

Универзитет у Београду – Хемијски факултет
Наставно-научно веће

Предмет: Образложење предлога теме докторске дисертације кандидата Давида Пирића, мастер физикохемијичара

Молим Наставно-научно веће Универзитета у Београду – Хемијског факултета да одобри израду докторске дисертације под радним насловом:

Тема: „Карактеризација липида хумане плазме у одабраним канцерима применом машинског учења, мултиваријантне анализе и анализе обогаћења података”

1. Научна област: Хемијске науке

Ужа научна област: Аналитичка хемија, Биоинформатика

2. Предмет научног истраживања

Предмет истраживања предвиђен овом докторском дисертацијом је квалитативна и квантитативна анализа липида крвне плазме особа оболелих од одређених врста канцера, и примена различитих метода мултиваријантне анализе и машинског учења у обради липидомичких података, као полазне основе за изградњу класификационих и других предиктивних модела, којима би се предвиђале карактеристике узорака оболелих и здравих особа само на основу профила липидних молекула из плазме. Увиди из класификационих модела у комбинацији са познавањем липида плазме би донели нова биохемијска сазнања о улози липидних молекула у формирању и развоју датих карцинома, те би указали на измене у метаболизму липида. Додатни предмет планиране докторске тезе је формирање претраживе квантитативне базе података која садржи све липиде одређене у хуманој плазми (и серуму).

Планирана истраживања у докторској тези била би подељена на три целине:

- 2.1. Прва целина би била експериментална студија, која подразумева липидомичку анализу узорака плазме 17 пацијенткиња оболелих од дифузног *B* крупноћелијског лимфома (енгл. *diffuse large B-cell lymphoma, DLBCL*), као и узорака здравих жена ($n = 21$), које чине контролну групу. Студија би била изведена у складу са Етичком дозволом Клиничко-болничког центра „Земун” од 16. априла 2024. године, и омогућила би идентификацију и квантификацију око 250 липида (фосфолипиди, лизофосфолипиди, плазмалогени, сфингомијелини, церамиди и диацилглицероли), путем најсавременије масено-спектрометријске технологије спрегнуте са течном хроматографијом ултра-високих перформанси. Добијени резултати би били анализирани статистичким и хеометријским методама, у циљу откривања панела липидних једињења различито заступљених у *DLBCL* и контролној групи (енгл. *differential abundant lipids, DALs*), као и развоја предиктивних мултиваријантних модела. Ови експерименти би представљали основу за каснија истраживања липидних врста као потенцијалних дијагностичких маркера овог *B*-ћелијског лимфома.
- 2.2. Другу целину ове докторске тезе чини потпуно теоријско истраживање, са циљем проналажења липидних врста, као потенцијалних биомаркера који би диференцирали узорке серума и плазме особа оболелих од одређених врста солидних канцера и здравих испитаника. Други циљ је креирање прогностичких модела обољења и осталих карактеристика везаних за одређене канцере на основу липидног профила плазме, применом различитих метода машинског учења. Да би били тачни, прогностички модели захтевају велики број узорака/испитаника, те би се у овом делу користили подаци, то јест

табеле са концентрацијама на стотине липидних молекула измерених у плазми стотина људи оболелих од веома заступљених облика канцера (као што су канцер дојке и простате), као и од канцера панкреаса који има јако ниску стопу преживљавања, а који су доступни у отвореном приступу. Ови подаци су објављени под лиценцом *CC BY 4.0* (<https://www.creativecommons.org/licenses/by/4.0>), која дозвољава коришћење података уз услов да се цитира извор. Неки од података о концентрацијама липида хумане плазме (или серума) над којим ће се применити биоинформатичка анализа су следећи:

- подаци за канцер бубрега (Jirásko, R., Idkowiak, J., Wolrab, D. *et al.* Altered plasma, urine, and tissue profiles of sulfatides and sphingomyelins in patients with renal cell carcinoma. *Cancers* **2022**, *14*, 4622. <https://doi.org/10.3390/cancers14194622>):

