

© 2024 Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Zaragoza.

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

TIP Revista Especializada en Ciencias Químico-Biológicas, 27: 1-17, 2024.

<https://doi.org/10.22201/fesz.23958723e.2024.652>

Quimioinformática, Inteligencia Artificial y la Química de Alimentos

Juan F. Avellaneda-Tamayo¹, Andrés Sánchez-Ruiz²,
Biljana Savic², José L. Medina-Franco^{1a} y Gonzalo Colmenarejo^{2b}

¹Grupo de Investigación DIFACQUIM, Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria 04510, Alcaldía Coyoacán, Ciudad de México, México

²Unidad de Bioestadística y Bioinformática, IMDEA Alimentación, CEI UAM+CSIC, 28049, Madrid, España. E-mails: ^amedinajl@unam.mx; ^bgonzalo.colmenarejo@imdea.org

RESUMEN

La Quimioinformática y la Inteligencia Artificial (IA) son disciplinas con métodos que, actualmente, contribuyen al desarrollo de varias áreas del conocimiento de la Química, por medio del almacenamiento, la organización, la búsqueda de datos químicos, el procesamiento y el modelado, para generar información y conocimiento a nivel molecular de las relaciones estructura-propiedad de los compuestos químicos existentes y determinar las propiedades de los nuevos, a partir del diseño con base en un perfil deseado. Todas estas técnicas se han utilizado en la Química de Alimentos, por lo que el objetivo de este artículo es analizar las bases de los métodos Quimioinformáticos y de la IA en su aplicación al estudio de los compuestos químicos de los alimentos.

Palabras clave: inteligencia artificial, química de alimentos, quimioinformática, relaciones estructura-propiedad, representación molecular.

Chemoinformatics, Artificial Intelligence, and Food Chemistry

ABSTRACT

In the last few years, computational methods covering Chemoinformatics and Artificial Intelligence (AI) have had a significant impact on the development of Chemistry in its multiple areas. Applications range from storage, organization, and searching of chemical data, to their processing and modeling to generate information and knowledge. Examples of large applications of chemoinformatic methods and AI in Chemistry are understanding at a molecular level the structure-property relationship of known compounds, the prediction of properties for new compounds, and the design of new chemical structures with a desired property profile. All these techniques have been applied to Food Chemistry. The objective of this review is then to present and discuss in a critical approach the methods of Chemoinformatics and AI, as well as their applications in the study of the chemical compounds present in foods.

Keywords: artificial intelligence, food chemistry, chemoinformatics, structure-property relationships, molecular representation.

Abreviaturas: IA: Inteligencia Artificial; InChI: Identificador Químico Internacional; PCA: Análisis de Componentes Principales; SMILES: Especificación de Introducción Lineal Molecular Simplificada; RNA: Redes Neuronales Artificiales; t-SNE: Encaje Estocástico de Vecinos t-Distribuido.

INTRODUCCIÓN

La Quimioinformática (Figura 1) es una disciplina cuyo objetivo es la aplicación de los recursos informáticos para resolver problemas en Química (Gasteiger, 2020; Miranda-Salas *et al.*, 2023). Estos problemas o aplicaciones abarcan frecuentemente: **i**) la falta de organización de la información en bases de datos moleculares (p. ej., quimiotecas) que facilite su búsqueda, acceso y actualización; este proceso es relevante y un desafío por el incremento exponencial de los datos químicos; **ii**) el análisis de las relaciones que hay entre las estructuras químicas y una o varias propiedades de interés (por ejemplo, el sabor, las propiedades biológicas, farmacéuticas o farmacológicas, etc.) de nuevos compuestos o de compuestos químicos existentes, y **iii**) la predicción de propiedades a partir de relaciones complejas. A su vez, la predicción de propiedades está directamente relacionada con otras dos acciones muy comunes en la Quimioinformática: **A**) diseñar nuevos compuestos *de novo*; y **B**) filtrar colecciones de compuestos conocidos para seleccionar los que tienen las propiedades deseadas; esta es una práctica empleada durante varios años, pero en constante mejora y optimización, conocida como cribado (del inglés, *screening*), tamizaje o evaluación computacional de bases de datos moleculares.

La “Informática de la Química de Alimentos” (*FoodInformatics*) es una subdisciplina definida como la aplicación de métodos informáticos para resolver problemas en el campo de la Química de Alimentos. Entre sus aplicaciones están la sistematización de la información en bases de datos de compuestos químicos alimentarios, así como el análisis y la predicción de sus relaciones estructura-propiedad (por ejemplo, el sabor, el olor y las actividades biológicas, etc.). Actualmente, se habla por ejemplo de un diseño racional de odorantes contenidos en los alimentos o que se pueden utilizar de forma segura como aditivos alimentarios. El campo de la Informática en la Química de Alimentos comenzó su desarrollo en las últimas décadas; las primeras revisiones datan del año 2009 (Martínez-Mayorga & Medina-Franco, 2009), así como un estudio quimioinformático de los compuestos GRAS (*Generally Recognized as Safe*, Aditivos de Alimentos considerados Seguros) (Martínez-Mayorga, Peppard, López-Vallejo, Yongye & Medina-Franco, 2013); sin embargo, es alrededor del año 2014 que empieza a formalizarse y está en vías de una consolidación (Martínez-Mayorga & Medina-Franco, 2014).

El objetivo del presente artículo es discutir las bases de la Quimioinformática, la Inteligencia Artificial (IA) (Tseng,



Figura 1. Aplicaciones de la química informática en la química de los alimentos. La Figura es de Creatividad Personal.

Chuang & Appell, 2023), y la Inteligencia Aumentada (Saldívar-González, Fernández-de Gortari & Medina-Franco, 2023) aplicadas al estudio de los compuestos químicos presentes en los alimentos.

CONCEPTOS GENERALES DEL ESTUDIO QUÍMICO DE LOS ALIMENTOS

Los primeros estudios modernos en el campo de la Química de Alimentos se remontan al siglo XIX, cuando algunos científicos de la época se interesaron en los componentes de los productos alimenticios. Entre ellos Justus von Liebig, quien, al realizar un análisis bioquímico, dividió a los alimentos en plásticos (nitrogenados o proteicos) y respiratorios (almidón, azúcar, gomas y grasas), reconoció su importancia en la alimentación y lo relevante de respetar determinadas proporciones. Desde entonces, el conocimiento sobre los alimentos continúa su progreso (Boatella, 2017). A continuación se da una breve descripción de los distintos grupos moleculares presentes en los alimentos.

Macronutrientes y nutrición

Los macronutrientes son compuestos químicos con aporte nutricional consumidos en cantidades significativas en la alimentación. Estos incluyen carbohidratos, proteínas y grasas, como fuentes de energía; son los componentes básicos de las células; además del agua, el oxígeno, la fibra y los elementos como el calcio, el magnesio, el sodio, el potasio y el cloro, fundamentales para la vida pero que no proporcionan energía. El análisis de los macronutrientes y su importancia en la nutrición humana ha sido objeto de una exhaustiva investigación científica, cuyo resultado es un entendimiento profundo de los elementos esenciales de la dieta que respaldan la vida y la salud; son los componentes dietéticos básicos de nuestra alimentación con una función central en la regulación de los procesos fisiológicos vitales.

La importancia de los macronutrientes va más allá de su valor calórico. Se extiende a las complejas vías metabólicas y las interacciones bioquímicas en el organismo. Por tanto, comprender en profundidad las propiedades químicas y estructurales de los macronutrientes resulta esencial para apreciar el impacto significativo que tienen en la nutrición y en la salud (Caballero, Trigo & Finglas, 2003).

Micronutrientes y productos naturales (metabolitos secundarios)

Los micronutrientes son minerales como el cobre, el zinc y el hierro, además de todas las vitaminas, que se consumen en pequeñas cantidades, pero su papel es crucial en la salud y el bienestar. Su presencia en los alimentos es un tema de gran interés en las ciencias de la nutrición y de la salud (Streit, 2023).

Los alimentos también contienen productos naturales, estos son compuestos químicos conocidos como metabolitos

secundarios, por lo general con actividad biológica e impacto en la salud. Los productos naturales atraen cada vez más la atención por su diversidad química y su bioactividad, lo que resulta en una exploración a profundidad tanto de sus funciones como de una posible aplicación. Estos esfuerzos buscan aprovechar su utilidad como ingredientes en los productos alimenticios, por sus compuestos bioactivos para el ámbito de la nutrición e incluso para el ámbito farmacéutico. Un ejemplo es la cafeína, cuyo quimiotipo se ha utilizado en el diseño de inhibidores para distintas áreas terapéuticas (Faudone, Arifi & Merk, 2021). Las funciones versátiles que tienen los productos naturales subrayan su importancia y las diversas oportunidades que ofrecen para mejorar la salud, la nutrición y el bienestar (Romano & Tatonetti, 2019; Saldívar-González, Aldas-Bulos, Medina-Franco & Plisson, 2022).

Biotransformaciones y metabolitos

La biotransformación es un proceso metabólico fundamental que ocurre principalmente en el hígado y que tiene como objetivo facilitar la eliminación de sustancias, ya sean de origen externo (exógenas) o interno (endógenas). Este proceso implica una serie de reacciones químicas que modifican las estructuras moleculares de estas sustancias. La biotransformación es esencial para regular y eliminar compuestos químicos del cuerpo, contribuyendo a mantener un equilibrio en la función y la salud del organismo (Almazroo, Miah & Venkataramanan, 2017).

El análisis y predicción del perfil de los metabolitos en diversos grupos de pacientes es esencial para obtener información de los factores que inciden en la salud, como la progresión de la enfermedad y el metabolismo de los fármacos. En este ámbito, Xusheng Chen y autores propusieron tres metabolitos de los ácidos grasos como biomarcadores relacionados con la reducción de la eficacia de los medicamentos basados en el platino para combatir el cáncer de vejiga. En situaciones como esta son necesarias las herramientas que ofrece la Quimioinformática para determinar las causas de la reducción de la efectividad del medicamento (Chen, Zhang, Liao & Zhao, 2023; Ghosh, Zhang, Ghosh & Kechris, 2020).

La relación entre los alimentos y la Epigenética

La Epigenética estudia la regulación de la expresión de los genes más allá de su secuencia. Los factores que afectan su regulación están relacionados con la etapa de desarrollo del organismo y sus condiciones de exposición, como variables ambientales, la alimentación o la actividad física (Dupont, Armant & Brenner, 2009).

