

The chemoinformatics revolution: teaching chemistry to robots

LA REVOLUCIÓN QUIMIOINFORMÁTICA: ENSEÑÁNDOLE QUÍMICA A LOS ROBOTS

José Raúl Montero Bastidas¹

En este artículo se describen los avances más recientes en la automatización de la química orgánica con un enfoque en la química medicinal. Diferentes plataformas han permitido la ejecución de cientos e incluso miles de reacciones en paralelo con el fin de acelerar el descubrimiento de un medicamento. Por otro lado, modelos de inteligencia artificial están siendo aplicados en química orgánica para diseñar experimentos con condiciones que aumenten la probabilidad de éxito. Estos avances en conjunto prometen revolucionar la síntesis orgánica y los campos de la industria que dependen de ella.

Palabras clave: Automatización, quimioinformática, reacciones en paralelo, HTE, aprendizaje automático, inteligencia artificial

This article describes recent advances in the automation of organic chemistry focused on medicinal chemistry. Different platforms have enabled the execution of hundreds and even thousands of reactions in parallel to accelerate the discovery of pharmaceuticals. On the other hand, artificial intelligence models are being applied in organic chemistry to design experiments with conditions that increase the probability of success. Together, these advances promise to revolutionize organic synthesis and the fields of industry that depend on it.

Keywords: Automation, cheminformatics, parallel reactions, HTE, machine learning, artificial intelligence

Recibido: 06 de noviembre de 2023


Aceptado en forma final: 04 de marzo de 2024

Cómo citar este artículo:

Montero-Bastidas, J.R.: "La revolución quimioinformática: enseñándole química a los robots". *Revista de Química*, 2024, 38 (1), 2-14

DOI: <https://doi.org/10.18800/quimica.202401.001>

¹ Merck & Co., Inc., West Point, PA 19486, USA

 <https://orcid.org/0000-0002-6325-8565>
jose.raul.montero.bastidas@merck.com



INTRODUCCIÓN

¿Cómo lucirá el laboratorio de química en el futuro? La automatización tanto de los procedimientos sintéticos, así como del diseño de experimentos sugiere que los robots se volverán nuestros compañeros de laboratorio. La automatización de la síntesis orgánica requiere la implementación de robots que puedan realizar reacciones, así como el desarrollo de algoritmos de aprendizaje automático que puedan plantear experimentos. Nos encontramos en un momento clave debido al avance de ambos campos en las últimas décadas. El uso de robots para ejecutar reacciones en paralelo se ha vuelto una práctica rutinaria en varios campos de la industria y los modelos de inteligencia artificial han comenzado a mostrar las ventajas que pueden traer a la sociedad como es el uso de ChatGPT.

El desarrollo de compuestos farmacéuticos depende de la química sintética y la automatización promete reducir el tiempo necesario para desarrollar una medicina. La pandemia del COVID-19 nos ha demostrado cuán vital es poder combatir una nueva enfermedad en un corto periodo de tiempo y esto solo se puede lograr mediante avances tecnológicos. A continuación, discutiremos el desarrollo de las nuevas plataformas para implementar reacciones químicas en paralelo así como de los algoritmos para predecir los resultados de estos experimentos con un enfoque en su aplicación en la química medicinal.

DESCUBRIMIENTO DE FÁRMACOS: UN RESUMEN

El descubrimiento de un nuevo compuesto farmacéutico no solo conlleva encontrar una molécula que sea efectiva contra una enfermedad sino también que cuente con el perfil de propiedades adecuado para convertirse en una medicina. Los candidatos a fármacos deben poseer ciertas propiedades fisicoquímicas y biológicas para que sean seguros de usar. Para lograr este objetivo, se sigue un proceso iterativo de Diseño, Síntesis, Test y Análisis (DSTA) (Figura 1).¹⁻³ El primer paso implica el diseño de compuestos con el potencial de tener actividad biológica contra el objetivo o diana biológica de interés. Una de las herramientas empleadas en este paso es el *docking* que nos permite modelar las interacciones de los compuestos diseñados con la diana de interés para ayudar a seleccionar los mejores candidatos. Además, usando modelos que puedan predecir la actividad biológica y propiedades fisicoquímicas en base a la estructura química se priorizan las propuestas que tengan mayor probabilidad de tener un efecto terapéutico deseable y sin efectos secundarios. Posteriormente, las moléculas son sintetizadas en el laboratorio y evaluadas en los ensayos biológicos. Con base en los resultados obtenidos de estos experimentos, las hipótesis son reevaluadas para diseñar mejores moléculas y comenzar un nuevo ciclo de DSTA.

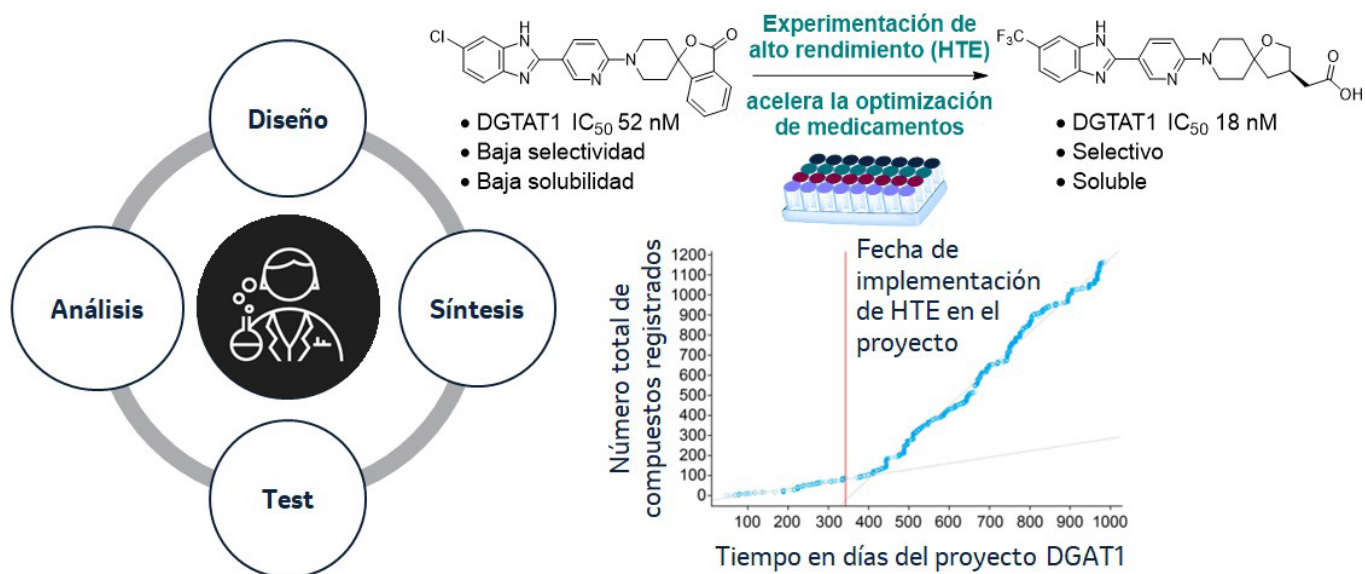


Figura 1. Implementación de HTE para acelerar el ciclo DSTA en el descubrimiento de fármacos. Imagen adaptada de Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. *The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. Acc. Chem. Res.* **2017**, *50* (12), 2976–2985.

SÍNTESIS ORGÁNICA EN PARALELO

La automatización de la química orgánica permite sintetizar varias moléculas en paralelo y acelerar significativamente el paso de síntesis en el ciclo de DSTA (Figura 1).¹⁻³ La síntesis de varias moléculas en paralelo se conoce como Experimentación de Alto Rendimiento o *High-Throughput Experimentation* (HTE) en inglés.²⁻⁴ Además de ahorrar tiempo, los robots usados en HTE permiten ejecutar reacciones en escalas de micro- y nanomoles reduciendo, por tanto, a cantidades mínimas la utilización de valiosos y costosos intermediarios que suelen ser sintetizados en múltiples pasos. Las plataformas de HTE se pueden aplicar con dos objetivos diferentes: optimizar las condiciones de una reacción o sintetizar múltiples moléculas en una librería. Ambos objetivos se pueden llevar a cabo al mismo tiempo siendo no excluyentes uno del otro. A continuación, discutiremos los recientes avances en HTE y las limitaciones que conlleva.

