

OD KOMPLEKSNIH ZMESI DO BIOAKTIVNIH MOLEKUL: NAPREDNE STRATEGIJE ZA ODKRIVANJE BIOAKTIVNIH SPOJIN NARAVNEGA IZVORA

FROM COMPLEX MIXTURES TO BIOACTIVE MOLECULES: ADVANCED STRATEGIES FOR THE DISCOVERY OF BIOACTIVE NATURAL COMPOUNDS

AVTORJA / AUTHORS:

asist. Tilen Čuš, mag. farm.

izr. prof. dr. Nina Kočevar Glavač, mag. farm.

*Univerza v Ljubljani, Fakulteta za farmacijo,
Aškerčeva 7, 1000 Ljubljana*

NASLOV ZA DOPISOVANJE / CORRESPONDENCE:

E-mail: tilen.cus@ffa.uni-lj.si



POVZETEK

Spojine naravnega izvora ostajajo pomemben vir novih zdravilnih učinkovin, a njihovo odkrivanje ovirajo delovno intenzivni postopki izolacije, ponovna izolacija že znanih spojin in njihova zahtevna strukturna identifikacija. Zaradi tega so se razvili številni novi sodobni pristopi, ki temeljijo na integraciji računalniških metod in analitskih podatkov. Pomembno vlogo ima jedrska magnetna resonanca, ki v kombinaciji z umetno inteligenco in drugimi računalniškimi metodami omogoča napovedovanje strukture spojin tudi v kompleksnih zmesih. Ključna sta tudi napredek v masni spektrometriji ter pojav molekulskega mreženja. Slednje omogoča učinkovito obdelavo, upodobitev velike količine podatkov masne spektrometrije ter hitro prepoznavanje strukturnega razreda spojin. Dodatno ga podpirajo še nove metode za proučevanje fragmentacije spojin, ki omogočajo napovedovanje strukture spojin iz njihovih masnih spektrov. Novi pristopi tako bolje izkoristijo obilico dobljenih podatkov in omogočajo hitrejšo, usmerjeno in natančnejšo identifikacijo, dereplikacijo in izolacijo bioaktivnih spojin v naravnih izvlečkih.

KLJUČNE BESEDE:

dereplikacija, masna spektrometrija, molekulsko mreženje, spojine naravnega izvora

ABSTRACT

Natural products remain an important source of new drug candidates, but their discovery is often hindered by time-consuming procedures, repeated isolation of known compounds, and their challenging structural identification. As a result, numerous modern approaches have been developed that integrate computational methods with analytical data. Nuclear magnetic resonance combined with artificial intelligence and other computational tools enables the structural prediction of compounds even in complex mixtures. Significant progress in mass spectrometry and the introduction of molecular networking has further advanced the field. The latter allows efficient processing and visualization of large volumes of mass spectrometry data and facilitates the rapid structural elucidation of compounds in complex mixtures. These met-

hods are further supported by new tools for studying compound fragmentation, which enable structural prediction directly from mass spectra. Together, these modern approaches make better use of available data and significantly accelerate the targeted and accurate identification, dereplication, and isolation of bioactive compounds from natural extracts.

KEY WORDS:

dereplication, mass spectrometry, molecular networking, natural products

1 UVOD

Rastline, glive in drugi organizmi v preteklosti niso bili le vir hrane, temveč so jih uporabljali tudi kot plačilno sredstvo, gradbeni material in celo vir umetniškega navdiha (1). Na področju farmacije so bili izolirane spojine ter izvlečki naravnega izvora dolgo vodilni vir novih zdravilnih učinkovin in zdravil. Med bolj znanimi sta gotovo analgetik morfin iz vrtnega maka (*Papaver somniferum*) ter antimalarik artemizinin iz sladkega pelina (*Artemisia annua*), za odkritje katerega so leta 2015 podelili Nobelovo nagrado (2, 3). Med spojine naravnega izvora ne uvrščamo le rastlinskih specializiranih metabolitov, temveč tudi številne spojine, izolirane iz živali, gliv, bakterij ter arhej. Primer je antibiotik penicilin iz gliv rodu *Penicillium* ali spongouridin iz spužve *Tectitethya crypta* (2, 4).