En el campo de la nutrición, la Epigenética tiene una importancia excepcional, pues es conocido que los nutrientes y componentes bioactivos de los alimentos pueden modificar los fenómenos epigenéticos y alterar la expresión de los genes a nivel transcripcional (Açar & Akbulut, 2023; Juárez-Mercado,

Avellaneda-Tamayo, Villegas-Quintero, Chávez-Hernández, López-López & Medina-Franco, 2024; Voruganti, 2023). Un ejemplo clásico de cómo la alimentación puede afectar a la Epigenética es el caso de las abejas. Este insecto utiliza la alimentación para definir el desarrollo entre una abeja reina y obrera: las abejas reinas son alimentadas con jalea real mientras que las obreras son alimentadas con una mezcla de polen y néctar. Asimismo, en humanos se sabe que compuestos como la vitamina B12 presente en carnes, pescado, huevo o leche, es uno de los cofactores esenciales en los procesos epigenéticos como la metilación (Choi & Friso, 2010).

Nutracéuticos

La Nutracéutica está en desarrollo y rápido crecimiento gracias a la creciente conciencia de los posibles beneficios de los suplementos dietéticos utilizados en conjunto con los tratamientos farmacológicos (Tambe, Jain, Amin, Mali & Cruz, 2023). Se denomina nutracéutico en este contexto a los compuestos químicos de alimentos o derivados que ofrecen beneficios fisiológicos, normalmente de tipo preventivo o coadyuvando a terapias farmacológicas (Chopra *et al.*, 2022).

Para algunas personas, la incorporación de los nutracéuticos en su dieta es una forma de lograr una alimentación saludable. No obstante, el desarrollo y el uso de los nutracéuticos forman parte de un debate más amplio sobre cómo las personas adaptan su patrón de alimentación en el contexto de condiciones ambientales cambiantes, estilos de vida en evolución y el creciente intercambio global de ideas y productos (Heinrich, 2019).

INTRODUCCIÓN A LA CIENCIA DE DATOS E IA PARA EL ESTUDIO QUÍMICO DE LOS ALIMENTOS

La IA comprende un vasto campo de trabajo acerca de la comprensión y la construcción de máquinas inteligentes, que calculen cómo actuar de forma efectiva y segura en una amplia gama de situaciones (Russell, Russell & Norvig, 2020). El primer trabajo en IA moderna es el de Warren McCulloch y Walter Pitts en 1943, sobre el modelado matemático de neuronas artificiales (McCulloch & Pitts, 1943). Desde entonces la IA continúa en evolución dividiéndose en subdisciplinas como: la IA simbólica, los sistemas expertos, el aprendizaje automático (*machine learning*) y más reciente, con la explosión en la capacidad de cálculo, el aprendizaje automático profundo (*deep learning*); actualmente todas muestran su gran utilidad y versatilidad a la hora de resolver una amplia gama de tareas. Este progreso se ha visto a su vez acompañado por un aumento exponencial en la cantidad, la variedad y la velocidad con la que se generan datos, dando lugar a lo que se conoce como *Big Data* (Zhu, 2020). Sin embargo, el crecimiento ha superado los recursos para procesar y analizar la información con eficacia, lo que deja en muchos casos a los métodos tradicionales limitados o insolventes, y plantea el reto de desarrollar metodologías más útiles.

El nacimiento de la Ciencia de Datos se da en este contexto por tratarse de una disciplina con herramientas y técnicas que extraen el conocimiento a partir de cantidades masivas de información, y con apoyo de la Estadística (Donoho, 2017). Por primera vez en los años sesenta John Tukey la definió (Tukey, 1962). Es un área con la ventaja competitiva que ofrece el aprendizaje automático para detectar patrones complejos y relaciones no lineales, llevando a término tareas de clasificación, regresión o de detección de anomalías (Donoho, 2017).

Es de reciente aparición el término “Inteligencia Aumentada” para referirse a la asociación entre los métodos y con las ventajas de funcionar con la creatividad, la intuición y el criterio humano (IA + IH) (Bazoukis, Hall, Loscalzo, Antman, Fuster & Armoundas, 2022; Saldívar-González *et al.*, 2023).

Ejemplos de aplicaciones recientes de la IA al estudio químico de los alimentos

En la primera década de este siglo, se empezó a popularizar el uso de técnicas de aprendizaje automático y Ciencia de Datos en el campo de la Química de Alimentos. Algunos ejemplos de esto son: i) el trabajo de un análisis descriptivo de las características moleculares del espacio químico de los compuestos responsables de los efectos sensoriales como el olor y el sabor (Ruddigkeit & Reymond, 2014); ii) el propuesto por H. Hopfer y autores, centrado en medir las alteraciones de las propiedades organolépticas del vino en función de la temperatura y el método de conservación (Hopfer, Buffon, Ebeler & Heymann, 2013). En el mismo ámbito se desarrolló una métrica para cuantificar las cualidades odorantes de los compuestos químicos con base en sus descriptores moleculares (Haddad, Khan, Takahashi, Mori, Harel & Sobel, 2008) y iii) la publicación de Sprous y Salemme, interesados en detectar las cualidades que compartían las moléculas responsables del sabor y de los fármacos (Sprous & Salemme, 2007).

A medida que avanzó la segunda década del año 2000, surgieron métodos de aprendizaje profundo con aplicación en investigación en la Química de Alimentos, de la mano de librerías de programación de libre acceso como *Tensorflow* o *Pytorch*; seguida de propuestas novedosas como la subdisciplina del emparejamiento de alimentos (*Food-pairing*). *FlavorGraph* es un buen ejemplo (Park, Kim, Kim, Spranger & Kang, 2021) por tratarse de un algoritmo de aprendizaje profundo cuya finalidad es sugerir combinaciones de productos alimenticios que sean equilibrados y aceptables al consumo, con base en su coocurrencia en recetas. En el ámbito culinario también se encuentra la publicación de Morales-Garzón y autores, quienes proponen un modelo basado en las representaciones de palabras (*word embeddings*) cuya finalidad es poder adaptar recetas de cocina a las preferencias del usuario (Morales-Garzón, Gómez-Romero & Martín-Bautista, 2021). Este tipo de funciones son muy útiles, por ejemplo, en la adaptación de alimentos a dietas

vegetarianas. En lo que se refiere a la evaluación del estado de deterioro de los alimentos, la revisión de Hassan Anwar y autores presenta varios métodos de aprendizaje automático y de aprendizaje profundo para detectar las condiciones de distintos alimentos como la carne, la leche, el pescado, el café, el té, la margarina y ciertos tipos de aceites comestibles (Anwar, Anwar & Murtaza, 2023). Este tipo de algoritmos es conocido popularmente como narices robóticas.

Sin embargo, como se verá, en esta revisión, no todos los tipos de IA en la Química de Alimentos son utilizados para tareas culinarias (Tseng *et al.*, 2023).

Bases de datos disponibles

Las bases de datos actuales son el resultado de un gran esfuerzo realizado durante la última década, ya que en años anteriores la información disponible (y gratuita) era más limitada (Martínez-Mayorga, Peppard, Ramírez-Hernández, Terrazas-Álvarez & Medina-Franco, 2014). Esto suponía un problema, ya que una de las principales dificultades al entrenar cualquier modelo de IA es la gran cantidad (y calidad) de datos que son requeridos (Martínez-Mayorga, Rosas-Jiménez, González-Ponce, López-López, Neme & Medina-Franco, 2024). En la actualidad existen numerosos repositorios de acceso abierto, y de acuerdo al

tipo de información que contienen, se dividen en tres grupos (Tseng *et al.*, 2023): **i)** bases de datos con información sobre la composición química de los alimentos, incluyendo nutrientes, compuestos bioactivos y compuestos tóxicos; **ii)** bases de datos con información sobre los compuestos químicos presentes en los alimentos, incluyendo su estructura química, propiedades y función; por último, **iii)** los repositorios con información de las moléculas que contribuyen al sabor y olor de los alimentos. La Tabla I resume las bases de datos públicas más conocidas en la Química de Alimentos.

A pesar de los grandes avances logrados en los últimos años, es notoria la necesidad de aumentar el contenido de las bases de datos confiables y de acceso abierto para acelerar el avance en el campo de la Quimioinformática de alimentos, y robustecer la simbiosis entre los trabajos en IA y el acceso abierto (Miljković & Medina-Franco, 2024). Esto se evidencia al compararlo con la cantidad y el tamaño de las bases de datos que existen para el estudio de los productos naturales (Sorokina & Steinbeck, 2020).

APLICACIONES DE LA QUIMIOINFORMÁTICA Y LA QUÍMICA COMPUTACIONAL

Tan amplio como la Química aplicada al estudio de los alimentos, o el avance de la Ciencia de Datos en las últimas décadas,

Tabla I. Ejemplos de bases de datos de acceso libre más comunes en la Química de Alimentos (modificado de Tseng *et al.*, 2023).

Base de datos	Tipos de base de datos	Descripción
FoodData Central	Composición de los alimentos.	Creada y en mantenimiento por el Departamento de Agricultura de los Estados Unidos. Debe su origen a la fusión de cinco bases de datos. Contiene información nutricional detallada de la composición de una amplia variedad de alimentos.
FooDB	Composición de alimentos y de los compuestos químicos.	Contiene datos de la composición y los compuestos químicos de los alimentos.
FlavorDB	Propiedades organolépticas.	Registra los compuestos que contribuyen a los sabores y aromas de los alimentos.
InFoods	Composición de los alimentos.	Permite obtener detalles de la composición de los alimentos segmentada por continentes y países.
EuroFIR	Composición de los alimentos.	Refiere al contenido energético de los macronutrientes, los minerales y las vitaminas presentes en los alimentos.
BitterDB	Compuestos químicos y propiedades organolépticas.	Aporta conocimiento de los compuestos amargos según diferentes criterios. Incluye otras funciones como la exploración de receptores humanos del sabor amargo, entre otras.
VirtualTaste	Compuestos químicos y propiedades organolépticas.	Además de ser una base de datos del sabor de los compuestos naturales y de las drogas, también tiene un modelo capaz de asignar un sabor a un compuesto químico.
ChEMBL/PubChem/ChemSpider	Compuestos químicos y actividades biológicas	Aunque no presenta información química específica de los compuestos alimenticios, los incluye, como son: su estructura, sus propiedades físicas, cinética molecular, actividad química e interacción con las proteínas, entre otras.

son las herramientas que presenta la Quimioinformática en su versatilidad, en su aplicación y en sus ejemplos de éxito. En esta sección se describen conceptos clave de la Quimioinformática, que son utilizadas en la Química de Alimentos.

¿Qué es la Quimioinformática?