I. Miniaturización de reacciones

El tiempo necesario para ejecutar decenas de reacciones químicas puede ser un factor limitante para la adaptación de HTE. La optimización de una reacción química

implica la evaluación de diferentes condiciones hasta lograr el rendimiento deseado. Con el fin de acelerar este proceso vía HTE, se han implementado arreglos de viales ya equipados con los reactivos empleados usualmente durante la optimización de una transformación química particular. Estos arreglos de viales se conocen como catalysis kits y usualmente consisten de 24 viales de 1 mL o 100 μL (Figura 2a).² Por ejemplo, el catalysis kit para optimizar el acoplamiento de C—O entre un alcohol y un haluro de arilo mediante fotoredox viene equipado con 3 fotocatalizadores en dos diferentes cantidades (2 y 5 % mol) y con un precatalizador de níquel en todos los viales.⁴ Actualmente, los catalysis kits son rutinariamente utilizados en la industria como una de las primeras herramientas para optimizar una reacción.

De hecho, hay disponibles plataformas comerciales para implementar 24 o 96 reacciones y son ampliamente usadas para montar reacciones en escalas de 25 μL a 1 mL. Usualmente se preparan soluciones stock de los reactivos que son dispensadas en cada vial con pipetas de múltiples canales. Sin embargo, para ejecutar reacciones en escalas incluso más pequeñas se necesitan equipos más sofisticados. Con este fin, se han adaptado los robots de transferencia de líquidos utilizados en los ensayos biológicos para implementar reacciones en escalas de nanomoles.⁵⁻⁷ Estos robots cuentan

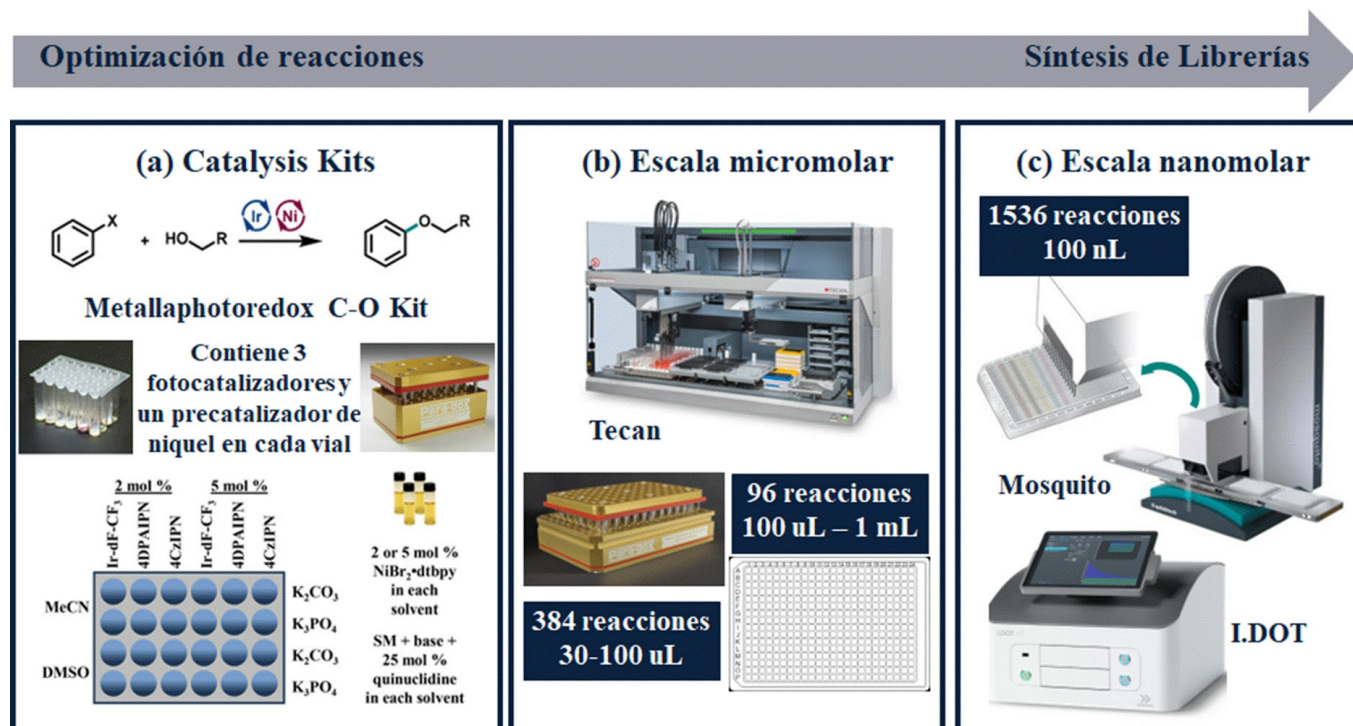


Figura 2. (a) Catalysis kit para optimizar un acoplamiento C—O. Imagen adaptada de Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Diego, J. E. de; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. *The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future*. *Org. Process Res. Dev.* **2019**, *23* (6), 1213–1242, Copyright © 2019 American Chemical Society (b) HTE para la implementación de reacciones en escala de micromoles, (c) HTE para la implementación de reacciones en escala de nanomoles.

con agujas para dispensar las soluciones en plataformas con 384 o 1536 compartimentos (**Figura 2a y 2b**, respectivamente) con volúmenes de reacción entre 1 μL y 80 μL . Para poder transferir las soluciones solo se pueden utilizar reactivos que se puedan disolver en el solvente de la reacción. Además, no se pueden utilizar solventes volátiles en escalas de nanolitros debido a las pérdidas por evaporación. Esto explica por qué la solubilidad de los diferentes reactivos es una de las mayores limitaciones si se quiere implementar HTE con éxito.⁸

2. Condiciones universales de reacción

El descubrimiento de un fármaco es de por sí un trabajo arduo que puede tomar años; por lo cual, la síntesis de las moléculas no debería ser una limitación. Encontrar las condiciones correctas para sintetizar una molécula puede tomar tiempo, más aún si se quiere sintetizar una librería

de compuestos. Para automatizar este proceso, se necesita un conjunto de condiciones de reacción generales que nos permitan acceder a cualquier espacio químico que deseemos. Por ello, la HTE se ha utilizado para encontrar tales condiciones generales para diferentes tipos de reacciones.

Se han desarrollado estrategias para determinar la generalidad de diferentes metodologías sintéticas. Las dos estrategias más conocidas son mediante la “evaluación de tolerancia” y el uso de “librerías químicas informativas”.⁹ La evaluación de tolerancia busca examinar el desempeño de la reacción en la presencia de varios grupos funcionales.¹⁰ Para ello, se añaden aditivos con diferentes grupos funcionales a la reacción y se evalúa cómo la presencia del aditivo influye en el rendimiento. Por ejemplo, en la **Figura 3a** se muestra la evaluación de tolerancia para un acoplamiento C–N en la presencia de un pirrol y de un alquino. La reacción tolera