<https://www.mdpi.com/2072-6694/14/19/4622/s1>, приступљено 10.03.2024;

- подаци за канцер бубрега, дојке и простате (Wolrab, D., Jirásko, R., Peterka, O. *et al.* Plasma lipidomic profiles of kidney, breast and prostate cancer patients differ from healthy controls. *Sci Rep* **2021**, *11*, 20322. <https://doi.org/10.1038/s41598-021-99586-1>):

https://www.static-content.springer.com/esm/art%3A10.1038%2Fs41598-021-99586-1/MediaObjects/41598_2021_99586_MOESM1_ESM.xlsx, приступљено 10.03.2024;

- подаци за канцер панкреаса (Wolrab, D., Jirásko, R., Cífková, E. *et al.* Lipidomic profiling of human serum enables detection of pancreatic cancer. *Nature Communications* **2022**, *13* (1), 124. <https://doi.org/10.1038/s41467-021-27765-9>):

https://www.static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27765-9/MediaObjects/41467_2021_27765_MOESM4_ESM.xlsx, приступљено 10.03.2024.

Преузети подаци би били основа за изградњу класификационих и других предиктивних модела за поменуте солидне канцере, применом надгледаних метода машинског учења, које би биле строго валидиране. Такође је планирано тестирање ненадгледаних метода као бинарних класификатора у исту сврху. У циљу извлачења биохемијских информација о функционалним променама липидома плазме који прати развој канцера, радило би се напредно испитивање података (енгл. *data mining*), у оквиру кога су планиране: 1) анализе обogaћења (енгл. *enrichment analysis*) *DALs* из плазме у смислу њихове припадности липидним (под)класама, броју угљеникових атома и двоструких веза у молекулу, садржају појединих масних киселина, унутарћелијској локацији и слично; 2) функционална анализа обogaћења (енгл. *pathway analysis*) којом се утврђују специфичне метаболичке путање у којима учествују *DALs*. За ово би се употребили алгоритми развијени код куће, паралелно са софтверским пакетима и Веб алаткама. Идеја је да пакети развијени у оквиру ове тезе отклоне недостатке постојећих Веб алатки, које представљају уско грло у анализи липидомичких података; 3) анализа публикованих радова на тему промене липидома у одабраним и сличним канцерима употребом алгоритама за напредно испитивање текстуалних података (енгл. *text mining*). Наиме, без напредног испитивања текстуалних података, карактеризација липидома код канцера се своди на пуку листу липида чије су концентрације значајно измењене у односу на здраву популацију.

2.3. У оквиру треће целине била би креирана јединствена претражива и квантитативна липидомичка база података специјализована за хуману плазму (и серум), која би садржавала све до сада идентификоване липидне молекуле у овим медијумима, на свим нивоима структурне резолуције, праћене свим номенклатурним системима прихваћеним у липидомичкој заједници, уз одговарајуће семантичке и биоинформатичке идентификаторе, који би водили ка постојећим базама (претежно мешовито метаболичко-липидомичким). Оваква база би омогућавала брже, лакше и поузданије претраживање липидних врста измерених у плазми/серуму и надоместила бројне недостатке и проблеме везане са претраживањем липида по постојећим базама (*LIPID MAPS*, *SwissLipid*, *ChEBI*, *PubChem*, *HMBD*, *ChemSpider*), као што су: да ниједна база не

садржи опсег концентрација липида, као ни све липидне врсте из хумане плазме, да у одређеним базама постоје различити идентификатори за једну исту липидну врсту, нејасноће око нивоа структурне резолуције липида унетих у базе, јер структура базе која није хијерархијска отежава проналажење изомера који произилазе из истог липида, итд. Направљена база би се касније могла интегрисати у *LIPID MAPS*, базу отвореног приступа, где би концентрације липида мерене у оквиру међународних липидомичких конзорцијума, биле доступне свима за поређење.