Spojine naravnega izvora so izjemno pomembne še danes. Po zadnjih ocenah je le tretjina za zdravljenje odobrenih majhnih molekul popolnoma sinteznega izvora. Drugo tretjino predstavljajo nespremenjene učinkovine naravnega izvora, preostalo so sintezno pridobljene spojine, ki oponašajo naravni substrat ali vsebujejo strukturne motive, značilne za spojine naravnega izvora (5). Kljub široki zastopanosti spojin naravnega izvora pa so v zadnjih letih številna farmacevtska podjetja izgubila zanimanje za njihove raziskave. Med januarjem 2014 in junijem 2025 je bilo tako med vsemi odobrenimi malimi molekulami le približno 12 % popolnoma naravnih spojin in njihovih derivatov (6). Možen razlog za to je razmah sintezne kemije, reševanja visoke zmogljivosti in strukturno podprtega načrtovanja, ki omogočajo hitro testiranje velikega števila spojin in njihovo nadaljnjo optimizacijo (7–10).

2 KEMIJSKI PROSTOR SPOJIN NARAVNEGA IZVORA

V primerjavi s konvencionalnimi sintezni spojini imajo spojine naravnega izvora pogosto bistveno drugačne lastnosti. Slednje so praviloma kompleksnejše, kar se odraža v višji molekularni masi, višjem številu sp^3 hibridiziranih ogljikovih atomov in več stereogenimi centri. Vsebujejo tudi več kisikovih atomov, akceptorjev in donorjev vodikovih vezi ter manj halogenov in dušikovih atomov. Hkrati so bolj rigidne in imajo nižji izračunan porazdelitveni koeficient oktanol-voda ($\log P$). Te razlike so lahko vir tako prednosti kot slabosti. Višja molekularna masa in rigidnost sta lahko koristni pri načrtovanju modulatorjev interakcij med proteini, a lahko hkrati vodita v njihovo slabšo biološko uporabnost (8).

Strukturne značilnosti naravnih spojin so tesno povezane s taksonomskim izvorom organizmov, saj imajo filogenetsko bolj oddaljene skupine organizmov pogosto različne biosintezne poti, ki proizvajajo kemijsko različne metabolite (11). S kemijskega vidika je izjemno zanimivo kraljestvo rastlin, saj po zadnjih ocenah vsebuje med 1,5 in 25,7 milijonov edinstvenih metabolitov (12). Zanje so značilne poliaromatske spojine z več atomi kisika, kot sta recimo resveratrol ali epigalokatehinalat. Med rastlinami so izjemno razširjeni tudi glikozidi in alkaloidi, ki jih v manjši meri najdemo tudi v vseh drugih kraljestvih. Pri bakterijah pogosteje najdemo spojine z dušikom in žveplom, makrociklične spojine, aminoglikozide in poliketide. Slednji so pogosti tudi v glivah. Terpeni in terpenoidi so prisotni v vseh kraljestvih, najštevilčnejši so v kraljestvu rastlin in izjemno redki med živalmi (11, 13).

3 ODKRIVANJE SPOJIN NARAVNEGA IZVORA

Postopek odkrivanja novih zdravilnih učinkovin se začne z izbiro organizma, kar je zaradi velike kompleksnosti bioloških sistemov pogosto zapleten korak. Živi organizmi iste vrste so med seboj lahko precej raznoliki. Na njihovo kemijsko sestavo navadno vplivajo okoljski dejavniki, rastni pogoji, razvojna faza organizma in tudi variabilnost v genetski zasnovi. Tako se lahko kemijska sestava iste vrste organizma bistveno razlikuje med gojenimi in divje rastočimi osebki, kar lahko pomembno omeji njihovo pridobivanje



(14). Slednje je lahko dodatno omejeno tudi zaradi nedostopnosti rastišč, nizkega števila osebkov ali različnih nacionalnih in mednarodnih zakonodajnih zahtev. Med njimi je še posebej pomemben protokol Nagoya, ki ureja pogoje, pod katerimi smejo raziskovalci in podjetja dostopati do naravnih genetskih virov. Protokol predvideva tudi pravično delitev koristi, pridobljenih z uporabo določenega organizma (15).

Po izbiri organizma sledi priprava izvlečkov, ki je izjemno pomembna, saj lahko z različnimi postopki ekstrakcije pripravimo izvlečke z bistveno drugačnim kemijskim profilom. V drugem koraku s testi *in vitro* ovrednotimo bioaktivnost izvlečkov. Testi *in vitro* pogosto niso prilagojeni za vrednotenje izvlečkov naravnega izvora, ki vsebujejo veliko količino različnih spojin, kar lahko vodi v lažno pozitivne ali negativne rezultate in zahteva dodatne teste za potrditev zaključkov osnovnih testiranj (16).