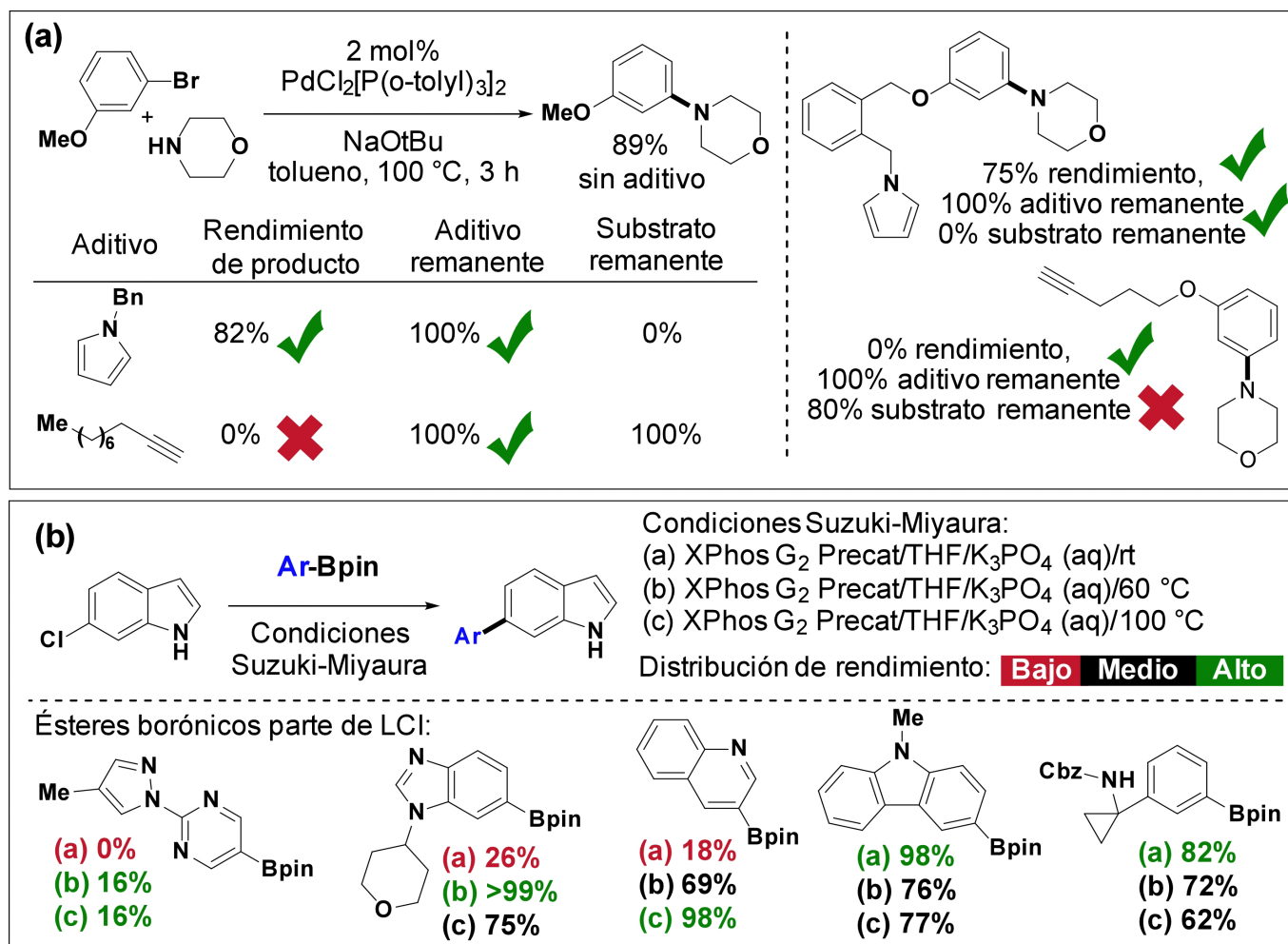


Figura 3. Estrategias para evaluar la generalidad de reacciones orgánicas: (a) Evaluación de tolerancia, (b) Librerías químicas informativas. Imagen adaptada de Montero Bastidas, J. R.; El Marrouni, A.; Chiriac, M. I.; Struble, T.; Kalyani, D. *Accelerating Drug Discovery by High-Throughput Experimentation; Medicinal Chemistry Reviews*; 2022; pp 443–463. Copyright © 2022 American Chemical Society. Abreviaturas: Bpin: boronato de pinacol; XPhos G₂: Cloro(2-diciclohexilfosfino-2',4',6'-trisisopropil-1,1'-bifenil)[2-(2'-amino-1,1'-bifenil)]paladio(II); LIC: Librerías químicas informativas.

el pirrol, pero no el alquino; este resultado permite predecir el éxito de la reacción en moléculas más complejas que contienen ambos grupos funcionales.¹⁰

Otra estrategia para evaluar la generalidad de una reacción es evaluar su desempeño en un grupo de sustratos representativos. La elección de este set de sustratos puede ser guiada por técnicas quimiinformáticas como clustering o los sustratos pueden ser seleccionados manualmente.¹¹ Para evaluar las reacciones en moléculas estructuralmente similares a compuestos farmacéuticos, se han seleccionado set de sustratos con estructuras comunes en medicinas conocidos como librerías químicas informativas.¹² En la **Figura 3b** se muestra la evaluación de tres condiciones para un acoplamiento cruzado Suzuki-Miyaura con cinco compuestos que se han extraído de una librería química informativa de ésteres borónicos.¹³ Se puede notar cómo las condiciones óptimas difieren en cada sustrato, necesitando cada uno de ellos una temperatura diferente para conseguir resultados satisfactorios.

3. HTE para nuevos tipos de reacciones

A pesar de las ventajas de ejecutar reacciones en paralelo, no todas las reacciones pueden ser implementadas por HTE. Por ejemplo, una limitación es que no se pueden usar reactivos sólidos que no sean solubles en el disolvente de reacción.⁸ Aunque existen mecanismos para dispensar sólidos,^{14,15} estos no son lo suficientemente precisos en cantidades de microgramos, especialmente cuando se usan

ligandos o catalizadores en centésimas de equivalentes. Una posible solución es recubrir los reactivos en perlas de vidrio, conocidas como *chembeads* (**Figura 4a**).^{16,17} El porcentaje de reactivo dentro de los *chembeads* es de aproximadamente 10% lo que permite dispensar la cantidad necesaria de los reactivos de forma exacta.

A medida que se desarrollan nuevos tipos de reacciones, también se deben adaptar nuevas plataformas para ejecutar estas reacciones en HTE. Por ejemplo, las reacciones de acoplamiento para formar enlaces Csp²-Csp³ o Csp³-Csp³ se han vuelto muy populares para incrementar el carácter sp³ en una molécula. Los compuestos con una fracción mayor de centros sp³ tienden a tener mejor solubilidad y estabilidad metabólica lo que aumenta sus probabilidades de convertirse en ingredientes farmacéuticos activos.¹⁸ El carbono sp³ suele ser instalado mediante radicales alquilo que pueden formarse fotoquímicamente o electroquímicamente. Para evaluar rápidamente qué sustituyente alquilo impartirá las mejores propiedades fisicoquímicas a una molécula se requieren plataformas que permitan ejecutar reacciones fotoquímicas y electroquímicas en HTE. Afortunadamente, existen reactores para ejecutar reacciones fotoquímicas en paralelo en los que se pueden variar la intensidad lumínica y la longitud de onda, parámetros que son fundamentales en estas reacciones (**Figura 4b**).¹⁹ A su vez, también existen plataformas HTE para reacciones electroquímicas que permiten controlar el voltaje y la corriente aplicada, así como evaluar diferentes electrodos de acuerdo a las condiciones óptimas a utilizar (**Figura 4c**).²⁰



Figura 4 Plataformas de HTE para ejecutar (a) reacciones con reactivos sólidos. Imagen adaptada de Gesmundo, N. J.; Tu, N. P.; Sarris, K. A.; Wang, Y. *ChemBeads-Enabled Photoredox High-Throughput Experimentation Platform to Improve C(sp²)-C(sp³) Decarboxylative Couplings. ACS Med. Chem. Lett.* **2023**, *14* (4), 521–529, Copyright © 2023 American Chemical Society (b) reacciones fotoquímicas en paralelo, (c) reacciones electroquímicas en paralelo.

4. Acelerando la purificación

La implementación de una reacción es solo el primer paso para sintetizar un compuesto. Una vez que la reacción ha concluido, es necesario un proceso de purificación para obtener el compuesto deseado libre de impurezas que puedan influir en su efectividad en el ensayo biológico. Este proceso suele ser costoso y lento más aún si se trata de una librería de 96 o 384 compuestos. Para suplir estas deficiencias, se han desarrollado tecnologías de purificación en paralelo compatibles con las plataformas de HTE antes mencionadas. Incluso reacciones en escalas de micromoles pueden ser purificadas para obtener cantidades suficientes para evaluar el compuesto en los ensayos biológicos.²¹ Los micromoles de productos no pueden cuantificarse con una simple balanza por lo que se usan otras técnicas de cuantificación, como *Charge Aerosol Detection* (CAD). Otra forma de acelerar la purificación de compuestos consiste en enfocarse en las moléculas que muestren actividad en los ensayos biológicos. Para ello, primero se realizan los ensayos biológicos directamente en las mezclas de reacción, esta estrategia se conoce como *Direct to Biology* (DtB).^{22,23} Sin embargo, no todas las reacciones y no todos los ensayos pueden implementarse en DtB y es necesario llevar a cabo una validación adecuada de la reacción que dé certeza de que ninguno de los reactivos ni productos de partida influyan en los resultados del ensayo. De esta manera, si la mezcla muestra una respuesta positiva se concluye que debe ser provocada por el producto de reacción.