3. Основне хипотезе

Основне хипотезе које би се истраживањима у оквиру ове докторске дисертације потврдиле или одбациле су следеће:

- 3.1. Постоје липидне врсте, класе и подкласе липида у хуманој плазми чија је концентрација значајно измењена код жена са *DLBCL* канцером;
- 3.2. Применом хемометријских метода и алгоритама машинског учења на сетове података добијене липидомиком из плазме особа са одабраним канцерима могуће је издвојити додатне или другачије липидне врсте од оних на које указује класична статистичка обрада података, а који би се могли довести у везу са патологијом одабраних канцера;
- 3.3. Применом метода машинског учења, укључујући и дубоко учење, могуће је изградити ефикасније бинарне класификационе моделе и предиктивне моделе за одабране врсте канцера, који превазилазе оне добијене линеарним методама, као што је *OPLS-DA*;
- 3.4. Синтезом информација добијених разноврсним методама напредног испитивања података, као што су анализе обогаћења, анализе метаболичких путања, и напредно испитивање текстуалних података научне литературе могуће је прикупити највише информација о измењеном липидому код одабраних врста канцера (енгл. *information gain*);
- 3.5. Прављење нове, претраживе и квантитативне базе података липида из хумане плазме (и серума) би убрзало анализу података добијених липидомиком ових медијума, и помогло да се уоче нови липидни молекули, који не постоје ни у једној бази, што би представљало основу за будући липидни атлас хумане плазме.

4. Циљ истраживања и очекивани резултати

Промене у липидном статусу се по правилу уочавају код пацијената оболелих од канцера, јер је канцер увек у спреси са променама у метаболизму липида [1]. Стога је познавање састава липида, пре свега у крвном серуму и плазми од изузетне важности у праћењу, дијагностици и раној детекцији одређених туморских болести [1]. Применом липидомике као аналитичке технике пре свега засноване на хроматографским методама спрегнутим са масеном детекцијом може се квалитативно и квантитативно окарактерисати липидом у једном биолошком систему, и остварити увид у биохемијске улоге свих липидних врста, њихове интеракције са другим биомолекулима и интеграцију у метаболичке и сигналне путеве [2]. Аутоматизацијом процеса екстракције липида и спрезањем течне хроматографије ултра високих перформанси (*UHPLC*) са високо осетљивим хибридном масеним спектрометрима, уз комерцијалну доступност стотине ултра-пречишћених липидних интерних стандарда, липидомика је све присутнија у изучавању канцера. У оваквим истраживањима се најчешће користи хумана плазма (или серум), због њене доступности, лакше екстракције липида (у односу на ткива) и веће хемијске стабилности липида из плазме.

Савремени липидомички ток омогућава одређивање и преко хиљаду различитих липидних молекула, са високом структурном резолуцијом, са лимитима квантификације на атомолском нивоу, што омогућава анализу до сада мало истраживаних липида [2]. Овакав аналитички ток резултира у великој количини података (енгл. *big data*). Типична „матрица” података садржи

концентрације неколико стотина (или више) липидних врста измерених у десетинама или стотинама узорака. Обрада, анализа и интерпретација оваквих података превазилазе оквире класичне статистичке анализе [3], те се морају користити технике напредне обраде података. Додатни проблем стварају и вредности које недостају у подацима, нелинеарност, непараметарска дистрибуција променљивих, хетероскедастичност, мултиколинеарност и сл. [4]. Такође, у студијама са малим бројем узорака у односу на број измерених параметара, често долази до предетерминације модела (енгл. *overfitting*) што представља велики проблем, нарочито за мултиваријантне анализе које су у најширој употреби у липидомици. Упркос томе, огромна већина аутора липидомичких публикација је задовољна комбинацијом класичне статистичке анализе, анализе главних компоненти (*PCA*) и (*O*)*PLS-DA*. Са друге стране, примена напредних нелинеарних метода, као што су регресионе шуме, метода потпорних вектора и других техника машинског учења, релативно отпорних на одступања у структури података, даје боље предиктивне моделе. Па ипак, поменути алгоритми машинског учења наилазе на изазове и захтевају ангажовање високих рачунарских ресурса [5]. Осим пар студија које користе дубљу анализу липидомичких резултата у канцеру, већина се ослања на квали/кванти анализе липида [2]. *FAIR* принципи управљања подацима (енгл. *findability* – могућност налажења, *accessibility* – доступност, *interoperability* – компатибилност, *reusability* – могућност поновне употребе) дају додатну вредност истраживању. Овако радећи могуће је користити литературно доступне сетове у циљу извлачења додатних биохемијских сазнања [6].