Za odkrivanje bioaktivnih spojin v izvlečkih pogosto uporabljamo bioaktivnostno vodeno frakcionacijo. V njej začetne izvlečke frakcioniramo s pomočjo kromatografskih metod ter vrednotimo bioaktivnost posameznih frakcij. Bioaktivne frakcije nato dodatno frakcioniramo in ponovno vrednotimo njihovo bioaktivnost. Postopek večkrat ponovimo, da dobimo čim bolj čiste spojine, ki jih nato ovrednotimo z metodami, kot so masna spektrometrija (MS), jedrska magnetna resonanca (NMR) in infrardeča spektroskopija (17). Bioaktivnostno vodena frakcionacija je zelo dolgotrajna, delovno intenzivna ter zahteva relativno veliko količino začetnega izvlečka. Pogosto težavo pri njej predstavlja tudi preferenčna izolacija bolj zastopanih spojin v izvlečkih, medtem ko manj zastopanih pogosto ne uspemo izolirati. Mnogokrat se soočamo tudi s ponovno izolacijo že znanih spojin, zaradi česar so se v zadnjem času razvile metode dereplikacije, katerih cilj je hitro prepoznavanje že znanih spojin v zmesih. Dodatno težavo lahko predstavlja tudi višja molekulska masa in višje število kiralnih centrov spojin naravnega izvora, kar oteži pravilno določitev njihovih struktur (14).

4 SODOBNE METODE IDENTIFIKACIJE, IZOLACIJE IN DEREPLIKACIJE NARAVNIH SPOJIN

4.1 BIOKEMOMETRIČNE METODE

Da bi se izognili izolaciji že znanih spojin in pospešili izolacijo ter identifikacijo spojin naravnega izvora, so znanstveniki

razvili številne metode dereplikacije, katerih cilj je hitro prepoznati že znane spojine v kompleksnih zmesih (17). Primer so kemometrične metode, ki iz dobljenih analiznih podatkov izluščijo kar se da veliko informacij. Kemometrične metode delimo na nenadzorovane in nadzorovane metode. Nenadzorovane metode so deskriptivne in jih uporabljamo za poenostavitev podatkov ter njihovo razvrščanje. Mednje uvrščamo analizo glavnih komponent, ki jo uporabljamo za zmanjšanje dimenzionalnosti podatkov, in hierarhično gručenje, ki podatke glede na medsebojno podobnost razvrsti v gruče. Za nadzorovane metode je značilno, da modelirajo odnos med neodvisno in odvisno spremenljivko. Med njih uvrščamo diskriminantno analizo in linearno ter multivariatno regresijo. Metode same po sebi ne omogočajo strukturne identifikacije spojin, lahko pa so izjemno koristne, kadar želimo ugotoviti, kateri signali pripadajo posamezni spojin, s čimer pospešimo izolacijo in strukturno identifikacijo spojin (17–19).

V literaturi najdemo številne uspešne primere uporabe kemometričnih metod za pomoč pri identifikaciji bioaktivnih spojin. Grienke in sod. so biokemometriko uporabili na podatkih NMR in MS za identifikacijo bioaktivnih lanostanskih triterpenov smrekove kresilače (*Fomitopsis pinicola*) ter njihovo nadaljnjo izolacijo (19), Kellogg in sod. pa so kemometriko skupaj z MS uporabili za identifikacijo in nadaljnjo izolacijo protimikrobnih učinkovin iz endofitskih gliv (18).

4.2 METODE NA OSNOVI JEDRSKE MAGNETNE RESONANCE

Določitev strukture spojin z veliko gotovostjo je mogoča z uporabo NMR v kombinaciji z MS. Za ta namen je potrebno izvesti veliko število tako eno- in dvodimenzionalnih poskusov (HSQC, COSY, DEPT, ^1H , ^{13}C , HMBC) na relativno veliki količini čim bolj čiste spojine. Določevanje struktur spojin s pomočjo NMR pa lahko predstavlja tudi velik izziv, še posebej pri spojinah naravnega izvora, saj imajo pogosto višjo molekulska maso in več stereogenih centrov (17).

