

AAALifeSci Course Companion Manual

*Zaawansowana chemia analityczna w naukach
biomedyczo-farmaceutycznych*

Redakcja
Pedro Domingues, Antonia García, Elżbieta Skrzydlewska



Zaawansowana chemia analityczna w naukach biomedyczo-farmaceutycznych

Redakcja

Pedro Domingues, Antonia Garcia, Elżbieta Skrzydlewska

Tytuł oryginału

Advanced Analytical Chemistry for Life Sciences

AACLifeSci

AACLifeSci Course companion manual

Projekt współfinansowany przez Komisję Unii Europejskiej.
Niniejsza publikacja odzwierciedla jedynie poglądy autorów, a Komisja nie ponosi odpowiedzialności za jakiegokolwiek wykorzystanie informacji zawartych w tej publikacji.

ISBN 978-83-951534-9-5

Niniejszy manual udostępniony jest na licencji Creative Commons Attribution-NonCommercial 3.0 License CC BY-NC 3.0. Redaktorami manuala są Pedro Domingues, Antonia García i Elżbieta Skrzydlewska.

SPIS TREŚCI

Wprowadzenie	7
I. Wstęp	7
II. Założenia kursu i efekty kształcenia	8
III. Forma i sposoby realizacji treści programowych przedmiotu	9
IV. Literatura dodatkowa	14
Moduł 1 – Zastosowanie metod łączących techniki separacyjne ze spektrometrią mas w naukach biomedycy i farmacji	15
I. Wprowadzenie	15
II. Założenia i efekty kształcenia	15
III. Spis treści	16
1. Spektrometria mas	17
1.1. Wprowadzenie	17
1.2. Metody jonizacji próbek biologicznych w spektrometrii mas	17
1.3. Typy analizatorów stosowanych w spektrometrii mas	20
1.4. Tandemowa spektrometria mas	25
1.5. Interpretacja widm masowych	26
2. Techniki separacyjne sprzężone ze spektrometrią mas	28
2.1. Chromatografia ciekłowa – spektrometria mas (LC-MS)	28
2.1.1. Wprowadzenie do LC-MS	29
2.1.2. Teoretyczne podstawy chromatografii ciekłowej	29
2.1.2.1. Proces rozdzielania chromatograficznego	30
2.1.2.2. Chromatografia w normalnym układzie faz (NP)	33
2.1.2.3. Chromatografia w odwróconym układzie faz (RP)	33
2.1.2.4. Chromatografia oddziaływań hydrofilowych (HILIC)	34
2.1.2.5. Chromatografia jonowymienna	34
2.1.3. Moduły systemu LC-MS	34
2.2. Chromatografia gazowa – spektrometria mas (GC-MS)	37
2.2.1. Wprowadzenie do GC-MS	37
2.2.2. Podstawy teoretyczne GC-MS	37
2.2.2.1. Proces chromatograficzny	37
2.2.2.2. Kolumny chromatograficzne	38
2.2.2.3. GC×GC/MS	39
2.2.2.4. Przygotowanie próbek do analizy chromatograficznej.	42
Derywatywacja.	
2.2.3. Połączenie GC-MS	46
2.3. Elektroforeza kapilarna – spektrometria mas (CE-MS)	47
2.3.1. Wstęp do CE-MS	47
2.3.2. Teoretyczne zasady techniki CE-MS	48
2.3.2.1. Kapilarna elektroforeza strefowa	51

2.3.2.2. Micelarna chromatografia elektrokinetyczna	51
2.3.2.3. Elektrochromatografia kapilarna	51
2.3.2.4. Kapilarna elektroforeza żelowa	52
2.3.2.5. Aparatura	52
3. Część praktyczna	54
3.1. Rejestracja widm masowych ESI-MS	54
3.2. Rejestracja widm masowych ESI-MS/MS	56
3.3. Interpretacja widm masowych	58
3.4. Wykorzystanie metody GC-MS w trybie pracy monitorowania pojedynczych jonów	58
3.5. Wykonanie analiz przy użyciu metody HPLC-MS oraz konfiguracja parametrów dla trybu data dependent MS/MS	65
Moduł 2 – Metabolomika	71
I. Wstęp do metabolomiki	71
II. Założenia i efekty kształcenia	71
III. Spis treści	72
1. Wstęp do metabolomiki	73
2. Podejścia analityczne w metabolomice	74
2.1. Cykl pracy w badaniach metabolomicznych	75
2.1.1. Projekt eksperymentu	76
2.1.2. Pobieranie próbek	76
2.1.3. Ekstrakcja metabolitów	76
2.1.4. Rozdział i detekcja	77
2.2. Kontrola jakości i procedura zapewnienia jakości w metabolomice	78
3. Przetwarzanie danych i identyfikacja metabolitów	79
3.1. Przetwarzanie danych koncepcja pipeline	80
3.1.1. Przetwarzanie surowych danych	80
3.1.2. Wstępne przetwarzanie danych	80
3.1.3. Wstępna obróbka danych	86
3.2. Identyfikacja metabolitów	87
3.2.1. Identyfikacja za pomocą GC-EI-MS	87
3.2.2. Identyfikacja metabolitów za pomocą LC lub CE-(ESI)MS: Praca z bazami danych	89
3.2.3. Bazy danych metabolitów, klasyfikacja w oparciu o wartość masy cząsteczkowej	90
3.2.4. Bazy danych metabolitów, klasyfikacja w oparciu o widma masowe	90
3.2.5. Informatyczne bazy danych	92
3.3. Analiza statystyczna	92
3.3.1. PCA	93
3.3.2. PLS-DA	94
3.3.3. OPLS-DA	94
3.3.4. Wykres-S, metoda Jack-knife i test projekcyjny VIP	94
3.3.5. Walidacja statystycznych modeli wielowymiarowych	95
4. Analiza danych	96

4.1. Od identyfikacji danych do szlaków biochemicznych: klasyfikacja danych w celu oceny istotności biologicznych.	96
4.1.1. Bazy danych szlaków biochemicznych	96
4.1.2. Analiza szlaków biochemicznych	97
4.1.3. Wzbogacanie metabolitów	97
4.1.4. Topologiczna analiza szlaków biologicznych	98
4.2. Walidacja biomarkerów: ustalenie istotności metabolitów	99
5. Zajęcia praktyczne z metabolomiki.	99
5.1. Metabolomika targetowa i nietargetowa	99
5.2. Wstępna obróbka danych przy użyciu platformy Workflow for Metabolomics (W4M)	100
5.2.1. Konwersja plików MS	100
5.2.2. Kontrola surowych danych	101
5.2.3. Prześlij dane do Galaxy	103
5.2.4. Wstępne przetwarzanie danych LC-MS za pomocą XCMS.	107
5.3. Statystyka	123
5.4. Komentarze	124
Moduł 3 – Lipidomika	126
I. Wprowadzenie	126
II. Założenia i efekty kształcenia	126
III. Spis treści	127
1. Wstęp do lipidomiki	127
1.1. Podejścia analityczne stosowane w lipidomice	130
1.1.1. Przygotowanie próbek: ekstrakcja i rozdział lipidów	130
1.1.2. Identyfikacja lipidów oparta na analizach MS: koncepcje	132
1.1.2.1. Nietargetowe podejścia lipidomiczne oparte na LC-MS	135
1.1.2.2. Targetowe podejścia lipidomiczne “Shotgun”	136
1.2. Ilościowa analiza lipidomiczna	137
1.3. Przetwarzanie danych w celu identyfikacji/ilościowej oceny lipidów	137
1.4. Część praktyczna	139
1.4.1. Analiza lipidomiczna profilu fosfolipidów monocytów linii THP1 z wykorzystaniem HILIC-LC-MS	139
1.5. Dodatkowe źródła wiedzy	142
Module 4 - Proteomika	143
I. Wstęp	143
II. Założenia i cele przedmiotu	143
III. Spis treści	144
1. Wstęp do proteomiki	145
2. Podejścia analityczne stosowane w proteomice	146
2.1. Przygotowanie próbek	146
2.2. Identyfikacja białek oparta na analizach MS: koncepcje	147
2.3. Metody spektrometrii mas stosowane do identyfikacji białek	148
2.4. Ilościowa analiza proteomiczna	149
2.5. Identyfikacja PTM	152

2.6. Identyfikacja wewnątrzkomórkowej lokalizacji białek	154
2.7. Identyfikacja interakcji między białkami	154
3. Przetwarzanie danych w celu identyfikacji białek	155
3.1. Bazy danych zawierające sekwencje białkowe	155
3.2. Przeszukiwanie białkowych baz danych	156
3.3. Analiza danych	156
4. Część praktyczna	157
4.1. Identyfikacja białek w materiale komórkowym z użyciem podejścia PMF i PFF za pomocą platformy MASCOT	157
4.2. Identyfikacja białek z eksperymentu LC-MS	162
4.2.1. Zamiana otrzymanych danych na pliki mzML z użyciem MSConverter (ProteoWizard)	163
4.2.2. Utworzenie bazy danych FASTA używając serwera Uniprot (SwissProt)	165
4.2.3. Wyszukiwarka: SearchGUI	166
4.2.4. Generowanie i ocena wyników: użycie PeptideShaker do wizualizacji peptydów i białek oraz do walidacji wyników. Analiza PTM	169
4.2.5. Analiza danych: informacja o białkach, analiza szlaków i ontologia genów	169

Wprowadzenie

Pedro Domingues¹, Antonia Garcia² Elzbieta Skrzydlewska³

¹*Mass Spectrometry Center, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal, QOPNA (Química Orgânica de Produtos Naturais e Agroalimentares)*

²*Centre for Metabolomics and Bioanalysis (CEMBIO), Facultad de Farmacia, Universidad San Pablo CEU; Boadilla del Monte, 28668 Madrid, Spain*

³*Department of Analytical Chemistry, Medical University of Białystok, 15-089 Białystok, Poland*

I. Wstęp

Obserwowany w obecnej dekadzie wykładniczy wzrost badań genomicznych sprzyja szybkiemu rozwojowi pokrewnych badań omicznych, ale opartych o spektrometrię mas. W związku z tym widoczny jest w ostatnich latach szybki rozwój badań w dziedzinie proteomiki. Jednocześnie do badań naukowych wprowadzana jest również lipidomika i metabolomika. To zróżnicowane tempo rozwoju nauk omicznych determinują różne czynniki, takie jak niewystarczający rozwój narzędzi bioinformatycznych, jak również brak wykwalifikowanych specjalistów w tych dziedzinach. Dlatego opracowano kurs AACLifeSci, którego celem jest zapewnienie zaawansowanych szkoleń dla doktorantów z nauk biomedycznych/farmaceutycznych w zakresie spektrometrii mas i łączonych technik separacyjnych stosowanych w metabolomice, lipidomice i proteomice.

Ostatnie dziesięciolecie to jednocześnie szybki rozwój chorób cywilizacyjnych, co jest uważane za główną przyczynę zwiększonej zachorowalności i śmiertelności w Europie. W tym kontekście badania jakościowe i ilościowe składników organizmu człowieka mające na celu zrozumienie metabolizmu, a także identyfikację biomarkerów chorobowych oraz proponowanie nowych podejść terapeutycznych są postrzegane, jako jeden z możliwych środków zaradczych. W związku z tym niezbędnymi narzędziami wykorzystywanymi do wyjaśnienia biochemicznych podstaw funkcjonowania organizmu poprzez charakterystykę i kwantyfikację puli cząsteczek biologicznych są: metabolomika, lipidomika i proteomika. Ponieważ pojedyncze związki uważa się obecnie za mało specyficzne biomarkery, podejście omiczne daje nowe możliwości bardziej szczegółowego wglądu w mechanizmy związane z rozwojem chorób. Takie podejście poprawia efektywność diagnostyczną, ale również ułatwia monitorowanie farmakoterapii, a także pozwala na personalizowanie terapii.

Nauczanie podejścia omicznego na uniwersytetach europejskich jest zróżnicowane. Podejście omiczne jest na ogół przekazywane studentom przez jednostki, które zajmują się zaawansowaną chemią analityczną [AAC] i zapewniają studentom elementy analizy instrumentalnej niezbędnej do badań biomedycznych. Jednak liczba wyspecjalizowanych jednostek i zespołów dydaktycznych mających dostęp do sprzętu niezbędnego do kompleksowego podejścia omicznego jest ograniczona. Brakuje również pracowników naukowo-dydaktycznych posiadających kwalifikacje do nauczania AAC jak również materiałów dydaktycznych i skutecznego systemu wymiany doświadczeń na różnych poziomach. Tak więc tylko w niektórych przypadkach doktoranci mogą być odpowiednio przeszkoleni w zakresie AAC oraz nabyć umiejętności niezbędnych do integracji AAC z potrzebami służby zdrowia, w tym szpitalnych laboratoriów diagnostycznych.

Wykorzystanie zaawansowanej chemii analitycznej w naukach badaniach biomedycznych i farmaceutycznych (ACCLifeSci) stało się podstawą projektu edukacyjno-naukowego koordynowanego przez profesor Elżbietę Skrzydlewską, dziekana Wydziału Farmaceutycznego MUB z Oddziałem Medycyny Laboratoryjnej i finansowanego w ramach programu Erasmus + i realizowanego dzięki partnerskiej współpracy trzech uniwersytetów:

- Uniwersytet Medyczny w Białymstoku, Polska, Wydział Farmaceutyczny z Oddziałem Medycyny Laboratoryjnej, Zakład Chemii Analitycznej (prof. Elżbieta Skrzydlewska)
- Uniwersytet w Aveiro, Portugalia, Wydział Chemii, Centrum Spektrometrii Masowej (prof. Pedro Domingues)
- Uniwersytet San Pablo-CEU, Hiszpania, (prof. Coral Barbás).

Projekt AACLifeSci został zrealizowany przy wsparciu Komisji Unii Europejskiej. Celem projektu było dostosowanie programów studiów doktoranckich trzech partnerskich uniwersytetów do potrzeb związanych z analizą biomolekularną poprzez wykorzystanie zaawansowanych technik analitycznych w zakresie omik poprzez wymianę doświadczeń na poziomie europejskim. Projekt ten polegał na przygotowaniu kadry akademickiej trzech partnerskich uniwersytetów do kształcenia doktorantów w stosowaniu technik AAC w metabolomice, lipidomice i proteomice wprowadzając modyfikacje w programach studiów i tworząc zasoby edukacyjne, w tym podręcznik i e-materiały dla doktorantów. W ramach projektu wykorzystano partnerskie działania edukacyjne /szkoleniowe/ i dydaktyczne, które obejmowały wymianę doświadczeń i dobrych praktyk, oraz tworzenie stabilnej i aktywnej sieci współpracy międzynarodowej, a także osobisty i społeczny rozwój uczestników.

Pełny kurs został podzielony na cztery moduły, co pozwala na uporządkowaną naukę w postaci wykładów i zajęć praktycznych. Pierwszy moduł kursu obejmuje główne aspekty biomolekularnej spektrometrii mas, w tym najnowsze osiągnięcia instrumentalne i ich zalety, a także interpretacji widm masowych w celu identyfikacji i potwierdzenia struktury związków. Przedstawione zostaną również podstawy wysokosprawnych technik separacyjnych sprzężonych ze spektrometrią mas, w tym mechanizmy rozdziału związków i oprzyrządowanie. Ten moduł zapewni podstawy analityczne, dzięki którym doktoranci będą mogli dalej rozwijać swoją wiedzę w dziedzinie nauk omicznych opartych o spektrometrię mas. Dalsze moduły omawiają podstawy badań metabolomicznych, lipidomicznych i proteomicznych. Moduły te przedstawiają metody analityczne, metody przetwarzania, analizy i procesowania danych dla każdej z omawianych metodologii. Zostaną również opisane niektóre z zasobów bioinformatycznych obecnie używanych, a które studenci będą mieli okazję przetestować na zajęciach praktycznych.

II. Założenia kursu i efekty kształcenia

Założenia

Celem kursu jest zapoznanie studentów z teoretycznymi i praktycznymi umiejętnościami z zakresu chemii, biochemii i bioinformatyki w odniesieniu do technik separacyjnych, w tym chromatograficznych i spektrometrii mas dedykowanych metabolomice, lipidomice i proteomice. Przedstawione będą teoretyczne podstawy metod separacyjnych i spektrometrii mas oraz szczegółowo zostaną omówione dziedziny omiczne. Studenci zapoznają się z różnymi aspektami omik i sposobami ich zastosowania w badaniach naukowych. Wyjaśnione zostanie praktyczne zastosowanie tych metod w nauce oraz ich zalety i ograniczenia. Praktyczne umiejętności będą głównie koncentrować się na wykorzystaniu ogólnodostępnych

baz danych i darmowych narzędzi informatycznych do przetwarzania i analizy otrzymanych danych omicznych.

Na zakończenie tego kursu studenci powinni umieć analizować naukowe problemy związane z erą post genomową i wiedzieć, jak zdobyć i jak przeprowadzić prostą analizę danych omicznych uzyskanych za pomocą spektrometrii masowej. Zrozumienie zasad, metod i zastosowań najnowszych technik omicznych w znaczący sposób pozwoli studentowi odnaleźć się w pracy zarówno w nauce jak i przemyśle.

Specyficzne efekty kształcenia:

Po ukończeniu kursu studenci będą:

1. dokonywać krytycznego przeglądu dostępnych rodzajów analizatorów masy i metod jonizacji z uwzględnieniem ich wad i zalet, w tym krytycznej analizy tandemowej spektrometrii mas;
2. rozpoznawać sposoby fragmentacji cząsteczek organicznych, głównie drobno-cząsteczkowych metabolitów peptydów i lipidów;
3. omawiać kompleksowo różne mechanizmy separacji sprzężonej z MS dla metabolomiki, lipidomiki i proteomiki;
4. znać i rozumieć różne metody przygotowania próbek i ich ograniczenia dla każdej z omicznych aplikacji;
5. identyfikować wszystkie etapy omicznych badań i ich zróżnicowane podejścia stosowane w metabolomice, lipidomice i proteomice;
6. omawiać wykorzystanie powszechnie dostępnego oprogramowania do ponownego przetwarzania danych i analizy szlaków metabolicznych;
7. posiadać niezbędne umiejętności w zakresie korzystania z zasobów ogólnodostępnych baz danych;
8. wyjaśniać nie specjalistom, jakich informacji w różnych dziedzinach Life Science mogą dostarczyć badania omiczne.

III. Forma i sposoby realizacji treści programowych przedmiotu

Kurs składa się z czterech niezależnych, ale ściśle powiązanych modułów:

Moduł 1 - Zastosowanie metod wykorzystujących połączenie technik separacyjnych i spektrometrii mas w naukach biomedyczo-farmaceutycznych

Moduł 2 - Metabolomika;

Moduł 3 - Lipidomika;

Moduł 4 - Proteomika.

Każdy z tych modułów analizuje główne teoretyczne i techniczne aspekty, które są niezbędne do lepszego zrozumienia podejścia omicznego w celu rozwiązania problemów nauk przyrodniczych. Wiedza i umiejętności, które będą nabyte podczas tego kursu zostały uznane za odpowiednie dla poziomu wiedzy na studiach doktoranckich.

Kurs składa się z różnych tematycznie zajęć praktycznych w ramach których studenci mogą zdobyć praktyczną wiedzę i umiejętności aplikacyjne niezbędne w każdym module. Każdy moduł jest niezależny. Jednak znajomość zagadnień związanych z modułem 1 - Techniki separacyjne i spektrometria mas w naukach biomedyczo-farmaceutycznych jest niezbędna do zrozumienia następnych modułów.

Program studiów doktoranckich na Uniwersytecie Medycznym w Białymstoku, Wydział Farmaceutyczny z OML, po wprowadzeniu założeń projektu AACLifeSci

Wydział Farmaceutyczny z Oddziałem Medycyny Laboratoryjnej Medycznej Uniwersytetu w Białymstoku oferuje studia doktoranckie w dziedzinie nauk medycznych, dyscyplina biologia medyczna i w dziedzinie nauk farmaceutycznych.

Program studiów doktoranckich w dziedzinie w naukach medycznych trwa cztery lata. Drugi rok obejmuje kurs „Nowoczesne techniki analityczne w naukach biomedycznych”, który został przedłużony z 15 do 30 godzin w oparciu o założenia projektu AACLifeSci. Kurs odpowiada 3 punktom ECTS.

Celem tego kursu jest dostarczenie studentom strategii i metod nowoczesnej i specjalistycznej analizy biochemicznej wykorzystywanej głównie w laboratoriach badawczych.

Po zakończeniu tego kursu studenci powinni:

- mieć wiedzę na temat technik przygotowania próbek stosowanych w metabolomice, lipidomice i proteomice;
- znać i umieć opisać współczesne techniki separacyjne, w tym techniki chromatograficzne i elektromigracyjne;
- znać metodykę badawczą stosowaną w analizie klasycznej i omicznej;
- wykazać się znajomością pojęć i praktycznych zastosowań biostatystycznej oceny wyników badań;
- umieć zaproponować technikę analityczną dla rozwiązania konkretnego problemu naukowego z dziedziny nauk biomedycznych;
- umieć przeprowadzić analizę metabolomiczną / lipidomiczną / proteomiczną;
- potrafić interpretować uzyskane wyniki za pomocą metod statystycznych;
- posiadać umiejętność wyciągania wniosków z badań w celu rozwiązania problemów.

UMB nauki biomedyczne	Całkowita liczba godz.	Spektrometria mas i techniki separacyjne	Metabolomika	Lipidomika	Proteomika
	Nowoczesne techniki analityczne w naukach biomedycznych				
Wykłady i workshopy	6h	2h	1h	1h	2h
Zajęcia praktyczne	24h	10h	4h	5h	5h
Praca własna studenta	57h	15h	10h	14h	18h
Całkowita praca studenta	87h	27h	15h	20h	25h

Program studiów doktoranckich w dziedzinie w naukach farmaceutycznych trwa również cztery lata. Drugi rok obejmuje kurs „Zaawansowane techniki analityczne w badaniach omicznych, który jest realizowany w ciągu 15 godzin w oparciu o zawartość AACLifeSci. Kurs odpowiada 1 punktowi ECTS. Jednak kurs ten jest poprzedzony kursem „Elementy nowoczesnej analizy farmaceutycznej”, dostarczającym znajomości technik separacyjnych i spektrometrii mas oraz analizy targetowej. Dlatego drugi kurs zawiera informacje tylko o

analizie omicznej. 30 godzin przypadających na ten kurs jest przeznaczonych na AACLifeSci, który odpowiada 4 punktom ECTS.

Celem tego kursu jest przedstawienie studentom strategii i metod analizy omicznej wykorzystywanych w laboratoriach badawczych.

Po zakończeniu tego kursu studenci powinni:

- znać metody przygotowania próbek biologicznych do analizy omicznej;
- znać metodologię badań stosowaną w analizie klasycznej i omicznej;
- wykazać się znajomością pojęć i praktycznych zastosowań biostatystycznej oceny wyników badań;
- umieć przeprowadzić analizę metabolomiczną / lipidomiczną / proteomiczną;
- potrafić interpretować uzyskane wyniki za pomocą metod statystycznych;
- posiadać umiejętność wnioskowania i wyciągania wniosków z badań w celu rozwiązania problemów.

UMB nauki farmaceutyczne	Całkowita liczba godz.	Spektrometria mas i techniki separacyjne	Metabolomika	Lipidomika	Proteomika
		Elementy współczesnej analityki farmaceutycznej	Zaawansowane techniki analityczne w badaniach omicznych		
Wykłady i workshopy	8h	5h	1h	1h	1h
Zajęcia praktyczne	22h	10h	4h	4h	4h
Praca własna studenta	30h	14h	3h	5h	6h
Całkowita praca studenta	58h	29h	8h	10h	11h

Program studiów doktoranckich Uniwersytetu CEU San Pablo w Madrycie po wprowadzeniu założeń projektu AACLifeSci

Wydział Farmaceutyczny Uniwersytetu San Pablo-CEU kształci magistrów „Odkrywanie leków”. Przyszły magister jest kształcony we współpracy trzech uniwersytetów: University Complutense, San Pablo-CEU i Alcalá University, które są zlokalizowane w prowincji Madryt. Wspomniane uniwersytety współpracują również w ramach studiów doktoranckich w zakresie chemii medycznej. W roku akademickim 2016/2017 rozpoczęto czwarty cykl tego typu studiów.

Program studiów magisterskich obejmuje jeden cykl akademicki z 60 punktami ECTS. W drugim semestrze realizowane są 30 godzinne zajęcia fakultatywne: Zaawansowane techniki bioanalityczne, które wycenione są na 3 punkty ECTS. Celem tego kursu jest dostarczenie studentom wiedzy i umiejętności w zakresie nowoczesnych technik bioanalizy, głównie opartych o spektrometrię mas sprzężoną z technikami separacyjnymi o wysokiej rozdzielczości. Ponadto studentom dostarczana jest wiedza na temat metod oczyszczania i wstępnego zateżnienia próbek. Zajęcia obejmują naukę teorii i praktyki poprzez pracę z rzeczywistymi próbkami w nowoczesnym laboratorium.

Po zakończeniu tego kursu studenci powinni:

- umieć wybrać i zastosować odpowiednie metody analityczne, biorąc pod uwagę podstawowe oprzyrządowanie, zakres i zastosowania metod stosowanych obecnie do charakterystyki jakościowej i analizy farmaceutyków oraz próbek biologicznych
- umieć wybrać i zastosować odpowiednią metodę analityczną do analizy aktywnych składników i ich metabolitów w próbkach biologicznych oraz analizy składników aktywnych których zawartość w końcowym produkcie farmaceutycznym jest bardzo mała;
- wskazać i scharakteryzować wszystkie etapy badania oraz ich różne podejścia stosowane w metabolomice.

CEU Odkrywanie leków	Całkowita liczba godzin	Spektrometria mas i techniki separacyjne	Chromatografia i EC w bioanalizie	Metabolomika	Oczyszczanie i zateżanie próbek
Zaawansowane techniki bioanalityczne					
Wykłady i workshopy	24h	11h	8h	3h	2h
Zajęcia praktyczne	6h	2h	2h		2h
Praca własna studenta	60h				
Całkowita praca studenta	90h				
ECTS	3				

Zajęcia z „Zawansowanych technik bioanalitycznych” będą prowadzone w oparciu o manual AACLifeSci w drugim semestrze roku akademickiego i obejmują 3 punkty ECTS.

