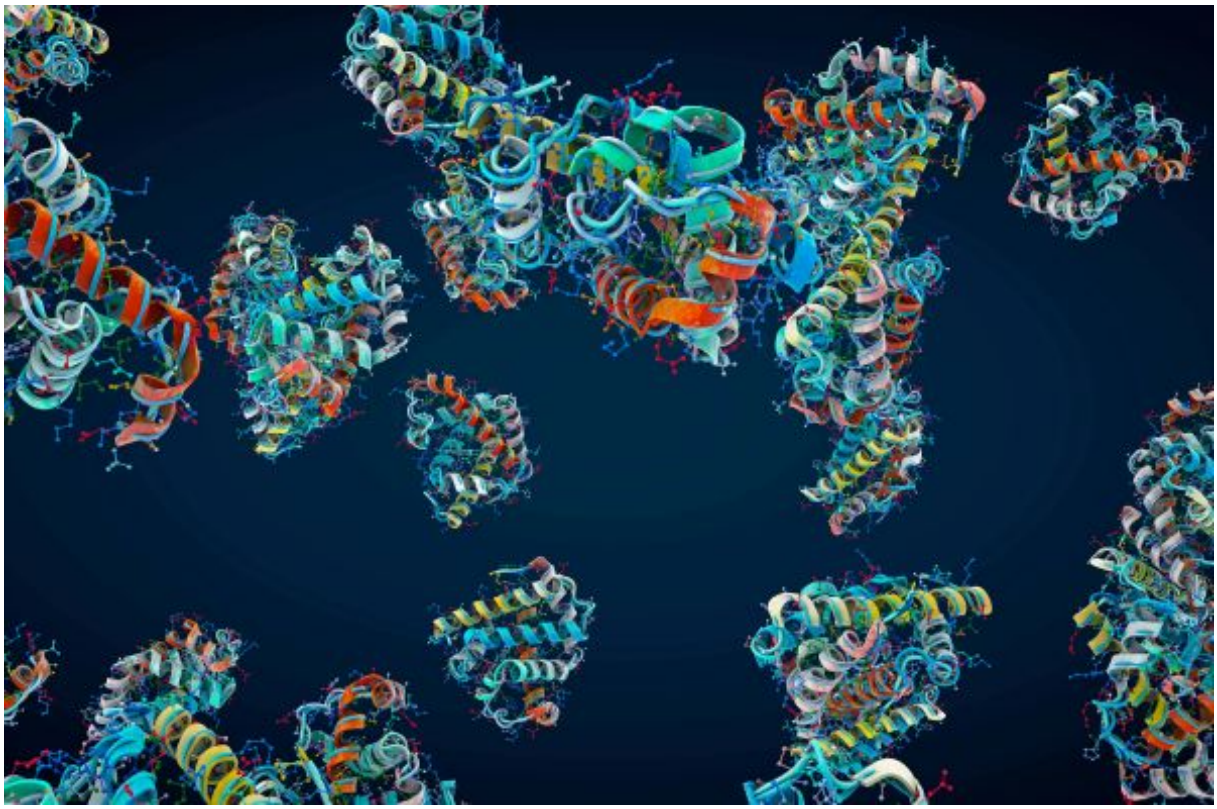

Crean una biblioteca abierta pionera para la identificacion de biomoleculas

La UOC y el ICFO presentan un recurso clave para hacer avanzar la investigacion de enfermedades como el cancer. El proyecto, basado en la tecnica analitica de la espectroscopia Raman, consiste en una base de datos en abierto para identificar biomoleculas que quiere crecer con la aportacion de la comunidad cientifica. 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 Fotonicas (ICFO) han creado una base de datos espectral Raman accesible y abierta a la comunidad cientifica que contiene 140 biomoleculas de los principales tipos, como acidos nucleicos, proteinas, lipidos, carbohidratos, etc. La espectroscopia Raman es una tecnica que permite analizar la composicion quimica y la estructura molecular de materiales mediante la interaccion de la luz con la materia, especificamente en el llamado fenomeno de dispersion Raman, descubierto por el fisico 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 tambien 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 biomedicas era la falta de datos espectrales abiertos de biomoleculas. Por eso nos propusimos crear una biblioteca accesible, estandarizada y util para la comunidad cientifica, que sirva de base para futuras investigaciones y aplicaciones clinicas", explica Marcelo Teran, primer autor del articulo.

En el proyecto, los investigadores han implementado dos algoritmos de busqueda que han mostrado una precision del 100 % tanto en el top 10 de la identificacion de moleculas -por ejemplo, el colageno- como en la identificacion del tipo de molecula -por ejemplo, proteina-, en cuanto a las mediciones de biomoleculas puras y a la hora de replicar los resultados de estudios anteriores.

Datos biomedicos en abierto para hacer avanzar la medicina

"La espectroscopia Raman permite analizar la composicion quimica de muestras de forma no invasiva, lo que es muy valioso en el ambito medico. Esta base de datos puede facilitar la identificacion precisa de biomoleculas y, en el futuro, contribuir a estudiar como varia su presencia en procesos biologicos como el cancer", asegura Teran. "La disponibilidad de datos biomedicos de calidad es esencial para avanzar en el desarrollo de soluciones basadas en inteligencia artificial. Esta necesidad fue el punto de partida de la investigacion", explica el investigador de la UOC.

Los investigadores recopilaron datos de espectros Raman de biomoleculas a partir de los articulos mas relevantes publicados en este ambito y desarrollaron un algoritmo con tecnicas clasicas de vision por computador para extraer automaticamente los datos. Uno de los retos de este trabajo fue el escaso numero de datos espectrales publicados en abierto, que superaron utilizando validaciones experimentales. "Nuestro trabajo ofrece una herramienta que puede ayudar a identificar de forma objetiva, rapida y estandarizada la composicion molecular a partir de su espectro Raman. Actualmente, esta identificacion se realiza mediante el analisis 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 solucion estandar que reduce el sesgo humano durante el analisis", afirma el doctorando.

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

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

"Todavia es poco comun que en los articulos cientificos se compartan los datos abiertamente, en especial en el ambito de la espectroscopia Raman. Esta falta de acceso a datos limita considerablemente la investigacion biomedica. Para que la IA pueda aplicarse con exito, requiere grandes volumenes de datos fiables y accesibles, y es aqui donde los proyectos de ciencia abierta tienen un papel clave", explica Marcelo Teran.
El objetivo es que, a medida que vaya ampliandose la base de datos, pueda impulsarse el entrenamiento de modelos de inteligencia artificial en el ambito del analisis molecular de muestras biologicas. Esto abrira la puerta a nuevas aplicaciones en diagnostico y monitorizacion 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 Economia 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 i¿½Grup reconeguti¿½ 2021 SGR 01456 Departament de Recer a i Universitats de la Generalitat de Catalunya.