciones iniciales que se usa para validar los pasos sintéticos propuestos (Struble et al., 2020). Es fundamental mencionar que, para que una herramienta de planificación sintética verdaderamente ayude a aumentar el conocimiento químico y acelere el proceso de síntesis, debe ser fácil de usar e intuitiva en su manejo e interpretación de resultados. Además, las rutas sugeridas deben derivar en materiales que puedan adquirirse en el mercado.

Otras aplicaciones de la IA en el proceso de síntesis incluyen la predicción de la accesibilidad sintética y sus costos relacionados (Thakkar et al., 2021). Una estimación de la accesibilidad sintética, o bien, ejecutar filtros relacionados con el coste de los reactivos, en principio, podría ayudar a filtrar más las bases de datos iniciales y/o priorizar compuestos.

Aunque ya existen sintetizadores automáticos como AutoSyn (Collins et al., 2020) y RoboRXN (Vaucher et al., 2020), se espera que en un futuro la automatización de la síntesis química tenga un papel protagónico en el proceso de descubrimiento de fármacos. Incluir esta tecnología puede ser un factor vital para aumentar la productividad de los proyectos de descubrimiento de fármacos, garantizar la reproducibilidad y la transferibilidad de los protocolos de síntesis de un laboratorio a otro y podrían hacer que la planeación y la síntesis de moléculas complejas sea cada vez más rutinaria.

Medicina personalizada: el futuro de la implementación de IA en el uso terapéutico de fármacos

En el diseño y desarrollo de fármacos clásico (ya sea dirigido a una o varios blancos terapéuticos) se asume que la respuesta clínica de los pacientes es aproximadamente la misma. Esto incluye el efecto terapéutico deseado, así como los efectos adversos que se quieren disminuir (discutidos en la sección 3.4). Sin embargo, es bien sabido que la respuesta a un tratamiento clínico puede variar, a veces en forma dramática, de paciente a paciente. Esta variabilidad entre la población está asociada a muchos efectos como factores genéticos, y epigenéticos. Todos estos factores son difíciles de considerar en forma sistemática desde el inicio en un programa de diseño de fármacos y, por tanto, dificulta aún más el uso terapéutico exitoso de los fármacos, especialmente en el caso de enfermedades complejas como cáncer y enfermedades neurodegenerativas. I-PREDICT (Sicklick et al., 2019) es un ejemplo de un estudio diseñado para optimizar ensayos clínicos que brindó tratamientos individualizados a pacientes con cáncer mediante la secuenciación del ADN tumoral y recomendaciones oportunas para el tratamiento individualizado con terapias combinadas. En este contexto, la adopción de IA y otras tecnologías innovadoras, y el uso de la gran información disponible en múltiples fuentes (e.g., farmacovigilancia), está permitiendo tratamientos específicos, más precisos y cambiando el sistema de salud hacia un futuro donde la medicina es personalizada, predictiva, preventiva y participativa (las '4P') (K. B. Johnson et al., 2021). Se espera que, durante las próximas décadas, la inclusión de datos más diversos que incluyan poblaciones poco estudiadas, la implementación de estrategias para minimizar la exclusión y el sesgo, y las nuevas tecnologías de IA tengan un impacto significativo en los tratamientos y en los resultados de los pacientes, particularmente en aquellos en áreas de la medicina con necesidades insatisfechas.

Consideraciones y recomendaciones de aspectos generales críticos de IA

TABLA 1.
 Consideraciones y recomendaciones para atender a puntos críticos generales en IA en el diseño de fármacos.

La Tabla 1 resume aspectos generales que, en la opinión de los autores, se deben considerar en IA en el contexto de investigación, diseño, desarrollo y uso terapéutico de fármacos. La lista de tópicos no es exhaustiva y remarca algunos puntos especialmente relevantes a tomar en cuenta cuando se aplican métodos de IA.

