

Image not found

Crean una biblioteca abierta pionera para la identificación de biomoléculas

La UOC y el ICFO presentan un recurso clave para hacer avanzar la investigación de enfermedades como el cáncer. El proyecto, basado en la técnica analítica de la espectroscopia Raman, consiste en una base de datos en abierto para identificar biomoléculas que quiere crecer con la aportación de la comunidad científica. Los resultados se publican en **Chemometrics and Intelligent Laboratory Systems**.

November 27, 2025

Investigadores de la Universitat Oberta de Catalunya (UOC) y del Instituto de Ciencias Fónicas (ICFO) han creado una base de datos espectral Raman accesible y abierta a la comunidad científica que contiene 140 biomoléculas de los principales tipos, como ácidos nucleicos, proteínas, lípidos, carbohidratos, etc. La espectroscopia Raman es una técnica que permite analizar la composición química y la estructura molecular de materiales mediante la interacción de la luz con la materia, específicamente en el llamado fenómeno de dispersión Raman, descubierto por el físico Chandrasekhara Venkata Raman en 1928. El trabajo, *Open Raman spectral library for biomolecule identification*, publicado en abierto en la revista *Chemometrics and Intelligent Laboratory Systems*, ha sido liderado por la UOC, y también han colaborado los investigadores del ICFO **Jose Javier Ruiz** y el **Dr. Pablo Loza-Alvarez**.

"Hasta ahora, una de las limitaciones del potencial de Raman en aplicaciones biomédicas era la falta de datos espectrales abiertos de biomoléculas. Por eso nos propusimos crear una biblioteca accesible, estandarizada y útil para la comunidad científica, que sirva de base para futuras investigaciones y aplicaciones clínicas", explica Marcelo Teran, primer autor del artículo.

En el proyecto, los investigadores han implementado dos algoritmos de búsqueda que han mostrado una precisión del 100 % tanto en el top 10 de la identificación de moléculas -por ejemplo, el colágeno- como en la identificación del tipo de molécula -por ejemplo, proteína-, en cuanto a las mediciones de biomoléculas puras y a la hora de replicar los resultados de estudios anteriores.

Datos biomédicos en abierto para hacer avanzar la medicina

"La espectroscopia Raman permite analizar la composición química de muestras de forma no invasiva, lo que es muy valioso en el ámbito médico. Esta base de datos puede facilitar la identificación precisa de biomoléculas y, en el futuro, contribuir a estudiar cómo varía su presencia en procesos biológicos como el cáncer", asegura Teran. "La disponibilidad de datos biomédicos de calidad es esencial para avanzar en el desarrollo de soluciones basadas en inteligencia artificial. Esta necesidad fue el punto de partida de la investigación", explica el investigador de la UOC.

Los investigadores recopilaron datos de espectros Raman de biomoléculas a partir de los artículos más relevantes publicados en este ámbito y desarrollaron un algoritmo con técnicas clásicas de visión por computador para extraer automáticamente los datos. Uno de los retos de este trabajo fue el escaso número de datos espectrales publicados en abierto, que superaron utilizando validaciones experimentales. "Nuestro trabajo ofrece una herramienta que puede ayudar a identificar de forma objetiva, rápida y estandarizada la composición molecular a partir de su espectro Raman. Actualmente, esta identificación se realiza mediante el análisis visual de los principales picos de los espectros y se compara con las referencias de la literatura. Nuestra herramienta puede agilizar este proceso al mismo tiempo que proporciona una solución estandar que reduce el sesgo humano durante el análisis", afirma el doctorando.

Una base de datos destinada a crecer con la aportación de la comunidad

De cara al futuro, los investigadores esperan que la comunidad científica contribuya a ampliar la base de datos, de modo que se convierta en una biblioteca colaborativa y de referencia para espectros Raman de biomoléculas.

"Todavía es poco común que en los artículos científicos se compartan los datos abiertamente, en especial en el ámbito de la espectroscopia Raman. Esta falta de acceso a datos limita considerablemente la investigación biomédica. Para que la IA pueda aplicarse con éxito, requiere grandes volúmenes de datos fiables y accesibles, y es aquí donde los proyectos de ciencia abierta tienen un papel clave", explica Marcelo Teran.

El objetivo es que, a medida que vaya ampliándose la base de datos, pueda impulsarse el entrenamiento de modelos de inteligencia artificial en el ámbito del análisis molecular de muestras biológicas. Esto abrirá la puerta a nuevas aplicaciones en diagnóstico y monitorización de enfermedades.

Referencia:

Teran, M., Ruiz, J. J., Loza-Alvarez, P., Masip, D. y Merino, D. Open Raman spectral library for biomolecule identification. En: Chemometrics and Intelligent Laboratory Systems. 2025. Vol. 264, 105476. DOI: <https://doi.org/10.1016/j.chemolab.2025.105476>

Agradecimientos:

The authors acknowledge funding from Fundacio CELLEX; Ministerio de Economía y Competitividad - Severo Ochoa programme for Centres of Excellence in R&D (CEX2019-000910-S); CERCA programme (999619436); Laserlab-Europe (871124); Ministerio de Ciencia e Innovacion PID2021-122807OB-C31 and PID2022-138721NBI00 projects funded by MCIN/AEI/10.13039/501100011033/FEDER, UE; CARET project. The SLN facility corresponds to a $\frac{1}{2}$ Grup reconegut $\frac{1}{2}$ 2021 SGR 01456 Departament de Recerca i Universitats de la Generalitat de Catalunya.