El desarrollo de la ciencia aporta una cantidad considerable de datos experimentales, de parámetros y de resultados, así como de cálculos teóricos. En Química, son ejemplos de lo anterior, las condiciones de una reacción y sus correspondientes resultados en rendimiento y especificidad, el desplazamiento químico de una señal en resonancia magnética nuclear para sustancias químicas homólogas, la energía libre de Gibbs calculada para un conjunto de átomos, o una actividad celular específica con relación a un estímulo. El almacenamiento, tratamiento y procesamiento de los datos es lo que permite extraer información de forma inductiva a partir de ellos. La búsqueda de patrones y el modelado estadístico necesita por tanto de la construcción y mantenimiento de las bases de datos. Estos procesos son a lo que se denomina Informática. Cuando la información está relacionada con la Química, se habla de Química Informática o *Quimioinformática*. Así mismo, por medio de la generalización de la información obtenida a partir de los datos, es posible la construcción de modelos y teorías, que se transforman en conocimiento. El término “Quimioinformática” lo usó por primera vez Frank Brown en el año 1998, definido como la “mezcla de recursos informáticos para transformar los datos químicos en información y la información en conocimiento, con la pretensión de tomar mejores decisiones y rápidamente” (Brown, 1998). Si bien, originalmente se aplicó de forma específica al diseño y descubrimiento de fármacos, actualmente se ha extendido a otras áreas como la síntesis orgánica, la Química Analítica y la Química de los Alimentos, entre otras aplicaciones (López-López, Bajorath & Medina-Franco, 2021).

La Quimioinformática se traslapa de forma importante con disciplinas como la Química Computacional, la Quimiometría, y la Bioinformática, pero utiliza representaciones específicas, así como métodos y herramientas propias que la diferencian de las demás. Los objetivos y alcances de la Quimioinformática son marcadamente distintos a los otros campos mencionados (López-López *et al.*, 2021).

Un desarrollo didáctico y en español de las herramientas abordadas en este artículo puede consultarse en su versión en línea (Saldívar-González *et al.*, 2024).

¿Cómo las computadoras ven e interpretan las moléculas?

Quienes desarrollan y emplean la Química, han armonizado la representación de los compuestos a lo largo de los años. Se han utilizado abreviaturas para simbolizar los átomos de un elemento en particular, ubicadas en vértices de grafos, así como líneas que representan la conectividad y las propiedades

del enlace químico, además de describir la disposición de los electrones en el enlace. Sin embargo, un problema más complejo es la representación molecular adecuada para que sea interpretada por las computadoras con el fin de que la modifiquen y la usen en diferentes tipos de cálculos (David, Thakkar, Mercado & Engkvist, 2020). Esta representación consiste en un proceso y lenguaje estandarizado y sistemático para ingresar la información, procesarla y almacenarla, cubrir diferentes propiedades de las moléculas como su conectividad, su disposición tridimensional, sus cargas y sus propiedades fisicoquímicas, entre otras (ver Figuras 2A a 2D). Las representaciones deben ser compactas para almacenar una gran cantidad de datos moleculares, que permitan el acceso rápido en las búsquedas entre grandes cantidades de información. A pesar de que las primeras aproximaciones a una representación legible por computadoras estuvieron enfocadas al trabajo con moléculas orgánicas pequeñas (menores a 1,000 Da = 1,000 g * mol⁻¹), actualmente se han hecho comunes las que representan macromoléculas como los macrociclos, los péptidos, las proteínas y las reacciones químicas. Esto favorece las aplicaciones computacionales como las ciencias ómicas, la Bioinformática y el modelado de proteínas para el desarrollo y descubrimiento de fármacos.

Grafos moleculares

Si bien la representación gráfica de las moléculas se realiza a través de notaciones o fórmulas condensadas, a través de múltiples formatos, la representación de las moléculas en las computadoras se da internamente en forma de grafos moleculares del tipo $G = (V, E)$. Esta aproximación coincide con la imagen habitual que tiene un químico de la estructura molecular. Los grafos moleculares están compuestos por nodos V y aristas E . Cada uno de los nodos determina la naturaleza del átomo, por ejemplo, a modo de número atómico, e incluye propiedades como el tipo de enlaces que puede adoptar, las coordenadas tridimensionales, la carga, la configuración isotópica o la quiralidad. Las aristas, por su parte representan la conexión entre los nodos del grafo y representa al enlace químico en las moléculas. Sin embargo, la naturaleza del enlace no está descrita por la arista en sí misma, sino por las propiedades del nodo. El orden de procesamiento del grafo es determinado a través de diferentes algoritmos de la teoría de grafos, que de forma consistente dan lugar a la misma molécula. Para esto, diferentes arquitecturas de algoritmos de lectura son implementados por los programas de procesamiento de información química disponibles en la actualidad (Bondy & Murty, 1976). La Figura 2A es un ejemplo de una representación bidimensional de tipo grafo para la molécula de ácido salicílico.

Si bien el grafo es un tipo de formato descriptivo y específico que representa la estructura de una molécula en aplicaciones computacionales, el tamaño de las moléculas con las que trabaja de forma óptima es limitado debido a un alto uso de recursos de cómputo (procesamiento) y de almacenamiento (memoria

lineal, en que se especifica con números el inicio y el final de una estructura cíclica, y usa letras minúsculas para las subestructuras aromáticas. Consiste en un sistema legible e intuitivo para las personas, pero no es unívoco por depender del orden en que se numeran los átomos en la estructura. Este formato permite especificar enlaces dobles y triples, quiralidad, multiplicidad de grupos funcionales y estados de carga formal. Además es legible por diferentes librerías quimioinformáticas, como *RDKit* (Landrum, s/f) y *MolVS* (Swain, s/f). Los SMILES pueden ser transformados a un formato unívoco por medio de diferentes métodos (Weininger, 1988). El más usado actualmente consiste en enumerar la estructura a partir del InChI, que sí es una descripción única.

El formato InChI apareció en el año 2006 como una respuesta al llamado de la Unión Internacional de Química Pura y Aplicada (IUPAC, por sus siglas en inglés) para la creación de una notación estándar de uso informático que fue realizado en el año 1949 (Division of Chemistry and Chemical Technology, 1964). Consiste en un sistema de código abierto para simbolizar moléculas en cadenas de texto de 1D. La codificación está dividida en capas (p. ej. la principal, de estereoquímica, de información isotópica) y cada capa en bloques (Heller *et al.*, 2015). En el caso de la capa principal, por ejemplo, contiene la información de la fórmula química, la conectividad atómica y los átomos de hidrógeno. En esta notación, cada código es único, a diferencia de los SMILES.

Los InChIKey son códigos de veintisiete caracteres que resumen a la capa principal, la de estereoquímica e isótopos, y los protones presentes, generados a partir de un proceso matemático del InChI, diseñados para ser identificadores y estructuras únicas de las moléculas (Southan, 2013). Sin embargo, se pueden presentar ambigüedades en su uso, por lo que no necesariamente es posible regresar al InChI a partir del InChIKey, y aún con menos frecuencia al grafo molecular (Pletnev, Erin, McNaught, Blinov, Tchekhovskoi & Heller, 2012).

Andamios moleculares

Este tipo deriva del término en inglés, *molecular scaffold*, definido como el sistema cíclico base de la molécula junto con sus conectores, y busca generalizar algunas características presentes en todo el grupo de compuestos químicos que lo contienen. Estas características comprenden propiedades fisicoquímicas, biológicas (p. ej. bioactividad), etc.

Se han desarrollado aproximaciones para el análisis de los andamios moleculares (Langdon, Brown & Blagg, 2011), la más usada es la de Bemis y Murcko (Figura 2A). Consiste en romper todos los enlaces alquílicos de la molécula, para conservar los sistemas de anillos, así como los conectores entre ellos; es decir que todas las cadenas laterales de los ciclos son eliminadas, quedando únicamente los sistemas cíclicos con sus conectores (Bemis & Murcko, 1996).

Propiedades y descriptores

Posterior a representar una estructura química para que sea registrada por una computadora, se describe numéricamente por medio de “descriptores” que capturan cuantitativamente parte de la información o características de la molécula. Por ejemplo, su tamaño, el volumen (total y polar), la flexibilidad, la energía, la solubilidad o la lipofilidad (ver Figura 2D). Si bien algunos de los descriptores se miden experimentalmente (por ejemplo, la solubilidad en agua), otros se calculan de forma computacional. Ejemplos de descriptores comunes usados en la Quimioinformática pueden clasificarse en: **A)** constitucionales; **B)** fisicoquímicos (Lipinski, Lombardo, Dominy & Feeney, 1997); **C)** cuánticos (Grisoni, Ballabio, Todeschini & Consonni, 2018; Todeschini, Consonni, Mannhold, Kubinyi & Folkers, 2010); **D)** huellas digitales moleculares (Yang, Cai, Zhao, Xie & Chen, 2022); y **E)** indicadores compuestos, como de compuesto tipo fármaco (Gleeson, 2008; Hughes *et al.*, 2008; Shultz, 2013, 2014; Tian, Wang, Li, Li, Xu & Hou, 2015; Veber, Johnson, Cheng, Smith, Ward & Kopple, 2002), de compuesto tipo producto natural (Ertl, Roggo & Schuffenhauer, 2008), y la perspectiva de desarrollo del indicador de compuesto tipo alimento.

Los diferentes descriptores para la medición y el cálculo dependen de los objetivos del estudio o del propósito por el que se van a utilizar (Sheridan & Kearsley, 2002); seleccionarlos está en función de la naturaleza de las moléculas (p. ej., moléculas orgánicas pequeñas, péptidos, macrociclos, compuestos inorgánicos, organometálicos, etc.) y la cantidad de moléculas por analizar (la duración del cálculo es proporcional al tamaño, a la cantidad de compuestos y a la capacidad de cómputo invertida) (Medina-Franco, Sánchez-Cruz, López-López & Díaz-Eufracio, 2022b). En la mayoría de las aplicaciones de la Quimioinformática están involucradas cientos, miles o millones de estructuras, por lo que se utilizan huellas digitales moleculares (Yang *et al.*, 2022).

Diversidad química

La diversidad química en las quimiotecas se emplea para cuantificar el grado de diferenciación de los compuestos presentes. Esto se logra a partir de diferentes métricas relacionadas con el concepto de *similitud química*. La similitud consiste en cuantificar las propiedades comunes entre un conjunto de moléculas y es dependiente de la forma en que a estas se les describe (representación molecular), y del tipo de métrica, de función o de algoritmo que se utilice en el estudio (Maggiora, Vogt, Stumpfe & Bajorath, 2014).

La similitud química es un concepto clave de la Quimioinformática. Se espera que compuestos químicos con estructuras similares posean similares propiedades y similares actividades biológicas frente a dianas terapéuticas (Johnson & Maggiora, 1990). En los últimos años se han desarrollado diversos métodos para calcular la similitud química o

molecular, relacionando el número de propiedades comunes entre dos moléculas y el total de propiedades diferentes. Este es el principio que se aplica en los coeficientes de Tanimoto (Tanimoto, *s/f*; Willett, Barnard & Downs, 1998), Tversky (Tversky, 1977) o Dice (Willett *et al.*, 1998), entre otros (Maggiara & Shanmugasundaram, 2004).