Zajęcia zostaną zorganizowane w następujący sposób

CEU	Całkowita liczba godzin	(1) Spektrometria mas & (2) techniki separacyjne	Metabolomika	Lipidomika	Proteomika
Wykłady i workshopy	22h	(1) 6h (2) LC:2h; GC:1h; CE:1h	4h	4h	4h
Zajęcia praktyczne	8h	2h	2h	2h	2h
Praca własna studenta	60h	30h	10h	10h	10h
Całkowita praca studenta	90h	42h	16h	16h	16h
ECTS	3	1.4	0.53	0.53	0.53

Program studiów doktoranckich Uniwersytetu w Aveiro w zakresie biochemii, po wprowadzeniu założeń projektowych AACLifeSci

Program studiów doktoranckich Uniwersytetu Aveiro w zakresie biochemii jest realizowany w ciągu czterech lat. Pierwszy rok obejmuje siedem przedmiotów, z których cztery są opcjonalne. Dwa z tych przedmiotów (zaawansowane kursy z biochemii I i z biochemii II) będą realizowane w oparciu o manual AACLifeSci. Te przedmioty składają się z trzech opcjonalnych modułów, z których każdy odpowiada 3 punktom ECTS. Aby ukończyć studia każdy student musi zaliczyć dwa z tych modułów. Oceniony będzie każdy

moduł, a ocena końcowa zostanie wystawiona na podstawie średniej ważonej oceny uzyskanej w dwóch modułach. Kursy i odpowiednie moduły to:

Zaawansowany kurs Biochemii I:

Celem tego kursu jest przekazanie studentom zaawansowanych umiejętności w zakresie strategii metod nowoczesnej i specjalistycznej analizy biochemicznej wykorzystywanej w laboratoriach badawczych i klinicznych oraz w przemyśle.

Po zakończeniu tego kursu studenci powinni:

- rozumieć podstawy i znać zastosowania spektrometrii mas i jej połączenia z technikami chromatograficznymi;
- umieć krytycznie oceniać metody stosowane w badaniach lipidomicznych i metabolomicznych;
- umieć zaplanować eksperymenty lipidomiczne lub metabolomiczne;
- umieć analizować dane lipidomiczne lub metabolomiczne.

Uniwersytet w Aveiro	Całkowita liczba godz.	Proteomika	Glikomika	Bioinformatyka
Wykłady i workshopy	20h	10h	10h	10h
Zajęcia praktyczne		5h	5h	5h
Praca własna studenta	150h	75h	75h	75h
Całkowita praca studenta	180h	90h	90h	90h
ECTS	6	3	3	3

Zaawansowany kurs Biochemii II

Kurs ma na celu dostarczenie studentom zaawansowanych umiejętności w zakresie strategii i metod nowoczesnej i specjalistycznej analizy biochemicznej w zakresie proteomiki, glikomiki i bioinformatyki.

Po zakończeniu kursu student musi umieć:

- krytycznie ocenić, które metody należy zastosować w celu rozwiązania konkretnych problemów związanych z proteomiką i glikomiką;
- zaplanować eksperymenty z proteomiką i glikomiką;
- analizować dane uzyskane w badaniach proteomicznych i glikomicznych;
- wybrać i zastosować odpowiednie algorytmy i narzędzia w tych aplikacjach biochemicznych.

Uniwersytet w Aveiro	Całkowita liczba godz.	Proteomika	Glikomika	Bioinformatyka
Wykłady i workshopy	20h	10h	10h	10h
Zajęcia praktyczne	10h	5h	5h	5h
Praca własna studenta	150h	75h	75h	75h
Całkowita praca studenta	180h	90h	90h	90h
ECTS	6	3	3	3

AACLifeSci zostanie w pełni zintegrowane z czterema modułami programu studiów doktoranckich w zakresie biochemii Uniwersytetu w Aveiro, z których każdy odpowiada 3 punktom ECTS, co w sumie daje 12 punktów ECTS.

Moduł 1- Chromatografia i Spektrometria mas dla nauk przyrodniczych (3 ECTS)

Moduł 2- Metabolomika (3 ECTS)

Moduł 3- Lipidomika (3 ECTS)

Moduł 4- Proteomika (3 ECTS)

IV. Literatura dodatkowa

Moduł I

1. H.J. Hübschmann. Handbook of GC-MS: Fundamentals and Applications (3rd Edition). Somerset, NJ, USA: Wiley, 2015. ProQuest ebrary. Web. 8, 2015.
2. L.R. Snyder, J.J. Kirkland, J.W. Dolan. Introduction to modern liquid chromatography. Wiley, 2011.
3. E. De Hoffmann, V. Stroobant. Mass spectrometry: principles and applications. Wiley, 2007.
4. D.R. Baker. Capillary electrophoresis. Wiley, 1995.
5. https://www.agilent.com/cs/library/usermanuals/public/G7100-90001_CEMSAAnalysis_ebook.pdf

Moduł II

1. A. García, J. Godzien, Á. López-González, C. Barbas. Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics. *Bioanalysis*, 2017, 9(1):99-130.

Moduł III

1. L. Feng, G.D. Prestwich (ed.) Functional Lipidomics. Taylor & Francis, Boca Raton, 2006.
2. D. Armstrong. Lipidomics: Volumes 1e 2: Methods and Protocols (Methods in Molecular Biology). Humana Press, 2009, 1.
3. U. Loizides-Mangold. On the future of mass-spectrometry-based lipidomics. *FEBS J.*, 2013, 280(12):2817-29.
4. R. Harkewicz, E.A. Dennis. Applications of mass spectrometry to lipids and membranes. *Ann. Rev. Biochem.*, 2011, 80:301-25.

Moduł IV

1. J. Lovric. Introducing Proteomics: From concepts to sample separation, mass spectrometry and data analysis, Wiley. ISBN: 978-0-470-03523-8, 2011.
2. R. Aebersold, M. Mann. Mass-spectrometric exploration of proteome structure and function, *Nat. Rev. Mol. Cell. Biol.*, 2015, 16(5):269-80.
3. A.M. Silva, R. Vitorino, M.R. Domingues, C.M. Spickett, P. Domingues. Post-translational Modifications and Mass Spectrometry Detection, *Free Radic. Biol. Med.*, 2013, 65:925-41.

Moduł 1

Zastosowanie metod wykorzystujących połączenie technik separacyjnych i spektrometrii mas w naukach biomedyczo-farmaceutycznych

Wojciech Łuczaj¹, Antonia Garcia², Agnieszka Gęgotek¹, Katarzyna Bielawska¹,
Elżbieta Skrzydlewska¹

¹Department of Analytical Chemistry, Medical University of Białystok, 15-089 Białystok, Poland

²Centre for Metabolomics and Bioanalysis (CEMBIO), Facultad de Farmacia, Universidad San Pablo CEU; Boadilla del Monte, 28668 Madrid, Spain

I. Wprowadzenie

Pierwszy moduł kursu ACCLifeSci dotyczy podstawowych zagadnień związanych z wykorzystaniem spektrometrii mas w badaniach bimolekularnych. W ramach tego modułu zostaną omówione najnowsze rozwiązania techniczne stosowane w spektrometrach mas z uwzględnieniem ich zalet i ograniczeń. Założeniem kursu jest również przekazanie wiedzy dotyczącej interpretacji widm masowych w celu identyfikacji związków i przewidywania ich struktury. Zostaną także przedstawione podstawowe zagadnienia dotyczące połączenia wysokosprawnych technik separacyjnych ze spektrometrią mas, oraz dokładnie scharakteryzowane poszczególne elementy wchodzące w skład zestawu LC-MS. Szczególna uwaga zostanie także poświęcona wyjaśnieniu mechanizmu rozdziału chromatograficznego. W związku z powyższym głównym założeniem modułu jest nabycie przez studentów wiedzy i umiejętności wykorzystywania podstawowych narzędzi analitycznych, co umożliwi im prowadzenie badań „omicznych” wykorzystujących spektrometrię mas.

II. Założenia i efekty kształcenia

Założenia

Głównym założonym celem jest przekazanie studentom wiedzy teoretycznej z zakresu chemii, biochemii, bioinformatyki oraz umożliwienie zdobycia praktycznych umiejętności przeprowadzania analiz opartych o metody łączące techniki separacyjne ze spektrometrią mas w przełożeniu na ich wykorzystanie w naukach o życiu. Omówione zostaną potencjalne możliwości wykorzystania tych połączeń w badaniach omicznych, z przedyskutowaniem ich zalet, ograniczeń oraz naukowych zastosowań. Rozwijane w ramach kursu praktyczne umiejętności oparte będą w dużym stopniu o ogólnodostępne bazy danych oraz darmowe programy informatyczne wykorzystywane w analizach omicznych.

Specyficzne efekty kształcenia:

Po ukończeniu kursu studenci powinni potrafić:

1. opisać podstawowe mechanizmy związane z procesem jonizacji związków, ich rozdziału oraz detekcji z wykorzystaniem spektrometrii mas;
2. opisać podstawowe założenia tandemowej spektrometrii mas, jej zastosowanie oraz dedykowaną aparaturę analityczną;
3. przeprowadzić identyfikację związków na podstawie widm masowych;
4. opisać zasadę rozdziału, optymalizacji i wykrywania związków przy użyciu chromatografii ciekłowej, gazowej oraz elektroforezy kapilarnej;

5. przewidzieć jak zmiana parametrów wpłynie na rozdział chromatograficzny prowadzony z wykorzystaniem tych metod;
6. opisać budowę i zasadę działania modułów wchodzących w skład zestawu stanowiącego połączenie modułu separacyjnego i spektrometru mas;
7. wybrać najodpowiedniejszą metodologię w celu efektywnego rozdziału i detekcji analizowanych związków;
8. dokonać wyboru zestawu aparatury analitycznej i uzasadnić ten wybór w oparciu o wydajność oraz wymagania jakie determinują poszczególne podejścia „omiczne”.

III. Spis treści

Moduł 1 – Zastosowanie metod łączących techniki separacyjne ze spektrometrią mas w naukach biomedyczo-farmaceutycznych

I. Wprowadzenie

II. Założenia i efekty kształcenia

III. Spis treści

1. Spektrometria mas

1.1. Wprowadzenie

1.2. Metody jonizacji próbek biologicznych w spektrometrii mas

1.3. Typy analizatorów stosowanych w spektrometrii mas

1.4. Tandemowa spektrometria mas

1.5. Interpretacja widm masowych

2. Techniki separacyjne sprzężone ze spektrometrią mas

2.1. Chromatografia cieczowa – spektrometria mas (LC-MS)

2.1.1. Wprowadzenie do LC-MS

2.1.2. Teoretyczne podstawy chromatografii cieczowej

2.1.2.1. Proces rozdziału chromatograficznego

2.1.2.2. Chromatografia w normalnym układzie faz (NP)

2.1.2.3. Chromatografia w odwróconym układzie faz (RP)

2.1.2.4. Chromatografia oddziaływań hydrofilowych (HILIC)

2.1.2.5. Chromatografia jonowymienna

2.1.3. Moduły systemu LC-MS

2.2. Chromatografia gazowa – spektrometria mas (GC-MS)

2.2.1. Wprowadzenie do GC-MS

2.2.2. Podstawy teoretyczne GC-MS

2.2.2.1. Proces chromatograficzny

2.2.2.2. Kolumny chromatograficzne

2.2.2.3. GC×GC/MS

2.2.2.4. Przygotowanie próbek do analizy chromatograficznej.

Derywatywacja.

2.2.3. Połączenie GC-MS

2.3. Elektroforeza kapilarna – spektrometria mas (CE-MS)

2.3.1. Wstęp do CE-MS

2.3.2. Teoretyczne zasady techniki CE-MS

2.3.2.1. Kapilarna elektroforeza strefowa

2.3.2.2. Micelarna chromatografia elektrokinetyczna

2.3.2.3. Elektrokromatografia kapilarna

2.3.2.4. Kapilarna elektroforeza żelowa

2.3.2.5. Aparatura

3. Część praktyczna

3.1. Rejestracja widm masowych ESI-MS

3.2. Rejestracja widm masowych ESI-MS/MS

3.3. Interpretacja widm masowych

3.4. Wykorzystanie metody GC-MS w trybie pracy monitorowania pojedynczych jonów

3.5. Wykonanie analiz przy użyciu metody HPLC-MS oraz konfiguracja parametrów dla trybu data dependent MS/MS

1. Spektrometria mas

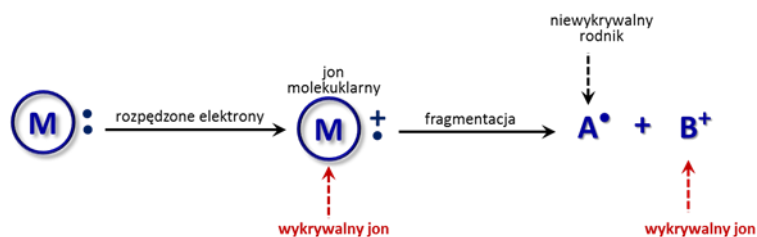
1.1. Wprowadzenie

W ostatnich latach dokonał się ogromny postęp związany z rozwojem spektrometrii mas (MS), jako techniki analitycznej. Dlatego obecnie technika ta jest szeroko stosowana do identyfikacji białek, lipidów i metabolitów drobnocząsteczkowych oraz ilościowej oceny ich zawartości w materiale biologicznym. Zastosowanie ulepszonych technik jonizacji próbek, fragmentacji związków oraz pojawienie się bardziej czułych i dokładniejszych analizatorów spowodowało, że spektrometria mas pozostaje w ścisłym związku z naukami o życiu. Do tego przyczynił się również znaczny rozwój technik separacyjnych stosowanych w połączeniu ze spektrometrią mas. Z uwagi na powyższe tematyka prowadzonych zajęć oparta będzie o zagadnienia związane z nowoczesną spektrometrią mas oraz jej praktycznym wykorzystaniu w połączeniu z różnymi technikami separacyjnymi.

Spektrometria mas jest wyspecjalizowaną techniką analityczną pozwalającą na dokładne określenie masy badanego związku, jego budowy strukturalnej jak również ilości w badanym materiale. Procedura analityczna oparta o technikę MS obejmuje kilka etapów począwszy od jonizacji próbki w źródle, przez rozdzielenie powstałych jonów w analizatorze w oparciu o stosunek ich masy do ładunku (m/z), następnie ich identyfikację w detektorze oraz interpretację widm masowych, stanowiących końcowy wynik przeprowadzonych analiz. Z kolei tandemowa spektrometria mas (MS/MS), dzięki sfragmentowaniu wyizolowanych jonów analitu, umożliwia uzyskanie dodatkowych szczegółowych informacji na temat ich struktury.

1.2. Metody jonizacji próbek biologicznych w spektrometrii mas

Istnieje kilka metod jonizacji związków organicznych umożliwiających ich analizę techniką MS. Jonizacji elektronowa (EI - Electron impact ionization) stanowi tradycyjną już metodę używaną do jonizacji lotnych związków, charakteryzujących się niską masą cząsteczkową. Związki jonizowane przy użyciu tej metody oprócz lotności powinny również charakteryzować się wysoką odpornością na zmiany temperatury. W chwili obecnej ten typ jonizacji stosowany jest przede wszystkim w połączeniu z chromatografią gazową (GC - gas chromatography).



Rycina 1.1. Jonizacja i fragmentacja w źródle typu EI.

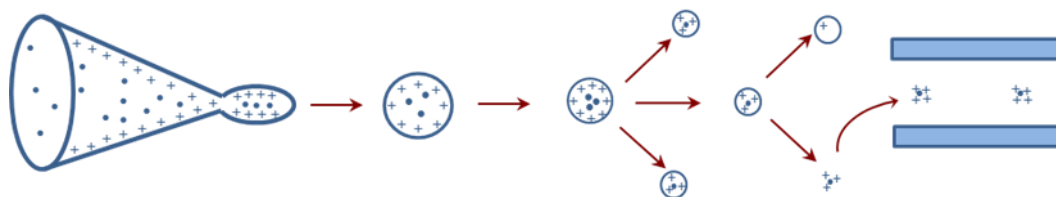
W metodzie tej dochodzi do jonizacji cząsteczki związku znajdującego się w fazie gazowej poprzez bombardowanie jej rozpędzonymi elektronami. Efektem procesu jest także fragmentacja związku, co z kolei umożliwi jego dokładną identyfikację dzięki istnieniu bibliotek widm fragmentacyjnych związków, uzyskanych przy użyciu tego typu jonizacji. Sam mechanizm jonizacji metodą EI polega na oderwaniu pojedynczego elektronu od cząsteczki związku, dzięki czemu powstaje kationorodnik (M^+), którego stosunek m/z liczbowo odpowiada masie molekularnej zjonizowanego związku (Rycina 1.1).

Podstawową wadą tego typu jonizacji jest fakt, że analizowane związki muszą być lotne, co w konsekwencji wyklucza analizę związków polarnych, nielotnych oraz labilnych termicznie, chyba, że zostaną one poddane derywatywacji mającej na celu uzyskanie ich lotnych pochodnych. Kolejnym podstawowym ograniczeniem jonizacji elektronowej jest fragmentacja związku, skutkiem czego jest brak obecności jonu molekularnego analizowanego związku na widmie masowym. W konsekwencji ważna informacja analityczna dotycząca dokładnej masy związku zostaje utracona. Z uwagi na tę niedogodność pod koniec lat osiemdziesiątych opracowano dwie nowe metody jonizacji typu elektrosprej (Electrospray ionization ESI) oraz laserowa jonizacja wspomaganą matrycą (MALDI - matrix-assisted laser desorption ionization). Obie metody w odróżnieniu od metody EI pozwalają na jonizację nielotnych oraz labilnych termicznie związków organicznych, dzięki czemu obecnie są one najczęściej stosowanymi metodami jonizacji związków wielkocząsteczkowych takich jak białka czy peptydy. Z uwagi na fakt, że głównym jonom powstającym w wyniku procesu jonizacji jest jon molekularny, a fragmentacja zachodzi w znikomym stopniu, MALDI i ESI zaliczane są do tzw. „miękkich” metod jonizacji (soft ionization techniques). Największą zaletą obu metod jest pominięcie etapu derywatywacji związków, co niezbędne jest w przypadku analiz techniką GC-MS.

Zarówno metoda ESI jak i MALDI różnią się znacznie pod względem mechanizmu jonizacji od metody EI. W przypadku ESI za powstawanie jonów odpowiedzialne jest wysokie napięcie przyłożone do metalowej kapilary, przez którą przepływa faza ruchoma zawierająca analizowany związek. Wysokie napięcie powoduje, że w źródle ESI powstaje rozproszony stożek naładowanych na powierzchni kropli (spray), które następnie ulegają desolwatacji, w wyniku czego krople stają się coraz mniejsze przeobrażając się w kilka znacznie mniejszych kropli dzięki tzw. „wyładowaniom Coulomba” będących skutkiem odpychania się jednoimiennych ładunków zgromadzonych na powierzchni kropli (Rycina 1.2). Efektem końcowym całego procesu jest wytworzenie naładowanych jonów analitu rozproszonych pod ciśnieniem atmosferycznym w fazie gazowej.

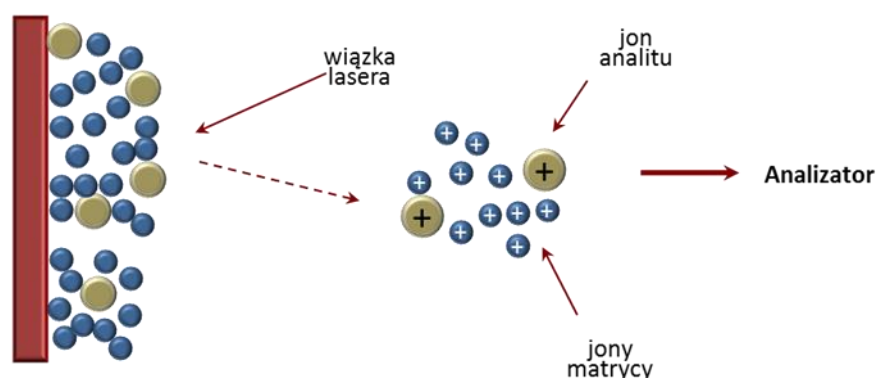
Główną zaletą ESI jest jonizacja zachodząca bezpośrednio w roztworze, dzięki czemu termolabilne związki nie ulegając jednoczesnej degradacji. Szybkość przepływu fazy

ruchomej ma istotny wpływ na efektywność jonizacji w źródle ESI. Najbardziej efektywna jonizacja ma miejsce, gdy prędkość przepływu fazy ruchomej w systemie stanowiącym połączenie chromatografu cieczowego ze spektrometrem mas, zawierającym źródło ESI (LC-ESI-MS) zawiera się w przedziale od 5 do 100 $\mu\text{L min}^{-1}$. Najczęściej stosowaną fazą ruchomą przy metodzie ESI jest mieszanina wody z acetonitrylem lub metanolem. Czyste rozpuszczalniki organiczne, jako ciecze niepolarne nie zapewniają wystarczająco efektywnego procesu jonizacji. Dodatek niewielkiej ilości kwasu octowego lub mrówkowego (0.1%-1%) do fazy ruchomej wspomaga jonizację ESI w trybie pozytywnym z uwagi na zwiększoną tendencję do protonowania związków. Z kolei dodatek soli (takich jak mrówczan amonu czy szczawian amonu) znacznie wspomaga zjawisko przechodzenia analitu z cząsteczek obojętnych w jony naładowane ujemnie, dzięki tworzeniu odpowiednich adduktów z jonami soli.



Rycina 1.2. Mechanizm jonizacji typu elektrosprej (ESI).

Jonizację MALDI powoduje napromieniowanie próbki, wcześniej poddanej krystalizacji w matrycy, wiązką lasera o odpowiedniej energii w postaci krótkiego nanosekundowego impulsu. Dokładny mechanizm jonizacji MALDI jest nadal przedmiotem wyjaśnień. Przypuszcza się, że obejmuje on desorpcję próbki oraz jej jonizację na zasadzie transferu protonu przy udziale związków matrycy. Głównymi etapami w procesie jonizacji metodą MALDI jest rozpuszczenie analitu w roztworze zawierającym związki matrycy, który następnie jest odparowywany. W kolejnym etapie, odbywającym się pod poróżnią, na skutek impulsu lasera (Rycina 1.3) dochodzi do desorpcji.



Rycina 1.3. Mechanizm laserowej jonizacji wspomaganiej matrycą (MALDI).

Kluczową rolę w procesie jonizacji metodą MALDI spełnia matryca, którą najczęściej stanowią małowcząsteczkowe związki organiczne, takie jak 2,5-dihydroksybenzoesowy kwas (DHB), kwas nikotynowy, synapinowy czy kwas 3,5-dimetoksy-4-hydroksycynamonowy. Dobór matrycy uzależniony jest od właściwości analitu oraz długości fali światła emitowanego przez laser, która najczęściej wynosi 337 nm. Należy podkreślić, że rodzaj matrycy decyduje o efektywności jonizacji w trybie pozytywnym bądź negatywnym. Przykładowo DHB nadaje się do analizy lipidów zarówno w trybie jonizacji pozytywnej jak i negatywnej, podczas gdy p-nitroanilina zapewnia większą czułość analiz fosfatydyloetanoloamin w trybie jonizacji negatywnej.

Zarówno jonizacja MALDI jak i ESI ma szereg wspólnych zalet, pośród których należy wymienić wysoką czułość umożliwiającą przeprowadzenie analiz próbek zawierających pikomolowe ilości analitu. Dodatkowo prosta obsługa tych źródeł powoduje, że są one powszechnie stosowane w przypadku aplikacji mających na celu przeprowadzenie szybkich analiz skринingowych. Jeśli chodzi o wady, to w przypadku jonizacji MALDI główną stanowi jednoczesna jonizacja cząsteczek analitu oraz związków matrycy, co utrudnia interpretację widm masowych w zakresie niskich wartości m/z (poniżej 800 Da), obejmującym sygnały pochodzące od jonów fragmentacyjnych, co w konsekwencji stwarza duży problem w dokładnym określeniu budowy strukturalnej analizowanych związków.

Innym typem "jonizacji miękkiej" wykorzystywanej obecnie do jonizacji związków nietlotnych jest jonizacja chemiczna pod ciśnieniem atmosferycznym APCI (atmospheric-pressure chemical ionization). W przypadku jonizacji APCI podobnie jak w ESI, strumień fazy ruchomej zawierającej analit ulega rozproszeniu z utworzeniem małych kropli jak opisano wyżej. Natomiast główną różnicę stanowi mechanizm formowania tych kropli, a w dalszym etapie również jonów. W przypadku APCI wymaga to zastosowania zarówno wysokiej temperatury jak i gazu nebulizującego. Uzyskany aerosol jest następnie poddawany działaniu pola elektrycznego znajdującego się w obszarze gdzie dochodzi do tzw. wyładowań koronowych (*corona discharge*), które inicjują szereg procesów takich jak wytworzenie plazmy gazowej oraz przeniesienie protonu, prowadzących w konsekwencji do wytworzenia jonów w fazie gazowej. Generalnie jonizacja APCI jest zalecana w przypadku związków mniej polarnych, natomiast wymóg stosowania dużych szybkości przepływu fazy ruchomej predysponuje do stosowania jej w systemach HPLC wyposażonych w klasyczne kolumny chromatograficzne. W tabeli 1.1. zebrano podstawowe dane dotyczące różnych metod jonizacji i porównano je ze sobą.

1.3. Typy analizatorów stosowanych w spektrometrii mas

Widmo masowe stanowi szereg sygnałów, dla których przypisana jest wartość stosunku m/z (masy do ładunku) jonów wytworzonych w procesie jonizacji związku. Otrzymanie widma masowego warunkuje możliwość rozdzielania jonów o różnych masach oraz określenia ich relatywnej ilości. W tym celu stosuje się różne technologie wykorzystywane przez analizatory MS, które różnią się rozdzielczością, dokładnością pomiaru masy, dynamicznym zakresem pomiaru oraz możliwością przeprowadzania tandemowych eksperymentów MS/MS lub brakiem takiej możliwości. Podstawowe parametry charakteryzujące analizator MS definiuje się następująco:

Rozdzielczość – termin ten w chemii analitycznej odnosi się do różnych obszarów zainteresowań i najczęściej definiuje możliwość rozdzielania dwóch sygnałów o bardzo zbliżonych właściwościach. Zbliżone właściwości sygnałów w spektrometrii mas odnoszą się

do zbliżonych wartości stosunku m/z jonów, dlatego rozdzielczość, jak i zdolność rozdzielcza (RP) w tej technice definiowana jest poniższymi wzorami:

$$R = m/\Delta m,$$

$$RP = m_1/(m_2 - m_1),$$

gdzie R - rozdzielczość, RP – zdolność rozdzielcza, m – wartość stosunku m/z i Δm – różnica mas, m_1 – wartość stosunku m/z lżejszego jonu, $(m_1 - m_2)$ – różnica w wartości stosunku m/z następujących po sobie jonów (wyrażona w Da) mierzona jako szerokość pików na specyficznym zdefiniowanej jego wysokości; zazwyczaj jest to szerokość w połowie wysokości pików (FWHM - full width at half maximum).