MODELOS DE INTELIGENCIA ARTIFICIAL EN SÍNTESIS ORGÁNICA

Hasta ahora hemos discutido los recientes avances tecnológicos en la automatización de síntesis orgánica que buscan optimizar la inversión no solo de tiempo sino también de recursos. La evolución de las plataformas de HTE ha permitido que más reacciones puedan ser miniaturizadas y que los productos puedan ser evaluados incluso antes de ser purificados. Sin embargo, aún quedan desafíos por resolver como la selección de la reacción y las condiciones correctas para sintetizar un compuesto en particular. Este tipo de decisiones se basan generalmente en experiencia previa o en una búsqueda detallada de reacciones similares en la literatura. Desafortunadamente, este proceso puede tomar tiempo o conllevar a elecciones subjetivas más aún si se evalúan todos los compuestos en una librería de más de cien moléculas. Por ello, la inteligencia artificial busca ayudar a automatizar la toma de decisiones e incluso a aumentar el porcentaje de éxito al sintetizar una molécula.

¿Qué le enseñamos a las computadoras?

Los algoritmos de inteligencia artificial se han

utilizado desde hace un par de décadas en química medicinal para predecir las propiedades fisicoquímicas de las moléculas, así como su afinidad a dianas biológicas no deseadas.²⁴ Las relaciones cuantitativas de estructura y actividad (QSAR, por sus siglas en inglés) generadas por los modelos permiten priorizar moléculas con un mayor potencial para convertirse en fármacos y acelerar el diseño en el ciclo DSTA. Se puede también acelerar la síntesis al construir modelos predictivos del éxito de las reacciones. Estos modelos pueden proponer qué tipo de transformación o reactivos usar así como estimar el rendimiento o selectividad de la reacción.

Corey fue un pionero en la construcción de algoritmos para generar rutas retrosintéticas aunque estos no estuvieron basados en aprendizaje automático.²⁵ El grupo de investigación de Corey recopiló un conjunto de reglas empíricas para proponer los pasos a seguir en la síntesis de una molécula. Recientemente, el grupo de Grzybowski siguió un enfoque similar logrando codificar manualmente más de 50 000 reglas retrosintéticas para construir el programa Synthia (formalmente, Chematica).²⁶ Las rutas sintéticas propuestas por Synthia han sido exitosamente replicadas en varios laboratorios logrando la síntesis de moléculas como la Umifenovir mostrada en la **Figura 5**.²⁷⁻³¹ Umifenovir es un antiviral usado para combatir el virus SARS-CoV1; por lo que fue propuesto como un candidato para tratar la enfermedad del COVID-19. Por este motivo, el laboratorio de Cernak utilizó Synthia para diseñar cuatro nuevas rutas sintéticas para producir Umifenovir.²⁸ Las cuatro rutas sintéticas fueron replicadas con éxito en el laboratorio y con solo ligeras modificaciones en las condiciones propuestas originalmente. Aunque desafortunadamente Umifenovir mostró una efectividad limitada contra el COVID-19 en los ensayos clínicos, este estudio demuestra como Synthia puede ayudar en el descubrimiento de medicinas.

Existen también algoritmos retrosintéticos construidos mediante inteligencia artificial, como ASKCOS desarrollado en el MIT con el apoyo del Consorcio de Aprendizaje Automático para la Síntesis y Descubrimiento de Farmacéuticos (MLPDS, por sus siglas en inglés).³² Al igual que Synthia, ASKCOS ha predicho satisfactoriamente rutas sintéticas reproducibles en el laboratorio, como es el caso de la síntesis de Safinamida.³³ Otros modelos de retrosíntesis populares incluyen AiZynthFinder e IBM RXN.^{34,35} Aparte de diseñar rutas retrosintéticas, los modelos de aprendizaje automático pueden predecir la selectividad de una reacción química de sintetizar el producto deseado. Por ejemplo, ASKCOS incluye algoritmos para predecir la regioselectividad en reacciones de funcionalización de enlaces C—H en anillos aromáticos y heteroaromáticos.^{36,37} Otra aplicación de los modelos de inteligencia artificial es en la automatización de procesos rutinarios en la síntesis orgánica. El uso de ChatGPT ha llamado la atención recientemente y el campo

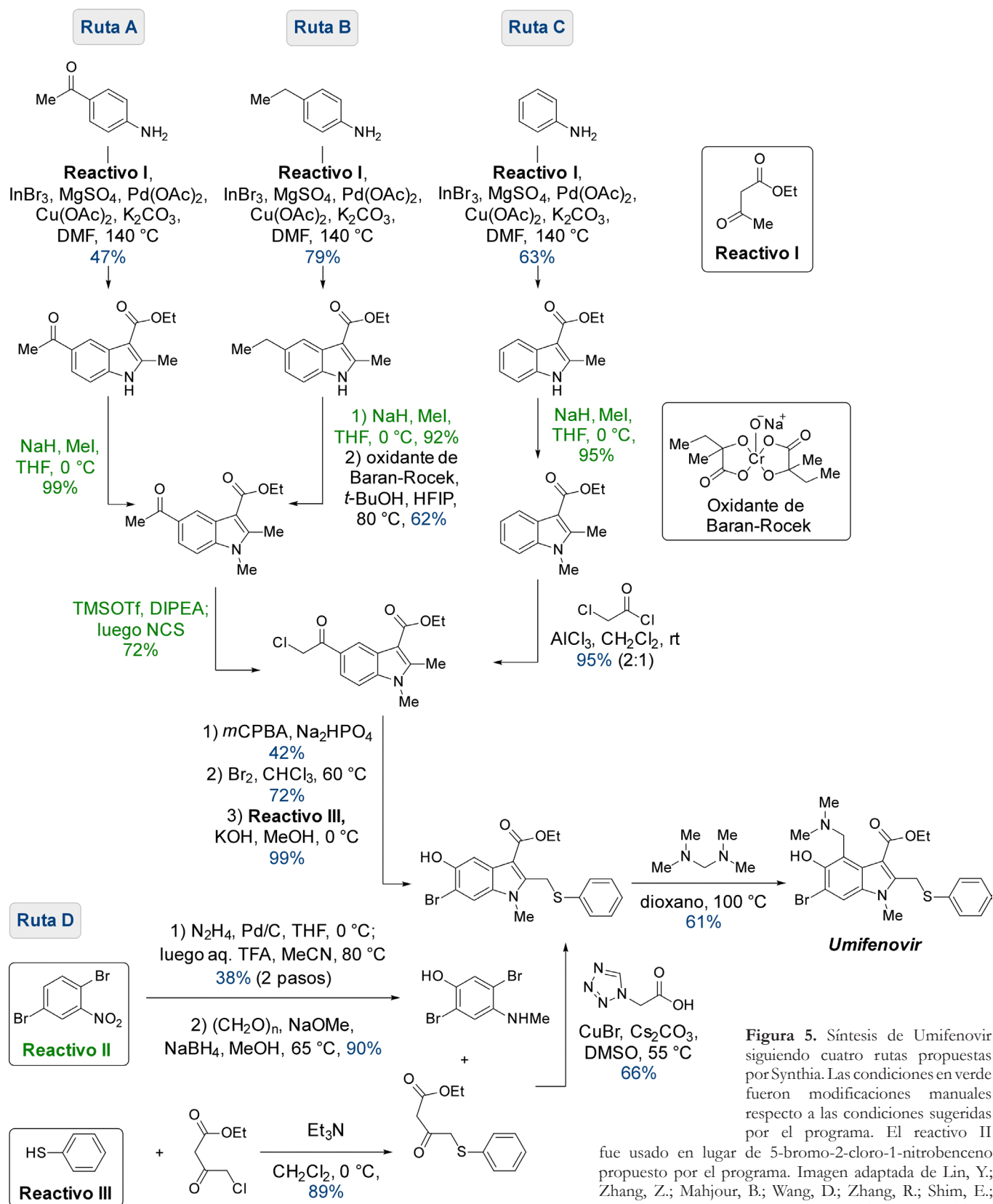


Figura 5. Síntesis de Umifenovir siguiendo cuatro rutas propuestas por Synthia. Las condiciones en verde fueron modificaciones manuales respecto a las condiciones sugeridas por el programa. El reactivo II fue usado en lugar de 5-bromo-2-cloro-1-nitrobenzoceno propuesto por el programa. Imagen adaptada de Lin, Y.; Zhang, Z.; Mahjour, B.; Wang, D.; Zhang, R.; Shim, E.; McGrath, A.; Shen, Y.; Brugger, N.; Turnbull, R.; Trice, S.; Jasty, S.; Cernak, T. *Reinforcing the Supply Chain of Umifenovir and Other Antiviral Drugs with Retrosynthetic Software. Nat. Commun.* **2021**, *12*, 7327.

de la química orgánica no es una excepción. Se ha utilizado ChatGPT para diseñar experimentos para HTE mostrando cómo este algoritmo puede entender el lenguaje químico.³⁸