Identifikacijo novih spojin ter določitev njihove strukture iz podatkov NMR lahko bistveno pospešijo in izboljšajo metode umetne inteligence. Primera sta metodi SMART 2.0 (*Small Molecule Accurate Recognition Technology*) in DP4+ (20, 21). SMART 2.0 je metoda, osnovana na konvolucijskih nevronskih mrežah, ki uporablja podatke tisočih spektrov HSQC in omogoča napovedovanje struktur spojin v kompleksnih zmesih (21). Metoda DP4+ je zasnovana na kvantnokemijskih izračunih, konformacijski in statistični analizi. Deluje tako, da za več možnih 3D-konformacij

določene molekule izračuna pričakovane NMR-sigle in jih primerja z eksperimentalno določenimi podatki. Razlike med spektri nato ovrednoti in napove, katera konformacija ali izomerna oblika je najbolj verjetna, pri čemer je končna odločitev o strukturi še vedno odvisna od raziskovalca in njegovega strokovnega znanja (20, 22).

Čeprav je ^1H NMR bistveno bolj občutljiv kot ^{13}C NMR, je dereplikacija posameznih metabolitov samo z ^1H NMR izjemno zahtevna in temelji na primerjavi dobljenih spektrov s spektri iz različnih podatkovnih baz, kar je dolgotrajno in pogosto neuspešno (17). Zaradi tega se v zadnjem času uveljavljajo metode, osnovane na ^{13}C NMR. Prvi tak pristop je CAMEL (*Caractérisation de mélanges*), ki temelji na spektrih ^{13}C NMR posameznih frakcij izvlečkov, dodatno podkrepjenih s podatki 2D NMR. Glede na kolinearnost posameznih signalov v ^{13}C NMR-spektrih se oblikujejo virtualni spektri, ki jih primerjamo s spektralno knjižnico znanih spojin, kar omogoča zgodnejše prepoznavanje že znanih spojin (23).

Drug pristop, temelječ na ^{13}C NMR, ki omogoča dereplikacijo spojin, je MixONat. Razvili so ga leta 2020 in temelji na klasičnih ^{13}C -spektrih, podprtih s poskusi DEPT-135 in DEPT-90. Eksperimentalno določene spektre nato primerjamo s podatkovno bazo in glede na ujemanja med spektri program nato predlaga možne strukture spojin. Metoda omogoča hitro prepoznavo znanih spojin v kompleksnih zmesih, s čimer se izognemo dolgotrajni izolaciji. V redkih primerih omogoča tudi razločevanje med stereoizomeri posameznih spojin (24).

4.3 METODE NA OSNOVI MASNE SPEKTROMETRIJE

Tekočinska kromatografija, sklopljena z masno spektrometrijo (LC-MS), je izjemno koristna pri izolaciji in identifikaciji novih spojin v kompleksnih vzorcih. Poleg določanja molekulske mase in kemijske sestave spojine lahko podatke o fragmentaciji uporabimo tudi za proučevanje struktur molekul. Kot rezultat takšne analize pridobimo kompleksne kromatograme s številnimi vrhovi in pripadajočimi masnimi spektri. Za njihovo analizo zato potrebujemo še dodatna orodja, kot je Mzmine, ki med drugim omogoča odstranjevanje šuma, iskanje in dekonvolucijo prekrivajočih se signalov, odstranjevanje izotopskih signalov ter določene statistične operacije (25). Masne spektre spojin lahko nato primerjamo z že znanimi spektri v različnih zbirkah, kot so LOTUS, COCONUT, "Dictionary of Natural Products" ali lastnimi podatkovnimi bazami (17). Na fragmentacijo spojin pomembno vplivajo tudi nastavitve masnega spektrometra

in ionski izvor. Posledično ima lahko ista spojina različne masne spektre, kar oteži primerjavo med spektri, pridobljenimi na različnih spektrometrih. Zaradi tega so se razvile številne metode *in silico* za proučevanje fragmentacije spojin, kot so MS-finder, CFM-ID in CSI:FingerID (26).