La ecuación 1 muestra el cálculo general de la similitud por medio del coeficiente de Tanimoto T_c , donde a y b son el número de características en consideración, propias de las moléculas A y B , y c es el número de características que comparten (Maggiara *et al.*, 2014). De acuerdo con la forma de la función, el índice de similitud o coeficiente de Tanimoto puede tomar valores entre 0 y 1, donde 0 significa ninguna característica en común, y 1 es la identidad entre las dos moléculas que se están comparando.

$$T_c(A, B) = c / ((a - c) + (b - c) + c) \quad (\text{ecuación 1})$$

En la descripción de bases de datos, es útil comparar la similitud acumulada (Figura 3A) de todos los compuestos presentes entre sí, con el fin de verificar su diversidad. Una base de datos diversa presenta valores cercanos a cero, por lo que la curva de similitud acumulada presenta un aumento rápido a valores pequeños. Las quimiotecas menos diversas tienen valores más grandes de similitud, por lo que la curva presenta una pendiente menos inclinada (Medina-Franco, Martínez-Mayorga, Bender & Scior, 2009).

Por medio del análisis de andamios moleculares, también es posible caracterizar la diversidad de estructuras presentes en una quimioteca. La presencia de pocos, o muchos núcleos base, y el porcentaje de compuestos de la base de datos que presenten

cada núcleo base, evidencian el objetivo de la preparación de la base de datos, ya sea para actuar sobre una diana biológica (regularmente baja diversidad de andamios moleculares) o hacia la mayor diversidad posible (con alta diversidad de andamios). Cada uno de estos ejemplos presenta una utilidad y retos propios (Langdon *et al.*, 2011). Para cuantificar la diversidad en función del contenido de andamios moleculares, se toman como base: la entropía de Shannon, que determina el grado de esparcimiento de los datos a lo largo de las diferentes categorías, y las curvas de recuperación de estructuras cíclicas (Figura 3B). En un sentido análogo a la visualización de la similitud acumulada, esta herramienta indica la fracción de la quimioteca que es cubierta por cierta fracción de los núcleos base presentes. Una ascendente diagonal indica una mayor diversidad y una vertical que la mayoría de los compuestos están densamente distribuidos en una pequeña fracción de los núcleos base presentes (Medina-Franco *et al.*, 2009).

Finalmente, una aplicación desarrollada por González-Medina y autores, consistió en determinar la diversidad global de una quimioteca con base en un consenso entre varios criterios (González-Medina, Prieto-Martínez, Owen & Medina-Franco, 2016). Así, es posible corregir el sesgo existente por la representación química empleada en la cuantificación de la diversidad. Esta técnica tiene en cuenta la diversidad de los andamios moleculares, de las huellas dactilares moleculares, así como de las propiedades fisicoquímicas, mediante una reducción de la dimensionalidad que resulta intuitiva en su interpretación, y útil en la clasificación de las bases de datos. Un ejemplo de uso de este método fue la caracterización realizada por Naveja y autores de los compuestos químicos de origen alimenticio presentes en *FooDB* (The Metabolomics Innovation Centre,

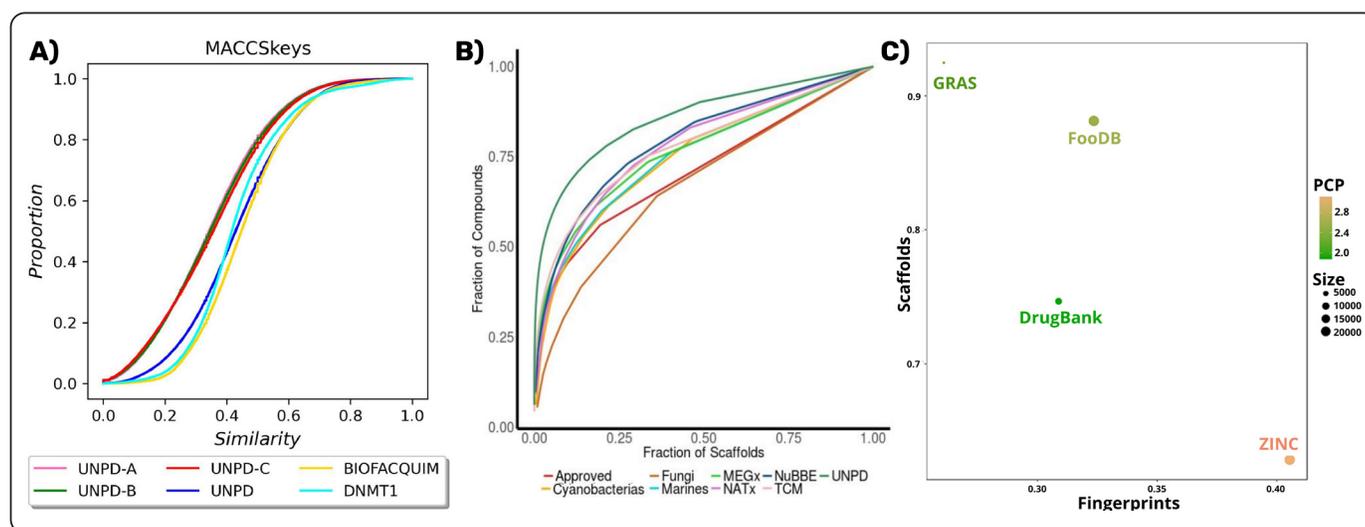


Figura 3. Ejemplos de métodos de visualización utilizados en el estudio de la diversidad química (A) Curvas de similitud acumulada (recuperado de Chávez-Hernández & Medina-Franco, 2023); (B) Curvas de recuperación de estructuras cíclicas (recuperado de Saldívar-González, Valli, Andricopulo, da Silva Bolzani & Medina-Franco, 2019); (C) Cuantificación y visualización de diversidad consenso (modificado de Naveja, Rico-Hidalgo & Medina-Franco, 2018).

2020; Naveja, Rico-Hidalgo & Medina-Franco, 2018) (ver Figura 3C).

Estudio del espacio químico y su visualización

En la Quimiinformática, una forma de describir, clasificar y visualizar la distribución de los compuestos químicos en una quimioteca, de acuerdo con diferentes propiedades fisicoquímicas, estructurales o mecanocuánticas, se conoce como “espacio químico”. Es decir, todo conjunto de N moléculas, caracterizado por M descriptores, es un espacio químico (Saldívar-González & Medina-Franco, 2022; Varnek & Baskin, 2011).

Por otro lado, un multiverso químico recién se describió como un conjunto de N moléculas, y al mismo tiempo por conjuntos M_i ($i = 1, \dots, j; j > 1$) de descriptores (Medina-Franco, Chávez-Hernández, López-López & Saldívar-González, 2022a). Es decir, que el multiverso químico es un conjunto de múltiples espacios químicos para un grupo de moléculas.

Adicionalmente, la representación de un conjunto de moléculas por medio de un grupo de descriptores permite visualizar cualitativa y cuantitativamente las relaciones que hay entre las moléculas. La Figura 4 es un ejemplo de diferentes visualizaciones de espacios y multiversos químicos. La Figura 4A ilustra un espacio químico usando la técnica de *ChemMaps*, para *FooDB* y otras quimiotecas (Naveja *et al.*, 2018). *ChemMaps* muestra parcialmente la base de datos de interés por medio del análisis de componentes principales (PCA) de diferentes descriptores moleculares de compuestos satélites, y los que están a su alrededor a través de la similitud pareada frente a los ya representados (Naveja & Medina-Franco, 2017). Este análisis indica que los compuestos de origen alimenticio presentan una mayor diversidad en términos de sus propiedades fisicoquímicas que los productos naturales y las aprobadas para uso clínico. También se puede inferir el carácter promisorio de los compuestos químicos de alimentos para llevar a cabo campañas de cribado con el fin de encontrar nuevas moléculas bioactivas, en la confianza de que son seguros para el consumo humano (Naveja *et al.*, 2018). En la Figura 4B se observa un multiverso químico (en este caso, dos espacios químicos alternos, cada uno de ellos con una representación molecular diferente) para dos quimiotecas de origen natural: la Base de Datos de Productos Naturales de América Latina (LaNaPDB) (Gómez-García *et al.*, 2023), la Colección de Productos Naturales Abiertos (COCONUT, por las siglas en inglés, *Collection of Open Natural Products*) y una quimioteca de moléculas aprobadas para su uso clínico por la FDA, con seis propiedades fisicoquímicas de interés farmacéutico, y una reducción de componentes por medio de PCA y Encaje Estocástico de Vecinos t-Distribuido (t-SNE) (Gómez-García *et al.*, 2023; Sorokina, Meseburger, Rajan, Yirik & Steinbeck, 2021; Van der Maaten & Hinton, 2008; Wishart *et al.*, 2018). Este análisis permitió demostrar la complejidad de las moléculas

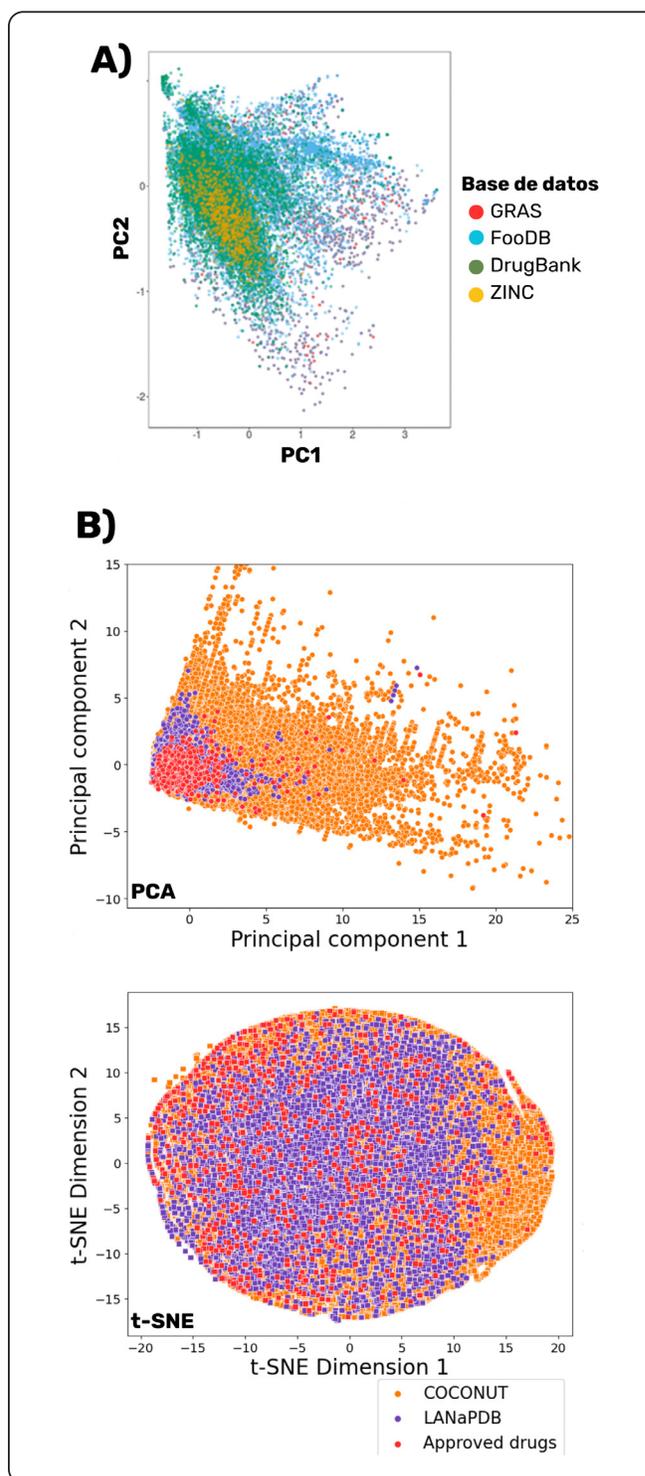


Figura 4. Ejemplo de representaciones visuales del espacio -y multiverso- químico de compuestos de alimentos. (A) *ChemMaps* comparativo entre *FooDB*, GRAS, productos naturales en ZINC y moléculas de uso clínico (recuperado de Naveja, Rico-Hidalgo & Medina-Franco, 2018); (B) PCA y t-SNE para LaNaPDB, COCONUT y medicamentos aprobados por la FDA, en términos de propiedades fisicoquímicas de interés farmacéutico (adaptado de Gómez-García *et al.*, 2023).