Aspectos críticos	Consideraciones y recomendaciones
Acceso a datos robustos y confiables	<ul style="list-style-type: none"> Fortalecer y mantener colaboraciones, consorcios y asociaciones entre grupos de investigación. La anotación de datos es inconsistente. Se requiere mejorar la cantidad y calidad de los datos de entrada para obtener modelos más confiables (Rodrigues, 2019). También se ha hecho énfasis en el uso de modelos de IA centrados en datos, en donde se busca diseñar sistemáticamente los datos necesarios para construir con éxito un sistema de IA (ver iniciativa data-centric IA). En otras palabras, se sugiere cambiar del paradigma de los datos a gran escala al de los buenos datos. Los datos utilizados para establecer estándares en proyectos de IA no consideran la diversidad poblacional, ni incluyen suficientes minorías. Se requiere promover y demostrar transparencia, ofreciendo información sobre la recolección (grupos y sectores bajo estudio) y sobre el uso de datos (compañías que financian los proyectos de IA).
Predicciones y usos “éticos” ^a de la información generada con IA	<ul style="list-style-type: none"> Balancear el dominio de aplicación con la confiabilidad de los mismos datos y las predicciones. Contextualizar las predicciones en función del posible sesgo en los datos y enfatizar en la incertidumbre que tienen los resultados. Utilizar los algoritmos de inteligencia artificial para predecir efectos deseables y benéficos a la sociedad (Urbina et al., 2022). La toma de decisiones por medio de algoritmos es susceptible a sesgos controlados por humanos y por datos. Se necesita fortalecer iniciativas como diversity.ai y generar un análisis algorítmico profundo para detectar, analizar, prevenir y combatir los prejuicios humanos y la discriminación racial, de género, de edad y étnicos por parte de los propios sistemas de aprendizaje profundo (Echaiz et al., 2021). Generar guías sobre el uso ético del aprendizaje automático en el espacio de la química, por ejemplo, en el manejo de datos relacionados con toxicidad (Blum, 2022). Fortalecer la formación ética de los estudiantes de ciencia e informática (Durrani, 2022; Wang et al., 2022).
Evaluación de modelos de IA	<ul style="list-style-type: none"> Establecer indicadores para evaluar el rendimiento de los enfoques de IA (Bender et al., 2022). Estandarizar indicadores para evaluar los descubrimientos valiosos (Walters y Barzilay, 2021).
Interpretabilidad de modelos de IA	<ul style="list-style-type: none"> Se debe proporcionar interpretaciones comprensibles del proceso de toma de decisiones realizado por enfoques de aprendizaje profundo (Jiménez-Luna et al., 2020). Para las aplicaciones en las que se requiere una comprensión mecanicista (por ejemplo, estudios QSAR), es esencial que los modelos de ML complejos ofrezcan una ventaja de rendimiento sobre los modelos más simples e interpretables (Rodríguez-Pérez y Bajorath, 2021).
Propiedad intelectual (PI)	<ul style="list-style-type: none"> La IA puede incentivar la innovación en diseño de fármacos, no obstante, aún no hay guías de PI para brindar una protección equilibrada para las obras e invenciones creadas por algoritmos computacionales (Griffen et al., 2018; Yanisky-Ravid y Jin, 2020).
Diversificación en las aplicaciones del diseño de fármacos	<ul style="list-style-type: none"> La IA puede proporcionar mejores modelos predictivos e investigaciones más ágiles para comprender los mecanismos de las enfermedades. En el caso de enfermedades raras y desatendidas, cuya fisiopatología a menudo está mal caracterizada, estas técnicas ofrecen una forma rápida de identificar hipótesis comprobables que pueden traducirse en la clínica.
Hacia la automatización (absoluta) del proceso de diseño de fármacos	<ul style="list-style-type: none"> Aunque hay una gran tendencia para automatizar (casi en su totalidad) el diseño de compuestos que son candidatos a fármaco, hay que considerar que no hay aprendizaje sin la reflexión y realimentación que aportan los seres Humanos.

^a Se ha cuestionado si las predicciones de modelos de IA pueden ser éticas o no. La ética es un factor humano que está a expensas de la subjetividad (e.g., moral, sesgos culturales, económicos y sociales) que pueden tener los datos (por ejemplo, iniciativa de [Beca de Ética GSK.ai-Stanford](#)).