5 MOLEKULSKO MREŽENJE

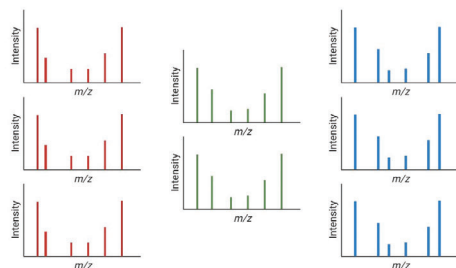
Poleg omenjenih pristopov se je v zadnjih letih kot izjemno uporabno izkazalo tudi molekulsko mreženje, osnovano na MS, ki so ga leta 2012 razvili Watrous in sod. ter ga uporabili za analizo metabolitov živih bakterijskih kolonij. Molekulsko mreženje omogoča učinkovito organizacijo in upodobitev velike količine podatkov, pridobljenih z MS, in predstavlja enega izmed vodilnih orodij za proučevanje spojin v kompleksnih vzorcih (27).

Molekulsko mreženje temelji na spoznanju, da strukturno podobne spojine fragmentirajo na podobne načine in tvorijo podobne masne spektre. Podobnost nato ovrednotimo, pri čemer najpogosteje uporabljamo kosinusno podobnost, ki upošteva vzorec signalov m/z in njihovo intenziteto. Poznamo tudi novejša metode, kot sta Spec2Vec in MS2DeepScore, ki temeljita na metodah umetne inteligence (28, 29).

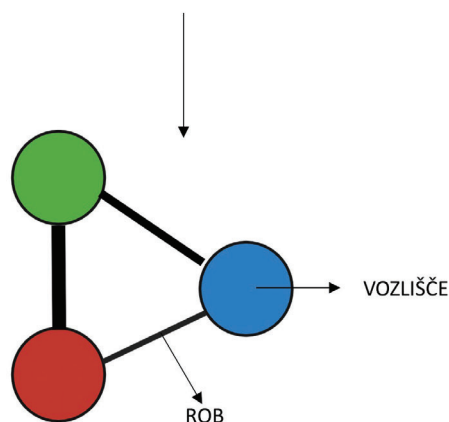
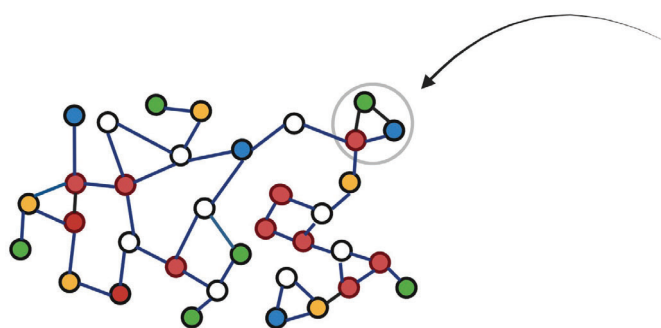
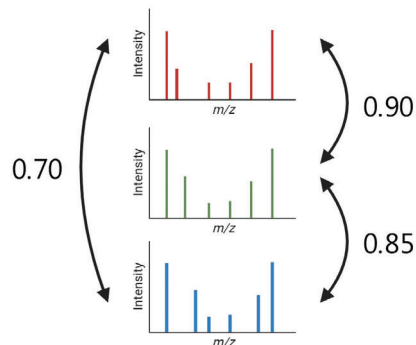
Klasično molekulsko mreženje (angl. *classical molecular networking*) poteka v dveh fazah. Na začetku vzorce spojin analiziramo z LC-MS, pri čemer moramo za vse spojine pridobiti fragmentacijske spektre. Pridobljene podatke nato naložimo na platformo "Global Natural Products Social Molecular Networking" (GNPS), kjer poteka izdelava molekulskih omrežij (30). Kot je prikazano na sliki 1, program sprva združi identične spektre v konsenzne spektre, s čimer se izogne, da se isti spektri pojavijo večkrat. S pomočjo prej omenjenih metod nato ovrednoti medsebojno podobnost konsenznih spektrov in ustvari molekulsko omrežje. Vsak konsenzni spekter je v omrežju predstavljen z vozliščem, ki s podobnimi vozlišči tvori povezave, imenovane robove, katerih debelina je odvisna od njihove podobnosti. Več podobnih vozlišč se tako združuje v gruče, ki nato tvorijo še večja omrežja (31, 32).

Kadar želimo hitro obdelati veliko količino masnih spektrov, je klasično molekulsko mreženje izjemno uporabno, a metoda ni sposobna razlikovati med položajnimi izomeri in stereoizomeri. Zaradi tega so Nothias in sod. razvili na značilkah osnovano molekulsko mreženje (angl. *feature*

1. Združevanje eksperimentalnih spektrov v konsenzne



2. Vrednotenje podobnosti med spektri



3. Izdelava molekulskega omrežja

Slika 1: Prikaz izdelave molekulskega omrežja. Iz eksperimentalno pridobljenih spektrov tvorimo konsenzne spektre in glede na njihovo medsebojno podobnost tvorimo molekulska omrežja. Konsenzni spektri so v molekulskem omrežju predstavljeni z vozlišči, povezanimi z robovi. Slika je bila ustvarjena s programom Biorender.