Dokładność (mass accuracy) – daje informację na temat tego czy zmierzona wartość zgodna jest z wartością rzeczywistą. Dokładność pomiaru masy w spektrometrii mas określana jest jako błąd pomiaru wyrażony w ppm i obliczany na podstawie poniższego wzoru:

Błąd pomiaru masy (mass error) = (wyliczona teoretycznie dokładna masa monoizotopowa – dokładna masa zmierzona) / wyliczona teoretycznie dokładna masa monizotopowa $\times 10^6$

Analizator zapewniający pomiar masy z dokładnością około ± 0.1 Da charakteryzuje się dokładnością pomiaru masy jednej jednostki (one unit). Zazwyczaj większa rozdzielczość instrumentu umożliwia dokładniejszy pomiar masy, co w praktyce oznacza lepsze możliwości spektrometru mas do identyfikacji związków.

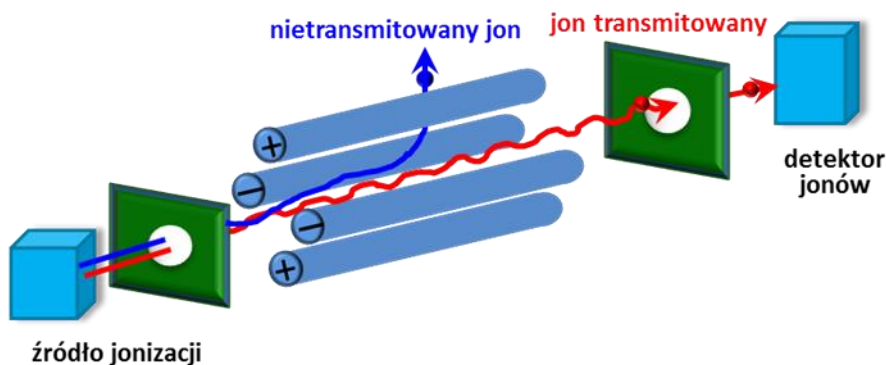
Tabela 1.1. Charakterystyka różnych technik jonizacji.

Technika jonizacji	Analit	Zakres mas	Sposób wprowadzania próbek	Zalety	Wady
EI	związki drobnocząsteczkowe niepolarne, lotne, stabilne termicznie,	<1 kDa	GC lub próbki ciekłe/stałe	analiza związków nierozpuszczalnych i niepolarnych, brak supresji jonów, łatwa do aplikacji z GC, istnienie bibliotek widm	analiza związków tylko lotnych - wymóg stabilności termicznej, związki o niskich masach cząsteczkowych jonizacja twarda
ESI	związki polarne: t.j. peptydy, białka, cukry, nukleotydy	<200 kDa	LC lub z roztworu	analiza związków termolabilnych wielkocząsteczkowych, wysoka czułość, łatwa do aplikacji z LC, jonizacja miękka	wymagana rozpuszczalność w polarnych rozpuszczalnikach, wrażliwa na obecność soli w próbce, supresja jonów
MALDI	związki polarne: t.j. peptydy, białka, cukry, nukleotydy	<500 kDa	związki rozpuszczalne w stałej matrycy	analiza związków termolabilnych, wielkocząsteczkowych, wysoka czułość, mniej wrażliwe na obecność soli w próbce, jonizacja miękka	szeroki wybór matryc, trudności w analizach ilościowych, supresja jonów
APCI	związki mniej polarne: t.j. peptydy, białka, cukry, nukleotydy	<1 kDa	LC lub z roztworu	wysokie szybkości przepływu, łatwa do aplikacji z LC, jonizacja miękka	wymagana rozpuszczalność analitów w rozpuszczalnikach polarnych, wrażliwa na obecność soli w próbce, supresja jonów

Liniowy zakres dynamiczny – definiowany jest jako zdolność spektrometru mas do rejestracji przyrostu sygnału warunkowanego zwiększającą się ilością analitu. Zakres liniowości ma kluczowe znaczenie jeśli chodzi o dokładność pomiarów, zwłaszcza w przypadku analiz ilościowych.

Istnieje kilka typów urządzeń używanych w spektrometrii mas, które umożliwiają separację jonów z uwagi na różną ich masę, zwanych analizatorami, lecz każdy z nich posiada zarówno swoje zalety jak wady. Analizatory mas wykorzystywane do analiz próbek biologicznych to kwadrupole, pułapki jonowe, analizatory czasu przelotu, oraz wprowadzone w ostatnich latach pułapki jonowe typu orbitrap.

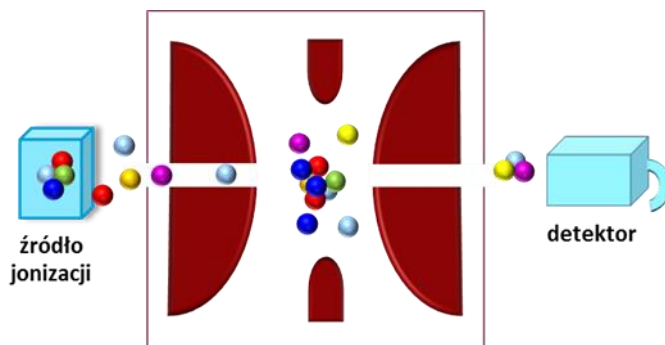
Kwadrupol (Q), jest zbudowany jak jego sama nazwa wskazuje z czterech równoległych ułożonych prętów, jak zostało to zilustrowane na rycinie 1.4. Do każdej z dwóch par prętów kwadrupola, ułożonych na przeciw siebie przyłożone jest zarówno napięcie stałe jak i zmienne. Zastosowanie specyficznych wartości obu napięć pozwala jonom o konkretnej wartości m/z zachować stabilną trajektorię podczas poruszania się wewnątrz analizatora, dzięki czemu swobodnie go opuszczają i trafiają do detektora, podczas gdy pozostałe jony będą usuwane z analizatora. Opisana sytuacja obrazuje podstawowy tryb pracy kwadrupola, zwany monitorowaniem wybranych jonów SIM (selected ion monitoring), który przy zapewnieniu odpowiedniego stosunku sygnału do szumu z powodzeniem stosowany jest w przypadku targetowych analizach ilościowych. Z drugiej strony analizator typu kwadrupola może być wykorzystywany do skanowania w tzw. pełnym zakresie mas (full-scan mode). Podczas tego trybu pracy na skutek zmiany w czasie stosunku przyłożonego stałego i zmiennego napięcia wszystkie jony powstałe w źródle są przepuszczone przez kwadrupol i trafiają do detektora w odpowiednim czasie. W wyniku skanowania w pełnym zakresie mas uzyskiwane jest widmo masowe, które w przypadku kwadrupola charakteryzuje się niską rozdzielczością.



Rycina 1.4. Schemat kwadrupolowego analizatora mas (Q).

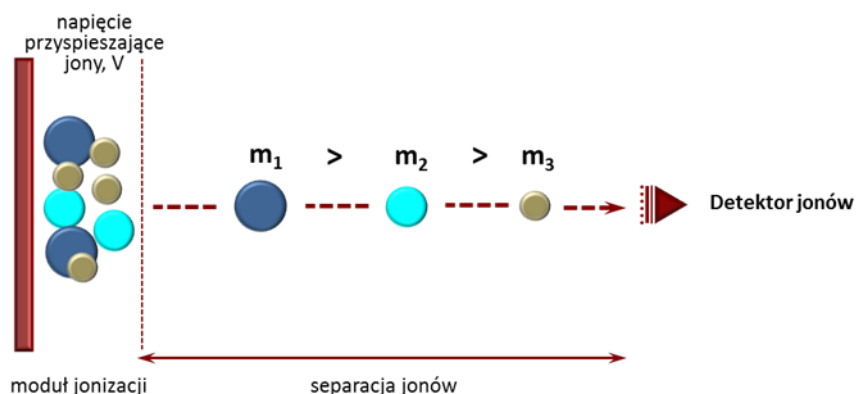
Pułapka jonowa (IT) (ion-trap) jest analizatorem mas, który jak sama jego nazwa wskazuje powoduje uwięzienie jonów wykorzystując do tego celu kwadrupolowe pole magnetyczne. Istnieją dwa typy pułapek jonowych 2D i 3D, które zatrzymują jony odpowiednio w płaszczyźnie dwuwymiarowej lub przestrzeni trójwymiarowej. Pułapki jonowe 3D, zwane także kwadrupolowymi pułapkami jonowymi składają się z czterech elektrod: dwóch pierścieniowych i dwóch eliptycznych połączonych z elektrodami. Przyłożenie do elektrod odpowiedniej wartości stałego napięcia powoduje zatrzymanie jonów (Rycina 1.5). Wszystkie jony wprowadzone do pułapki ze źródła jonizacji są w niej „uwięzione” na skutek działania trójwymiarowego kwadrupolowego, zmiennego pola elektrycznego. W wyniku przyłożenia odpowiedniej wartości napięcia do elektrod

pierścieniowych jony o specyficznej wartości stosunku m/z nie są zatrzymywane, opuszczają pułapkę i trafiają do detektora. Zmiana wartości napięcia skutkuje tym, że jony o innych wartościach m/z trafiają do detektora. W przypadku pułapki jonowej typu 2D, tzw. liniowej pułapki jonowej, jony więzione są na skutek działania zmiennego pola elektrycznego wytworzonego w płaszczyźnie dwuwymiarowej. Wewnątrz pułapki 2D hamujący potencjał jest przyłożony do elektrod na początku i końcu kwadrupola w celu ograniczenia ruchów jonów. Oba analizatory typu pułapki jonowej są instrumentami o niskiej rozdzielczości. Jednak liniowa pułapka jonowa charakteryzuje się większymi możliwościami w zakresie ilości zatrzymywanych jonów, jak również szybkości skanownia. W przypadku obu analizatorów niezbędny jest hel, jako gaz obojętny, zapewniający odpowiednią trajektorię ruchu jonów, jak również w celu przeprowadzenia eksperymentów MS^n .



Rycina 1.5. Schemat analizatora mas typu pułapki jonowej 3D.

Analizator czasu przelotu (TOF; time-of-flight) (Rycina 1.6) wykorzystuje fakt, że prędkość jonów powstałych w źródle i obdarzonych taką samą energią kinetyczną (E_k), to ich prędkość będzie zależna od ich wartości stosunku m/z , co ujmuje następująca zależność ($E_k = mv^2/2z$). W praktyce to oznacza, że czas jaki potrzebują te jony na pokonanie drogi wewnątrz rury analizatora czasu przelotu, również będzie zależał od ich stosunku m/z . W konsekwencji po upływie czasu, w trakcie którego wszystkie jony dotrą do detektora uzyskiwane jest kompletne widmo masowe analizowanej próbki.



Rycina 1.6. Mechanizm separacji jonów w analizatorze mas czasu przelotu (TOF).

Najważniejszą zaletą analizatora TOF jest znacznie wyższa rozdzielczość w porównaniu do potrójnego kwadrupola jak również pułapki jonowej, co umożliwia bardziej dokładny pomiar stosunku m/z . Rozdzielczość w przypadku analizatora TOF determinowana jest możliwością zmierzenia nawet znikomych różnic w czasie, w których do detektora trafiają jony o zbliżonej wartości m/z . Możliwość pomiaru tych rozbieżności w czasie można zwiększyć poprzez

zwiększenie długości rury czasu przelotu, co w konsekwencji wydłuży czas jaki jony potrzebują na pokonanie dystansu od źródła do detektora. Innym rozwiązaniem pozwalającym na zwiększenie rozdzielczości jest zastosowanie w analizatorach TOF zwierciadeł, zwanych reflektromi, umieszczanych na końcu rury czasu przelotu. Wiązka jonów jest odbijana od reflektromu wydłużając w ten sposób dwukrotnie drogę wędrówki jonów. Reflektrom dodatkowo zwiększa rozdzielczość analizatora poprzez zmniejszenie rozproszenia energii kinetycznej. Inną dodatkową zaletą analizatorów TOF jest duża szybkość skanowania, którą ogranicza jedynie czas jaki jest potrzebny najcięższym jonom na dotarcie do detektora. Kolejną istotną zaletą liniowego analizatora TOF jest brak ograniczenia zakresu mierzonych mas, co jednocześnie predysponuje użycie tego analizatora do analiz np. nietkniętych, całych białek.

Na przestrzeni ostatnich lat opracowano nowy typ analizatora nazwanego *orbitrap*, który stanowi modyfikację pułapki jonowej. Analizator typu orbitrap, składa się z dwóch elektrod, centralnej - wewnętrznej oraz zewnętrznej, których kształt przypomina zwięzające się beczki. Jony więzione są wewnątrz analizatora wokół centralnej elektrody, natomiast wartości m/z obliczane są na podstawie transformacji Fouriera w oparciu o częstotliwość oscylacji zatrzymywanych jonów. Ten typ analizatora charakteryzuje się wysoką rozdzielczością rzędu powyżej 500,000 FWHM (full width at half maximum), wysoką dokładnością pomiaru masy (<1pp z wewnętrzną kalibracją) oraz wysoką czułością. Jako podsumowanie, tabela 1.2 przedstawia porównanie podstawowych parametrów różnych analizatorów stosowanych w spektrometrii mas.

Tabela 1.2. Porównanie parametrów różnych analizatorów mas.

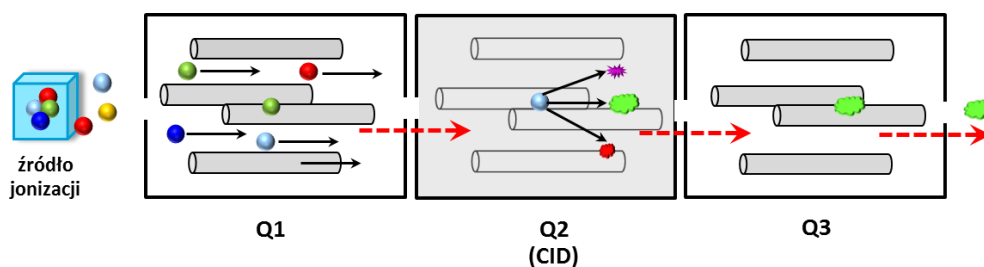
Analizator	Q	IT	TOF	Orbitrap
Szybkość zbierania danych (Hz)	2-10	2-10	10-100	1-18
Dokładność pomiaru (ppm)	niska	niska	1-10 ^a ppm	1-5 ppm
Zakres pomiaru (m/z)	<3000	<6000	<100,000 ^a	<6000
Rozdzielczość	jednostka	jednostka	<50,000 ^a	<500,000 ^b
Zalety	łatwa aplikacja różnych źródeł jonizacji, szerszy zakres dynamiczny	łatwa aplikacja różnych technik jonizacji, MS ⁿ	duża szybkość skanowania, szeroki zakres mas, duża dokładność pomiaru masy, wysoka rozdzielczość	duża dokładność pomiaru masy, możliwość analizy w dwóch polaryzacjach jednocześnie, wysoka czułość, wysoka rozdzielczość
Wady	niska rozdzielczość, mała dokładność pomiaru masy, wąski zakres mas, mała szybkość skanowania, MS/MS wymaga kilku analizatorów	niska rozdzielczość, mała dokładność pomiaru masy, wąski zakres mas, mała szybkość skanowania,	węższy zakres dynamiczny w porównaniu do Q	mniejsza szybkość skanowania w porównaniu z QTOF, węższy zakres dynamiczny w porównaniu do Q

a) w hybrydowej/reflektromowej konfiguracji; b) rozdzielczość robocza (FWHM)

1.4. Tandemowa spektrometria mas

Tandemowa spektrometria mas (MS/MS) jest definiowana jako technika, która obejmuje jednocześnie dwa eksperymenty MS, pomiędzy którymi ma miejsce fragmentacja jonów. Analizy MS/MS mogą być prowadzone jedynie przy wykorzystaniu tandemowych spektrometrów mas, złożonych z dwóch oddzielnych analizatorów mas. Oba moduły analizatorów przedzielone są komorą zawierającą gaz kolizyjny, w której dochodzi do fragmentacji jonów metodą dysocjacji indukowanej kolizyjnie CID (collision-induced dissociation). Tandemowe eksperymenty MS wykorzystywane są przede wszystkim do targetowych analiz ilościowych, w których duża dokładność pomiaru masy jak i wysoka rozdzielczość nie jest wymagana, jak również analiz jakościowych mających na celu identyfikację związków.

Tandemowym spektrometrem mas najczęściej wykorzystywanym do targetowych oznaczeń ilościowych jest *potrójny kwadrupol (QqQ)*. Ten spektrometr masowy jak sama nazwa wskazuje składa się z trzech analizatorów kwadrupolowych (Rycina 1.7), z czego w rzeczywistości drugi kwadrupol (Q2) stanowi komora kolizyjna (q), w której dochodzi do dysocjacji jonów indukowanej kolizją z cząsteczkami gazu obojętnego (CID). Zatem w drugim module (Q2) potrójnego kwadrupola, jony wybrane z pierwszego kwadrupola (Q1), ulegają fragmentacji, a powstałe jony fragmentacyjne transmitowane są do trzeciego kwadrupola (Q3).



Rycina 1.7. Tandemowa spektrometria mas w oparciu o działanie potrójnego kwadrupola.

Oprócz potrójnego kwadrupola innymi powszechnie stosowanymi tandemowymi spektrometrami mas, są spektrometry hybrydowe, wykorzystujące połączenie różnego rodzaju analizatorów. Typowym przykładem hybrydowego spektrometru mas jest połączenie kwadrupola z analizatorem czasu przelotu (QqTOF) (z ang. quadrupole time-of-flight) czy połączenia kwadrupola z analizatorem orbitrap (QqOrbitrap). Przykładem innego znanego tandemowego spektrometru mas jest pułapka jonowa, w której wszystkie etapy analizy w trybie MSMS odbywają się w obrębie jednego analizatora (pułapki jonowej). W przypadku pułapek jonowych wykorzystywane są inne mechanizmy fragmentacji, przede wszystkim oparte o zjawisko dysocjacji związanej z przeniesieniem elektronu (ETD - electron-transfer dissociation) lub dysocjacji będącej wynikiem wysokoenergetycznych kolizji (HCD - higher-energy collisional dissociation).

Wyróżnia się cztery główne typy analiz w trybie MS/MS, które można przeprowadzić przy użyciu tandemowych spektrometrów mas, z czego trzy z nich możliwe są tylko przy zastosowaniu potrójnego kwadrupola:

Skanowanie jonów wtórnych (product-ion scan): w pierwszym module tandemowego spektrometru mas (Q1) wybierane są jony molekularne oznaczanego związku. Wyselekcjonowane jony poddawane są następnie fragmentacji w komorze kolizyjnej (q) na skutek zderzeń z cząsteczkami gazu kolizyjnego. W trzecim kwadrupolu (Q3) uzyskiwane

jest widmo fragmentacyjne w wyniku skanowania jonów potomnych powstałych w komorze kolizyjnej. Ten typ analizy może być przeprowadzony przy użyciu wszystkich dostępnych rodzajów tandemowych spektrometrów mas.

Skanowanie jonów pierwotnych/macierzystych (z ang. precursor-ion scan): trzeci kwadrupol (Q3) przepuszcza jedynie jony fragmentacyjne o znanych wartościach m/z , podczas gdy pierwszy kwadrupol (Q1) pracuje w trybie skanowania w zadeklarowanym zakresie mas przepuszczając wszystkie jony prekursorów, które w kolejnym etapie ulegają fragmentacji w komorze kolizyjnej (q). Detektor rejestruje sygnał, tylko wtedy gdy jon przepuszczony przez moduł Q1 po sfragmentowaniu daje jon fragmentacyjny selektywnie przepuszczany przez moduł Q3. Skanowanie jonów pierwotnych możliwe jest tylko przy użyciu tandemowego spektrometru mas typu potrójnego kwadrupola.

W trybie skanowania *utrąty cząsteczki obojętnej (neutral loss scan mode)*, oba kwadrupole Q1 i Q3 pracują w trybie skanowania, przepuszczając jony tylko wtedy gdy różnica w ich wartościach m/z odpowiada masie neutralnej cząsteczki utraconej w wyniku fragmentacji w komorze kolizyjnej (q). Jony potomne, które powstały na skutek utraty innej masy niż masa wybranej neutralnej cząsteczki np. wody, nie są przepuszczane przez Q3 do detektora. Ten typ tandemowego eksperymentu przeprowadzić można tylko przy użyciu potrójnego kwadrupola.

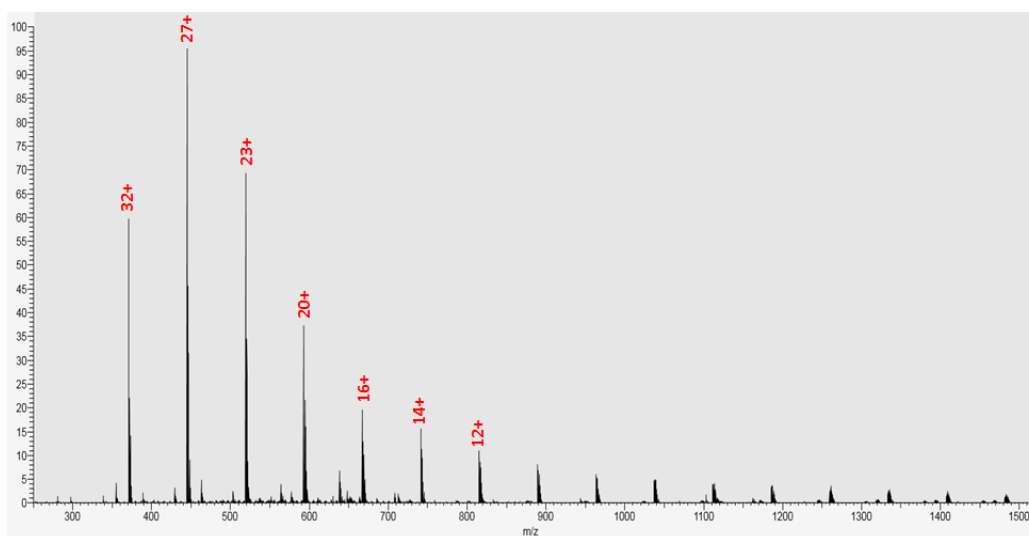
Monitorowanie reakcji wielokrotnych/następczych MRM (multiple reactions monitoring) oraz *monitorowanie wybranych reakcji fragmentacji SRM (selected reaction monitoring)* są typowymi trybami pracy tandemowego spektrometru mas, wykorzystywanymi do analiz ilościowych. W trybie SRM monitorowane jest przejście jonu molekularnego w jon potomny powstały w wyniku fragmentacji, wyselekcjonowany jako specyficzny fragment cząsteczki jonu molekularnego. Taki tryb pracy uzyskiwany jest w chwili gdy w poszczególnych kwadrupolach przepuszczane są tylko pojedyncze, wyselekcjonowane dwa jony. Jon prekursora wybierany jest przez moduł Q1, a charakterystyczny jon fragmentacyjny przez moduł kwadrupola Q3. Zasada pracy spektrometru w trybie MRM jest taka sama jak w trybie SRM, z tą różnicą, że monitorowanych jest więcej niż jedna para jonów prekursor – charakterystyczny fragment. Połączenie rozdziału chromatograficznego metodą chromatografii cieczowej z identyfikacją związków przy użyciu tandemowego spektrometru mas pracującego w trybie SRM/MRM pozwala na znaczne zwiększenie specyficzności pomiaru dzięki wyeliminowaniu par prekursor – charakterystyczny fragment, odpowiadających jonom pochodzącym od innych substancji obecnych w analizowanej próbce. Podstawową wadą tego trybu pracy jest brak możliwości jego zastosowania do globalnych analiz specyficznych związków, gdy nie można przyporządkować im pary charakterystycznych jonów prekursor - fragment. Z kolei detekcja substancji przy zastosowaniu trybu MRM ma znaczną przewagę nad trybem SIM z uwagi na znacznie wyższą czułość, selektywność oraz powtarzalność oznaczeń. Dlatego w chwili obecnej tryb MRM jest powszechnie stosowanym trybem w celu prowadzenia analiz ilościowych z udziałem potrójnego kwadrupola, jako detektora masowego.

1.5. Interpretacja widm masowych

W tym rozdziale omówione zostaną podstawowe zasady interpretacji widm masowych otrzymanych przy użyciu metod jonizacji opisanych w rozdziale 1.2. Na wstępie należy podkreślić, że na widmach masowych uzyskanych w wyniku jonizacji elektronowej obecne są

jedynie jony naładowane dodatnio, co odróżnia te widma od widm uzyskanych technikami MALDI, ESI czy APCI, przy użyciu których możliwa jest również jonizacja negatywna.

Specyficzną cechą jonizacja ESI, jest tworzenie jonów naładowanych wielokrotnie. Liczba powstających ładunków na cząsteczce związku zależy od jego struktury, a konkretnie liczby grup, w które mogą przyjąć lub utracić proton podczas odpowiednio pozytywnej lub negatywnej jonizacji. Zjawisko to ma kluczowe znaczenie podczas interpretacji widm masowych związków wielkocząsteczkowych takich jak np. natywne białka (Rycina 1.8), podczas gdy związki o niskich masach cząsteczkowych zazwyczaj posiadają znacznie mniej potencjalnych miejsc, gdzie mogłoby dojść do protonacji/deprotonacji. Dlatego widma masowe związków o niskich masach cząsteczkowych z reguły nie zawierają sygnałów pochodzących od jonów wielokrotnie naładowanych.

