



RETROSÍNTESIS, SU RELEVANCIA EN QUÍMICA ORGÁNICA Y EN EL DESARROLLO DE FÁRMACOS

La retrosíntesis, introducida por E.J. Corey en los años 60, constituye una herramienta fundamental en la química sintética. Su enfoque sistemático permite diseñar rutas descomponiendo moléculas objetivo complejas en precursores más simples, facilitando la síntesis química desde materiales de partida accesibles. Este proceso no solo es esencial en la fabricación de productos farmacéuticos, sino que también impulsa el desarrollo de materiales avanzados con aplicaciones tecnológicas y biomédicas.

PABLO TALAVANTE,
Altenea Biotech.

GUILLERMO MARCOS AYUSO,
Altenea Biotech.

En el contexto del desarrollo de fármacos, la retrosíntesis es un aliado indispensable para abordar la creciente demanda de moléculas novedosas, especialmente ante la emergencia de enfermedades complejas como el Alzheimer o cáncer entre otras. En este ámbito, la exploración del vasto espacio químico para identificar estructuras viables y escalables plantea desafíos significativos, que se ven exacerbados por la limitada disponibilidad de recursos y el tiempo necesario para implementar rutas tradicionales. Históricamente, la retrosíntesis dependió de la experiencia y la interpretación de los científicos, quienes se apoyaban en reacciones conocidas y análisis manuales. Sin embargo, esta aproximación presenta limitaciones intrínsecas, como el alto coste en términos de tiempo y esfuerzo humano, además de una capacidad restringida para explorar rutas sintéticas en un espacio químico cada vez más vasto y multidimensional.

La incorporación de algoritmos basados en inteligencia artificial (IA) ha revolucionado este campo. Métodos

recientes, como las redes neuronales profundas, permiten analizar grandes bases de datos químicas y predecir rutas sintéticas de manera más eficiente. Estos algoritmos no solo ofrecen un rendimiento superior en problemas de retrosíntesis de un solo paso, sino que también pueden integrarse con técnicas de búsqueda para identificar caminos sintéticos complejos con mayor precisión.

Además, la capacidad de los algoritmos de IA para procesar datos estructurales y de reacción ha generado una transición hacia una retrosíntesis automatizada y escalable. Por ejemplo, herramientas como las redes neuronales basadas en grafos son capaces de inferir transformaciones químicas relevantes incluso en moléculas no reportadas previamente, superando la dependencia de reglas predefinidas.

A la hora de modelar un problema de este tipo de forma computacional, existen varias formas de plantear su solución, pero todas pasan por resolver dos aspectos: la retrosíntesis en un paso y la retrosíntesis en múltiples pasos. El primero de ellos, consiste en, dada una molécula