en LaNaPDB, y que tienen propiedades similares a las de COCONUT, y su multiverso químico se traslapa parcialmente con el de los medicamentos aprobados.

REVISIÓN DE LAS APLICACIONES Y LAS TENDENCIAS DE LA QUÍMICA INFORMÁTICA Y LA INTELIGENCIA ARTIFICIAL EN EL ESTUDIO DE LOS ALIMENTOS

La aplicación de las técnicas de la computación, descritas en secciones anteriores (Quimioinformática, IA y Ciencia de Datos), a los compuestos de alimentos se ha ido extendiendo por múltiples vertientes en los últimos años, lo cual ha dado lugar a la acuñación del término “*FoodInformatics*” para describir este área de investigación. El vocablo dio título a un volumen publicado en el año 2014, con distintos trabajos en el tema (Martínez-Mayorga & Medina-Franco, 2014). Revisiones posteriores relacionadas a esta materia se publicaron en el año 2018 (Peña-Castillo, Méndez-Lucio, Owen, Martínez-Mayorga & Medina-Franco, 2018), y más recientemente, aunque centrado en las bases de datos, en el año 2023 (Tseng *et al.*, 2023).

La temática es de interés por su aplicación: **i**) para el diseño de aditivos de alimentos, como conservadores (o conservantes), colorantes, odorantes, y saborizantes, así como en el diseño de nuevos ingredientes bioactivos para alimentos funcionales **ii**) para ayudar en la identificación del modo de acción de compuestos bioactivos de alimentos en la salud, por ejemplo, con un efecto positivo para la salud (preventivos y/o curativos), o bien para efectos indeseados (toxicidad, e interacción con los fármacos, etc.). También se han realizado estudios en los que se analiza por computadora la interacción de estas moléculas con los receptores humanos. Finalmente, la utilización de estas moléculas como fuente de quimiotipos para el diseño de fármacos, la cafeína, por ejemplo, es un paradigma pues ha sido la base para el descubrimiento de fármacos dirigidos a dianas moleculares como receptores de adenosina y fosfodiesterasas. Un ejemplo más es la celebración en el último congreso de la Sociedad Americana de Química (ACS, otoño de 2023), de dos simposios sobre: “*Advances in Food Chemical Informatics, Knowledge Bases and Databases*”, y “*Artificial Intelligence (AI) applications for Food and Agriculture*”.

En las secciones anteriores se mencionan los espacios químicos que están en las bases de datos de compuestos de alimentos como *FooDB* (The Metabolomics Innovation Centre, 2020), con probabilidad de ser la base de datos de compuestos de alimentos pública más completa. Naveja y autores la consultaron cuando contenía unos 24K compuestos, y observaron: **i**) que las moléculas presentaban una alta diversidad, tanto estructural como en sus propiedades fisicoquímicas, **ii**) complejidad estructural, **iii**) una proporción alta (~30%) de moléculas acíclicas, **iv**) que las moléculas con anillos eran principalmente monocíclicas, **v**) que comparten andamios moleculares típicos de los productos naturales (Naveja *et al.*, 2018). Se utilizó adicionalmente el método *ChemMaps* para

ver el espacio químico (Naveja & Medina-Franco, 2017). En una actualización de este análisis, Avellaneda-Tamayo y autores ampliaron los resultados con la versión más reciente de *FooDB* y obtuvieron más de 70K compuestos con las siguientes características: **i**) lipídicos que desplazan la tendencia de los diferentes descriptores constitucionales y fisicoquímicos hacia compuestos no polares, **ii**) una baja diversidad estructural entre las moléculas, **iii**) un traslape significativo entre los compuestos presentes en los alimentos y los aprobados para uso clínico, **iv**) un análisis quimioinformático de los compuestos de alimentos disponibles comercialmente (Avellaneda-Tamayo, Chávez-Hernández, Prado-Romero & Medina-Franco, 2024).

El perfil ácido/base de las moléculas en *FooDB* fue analizado por Santibáñez-Morán y autores con el siguiente resultado: **i**) un 41% de moléculas neutras, alrededor del doble que en fármacos aprobados, **ii**) un porcentaje comparativamente muy bajo de moléculas monobásicas 5% vs. 28% de los fármacos, **iii**) entre los grupos funcionales, los ácidos frecuentes fueron los fenoles (16%), los fosfatos (17%) y los carboxilatos (17%) (Santibáñez-Morán, Rico-Hidalgo, Manallack & Medina-Franco, 2019).

En el campo del mecanismo de acción de compuestos de alimentos, más recientemente se ha investigado mediante métodos computacionales la presencia de agregadores y subestructuras “problemáticas” (del inglés *nuisance substructures*) en la *FooDB*. Los **primeros (agregadores) son compuestos que se agregan** formando coloides que desnaturalizan proteínas, resultando en falsos positivos en los ensayos bioquímicos en los que se prueban. **Las segundas (subestructuras “problemáticas”)** son **partes de moléculas** que se ha visto que están asociadas a promiscuidad y falsos positivos en ensayos por distintos motivos: unión covalente (Kaya & Colmenarejo, 2020) identificaron en el *FooDB*, con 26K compuestos de alimentos, subestructuras problemáticas utilizando distintos filtros quimioinformáticos (Baell & Holloway, 2010; Blake, 2005; Hann, Hudson, Lewell, Lifely, Miller & Ramsden, 1999), así como agregadores con la herramienta *Aggregator Advisor* (Irwin *et al.*, 2015). También identificaron una gran cantidad de Panaceas Metabólicas Inválidas (del inglés, *Invalid Metabolic Panaceas*), que son productos naturales conocidos por su capacidad biomolecular para interactuar con múltiples dianas biológicas (Bisson, McAlpine, Friesen, Chen, Graham & Pauli, 2016). Posteriormente este análisis se repitió en 71,000 moléculas, con una versión actualizada del *FooDB* y un modelo de aprendizaje automático, el *Scam Detective*, para identificar los agregadores (Sánchez-Ruiz & Colmenarejo, 2021). En el mismo trabajo también se analizaron las propiedades fisicoquímicas y la distribución de los andamios moleculares, el resultado fue ver patrones muy diferentes en los subgrupos de acilgliceroles y no-acilgliceroles de las moléculas de alimentos, tanto entre ellos como al compararlos con las moléculas del *DrugBank* (Wishart *et al.*, 2018).

En otro trabajo relacionado de los mismos autores se empleó el método de predicción de dianas basado en el ligando SEA+TC (*Similarity Ensemble Approach-maximum Tanimoto Coefficient*) con el fin de proponer interacciones entre los compuestos de alimentos presentes en el *FooDB* y las proteínas humanas (Sánchez-Ruiz & Colmenarejo, 2022). También llevaron a cabo un análisis sistemático de todas las interacciones de las moléculas de alimentos con las proteínas humanas de las que sólo el 1.6% se tenía evidencia publicada; la mayoría estaban concentradas en tres grupos químicos: fenilpropanoides y policétidos, compuestos organoheterocíclicos, y bencenoides. La utilización de SEA+TC permitió prever interacciones potenciales para 64% de los compuestos en *FooDB*, que se hicieron disponibles como material suplementario de la publicación. Esta lista de más de 88,000 predicciones representa un conjunto de hipótesis para la realización de ensayos rápidos efectuados por grupos experimentales y con una alta probabilidad de confirmación. De hecho, mediante análisis retrospectivos de *ChEMBL* y acoplamiento molecular (*docking*) los autores lograron validar setenta y cinco de estas interacciones. Asimismo, se realizó un análisis de andamios enriquecidos por clase química y grupo de dianas.

En este mismo marco de interacciones moleculares, pero entre fármacos y compuestos de alimentos, *FDMine* es un buen ejemplo de modelo de grafos para predecir posibles interacciones entre los fármacos y los alimentos, con base en el *FooDB* y *DrugBank*. Este tipo de predicciones pueden ser útiles para evaluar posibles efectos sinérgicos, en que un alimento mejore el efecto de un fármaco, o lo opuesto sea en detrimento de la efectividad del medicamento (Rahman, Vadrev, Magana-Mora, Leyman & Soufan, 2022).

Un modelo de aprendizaje profundo, *DeepDDI*, se puso a prueba y se utilizó para predecir las interacciones fármaco-fármaco y fármaco-alimento (Ryu, Kim & Lee, 2018). En este caso, el modelo se probó a partir de los datos del *DrugBank*, y fue aplicado para predecir las interacciones fármaco-alimento.

Otro trabajo se basó en la previsión, por aprendizaje automático, de las interacciones entre compuestos GRAS, los ingredientes inactivos, la glicoproteína-P (Pgp) y la uridina difosfatoglucuronosiltransferasa-2B7 (UGT2B7) en las propiedades farmacocinéticas de los fármacos.