La optimización de una reacción química puede ser un proceso laborioso. Afortunadamente, los algoritmos de optimización Bayesiana han sido usados con éxito por diferentes grupos para optimizar reacciones químicas. Por ejemplo, el grupo de Doyle desarrolló el programa EDBO+ para optimizar tanto el rendimiento como la selectividad de una reacción.^{39,40} En particular, se comparó el desempeño del algoritmo con el desempeño de químicos expertos en optimizar el rendimiento de un acoplamiento de un bromuro de arilo con un imidazol. Aunque los químicos lograron mejores rendimientos en las primeras iteraciones, EDBO+ logró alcanzar más de 98% de rendimiento en menos reacciones. Esta diferencia se puede atribuir a las diferentes estrategias utilizadas. Los químicos tendemos a evaluar condiciones similares a aquellas que nos han dado mejores resultados hasta el momento (explotación) lo cual puede estancarnos en un “máximo local”. Por otro lado, los algoritmos de optimización Bayesiana evalúan también condiciones muy diferentes a las exploradas hasta el momento (exploración) lo que permite llegar a un “máximo global”. De esta manera, los métodos utilizados por los algoritmos de inteligencia artificial pueden

inspirar a los químicos a reevaluar las estrategias utilizadas en la síntesis de moléculas.

El grupo de Lapkin también desarrolló un algoritmo de optimización Bayesiana que se caracteriza por utilizar información previa de reacciones similares para optimizar la síntesis de nuevos compuestos (**Figura 6**).^{41,42} El modelo fue aplicado para optimizar la síntesis de cuatro compuestos **1-4** con el mismo tipo de transformación: un acoplamiento intramolecular mediado por la activación de un enlace C–H. El modelo fue entrenado inicialmente con 16 reacciones con diferentes condiciones para sintetizar **1**. Las siguientes reacciones fueron propuestas por el algoritmo; el cual necesitó solo 7 experimentos adicionales para alcanzar el rendimiento óptimo (línea azul). Utilizando los datos recolectados en la optimización de **1**, el algoritmo fue usado para optimizar la síntesis de **2** (línea roja) y posteriormente de **3** (línea naranja) requiriendo solo 11 y 5 experimentos respectivamente. La síntesis de **4** desafía al algoritmo a sintetizar una lactama de seis miembros en contraste a las de cinco miembros sintetizadas previamente. Aún así, solo le tomó al algoritmo 9 experimentos para llegar a un rendimiento óptimo de **4** (línea verde). Estos ejemplos demuestran cómo datos recolectados previamente pueden ayudar a acelerar la optimización Bayesiana de reacciones químicas.

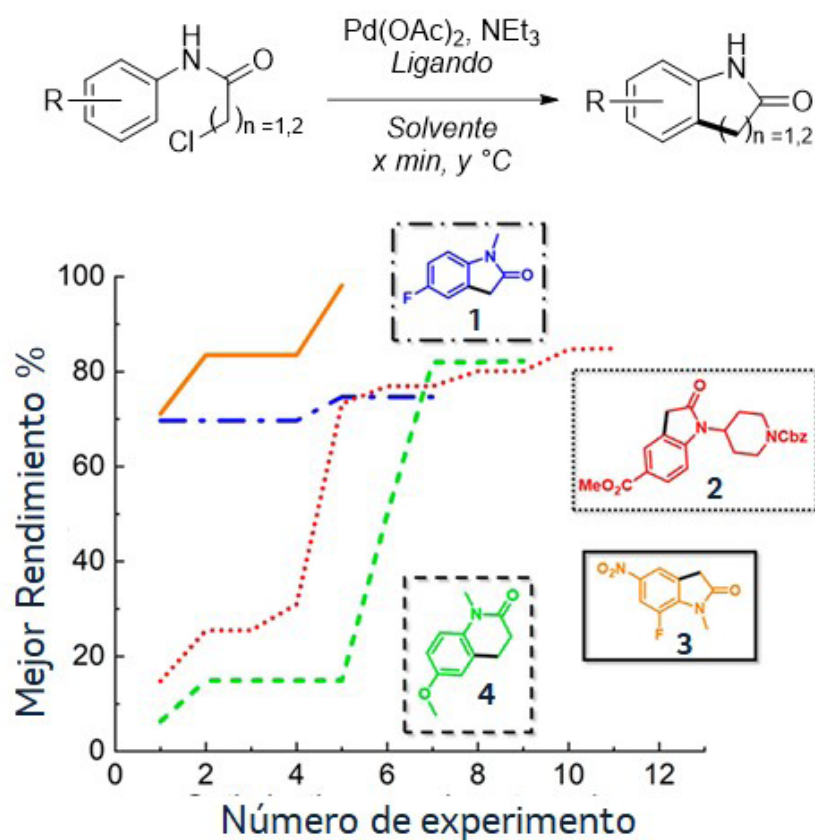


Figura 6. Optimización bayesiana de rendimientos de reacción para un acoplamiento intramolecular vía la activación de un enlace C–H en cuatro compuestos. Moléculas enumeradas en orden de optimización. Imagen adaptada de Taylor, C. J.; Felton, K. C.; Wigh, D.; Jeraal, M. I.; Grainger, R.; Chessari, G.; Johnson, C. N.; Lapkin, A. A. *Accelerated Chemical Reaction Optimization Using Multi-Task Learning*. *ACS Cent. Sci.* **2023**, *9*, 957–968.

¿Cómo se entrenan los modelos?

Primer dilema: ¿cantidad o calidad?

Una pregunta frecuente consiste en determinar qué factor resulta más relevante: la calidad o la cantidad de datos. Para construir modelos de inteligencia artificial más exactos se necesitan ambos. Desafortunadamente, las bases de datos con las que contamos actualmente debatiblemente carecen de uno de los dos factores. Por un lado, las reacciones reportadas en la literatura, aunque repleta de diferentes ejemplos, no han sido recolectadas en un formato entendible por las computadoras. Las bases de datos como Reaxys, muy útiles para encontrar referencias, han hecho el mejor esfuerzo para convertir las condiciones de reacción reportadas en los artículos a un formato que pueda ser asimilado por la computadora. Desafortunadamente, este proceso no es perfecto y muchas veces se omiten disolventes o reactivos claves para la reacción. A pesar de estas limitaciones, muchos modelos como ASKCOS han sido construidos con esta base de datos. Para resolver este problema se han creado iniciativas para incentivar a los autores a compartir la información de sus artículos en nuevas bases de datos que almacenen la información de forma consistente. Esta nueva base de datos permite que la información sea encontrable, accesible, interoperable y reusable (FAIR, por sus siglas en inglés) para permitir su uso en la construcción de modelos.

Otro problema de las bases de datos de la literatura se observa en la distribución del rendimiento de las reacciones. Para entrenar adecuadamente a los algoritmos de inteligencia artificial se requieren datos que reflejen representativamente todo el espacio químico. Los artículos científicos suelen reportar solo reacciones con rendimientos de moderados a altos y excluyen sustratos que fallan en las reacciones. La carencia de datos fallidos provoca que los modelos predigan que casi siempre todas las reacciones tengan éxito, lo cual está lejos de la realidad. Este paradigma ha iniciado una discusión sobre la diversidad de sustratos que deben ser reportados en las metodologías sintéticas.^{43,44} La revolución quimioinformática está cambiando la perspectiva sobre cómo reportar, guardar y analizar la literatura científica.

Datos más representativos para entrenar los modelos pueden ser recolectados vía HTE. Sin embargo, los modelos construidos con datos de HTE suelen predecir efectivamente solo compuestos similares a los estudiados. Más aún, las reacciones evaluadas son con sustratos disponibles comercialmente, que son más simples que las moléculas con las que se trabaja en química medicinal. Para que un modelo sea útil debe poder predecir resultados de reacciones muy diferentes que con las que fue entrenado; en otras palabras, los algoritmos deben ser extrapolables o generalizables. Los modelos con datos HTE no suelen ser extrapolables a menos

que el espacio químico sea limitado previamente, como sucede con EDBO+ donde las condiciones a evaluar son previamente definidas.

Segundo dilema: ¿modelos simples o complejos?

