



Creen una biblioteca oberta pionera per a la identificacio de biomolecules

La UOC i l'ICFO presenten un recurs clau per fer avançar la investigació de malalties com el càncer. El projecte, basat en la tècnica analítica de la espectroscòpia Raman, consisteix en una base de dades en obert per a identificar biomolecules i busca créixer amb l'aportació de la comunitat científica. Els resultats es publiquen en *Chemometrics and Intelligent Laboratory Systems*.

November 27, 2025

Investigadors de la Universitat Oberta de Catalunya (UOC) i de l'Institut de Ciències Fòniques (ICFO) han creat una base de dades espectral Raman accessible i oberta a la comunitat científica que conte 140 biomolecules dels tipus principals, com àcids nucleics, proteïnes, lípids, carbohidrats, etc. L'espectroscòpia Raman és una tècnica que permet analitzar la composició química i l'estructura molecular de materials mitjançant la interacció de la llum amb la matèria, específicament en l'anomenat fenomen de dispersió Raman, descobert pel físic Chandrasekhara Venkata Raman el 1928.

L'article, *Open Raman spectral library for biomolecule identification*, publicat en obert a la

revista Chemometrics and Intelligent Laboratory Systems, ha estat liderat per la UOC, amb la col·laboració dels investigadors de l'ICFO **Jose Javier Ruiz** i el **Dr. Pablo Loza-Alvarez**.

"Fins ara, una de les limitacions del potencial de Raman en aplicacions biomèdiques era la falta de dades espectrals obertes de biomolècules. Per això ens vam proposar crear una biblioteca accessible, estandaritzada i útil per a la comunitat científica, que serveixi de base per a futures recerques i aplicacions clíniques", explica Marcelo Teran, primer autor de l'article.

En el projecte, els investigadors han implementat dos algorismes de cerca que han mostrat una precisió del 100 % tant en el top 10 de la identificació de molècules -per exemple, el col·lagen- com en la identificació del tipus de molècula -per exemple, proteïna-, pel que fa als mesuraments de biomolècules pures i a l'hora de replicar els resultats d'estudis anteriors

Dades biomèdiques en obert per fer avançar la medicina

"L'espectroscòpia Raman permet analitzar la composició química de mostres de manera no invasiva, la qual cosa és molt valuosa en l'àmbit mèdic. Aquesta base de dades pot facilitar la identificació precisa de biomolècules i, en el futur, contribuir a estudiar com varia la seva presència en processos biològics com el càncer", assegura Teran. "La disponibilitat de dades biomèdiques de qualitat és essencial per avançar en el desenvolupament de solucions basades en intel·ligència artificial. Aquesta necessitat va ser el punt de partida de la recerca" explica l'investigador de la UOC

Els investigadors van recopilar dades d'espectres Raman de biomolècules a partir dels articles més rellevants publicats en aquest àmbit i van desenvolupar un algorisme amb tècniques clàssiques de visió per ordinador per extreure'n automàticament les dades. Un dels reptes d'aquest treball va ser el baix nombre de dades espectrals publicades en obert que van superar utilitzant validacions experimentals. "El nostre treball ofereix una eina que pot ajudar a identificar de manera objectiva, ràpida i estandaritzada la composició molecular a partir del seu espectre Raman. Actualment, aquesta identificació es fa mitjançant l'anàlisi visual dels pics principals dels espectres i es compara amb les referències de la literatura. La nostra eina pot agilitzar aquest procés al mateix temps que proporciona un solució estàndard que redueix el biaix humà durant l'anàlisi", afirma el doctorand

Una base de dades destinada a créixer amb l'aportació de la comunitat

De cara al futur, els investigadors esperen que la comunitat científica contribueixi a ampliar la base de dades, de manera que es converteixi en una biblioteca col·laborativa i de referència per a espectres Raman de biomolècules

"Encara és poc comú que en els articles científics es comparteixin les dades obertament, especialment en l'àmbit de l'espectroscòpia Raman. Aquesta manca d'accés a dades limita considerablement la recerca biomèdica. Perquè la IA pugui aplicar-se amb èxit, necessitem grans volums de dades fiables i accessibles, i és aquí on els projectes de ciència oberta tene

un paper clau", explica Marcelo Teran

L'objectiu es que, a mesura que es vagi ampliant la base de dades, es pugui impulsar l'entrenament de models d'intel·ligència artificial en l'àmbit de l'anàlisi molecular de mostres biològiques. Això obrirà la porta a noves aplicacions en diagnòstic i monitoratge de malalties

Referència:

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>

Agraïments:

The authors acknowledge funding from Fundació 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 Innovación PID2021-122807OB-C31 and PID2022-138721NBI00 projects funded by MCIN/AEI/10.13039/501100011033/FEDER, UE; CARET project. The SLN facility corresponds to a 1/2 Grup reconegut i 1/2 2021 SGR 01456 Departament de Recerca i Universitats de la Generalitat de Catalunya.