Conclusiones

En el descubrimiento y desarrollo moderno de fármacos se utiliza una combinación de enfoques computacionales, experimentales y clínicos para identificar nuevos candidatos a fármacos. Hasta el momento, no existe una solución única que garantice un descubrimiento exitoso de fármacos, por lo tanto, es importante permanecer receptivo a las nuevas tecnologías. La llamada “inteligencia aumentada” (IA + IH), tiene aplicaciones en el proceso de descubrimiento y desarrollo de fármacos, en la toma de decisiones, y también en el desarrollo de terapias más efectivas a partir de fármacos ya aprobados (medicina de precisión o personalizada). Aún quedan retos que enfrentar en la implementación de la IA en el proceso de descubrimiento, diseño y uso terapéutico de fármacos, como la falta de estándares en los datos generados y en las bases de datos disponibles, y la falta de métodos robustos y estandarizados de validación. La variabilidad en la respuesta entre los pacientes a un mismo tratamiento terapéutico, en especial en enfermedades complejas, impone otro reto. Retos adicionales por abordar son la falta de transparencia y colaboración de datos (que es un balance o disyuntiva entre la ciencia abierta y la propiedad intelectual); la falta de personal entrenado interdisciplinariamente y la comunicación humana eficiente entre los grupos o equipos de investigación multidisciplinarios. No obstante, los métodos de IA ofrecen una oportunidad para diversificar aplicaciones en diseño de fármacos, proporcionar mejores modelos predictivos, agilizar las investigaciones y ayudar a proponer tratamientos efectivos en forma personalizada. La validación experimental y adecuada de los resultados de IA que hay hasta ahora permitirá retroalimentar y mejorar los modelos existentes y establecerá las bases a nuevos paradigmas.

Se espera que la integración adecuada de métodos básicos y aquellos más sofisticados de bioinformática, computación, nanotecnología y farmacogenómica conduzca a la siguiente etapa de avances en el descubrimiento y uso terapéutico exitoso de fármacos.

Agradecimientos

FISG agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca otorgada, no 848061.

Referencias

- My Intelligent Machine. (2021). Augmented intelligence at the service of personalised medicine: how artificial intelligence is accelerating drug discovery and development. *Springer Nature*, consultada en julio 15, 2022. <https://www.nature.com/articles/d43747-021-00043-9>
- Bajorath, J. (2022). Deep machine learning for computer-aided drug design. *Frontiers in Drug Discovery*, 2. 829043.
- Batool, M., Ahmad, B., y Choi, S. (2019). A Structure-Based Drug Discovery Paradigm. *International Journal of Molecular Sciences*, 20(11),2783.
- Bazoukis, G., Hall, J., Loscalzo, J., Antman, E. M., Fuster, V., y Armoundas, A. A. (2022). The inclusion of augmented intelligence in medicine: A framework for successful implementation. *Cell Reports. Medicine*, 3(1), 100485.