Una razón del *boom* de inteligencia artificial es el avance en los métodos de optimización de los algoritmos basados en redes neuronales. Estos algoritmos suelen ser muy útiles cuando se tienen grandes bases de datos como las usadas para entrenar modelos de predicción de imágenes o de lenguaje como ChatGPT. A pesar de las décadas que llevamos estudiando la química orgánica solo hemos accedido a un pequeño porcentaje del espacio químico existente. Solo en el contexto de química medicinal, el espacio químico se ha calculado entre 10^{13} a 10^{180} moléculas de las cuales solo hemos sintetizado menos del 1%.⁴⁵ Más aún, si consideramos que cada reacción depende de múltiples parámetros (reactivos, disolvente, temperatura, estequiometría, etc.) nos encontramos con una cantidad abismal de posibilidades que abarcar. Esto implica que la cantidad de datos con la que contamos actualmente sea limitada y explica por qué modelos más simples que las redes neuronales han sido más exactos en algunos casos. Una estrategia para explorar este espacio químico tan vasto es utilizar modelos de optimización bayesiana o de aprendizaje activo para recolectar solo los datos que sean necesarios como fue mostrado anteriormente (Figura 6).

Tercer dilema: ¿cómo validar los modelos?

La validación de los algoritmos es clave para entender cómo y cuándo se debe aplicar un modelo. Esto se puede conocer evaluando el modelo en situaciones que se asemejen a los casos reales en los que se usará. El primer paso al construir un algoritmo es dividir los datos en dos grupos: uno que se usará para entrenar el modelo y el otro para evaluarlo en datos no vistos por el mismo. La forma más sencilla es dividir de forma aleatoria los datos; no obstante, esta evaluación no ofrecería información de la predictibilidad del modelo en situaciones reales. Por ello, se deben considerar también otras maneras de dividir los datos de acuerdo al problema a tratar. Por ejemplo, el grupo de Denmark entrenó un modelo para predecir la selectividad en la adición de un tiol a una imina en la presencia de diferentes catalizadores.⁴⁶ El modelo tuvo buenos resultados en el set de prueba con una división aleatoria de los datos. Sin embargo, en un contexto real es muy probable que las primeras reacciones que se ejecuten no lleguen a un porcentaje de exceso enantiomérico (% ee) alto. Por ello, el modelo también fue entrenado solo con las reacciones por debajo de 80% ee y satisfactoriamente tuvo buena predicción en el resto de reacciones con niveles más altos de % ee.

¿Cuándo y cómo usar los modelos?

Los modelos de inteligencia artificial deben ser usados con cautela debido a todos los factores que toman lugar durante su construcción y optimización. Los modelos no tienen que excluir necesariamente y, mucho menos, reemplazar al químico especialista en síntesis. Al contrario, los modelos deben ser usados como una herramienta y fuente de ideas para diseñar los experimentos especialmente en los casos con mayor incertidumbre. Los modelos de retrosíntesis pueden guiarnos para decidir qué enlaces formar primero y los modelos que proponen condiciones de reacción nos pueden inspirar en que parámetros evaluar.

Por ejemplo, al diseñar la síntesis de una molécula compleja es más seguro utilizar reacciones robustas en los últimos pasos. Las reacciones que dependen de características específicas del sustrato generan un riesgo que podría ser mitigado por algoritmos de inteligencia artificial. Esto ha sido demostrado por el grupo de Newhouse en la universidad de Yale al lograr una síntesis concisa de clovan-2,9-diona (**Figura 7**).^{47,48} En el caso de las reacciones de ciclación radicalaria 6-endo-trig, aunque funcionan en determinadas ocasiones, resulta difícil anticipar en qué ocasiones tendrán éxito. Por ello, el grupo de Newhouse entrenó un modelo de inteligencia artificial para predecir el rendimiento de una reacción mediada por una ciclación radicalaria 6-endo-trig. El modelo entrenado con la base de datos de Reaxys fue usado para predecir la desconexión más favorable para esta transformación. En efecto, la reacción fue exitosa con un rendimiento similar al predicho, logrando la síntesis de clovan-2,9-diona en solo cinco pasos.

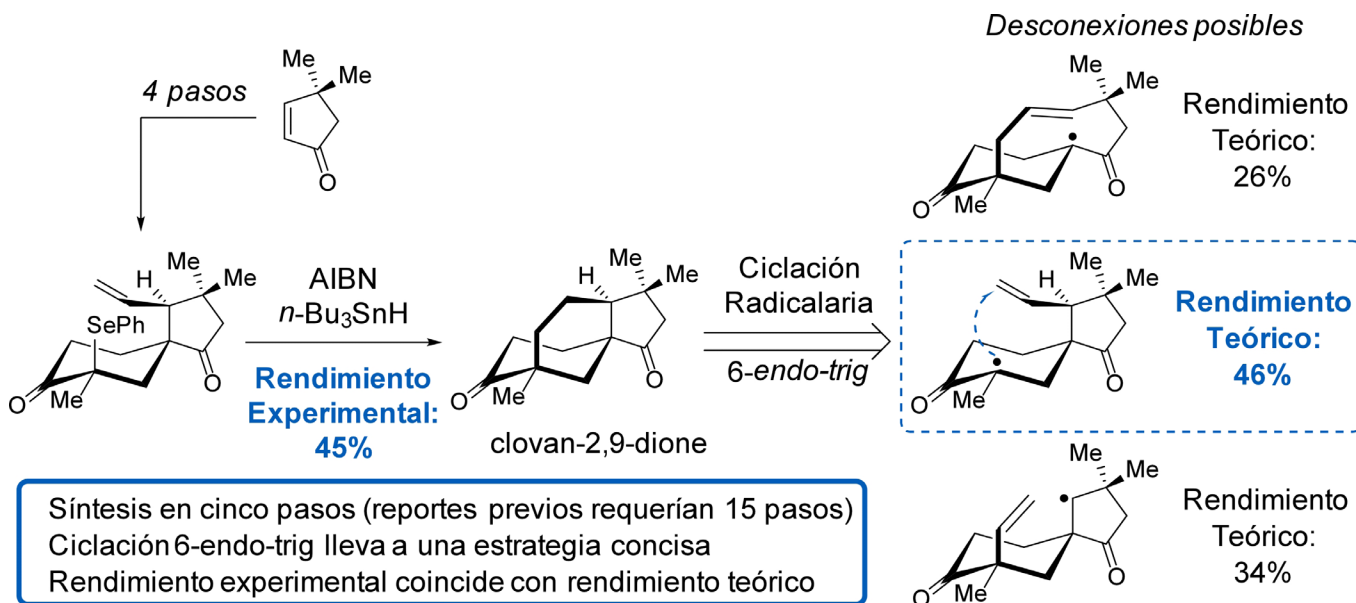


Figura 7. Síntesis concisa de clovan-2,9-dione guiado por un modelo de inteligencia artificial para predecir el rendimiento de ciclaciones radicalarias 6-endo-trig. Imagen adaptada de Zhang, P.; Eun, J.; Elkin, M.; Zhao, Y.; Cantrell, R. L.; Newhouse, T. R. *Neural Network Model Informs Total Synthesis of Clovane Sesquiterpenoids*. *ChemRxiv* 2021.

CONCLUSIONES

La automatización de la síntesis orgánica está en constante evolución, tanto en la invención de nuevas plataformas para ejecutar reacciones en paralelo, como en la construcción de algoritmos de inteligencia artificial para asistir con el diseño de experimentos. La miniaturización de las reacciones ha sido posible incluso en escalas de nanomoles permitiendo ejecutar más de mil reacciones al mismo tiempo usando mínimas cantidades de sustratos. Incluso se han desarrollado plataformas para ejecutar reacciones fotoquímicas y electroquímicas, y tecnologías para dispensar reactivos sólidos lo que ha sido una limitación de HTE hasta el momento. A su vez, se están implementando modelos de aprendizaje automático en síntesis orgánica para determinar qué reacción y qué condiciones utilizar. Los algoritmos pueden ser usados en casos de incertidumbre como en elegir qué enlace formar en el último paso al crear una molécula compleja. Los avances en HTE y en inteligencia artificial están cambiando la perspectiva de cómo sintetizar moléculas y prometen acelerar este proceso.

El entrenamiento de modelos de aprendizaje automático requiere una fuente masiva de datos. La recolección de datos experimentales implica una inversión significativa de recursos y tiempo, especialmente si no se cuentan con tecnologías como HTE. Por este motivo, se han creado iniciativas para organizar la información de las reacciones de artículos científicos en una plataforma abierta para todo el público como es el caso de la Open Reaction Database.⁴⁹ Esto, además, permite la estandarización de los

datos creando fuentes más confiables que las disponibles actualmente. El entrenamiento de los modelos también requiere experiencia en varias materias desde la química orgánica pasando por la quimioinformática hasta las ciencias de la computación. Por ello, las colaboraciones entre distintos grupos con experiencia en estos campos resultan ser muy fructíferas como es el Centro de Síntesis Asistida por Computadoras (C-CAS, por sus siglas en inglés) financiada por la Fundación Nacional de Ciencia de EEUU (NSF, por sus siglas en inglés). También existen colaboraciones entre diferentes industrias e instituciones académicas con el fin de delinear claramente los objetivos que se deben buscar en los modelos, así como también financiar investigaciones relacionadas.

