



EL PAPEL DE LA INTELIGENCIA ARTIFICIAL EN LA PREDICCIÓN DE PROPIEDADES ESTRUCTURALES EN LA SÍNTESIS DE FÁRMACOS

Es innegable que nos encontramos ante una nueva revolución basada en la Inteligencia Artificial (IA) cuya relevancia será comparable a la de la primera revolución industrial. Su alcance se prevé notable en la mayoría de los ámbitos de la vida moderna, forzándonos a adaptarnos, aprender e interpretar estas nuevas herramientas.



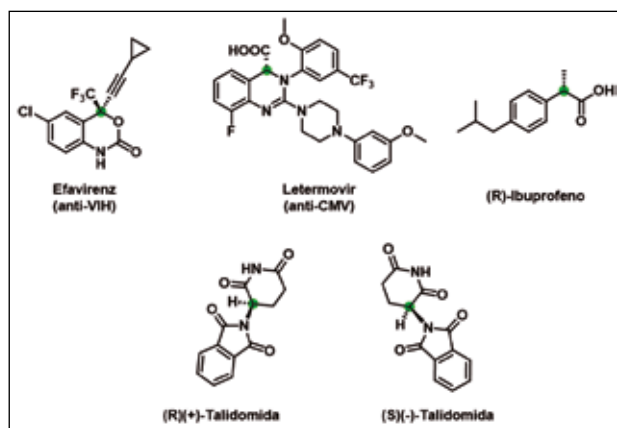
GUILLERMO MARCOS,
Research Scientist de Altenea Biotech en Parque Científico de Madrid.

Una de las disciplinas que ha experimentado grandes cambios en la forma de operar ha sido el diseño y descubrimiento de nuevos fármacos. Este campo, primordial para poder enfrentarnos a las enfermedades contemporáneas que afectan a la sociedad así como a las potenciales epidemias futuras, disfruta de las mejoras que la IA es capaz de ofrecer, desde la reducción de tiempos en el diseño/optimización, reposicionamiento de fármacos, el diseño de *novos* de moléculas potencialmente relevantes para la industria y evidentemente, la reducción de coste y ahorro energético de todos los procesos y por lo tanto lograr conseguir un proceso mas sostenible.

Recientemente, la rama de la inteligencia artificial conocida como *Machine Learning* (ML) o aprendizaje automático, está resultando de vital importancia en uno de los campos con más repercusión en el desarrollo de un nuevo fármaco, su síntesis. Mucho se ha hablado en los últimos años, de cómo los algoritmos son capaces de ayudar en la aparición de nuevas moléculas potencialmente exitosas en la lucha contra enfermedades, pero ha quedado relegado a un papel secundario las posibilidades que ofrece la inteligencia artificial a los procedimientos puramente experimentales. El aprendizaje automático o ML es un campo que explora algoritmos capaces de mejorar automáticamente a través de la experiencia. Aunque sus raíces

se remontan a la informática y la estadística, el ML se ha extendido rápidamente a diversas disciplinas impulsado por sus ventajas en el análisis y la predicción de fenómenos complejos. Esta expansión ha abierto nuevas vías para la exploración científica, yendo más allá de los métodos convencionales de ensayo y error. Más concretamente, la adopción de técnicas de ML en química introduce flujos de trabajo de descubrimiento digital que permiten a los investigadores ahorrar tiempo, recursos y reducir residuos en áreas de investigación como la catálisis, la síntesis orgánica o en numerosos procesos industriales.

Dentro de la parte sintética de una nueva molécula, es de vital trascendencia controlar las propiedades estructurales a lo largo de su camino de construcción, causantes de los enantiómeros y regioisómeros. Muchas de las moléculas orgánicas presentan la propiedad de ser quirales, y por ello, existen en dos formas o enantiómeros, que son imagen especular una de otra, pero no son superponibles. El ejemplo más claro son las palmas de nuestras manos, donde una es la imagen especular de la otra, y es necesario enfrentarse para que sean superponibles. Los fármacos enantioméricamente puros tienen ventajas sobre las mezclas racémicas, ya que se unen con enzimas, hormonas y otras proteínas de forma estereoespecífica, lo que conduce a buena eficacia sin efectos adversos. La



Ejemplos de fármacos enantioméricamente definidos.

talidomida es uno de los ejemplos más claros de la importancia del estudio de la racemización de los fármacos en el ser humano, ya que, al convertirse en su otro enantiómero, resulta tener propiedades tóxicas, y fue responsable de provocar malformaciones embrionarias en las embarazadas. En la actualidad, los distintos organismos reguladores de los fármacos de la mayoría de los países prohibieron la comercialización de fármacos racémicos. Esto hace imprescindible el control estructural de todas las moléculas que se comercializan con fines terapéuticos. Para ello, en la química sintética más tradicional, se emplean catalizadores con ciertas propiedades, capaces de generar diferentes ratios de las distintas estructuras y cuyos resultados se obtienen mediante prueba y error. Es en este campo, es donde las herramientas de IA tienen un gran potencial para ayudar a los investigadores a obtener metodología sintética más eficaz para obrar dichas moléculas.