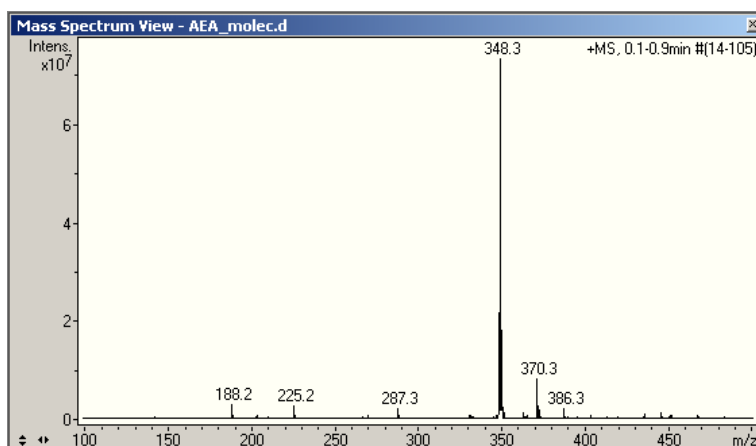


Rycina 1.8. Widmo ESI-MS tioredoksyny (12 KDa), jako przykład widma związku ulegającego wielokrotnej jonizacji [Q-Orbitrap, Thermo Fisher, Q Exactive Plus].

Na widmach masowych ESI związków o niskich masach cząsteczkowych obdarzonych tylko jednym ugrupowaniem polarnym, jako sygnał o największej intensywności obserwowany jest głównie jon molekularny, odpowiadający zazwyczaj jego protonowanej $[M+H]^+$ lub deprotonowanej $[M-H]^-$ formie, dzięki czemu budowę strukturalną takich związków jest łatwo przewidzieć (Rycina 1.9). Z uwagi na fakt, że niektóre związki jonizują lepiej w trybie negatywnym niż pozytywnym i odwrotnie, uzasadnione jest przeprowadzić ich analizy w obu trybach jonizacji. Dodatkowo możliwość tworzenia adduktów z takimi jonami jak Na^+ , K^+ , H_3O^+ , NH_4^+ oraz związkami dodanymi do fazy ruchomej (np. solami w postaci mrówczanów i octanów) musi być bezwzględnie brana pod uwagę podczas interpretacji sygnałów na widmie ESI. Widma masowe otrzymane w wyniku jonizacji APCI i MALDI są podobne do widm ESI, ponieważ zazwyczaj jony molekularne są obserwowane jako główne jony, jak również prawdopodobieństwo tworzenia przez nie jonów wielokrotnie naładowanych jest również małe.

W dalszej części rozdziału zostanie omówionych kilka aspektów wykorzystywanych podczas interpretacji widm masowych. Na widmach masowych można obserwować sygnały pochodzące od innych jonów niż jony molekularne związków. Sygnały te pochodzą od jonów fragmentacyjnych i ich intensywność zależy od wielu czynników takich jak metoda jonizacji, typ fragmentacji czy obecność nietrwałych/labilnych ugrupowań w obrębie cząsteczki

związku. W wyniku analiz wykorzystujących miękką jonizację zazwyczaj na widmie nie obserwowane są jony fragmentacyjne, chyba że są prowadzone przy użyciu tandemowego spektrometru mas. Z drugiej jednak strony podczas jonizacji elektronowej w analizowane



Rycina 1.9. Widmo masowe ESI dla anandamidu, otrzymane w trybie pozytywnej jonizacji; sygnał przy m/z 348.3 odpowiada dodatnio naładowanemu jonowi molekularnemu $[M+H]^+$ anandamidu [Pułapka jonowa, Bruker Daltonics, Esquire 6000].

związki w znacznym stopniu ulegają w źródle fragmentacji. Należy podkreślić, że sposób fragmentacji związków należących do tej samej grupy (np. białek czy lipidów) jest bardzo podobny. Dlatego wiedza na temat specyficznych mechanizmów fragmentacji związków z danej klasy jest bardzo istotna do ich identyfikacji na podstawie widm masowych. Przykładowo związki zawierające w swej strukturze łańcuchy węglowodorowe podczas jonizacji elektronowej dają na widmie masowym serię sygnałów różniących się o 14 Da, co odpowiada grupie $(-CH_2-)$. Z kolei utrata cząsteczki wody oznacza obniżenie o 18 Da wartości m/z jonu molekularnego danej substancji, z której jest ona eliminowana. Jest to efekt często obserwowany w przypadku związków bogatych w grupy hydroksylowe. Oprócz przytoczonych przykładów istnieje wiele innych specyficznych dla różnych grup związków mechanizmów fragmentacji, które z powodzeniem są wykorzystywane do ich identyfikacji.

2. Techniki separacyjne sprzężone ze spektrometrią mas

2.1. Chromatografia cieczowa - spektrometria mas (LC-MS)

W chromatografii cieczowej związki identyfikowane są na podstawie ich czasów retencji (t_R). Niestety w przypadku bardzo złożonych próbek zawierających związki o zbliżonych wartościach czasu retencji, rozdział chromatograficzny nie jest wystarczający do ich identyfikacji. W takiej sytuacji istnieje potrzeba dostarczenia większej ilości informacji, poprzez zastosowanie dodatkowej techniki, która w połączeniu z rozdziałem chromatograficznym zapewni wymierny efekt. Ponieważ spektrometria mas w chwili obecnej stanowi najbardziej czułą technikę analityczną, jej połączenie z chromatografią cieczową (LC-MS) w celu identyfikacji związków jest bardzo porządane. Jedną z podstawowych zalet takiego połączenia jest zmniejszenie zjawiska supresji jonów, dzięki czemu zwiększa się liczba wykrywanych substancji w porównaniu do sytuacji, w której analizowana próbka wprowadzana jest bezpośrednio z pompy strzykawkowej do źródła jonów bez rozdzielania

chromatograficznego zawartych w niej związków. Spektrometria mas mimo wykorzystania do analizy bardzo małej ilości analitu, zazwyczaj rzędu piko, a nawet femtogramów jest w stanie dostarczyć jednoczesnej informacji nie tylko na temat masy molekularnej związku, ale również jego budowy strukturalnej. Dodatkowo MS zapewnia najwyższą selektywność z uwagi na możliwość śledzenia w trakcie analizy procesu fragmentacji jonów poszczególnych związków. Dlatego też z uwagi na powyższe zalety połączenie LC-MS stanowi idealną technikę stosowaną zarówno do analiz ilościowych jak i jakościowych.

2.1.1. Wprowadzenie do LC-MS

Połączenie rozdziału chromatograficznego mieszaniny analizowanych związków z możliwościami spektrometrii mas do ich bezbłędnej identyfikacji posiada wiele zalet. LC-MS pozwala na rozróżnienie dwóch różnych związków posiadających ten sam czas retencji t_R , dzięki różnicom wynikającym z ich widm masowych. Unikalna specyficzność i selektywność techniki LC-MS pozwala uzyskać precyzyjne wyniki analiz ilościowych, co w przypadku zastosowania tylko chromatografii cieczowej byłoby niemożliwe. “Miękkie” techniki jonizacji (rozdział 1.2), które najczęściej są stosowane w połączeniach LC-MS zapewniają analizę bardzo licznej grupy związków, z uwzględnieniem zarówno metabolitów o niskich masach cząsteczkowych, leków (o masie poniżej 1000 Da) jak również wielkocząsteczkowych polimerów, których masa może być większa niż 10 000 Da. Spektrometr mas jest jedynym narzędziem analitycznym pozwalającym na jednoczesne określenie dokładnej masy związków jak i jego budowy strukturalnej, umożliwiając w ten sposób bezbłędą jego identyfikację. Dodatkowo wysoka selektywność i specyficzność jaką charakteryzuje się spektrometr mas pozwalają na przeprowadzenie dokładnych i precyzyjnych analiz ilościowych.

Pomimo wielu zalet połączenia chromatografii cieczowej ze spektrometrią mas istnieje wiele aspektów, które należy wziąć pod uwagę przed przystąpieniem do analiz z wykorzystaniem takiego systemu. Przykładowo różne związki posiadają odmienną zdolność do jonizacji w danych warunkach analizy, co determinuje różną czułość oznaczeń pomimo tych samych ilości związków zawartych w analizowanej próbce. Dodatkowym ograniczeniem dla analiz prowadzonych z użyciem połączenia LC-MS jest zjawisko zwane supresją jonów (ion suppression) wynikające z tłumienia sygnału pochodzącego od analitu na skutek obecności innych związków wymywanych z kolumny chromatograficznej w tym samym czasie, co w konsekwencji zmniejsza efektywność jonizacji analitu.

2.1.2. Teoretyczne podstawy chromatografii cieczowej

Chromatografię można rozpatrywać jako proces rozdziału związków zawartych w mieszaninie umożliwiający ich wykrycie oraz identyfikację. Identyfikacja opiera się o chromatograficzne parametry retencyjne. Według IUPAC - Międzynarodowej Unii Chemii Czystej i Stosowanej (The International Union of Pure and Applied Chemistry) Union of Pure and Applied Chemistry (IUPAC) pojęcie chromatografii definiowane jest następująco: “Chromatografia to fizyczna metoda rozdziału związków znajdujących się w dwóch różnych fazach, z których jedna jest nieruchoma (stanowi fazę stacjonarną), a druga (faza ruchoma) porusza się w określonym kierunku. Faza ruchoma opisywana jest jako “poruszająca faza przechodząca w określonym kierunku przez lub po powierzchni fazy stacjonarnej”. Faza ruchoma może być zarówno cieczą, gazem lub płynem nadkrytycznym, natomiast fazę stacjonarną może stanowić ciało stałe, żel bądź ciecz.

W skład typowego systemu chromatograficznego wchodzi cztery główne elementy: podajnik próbek, faza ruchoma, faza stacjonarna oraz detektor. Powszechnie wykorzystywanych jest kilka technik chromatograficznych, dlatego w zależności od ich rodzaju znaczenie poszczególnych elementów składowych systemu może być inne. Podstawowym zadaniem podajnika próbek jest wprowadzenie analitu do fazy ciekłej, będącej odpowiednim układem rozpuszczalników. Najczęściej stosowanym rodzajem podajnika próbek jest pętla iniekcyjna. Faza ruchoma i stacjonarna są kluczowymi elementami każdego układu chromatograficznego odpowiedzialnymi za proces rozdziału chromatograficznego. W systemach wysokosprawnej/wysokociśnieniowej chromatografii cieczowej (HPLC), faza ruchoma, którą stanowi rozpuszczalnik, bądź układ rozpuszczalników, przepływa w systemie ze stałą prędkością i pod wysokim ciśnieniem (zazwyczaj do 400 barów). Faza stacjonarna natomiast upakowana jest wewnątrz kolumny chromatograficznej w sposób umożliwiający wytrzymanie wysokich ciśnień.

2.1.2.1. Proces rozdziału chromatograficznego

Podczas analiz prowadzonych techniką HPLC wykorzystywane są różne mechanizmy retencyjne mają miejsce. Niewątpliwie podstawowe oddziaływania oparte są o polarność rozdzielanych związków. Do rozdzielania związków zawartych w mieszaninie dochodzi wtedy gdy wykazują one inne powinowactwo zarówno do fazy stacjonarnej jak i ruchomej. W takiej sytuacji czas jaki jest potrzebny na przebycie drogi w systemie od momentu nastrzyku do wykrycia przez detektor dla każdego związku będzie inny. Jeśli wszystkie związki obecne w próbce wykazują powinowactwo tylko do fazy ruchomej i nie oddziałują z fazą stacjonarną, nie dojdzie do ich rozdzielania ponieważ wszystkie w krótkim i tym samym czasie dotrą do detektora. Natomiast w sytuacji gdy wszystkie związki obecne w próbce silnie oddziałują z fazą stacjonarną, nie wykazując przy tym żadnego powinowactwa do fazy ruchomej, zostaną zatrzymane na kolumnie i nie dotrą do detektora. Z tego powodu wiedza dotycząca właściwości chemicznych analizowanych związków jest kluczowa w aspekcie przeprowadzenia ich rozdziału chromatograficznego. Jeśli właściwości te są dobrze scharakteryzowane, odpowiedni dobór warunków chromatograficznych np. właściwości zastosowanej fazy stacjonarnej bądź ruchomej umożliwi uzyskanie satysfakcjonującego rozdziału chromatograficznego. Efektywność rozdziału chromatograficznego jest zależna od kilku parametrów zawartych w tabeli 1.3.

Czas retencji (t_R), jest to czas potrzebny na wymycie analitu z kolumny do momentu jego pojawienia się w detektorze. Z uwagi na to, że długość kolumny oraz prędkość przepływu fazy ruchomej mają wpływ na t_R , został wprowadzony inny parametr zwany *współczynnikiem pojemnościowym (k)*. Parametr ten opisuje zależność czasu retencji analitu od specyficznego czasu jaki upływa do momentu pojawienia się w detektorze substancji nie oddziałującej z wypełnieniem kolumny (t_0). Czas t_0 nazywany jest często *czasem martwym kolumny*. Różnica pomiędzy t_R i t_0 nazywana jest z kolei *zredukowanym czasem retencji*. Dlatego zależność współczynnika pojemnościowego od zredukowanego czasu retencji opisana jest poniższym równaniem:

$$k = (t_R - t_0)/t_0$$

Należy zwrócić uwagę na to, że aby uzyskać wystarczającą rozdzielczość chromatograficzną wartości k rozdzielanych substancji powinny być w przedziale od 1 do 10. Efektywność rozdziału dwóch substancji A i B warunkuje *współczynnik podziału (α)*, który z kolei określany jest stosunkiem ich współczynników pojemnościowych k_B i k_A :

$$\alpha = k_B/k_A$$

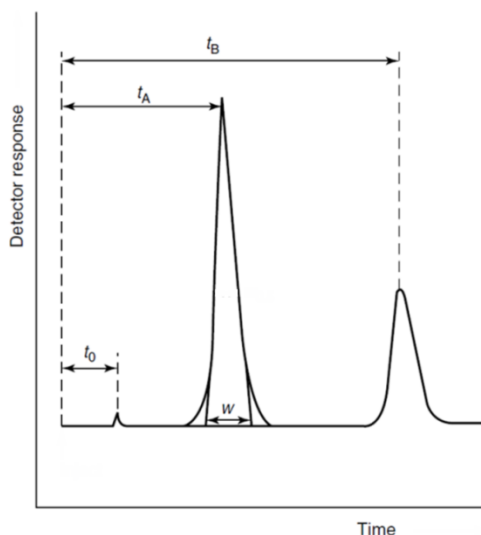
Tabela 1.3. Parametry warunkujące rozdział chromatograficzny.

Parametr	Symbol	Definicja
Czas retencji	t_R	czas jaki potrzebny jest na wymycie analitu do momentu pojawienia się w detektorze
Czas martwy	t_0	czas jaki upływa do momentu pojawienia się w detektorze substancji nie oddziałującej z wypełnieniem kolumny
Zredukowany czas retencji	$(t_R - t_0)$	różnica pomiędzy czasem retencji a czasem martwym
Współczynnik pojemnościowy	k	stosunek zredukowanego czasu retencji do czasu martwego
Selektywność (współczynnik podziału)	α	stosunek współczynników pojemnościowych rozdzielanych substancji
Rozdzielczość	R	charakteryzuje rozdzielenie dwóch pików chromatograficznych
Liczba pólek teoretycznych	N	liczba niewielkich pasm rozmieszczonych wzdłuż kolumny, w których dochodzi do równowagowego podziału substancji między fazę stacjonarną i ruchomą

Rozdział dwóch substancji ma miejsce wtedy gdy istnieje możliwość detekcji jednego z nich w obecności drugiego. Możliwość taką opisuje inny parametr chromatograficzny – *rozdzielczość*, opisany poniższym równaniem:

$$R = (t_B - t_A) / 0.5(w_A + w_B)$$

w którym w_A i w_B oznaczają szerokości u podstawy pików odpowiadających rozdzielanym substancjom A i B. Graficzne zilustrowanie parametrów chromatograficznych opisanych wyżej przedstawia rycina 1.10.



Rycina 1.10. Parametry rozdziału chromatograficznego w chromatografii cieczowej: t_0 , czas martwy; t_A i t_B , czasy retencji substancji A i B; w , szerokość pików.

Jeśli chodzi o kolumnę chromatograficzną, jej sprawność jest podstawowym parametrem decydującym o jej właściwościach chromatograficznych. Sprawność kolumny, znana także pod pojęciem efektywności zależy od *liczby pól teoretycznych* (N). Parametr ten z kolei odnosi się do liczby niewielkich pasm rozmieszczonych wzdłuż kolumny, w których dochodzi do stanu równowagi opisującego podziału substancji między fazę stacjonarną i ruchomą. Liczba pól zależy od długości kolumny (L), podczas gdy inny parametr *wysokość pulki* (H), definiowany jako efektywność kolumny przeliczaną na jednostkę długości od L nie zależy. Generalna zasada w chromatografii opiera się na zależności, że im mniejsza wysokość pulki tym lepsza rozdzielczość chromatograficzna jest uzyskiwana. Podobnie większa liczba pól teoretycznych jest również jak najbardziej pożądana. Zależność pomiędzy dwoma wspomnianymi parametrami opisuje wzór:

$$H = L/N$$

Właściwości fazy ruchomej również odgrywają istotną rolę w procesie rozdzielania chromatograficznego. Jeśli faza ruchoma zapewnia wystarczający rozdział związków polarnych, związki mniej polarne będą dłużej zatrzymywane na kolumnie i odwrotnie. Natomiast nie zawsze istnieje możliwość uzyskania dostatecznego rozdzielania z wykorzystaniem fazy ruchomej składającej się z jednego rozpuszczalnika. Zazwyczaj w takiej sytuacji należy zastosować układ dwóch lub więcej rozpuszczalników. Rozdzielenie związków uzyskane przy zastosowaniu stałego składu fazy ruchomej niezmiennego się podczas trwania całej analizy odnosi się do *elucji izokratycznej*. Natomiast, gdy skład fazy ruchomej ulega zmianie w trakcie analizy, mamy wówczas do czynienia z *elucją gradientową*. W przypadku elucji gradientowej ilość rozpuszczalnika odpowiedzialnego za lepsze wymywanie związków z kolumny zwiększa się w czasie trwania analizy.

Pomimo, możliwości zmiany omówionych wyżej parametrów chromatograficznych, nie zawsze udaje się uzyskać efektywny rozdział wszystkich składników zawartych w analizowanej próbce. Z tego powodu w zależności od charakteru analizowanych związków oraz polarności zarówno fazy stacjonarnej jak i ruchomej, w praktyce stosuje się kilka układów chromatograficznych scharakteryzowanych w tabeli 1.4.

Tabela 1.4. Porównanie różnych układów chromatografii ciekłej.

Układ LC	Faza ruchoma	Faza stacjonarna	Rodzaj rozdzielanych związków
NP	rozpuszczalniki organiczne: dichlorometan, octan etylu	krzemionka, aminowa, cyjankowa, diol	związki organiczne nie rozpuszczalne w wodzie
RP	woda/rozpuszczalnik organiczny	C18, C8, C4, cyjankowa, aminowa	związki niepolarne, słabe kwasy i zasady
HILIC	acetonitryl z wodą, dodatki substancji jonowych	polarna, czysta krzemionka	związki polarne
IEC	wodne roztwory buforowe,	anionowa lub kationowa, żywice jonowymiennie	związki jonowe, jony nieorganiczne

2.1.2.2. Chromatografia w normalnym układzie faz (NP)

Rozdział chromatograficzny w normalnym układzie faz uzyskiwany na skutek polarnych oddziaływań rozdzielnych związków z fazą stacjonarną, stanowiącej wypełnienie kolumny. Pojęcie “normalny” układ faz odnosi się do układu, w którym faza stacjonarna jest polarna, natomiast faza ruchoma jest niepolarna. Podczas procesu chromatograficznego w normalnym układzie faz, związki bardziej polarne są dłużej zatrzymywane na kolumnie, natomiast dzięki zwiększaniu polarności rozpuszczalników, używanych w roli fazy ruchomej, możliwe jest skrócenie czasów retencji rozdzielanych substancji. Im bardziej polarne rozpuszczalniki zostaną użyte do analizy, tym szybciej związki zostaną wymyte z kolumny.

Niemodyfikowana krzemionka z wolnymi grupami silanolowymi na jej powierzchni, jako grupami funkcyjnymi jest podstawowym wypełnieniem kolumn powszechnie stosowanym w chromatografii w układzie NP. Fazę ruchomą z kolei najczęściej stanowią takie rozpuszczalniki jak dichlorometan, octan etylu, lub inne nie mieszające się z wodą rozpuszczalniki. Z praktycznego punktu widzenia, głównym powodem stosowania chromatografii NP jest oddzielenie frakcji związków polarnych od związków niepolarnych, poprzez zwiększenie retencji polarnych substancji przy jednoczesnym wymyciu z kolumny związków hydrofobowych, które nie są na niej zatrzymywane. Należy również podkreślić, że chromatografia NP ma zastosowanie do rozdzielania izomerów geometrycznych oraz różniących się położeniem podstawników.

2.1.2.3. Chromatografia w odwróconym układzie faz (RP)

Chromatografia w odwróconym układzie faz (reversed-phase) stanowi najprawdopodobniej podstawę około 60% metod analitycznych wykorzystujących technikę HPLC. W układzie RP rozdział chromatograficzny ma miejsce dzięki oddziaływaniom związków zawartych w próbce z niepolarną fazą stacjonarną oraz polarną fazą ruchomą. Faza ruchoma jest znacznie bardziej polarna w porównaniu do fazy stacjonarnej, dlatego z kolumny chromatograficznej związki polarne wymywane są w pierwszej kolejności, a następnie związki mniej polarne. W chromatografii w układzie RP rozdział chromatograficzny oparty jest o różnice w hydrofobowości obecnych w analizowanej mieszaninie składników.

Istnieje kilka rodzajów faz stacjonarnych wykorzystywanych w wykosprawnej chromatografii cieczowej RP. Z reguły wypełnienie kolumn chromatograficznych RP stanowi krzemionka, na powierzchni której są związane kowalencyjnie długołańcuchowe węglowodory. Najbardziej uniwersalną i zarazem najbardziej popularną fazą stacjonarną jest faza zawierająca grupy oktadecysilanowe (C18 lub ODS). Natomiast typową fazą ruchomą stosowaną w chromatografii RP jest układ dwóch rozpuszczalników, z czego jeden stanowi woda, a drugi rozpuszczalnik organiczny, w roli którego najczęściej stosowany jest metanol lub acetonitryl. Chromatografia cieczowa w układzie RP jest uniwersalną techniką z uwagi na fakt, że umożliwia podczas jednej analizy jednoczesny rozdział niepolarnych związków i substancji o charakterze jonowym.

Jeśli chodzi o analizę próbek biologicznych chromatografia NP i RP wykorzystywane są w odmienny sposób. Przykładowo rozdział fosfolipidów na klasy przy użyciu kolumn NP, możliwy jest z uwagi na obecność w obrębie ich struktury tzw. „głów” polarnych (polar head), podczas gdy różnica w hydrofobowości ich łańcuchów kwasów tłuszczowych, determinuje ich rozdział przy zastosowaniu kolumn RP. Ponadto gradientowa chromatografia

cieczowa w układzie RP jest z powodzeniem wykorzystywana w metabolomice do profilowania metabolitów.

2.1.2.4. Chromatografia oddziaływań hydrofilowych (HILIC)

Chromatografia oddziaływań hydrofilowych (HILIC) (hydrophilic interaction chromatography) jest rozwinięciem chromatografii w układzie faz normalnych. W HILIC stosuje się polarne wypełnienia kolumn chromatograficznych oraz układ rozpuszczalników typowy dla chromatografii RP, najczęściej acetonitryl z niewielką ilością wody. Do fazy ruchomej dodawane są sole octanu amonu lub mrówczanu amonu w celu zwiększenia jej polarności oraz siły jonowej, natomiast wypełnienie kolumn chromatograficznych typu HILIC stanowi czysta krzemionka, lub krzemionka ze związanymi na jej powierzchni grupami aminowymi bądź amidowymi.

Mechanizm rozdziału w chromatografii HILIC jest złożony i oparty o zjawisko ekstrakcji w układzie rozpuszczalnik - rozpuszczalnik na granicy faz wytworzonej pomiędzy warstwą wody na powierzchni polarnego wypełnienia a rozpuszczalnikiem organicznym fazy ruchomej. Rozdział chromatograficzny z użyciem techniki HILIC polega na tym, że związki polarne zatrzymywane są na kolumnie, podczas gdy związki mniej polarne obecne są przede wszystkim w fazie ruchomej. Zastosowanie rozpuszczalników organicznych w chromatografii typu HILIC, jest korzystne w aspekcie jej połączenia ze spektrometrią mas. Rozpuszczalniki organiczne obecne w fazie ruchomej dodatkowo zwiększają czułość analiz MS dzięki ograniczeniu zjawiska supresji jonów.

2.1.2.5. Chromatografia jonowymienna

Chromatografia jonowymienna (IEC) (ion-exchange chromatography) zapewnia rozdział jonów lub substancji wykazujących zdolność do przechodzenia w jony. W chromatografii tej jako faza stacjonarna wykorzystywane są materiały zawierające jonowe grupy funkcyjne, obdarzone przeciwnym ładunkiem aniżeli jony rozdzielanych substancji. Dlatego anionowe wypełnienie kolumny chromatograficznej stosowane jest przy rozdziale mieszaniny kationów lub związków naładowanych dodatnio, podczas gdy wypełnienie kationowe zapewnia rozdział mieszaniny anionów lub związków naładowanych ujemnie. W chromatografii IEC faza ruchoma jest wysoce polarna, dlatego najczęściej stanowi ją woda z dodatkiem soli lub wodny roztwór buforowy. Zwiększanie siły jonowej fazy ruchomej jest podstawowym czynnikiem odpowiedzialnym za wymywanie rozdzielanych substancji z kolumny.

Chromatografia jonowymienna jest stosowana do rozdziału zarówno drobnocząsteczkowych jak i nieco większych związków o znaczeniu biologicznym takich jak aminokwasy, kwasy karboksylowe czy aminy. Niestety należy pamiętać o tym, że z uwagi na zjawisko supresji jonów skutecznie wspomagane dużą mocą jonową fazy ruchomej IEC, bezpośrednie połączenie tej techniki separacyjnej ze spektrometrią mas jest trudne do realizacji.