Figure 1: Construction of molecular networks. Consensus spectra are generated from experimentally acquired spectra, and based on their mutual similarity, molecular networks are generated. Consensus spectra are represented as nodes connected by edges. Created in Biorender.

based molecular networking, FBMN), ki poleg podatkov o fragmentaciji spojin upošteva tudi njihove retencijske čase. Pristop na značilkah osnovanega molekulskega mreženja je kompleksnejši in potrebuje dodatno obdelavo podatkov, a omogoča razlikovanje med izomeri, ki imajo različen retencijski čas (25, 33). Klasično molekulsko mreženje in na značilkah osnovano molekulsko mreženje sta najbolj uveljavljeni metodi molekulskega mreženja. Med drugim so ju uporabili za izolacijo in identifikacijo protiglivne učinkovine tanamicin iz bakterije *Pseudomonas aeruginosa* ter za proučevanje bromopirolnih alkaloidov v spužvi *Agelas dispar* (27, 34).

Ob fragmentaciji lahko ista spojina tvori adukte z različnimi ioni, kot je recimo $[M + H]^+$ ali $[M + Na]^+$. Posledično ima

lahko ista spojina več spektrov, kar oteži interpretacijo molekulskih omrežij. Zaradi tega so Schmid in sod. razvili na ionski identiteti osnovano molekulsko mreženje (angl. *ion identity molecular networking*, IIMN), ki upošteva vpliv različnih ionskih aduktov spojin, s čimer bistveno poenostavijo zgrajena omrežja (35). Ta metoda je precej podobna na značilkah osnovanemu molekulskemu mreženju – masne spektre prav tako razvrsti še po retencijskem času in pri spektrih, ki se pojavljajo pri podobnih ali enakih časih, z algoritmi preveri, kateri od njih bi lahko pripadali isti spojini. Metodo so med drugim uporabili za identifikacijo brevipolidov in metoksiliranih flavonoidov iz *Hyptis brevipes* in *Artemisia* sp., ki zavirajo proonkogeno signaliziranje v humanih celicah melanoma (36).

Za iskanje bioaktivnih spojin je še posebej pomembno na bioaktivnosti osnovano molekulsko mreženje (angl. *bioactivity based molecular networking*, BBMN), pri katerem podatke o fragmentacijskih spektrih združimo s podatki o bioaktivnosti. Metoda bistveno olajša identifikacijo bioaktivnih spojin v kompleksnih zmesih in pospeši njihovo izolacijo (37). Na bioaktivnosti osnovano molekulsko mreženje so med drugim uspešno uporabili za identifikacijo in izolacijo bioaktivnih spojin proti virusu SARS-CoV-2 ter virusu Chikungunya (37, 38).

Poznamo še številne druge metode molekulskega mreženja, ki omogočajo tvorjenje molekulskih omrežij glede na strukturne in substrukturne lastnosti molekul, biosintezni izvor ali kemijski razred spojin. Slednje so še posebej uporabne, kadar želimo proučevati različne biosintezne poti ali kadar ne moremo identificirati spojine, a lahko z njimi napovemo, v kateri strukturni razred spadajo (32).

6 SKLEP

Spojine naravnega izvora so skozi zgodovino predstavljale izjemen vir novih spojin vodnic in terapevtskih učinkovin. Kljub temu je v zadnjih letih zanimanje farmacevtske industrije za njihovo proučevanje bistveno upadlo. Tradicionalne strategije odkrivanja naravnih spojin so namreč dolgotrajne, delovno intenzivne in pogosto vodijo v izolacijo že znanih spojin. Nov zagon tega področja so prinesle številne nove metode dereplikacije, ki združujejo napredne analitske metode z računalniškimi orodji, katerih cilj je zgodnja identifikacija bioaktivnih spojin. Med najpomembnejšimi so različne metode *in silico* ter biokemometrične metode, osnovane na podatkih MS in NMR, ki omogočajo učinkovito analizo velikega števila kompleksnih podatkov ter povezavo spektralnih podatkov z bioaktivnostjo. Med njimi izstopa molekulsko mreženje, ki na podlagi podobnosti masnih spektrov spojin omogoča učinkovito vizualizacijo in identifikacijo spojin v kompleksnih zmesih. Nove metode so bistveno hitrejše in omogočajo natančnejšo določitev strukture in usmerjeno izolacijo bioaktivnih molekul. Še posebej koristne so pri proučevanju struktur kompleksnih molekul, s čimer odpirajo nove možnosti za njihovo uporabo. Menimo, da bodo za uspešnejše in hitrejšo odkrivanje novih učinkovin v prihodnje ključni nadaljnji razvoj visokozmogljivih tehnologij, boljša povezljivost podatkovnih baz in napredek računalniških