En los últimos años, la aplicación del ML en catálisis homogénea ha surgido como un enfoque prometedor para acelerar el descubrimiento, la comprensión y el diseño de sistemas catalíticos. Este campo de la catálisis ha sido galardonado en las últimas décadas con varios premios Nobel por su enorme importancia, especialmente en el desarrollo de procesos químicos sostenibles y eficientes. Una característica útil del ML en la catálisis homogénea es su capacidad para acelerar el diseño y la optimización de catalizadores, fundamentales para el correcto devenir de las reacciones químicas. Mediante el cribado virtual de posibles candidatos, los investigadores pueden identificar catalizadores prometedores para su validación experimental, minimizando así el número de experimentos necesarios para alcanzar resultados óptimos de reacción. Los algoritmos de ML han demostrado ser herramientas valiosas para predecir la reactividad y la selectividad.

Utilizando amplias bases de datos de reacciones para entrenar modelos predictivos, los científicos pueden anticipar los resultados de diversos sistemas catalíticos y, por tanto, diseñar catalizadores con la selectividad y el regio/estereocontrol deseados, vitales en la síntesis farmacéutica. Mediante el análisis de perfiles cinéticos, estructuras moleculares y condiciones de reacción, los algoritmos de ML pueden identificar patrones y correlaciones, arrojando luz sobre las vías subyacentes que rigen la actividad

catalítica. Los predictores de ML se integran frecuentemente con herramientas de quimio-informática, caracterización molecular o estructural y validación experimental. En este contexto, la quimio-informática y los métodos basados en la *Mecánica Cuántica* se emplean habitualmente para caracterizar moléculas en descriptores adecuados para generar predicciones. Estos descriptores, pretender definir de la forma más precisa posible, como es una molécula de una forma matemática. Los más comunes, son los descriptores que describen tridimensionalmente, mediante coordenadas, ángulos y vectores, la disposición relativa de una molécula en el espacio, así como la carga o densidad electrónica de cada uno de los átomos que la conforman. A pesar de ser adiciones relativamente recientes a los modelos de ML, estos métodos de predicción están ganando terreno, con un número creciente de publicaciones que demuestran su potencial en el descubrimiento de catalizadores. De este modo, con los resultados obtenidos durante los años, estos algoritmos de aprendizaje automático son capaces de generar predicciones muy acertadas sobre características estructurales tan complejas como la relación de enantiómeros o de regioisómeros.

Los modelos de ML se dividen en dos grandes grupos, aprendizaje supervisado y no supervisado. Dentro del aprendizaje supervisado, encontramos modelos de clasificación y modelos de regresión, y dentro del aprendizaje no supervisado, encontramos los modelos de clasificación o *clustering*. Algunos de los modelos más empleados dentro de esta área, serían la Regresión lineal, Regresión logística, Árboles de decisión, Random Forest, XGBoost o Gradient Boosting. Por otro lado, los modelos basados en redes neuronales no son tan comunes, debido a que necesitan de grandes cantidades de datos para poder predecir correctamente. Por desgracia, el campo de la generación de estructuras y predicción de propiedades estructurales tiene la limitación de que el número de datos disponible no es tan elevado como en otras líneas de investigación, ya que, en su mayoría, han de ser obtenidos mediante la más pura experimentación. Pese a ello, los resultados que se están obteniendo en la actualidad, son de gran utilidad, permitiendo acceder a moléculas estructuralmente puras.

Para finalizar, podemos concluir que al igual que en otros ámbitos, la inteligencia artificial, y en especial los modelos de *Machine Learning*, están resultando ser una auténtica revolución en la predicción de propiedades relacionadas con la síntesis química. Esto está permitiendo acceder a predicciones de todo tipo, desde mejoras en los rendimientos de reacción, como en una mayor selectividad, lo que se traduce en una disminución del tiempo de síntesis, un abaratamiento de los costes, un menor uso y más eficiente de la energía y, en definitiva, un acceso más veloz a fármacos potencialmente exitosos.