Такође, развој функционалне липидомике заостаје за развојем аналитичких платформи, пре свега услед недостатака одговарајућих биоинформатичких алата. Потенцијална решења су побољшане и нове Веб алатке за дубоку анализу липидомичких података, као и синтеза информација добијених појединачним алатима. Стартна позиција би била тзв. анализа обogaћења (енгл. *enrichment analysis* или *over-representation analysis*) [7], која представља анализу комбинација основних липидних варијабли, које прате пут метаболичких реакција, биосинтезе и разградње појединачних класа липида. Један од најбитнијих аспеката функционалне липидомике је детекција метаболичких и сигналних путева погођеним одређеним канцером и утврђивање природе измена на њима. То омогућује функционална анализа обogaћења или анализа путања (енгл. *functional enrichment analysis* или *pathway analysis*) [8]. Међутим, софтверске алатке намењене овом задатку (нпр. *LIPEA* и *BioPAN*) су далеко од идеалних, а највећа мана им је што се излазни резултати разликују у зависности од липидних варијабли одабраних за унос и избора базе метаболичких путања. На анализу обogaћења утиче још и динамички опсег концентрација липида, који се протеже преко четири реда величине. Употреба неконзистентне номенклатуре липида такође је велики проблем за анализу липидних путања упркос томе што је *LIPID MAPS* препоручио јединствени стандардни систем за означавање липидних структура [9]. Овај хијерархијски систем прати ниво структурне резолуције липида који зависи од примењене масено-спектрометријске методе. Слично као и код софтвера за анализе обogaћења, често се догађа да Веб алатка не препознаје имена појединих унетих липида, упркос стандардном имену, јер многи липиди не постоје у бази података са којом је дата Веб алатка спрегнута. С друге стране, сваки липидни молекул, са комплетно одређеном структуром, има јединствени хемоинформатички идентификатор (енгл. *InChI key*), али то нема практичну примену у софтверским алаткама, јер већина липидомичких студија идентификује липиде на нижем нивоу структурне резолуције, које немају јединствен идентификатор [10]. Ови проблеми би се могли решити прављењем нове базе липидних структура са јединственим идентификаторима за све липидне врсте на свим нивоима структурне резолуције, која би касније могла да послужи као основа за прављење липидног атласа.

Стога су научни циљеви ове дисертације:

Идентификација липида, и њихових класа и подкласа различито заступљених у плазми оболелих од *DLBCL* канцера у односу на контролну групу, те формирање класификационих модела који би омогућили разликовање узорака крвне плазме жена са *DLBCL* канцером од контролних узорака и помогли у његовој дијагностици, раном откривању и лечењу.

Други циљ је изградња класификационих и прогностичких модела за одабране типове солидних канцера (нпр. бубрега, дојке, простате и панкреаса), на основу литературно доступних липидних профила плазме (или серума), применом различитих хеометријских метода и алгоритама машинског учења, који су уско повезани са претходним циљем – дијагностика, лечење и рано откривање. Трећи циљ је развој метода за евалуацију и поређење модела добијених различитим хеометријским методама.