metod. Takšni integrirani pristopi omogočajo bolj usmerjeno iskanje naravnih bioaktivnih spojin ter pomembno podpirajo razvoj sodobne farmacevtske industrije.

7 LITERATURA

1. Balick MJ, Cox PA. *Plants, People, and Culture. Plants, People, and Culture: The science of ethnobotany*. 2nd ed. New York (NY): Garland Science; 2020.
2. Dias DA, Urban S, Roessner U. A Historical overview of natural products in drug discovery. *Metabolites*. 2012 Apr 16;2(2):303–36.
3. Tu Y. Artemisinin — A gift from traditional Chinese medicine to the World (Nobel Lecture). *Angw Chem Int Ed Engl*. 2016 Aug 22;55(35):10210–26.
4. Pham J V., Yilma MA, Feliz A, Majid MT, Maffettone N, Walker JR, et al. A review of the microbial production of bioactive natural products and biologics. *Front Microbiol*. 2019 Jun 20;10:1404.
5. Newman DJ, Cragg GM. natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod* 2020 Mar 27;83(3):770–803.
6. Butler MS, Capon RJ, Blaskovich MAT, Henderson IR. Natural product-derived compounds in clinical trials and drug approvals. *Nat. Prod. Rep.* 2025; (Advance Article) DOI:10.1039/D5NP00031A.
7. Cummins LL, Chen S, Blyn LB, Sannes-Lowery KA, Drader JJ, Griffey RH, et al. Multitarget affinity/specificity screening of natural products: Finding and characterizing high-affinity ligands from complex mixtures by using high-performance mass spectrometry. *J Nat Prod*. 2003 Sep;66(9):1186–90.
8. Feher M, Schmidt JM. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci*. 2003 Jan-Feb;43(1):218–27.
9. Stratton CF, Newman DJ, Tan DS. Cheminformatic comparison of approved drugs from natural product versus synthetic origins. *Bioorg Med Chem Lett*. 2015 Nov 1;25(21):4802–7.
10. Wu K, Karapetyan E, Schloss J, Vadgama J, Wu Y. *Advancements in small molecule drug design: A structural perspective*. *Drug Discov Today*. 2023 Oct;28(10):103730.
11. Capecchi A, Reymond JL. Classifying natural products from plants, fungi or bacteria using the COCONUT database and machine learning. *J Cheminform*. 2021 Oct 18;13(1):82.
12. Hart CE, Gadiya Y, Kind T, Krettl CA, Gaetz M, Misra BB, et al. Defining the limits of plant chemical space: challenges and estimations. *GigaScience*. 2025 Jan 6;14:giaf033. DOI: 10.1093/gigascience/giaf033
13. Boldini D, Ballabio D, Consonni V, Todeschini R, Grisoni F, Sieber SA. Effectiveness of molecular fingerprints for exploring the chemical space of natural products. *J Cheminform*. 2024 Mar 25;16(1):35.
14. Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, et al. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov*. 2021 Mar;20(3):200–216.
15. Burton G, Evans-Illidge EA. Emerging R and D law: The Nagoya protocol and its implications for researchers. *ACS Chem Biol*. 2014 Mar 21;9(3):588–91.