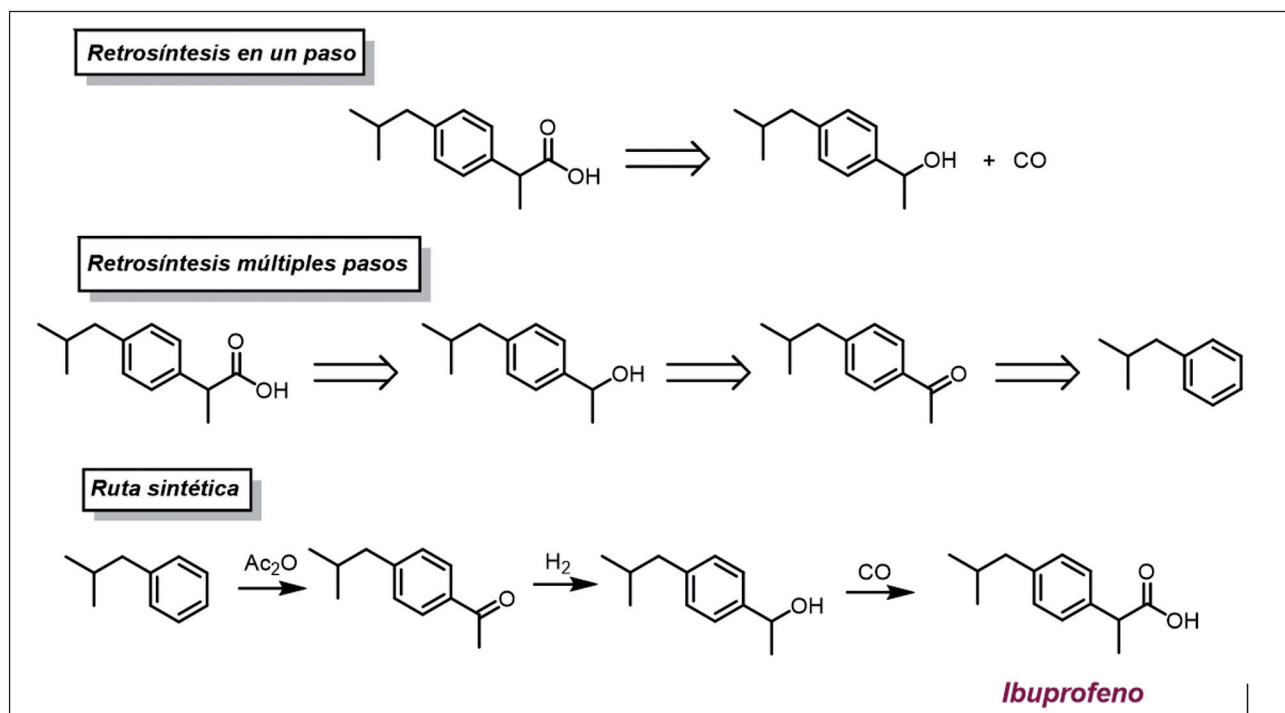


Figura 1. Retrosíntesis y ruta sintética de la molécula de ibuprofeno.

objetivo, obtener sus precursores inmediatos. El segundo, una vez resuelto el primero problema, se basa en obtener una ruta retrosintética completa, hasta llegar a productos comerciales, lo más eficiente posible.

Esta noción de eficiencia es otro de los puntos clave a la hora de desarrollar este tipo de algoritmos debido a que no existe una definición formal, pudiendo depender ésta del coste monetario de los productos de partida, del número de pasos de síntesis, del tiempo que consume cada reacción o del rendimiento de estas. Incluso, aunque se definiera una función de coste en base a estas variables, éstas pueden ser dinámicas (como la fluctuación del precio de los productos), pueden no tenerse acceso a un valor real y deben, por tanto, ser estimadas (como el tiempo de reacción o los rendimientos) o incluso depender del contexto (no sería lo mismo un contexto académico que uno industrial). Este tipo de problema recuerda a algo parecido a una partida de ajedrez, donde existe la necesidad de planificación y donde una jugada puede parecer buena a priori, pero puede entorpecernos cuando la partida haya avanzado más.

Para entrenar modelos de inteligencia artificial que ayuden a diseñar síntesis químicas en un solo paso, necesitamos bases de datos con reacciones conocidas. El aprendizaje se basa en la definición de patrones de reacción. Un patrón de reacción es una regla general que describe

cómo ciertos grupos químicos pueden transformarse sin necesidad de analizar toda la estructura molecular. Existen tres enfoques principales:

- Basados en patrones** (template-based): utilizan reglas predefinidas para predecir cómo una molécula puede descomponerse en sus precursores.
- Basados en semi-patrones** (semi-template-based): Combinan reglas predefinidas con modelos más flexibles.
- Sin patrones** (template-free): Se basan en aprendizaje automático, generalmente deep learning, sin depender de reglas establecidas.

El primer enfoque, basados en patrones, se basa en la identificación de estructuras dentro de la molécula para posteriormente poder aplicar reglas inversas para predecir los compuestos precursores. La desventaja de esta estrategia es la cantidad de patrones posibles ya que puede ser enorme (decenas de miles), lo que complica tanto el entrenamiento del modelo como su capacidad de hacer predicciones rápidas.

Para solucionar este problema, los modelos basados en semi-patrones utilizan una versión más flexible de los patrones de reacción. Esto permite que, aunque una reacción específica no esté en los datos de entrenamiento, el modelo pueda reconocer su esencia y predecir sus precursores.

