

Omics Automation using Large Language Models

Safe, scalable flow chemistry for domestic Lorazepam production with >99% purity.

Researchers at Purdue have created a comprehensive automated workflow generation software (dubbed CLAW) to assist with the study and rapid identification of lipids and their isomerization using mass spectroscopy (MS) and liquid chromatography (LC). Current MS and multiple reaction monitoring (MRM)-based profiling methods cannot differentiate isomeric lipids with carbon-carbon double bonds (C=C). Without a way to differentiate lipid isomers with C=Cs, the roles and behaviors of lipids cannot be fully understood, including possible new disease targets and biomarkers. For example, the ability to distinguish between different unsaturated lipid isomers has proven useful for differentiating between healthy and cancerous cells. Use of ozone electrospray ionization (OzESI) and subsequent MS analysis has been used to determine the isomerization of lipids with C=Cs. This method is successful, but the lipid samples and MS data can quickly become unmanageable in high throughput lipidomic assays, necessitating the creation of automated lipidomic OzESI-MS workflows.

The researchers designed CLAW such that once the lipid samples have been labeled and entered into the software, a workflow will be generated for the researcher that involves MS measurement of samples before and after ozonolysis with OzESI in order to determine possible C=C placement and isomerization. Post-MS measurement, lipid fragments are run through LC to ascertain their identity based off their relative retention times. Using the collected data from the MS and LC experiments, the system automatically references known lipid MS and LC data from literature to determine the identity and isomerization of lipid samples. Additionally, the user can interact with CLAW through a GUI or a novel, large language model-enabled natural language user interface (LUI) in which the user can communicate with custom AI agents to help manipulate and analyze lipidomic data. The added functionality provided by the LUI allows users who may not have the proper programming experience to more easily process and manipulate the data generated from experiments using CLAW.

Technology ID
2024-CHOP-70566

Category
Pharmaceuticals/Drug Discovery
& Development

View online page



Technology Validation:

To verify the ability of the CLAW system to accurately identify the lipids and their isomerization, the researchers measured the lipid profiles of canola oil at different stages of the refinement process. Triacylglycerols (TG) were targeted for measurement due to their high abundance in canola oil and their C=C placement variability. To measure the lipid profile of TGs for each sample, each sample was split in two, and one of each pair of samples were processed by OzESI. Each pair of samples were then measured with MRM to determine their mass spectra, finally, ion fragments resulting from MRM measurement were run through LC to measure their retention times to ensure accurate lipid identification. Using CLAW's statistical analysis of the MRM and LC data, it was found that isomeric ratios of TGs with fatty acids 18:1-n-9 and 18:1-n-7 remained relatively similar across the different stages of canola oil refinement process. To confirm this result, the researchers manually analyzed the MRM and LC data and found their results to be consistent with that of CLAW's with a standard deviation of 0.75 between both methods.

To test the ability of CLAW to evaluate lipid profiles in a biological context, the researchers collected intracellular lipid droplets (LD) from specific brain regions of 5xFAD (an Alzheimer's disease model) and age-matched WT male mice. Similar to the process described above, the samples were split in pairs, with one sample from each pair getting processed with OzESI. All samples were then run through the MRM system and subsequently LC. After lipid identification and bioinformatic analysis by the CLAW system, lipid profiles were constructed for different regions of the 5xFAD and WT mice. They discovered that fatty acid 24:6, 24:5, and phosphatidylglycerol 32:5 were downregulated in all 5xFAD samples, conversely, 22:3 and 22:2 campestral ester, 22:0 and 22:1 cholesteryl ester were significantly upregulated in 5xFAD mice.

Advantages:

- Significantly reduced analysis time
- System can determine isomerization of lipids
- Can be used for high throughput applications
- Novel use of LUI for easier data processing and manipulation

Applications:

Explore other available products test at [The Office of Technology Commercialization Online Licensing Store](#)

- Characterization and identification of lipids and their isomerization
- Search for novel lipid-based disease biomarkers
- R&D for lipid nanoparticles

TRL: Pharmaceuticals

Intellectual Property:

Provisional-Gov. Funding, 2024-02-16, United States

PCT-Gov. Funding, 2025-02-17, WO

Keywords: AI Agent, Automation, Chemistry and Chemical Analysis, Computer Technology, Large Language Model, Lipidome, Lipids, LLM, Mass Spectroscopy