2.1.3. Moduły systemu LC-MS

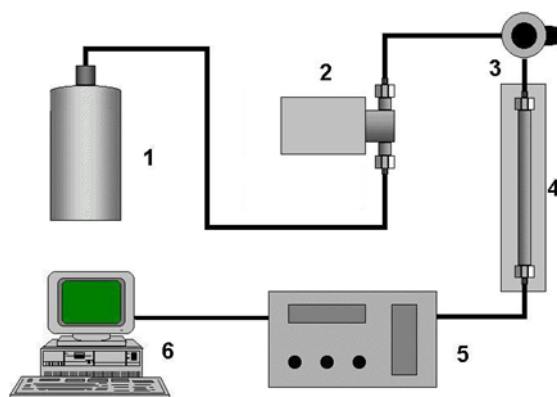
Schemat blokowy typowego systemu chromatografii cieczowej przedstawiono na rycinie 1.11. Zawiera on następujące moduły:

a) Pompa

Głównym zadaniem pompy, jako serca każdego systemu chromatograficznego, jest zapewnienie stabilnego przepływu fazy ruchomej, którego szybkość warunkują parametry

kolumny chromatograficznej, rodzaj źródła jonizacji oraz typ spektrometru mas tworzących razem połączanie LC-MS. Przykładowo prędkość fazy ruchomej rzędu 0.2 – 0.4 mL/min jest zazwyczaj stosowana w systemach wyposażonych w źródło ESI i kolumnę UPLC o średnicy 2.1 mm, podczas gdy szybkość przepływu 1 ml/min. jest preferowana w przypadku i źródła APCI i konwencjonalnej kolumny HPLC o średnicy 4.6 mm.

W przypadku połączenia LC-MS istotne jest, aby pompa nie generowała żadnej pulsacji, co utrudniałoby odczyt i interpretację wyników otrzymywanych w postaci chromatogramów. Ponadto, powstawaniu pęcherzyków powietrza w fazie ruchomej zapobiega degazer, stanowiący dodatkowy moduł chromatografu cieczowego każdego systemu LC-MS.



Rycina 1.11. Blokowy diagram chromatografu cieczowego (1 – Zbiornik fazy ruchomej, 2 - pompa, 3 – automatyczny podajnik próbek (autosampler), 4 – kolumna chromatograficzna, 5 - detektor, 6 – rejestrator/komputer).

b) Automatyczny podajnik próbek (autosampler)

Obecnie w chromatografach cieczowych stosowany jest jeden typ podajnika próbek, w postaci iniekcyjnej lub zaworu dozującego (*loop injector* lub *valve injector*). Proces dozowania polega na tym, że próbka przy użyciu strzykawki wprowadzana jest do strumienia fazy ruchomej wypełniając jednocześnie pętlę dozującą wykalibrowaną na dokładnie określoną objętość. Kiedy pętla jest wypełniona, faza ruchoma pomija ją i jest pompowana przez zawór do kolumny w celu jej zrównoważenia. W chwili dozowania (nastrzyku) faza ruchoma przepływa przez pętlę, wymywając jej zawartość i przenosząc próbkę na kolumnę. Parametry, którymi powinien się charakteryzować każdy podajnik próbek to duża powtarzalność, dokładność dozowania oraz konstrukcja zapobiegająca powstawaniu pęcherzyków powietrza w jego układzie.

c) Kolumny

Kolumny chromatograficzne stosowane w HPLC różnią się między sobą przede wszystkim średnicą, długością oraz rozmiarem ziaren fazy stacjonarnej. Typowe kolumny HPLC są zdolne wytrzymać ciśnienie fazy ruchomej przepływającej przez wypełnienie rzędu 400 barów. Ich średnica wynosi 4.6 mm, długość od 100 do 250 mm, natomiast rozmiar ziaren zawiera się w przedziale od 3 do 5 μm . Skonstruowanie i wprowadzenie na rynek pomp wykorzystywanych w ultrasprawnej chromatografii cieczowej (UPLC), zdolnych do pracy pod ciśnieniem 1200 barów, pozwoliło na zmniejszenie średnicy ziaren fazy stacjonarnej, co z kolei znacznie zwiększyło możliwości rozdzielcze kolumn UPLC w porównaniu z tradycyjnymi kolumnami HPLC. Dodatkową korzyścią, wynikającą ze

zmniejszenia wielkości ziarna, jest zmniejszenie szybkości przepływu fazy ruchomej, co w praktyce przekłada się na zwiększenie czułość oznaczeń, oraz obniżenie limitu detekcji, zwłaszcza w przypadku systemów nano-LC, dla których już śladowe ilości analitu są wystarczające do przeprowadzenia analizy. Znanych jest wiele typów i rodzajów kolumn, których charakterystykę przedstawia tabela 1.5.

Tabela 1.5. Rodzaje kolumn HPLC i ich nomenklatura.

Typ kolumny	Średnica wewnętrzna (i.d.)	Optymalna prędkość przepływu
nano-HPLC (z ang. <i>nanobore</i>)	0.075 mm	0.3 µL/min.
kapilarna	0.5 mm 1 mm	10 µL/min. 50 µL/min.
mikrokolumna	2.1 mm	250 µL/min.
HPLC o zmniejszonym ziarnie (z ang. <i>narrow(small)-bore column</i>)	3 mm	500 µL/min.
HPLC	4.6 mm	1250 µL/min.

Literatura:

G. Rozing. Trends in HPLC column formats - microbore, nanobore and smaller. *LC GC EUROPE*, 2003 16.6A:14-19.

S.R. Wilson, T. Vehus, H.S. Berg, E. Lundanes. Nano-LC in proteomics: recent advances and approaches. *Bioanalysis*, 2015, 7(14):1799-1815.

N. Gray, M.R. Lewis, R.S. Plumb, I.D. Wilson, J.K. Nicholson. High-throughput microbore UPLC-MS metabolic phenotyping of urine for large-scale epidemiology studies. *J. Proteome Res.*, 2015, 14(6):2714-2721.

d) Detektory

Detektory stosowane w chromatografii cieczowej charakteryzuje kilka parametrów takich jak selektywność, czułość oraz limit detekcji. *Selektywność* definiowana jest jako zdolność detektora do wykrycia substancji przy wyeliminowaniu sygnału pochodzącego od rozpuszczalników oraz innych substancji zawartych w analizowanej próbce. *Czułość*, ten termin często stosowany jest zamiennie z *limitem detekcji*, który określa najmniejszą ilość substancji jaką detektor jest w stanie wykryć w danych warunkach analizy oraz wystarczającą na jej ilościowe oznaczenie. Natomiast czułość charakteryzuje wielkość sygnału, jako odpowiedź detektora w chwili, gdy oznaczana substancja zostanie w nim wykryta. Opisanie parametry charakteryzują każdą metodę analityczną. Należy pamiętać o tym, że wyższa czułość w praktyce nie przekłada się na obniżenie limitu detekcji, czy też granicy wykrywalności, ponieważ sygnały pochodzące od zanieczyszczeń obecnych w próbce również charakteryzować się będą wyższą intensywnością.

W chromatografii cieczowej są stosowane różne rodzaje detektorów wliczając w to detektor absorpcyjny UV, fluorescencyjny, elektrochemiczny, konduktometryczny, refraktometryczny czy detektor masowy. Z analitycznego punktu widzenia każdy z wymienionych detektorów posiada swoje specyficzne zalety oraz wady, dlatego obecnie najczęściej w połączeniu z chromatografem cieczowym stosowany jest detektor masowy z detektorem UV jako pomocniczym. Stosowanie tylko detektora UV ma sens jedynie wtedy,

gdy długość fali odpowiadająca maksimum absorpcji analizowanego związku (λ_{\max}) jest znana. Większość związków organicznych absorbuje światło przy tej samej długości fali, dlatego detektor UV charakteryzuje niską selektywność. Niewątpliwie spektrometria mas, jest techniką zapewniającą kompletną informację analityczną, zwłaszcza w przypadku analizy próbek o złożonej matrycy. Detekcja MS wymaga minimalnej ilości analitu, rzędu pikogramów, aby dostarczyć dokładną informację na temat jego masy oraz struktury. Spektrometria mas stanowi zatem idealną technikę zarówno do ilościowych jak i jakościowych oznaczeń, dlatego w chwili obecnej detektor MS jest detektorem z wyboru stosowanym w połączeniu z chromatografią cieczową.

2.2. Chromatografia gazowa – spektrometria mas (GC-MS)

2.2.1. Wprowadzenie do GC-MS

Chromatografia gazowa sprzężona ze spektrometrią mas stanowi połączenie dwóch zaawansowanych technik instrumentalnych. Często wykorzystywana jest w analizie związków organicznych i uznawana jest za najskuteczniejszą metodę identyfikacji związków organicznych w złożonych matrycach, w szerokim zakresie stężeń. Jest to szczególnie korzystne w przypadku analizy związków o małej masie cząsteczkowej oraz o średniej lub niskiej polarności, o temperaturze wrzenia poniżej 350-400⁰C i występujących na poziomie ppb-ppm. Rozwój w technologii wypełnień kolumn chromatograficznych zmierza obecnie do tego, aby możliwe było wykorzystanie GC do analizy związków wysokowrzących (o temperaturze wrzenia powyżej 500⁰C). Możliwości analityczne GC-MS w dużym stopniu zależą jednak od spektrometrii mas. Wraz ze zmniejszającym się poziomem wykrywalności i rosnącą liczbą poszukiwanych związków, analiza danych może być kłopotliwa i wymaga dalszego rozwoju. Technika ta charakteryzuje się ponadto wysoką selektywnością i bardzo wysoką czułością, zapewniając szeroki zakres zastosowań. Jest to metoda ważna do monitorowania zarówno w analizach środowiskowych, jak i biochemii. Metabolomiczne podejście oparte na GC-MS jest szeroko stosowane w celu zrozumienia procesów prowadzących do rozwoju chorób takich jak m.in. nowotwory czy choroby układu krążenia.

2.2.2. Podstawy teoretyczne GC-MS

2.2.2.1. Proces chromatograficzny

Chromatografia gazowa wykorzystywana jest do oznaczania składu próbek zarówno gazowych jak i ciekłych, ale również substancji zaadsorbowanych na powierzchni sorbentu (np. mikroekstrakcja do fazy stałej (SPME)). Zasada rozdziału składników w GC polega na przeniesieniu substancji (w postaci pary) w strumieniu gazu nośnego (faza ruchoma) przez kolumnę kapilarną pokrytą wewnątrz cienką warstwą polimerowej fazy stacjonarnej. Podczas wprowadzania analizowanej mieszaniny do GC, w dozowniku na skutek działania wysokiej temperatury (200-300⁰C) składniki próbki ulegają odparowaniu i po zmieszaniu z gazem nośnym (Ar, He, N₂ lub H₂) przenoszone są na kolumnę chromatograficzną.

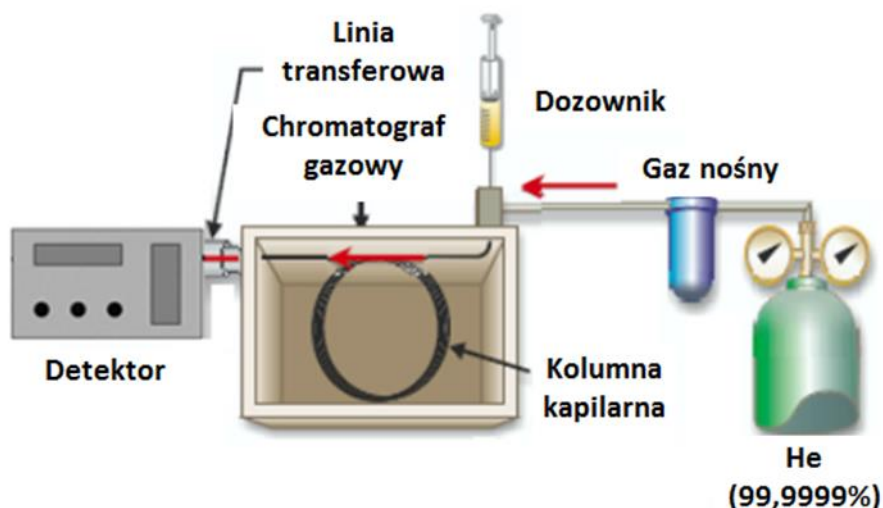
Na kolumnie anality ulegają rozdzielaniu pomiędzy fazę ruchomą (strumień gazu nośnego) a fazę stacjonarną (powłoka polimerowa) w zależności, głównie od ich struktury chemicznej. Prędkość przepływu analitów w kolumnie jest funkcją ich podziału pomiędzy dwie fazy pozostające w równowadze, przy czym składniki mające większe powinowactwo do fazy stacjonarnej przemieszczają się wolniej od składników mających większe

powinowactwo do fazy ruchomej. Rozdział składników jest zatem wynikiem różnego tempa migracji spowodowanej różnymi wartościami współczynnika podziału K_S . Współczynnik podziału można wyrazić równaniem Nernsta:

$$K_S = C_L/C_G$$

gdzie C_L i C_G wyrażają stężenie analitu w fazie ciekłej (stacjonarnej) i gazowej (ruchomej). Wyższe powinowactwo do fazy stacjonarnej oznacza wyższą wartość współczynnika K_S , co skutkuje późniejszym opuszczeniem kolumny, czyli dłuższym czasem retencji (t_R).

Składniki próbki rozdzielone na kolumnie, kolejno docierają do detektora, generując sygnał elektryczny, który jest proporcjonalny do ilości związku. Tradycyjne detektory stosowane w GC reagują na różnice w fizykochemicznych właściwościach gazu nośnego i gazu, w którym obecna jest eluowana z kolumny substancja, np. jonizacja w płomieniu, wychwytywanie elektronów lub przewodność cieplna. Zarejestrowane zmiany mogą być proporcjonalne do stężenia lub do natężenia przepływu masowego analitu w strumieniu gazu nośnego. Ważną cechą detektora jest wysoka czułość, wykrywalność (granica wykrywalności), wysoka stabilność sygnału, powtarzalny sygnał linii podstawowej i szeroki zakres liniowości. Otrzymane dane są przetwarzane w celu wygenerowania wykresu zmienności sygnału detektora w czasie (chromatogram). Schemat systemu GC, ilustrujący jego główne składniki, przedstawiono na rycinie 1.12.



Rycina 1.12. Schemat chromatografu gazowego (*de.leco-europe.com w modyfikacji własnej*)

2.2.2.2. Kolumny chromatograficzne

Odpowiedni dobór kolumny ma decydujący wpływ na jakość rozdziału składników mieszaniny, co jest wynikiem analizy chromatograficznej. W GC stosowane są dwa typy kolumn: pakowane (analityczne, mikropakowane) i o przekroju otwartym, często nazywane kapilarami (kapilarne, mikrokapilarne). Kolumny pakowane napełniane są fazą stacjonarną, a w przypadku kolumn kapilarnych faza stacjonarna osadzona jest na ich wewnętrznych ściankach. W kolumnach pakowanych o rozdziale chromatograficznym składników próbki decydują cząstki adsorbentu (m.in.: adsorbenty węglowe, krzemionka, tlenek glinu, sita molekularne, porowate polimery syntetyczne), natomiast w przypadku kolumn kapilarnych – ciekła faza stacjonarna (m.in.: silikony, skwalen, glikol polietylenowy).

Chromatografia gazowa adsorpcyjna (kolumny pakowane) jest obecnie rzadziej stosowana niż podziałowa (kolumny kapilarne), ze względu na niższą powtarzalność uzyskiwanych wyników, dłuższe czasy retencji oraz wygląd uzyskiwanych pików (piki "ogonujące") oraz znacznie niższą efektywność separacji. Są one głównie wykorzystywane do analizy gazów i substancji o małych masach cząsteczkowych i dużej lotności. Częściej stosowane są kolumny kapilarne wypełnione ciekłą fazą stacjonarną. Ciekłe fazy stacjonarne powinny być obojętne chemicznie względem rozdzielanych składników, zdolne do ich rozpuszczania, wysoce selektywne względem składników mieszaniny, powinny mieć niską lotność i wysoką stabilność termiczną w warunkach pracy kolumny GC. Są one szczególnie przydatne do rozdzielenia składników mało lotnych i wielkocząsteczkowych.

Najważniejszą cechą ciekłych faz stacjonarnych decydującą o ich właściwościach rozdzielczych jest polarność. Efektywność rozdzielenia na kolumnie kapilarnej zależy również od wymiarów kolumny (tj. długość-L i średnica wewnętrzna-ID) oraz grubości fazy stacjonarnej. Standardowa długość handlowo dostępnych kolumn kapilarnych wynosi 15, 30 i 60 m, natomiast średnica: 0,15; 0,25; 0,32 i 0,53 mm.

Polarność rozdzielanych związków decyduje o wyborze ciekłej fazy stacjonarnej. Zgodnie z ogólną regułą, w przypadku rozdzielania związków niepolarnych, należy stosować fazę niepolarną, a do rozdzielania związków polarnych - fazę o wysokiej polarności. Na fazach niepolarnych (tj. skwalen, 100% dimetylopolisiloksan, (5%-fenylo)-dimetylopolisiloksan), wymywanie analitów odbywa się zgodnie z ich lotnością, tj. wraz ze wzrostem ich temperatury wrzenia. W przypadku kolumn polarnych (np. glikol polietylenowy (Carbowax), (70% cyjanopropylo)-fenylometylopolisiloksan) czynnikiem determinującym rozdzielanie analitów jest ich polarność (moment dipolowy - μ) i siła oddziaływań z fazą stacjonarną (dipol, wiązania wodorowe). Polarne kolumny, takie jak np. cyjanopropylofenylometylopolisiloksan charakteryzują się również wysoką selektywnością, co jest związane z wysoką zawartością grup fenolowych i cyjanowych (Tabela 1.6). Do niektórych celów mogą być stosowane inne fazy, tj. np.: alkohole, sacharydy, związki nitrowe, aminy, sole kwasów nieorganicznych (np. azotan srebra). Istnieją także fazy, których mechanizm działania rozdzielczego obejmuje różne inne oddziaływania analitów z poszczególnymi cząstkami fazy stacjonarnej lub uporządkowaną strukturą, uzyskując znaczną selektywność rozdzielenia. Należą do nich m.in. chiralne fazy stacjonarne, fazy cyklodekstrynowe i ciekłokrystaliczne.

Do niektórych celów mogą być stosowane inne fazy, tj. np.: alkohole, sacharydy, związki nitrowe, aminy, sole kwasów nieorganicznych (np. azotan srebra). Istnieją także fazy, których mechanizm działania rozdzielczego obejmuje różne inne oddziaływania analitów z poszczególnymi cząstkami fazy stacjonarnej lub uporządkowaną strukturą, uzyskując znaczną selektywność rozdzielenia. Należą do nich m.in. chiralne fazy stacjonarne, fazy cyklodekstrynowe i ciekłokrystaliczne.

2.2.2.3. GC×GC/MS

Wśród technik chromatograficznych chromatografia gazowa odgrywa ważną rolę ze względu m.in. na wysoką rozdzielczość, elastyczność i szeroki zakres zastosowań. Zawodzi jednak w analizie bardzo złożonych próbek ze względu na ograniczoną pojemność pików (liczba pików chromatograficznych, które mogą być całkowicie rozdzielone przy użyciu danej kolumny chromatograficznej). Podczas migracji w kolumnie pasma chromatograficzne ulegają poszerzeniu. Skutkiem tego jest to, że ilość pasm, które mogą być całkowicie rozdzielone na wylocie kolumny jest ograniczona. Ograniczenia tego nie da się uniknąć

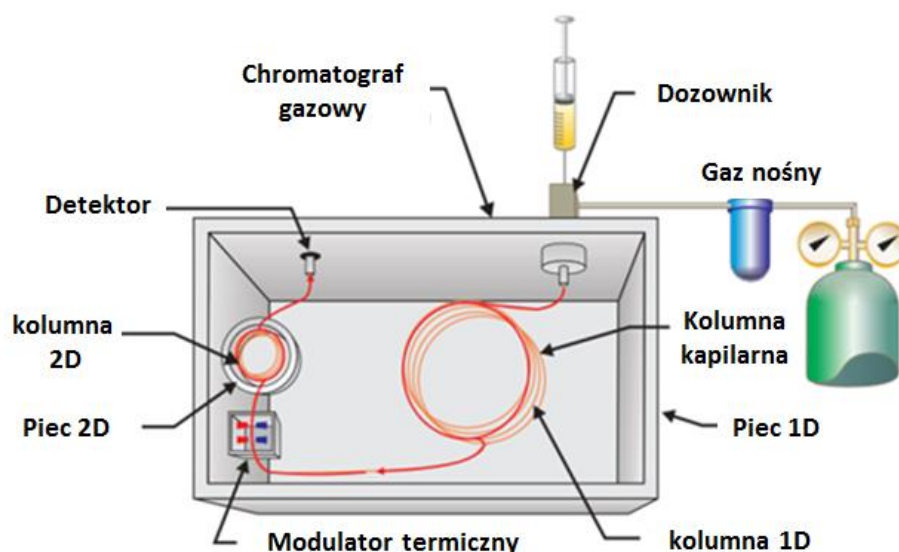
poprzez modyfikację parametrów chromatograficznych a jedynym skutecznym rozwiązaniem jest użycie dwuwymiarowego rozdzielania chromatograficznego (GC×GC), w którym anality częściowo rozdzielone na pierwszej kolumnie ponownie ulegają rozdzielaniu, na podstawie innego mechanizmu, na drugiej kolumnie GC.

Tabela 1.6. Klasyfikacja faz stacjonarnych w GC.

	FAZA STACJONARNA	ZASTOSOWANIE	TEMP [°C]
Wzrost polarności	dimetylopolisiloksan	niska selektywność rozdzielenie związków na podstawie ich temperatur wrzenia (lotności), bardzo wysoka odporność termiczna, zalecana do analiz środowiskowych, analizy śladów, metody EPA, pestycydów, PCB, analizy żywności i leków, najczęściej używana faza w GC	50-325
	5% fenylo- dimetylopolisiloksan	nieznacznie wyższa selektywność od kolumn dimetylopolisiloksanowych bardzo wysoka odporność termiczna, doskonała obojętność wobec związków zasadowych, standardowa faza z dużym zakresem zastosowań: analizy środowiskowe, analiza śladów, metody EPA, analiza pestycydów, PCB, żywności i leków,	50-325
	50% fenylo- dimetylopolisiloksan	wyższa selektywność od kolumn dimetylopolisiloksanowych, dobra odporność termiczna, przeznaczona do oznaczeń biomedycznych tj. analiza steroidów, cukrów standardowa faza do analiz środowiskowych, analizy ultra-śladów, metody EPA, analizy pestycydów, PCB, żywności i leków,	40-325
	6% cyjanopropylo- 94% fenylo- dimetylopolisiloksan	selektywność nominalna dla związków spolaryzowanych i polarnych, dobra odporność termiczna, faza ogólnego zastosowania, ale szczególnie nadaje się do zaawansowanych analiz środowiskowych (np. metody EPA dla WWA, PCB i pestycydów)	20-240
	7% cyjanopropylo- 7% fenylo- dimetylopolisiloksan	wysoka selektywność, dobra odporność termiczna, przeznaczona do analizy pochodnych cukrów i analiz środowiskowych rekomendowana do analizy alkoholi, PCB	40-280
	25% cyjanopropylo- 25% fenylo- dimetylopolisiloksan	wysoka selektywność, faza polarna, dobra odporność termiczna, dedykowana do analizy związków polarnych (np. kwasy tłuszczowe, sterole, cukry w postaci octanów alditoli)	40-240
	glikol polietylenowy	unikalna selektywność dla cząsteczek zdolnych do tworzenia wiązań wodorowych, ograniczona stabilność termiczna, dedykowana do analizy próbek utlenionych, podatnych na degradację pod wpływem tlenu, zalecana do analizy rozpuszczalników i alkoholi, jak również roztworów wodnych, nie zalecana do analizy pochodnych sililowych	20-260
	>70% cyjanopropylo- dimetylopolisiloksan	wysoka selektywność, faza polarna, dobra odporność termiczna, rekomendowana do analizy złożonych prób środowiskowych oraz FAMES	20-260
	>90% cyjanopropylo- polisiloksan	wysoka selektywność, faza o wysokiej polarności, dobra odporność termiczna, dedykowana do analizy FAMES, dioksyn, furanów i szybkiego rozdzielania związków aromatycznych, perfum, produktów petrochemicznych i innych związków, które trudno rozdzielić za pomocą konwencjonalnych kolumn	40-280

Od 1991 roku, w którym zaproponowane zostało pierwsze praktyczne rozwiązanie problemu kompletnej dwuwymiarowej chromatografii gazowej, nastąpił gwałtowny rozwój techniki GC×GC. Obecnie GC×GC jest powszechnie stosowana w wielu różnych dziedzinach, obejmując szeroki zakres zastosowań. Kompletna dwuwymiarowa chromatografia gazowa sprzężona ze spektrometrią mas (GC×GC/MS) jest techniką, która zapewnia rozdział i pełny profil rozkładu mas rozdzielonych związków, w przestrzeni dwuwymiarowej. Mechanizmy rozdziału w obu wymiarach powinny być całkowicie niezależne od siebie, tzn. ortogonalne. GC×GC posiada wszystkie zalety jednowymiarowej ("normalnej") chromatografii gazowej: jest łatwa do zautomatyzowania i zapewnia analizę wykrytych związków z wysoką czułością. Ponadto, rozdział w 2D daje dużo większą ilość informacji, co pozwala na łatwiejszą i bardziej pewną identyfikację nieznanymi substancji a piki związków należących do danego szeregu homologicznego zazwyczaj układają się wzdłuż linii prostych na dwuwymiarowej płaszczyźnie retencji, co pozwala na przyporządkowanie nieznanego związku do określonej klasy chemicznej tylko na podstawie jego położenia na płaszczyźnie retencji.

Początkowo składniki wprowadzonej próbki ulegają rozdzielaniu na kolumnie kapilarnej o wysokiej rozdzielczości (1D), podobnie jak w klasycznej chromatografii gazowej. Po opuszczeniu pierwszej kolumny eluent nie jest kierowany bezpośrednio do detektora, ale do modulatora, gdzie ulega koncentracji w postaci wąskich frakcji, w regularnych, krótkich odstępach czasu, krótszych od pierwotnej szerokości piku (tzw. okres modulacji). Następnie, zebrany materiał nastrzykiwany jest do drugiej kolumny w celu szybkiego rozdzielania składników w drugim wymiarze (2D) (rycina 1.13). Po nastrzyknięciu próbki, w modulatorze „zbierana” jest kolejna frakcja. Jednocześnie, na drugiej kolumnie rozdzielane są składniki poprzedniej frakcji. Mechanizmy rozdziału na obu kolumnach są od siebie niezależne. Dzięki temu możliwe jest rozdzielanie składników, które koeluowały ze sobą na wylocie pierwszej kolumny. Związki opuszczające drugą kolumnę kierowane są do detektora, który rejestruje szereg bardzo krótkich, np. 3 sekundowych chromatogramów. Proces zbierania frakcji i dozowania próbki do kolumny 2D powtarza się do końca analizy. Pozwala to na zachowanie rozdzielania pasm w kolumnie 1D i umożliwia uzyskanie dodatkowego rozdzielania w kolumnie 2D. Możliwe jest to dzięki zastosowaniu odpowiednio wysokiej częstotliwości próbkowania.