У оквиру другог дела истраживања посебан циљ је синтеза биохемијских информација о природи измена на липидому плазме, као што је на пример активација/репресија специфичних ензима учесника у метаболизму липида код одабраних врста канцера. У овом смислу би се примениле анализе обogaћења и анализа липидних метаболичких путања, уз дубоко истраживање података доступних из литературе о измењеном липидому код тих и сличних типова канцера.

Циљ трећег дела истраживања је да се направи претражива квантитативна база података липидних молекула из хумане плазме (и серума), која би представљала основу за липидни атлас хумане плазме.

Очекује се да предложена докторска теза оствари све задате циљеве, да се идентификују липидне врсте измењене у одабраним канцерима, и да се на основу липидомичких сетова података изграде класификациони и предиктивни модели, који ћи имати високу тачност, прецизност, осетљивост и специфичност у раздвајању узорака плазме/серума пацијената оболелих од канцера и контролних узорака. Посебно се очекује да нелинеарне методе машинског учења омогуће успешније предиктивне моделе са високим степеном тачности класификације канцерских и контролних узорака. Даље, очекује се да анализе обogaћења и анализе метаболичких и сигналних путања омогуће утврђивање измењених биохемијских процеса код испитиваних канцера. На крају, очекује се да нова претражива квантитативна база липида из хумане плазме нађе своје место међу Веб алаткама и буде инкорпорисана у највећу липидну базу података – *LIPID MAPS* (<https://www.lipidmaps.org>).

5. Методе истраживања

5.1. У првом делу истраживања примениће се методе екстракције липида из хумане плазме пацијенткиња оболелих од *DLBCL* канцера и контролне групе, методом по *Matyash*-у, а добијени екстракти ће се анализирати применом ултра-ефикасне течне хроматографије повезане са хибридним масеним спектрометром, додатно опремљеним инструментом који мери јонску покретљивост (*UHPLC-TIMS-TOF-MS*). Резултат ове анализе ће бити липидомички сет података са концентрацијама липидних врста који би био анализиран различитим биоинформатичким методама.

5.2. У другом делу истраживања би се липидомички сетови података преузети из литературе и извора са отвореним приступом за одабране типове солидних канцера (бубрега, дојке, простате и панкреаса), анализирали применом:

5.2.1. Основне батерије статистичких метода за експлоративну анализу, визуелизацију података, трансформацију варијабли, и уклањања дефеката у структури података, те параметријских и непараметријских тестова за тестирање значајности зависних и независних података.

5.2.2. Хеометријских техника за ненадгледано и надгледано линеарно моделовање у циљу препознавања образаца, класификације и предвиђања својстава узорака плазме и серума пацијената оболелих од одређених солидних тумора попут: анализе главних компоненти (енгл. *principal component analysis*, *PCA*),

стохастичког уметања *t*-дистрибуираних суседа (енгл. *t-distributed stochastic neighbor embedding*, *t-SNE*), линеарне дискриминантне анализа (*LDA*), дискриминантне анализе методом најмањих квадрата (*PLS-DA*), дискриминантне анализе методом пројекција на латентне структуре (*OPLS-DA*) и сл.

5.2.3. Напредних метода машинског учења са циљем развоја предиктивних модела високих перформанси попут: регресионих шума (енгл. *random forest*), метода потпорних вектора (енгл. *support vector machines*), метода екстремног градијентног појачања (енгл. *extreme gradient boosting*); само-организујуће мапе (енгл. *self-organizing maps*, *SOM*) и неуронске мреже (нпр. вишеслојни перцептрони, енгл. *multiple layer perceptron*, *MLP*);

5.2.4. Анализа обogaћења и анализа путања: путем Фишеровог и хипергеометријског теста примењених на изведене липидне варијабле добијене комбинацијама основних липидних варијабли на основу припадности липидним (под)класама, садржаја засићених, незасићених и полинезасићених масних киселина у фосфолипидима, на основу укупног броја угљеникових атома у масним киселинама и броја двоструких веза. Такође би се испробале и Веб алатке за анализу обogaћења.