Una gran oportunidad, y otra razón detrás de la explosión de la inteligencia artificial, es que cualquier persona puede desarrollar un modelo. El lenguaje de programación Python está jugando un papel clave en el avance del aprendizaje automático en química orgánica. Python se caracteriza por ser un lenguaje fácil de usar y existen muchas herramientas online gratuitas para aprender cómo funciona. Además, diversas plataformas abiertas al público se encuentran disponibles para ejecutar Python, como son Jupyter Notebooks y Colab financiada por Google. La gran

comunidad detrás de Python permite el intercambio de códigos, el mantenimiento y la mejora de este lenguaje de programación.

Los modelos de inteligencia artificial nos pueden guiar para decidir qué tipo de reacciones y qué condiciones usar para sintetizar los compuestos. HTE puede ser usado seguidamente para sintetizar las moléculas y, posteriormente, plataformas de micropurificación en paralelo nos permitirán purificar los productos. Sin embargo, antes de purificar las reacciones debemos analizar el resultado de las reacciones lo cual puede tomar considerable tiempo. Por ello, se han creado algoritmos para el análisis automático de reacciones en HTE con Python.⁵⁰ La inclusión de estos algoritmos completa el ciclo para la síntesis y reoptimización de reacciones dirigidas por inteligencia artificial (**Figura 8**). Aún queda camino por recorrer para mejorar cada parte del rompecabezas, pero los avances de los últimos años han empezado a mostrar cómo lucirá el laboratorio químico del futuro.

AGRADECIMIENTOS

Se agradece a Álvaro Gutiérrez Bonet, Pedro R. de Jesús Cruz, Abdellatif El Marrouni y Helen Mitchell por revisar este artículo antes de su envío.



Figura 8. Automatización de la síntesis orgánica: del diseño a la purificación de moléculas.

REFERENCIAS

- Cernak, T.; Gesmundo, N. J.; Dykstra, K.; Yu, Y.; Wu, Z.; Shi, Z. C.; Vachal, P.; Sperbeck, D.; He, S.; Murphy, B. A.; Sonatore, L.; Williams, S.; Madeira, M.; Verras, A.; Reiter, M.; Lee, C. H.; Cuff, J.; Sherer, E. C.; Kuethe, J.; Goble, S.; Perrotto, N.; Pinto, S.; Shen, D. M.; Nargund, R.; Balkovec, J.; DeVita, R. J.; Dreher, S. D. *Microscale High-Throughput Experimentation as an Enabling Technology in Drug Discovery: Application in the Discovery of (Piperidinyl)Pyridinyl-1H-benzimidazole Diacylglycerol Acyltransferase 1 Inhibitors*. *J. Med. Chem.* **2017**, *60* (9), 3594–3605.
- Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. *The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis*. *Acc. Chem. Res.* **2017**, *50* (12), 2976–2985.
- Montero Bastidas, J. R.; Marrouni, A. E.; Chiriack, M. I.; Struble, T.; Kalyani, D. *Accelerating Drug Discovery by High-Throughput Experimentation; Medicinal Chemistry Reviews; 2022; pp 443–463*.
- Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Diego, J. E. de; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. *The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future*. *Org. Process Res. Dev.* **2019**, *23* (6), 1213–1242.
- Lin, S.; Dikler, S.; Blincoe, W. D.; Ferguson, R. D.; Sheridan, R. P.; Peng, Z.; Conway, D. V.; Zawatzky, K.; Wang, H.; Cernak, T.; Davies, I. W.; DiRocco, D. A.; Sheng, H.; Welch, C. J.; Dreher, S. D. *Mapping the Dark Space of Chemical Reactions with Extended Nanomole Synthesis and MALDI-TOF MS*. *Science* **2018**, *361* (6402).
- Osipyanyan, A.; Shaabani, S.; Warmerdam, R.; Shishkina, S. V.; Boltz, H.; Dömling, A. *Automated, Accelerated Nanoscale Synthesis of Iminopyrrolidines*. *Angew. Chem. Int. Ed.* **2020**, *59* (30), 12423–12427.
- Mahjour, B.; Shen, Y.; Cernak, T. *Ultra-high-Throughput Experimentation for Information-Rich Chemical Synthesis*. *Acc. Chem. Res.* **2021**, *54* (10), 2337–2346.
- Gesmundo, N.; Dykstra, K.; Douthwaite, J. L.; Kao, Y. T.; Zhao, R.; Mahjour, B.; Ferguson, R.; Dreher, S.; Sauvagnat, B.; Sauri, J.; Cernak, T. *Miniaturization of Popular Reactions from the Medicinal Chemists' Toolbox for Ultra-high-Throughput Experimentation*. *Nat. Synth.* **2023**, 1–10.
- Gensch, T.; Glorius, F. *The Straight Dope on the Scope of Chemical Reactions*. *Science* **2016**, *352* (6283), 294–295.
- Collins, K. D.; Glorius, F. *A Robustness Screen for the Rapid Assessment of Chemical Reactions*. *Nat. Chem.* **2013**, *5* (7), 597–601.
- Kariofillis, S. K.; Jiang, S.; Żurański, A. M.; Gandhi, S. S.; Alvarado, J. I. M.; Doyle, A. G. *Using Data Science To Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed Cross-Coupling with Acetals as Alcohol-Derived Radical Sources*. *J. Am. Chem. Soc.* **2022**, *144* (2), 1045–1055.
- Dreher, S. D.; Krska, S. W. *Chemistry Informer Libraries: Conception, Early Experience, and Role in the Future of Cheminformatics*. *Acc. Chem. Res.* **2021**, *54* (7), 1586–1596.
- Kutchukian, P. S.; Dropinski, J. F.; Dykstra, K. D.; Li, B.; DiRocco, D. A.; Streckfuss, E. C.; Campeau, L. C.; Cernak, T.; Vachal, P.; Davies, I. W.; Krska, S. W.; Dreher, S. D. *Chemistry Informer Libraries: A Cheminformatics Enabled Approach to Evaluate and Advance Synthetic Methods*. *Chem. Sci.* **2016**, *7* (4), 2604–2613.
- Bahr, M. N.; Morris, M. A.; Tu, N. P.; Nandkeolyar, A. *Recent Advances in High-Throughput Automated Powder Dispensing Platforms for Pharmaceutical Applications*. *Org. Process Res. Dev.* **2020**, *24* (11), 2752–2761.
- Bahr, M. N.; Damon, D. B.; Yates, S. D.; Chin, A. S.; Christopher, J. D.; Cromer, S.; Perrotto, N.; Quiroz, J.; Rosso, V. *Collaborative Evaluation of Commercially Available Automated Powder Dispensing Platforms for High-Throughput Experimentation in Pharmaceutical Applications*. *Org. Process Res. Dev.* **2018**, *22* (11), 1500–1508.
- Gesmundo, N. J.; Tu, N. P.; Sarris, K. A.; Wang, Y. *ChemBeads-Enabled Photoredox High-Throughput Experimentation Platform to Improve C(sp²)-C(sp³) Decarboxylative Couplings*. *ACS Med. Chem. Lett.* **2023**, *14* (4), 521–529.
- Tu, N. P.; Dombrowski, A. W.; Goshu, G. M.; Vasudevan, A.; Djuric, S. W.; Wang, Y. *High-Throughput Reaction Screening with Nanomoles of Solid Reagents Coated on Glass Beads*. *Angew. Chem. Int. Ed.* **2019**, *58* (24), 7987–7991.
- Lovering, F.; Bikker, J.; Humblet, C. *Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success*. *J. Med. Chem.* **2009**, *52* (21), 6752–6756.
- Buglioni, L.; Raymenants, F.; Slattery, A.; Zondag, S. D. A.; Noël, T. *Technological Innovations in Photochemistry for Organic Synthesis: Flow Chemistry, High-Throughput Experimentation, Scale-up, and Photoelectrochemistry*. *Chem. Rev.* **2022**, *122* (2), 2752–2906.
- Rein, J.; Annand, J. R.; Wismer, M. K.; Fu, J.; Siu, J. C.; Klapars, A.; Strotman, N. A.; Kalyani, D.; Lehnher, D.; Lin, S. *Unlocking the Potential of High-Throughput Experimentation for Electrochemistry with a Standardized Microscale Reactor*. *ACS Cent. Sci.* **2021**, *7* (8), 1347–1355.
- Dykstra, K. D.; Streckfuss, E.; Liu, M.; Liu, J.; Yu, Y.; Wang, M.; Kozłowski, J. A.; Myers, R. W.; Buevich, A. V.; Maletic, M. M.; Vachal, P.; Krska, S. W. *Synthesis of HDAC Inhibitor Libraries via Microscale Workflow*. *ACS Med. Chem. Lett.* **2021**, *12* (3), 337–342.
- Gesmundo, N. J.; Sauvagnat, B.; Curran, P. J.; Richards, M. P.; Andrews, C. L.; Dandliker, P. J.; Cernak, T. *Nanoscale Synthesis and Affinity Ranking*. *Nature* **2018**, *557* (7704), 228–232.
- Hendrick, C. E.; Jorgensen, J. R.; Chaudhry, C.; Strambeanu, I. I.; Brazeau, J. F.; Schiffer, J.; Shi, Z.; Venable, J. D.; Wolkenberg, S. E. *Direct-to-Biology Accelerates PROTAC Synthesis and the Evaluation of Linker Effects on Permeability and Degradation*. *ACS Med. Chem. Lett.* **2022**, *13* (7), 1182–1190.
- Soares, T. A.; Nunes-Alves, A.; Mazzolari, A.; Ruggiu, F.; Wei, G. W.; Merz, K. *The (Re)-Evolution of Quantitative Structure–Activity Relationship (QSAR) Studies Propelled by the Surge of Machine Learning Methods*. *J. Chem. Inf. Model.* **2022**, *62* (22), 5317–5320.
- Corey, E. J.; Wipke, W. T. *Computer-Assisted Design of Complex Organic Syntheses*. *Science* **1969**, *166*, 178–192.
- Grzybowski, B. A.; Szymkuć, S.; Gajewska, E. P.; Molga, K.; Dittwald, P.; Wolos, A.; Klucznik, T. *Chematica: A Story of Computer Code That Started to Think like a Chemist*. *Chem* **2018**, *4* (3), 390–398.
- Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E. P.; Touthkine, A.; Dittwald, P.; Startek, M. P.; Kirkovits, G. J.; Roszak, R.; Adamski, A.; Sieredzińska, B.; Mrksich, M.; Trice, S. L. J.; Grzybowski, B. A. *Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory*. *Chem* **2018**, *4* (3), 522–532.
- Lin, Y.; Zhang, Z.; Mahjour, B.; Wang, D.; Zhang, R.; Shim, E.; McGrath, A.; Shen, Y.; Brugger, N.; Turnbull, R.; Trice, S.; Jasty, S.; Cernak, T. *Reinforcing the Supply Chain of Umifenovir*

