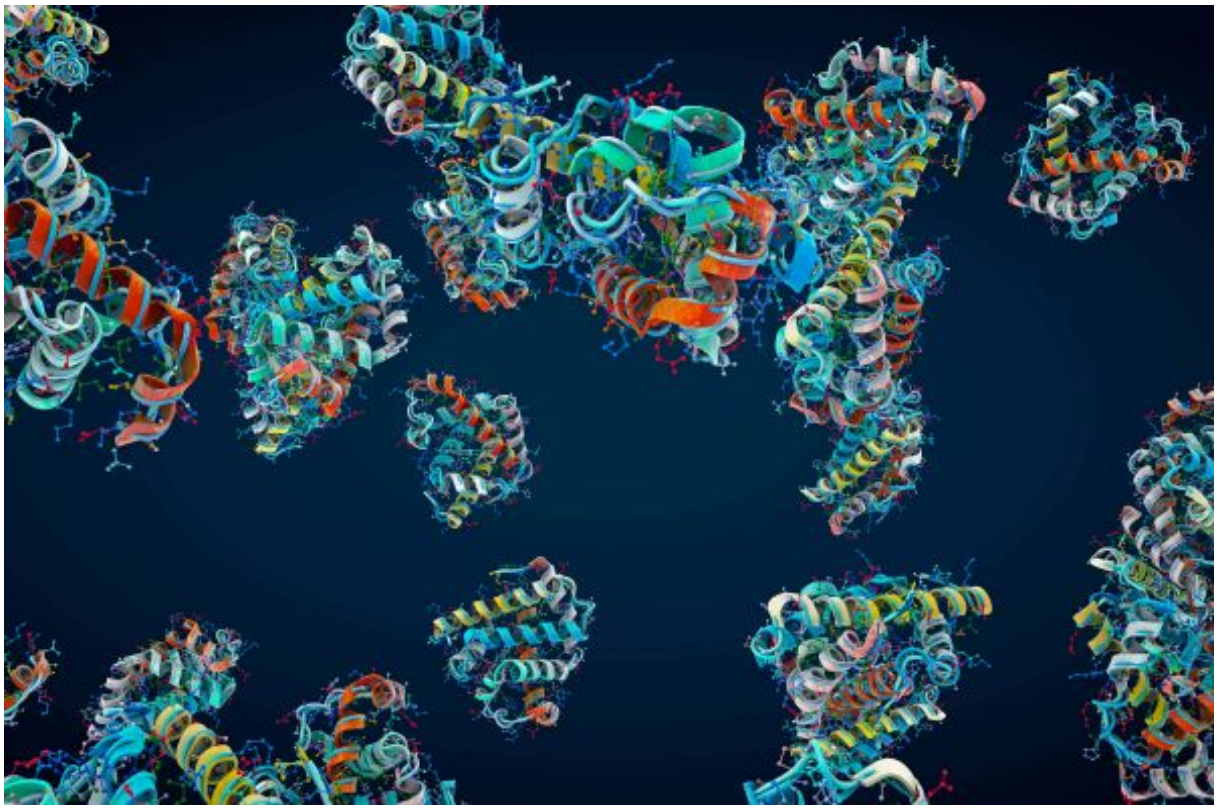

Creen una biblioteca oberta pionera per a la identificacio de biomolecules

La UOC i l'ICFO presenten un recurs clau per fer avancar la investigacio de malalties como el cancer. El projecte, basat en la tecnica analitica de la espectroscopia Raman, consisteix en una base de dades en obert per a identificar biomolecules i busca creixer amb l'aportacio de la comunitat cientifica. 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 Ciencies Fotoniques (ICFO) han creat una base de dades espectral Raman accessible i oberta a la comunitat cientifica que conte 140 biomolecules dels tipus principals, com acids nucleics, proteines, lipids, carbohidrats, etc. L'espectroscopia Raman es una tecnica que permet analitzar la composicio quimica i l'estructura molecular de materials mitjancant la interaccio de la llum amb la materia, especificament en l'anomenat fenomen de dispersio Raman, descobert pel fisic 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·laboracio 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 biomediques era la falta de dades espectrals obertes de biomolecules. Per aixo ens vam proposar crear una biblioteca accessible, estandarditzada i util per a la comunitat cientifica, que serveixi de base per a futures recerques i aplicacions cliniques", explica Marcelo Teran, primer autor de l'article.
En el projecte, els investigadors han implementat dos algoritmes de cerca que han mostrat una precisio del 100 % tant en el top 10 de la identificacio de molecules -per exemple, el col·lagen- com en la identificacio del tipus de molecula -per exemple, proteina-, pel que f als mesuraments de biomolecules pures i a l'hora de replicar els resultats d'estudis anteriors

Dades biomediques en obert per fer avancar la medicina
"L'espectroscopia Raman permet analitzar la composicio quimica de mostres de manera no invasiva, la qual cosa es molt valuosa en l'ambit medic. Aquesta base de dades pot facilitar la identificacio precisa de biomolecules i, en el futur, contribuir a estudiar com varia la seva presencia en processos biologics com el cancer", assegura Teran. "La disponibilitat de dades biomediques de qualitat es essencial per avancar en el desenvolupament de solucions basades en intel·ligencia 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 biomolecules a partir del articles mes rellevants publicats en aquest ambit i van desenvolupar un algoritme am tecniques classiques de visio per computador per extreure'n automaticament les dades. U 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 qu pot ajudar a identificar de manera objectiva, rapida i estandarditzada la composici molecular a partir del seu espectre Raman. Actualment, aquesta identificacio es fa mitjancan l'analisi visual dels pics principals dels espectres i es compara amb les referencies de l literatura. La nostra eina pot agilitzar aquest proces al mateix temps que proporciona un solucio estandard que redueix el biaix huma durant l'analisi", afirma el doctorand

Una base de dades destinada a creixer amb l'aportacio de la comunitat
De cara al futur, els investigadors esperen que la comunitat cientifica contribueixi a ampliar la base de dades, de manera que es converteixi en una biblioteca col·laborativa i d referencia per a espectres Raman de biomolecules
"Encara es poc comu que en els articles cientifics es comparteixin les dades obertament, e especial en l'ambit de l'espectroscopia Raman. Aquesta manca d'acces a dades limit considerablement la recerca biomedica. Perque la IA pugui aplicar-se amb exit, necessit grans volums de dades fiables i accessibles, i es aqui on els projectes de ciencia oberta tene

un paper clau", explica Marcelo Teran

L'objectiu es que, a mesura que es vagi ampliant la base de dades, es pugui impulsa l'entrenament de models d'intel·ligencia artificial en l'ambit de l'analisi molecular de mostr s biologiques. Aixo obrira la porta a noves aplicacions en diagnostic i monitoratge de malaltie

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

Agraiments:

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.