5.3. У трећем делу истраживања користиће се различите технике прикупљања, напредне анализе, организације и управљања подацима (релацијског система за управљање базама података *MySQL*, програмских језика *PHP* и *Javascript*, као и *HTML*) са циљем прављење интерне претраживе базе липида у машински читљивом формату, која би садржала укупност липидних врста, које су до сада детектоване и/или квантификоване у хуманој плазми (и серуму), на свим нивоима структурне резолуције, праћене свим номенклатурним системима прихваћеним у липидомичкој заједници, уз одговарајуће семантичке и биоинформатичке идентификаторе, који би водили ка постојећим базама (претежно мешовито метаболичко-липидомичким). Такође ће се изводити анализа публикованих радова на тему промене липида у одабраним канцерима употребом метода напредног испитивања текстуалних података, као и њиховом визуелизацијом (нпр. облак речи, енгл. *word cloud*).

У овој докторској тези биће коришћени следећи програми и софтверски алати:

- R програмски језик и његови пакети (<https://www.r-project.org>, <https://www.bioconductor.org>) као главни софтвер за статистику и хеометрију и као помоћни софтвер за напредне технике машинског учења; и укључиваће примену следећих пакета: за униваријантну статистику ‘*car*’, ‘*onewaytests*’, ‘*exactRankTests*’, ‘*coin*’, ‘*perm*’; корелације: ‘*Hmisc*’, ‘*ggcorrplot*’; за хеометријске моделе: ‘*factoextra*’, ‘*ropls*’; за методе машинског учења: ‘*party*’, ‘*randomForest*’, ‘*e1071*’; за визуелизацију: ‘*ggplot2*’, ‘*gridExtra*’, ‘*ggpubr*’, ‘*igraph*’, ‘*rgl*’; за манипулацију сликама: ‘*magick*’, ‘*imager*’;
- *python* програмски језик (<https://www.python.org>) и пакети из *Anaconda* окружења за науку о подацима засновани на овом програмском језику (<https://www.anaconda.com>) који укључују већину најчешће коришћених *python* библиотека за хеометрију и машинско учење, као што су: ‘*scikit-learn*’, ‘*Keras*’, ‘*Matplotlib*’, ‘*pandas*’, ‘*NumPy*’, ‘*PyTorch*’; ове библиотеке би биле коришћење као помоћни софтвер за статистику и хеометрију и као главни софтвер за напредне технике машинског учења;
- слободни софтвер *JASP* (<https://www.jasp-stats.org>) и *jamovi* (<https://www.jamovi.org>) за хеометрију и машинског учења;
- комерцијални статистички софтвер: *SPSS* (<https://www.ibm.com/spss>), *GraphPad* (<https://www.graphpad.com>) и *SIMCA* (<https://www.sartorius.com/en/products/process-analytical-technology/data-analytics-software/mvda-software/simca>);
- *VOSviewer* (<https://www.vosviewer.com>) за напредно испитивање текстуалних података и њихову визуелизацију.

- Веб алат *MetaboAnalyst* (<https://www.metaboanalyst.ca>) за анализу липида, анализу обогаћења и анализу путања;
- базе отвореног приступа: *LIPID MAPS* (<https://www.lipidmaps.org>), *SwissLipid* (<https://www.swisslipids.org>), *ChEBI* (<https://www.ebi.ac.uk/chebi>), *PubChem* (<https://www.pubchem.ncbi.nlm.nih.gov>), *HMDB* (<https://www.hmdb.ca>), *ChemSpider* (<https://www.chemspider.com>);
- Веб алати: *LipidSig* (<http://www.chenglab.cmu.edu.tw/lipidsig>) за анализу липидомичких података, *LIPEA* (<https://www.hyperlipea.org>) за анализу путања, и *LION/web* (<http://www.lipidontology.com>) за анализу обогаћења.