En otra área alternativa, además de volver a sobresalir los compuestos de alimentos como una fuente de fármacos, y el uso de la cafeína como homotipo, también destaca la utilidad de estos compuestos en el combate contra la *COVID19* vía identificación de nuevos quimiotipos basados en los compuestos de alimentos (Juárez-Mercado *et al.*, 2023). Por ejemplo, varios compuestos de la *FooDB* se propusieron como inhibidores de la proteasa principal del *SARS-CoV-2* a partir de estudios de acoplamiento y de dinámica molecular

(Masand, Sk, Kar, Rastija & Zaki, 2021). En otro trabajo, se utilizó una combinación de varios métodos de cribado virtual para identificar potenciales inhibidores de la misma proteína, en este caso a partir no sólo de *FooDB*, sino también de la base de datos de la “materia química oscura” (*Dark Chemical Matter*, o DCM, conjunto de compuestos de baja o nula actividad reportada) (Santibáñez-Morán, López-López, Prieto-Martínez, Sánchez-Cruz & Medina-Franco, 2020). Para el caso del cáncer, el modelo de aprendizaje automático *HyperFoods* se desarrolló y aplicó a los compuestos de alimentos con el fin de identificar moléculas con posible actividad antitumoral (Veselkov *et al.*, 2019). Una reciente creación es la base de datos anotada de libre acceso con compuestos de alimentos y cuya actividad está publicada con dianas epigenéticas, en la “*Epi Food Chemical Database*” (Juárez-Mercado *et al.*, 2024). Las moléculas pertenecían a la *FooDB* o a bases de datos de productos naturales como Open Natural ProdUcTs y *COCONUT*.

Para finalizar, hay dos trabajos con una orientación químico-analítica. Uno de ellos consiste en una revisión del campo del análisis metabólico de compuestos de alimentos, o *foodomics*, a la luz de la aplicación en el mismo de nuevas aproximaciones de aprendizaje automático (Jiménez-Carvelo & Cuadros-Rodríguez, 2021). El segundo es un trabajo que usó el método de aprendizaje automático para desarrollar un clasificador multi etiqueta de olores de tres alimentos: queso, licor, y aceite comestible, con datos de lecturas de un conjunto de sensores quimioresistivos (Schroeder *et al.*, 2019).

CONCLUSIONES Y PERSPECTIVAS

En este artículo se mencionaron conceptos, procedimientos y ejemplos de uso propios de la informática, con sus respectivas aplicaciones en la Química y especialmente en la Química de Alimentos, como son: la Quimioinformática, la Ciencia de Datos y la Inteligencia Artificial. La conclusión es que el campo de la Informática de la Química de Alimentos (*FooInformatics*), especialmente en combinación con la Inteligencia Artificial y la Ciencia de Datos, abre perspectivas para utilizar a los compuestos de alimentos con diferente fin en: **i)** dietas más saludables y personalizadas, **ii)** alimentos con una composición óptima, para protección de la salud, cuidando el sabor, el olor, el color y la conservación, **iii)** diseño de nuevos fármacos y nutracéuticos, **iv)** minimización de interacción con fármacos, etc. El uso de este tipo de herramientas en la computación permite el ahorro de tiempo y dinero en la experimentación, así como en la racionalización y en la compresión de una gran cantidad de datos, para futuras investigaciones. Los modelos desarrollados con estas técnicas permiten la realización de “experimentos en la computadora” y la posibilidad de nuevas hipótesis con aplicación en las áreas antes citadas, además de la creación de nuevos diseños. Sin embargo, existen retos como mejorar los datos que son confiables, un indicador quimioinformático “tipo-alimento” para identificar estructuras moleculares de uso

alimentario y un diseño molecular computacional de compuestos de alimentos.

AGRADECIMIENTOS

Este artículo está financiado por el proyecto PID2021-127318OB-I00 del Ministerio de Ciencia e Innovación de España: MCIN/AEI/10.13039/501100011033 y por “FEDER Una manera de hacer Europa”.

Juan F. Avellaneda Tamayo agradece al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT), México, por la beca de maestría número 1270553. Andrés Sánchez Ruiz agradece a la Comunidad de Madrid por el contrato sujeto al proyecto PEJ-2020-AIBMD-19384, y por el contrato de investigador predoctoral PIPF-2022-SAL-GL-26278, que han financiado este artículo. Biljana Savic agradece al programa EIT-Food RIS Fellowship por una beca para una estancia en IMDEA Alimentación.

REFERENCIAS

- Açar, Y. & Akbulut, G. (2023). Nutritional Epigenetics and Phytochemicals in Cancer Formation. *Journal of the American Nutrition Association*, **42**(7), 700–705. <https://doi.org/10.1080/27697061.2022.2147106>
- Almazroo, O. A., Miah, M. K. & Venkataramanan, R. (2017). Drug metabolism in the liver. *Clinics in Liver Disease*, **21**(1), 1–20. <https://doi.org/10.1016/j.cld.2016.08.001>
- Anwar, H., Anwar, T. & Murtaza, S. (2023). Review on food quality assessment using machine learning and electronic nose system. *Biosensors and Bioelectronics: X*, **14**, 100365. <https://doi.org/10.1016/j.biosx.2023.100365>
- Avellaneda-Tamayo, J. F., Chávez-Hernández, A. L., Prado-Romero, D. L. & Medina-Franco, J. L. (2024). Chemical Multiverse and Diversity of Food Chemicals. *Journal of Chemical Information and Modeling*, **64**(4), 1229–1244. <https://doi.org/10.1021/acs.jcim.3c01617>
- Baell, J. B. & Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry*, **53**(7), 2719–2740. <https://doi.org/10.1021/jm901137j>
- Bazoukis, G., Hall, J., Loscalzo, J., Antman, E. M., Fuster, V. & Armoundas, A. A. (2022). The inclusion of augmented intelligence in medicine: A framework for successful implementation. *Cell Reports. Medicine*, **3**(1), 100485. <https://doi.org/10.1016/j.xcrm.2021.100485>
- Bemis, G. W. & Murcko, M. A. (1996). The properties of known drugs. 1. Molecular frameworks. *Journal of Medicinal Chemistry*, **39**(15), 2887–2893. <https://doi.org/10.1021/jm9602928>
- Bisson, J., McAlpine, J. B., Friesen, J. B., Chen, S-N., Graham, J. & Pauli, G. F. (2016). Can invalid bioactives undermine natural product-based drug discovery? *Journal of Medicinal Chemistry*, **59**(5), 1671–1690. <https://doi.org/10.1021/acs.jmedchem.5b01009>
- Blake, J. F. (2005). Identification and evaluation of molecular properties related to preclinical optimization and clinical fate. *Medicinal Chemistry*, **1**(6), 649–655. <https://doi.org/10.2174/157340605774598081>
- Boatella, J. (2017). Relaciones nutricionales: del equivalente nutritivo a las listas de intercambio Nutritional relations: from nutritional equivalent to exchange lists. *Revista Española de Nutrición Humanay Dietética*, **21**(3), 300–309. <https://doi.org/10.14306/renhyd.21.3.195>
- Bondy, J. A. & Murty, U. S. R. (1976). *Graph theory with applications*. Ontario: American Elsevier Publishing Company. <https://doi.org/10.1137/1021086>
- Brown, F. K. (1998). Chapter 35. Chemoinformatics: What is it and how does it impact drug discovery. *Annual Reports in Medicinal Chemistry*, **33**, 375–384. [https://doi.org/10.1016/S0065-7743\(08\)61100-8](https://doi.org/10.1016/S0065-7743(08)61100-8)
- Caballero B., Trigo L. C. & Finglas P. M. (Eds). (2003). *Encyclopedia of food sciences and nutrition*. Amsterdam: Academic Press.
- Chávez-Hernández, A. L. & Medina-Franco, J. L. (2023). Natural products subsets: generation and characterization. *Artificial Intelligence in the Life Sciences*, **3**, 100066. <https://doi.org/10.1016/j.aills.2023.100066>
- Chen, X., Zhang, Z., Liao, W. & Zhao, Y. (2023). Assessment tool based on fatty acid metabolic signatures for predicting the prognosis and treatment response in bladder cancer. *Heliyon*, **9**(12), e22768. <https://doi.org/10.1016/j.heliyon.2023.e22768>
- Choi, S-W. & Friso, S. (2010). Epigenetics: a new bridge between nutrition and health. *Advances in Nutrition*, **1**(1), 8–16. <https://doi.org/10.3945/an.110.1004>
- Chopra, A. S., Lordan, R., Horbańczuk, O. K., Atanasov, A. G., Chopra, I., Horbańczuk, J. O., Jóźwik, A., Huang, L., Pirgozliev, V., Banach, M., Battino, M. & Arkells, N. (2022). The current use and evolving landscape of nutraceuticals. *Pharmacological Research: The Official Journal of the Italian Pharmacological Society*, **175**, 106001. <https://doi.org/10.1016/j.phrs.2021.106001>
- David, L., Thakkar, A., Mercado, R. & Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, **12**(1), 56. <https://doi.org/10.1186/s13321-020-00460-5>
- Division of Chemistry and Chemical Technology. (1964). *Survey of Chemical Notation Systems*. National Academies Press.
- Donoho, D. (2017). 50 Years of data science. *Journal of computational and graphical statistics*, **26**(4), 745–766. <https://doi.org/10.1080/10618600.2017.1384734>
- Dupont, C., Armant, D. R. & Brenner, C. A. (2009). Epigenetics: definition, mechanisms and clinical perspective. *Seminars in Reproductive Medicine*, **27**(5), 351–357. <https://doi.org/10.1055/s-0029-1237423>
- Ertl, P., Roggo, S. & Schuffenhauer, A. (2008). Natural product-likeness score and its application for prioritization of