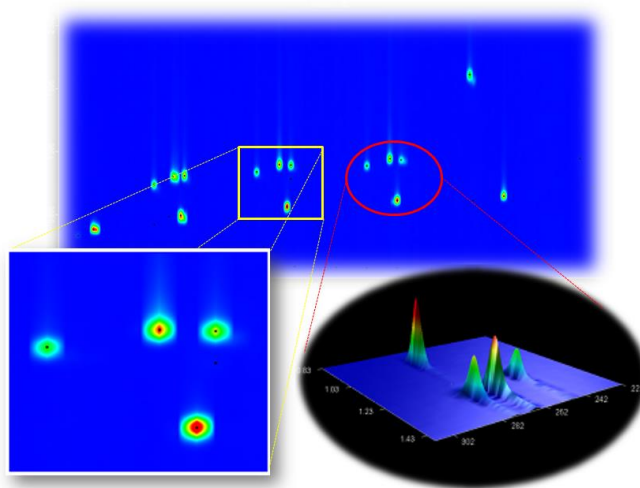


Rycina 1.13. Schemat system do GC×GC (*de.leco-europe.com*)

Kolumny chromatograficzne stosowane w obu wymiarach powinny charakteryzować się różnymi mechanizmami retencji a także różną selektywnością. Zastosowanie kombinacji kolumny niepolarniej i polarnej (NP/P) determinuje rozdzielanie związków mających taką samą temperaturę wrzenia na podstawie różnicy ich polarności. Rozdzielenie w wymiarze 2D uzyskuje się dzięki różnicom we współczynnikach aktywności substancji rozpuszczonych a fazą polarną. Umożliwia to rozdzielanie związków o różnej polarności. Poza zwiększoną pojemnością pików, zaletą GC×GC jest także dodatkowy wymiar separacji, co wpływa na zmniejszenie przeciążenia eluowanych pików, co pozwala na rozdzielanie związków na niższych poziomach stężeń.

Chromatogramy otrzymane przy użyciu GC×GC (2D) zawierają znacznie więcej informacji niż GC (1D), co umożliwia łatwiejszą i bardziej pewną identyfikację nieznanymi substancji. W konsekwencji wymagane jest zastosowanie bardzo sprawnego detektora, który powinien mieć możliwość zbierania danych z częstotliwością co najmniej 50 Hz. Aby dokładnie odtworzyć kształt pików, powinien on być próbkowany co najmniej 10 razy, a piki wymywane z drugiej kolumny mają bardzo małą szerokość u podstawy, od 150 do 400 ms. Wybór detektorów chromatograficznych, które mogłyby spełniać to kryterium jest bardzo ograniczony. Najczęściej stosowanymi detektorami w chromatografii GC×GC są: FID (maks. 300 Hz), μ -ECD (50-100 Hz) i TOFMS (maks. 500 Hz).

Podsumowując, płaszczyzna retencji ma większą pojemność pikową niż oś retencji, a zatem może pomieścić i rozdzielić bardzo złożone mieszaniny związków, m.in. w próbkach biologicznych o złożonym profilu czy też profile metabolitów. Ogólnie rzecz biorąc, możliwości wykorzystania GC×GC można podzielić na trzy obszary: „fingerprints” bardzo złożonych matryc, analizy targetowe oraz identyfikacja nieznanymi związków. Przykład chromatogramu 2D przedstawiono na rycinie 1.14.



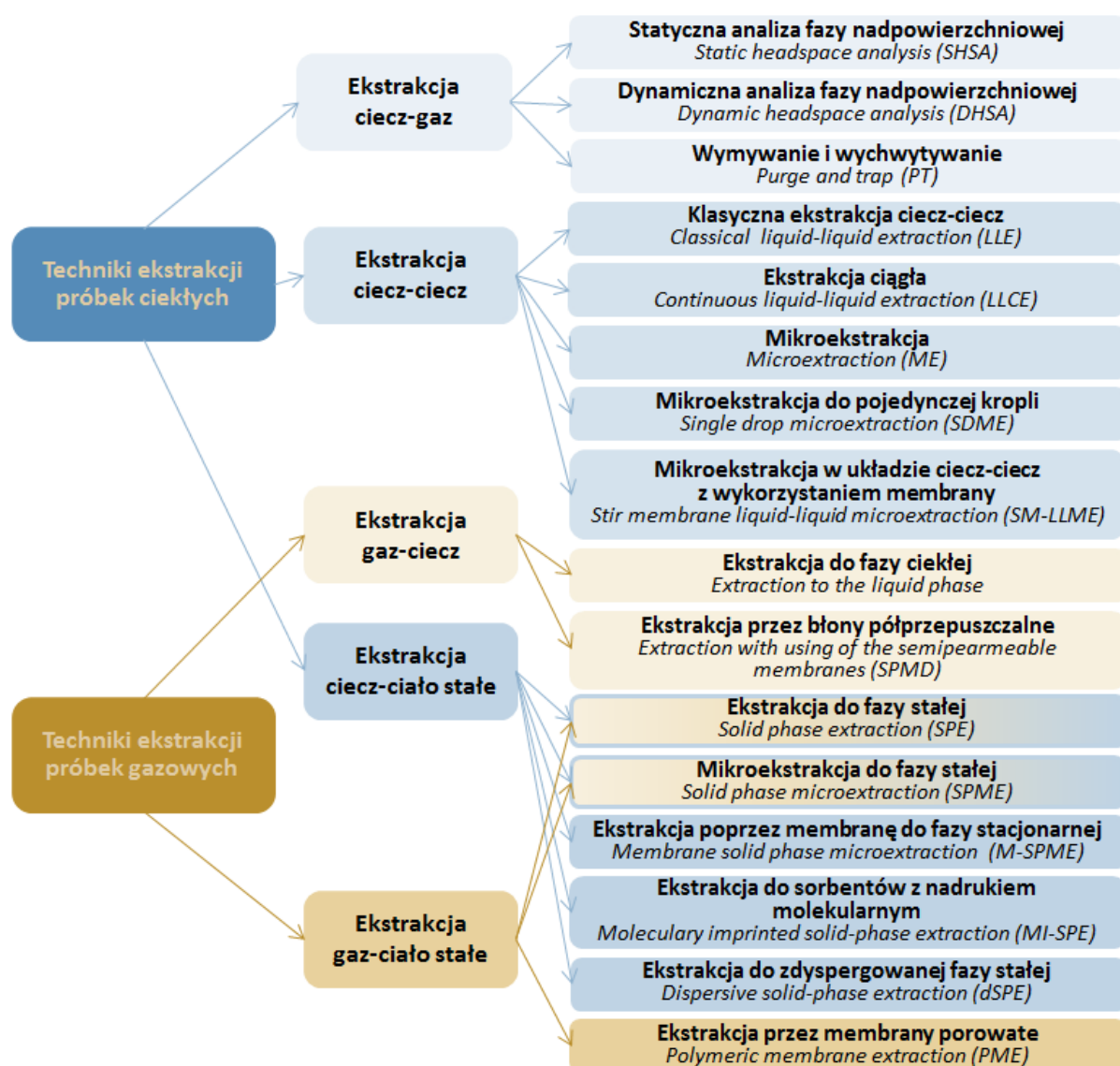
Rycina 1.14. Chromatogram 2D GC mieszaniny związków lotnych (alkany, alkohole, aldehydy, ketony, amina) [Pegasus 4D; Leco].

2.2.2.4. Przygotowanie próbek do analizy chromatograficznej. Derywatyżacja.

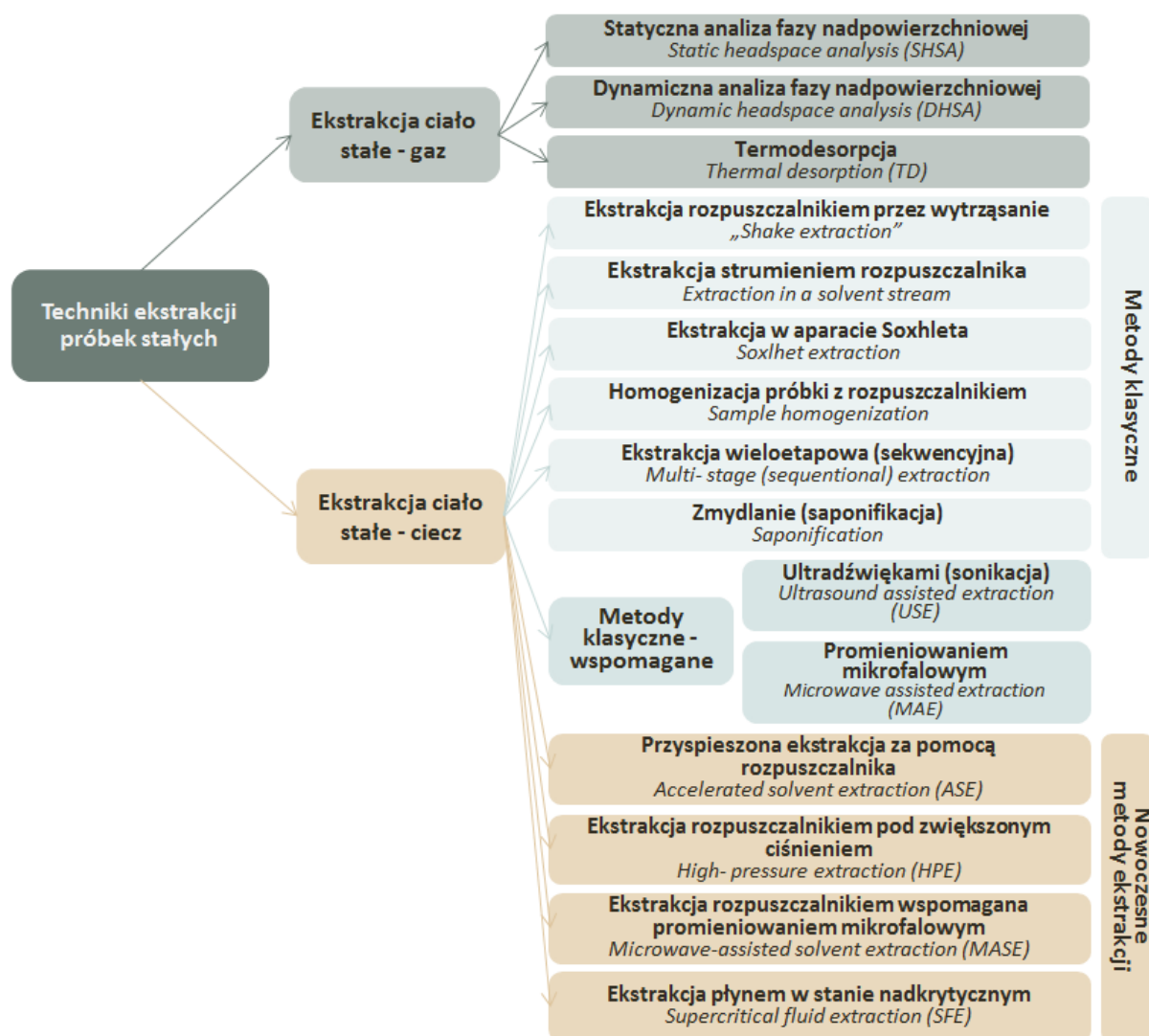
Najważniejszym i zarazem najtrudniejszym etapem w analizie chromatograficznej, decydującym o dokładności wyniku, szczególnie w przypadku próbek o złożonych matrycach, jest przygotowanie próbki. Przygotowanie próbki do analizy GC obejmuje zwykle

następujące etapy: oczyszczanie, ekstrakcję, derywatyzację i zateżenie oznaczanych substancji. Oczywistym wyzwaniem w przygotowaniu próbki do analizy GC i GC-MS jest pozbowienie jej składników nielotnych, oraz wzbogacanie próbki w składniki śladowe. Nie ma jednolitej i uniwersalnej metody przygotowania próbki, która byłaby odpowiednia dla wszystkich analizowanych substancji. Przygotowywanie próbki zależy od wielu czynników, ale głównie od jej stanu skupienia.

Przed analizą GC, do ekstrakcji lotnych i średnio lotnych związków z materiału biologicznego wykorzystywane są różne metody, tj. m.in.: klasyczna ekstrakcja ciecz-ciecz (LLE), ciągła ekstrakcja ciecz-ciecz (LLCE), ekstrakcja płynem w stanie nadkrytycznym (SFE), ciśnieniowa ekstrakcja za pomocą cieczy (PLE), ekstrakcja za pomocą rozpuszczalnika wspomaganą promieniowaniem mikrofalowym (MAE) i ekstrakcja za pomocą rozpuszczalnika wspomaganą ultradźwiękami (USE) (ryciny 1.15 i 1.16).



Rycina 1.15. Metody ekstrakcji próbek gazowych i ciekłych.



Rycina 1.16. Metody ekstrakcji próbek stałych.

Większość z nich została w pełni zautomatyzowana. Zastosowanie ekstrakcji SPE pozwala na znaczne skrócenie czasu przygotowania próbki, zredukowanie objętości zużytych rozpuszczalników a tym samym i niebezpiecznych odpadów oraz umożliwia, zarówno oddzielenie analitu od matrycy, jak i jego zateżnienie. W metodzie tej anality ekstrahowane są w układzie ciecz-ciało stałe, z wykorzystaniem podziału analitu pomiędzy próbką ciekłą a sorbent. Technika ta stała się najczęściej stosowaną techniką izolacji i/lub wzbogacania analitów z próbek ciekłych.

Zasada rozdziału techniką SPE zależy głównie od natury sorbentu. Interakcje pomiędzy analitem a cząsteczkami polarnego adsorbentu (np. żel krzemionkowy, tlenek glinu) polegają głównie na tworzeniu wiązań wodorowych, jak również wynikają z obecności oddziaływań typu dipol-dipol, czy też dipol-dipol indukowany (siły van der Waalsa). Ten rodzaj ekstrakcji oparty jest na tych samych zasadach co chromatografia adsorpcyjna. Gdy zaś sorbent stanowi krzemionka ze związaną chemicznie grupą polarną, tj. np. grupa aminowa (normalny układ

faz) lub niepolarną, jak np. grupa oktadecylowa (odwrócony układ faz), zasada rozdzielania jest taka sama jak w przypadku chromatografii podziałowej. Alternatywą dla SPE są techniki mikroekstrakcyjne, a mianowicie SPME (mikroekstrakcja do fazy stałej) i LPME (mikroekstrakcja do fazy ciekłej). Ekstrakcja SPME polega na sorpcji cząstek związków organicznych na cienkiej, cylindrycznej warstwie fazy stacjonarnej powlekającej włókno wykonane ze szkła lub kwarcu. Z kolei LPME służy do wstępnego przygotowania próbki z użyciem jedynie mikrolitrowych ilości rozpuszczalników (kilka μL). Ważną zaletą metod mikroekstrakcyjnych jest niski limit oznaczalności (5-50 ppt) i krótki czas przygotowania próbki (ok. 10 minut). Do analizy związków lotnych stosuje się również metodę analizy fazy nadpowierzchniowej (HS - headspace) czy też dynamiczną wersję metody analizy fazy nadpowierzchniowej, zwanej metodą „purge and trap” (PT).

Bezpośrednia analiza mieszaniny związków w chromatografii gazowej jest trudna ze względu na oddziaływania pomiędzy związkami, ale również interakcje pomiędzy związkami a fazą stacjonarną kolumny chromatograficznej. Interakcje te mogą zaburzać rozdzielanie pików i wpływać na ich „ogonowanie”, co sprawia, że identyfikacja analitów jest utrudniona. Upochodnienie oznaczanych związków (derywatywacja) w znaczący sposób wpływa na ograniczenie możliwości interakcji zakłócających analizę. Derywatywacja jest stosowana w celu zwiększenia lotności analizowanych substancji, zwiększenia stabilności termicznej analitów, aby uniknąć ich rozpadu podczas procesu chromatograficznego, jak również w celu zwiększenia czułości i specyficzności oznaczeń poprzez wprowadzenie odpowiednich grup funkcyjnych lub poprzez blokowanie grup funkcyjnych analitów. Do derywatywacji związków organicznych zwykle stosuje się reakcję alkilowania, acylowania lub silylowania analitów zawierających grupy funkcyjne z aktywnymi atomami wodoru, tj. np.: -COOH, -OH, -NH i -SH (tabela 1.7).

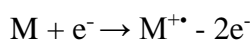
Najczęstszym sposobem derywatywacji stosowanym w GC jest silylowanie. Do tego celu najczęściej wykorzystywane są następujące odczynniki: trimetylochlorosilan (TMCS), N,O-bis(trimetylosililo)trifluoroacetamid (BSTFA), N-metylo-N-(trimetylosililo)trifluoroacetamid (MSTFA), N-metylo-N-t-butylodimetylosililotrifluoroacetamid (MTBSTFA). Związki zawierające kilka różnych grup funkcyjnych mogą wymagać wielu etapów derywatywacji.

Tabela 1.7. Sposoby derywatywacji stosowane w GC.

Rodzaj derywatywacji	Grupa funkcyjna Typ związku	Pochodna	Odczynnik
Sililowanie	-OH -alkohole, fenole -CO -ketony, steroidy -COOH -aminokwasy, kwasy tłuszczowe, steroidy -(CH ₂ OH) _n -cukry -NH, -NH ₂ -aminy, nukleozydy -SH -merkaptany -CONH, -CONH ₂ -imidy, białka	Etery trimetylosililowe Amidy trimetylosililowe	Bis(trimetylosililo)acetamid (BSA) Bis (trimetylosililo)trifluoroacetamid (BSTFA) Heksametylodisilazan (HMDS) N- metylo -N-(t-butylo-dimetylosililo)trifluoroacetamid (MTBSTFA) N- metylo-trimetylo-sililo-trifluoroacetamid (MSTFA) Trimetylochlorosilan (TMCS) Trimetylosililodietyloamina (TMS-DEA) Trimetylosililoimidazol (TMSI)
Alkilowanie	-OH -alkohole, fenole -CO -aldehydy -COOH -aminokwasy, kwasy tłuszczowe -NH, -NH ₂ -aminy, aminocukry -CONH - amidy -SH -merkaptany	Estry metylowe (DMF) Trifluorooctany (TFAA) Estry metylowe (BF ₃ -metanol) Etery pentafluorobenzylu (PFBBr) Metyloamidy (TMAH) Estry metylowe (DMF)	Bromek benzylu Trifluorek boru (BF ₃) w metanolu lub butanolu Diazometan (N ₂ CH ₂) Dimetyloformamid (DMF) Bromek pentafluorobenzylu (PFBBr) Chlorowodorek pentafluorobenzylhydroksylaminy (PFBHA) Wodorotlenek tetrabutylamonowy (TBH) Bezwodnik trifluoroctowy (TFAA) Wodorotlenek trimetyloaniliny (TMAH)
Acylowanie	-OH -alkohole, fenole -(CH ₂ OH) _n -cukry -NH, -NH ₂ -aminy -CONH -amidy -SH -merkaptany	Pentafluoropropioniany (PFPA) Trifluoroacetamidy (TFAI) Trifluoroacetamidy (MBTFA) Trifluoroacetamidy (TFAA) Etery trimetylosililowe (MBTFA)	Bezwodnik kwasu heptafluoromasłowego (HFBA) N-Metylo-bis(trifluoroacetamid) (MBTFA) Chlorek pentafluorobenzoilu (PFBCl) Pentafluoropropanol (PFPOH) Pentafluoropropanyloimidazol (PFPI) Bezwodnik kwasu pentafluoropropionowego (PFPA) Bezwodnik kwasu trifluoroctowego (TFAA) Trifluoroacetyloimidazol (TFAI)

2.2.3 Połączenie GC-MS

Istotą analizy GC-MS jest proces jonizacji cząsteczki, a następnie jej fragmentacja. Najczęściej stosowaną metodą jonizacji w GC-MS jest jonizacja elektronowa (EI) (opisana w sekcji 1.2). W celu uzyskania powtarzalnych widm MS, energia EI wynosi zazwyczaj 70 eV. W źródle EI cząsteczki lub atomy analitu tracą elektrony walencyjne, w wyniku czego powstają jony dodatnie:



Oprócz standardowej metody jonizacji - EI, w GC-MS można również stosować źródło jonów z jonizacją chemiczną (CI). Jest to stosunkowo łagodna metoda jonizacji, pozwala ona na zmniejszenie stopnia fragmentacji cząsteczki i zaobserwowanie jonu molekularnego analitu. Powtarzalność metody jest jednak gorsza niż w przypadku jonizacji EI, dlatego też jej

stosowanie nie jest tak rozpowszechnione. Jony powstałe w komorze jonizacyjnej są następnie przyspieszane w polu elektrostatycznym, skąd kierowane są przez szczelinę ogniskującą do analizatora mas pracującego pod wysoką próżnią.

Analizatory mas oddzielają naładowane jony w zależności od ich stosunku m/z , rejestrując masę oraz liczebność tych jonów (intensywność sygnałów). W urządzeniach GC-MS stosuje się wiele typów analizatorów, w tym m.in. tandemowe spektrometry mas (omówione wcześniej). GC-MS jest ważnym narzędziem identyfikacji i ilościowej analizy metabolitów ze względu na powtarzalność uzyskiwanych widm fragmentacyjnych. Stosowanie GC ograniczone jest do analizy związków lotnych, jednak duża część niskocząsteczkowych metabolitów mieści się w zakresie rozdziału chromatograficznego GC. Większość metabolitów posiada jednak polarne grupy funkcyjne, co stanowi utrudnienie dla uzyskania prawidłowej retencji i satysfakcjonującego rozdzielania składników, ze względu na kształt ich pików (piki niesymetryczne, „ogonujące”) i/lub słaby odzysk na skutek ich absorpcji na kolumnie. Z tego względu zazwyczaj konieczne jest zabezpieczenie polarnych grup funkcyjnych metabolitów poprzez zastosowanie chemicznej derywatywacji.

Literatura:

H.J. Hübschmann, Handbook of GC-MS: Fundamentals and Applications (3rd Edition). Somerset, NJ, USA: Wiley, 2015. ProQuest ebrary. Web. 8 July 2015.

X. Guo, Advances in Gas Chromatography, Chapter 4: “Gas Chromatography in Metabolomics Study” by Yunping Qiu and Deborah Reed. InTech, 2014.

2.3. Elektroforeza kapilarna – spektrometria mas (CE-MS)

2.3.1. Wstęp do CE-MS

Elektroforeza kapilarna sprzężona ze spektrometrią mas, CE-MS, stanowi połączenie techniki separacyjnej o wysokiej efektywności i rozdzielczości oraz spektrometrii mas i jest uważana za metodę ortogonalną dla GC lub LC. Wprowadzono ją pod koniec lat osiemdziesiątych i może dostarczyć istotnych informacji o składzie niskocząsteczkowych, polarnych związków oraz obdarzonych ładunkiem elektrycznym metabolitów. W rezultacie CE-MS wykorzystuje się w różnorodnych dziedzinach badań metabolomicznych, takich jak badania biomedyczne, kliniczne oraz roślin. Próbkę krwi (surowicy i osocza), moczu i innych płynów biologicznych, kultur komórkowych, tkanek i roślin można analizować za pomocą CE-MS z minimalną obróbką próbki. Wiele różnych typów analitów, w tym aminokwasy, krótkołańcuchowe kwasy organiczne (SCFA), pochodne kwasów tłuszczowych (acylo-karnityny), nukleotydy, fosforany węglowodanów, nukleotydy i metabolizm poliamin, są często analizowane za pomocą CE-MS. Ponadto, można również analizować małe peptydy i białka (po hydrolizie lub natywne białka).

Zalety techniki CE-MS

- Technika CE-MS doskonale nadaje się do analizy próbek w roztworach wodnych. Co więcej, można analizować próbki o wysokim stężeniu soli, takie jak jak pożywki hodowlane lub ekstrakty komórkowe.
- Wymagana jest mała objętość próbki. Jest to bardzo ważne w wielu eksperymentach, gdyż dostępna jest tylko minimalna ilość próbki (małe rozmiary zwierząt, łyż, płyn mózgowo-rdzeniowy i inne płyny biologiczne itp.).
- Odpowiednia dla analitów hydrofilowych.

- Wydajna separacja i ortogonalny mechanizm dla technik chromatograficznych.
- Po analizie, kapilarę można bardzo łatwo przepłukać i skondycjonować przed następną analizą.
- Zielona technika analityczna. Zużycie odczynników i rozpuszczalników organicznych jest bardzo niskie, dużo niższe niż w technice LC.

Wady techniki CE-MS

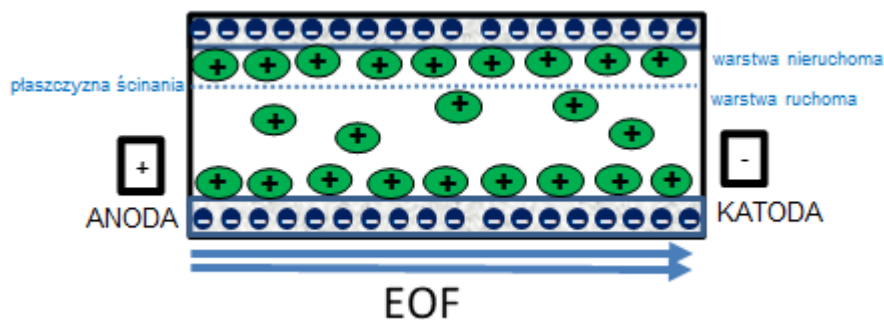
- Niska powtarzalność i stabilność CE w odniesieniu do czasów migracji i odpowiedzi detektora.
- Niska czułość ze względu na ilość próbki wprowadzanej do systemu, około nL.
- Najmniej stabilne połączenie.
- Brak dostępnych bibliotek widm. Identyfikacja nieznanymi związków w oparciu o własne biblioteki widm i bazy danych.