- and Other Antiviral Drugs with Retrosynthetic Software. *Nat. Commun.* **2021**, *12*, 7327.
29. Mikulak-Klucznik, B.; Gołębiowska, P.; Bayly, A. A.; Popik, O.; Klucznik, T.; Szymkuć, S.; Gajewska, E. P.; Dittwald, P.; Staszewska-Krajewska, O.; Beker, W.; Badowski, T.; Scheidt, K. A.; Molga, K.; Mlynarski, J.; Mrksich, M.; Grzybowski, B. A. Computational Planning of the Synthesis of Complex Natural Products. *Nature* **2020**, *588* (7836), 83–88.
 30. Lin, Y.; Zhang, R.; Wang, D.; Cernak, T. Computer-Aided Key Step Generation in Alkaloid Total Synthesis. *Science* **2023**, *379* (6631), 453–457.
 31. Hardy, M. A.; Nan, B.; Wiest, O.; Sarpong, R. Strategic Elements in Computer-Assisted Retrosynthesis: A Case Study of the Pukeanane Natural Products. *Tetrahedron* **2022**, *104*, 132584.
 32. 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.; Robinson, R. I.; Salentin, S.; Xing, L.; Jaakkola, T.; Green, William. H.; Barzilay, R.; Coley, C. W.; Jensen, K. F. Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *J. Med. Chem.* **2020**, *63* (16), 8667–8682.
 33. Coley, C. W.; III, D. A. T.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* **2019**, *365*, 557.
 34. Genheden, S.; Thakkar, A.; Chadimová, V.; Reymond, J. L.; Engkvist, O.; Bjerrum, E. AiZynthFinder: A Fast, Robust and Flexible Open-Source Software for Retrosynthetic Planning. *J. Cheminformatics* **2020**, *12* (1), 70.
 35. Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V. H.; Haeuselmann, R. A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. Predicting Retrosynthetic Pathways Using Transformer-Based Models and a Hyper-Graph Exploration Strategy. *Chem. Sci.* **2020**, *11* (12), 3316–3325.
 36. Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F. Regio-Selectivity Prediction with a Machine-Learned Reaction Representation and on-the-Fly Quantum Mechanical Descriptors. *Chem. Sci.* **2020**, *12* (6), 2198–2208.
 37. Struble, T. J.; Coley, C. W.; Jensen, K. F. Multitask Prediction of Site Selectivity in Aromatic C–H Functionalization Reactions. *React. Chem. Eng.* **2020**, *5* (5), 896–902.
 38. Mahjour, B.; Hoffstadt, J.; Cernak, T. Designing Chemical Reaction Arrays Using Phactor and ChatGPT. *Org. Process Res. Dev.* **2023**, *27* (8), 1510–1516.
 39. Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96.
 40. Torres, J. A. G.; Lau, S. H.; Anchuri, P.; Stevens, J. M.; Tabora, J. E.; Li, J.; Borovika, A.; Adams, R. P.; Doyle, A. G. A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J. Am. Chem. Soc.* **2022**, *144* (43), 19999–20007.
 41. Taylor, C. J.; Felton, K. C.; Wigh, D.; Jeraal, M. I.; Grainger, R.; Chessari, G.; Johnson, C. N.; Lapkin, A. A. Accelerated Chemical Reaction Optimization Using Multi-Task Learning. *ACS Cent. Sci.* **2023**, *9*, 957–968.
 42. Felton, K. C.; Rittig, J. G.; Lapkin, A. A. Summit: Benchmarking Machine Learning Methods for Reaction Optimisation. *Chem. Methods* **2021**, *1* (2), 116–122.
 43. Maloney, M. P.; Coley, C. W.; Genheden, S.; Carson, N.; Helquist, P.; Norrby, P. O.; Wiest, O. Negative Data in Data Sets for Machine Learning Training. *J. Org. Chem.* **2023**, *88* (9), 5239–5241.
 44. Kozłowski, M. C. On the Topic of Substrate Scope. *Org. Lett.* **2022**, *24* (40), 7247–7249.
 45. Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the Size of Drug-like Chemical Space Based on GDB-17 Data. *J. Comput.-Aided Mol. Des.* **2013**, *27* (8), 675–679.
 46. Henle, A. F. Z. J. J.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, No. 363.
 47. Zhang, P.; Eun, J.; Elkin, M.; Zhao, Y.; Cantrell, R. L.; Newhouse, T. R. A Neural Network Model Informs the Total Synthesis of Clovane Sesquiterpenoids. *Nat. Synth.* **2023**, *2* (6), 527–534.
 48. Newhouse, T. R.; Zhang, P.; Eun, J.; Elkin, M.; Zhao, Y.; Cantrell, R. L. A Neural Network Model Informs Total Synthesis of Clovane Sesquiterpenoids. *ChemRxiv* **2021**.
 49. Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143* (45), 18820–18826.
 50. Mason, J.; Wilders, H.; Fallon, D. J.; Thomas, R. P.; Bush, J. T.; Tomkinson, N. C. O.; Rianjongdee, F. Automated LC-MS Analysis and Data Extraction for High-Throughput Chemistry. *Digit. Discov.* **2023**, *2* (6), 1894–1899.

BIBLIOGRAFÍA ESENCIAL

Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **2017**, *50* (12), 2976–2985.

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.; Robinson, R. I.; Salentin, S.; Xing, L.; Jaakkola, T.; Green, William. H.; Barzilay, R.; Coley, C. W.; Jensen, K. F. Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *J. Med. Chem.* **2020**, *63* (16), 8667–8682.

Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143* (45), 18820–18826.