16. Henrich CJ, Beutler JA. Matching the power of high throughput screening to the chemical diversity of natural products. *Nat Prod Rep.* 2013;30(10).
17. Meunier M, Schinkovitz A, Derbré S. Current and emerging tools and strategies for the identification of bioactive natural products in complex mixtures. *Nat Prod Rep.* 2024 Nov 13;41(11):1766–1786
18. Kellogg JJ, Todd DA, Egan JM, Raja HA, Oberlies NH, Kvalheim OM, et al. Biochemometrics for natural products research: Comparison of data analysis approaches and application to Identification of bioactive compounds. *J Nat Prod.* 2016 Feb 26;79(2):376–86.
19. Grienke U, Foster PA, Zwirchmayr J, Tahir A, Rollinger JM, Mikros E. 1H NMR-MS-based heterocovariance as a drug discovery tool for fishing bioactive compounds out of a complex mixture of structural analogues. *Sci Rep.* 2019 Jul 31;9(1):11113.
20. Franco BA, Luciano ER, Sarotti AM, Zanardi MM. DP4+App: Finding the best balance between computational cost and predictive capacity in the structure elucidation process by DP4+. *factors analysis and automation.* *J Nat Prod.* 2023;86(10).
21. Reher R, Kim HW, Zhang C, Mao HH, Wang M, Nothias LF, et al. A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. *J Am Chem Soc.* 2020 Mar 4;142(9):4114–4120.
22. Grimblat N, Zanardi MM, Sarotti AM. Beyond DP4: An improved probability for the stereochemical assignment of isomeric compounds using quantum chemical calculations of NMR Shifts. *J Org Chem.* 2015 Dec 18;80(24):12526–34.
23. Bakiri A, Plainchont B, de Paulo Emerenciano V, Reynaud R, Hubert J, Renault JH, et al. computer-aided dereplication and structure elucidation of natural products at the university of Reims. *Mol Inform.* 2017 Oct;36(10).
24. Bruguère A, Derbré S, Dietsch J, Leguy J, Rahier V, Pottier Q, et al. MixONat, a software for the dereplication of mixtures based on ¹³C NMR spectroscopy. *Anal Chem.* 2020 Jul 7;92(13):8793–8801.
25. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics.* 2010 Jul 23;11:395.
26. Blaženović I, Kind T, Torbašinović H, Obrenović S, Mehta SS, Tsugawa H, et al. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: Database boosting is needed to achieve 93% accuracy. *J Cheminform.* 2017 May 25;9(1):32.
27. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, et al. Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A.* 2012 Jun 26;109(26):E1743–52.
28. Huber F, van der Burg S, van der Hooft JJJ, Ridder L. MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra. *J Cheminform.* 2021 Oct 29;13(1):84.
29. Huber F, Ridder L, Verhoeven S, Spaaks JH, Diblen F, Rogers S, et al. Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships. *PLoS Comput Biol.* 2021 Feb 16;17(2):e1008724
30. Wang M, Carver JJ, Phelan V V., Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 2016 Aug 9;34(8):828–837.
31. Aron AT, Gentry EC, McPhail KL, Nothias LF, Nothias-Esposito M, Bouslimani A, et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat Protoc.* 2020 Jun;15(6):1954–1991.
32. Qin GF, Zhang X, Zhu F, Huo ZQ, Yao QQ, Feng Q, et al. MS/MS-based molecular networking: an efficient approach for natural products dereplication. *Molecules.* 2022 Dec 24;28(1):157.
33. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods.* 2020 Sep;17(9):905–908.
34. Freire VF, Gubiani JR, Spencer TM, Hajdu E, Ferreira AG, Ferreira DAS, et al. Feature-based molecular networking discovery of bromopyrrole alkaloids from the marine sponge *agelas dispar*. *J Nat Prod.* 2022 May 27;85(5):1340–1350.
35. Schmid R, Petras D, Nothias LF, Wang M, Aron AT, Jagels A, et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nat Commun.* 2021 Jun 22;12(1):3832.
36. Hell T, Rutz A, Dürr L, Dobrzyński M, Reinhardt JK, Lehner T, et al. Combining activity profiling with advanced annotation to accelerate the discovery of natural products targeting oncogenic signaling in melanoma. *J Nat Prod.* 2022 Jun 24;85(6):1540–1554.
37. Nothias LF, Nothias-Esposito M, Da Silva R, Wang M, Protsyuk I, Zhang Z, et al. Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation. *J Nat Prod.* 2018 Apr 27;81(4):758–767.
38. Wasilewicz A, Bojkova D, Beniddir MA, Cinatl J, Rabenau HF, Grienke U, et al. Molecular networking unveils anti-SARS-CoV-2 constituents from traditionally used remedies. *J Ethnopharmacol.* 2024 Jan 30;319(Pt 2):117206.