2.3.2. Teoretyczne zasady techniki CE-MS

Rozdział w elektroforezie kapilarnej opiera się na separacji ze względu na ładunek elektryczny w kapilarze krzemionkowej, zwykle o średnicy 50 μm i długości 100 cm. Ruch analitów wzdłuż kapilary wypełnionej przewodzącym ciełym ośrodkiem pod wpływem pola elektrycznego opiera się na stosunku ładunku do masy, co nadaje tej technice unikalne właściwości, zarówno z jej zaletami jak i wadami. Najczęściej stosowane są odsłonięte kapilary krzemionkowe, ale także stosujemy kapilary otoczone powłoką (obojętne lub dodatnie powłoki). Powszechnie używanymi terminami są: elektroferogramy zamiast chromatogramy i czasy migracji zamiast czasów retencji. Wykres odpowiedzi detektora względem czasu migracji nazywany jest elektroferogramem.

Zasady separacji

Dwie siły regulują analizę CE: przepływ elektroosmotyczny (EOF) i ruchliwość elektroforetyczna (EM). Przepływ elektroosmotyczny jest konsekwencją ładunku powierzchniowego na wewnętrznej ścianie kapilary. W przypadku odsłoniętych kapilar, grupy silanolowe (Si-OH) na powierzchni są częściowo zjonizowane ujemnie ze względu na naładowane grupy silanowe (Si-O⁻) przy pH > 2. Ujemne ładunki przyciągają dodatnie ładunki buforu w kapilarze tworząc dwie wewnętrzne warstwy, jak pokazano na rycinie 1.17. "Warstwa nieruchoma" składa się ze ściśle ułożonych ładunków dodatnich, a "warstwa ruchoma" zawiera ładunki nie tak ściśle ułożone, bardziej rozproszone. Po przyłożeniu pola elektrycznego ruchoma warstwa kationów przemieszcza się w kierunku elektrody ujemnej, kationy są solwatowane i tym samym powoduje przemieszczenie całego roztworu buforu. Rezultatem jest przepływ roztworu buforowego w kierunku elektrody ujemnej, czyli jest to przepływ elektroosmotyczny. Ten rodzaj analizy nazywa się normalną elektroforezą kapilarną. Istnieje również inny tryb, odwrotna elektroforeza kapilarna, w którym zmieniona jest polarność elektrod. W tym przypadku elektrodą wlotową jest katoda, a przepływ powinien zostać stłumiony lub odwrócony, na przykład przez dodanie czwartorzędowych amin lub przez zastosowanie kapilar o dodatnim ładunku.



Rycina 1.17. Przepływ elektroosmotyczny w kapilarze.

Między dwiema warstwami znajduje się płaszczyzna ścinania, która charakteryzuje się różnicą potencjałów, zwaną potencjałem zeta (ζ), z powodu braku równowagi elektrycznej ładunków. Potencjał zeta jest proporcjonalny do przepływu elektroosmotycznego i grubości podwójnej warstwy i odwrotnie proporcjonalny do stałej dielektrycznej buforu, jak opisano w równaniu:

$$\zeta = 4\pi\delta e/\epsilon$$

gdzie δ jest to grubość podwójnej warstwy rozproszonej, e jest to ładunek na jednostkę powierzchni, a ϵ jest to stała dielektryczna buforu.

Na prędkość EOF wpływają różne parametry, takie jak pH buforu, napięcie i temperatura. Wzrost pH buforu powoduje wzrost EOF z powodu dysocjacji Si-OH i powstaniu grup Si-O⁻ na wewnętrznej ścianie kapilary, co zwiększa potencjał zeta (ζ). Pole elektryczne (napięcie na jednostkę długości) ma znaczący wpływ na prędkość przepływu elektroosmotycznego. Im wyższe pole elektryczne, tym większa prędkość przepływu elektroosmotycznego i elektroforetycznego oraz krótszy czas analizy. Jednak wysokie napięcie może wytwarzać ciepło Joule'a, które zwiększa temperaturę w kapilarze. Ważna jest kontrola tego parametru, ponieważ wysoka temperatura może niszczyć i denaturować próbki, a ponadto gradient lepkości buforu może być uformowany prostopadle w stosunku do kapilary zwiększając dyfuzję i dyspersję. Równanie Smoluchowskiego określa elektroosmotyczną prędkość przepływu:

$$v_{\text{eof}} = \epsilon\zeta E/4\pi\eta$$

gdzie v_{eof} określa elektroosmotyczną prędkość przepływu, ϵ jest to stała dielektryczna buforu, E określa przyłożone pole elektryczne wyrażone w voltach/cm, η jest to lepkość buforu, ζ jest to potencjał zeta mierzony na płaszczyźnie ścinania.

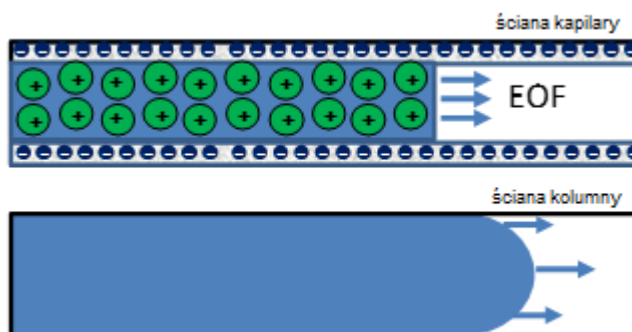
Bufor posiada ruchliwość elektroosmotyczną (μ_{eof}), która zależy od charakterystyki buforu czyli stałej dielektrycznej i lepkości i jest niezależna od przyłożonego pola elektrycznego, jak opisano w równaniu:

$$\mu_{\text{eof}} = \epsilon\zeta/4\pi\eta$$

μ_{eof} jest to ruchliwość elektroosmotyczna buforu, ϵ jest to stała dielektryczna buforu, η jest to lepkość buforu, ζ jest to potencjał zeta mierzony na płaszczyźnie ścinania.

Elektroosmotyczny profil przepływu różni się od profilu przepływu cieczy pod ciśnieniem (jak w HPLC). Na rycinie 1.18, na górnym diagramie pokazano płaski profil przepływu w CE, który jest spowodowany ruchem wszystkich ładunków w kierunku elektrod. Drugi schemat ilustruje paraboliczny profil przepływu w HPLC ze względu na tarcie na

ściankach kolumny, które spowalnia przepływ bliżej boków kolumny. Zaletą płaskiego profilu przepływu jest to, że wszystkie cząsteczki substancji rozpuszczonej mają tę samą prędkość wewnątrz kapilary, powstają wąskie piki o wysokiej wydajności. W parabolicznym profilu przepływu w HPLC cząsteczki substancji rozpuszczonej w środku kolumny poruszają się szybciej niż te bliżej ścianki kolumny, dając względnie szerokie piki.



Rycina 1.18. Diagram przepływu w CE oraz w HPLC.

Ruchliwość elektroforetyczna jest dodatkową siłą, która obecna jest tylko w naładowanych cząsteczkach i może przyciągać jony w kierunku elektrody z przeciwnym ładunkiem. Prędkość, z którą substancja rozpuszczona porusza się zgodnie z ładunkiem i pod wpływem przyłożonego pola elektrycznego, wynika z równania:

$$v_{ep} = \mu_{ep} E$$

v_{ep} jest to prędkość elektroforetyczna substancji rozpuszczonej, μ_{ep} jest to elektroforetyczna ruchliwość substancji rozpuszczonej, a E jest to pole elektryczne.

Ruchliwość elektroforetyczna (μ_{ep}) jest wprost proporcjonalna do ładunku substancji rozpuszczonej i odwrotnie proporcjonalna do wielkości cząsteczek, a także do lepkości ośrodka elektroforetycznego, jak opisano w następującym równaniu:

$$\mu_{ep} = Q / 6\pi r \eta$$

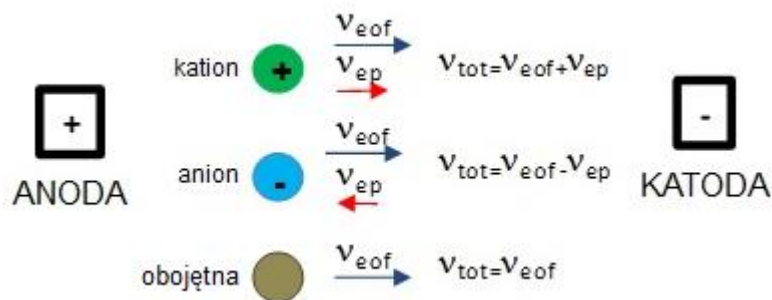
μ_{ep} jest to elektroforetyczna ruchliwość substancji rozpuszczonej, Q jest to ładunek substancji rozpuszczonej, r promień jonowy substancji rozpuszczonej, η lepkość ośrodka.

Ostatnie równanie dowodzi, że powodem, dla którego małe naładowane cząsteczki poruszają się przez kapilarę szybciej niż duże cząsteczki o niższym ładunku, jest ich różna ruchliwość elektroforetyczna (μ_{ep}).

Całkowitą prędkość substancji rozpuszczonej (v_{tot}) w kapilarze określa równanie:

$$v_{tot} = v_{ep} + v_{eof}$$

W elektroforezie kapilarnej o normalnej polarności wszystkie cząsteczki mają taką samą elektroosmotyczną prędkość przepływu (v_{eof}), co pokazano na rycinie 1.19, ale wymywają się w różnym czasie. Kationy najpierw eluują, ponieważ posiadają prędkość elektroforetyczną (v_{ep}) i migrują w tą samą stronę co przepływ elektroosmotyczny, obojętne cząsteczki mają tę samą prędkość co przepływ elektroosmotyczny, ponieważ nie są naładowane, i ostatecznie, aniony eluują ostatnie, ponieważ kierunek prędkości elektroforetycznej (v_{ep}) jest skierowany w stronę przeciwnej elektrody (anody).



Rycina 1.19. Całkowita ruchliwość substancji rozpuszczonej w kapilarze.

Liczne czynniki wpływają na rozdział elektroforetyczny: pH buforu separacyjnego, który wpływa na ładunek i promień solwatowanych cząsteczek, stężenie buforu gdy EOF zmniejsza się z pierwiastkiem kwadratowym stężenia buforu, lepkość ośrodka rozdzielczego, temperatura, zastosowane pole elektryczne itp.

W CE można wyróżnić następujące typy separacji: kapilarna elektroforeza strefowa (CZE), micelarna chromatografia elektrokinetyczna (MEKC), elektrochromatografia kapilarna (CEC), kapilarna elektroforeza żelowa (CGE). CZE jest odpowiednia dla związków polarnych i naładowanych i jest najczęściej stosowaną techniką CE, gdy jest sprzężona z MS. CE zastosować możemy również do wykrywania związków obojętnych przy użyciu metody MECK. Inne dodatki mogą modyfikować selektywność rozdziału, na przykład cyklodekstryny w celu rozdziału chiralnego.

2.3.2.1. Kapilarna elektroforeza strefowa

Jest to najprostsza i najbardziej rozpowszechniona metoda rozdziału przy udziale CE. Rozdziałowi ulegają tylko związki, które posiadają ładunek elektryczny. Anality przemieszczają się od jednego końca kapilary do drugiego zgodnie z sumą wektorów: ruchliwości elektroforetycznej i przepływu elektroosmotycznego.

2.3.2.2. Micelarna chromatografia elektrokinetyczna

Wprowadzona przez Terabe w 1984 roku w celu rozdziału związków obojętnych zamiast metody CZE. MEKC zawiera odpowiednie naładowane detergenty w buforze separacyjnym w stężeniu dostatecznie wysokim, aby utworzyć micelle. Detergenty mogą być kationowe, anionowe, niejonowe lub dwubiegunowe, ale w połączeniu z CE-MS muszą być lotne. Te makromolekularne agregaty z hydrofobowym rdzeniem wewnętrznym i powierzchnią polarną mogą oddziaływać z analitem za pomocą innego mechanizmu, czyli siły podziału i siły granicznej. Im bardziej hydrofobowe związki tym silniej oddziałują z micelą i obniżają prędkość wewnątrz kapilary. Jeśli nie wchodzi w interakcje z micelą, cząsteczka analitu migruje wraz z EOF. We wszystkich przypadkach zmiany stężenia buforu, pH, temperatura lub stosowanie dodatków, takich jak mocznik, jony metali lub selektory chiralne, można również stosować w celu zmiany selektywności. Podobnie jak w CZE, wszystkie modyfikatory, takie jak metanol, acetonitryl i 2-propanol, zostały z powodzeniem zastosowane.

2.3.2.3. Elektrochromatografia kapilarna

Jest to forma zminiaturyzowanej chromatografii cieczowej. Kapilary wypełnione są cząsteczkami fazy odwróconej na bazie krzemionki typu HPLC, 1-5 μm , jako faza

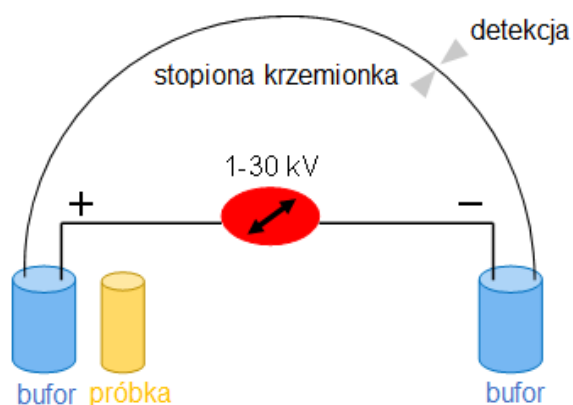
stacjonarna. Poza tym, stosuje się konwencjonalne fazy ruchome dla rozdzału typu RP, takie jak mieszaniny rozpuszczalników organicznych, rozpuszczalnik/wodne mieszaniny buforów. Obowiązuje mechanizm podziału substancji rozpuszczonych pomiędzy fazę ruchomą i stacjonarną, a przy naładowanych metabolitach dodatkowa prędkość elektroforetyczna modyfikuje rozdział. W systemie wymagana jest dodatkowa pompa w celu przepływu fazy ruchomej przez wypełnioną kapilarę. Profil przepływu EOF w kapilarze jest identyczny z profilem przepływu w CZE.

2.3.2.4. Kapilarna elektroforeza żelowa

Szeroko stosowana w badaniach biologii molekularnej dla cząsteczek różniących się tylko wielkością, a nie stosunkiem masy do ładunku, takich jak białka i kwasy nukleinowe DNA. Kapilara wypełniona jest żelem lub roztworem lepkim. Polimery w CGE mogą być kowalencyjnie usieciowane, związane wodorowo lub po prostu rozpuszczone w roztworach buforowych. Przepływ EOF jest często tłumiony, migracja analitów odbywa się wyłącznie za pomocą ruchliwości elektroforetycznej. W lepkim lub polimerowym medium sieciowym ruch analitów jest utrudniony, dłuższe cząsteczki migrują dłużej niż krótsze cząsteczki. W celu uzyskania bardziej jednolitego stosunku ładunku do masy w białkach, zwykle wykonuje się wstępne traktowanie detergentem, SDS. Zazwyczaj, białka wiążą się ze stałą liczbą cząsteczek SDS na jednostkę długości podobnie jak w elektroforezie żelowej (SDS-poliakryloamid).

2.3.2.5. Aparatura

System elektroforezy kapilarnej składa się z kapilary, fiolki wlotowej i wylotowej, fiolki na próbki, układu chłodzącego zdolnego do eliminacji ciepła wytwarzanego w kapilarze, dwóch elektrod (anoda i katoda), źródła zasilania i detektora, jak to pokazano na rycinie 1.20. Kapilarę, fiolkę wlotową i wylotową napełnia się tym samym buforem, który zwykle jest roztworem wodnym. Anoda i katoda są zwykle elektrodami palladowymi, podłączonymi do źródła zasilania, a detektor zwykle umieszcza się w pobliżu końca kapilary (blisko fiolki wylotowej).



Rycina 1.20. Schemat elektroforezy kapilarnej.

Kapilary ze stopionej krzemionki pokryte powłoką poliimidową stosuje się w technice CE-MS. Możemy stosować kapilary nieosłonięte lub wewnętrzna ściana kapilary może być pokryta polimerami (obojętnymi lub dodatnimi), aby zredukować elektroosmozę lub nawet ją

odwrócić. Kapilara umieszczona jest pomiędzy dwiema fiolkami, także pomiędzy elektrodami i próbką. Pobieranie próbki odbywa się poprzez zastąpienie fiolki buforu, fiolką z próbką i nastrzyknięciem próbki za pomocą siły grawitacyjnej, zastosowanie ciśnienia zewnętrznego (hydrodynamiczne) lub zastosowanie pola elektrycznego (nastrzyknięcie elektrokinetyczne). Następnie, kapilarę umieszcza się w fiolce buforowej, i przykłada odpowiednie napięcie w celu rozdziału składników. Dostępne są różne układy chłodzenia zdolne do zmniejszania ciepła wytwarzanego wewnątrz kapilary dzięki efektowi Joule'a.

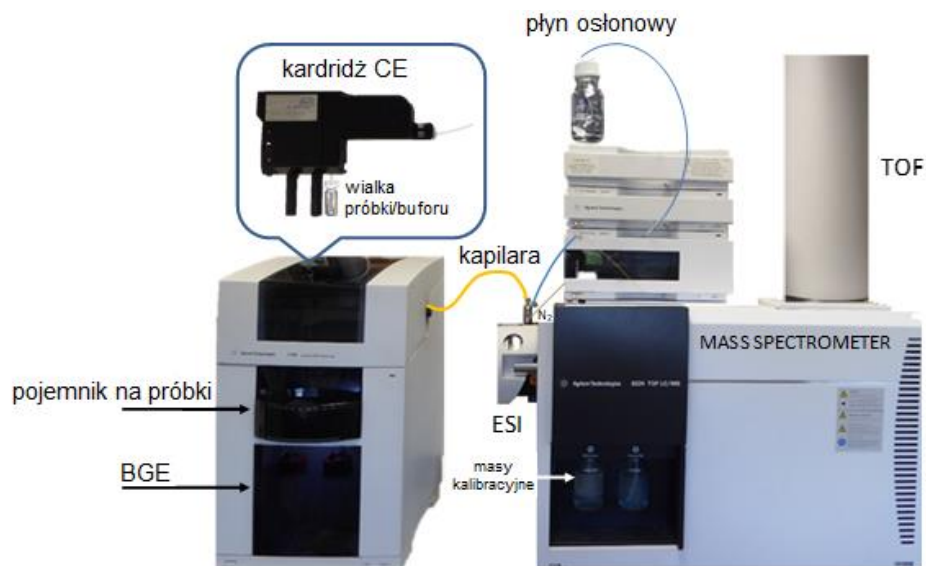
Najczęściej stosowanymi systemami detekcji w elektroforezie kapilarnej są detektor ultrafioletowy/widzialny (UV-VIS), detektor fotodiody (PDA), fluorescencyjny indukowany laserem (LIF) i spektrometr mas (MS) (Tabela 1.8). Zdecydowanie największe możliwości detekcji związane są z techniką CE-MS, ze względu na informacje o czasie migracji i widmie mas związków. W przypadku stosowania detektorów spektrometrii mas najczęściej stosowanym źródłem jonizacji jest jonizacja z elektrosprej (ESI).

Tabela 1.8. Różnice pomiędzy detektorami w CE.

Detektor	Zalety	Charakterystyka	Limit detekcji (M)
UV/VIS	możliwość bezpośredniej i pośredniej detekcji, często stosowany detektor	uniwersalny	10^{-3} 10^{-6} związki aromatyczne
LIF	czuły i selektywny, używany do związków fluorescencyjnych lub pochodnych	selektywny	$10^{-6} - 10^{-9}$
MS	informacje jakościowe i ilościowe, wysoka czułość i selektywność	uniwersalny selektywny	$\sim 10^{-5}$ (zależy od rodzaju detektora MS i rodzaju metabolitów)

W połączeniu z innymi technikami detekcji, takimi jak UV lub fluorescencyjnymi, wlot i wylot kapilary zanurza się w fiolkach z buforem separacyjnym (BGE). Jednakże, gdy stosuje się technikę MS jako detektor, fiolkę wylotową zastępuje się wejściem do MS. W takim przypadku konieczne jest użycie interfejsu, który pasuje do dwóch napięć i utrzymuje stały strumień cieczy (spray) przy bardzo małej objętości płynu (Rycina 1.21). W CE-ESI-MS objętość wprowadzona do kapilary wynosi około kilku nanolitrow. W celu utrzymania stałego przepływu i sprayu, powszechne jest stosowanie *cieczy osłonowej* (SL). Przepływ cieczy osłonowej, w zakresie od nl/min do μ l/min, pozwala na utworzenie stabilnego sprayu i utworzenie jonów, zwykle skład cieczy osłonowej to mieszanina woda-metanol w stosunku 1:1 z 0,1% kwasem octowym lub kwasem mrówkowym. Osiągnięcie stabilnej pracy w trybie elektrorozpylania często zależy od zrównoważenia wielu parametrów, takich jak pozycja kapilary, szybkość przepływu cieczy osłonowej i jej skład, przepływ gazu osłonowego oraz warunki ESI. Zastosowanie interfejsów, które nie posiadają cieczy osłonowej i wykazują wyższą czułości, związana jest z niską stabilnością analizy.

Wszystkie dostępne na rynku spektrometry mas mogą być połączone do CE, jednak największą liczbę publikacji posiada połączenie CE-TOF (analyzer czasu przelotu). Wynika to głównie z dużej szybkości skanowania (cyklu pracy) koniecznej do uzyskania wystarczającej liczby punktów wzdłuż piku, ale także z powodu dokładności mas i wysokiej rozdzielczości.



Rycina 1.21. Graficzna prezentacja CE połączonej z MS, z wyszczególnieniem najważniejszych części w obu aparatach.

Literatura:

D.R. Baker. Capillary electrophoresis. Wiley, 1995.

S. Terabe, K. Otsuka, T. Ando. Electrokinetic chromatography with micellar solution and open-tubular capillary. *Anal. Chem.*, 1985, 57 (4):834-841.

T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, T. Nishioka. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res.*, 2003, 2(5):488-94.

A. García, S. Naz, C. Barbas. Metabolite fingerprinting by capillary electrophoresis-mass spectrometry. *Methods Molecular Biology*, 2014, 1198:107-23.

A. García, J. Godzien, Á. López-González, C. Barbas. Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics based on GC-(EI)-Q-MS. *Anal. Chim. Acta*, 2015, 900:21-35.

3. Część praktyczna

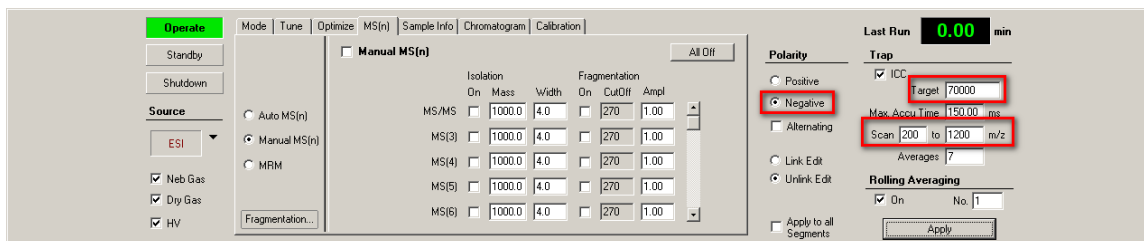
3.1. Rejestracja widm masowych ESI-MS

Po szczegółowym omówieniu zasady działania i budowy analizatora mas typu pułapki jonowej wyposażonej w źródło jonizacji typu ESI, zostanie przeprowadzona procedura, mająca na celu uzyskani widma ESI-MS. Przy wykorzystaniu pompy strzykawkowej stałoprzepływowej próbka zawierająca związek o dokładnie znanej masie molekularnej (np. rutyny lub innego dostępnego komercyjnie standardu związku drobnocząsteczkowego) zostanie bezpośrednio wprowadzona na źródło jonizacji ESI. Analiza zostanie przeprowadzona zarówno w negatywnym jak i pozytywnym trybie jonizacji, aby ocenić, który z nich zapewnia efektywniejszą jonizację analizowanego związku, co warunkuje jego budowa strukturalna.

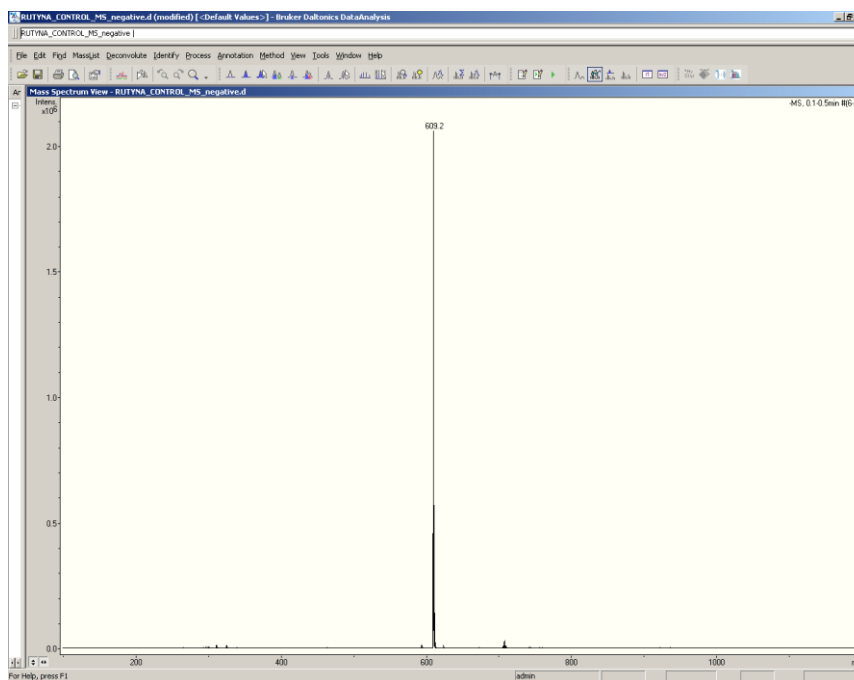
W kolejnym etapie zajęć zostanie omówiona zasada działania oraz budowa analizatora typu orbitrap wyposażonego w źródło jonizacji nanoESI. Następnie z użyciem tego analizatora i pompy strzykawkowej zostanie przeprowadzona analiza roztworu komercyjnie dostępnego standardu peptydu (np. tioredoksyny). W wyniku przeprowadzonej analizy studenci otrzymają widmo ESI-MS charakterystyczne dla związku, który ulega w źródle ESI wielokrotnej jonizacji. Uzyskane widma masowe ESI-MS zostaną zapisane, a następnie szczegółowo omówione i poddane wnikliwej interpretacji w punkcie 3.3.