Para aplicar este enfoque, primero es necesario identificar el centro de reacción, es decir, la parte de la molécula donde ocurre la transformación. Luego se aplican los

A pesar de los avances en química sintética, sigue existiendo el desafío de mantener las bases de datos actualizadas para reflejar nuevas reacciones y métodos emergentes. Esto requiere no solo esfuerzos de digitalización y análisis, sino también la integración de conocimientos expertos para mejorar la calidad de los datos.

semi-patrones a este centro de reacción para generar las moléculas precursoras. La gran ventaja de este método es que lo hace mucho más eficiente, reduciendo drásticamente la cantidad de opciones a considerar, de decenas de miles a solo unos pocos cientos. Sin embargo, el desafío radica en mantener la precisión química, ya que el uso de reglas más generales aumenta el número de posibles soluciones para cada retrosíntesis.

Los modelos sin patrones también comienzan identificando el centro de reacción, sin embargo, utilizan modelos generativos basados en inteligencia artificial, similares a un ChatGPT entrenado para predecir reacciones químicas.

Estos modelos aprenden a partir de ejemplos de reacciones reales y generan precursores moleculares intentando mantener la coherencia química. Sin embargo, al no seguir reglas estrictas, pueden ser más creativos al proponer nuevas rutas, pero también cometer errores, incluso rompiendo principios fundamentales como el mapeo correcto de los átomos entre los productos y sus precursores.

En general, hay un equilibrio entre precisión y creatividad en los métodos de retrosíntesis:

a) **Modelos basados en patrones** (template-based): Son los más precisos, pero limitados a las reacciones en sus datos de entrenamiento.

b) **Modelos semi-template-based**: Logran un balance entre precisión y flexibilidad.

c) **Modelos sin patrones** (template-free): Pueden generar rutas innovadoras, pero con mayor riesgo de errores.

El éxito de estos algoritmos depende en gran medida de la calidad y cantidad de datos disponibles. Aunque existen repositorios públicos como USPTO, Reaxys o SciFinder, presentan problemas como errores en las entradas, datos incompletos y falta de información sobre reacciones fallidas, lo que dificulta el aprendizaje de los modelos de IA.

A pesar de los avances en química sintética, sigue existiendo el desafío de mantener las bases de datos actualizadas para reflejar nuevas reacciones y métodos emergentes. Esto requiere no solo esfuerzos de digitalización y análisis, sino también la integración de conocimientos expertos para mejorar la calidad de los datos. Además, es fundamental filtrar metodologías ineficientes, contaminantes o que utilicen compuestos peligrosos, asegurando un enfoque más sostenible en la química computacional.

Referencias

- Strieth-Kalthoff, F.; Szymkuć, S.; Molga, K.; Aspuru-Guzik, A.; Glorius, F.; Grzybowski, B. A. Artificial Intelligence for Retrosynthetic Planning Needs Both Data and Expert Knowledge. *J. Am. Chem. Soc.* 2024, 146, 11005–11017. <https://doi.org/10.1021/jacs.4c00338>
- Zhong, Z.; Song, J.; Feng, Z.; Liu, T.; Jia, L.; Yao, S.; Hou, T.; Song, M. Recent Advances in Artificial Intelligence for Retrosynthesis. *arXiv* 2023, arXiv:2301.05864. <https://doi.org/10.48550/arXiv.2301.05864>
- Reker, D.; De Fabritiis, G. Computational Retrosynthesis and Reaction Prediction Using Machine Learning. *Nat. Commun.* 2023, 14, 41698. <https://doi.org/10.1038/s41467-023-41698-5>
- Jiang, Y.; Yu, Y.; Kong, M.; Mei, Y.; Yuan, L.; Huang, Z.; Kuang, K.; Wang, Z.; Yao, H.; Zou, J.; Coley, C. W.; Wei, Y. Artificial Intelligence for Retrosynthesis Prediction. *Engineering* 2023, 25, 32–50. <https://doi.org/10.1016/j.eng.2022.04.021>
- Gao, Z.; Tan, C.; Wu, L.; Li, S. Z. SemiRetro: Semi-template Framework Boosts Deep Retrosynthesis Prediction. *arXiv* 2022, arXiv:2202.08205. <https://arxiv.org/abs/2202.08205>
- Chen, B.; Li, C.; Dai, H.; Song, L. Retro*: Learning Retrosynthetic Planning with Neural Guided A* Search. *arXiv* 2020, arXiv:2006.15820. <https://arxiv.org/abs/2006.15820>