- compound libraries. *Journal of Chemical Information and Modeling*, **48**(1), 68–74. <https://doi.org/10.1021/ci700286x>
- Faudone, G., Arifi, S. & Merk, D. (2021). The medicinal chemistry of caffeine. *Journal of Medicinal Chemistry*, **64**(11), 7156–7178. <https://doi.org/10.1021/acs.jmedchem.1c00261>
- Gasteiger, J. (2020). Chemistry in times of artificial intelligence. *Chemphyschem: A European Journal of Chemical Physics and Physical Chemistry*, **21**(20), 2233–2242. <https://doi.org/10.1002/cphc.202000518>
- Ghosh, T., Zhang, W., Ghosh, D. & Kechris, K. (2020). Predictive modeling for metabolomics data. En S. Li (Ed.), *Computational Methods and Data Analysis for Metabolomics* (pp. 313–336). New York: Humana. https://doi.org/10.1007/978-1-0716-0239-3_16
- Gleeson, M. P. (2008). Generation of a set of simple, interpretable ADMET rules of thumb. *Journal of Medicinal Chemistry*, **51**(4), 817–834. <https://doi.org/10.1021/jm701122q>
- Gómez-García, A., Jiménez, D. A. A., Zamora, W. J., Barazorda-Ccahuana, H. L., Chávez-Fumagalli, M. Á., Valli, M., Andricopulo, A. D., Bolzani, V. da S., Olmedo, D. A., Solís, P. N., Núñez, M. J., Rodríguez Pérez, J. R., Valencia Sánchez, H. A., Cortés Hernández, H. F. & Medina-Franco, J. L. (2023). Navigating the chemical space and chemical multiverse of a unified Latin American natural product database: LANaPDB. *Pharmaceuticals*, **16**(10), 1388. <https://doi.org/10.3390/ph16101388>
- González-Medina, M., Prieto-Martínez, F. D., Owen, J. R. & Medina-Franco, J. L. (2016). Consensus diversity plots: a global diversity analysis of chemical libraries. *Journal of Cheminformatics*, **8**, 63. <https://doi.org/10.1186/s13321-016-0176-9>
- Grisoni, F., Ballabio, D., Todeschini, R. & Consonni, V. (2018). Molecular descriptors for structure–activity applications: a hands-on approach. En O. Nicolotti (Ed.), *Computational toxicology: methods and protocols* (pp. 3–53). New York: Humana Press. https://doi.org/10.1007/978-1-4939-7899-1_1
- Haddad, R., Khan, R., Takahashi, Y. K., Mori, K., Harel, D. & Sobel, N. (2008). A metric for odorant comparison. *Nature Methods*, **5**, 425–429. <https://doi.org/10.1038/nmeth.1197>
- Hann, M., Hudson, B., Lewell, X., Lifely, R., Miller, L. & Ramsden, N. (1999). Strategic pooling of compounds for high-throughput screening. *Journal of Chemical Information and Computer Sciences*, **39**(5), 897–902. <https://doi.org/10.1021/ci990423o>
- Hanson, B. (2017, junio 12). 2.2.2: Anatomy of a MOL file. Chemistry LibreTexts; Libretxts. [https://chem.libretexts.org/Courses/University_of_Arkansas_Little_Rock/ChemInformatics_\(2017\)%3A_Chem_4399_5399/2.2%3A_Chemical_Representations_on_Computer%3A_Part_II/2.2.2%3A_Anatomy_of_a_MOL_file](https://chem.libretexts.org/Courses/University_of_Arkansas_Little_Rock/ChemInformatics_(2017)%3A_Chem_4399_5399/2.2%3A_Chemical_Representations_on_Computer%3A_Part_II/2.2.2%3A_Anatomy_of_a_MOL_file)
- Heinrich, M. (2019). Nutraceutical. En *Encyclopedia Britannica*. <https://www.britannica.com/science/nutraceutical>
- Heller, S. R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. (2015). InChI, the IUPAC international chemical identifier. *Journal of Cheminformatics*, **7**, 23. <https://doi.org/10.1186/s13321-015-0068-4>
- Hopfer, H., Buffon, P. A., Ebeler, S. E. & Heymann, H. (2013). The combined effects of storage temperature and packaging on the sensory, chemical, and physical properties of a Cabernet Sauvignon wine. *Journal of Agricultural and Food Chemistry*, **61**(13), 3320–3334. <https://doi.org/10.1021/jf3051736>
- Hughes, J. D., Blagg, J., Price, D. A., Bailey, S., Decrescenzo, G. A., Devraj, R. V., Ellsworth, E., Fobian, Y. M., Gibbs, M. E., Gilles, R. W., Greene, N., Huang, E., Krieger-Burke, T., Loesel, J., Wager, T., Whiteley, L. & Zhang, Y. (2008). Physicochemical drug properties associated with in vivo toxicological outcomes. *Bioorganic & Medicinal Chemistry Letters*, **18**(17), 4872–4875. <https://doi.org/10.1016/j.bmcl.2008.07.071>
- Irwin, J. J., Duan, D., Torosyan, H., Doak, A. K., Ziebart, K. T., Sterling, T., Tumanian, G. & Shoichet, B. K. (2015). An aggregation advisor for ligand discovery. *Journal of Medicinal Chemistry*, **58**(17), 7076–7087. <https://doi.org/10.1021/acs.jmedchem.5b01105>
- Jiménez-Carvelo, A. M. & Cuadros-Rodríguez, L. (2021). Data mining/machine learning methods in foodomics. *Current Opinion in Food Science*, **37**, 76–82. <https://doi.org/10.1016/j.cofs.2020.09.008>
- Johnson, M. A. & Maggiora, G. M. (1990). *Concepts and applications of Molecular Similarity*. New York: John Wiley & Sons. <https://doi.org/10.1002/jcc.540130415>
- Juárez-Mercado, K. E., Avellaneda-Tamayo, J. F., Villegas-Quintero, H., Chávez-Hernández, A. L., López-López, C. D. & Medina-Franco, J. L. (2024). Food chemicals and epigenetic targets: an epi food chemical database. *ACS Omega*, <https://doi.org/10.1021/acsomega.4c03321>
- Kaya, I. & Colmenarejo, G. (2020). Analysis of nuisance substructures and aggregators in a comprehensive database of food chemical compounds. *Journal of Agricultural and Food Chemistry*, **68**(33), 8812–8824. <https://doi.org/10.1021/acs.jafc.0c02521>
- Landrum, G. (s/f). RDKit: Open-source cheminformatics software. Recuperado el 31 de octubre de 2023, de <https://www.rdkit.org/>
- Langdon, S. R., Brown, N. & Blagg, J. (2011). Scaffold diversity of exemplified medicinal chemistry space. *Journal of Chemical Information and Modeling*, **51**(9), 2174–2185. <https://doi.org/10.1021/ci2001428>
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, **23**(1-3), 3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
- López-López, E., Bajorath, J. & Medina-Franco, J. L. (2021). Informatics for chemistry, biology, and biomedical sciences.

- Journal of Chemical Information and Modeling*, **61**(1), 26–35. <https://doi.org/10.1021/acs.jcim.0c01301>
- Maggiore, G. M. & Shanmugasundaram, V. (2004). Molecular similarity measures. En Bajorath, J. (ed) *Chemoinformatics and computational chemical biology. Methods in molecular biology*. (pp. 1–50). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-60761-839-3_2
- Maggiore, G., Vogt, M., Stumpfe, D. & Bajorath, J. (2014). Molecular similarity in medicinal chemistry. *Journal of Medicinal Chemistry*, **57**(8), 3186–3204. <https://doi.org/10.1021/jm401411z>
- Martínez-Mayorga, K. & Medina-Franco, J. L. (2009). Chapter 2 Chemoinformatics-applications in food chemistry. En Taylor S. (ed), *Advances in Food and Nutrition Research*, (Vol. 58, pp. 33–56). Burlington: Academic Press - Elsevier Inc. [https://doi.org/10.1016/S1043-4526\(09\)58002-3](https://doi.org/10.1016/S1043-4526(09)58002-3)
- Martínez-Mayorga, K. & Medina-Franco, J. L. (eds) (2014). *Foodinformatics: applications of chemical information to food chemistry*. Cham: Springer. <https://doi.org/10.1007/978-3-319-10226-9>
- Martínez-Mayorga, K., Peppard, T. L., López-Vallejo, F., Yongye, A. B. & Medina-Franco, J. L. (2013). Systematic mining of generally recognized as safe (GRAS) flavor chemicals for bioactive compounds. *Journal of Agricultural and Food Chemistry*, **61**(31), 7507–7514. <https://doi.org/10.1021/jf401019b>
- Martínez-Mayorga, K., Peppard, T. L., Ramírez-Hernández, A. I., Terrazas-Álvarez, D. E. & Medina-Franco, J. L. (2014). Chemoinformatics analysis and structural similarity studies of food-related databases. En Martínez-Mayorga, K. & Medina-Franco, J. L. (Eds.), *Foodinformatics: applications of chemical information to food chemistry* (pp. 97–110). Cham: Springer. https://doi.org/10.1007/978-3-319-10226-9_3
- Martínez-Mayorga, K., Rosas-Jiménez, J. G., González-Ponce, K., López-López, E., Neme, A. & Medina-Franco, J. L. (2024). The pursuit of accurate predictive models of the bioactivity of small molecules. *Chemical Science*, **15**(6), 1938–1952. <https://doi.org/10.1039/D3SC05534E>
- Masand, V. H., Sk, M. F., Kar, P., Rastija, V. & Zaki, M. E. A. (2021). Identification of food compounds as inhibitors of SARS-CoV-2 main protease using molecular docking and molecular dynamics simulations. *Chemometrics and Intelligent Laboratory Systems*, **217**, 104394. <https://doi.org/10.1016/j.chemolab.2021.104394>
- McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, **5**(4), 115–133. <https://doi.org/10.1007/BF02478259>
- Medina-Franco, J. L., Chávez-Hernández, A. L., López-López, E. & Saldívar-González, F. I. (2022a). Chemical multiverse: an expanded view of chemical space. *Molecular Informatics*, **41**(11), 2200116. <https://doi.org/10.1002/minf.202200116>
- Medina-Franco, J. L., Sánchez-Cruz, N., López-López, E. & Díaz-Eufracio, B. I. (2022b). Progress on open chemoinformatic tools for expanding and exploring the chemical space. *Journal of Computer-Aided Molecular Design*, **36**(5), 341–354. <https://doi.org/10.1007/s10822-021-00399-1>
- Medina-Franco, J., Martínez-Mayorga, K., Bender, A. & Scior, T. (2009). Scaffold diversity analysis of compound data sets using an entropy-based measure. *QSAR & Combinatorial Science*, **28**(11–12), 1551–1560. <https://doi.org/10.1038/s43246-024-00487-3>
- Miljković, F. & Medina-Franco, J. L. (2024). Artificial intelligence-open science symbiosis in chemoinformatics. *Artificial Intelligence in the Life Sciences*, **5**, 100096. <https://doi.org/10.1016/j.ailsci.2024.100096>
- Miranda-Salas, J., Peña-Varas, C., Valenzuela Martínez, I., Olmedo, D. A., Zamora, W. J., Chávez-Fumagalli, M. A., Azevedo, D. Q., Castilho, R. O., Maltarollo, V. G., Ramírez, D. & Medina-Franco, J. L. (2023). Trends and challenges in chemoinformatics research in Latin America. *Artificial Intelligence in the Life Sciences*, **3**, 100077. <https://doi.org/10.1016/j.ailsci.2023.100077>
- Morales-Garzón, A., Gómez-Romero, J. & Martín-Bautista, M. J. (2021). A word embedding-based method for unsupervised adaptation of cooking recipes. *IEEE Access*, **9**, 27389–27404. <https://doi.org/10.1109/ACCESS.2021.3058559>
- Naveja, J. J. & Medina-Franco, J. L. (2017). ChemMaps: towards an approach for visualizing the chemical space based on adaptive satellite compounds. *F1000Research*, **6**, 1134. <https://doi.org/10.12688/f1000research.12095.2>
- Naveja, J. J., Rico-Hidalgo, M. P. & Medina-Franco, J. L. (2018). Analysis of a large food chemical database: chemical space, diversity, and complexity. *F1000Research*, **7**, 993. <https://doi.org/10.12688/f1000research.15440.2>
- Park, D., Kim, K., Kim, S., Spranger, M. & Kang, J. (2021). FlavorGraph: a large-scale food-chemical graph for generating food representations and recommending food pairings. *Scientific Reports*, **11**(1), 931. <https://doi.org/10.1038/s41598-020-79422-8>
- Peña-Castillo, A., Méndez-Lucio, O., Owen, J. R., Martínez-Mayorga, K. & Medina-Franco, J. L. (2018). Chemoinformatics in Food Science. En Engel T. & Gasteiger J. (Eds.), *Applied Chemoinformatics* (pp. 501–525). Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527806539.ch10>
- Pletnev, I., Erin, A., McNaught, A., Blinov, K., Tchekhovskoi, D. & Heller, S. (2012). InChIKey collision resistance: an experimental testing. *Journal of Cheminformatics*, **4**(1), 39. <https://doi.org/10.1186/1758-2946-4-39>
- Rahman, M. M., Vadrev, S. M., Magana-Mora, A., Levman, J. & Soufan, O. (2022). A novel graph mining approach to predict and evaluate food-drug interactions. *Scientific Reports*, **12**(1), 1061. <https://doi.org/10.1038/s41598-022-05132-y>
- Romano, J. D. & Tatonetti, N. P. (2019). Informatics and computational methods in natural product drug discovery:

- a review and perspectives. *Frontiers in Genetics*, **10**, 368. <https://doi.org/10.3389/fgene.2019.00368>
- Ruddigkeit, L. & Reymond, J.-L. (2014). The chemical space of flavours. En K. Martínez-Mayorga & J. L. Medina-Franco (Eds.), *Foodinformatics: applications of chemical information to food chemistry* (pp. 83–96). Cham: Springer. https://doi.org/10.1007/978-3-319-10226-9_2
- Russell, S. J., Russell, S. & Norvig, P. (2020). *Artificial intelligence: a modern approach*. Hoboken, NJ: Pearson.
- Ryu, J. Y., Kim, H. U. & Lee, S. Y. (2018). Deep learning improves prediction of drug-drug and drug-food interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(18), E4304–E4311. <https://doi.org/10.1073/pnas.1803294115>
- Saldívar-González, F. I. & 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. <https://doi.org/10.1080/17460441.2022.2084608>
- Saldívar-González, F. I., Aldas-Bulos, V. D., Medina-Franco, J. L. & Plisson, F. (2022). Natural product drug discovery in the artificial intelligence era. *Chemical Science*, **13**(6), 1526–1546. <https://doi.org/10.1039/D1SC04471K>
- Saldívar-González, F. I., Fernández-de Gortari, E. & Medina-Franco, J. L. (2023). Inteligencia artificial en el diseño de fármacos: hacia la inteligencia aumentada. *Educación Química*, **34**(2), 17–25. <https://doi.org/10.22201/fq.18708404e.2023.2.83233>
- Saldívar-González, F. I., Prado-Romero, D. L., Cedillo-González, R., Chávez-Hernández, A. L., Avellaneda-Tamayo, J. F., Gómez-García, A., Juárez-Rivera, L. & Medina-Franco, J. L. (2024). A spanish chemoinformatics GitBook for chemical data retrieval and analysis using Python programming. *Journal of Chemical Education*, <https://doi.org/10.1021/acs.jchemed.4c00041>
- Saldívar-González, F. I., Valli, M., Andricopulo, A. D., da Silva Bolzani, V. & Medina-Franco, J. L. (2019). Chemical space and diversity of the NuBBE database: a chemoinformatic characterization. *Journal of Chemical Information and Modeling*, **59**(1), 74–85. <https://doi.org/10.1021/acs.jcim.8b00619>
- Sánchez-Ruiz, A. & Colmenarejo, G. (2021). Updated prediction of aggregators and assay-interfering substructures in food compounds. *Journal of Agricultural and Food Chemistry*, **69**(50), 15184–15194. <https://doi.org/10.1021/acs.jafc.1c05918>
- Sánchez-Ruiz, A. & Colmenarejo, G. (2022). Systematic analysis and prediction of the target space of bioactive food compounds: filling the chemobiological gaps. *Journal of Chemical Information and Modeling*, **62**(16), 3734–3751. <https://doi.org/10.1021/acs.jcim.2c00888>
- Santibáñez-Morán, M. G., López-López, E., Prieto-Martínez, F. D., Sánchez-Cruz, N. & Medina-Franco, J. L. (2020). Consensus virtual screening of dark chemical matter and food chemicals uncover potential inhibitors of SARS-CoV-2 main protease. *RSC Advances*, **10**(42), 25089–25099. <https://doi.org/10.1039/D0RA04922K>
- Santibáñez-Morán, M. G., Rico-Hidalgo, M. P., Manallack, D. T. & Medina-Franco, J. L. (2019). The acid/base profile of a large food chemical database. *Molecular Informatics*, **38**(6), 1800171. <https://doi.org/10.1002/minf.201800171>
- Schroeder, V., Evans, E. D., Wu, Y.-C. M., Voll, C.-C. A., McDonald, B. R., Savagatrup, S. & Swager, T. M. (2019). Chemiresistive sensor array and machine learning classification of food. *ACS Sensors*, **4**(8), 2101–2108. <https://doi.org/10.1021/acssensors.9b00825>
- Sheridan, R. P. & Kearsley, S. K. (2002). Why do we need so many chemical similarity search methods? *Drug Discovery Today*, **7**(17), 903–911. [https://doi.org/10.1016/S1359-6446\(02\)02411-X](https://doi.org/10.1016/S1359-6446(02)02411-X)
- Shultz, M. D. (2013). Setting expectations in molecular optimizations: strengths and limitations of commonly used composite parameters. *Bioorganic & Medicinal Chemistry Letters*, **23**(21), 5980–5991. <https://doi.org/10.1016/j.bmcl.2013.08.029>
- Shultz, M. D. (2014). Improving the plausibility of success with inefficient metrics. *ACS Medicinal Chemistry Letters*, **5**(1), 2–5. <https://doi.org/10.1021/ml4004638>
- Sorokina, M. & Steinbeck, C. (2020). Review on natural products databases: where to find data in 2020. *Journal of Cheminformatics*, **12**(1), 20. <https://doi.org/10.1186/s13321-020-00424-9>
- Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A. & Steinbeck, C. (2021). COCONUT online: collection of open natural products database. *Journal of Cheminformatics*, **13**(1), 2. <https://doi.org/10.1186/s13321-020-00478-9>
- Southan, C. (2013). InChI in the wild: an assessment of InChIKey searching in Google. *Journal of Cheminformatics*, **5**(1), 10. <https://doi.org/10.1186/1758-2946-5-10>
- Sprou, D. G. & Salemme, F. R. (2007). A comparison of the chemical properties of drugs and FEMA/FDA notified GRAS chemical compounds used in the food industry. *Food and Chemical Toxicology*, **45**(8), 1419–1427. <https://doi.org/10.1016/j.fct.2007.02.004>
- Streit, L. (2023). *Micronutrients: types, functions, benefits and more*. Healthline. <https://www.healthline.com/nutrition/micronutrients>
- Swain, M. (s/f). *MolVS: Molecule Validation and Standardization — MolVS 0.1.1 documentation*. Recuperado el 31 de octubre de 2023, de <https://molvs.readthedocs.io/en/latest/>
- Tambe, S., Jain, D., Amin, P., Mali, S. N. & Cruz, J. N. (2023). Chapter 8 - Advances in the development of a 3D-printed nutraceutical delivery platform. En Inamuddin, Altalhi, T. & Cruz, J. N. (Eds.), *Nutraceuticals* (pp. 193–222). Londres: Academic Press. <https://doi.org/10.1016/B978-0-443-19193-0.00008-3>
- Tanimoto, T. T. (s/f). IBM internal report, 17th November. *IBM Company: Armonk, NY, USA*.

- The Metabolomics Innovation Centre. (2020, abril 07). *FoodDB*. <https://foodb.ca/>
- Tian, S., Wang, J., Li, Y., Li, D., Xu, L. & Hou, T. (2015). The application of in silico drug-likeness predictions in pharmaceutical research. *Advanced Drug Delivery Reviews*, **86**, 2–10. <https://doi.org/10.1016/j.addr.2015.01.009>
- Todeschini, R., Consonni, V., Mannhold, R., Kubinyi, H. & Folkers, G. (2010). *Molecular descriptors for chemoinformatics*. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527628766>
- Tseng, Y. J., Chuang, P.-J. & Appell, M. (2023). When machine learning and deep learning come to the big data in food chemistry. *ACS Omega*, **8(18)**, 15854–15864. <https://doi.org/10.1021/acsomega.2c07722>
- Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics*, **33(1)**, 1–67. <https://doi.org/10.1214/aoms/1177704711>
- Tversky, A. (1977). Features of similarity. *Psychological Review*, **84(4)**, 327–352. <https://doi.org/10.1037/0033-295X.84.4.327>
- Van der Maaten, L. & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, **9(86)**, 2579–2605. <http://jmlr.org/papers/v9/vandermaaten08a.html>
- Varnek, A. & Baskin, I. I. (2011). Chemoinformatics as a theoretical chemistry discipline. *Molecular Informatics*, **30(1)**, 20–32. <https://doi.org/10.1002/minf.201000100>
- Veber, D. F., Johnson, S. R., Cheng, H.-Y., Smith, B. R., Ward, K. W. & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, **45(12)**, 2615–2623. <https://doi.org/10.1021/jm020017n>
- Veselkov, K., González, G., Aljifri, S., Galea, D., Mirnezami, R., Youssef, J., Bronstein, M. & Laponogov, I. (2019). HyperFoods: machine intelligent mapping of cancer-beating molecules in foods. *Scientific Reports*, **9(1)**, 9237. <https://doi.org/10.1038/s41598-019-45349-y>
- Voruganti, V. S. (2023). Precision Nutrition: Recent Advances in Obesity. *Physiology*, **38(1)**, 42–50. <https://doi.org/10.1152/physiol.00014.2022>
- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, **28(1)**, 31–36. <https://doi.org/10.1021/ci00057a005>
- Willett, P., Barnard, J. M. & Downs, G. M. (1998). Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*, **38(6)**, 983–996. <https://doi.org/10.1021/ci9800211>
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C. & Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, **46(D1)**, D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
- Yang, J., Cai, Y., Zhao, K., Xie, H. & Chen, X. (2022). Concepts and applications of chemical fingerprint for hit and lead screening. *Drug Discovery Today*, **27(11)**, 103356. <https://doi.org/10.1016/j.drudis.2022.103356>
- Zhu, H. (2020). Big data and artificial intelligence modeling for drug discovery. *Annual Review of Pharmacology and Toxicology*, **60**, 573–589. <https://doi.org/10.1146/annurev-pharmtox-010919-023324>