- Bender, A., Schneider, N., Segler, M., Patrick Walters, W., Engkvist, O., y Rodrigues, T. (2022). Evaluation guidelines for machine learning tools in the chemical sciences. *Nature Reviews Chemistry*, 6(6), 428–442.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., y Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242.
- Bhatarai, B., Walters, W. P., Hop, C. E. C. A., Lanza, G., y Ekins, S. (2019). Opportunities and challenges using artificial intelligence in ADME/Tox. *Nature Materials*, 18(5), 418–422.
- Blum, M.-M. (2022). No chemical killer AI (yet). *Nature Machine Intelligence*, 4(6), 506–507.
- Butina, D., Segall, M. D., y Frankcombe, K. (2002). Predicting ADME properties in silico: methods and models. *Drug Discovery Today*, 7(11), S83–S88.
- Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D., y Carpenter, A. E. (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nature Reviews. Drug Discovery*, 20(2), 145–159.
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., y Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today*, 23(6), 1241–1250.
- Collins, N., Stout, D., Lim, J.-P., Malerich, J. P., White, J. D., Madrid, P. B., Latendresse, M., Krieger, D., Szeto, J., Vu, V.-A., Rucker, K., Deleo, M., Gorf, Y., Krummenacker, M., Hokama, L. A., Karp, P., y Mallya, S. (2020). Fully Automated Chemical Synthesis: Toward the Universal Synthesizer. *Organic Process Research y Development*, 24(10), 2064–2077.
- David, L., Thakkar, A., Mercado, R., y Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, 12(1), 56.
- Durrani, J. (2022). Drug discovery AI that developed new nerve agents raises difficult questions. *Chemistry World*, consultada en julio 15, 2022. <https://bit.ly/40jQdYr>
- Epstein, Z., Levine, S., Rand, D. G., y Rahwan, I. (2020). Who Gets Credit for AI-Generated Art? *iScience*, 23(9), 101515.
- Echaiz, L. F., Neupane, B., Rivera Lam, M., Sibal, P., UNESCO, y Hu, X. (2021). *El Aporte de la Inteligencia Artificial y las TIC Avanzadas a las Sociedades del Conocimiento: Una Perspectiva de Derechos, Apertura, Acceso y Múltiples Actores*. UNESCO Publishing.
- Gartner glossary. (s.f.). Augmented intelligence. En *Gartner glossary*, consultada en julio 15, 2022. <https://gtnr.it/3nmf7YW>
- Gasteiger, J. (2016). Chemoinformatics: Achievements and Challenges, a Personal View. *Molecules*, 21(2), 151.
- Gasteiger, J. (2020). Chemistry in Times of Artificial Intelligence. *Chemphyschem: A European Journal of Chemical Physics and Physical Chemistry*, 21(20), 2233–2242.
- Gawehn, E., Hiss, J. A., y Schneider, G. (2016). Deep Learning in Drug Discovery. *Molecular Informatics*, 35(1), 3–14.

- Griffen, E. J., Dossetter, A. G., Leach, A. G., y Montague, S. (2018). Can we accelerate medicinal chemistry by augmenting the chemist with Big Data and artificial intelligence? *Drug Discovery Today*, 23(7), 1373–1384.
- Gupta, A., Müller, A. T., Huisman, B. J. H., Fuchs, J. A., Schneider, P., y Schneider, G. (2018). Generative Recurrent Networks for De Novo Drug Design. *Molecular Informatics*, 37(1-2), 1700111.
- Gupta, M., Azumaya, C. M., Moritz, M., Pourmal, S., Diallo, A., Merz, G. E., Jang, G., Bouhaddou, M., Fossati, A., Brilot, A. F., Diwanji, D., Hernandez, E., Herrera, N., Kratochvil, H. T., Lam, V. L., Li, F., Li, Y., Nguyen, H. C., Nowotny, C., ... Verba, K. A. (2021). CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. *bioRxiv*. <https://doi.org/10.1101/2021.05.10.443524>
- Hadley, C. (2004). Biologists think bigger. *EMBO Reports*, 5(3), 236–238.
- Hansch, C., y Fujita, T. (1964). P- σ - π analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, 86(8), 1616–1626.
- Jiménez-Luna, J., Grisoni, F., y Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10), 573–584.
- Jiménez-Luna, J., Grisoni, F., Weskamp, N., y Schneider, G. (2021). Artificial intelligence in drug discovery: recent advances and future perspectives. *Expert Opinion on Drug Discovery*, 16(9), 949–959.
- Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., y Snowdon, J. L. (2021). Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical and Translational Science*, 14(1), 86–93.
- Johnson, S. R. (2008). The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *Journal of Chemical Information and Modeling*, 48(1), 25–26.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.
- Kolluri, S., Lin, J., Liu, R., Zhang, Y., y Zhang, W. (2022). Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review. *The AAPS Journal*, 24(1), 19.
- Kumar, V., Chandra, S., y Siddiqi, M. I. (2014). Recent advances in the development of antiviral agents using computer-aided structure based approaches. *Current Pharmaceutical Design*, 20(21), 3488–3499.
- Lagunin, A., Stepanchikova, A., Filimonov, D., y Poroikov, V. (2000). PASS: prediction of activity spectra for biologically active substances. *Bioinformatics*, 16(8), 747–748.
- Lamberti, M. J., Wilkinson, M., Donzanti, B. A., Wohlhieter, G. E., Parikh, S., Wilkins, R. G., y Getz, K. (2019). A Study on the Application and Use of Artificial Intelligence to Support Drug Development. *Clinical Therapeutics*, 41(8), 1414–1426.