6. Литература

- [1] Butler, L. M.; Perone, Y.; Dehairs, J.; Lupien, L. E.; de Laat, V.; Talebi, A.; Loda, M.; Kinlaw, W. B.; Swinnen, J. V. Lipids and Cancer: Emerging Roles in Pathogenesis, Diagnosis and Therapeutic Intervention. *Adv Drug Deliv Rev* 2020, 159, 245–293. <https://doi.org/10.1016/j.addr.2020.07.013>.
- [2] Han, X.; Gross, R. W. The Foundations and Development of Lipidomics. *J Lipid Res* 2022, 63 (2), 100164. <https://doi.org/10.1016/j.jlr.2021.100164>.
- [3] Kvasnička, A.; Najdekr, L.; Dobešová, D.; Pisklákova, B.; Ivanovová, E.; Friedecký, D. Clinical Lipidomics in the Era of the Big Data. *Clin Chem Lab Med* 2023, 61 (4), 587–598. <https://doi.org/10.1515/cclm-2022-1105>.
- [4] Ni, Z.; Wölk, M.; Jukes, G.; Mendivelso Espinosa, K.; Ahrends, R.; Aimo, L.; Alvarez-Jarreta, J.; Andrews, S.; Andrews, R.; Bridge, A.; *et al.* Guiding the Choice of Informatics Software and Tools for Lipidomics Research Applications. *Nat Methods* 2023, 20 (2), 193–204. <https://doi.org/10.1038/s41592-022-01710-0>.
- [5] Mohammed, M. A.; Abdulkareem, K. H.; Dinar, A. M.; Zapirain, B. G. Rise of Deep Learning Clinical Applications and Challenges in Omics Data: A Systematic Review. *Diagnostics (Basel)* 2023, 13 (4). <https://doi.org/10.3390/diagnostics13040664>.
- [6] Kopczynski, D.; Krause, D.; Al Machot, Fadi; Schwudke, D.; Hoffmann, N.; Ahrends, R. The Past and Future of Lipidomics Bioinformatics. In *Mass Spectrometry for Lipidomics*; 2023; pp 271–290. <https://doi.org/10.1002/9783527836512.ch10>.
- [7] Wieder, C.; Frainay, C.; Poupin, N.; Rodríguez-Mier, P.; Vinson, F.; Cooke, J.; Lai, R. P.; Bundy, J. G.; Jourdan, F.; Ebbels, T. Pathway Analysis in Metabolomics: Recommendations for the Use of over-Representation Analysis. *PLoS Comput Biol* 2021, 17 (9), e1009105. <https://doi.org/10.1371/journal.pcbi.1009105>.
- [8] García-Campos, M. A.; Espinal-Enríquez, J.; Hernández-Lemus, E. Pathway Analysis: State of the Art. *Frontiers in Physiology* 2015, 6 (DEC). <https://doi.org/10.3389/fphys.2015.00383>.
- [9] Liebisch, G.; Fahy, E.; Aoki, J.; Dennis, E. A.; Durand, T.; Ejsing, C. S.; Fedorova, M.; Feussner, I.; Griffiths, W. J.; Köfeler, H.; *et al.* Update on LIPID MAPS Classification, Nomenclature, and Shorthand Notation for MS-Derived Lipid Structures. *J Lipid Res* 2020, 61 (12), 1539–1555. <https://doi.org/10.1194/jlr.S120001025>.
- [10] Witting, M.; Malik, A.; Leach, A.; Bridge, A.; Aimo, L.; Conroy, M. J.; O'Donnell, V. B.; Hoffmann, N.; Kopczynski, D.; Giacomoni, F.; *et al.* Challenges and Perspectives for Naming Lipids in the Context of Lipidomics. *Metabolomics* 2024, 20 (1), 15. <https://doi.org/10.1007/s11306-023-02075-x>.