- Lenselink, E. B., ten Dijke, N., Bongers, B., Papadatos, G., van Vlijmen, H. W. T., Kowalczyk, W., IJzerman, A. P., y van Westen, G. J. P. (2017). Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. In *Journal of Cheminformatics*, 9,45.
- López-López, E., Bajorath, J., y Medina-Franco, J. L. (2021). Informatics for Chemistry, Biology, and Biomedical Sciences. *Journal of Chemical Information and Modeling*, 61(1), 26–35.
- Maggiora, G., Vogt, M., Stumpfe, D., y Bajorath, J. (2014). Molecular similarity in medicinal chemistry. *Journal of Medicinal Chemistry*, 57(8), 3186–3204.
- Mak, K.-K., y Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, 24(3), 773–780.
- Maltarollo, V. G., Gertrudes, J. C., Oliveira, P. R., y Honorio, K. M. (2015). Applying machine learning techniques for ADME-Tox prediction: a review. *Expert Opinion on Drug Metabolism y Toxicology*, 11(2), 259–271.
- Masoudi-Sobhanzadeh, Y., y Masoudi-Nejad, A. (2020). Synthetic repurposing of drugs against hypertension: a datamining method based on association rules and a novel discrete algorithm. In *BMC Bioinformatics*, 21, 313.
- Méndez-Lucio, O., Baillif, B., Clevert, D.-A., Rouquié, D., y Wichard, J. (2020). De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nature Communications*, 11(1), 10.
- Miljković, F., Rodríguez-Pérez, R., y Bajorath, J. (2021). Impact of Artificial Intelligence on Compound Discovery, Design, and Synthesis. *ACS Omega*, 6 (49). 33293–33299.
- Noguchi, Y., Ueno, A., Otsubo, M., Katsuno, H., Sugita, I., Kanematsu, Y., Yoshida, A., Esaki, H., Tachi, T., y Teramachi, H. (2018). A New Search Method Using Association Rule Mining for Drug-Drug Interaction Based on Spontaneous Report System. *Frontiers in Pharmacology*, 9, 197.
- Olascoaga-Del Angel, K.S., Konigsberg-Fainstein, M., Pérez-Villanueva, J., López Díaz-Guerrero N.E. (2022). Uso de la inteligencia artificial en la investigación para el reposicionamiento de fármacos. En *Revista Especializada en Ciencias Químico-Biológicas*, 25, 1-17.
- Özdemir, V., y Hekim, N. (2018). Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, “The Internet of Things” and Next-Generation Technology Policy. *Omics: A Journal of Integrative Biology*, 22(1), 65–76.
- Petsko, E. (2019, April 19). Springer nature has published the first AI-written textbook. *Mental Floss*. <https://bit.ly/3JM2KNp>
- Popova, M., Isayev, O., y Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science Advances*, 4(7), eaap7885.
- Priya, S., Tripathi, G., Singh, D. B., Jain, P., y Kumar, A. (2022). Machine learning approaches and their applications in drug discovery and design. *Chemical Biology y Drug Design*, 100(1), 136-153.

- Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., y Pande, V. (2015). Massively Multitask Networks for Drug Discovery. *In arXiv [stat.ML]*. <http://arxiv.org/abs/1502.02072>
- Reddy, A. S., y Zhang, S. (2013). Polypharmacology: drug discovery for the future. *Expert Review of Clinical Pharmacology*, 6(1), 41–47.
- Rodrigues, T. (2019). The good, the bad, and the ugly in chemical and biological data for machine learning. *Drug Discovery Today. Technologies*, 32-33, 3–8.
- Rodrigues, T., y Bernardes, G. J. L. (2020). Machine learning for target discovery in drug development. *Current Opinion in Chemical Biology*, 56, 16–22.
- Rodríguez-Pérez, R., y Bajorath, J. (2021). Explainable machine learning for property predictions in compound optimization. *Journal of Medicinal Chemistry*, 64(24), 17744–17752.
- Roncaglioni, A., Lombardo, A., y Benfenati, E. (2022). The VEGAHUB Platform: The Philosophy and the Tools. *Alternatives to Laboratory Animals: ATLA*, 50(2), 121–135.
- Saldívar-González, F. I., Aldas-Bulos, V. D., Medina-Franco, J. L., y Plisson, F. (2022). Natural product drug discovery in the artificial intelligence era. *Chemical Science*, 13(6), 1526–1546.
- Saldívar-González F. I. y Medina-Franco J. L. (2021). Nuevos usos a medicamentos: el reposicionamiento de fármacos y su importancia en crisis sanitarias. *Entorno; UDLAP*, 14, 4-15.
- Saldívar-González, F. I., y Medina-Franco, J. L. (2022). Approaches for enhancing the analysis of chemical space for drug discovery. *Expert Opinion on Drug Discovery*, 17(7), 789-798.
- Schneider, G. (2018). Automating drug discovery. *Nature Reviews. Drug Discovery*, 17(2), 97–113.
- Schneider, G., y Clark, D. E. (2019). Automated DE Novo drug design: Are we nearly there yet? *Angewandte Chemie*, 58(32), 10792–10803.
- Sicklick, J. K., Kato, S., Okamura, R., Schwaederle, M., Hahn, M. E., Williams, C. B., De, P., Krie, A., Piccioni, D. E., Miller, V. A., Ross, J. S., Benson, A., Webster, J., Stephens, P. J., Lee, J. J., Fanta, P. T., Lippman, S. M., Leyland-Jones, B., y Kurzrock, R. (2019). Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nature Medicine*, 25(5), 744–750.
- Struble, T. J., Alvarez, J. C., Brown, S. P., Chytil, M., Cisar, J., Desjarlais, R. L., Engkvist, O., Frank, S. A., Greve, D. R., Griffin, D. J., Hou, X., Johannes, J. W., Kreatsoulas, C., Lahue, B., Mathea, M., Mogk, G., Nicolaou, C. A., Palmer, A. D., Price, D. J., ... Jensen, K. F. (2020). Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *Journal of Medicinal Chemistry*, 63(16), 8667–8682.
- Tan, R. K., Liu, Y., y Xie, L. (2022). Reinforcement Learning for Systems Pharmacology-Oriented and Personalized Drug Design. *Expert Opinion on Drug Discovery*. <https://doi.org/10.1080/17460441.2022.2072288>

- Thakkar, A., Chadimová, V., Bjerrum, E. J., Engkvist, O., y Reymond, J.-L. (2021). Retrosynthetic accessibility score (RAscore) - rapid machine learned synthesizability classification from AI driven retrosynthetic planning. *Chemical Science*, 12(9), 3339–3349.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., ... Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873), 590–596.
- Urbina, F., Lentzos, F., Invernizzi, C., y Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence*, 4(3), 189–191.
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., y Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews. Drug Discovery*, 18(6), 463–477.
- Vaucher, A. C., Zipoli, F., Geluykens, J., Nair, V. H., Schwaller, P., y Laino, T. (2020). Automated extraction of chemical synthesis actions from experimental procedures. *Nature Communications*, 11(1), 3601.
- Walters, W. P. (2019). Virtual chemical libraries. *Journal of Medicinal Chemistry*, 62(3), 1116–1124.
- Walters, W. P., y Barzilay, R. (2021). Critical assessment of AI in drug discovery. *Expert Opinion on Drug Discovery*, 16(9), 937–947.
- Wang, L., Zhou, L., Yang, W., y Yu, R. (2022). Deepfakes: A new threat to image fabrication in scientific publications? *Patterns*, 3(5), 100509.
- Wouters, O. J., McKee, M., y Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA: The Journal of the American Medical Association*, 323(9), 844–853.
- Yanisky-Ravid, S., y Jin, R. (2020). Summoning a New Artificial Intelligence Patent Model: In the Age of Pandemic. *SSRN*, 3619069.
- Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R. R., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B. A., Lee, L. H., Soll, R., Madge, D., ... Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9), 1038–1040.