Eksperyment ESI-MS:

1. Przygotuj roztwór rutyny o stężeniu 1 mg/ml (masa molekularna 610.52) w DMSO.
2. Rozcieńcz roztwór z punktu 1 mieszaniną ACN/H₂O (60/40, v/v) do uzyskania stężenia 10 µg/ml – (roztwór rutyny w 1% DMSO).
3. Dlaczego wymagane jest 1% stężenie DMSO?
4. Przy użyciu pompy strzykawkowej wprowadź z prędkością przepływu 150 µL/h przygotowany roztwór bezpośrednio na źródło jonizacji.
5. Zarejestruj widmo masowe przez minimum 30 sekund po uprzednim ustawieniu parametrów zbierania jak pokazano na poniższej rycinie.



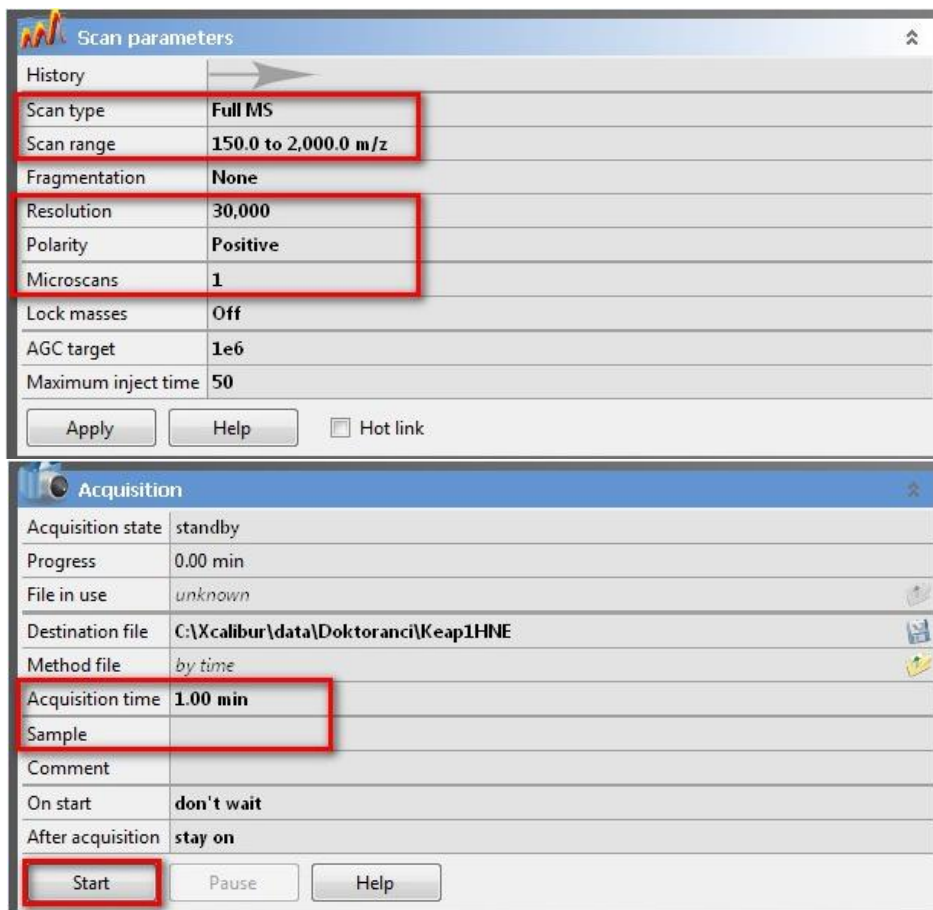
6. Odszukaj jon molekularny rutyny oraz uzasadnij wartość wskazanego stosunku m/z



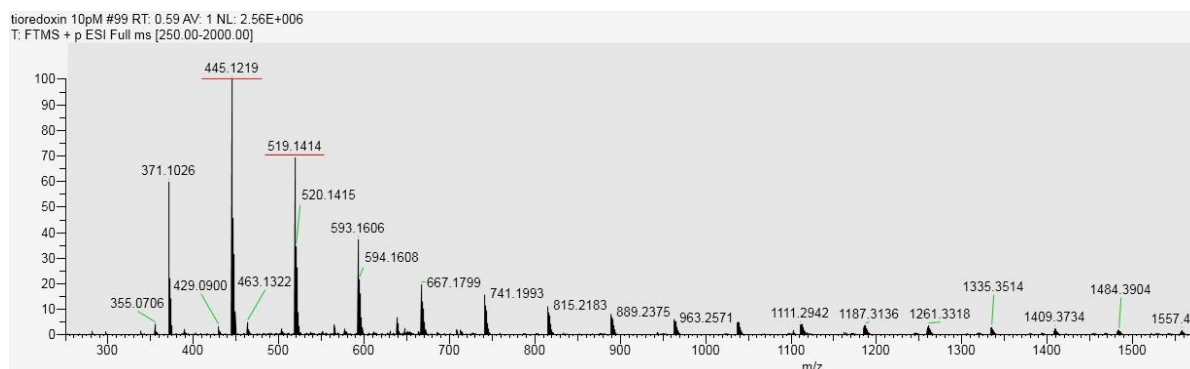
Widmo ESI-MS rutyny otrzymane w trybie jonizacji negatywnej

Eksperyment ESI-OrbiTrap MS:

1. Przygotuj roztwór tioredoksyny (masa molowa 12kDa) o stężeniu 10µM.
2. Rozcieńcz uzyskany roztwór mieszaniną ACN/H₂O (50/50 V/V) + 0.1% FA w celu uzyskania roztworu tioredoksyny o stężeniu 10pM.
3. Dlaczego roztwór zawierający dodatek 0.1% FA jest sporządzany?
4. Wprowadź przygotowany roztwór na źródło ESI przy użyciu pompy strzykawkowej z prędkością przepływu 2µL/min.
5. Zarejestruj widmo masowe w ciągu 60 sekund używając parametrów podanych niżej.



6. Otwórz w programie zarejestrowane widmo. Znając całkowitą masę tioredoksyny (12kDa), oblicz i przypisz ładunek poszczególnym jonom występującym na widmie.



e.g. ion 445.12 $m/z = (M+z)/z$
 $445.12 = (12000+z)/z$
 $z = 12000/444.12$
 $z = 27$

ion 519.14 $m/z = (M+z)/z$
 $519.14 = (12000+z)/z$
 $z = 12000/518.12$
 $z = 23$

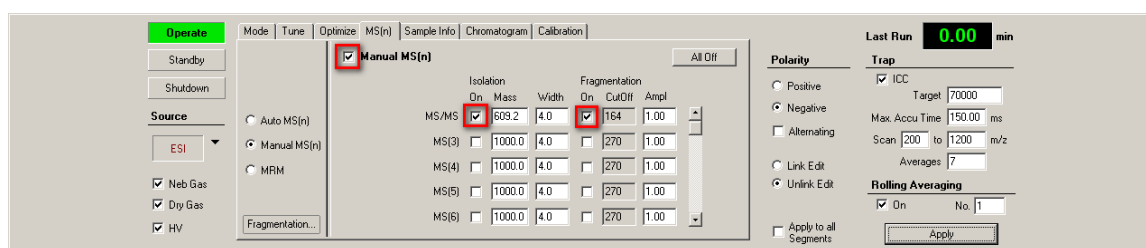
3.2. Rejestracja widm masowych ESI-MS/MS

W tej części zajęć praktycznych ta sama próbka roztworu rutyny użyta w punkcie 3.1 zostanie poddana tandemowej analizie masowej (MS/MS) z użyciem pompy strzykawkowej

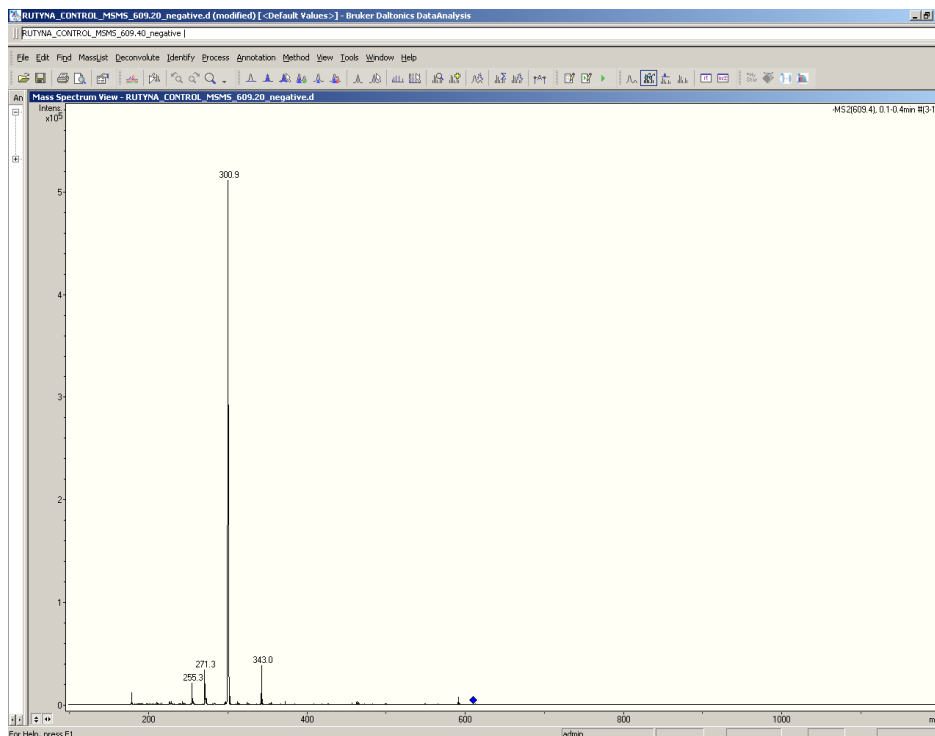
oraz pułapki jonowej wyposażonej w źródło jonizacji ESI. Na tym etapie, studenci używając specjalistycznego oprogramowania obsługującego spektrometr mas wprowadzą odpowiednie ustawienia w celu wyselekcjonowania wybranych jonów o odpowiednim stosunku m/z i poddania ich fragmentacji z uzyskaniem widm fragmentacyjnych ES-MS/MS. Dzięki otrzymanym widmom masowym studenci nabędą umiejętność wyselekcjonowania jonów prekursora oraz jego jonów potomnych w celu ewentualnego przeprowadzenia typowych targetowych analiz ilościowych w trybie monitorowania par jonowych MRM. Uzyskane widma masowe ESI-MS/MS zostaną zarejestrowane, a następnie szczegółowo omówione i poddane wnikliwej interpretacji w punkcie 3.3.

Eksperyment ESI-MS/MS:

1. Przy użyciu pompy strzykawkowej wprowadź z prędkością przepływu 150 $\mu\text{L/h}$ przygotowany w punkcie 3.1 roztwór rutyny bezpośrednio na źródło jonizacji.
2. W zakładce MSn zaznacz opcję „isolation” i „fragmentation” oraz wpisz wartość m/z jonu molekularnego rutyny wyselekcjonowanego w eksperymencie LC-MS.



3. Zarejestruj widmo masowe przez minimum 30 sekund po uprzednim ustawieniu parametrów zbierania jak pokazano na powyższej rycinie.



Widmo ESI-MS/MS rutyny otrzymane w trybie jonizacji negatywnej.

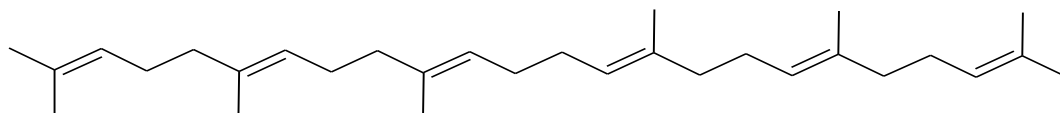
4. Przeanalizuj dokładnie otrzymane widmo ESI-MS/MS.
5. Które wartości m/z uzyskanych jonów wybrał(a)byś do ilościowej analizy w trybie MRM?
6. Dlaczego?

3.3. Interpretacja widm masowych

Studenci na podstawie otrzymanych w punktach 3.1 i 3.2 widm masowych dokonają identyfikacji piku molekularnego, a następnie biorąc pod uwagę strukturę analizowanego związku, na podstawie wartości m/z przedyskutują wybór charakterystycznych dla niego jonów fragmentarycznych. Dodatkowo widma zostaną poddane wnikliwej analizie pod kątem tworzenia adduktów z jonami sodu lub potasu i ewentualnego tworzenia jonów wielokrotnie naładowanych.

3.4. Wykorzystanie metody GC-MS w trybie pracy monitorowania pojedynczych jonów

Celem ćwiczenia jest praktyczne zapoznanie studentów z techniką chromatografii gazowej sprzężonej ze spektrometrią mas, na przykładzie jakościowego i ilościowego oznaczania typowego związku nietłotnego (np. skwalenu jako składnika olejów jadalnych, tj. olej amarantusowy, olej z wiesiołka, olej rokitnikowy) z wykorzystaniem metody GC-SIM MS.



Skwalen – $C_{30}H_{50}$, (6*E*,10*E*,14*E*,18*E*)-2,6,10,15,19,23-heksametylotetrakoza-2,6,10,14,18,22-heksen

a) Materiały

Przygotowanie wzorców

- Przygotuj serię roztworów wzorcowych skwalenu w następujących stężeniach: 0.5 mg/mL, 1.0 mg/mL, 2 mg/mL, 5 mg/mL i 10 mg/mL w heksanie

Przygotowanie próbek

- 10 μ L oleju przenieść do 2mL fiolki GC
- dodać 40 μ L odczynnika silylującego (BSTFA:TMCS; 9:1)
- fiolki szczelnie zamknąć i ogrzewać w temp. 80°C przez 1 godzinę a następnie schłodzić do temperatury pokojowej
- nastrzyknąć 1 μ L próbki

b) Warunki eksperymentu

Do analizy próbek wykorzystany zostanie chromatograf gazowy sprzężony ze spektrometrem mas GC/MS system with 7890A GC – 7000 quadrupole MS/MS (Agilent Technologies, Palo Alto, CA). Szczegółowe parametry metody podano poniżej:

GC

Kolumna kapilarna - HP-5ms (0.25-mm; 0.25- μ m, 30-m)

Przepływ gazu przez kolumnę (He) - 1 mL/min

ALS Inlets Columns Oven Aux Heaters Events Signals Configuration Counters Readiness GC Calculators

Selection

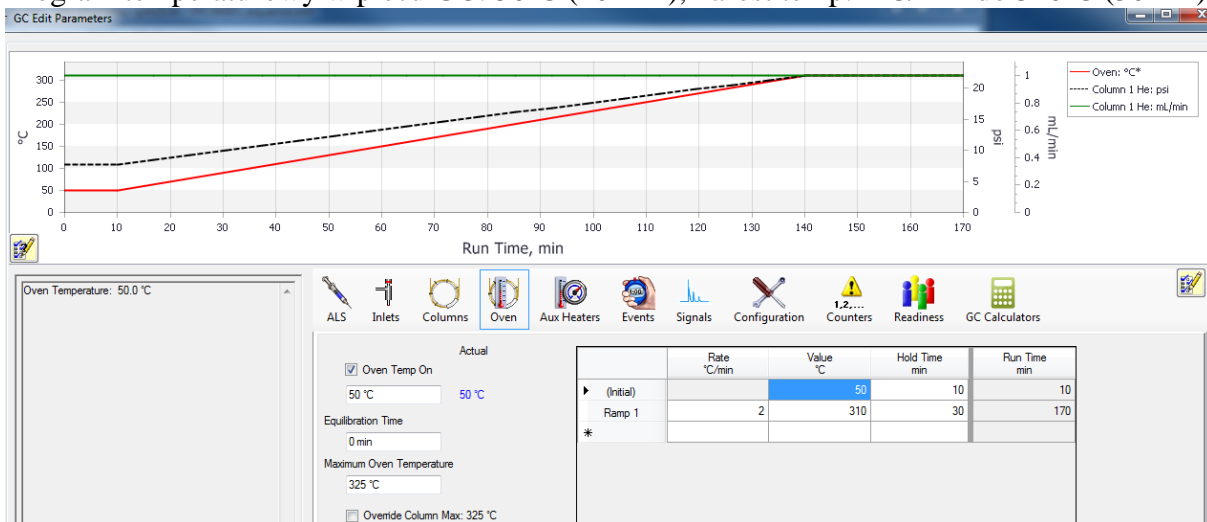
1	Agilent 19091S-433. 1371.48429 HP-5MS 5% Phenyl Methyl Silox -60 °C-325 °C (325 °C): 30 m x 250 μm x 0.25 μm In: Front SS Inlet He Out: MSD
	QQQ Collision Cell EPC

Control Mode On

	Setpoint	Actual
Flow	1 mL/min	1 mL/min
Pressure	7.6522 psi	7.7 psi
Average Velocity	36.445 cm/sec	(Initial): 0 min He @ 50 °C Oven Out: MSD 30 m x 250 μm x 0.25 μm
Holdup Time	1.3719 min	

Constant Flow

Program temperaturowy w piecu GC: 50°C (10 min), narost temp. 2°C/min do 310°C (30min)



Temperatura dozownika (S/SL): 270°C, split – 20:1

ALS Inlets Columns Oven Aux Heaters Events Signals Configuration Counters Readiness GC Calculators

Split-Splitless Inlet

	Setpoint	Actual
<input checked="" type="checkbox"/> Heater:	270 °C	270 °C
<input checked="" type="checkbox"/> Pressure:	7.6522 psi	7.7 psi
Total Flow:	24 mL/min	24 mL/min
<input checked="" type="checkbox"/> Septum Purge Flow:	3 mL/min	3 mL/min

Septum Purge Flow Mode: Standard

Gas Saver: On
20 mL/min After: 3 min

Mode: Split Split Ratio: 20 :1 Split Flow 20 mL/min

Temperatura linii transferowej - 260°C

ALS Inlets Columns Oven Aux Heaters Events Signals Configuration Counters Readiness GC Calculators

Thermal Aux 2 (MSD Transfer Line)

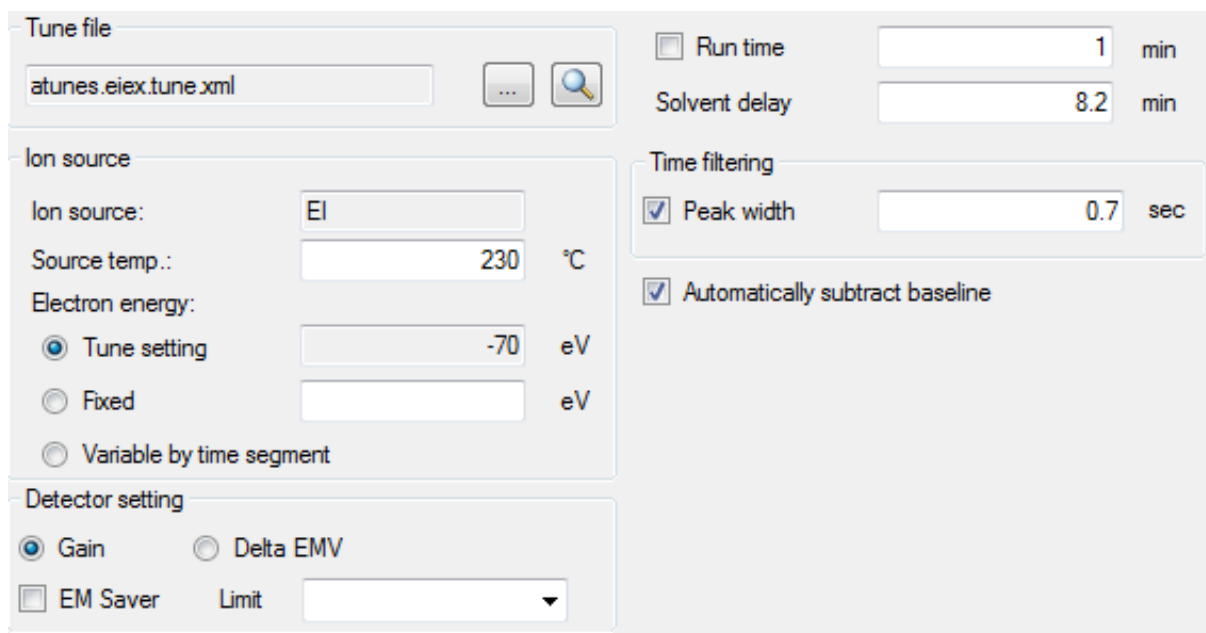
Actual

On
260 °C 260 °C

MS

Źródło jonów (EI) - 230°C

Energia jonizacji – 70 eV



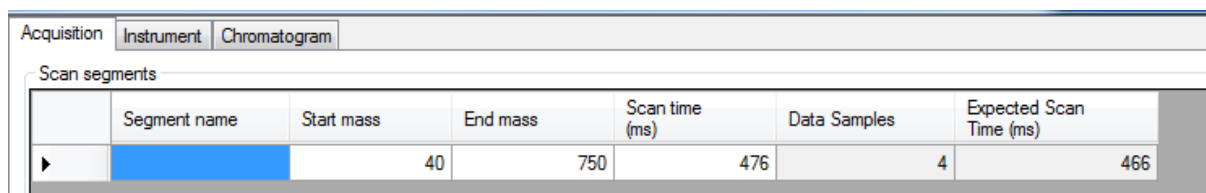
The screenshot shows the 'MS SIM/Scan Parameters' dialog box. It is divided into several sections:

- Tune file:** A text box containing 'atunes.eiex.tune.xml' with browse and search icons.
- Ion source:** Includes fields for 'Ion source:' (EI), 'Source temp.:' (230 °C), and 'Electron energy:' with radio buttons for 'Tune setting' (-70 eV), 'Fixed', and 'Variable by time segment'.
- Detector setting:** Includes radio buttons for 'Gain' (selected) and 'Delta EMV', and a checkbox for 'EM Saver' with a 'Limit' dropdown menu.
- Run time:** A checkbox and a text box set to '1 min'.
- Solvent delay:** A text box set to '8.2 min'.
- Time filtering:** Includes a checked checkbox for 'Peak width' (0.7 sec) and another checked checkbox for 'Automatically subtract baseline'.

W okienku "MS SIM/Scan Parameters", scharakteryzowane są warunki pracy spektrometru mas. "Tryb skanowania" używany jest w celu zarejestrowania całkowitego prądu jonowego (TIC - *total ion current*) natomiast „tryb SIM” w celu monitorowania wybranych jonów (SIM – *selected ion monitoring*).

Tryb skanowania

Pełny tryb skanowania stosowany jest w celu uzyskania pełnego spektrum analizowanych związków w zakresie m/z 40-750.

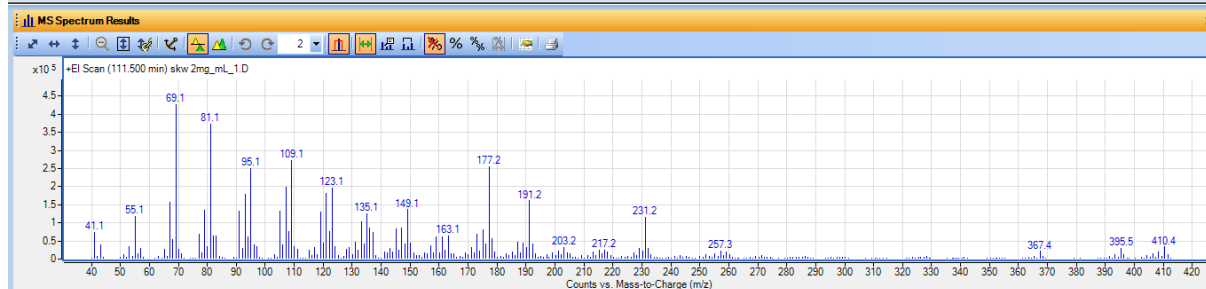
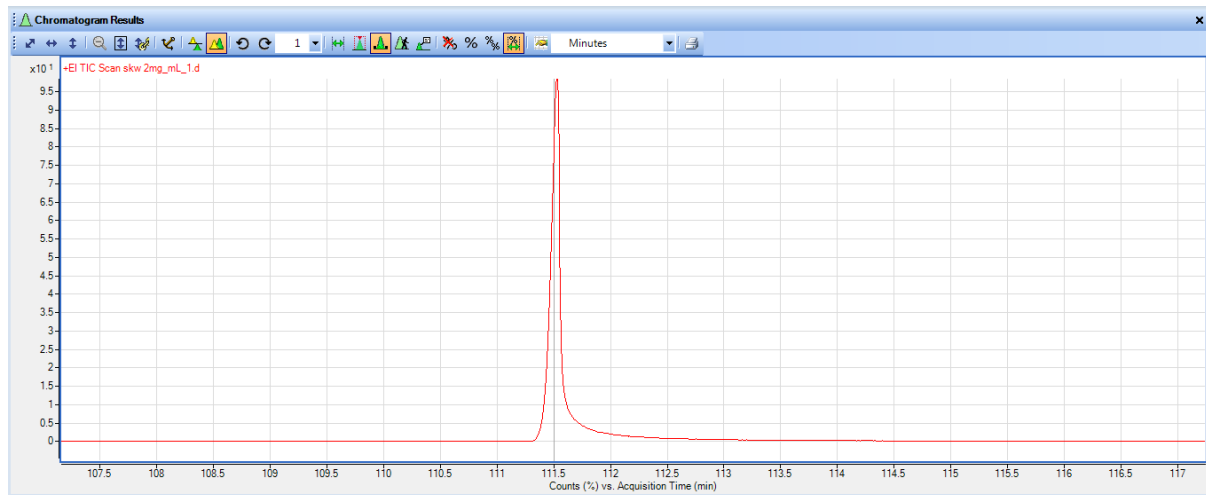


The screenshot shows a table titled 'Scan segments' with the following data:

Segment name	Start mass	End mass	Scan time (ms)	Data Samples	Expected Scan Time (ms)
	40	750	476	4	466

c) Analiza jakościowa

1. Zarejestruj chromatogram standardu skwalenu o stężeniu np. 2 mg/ml (w heksanie) wykorzystując powyższą metodę skanowania.

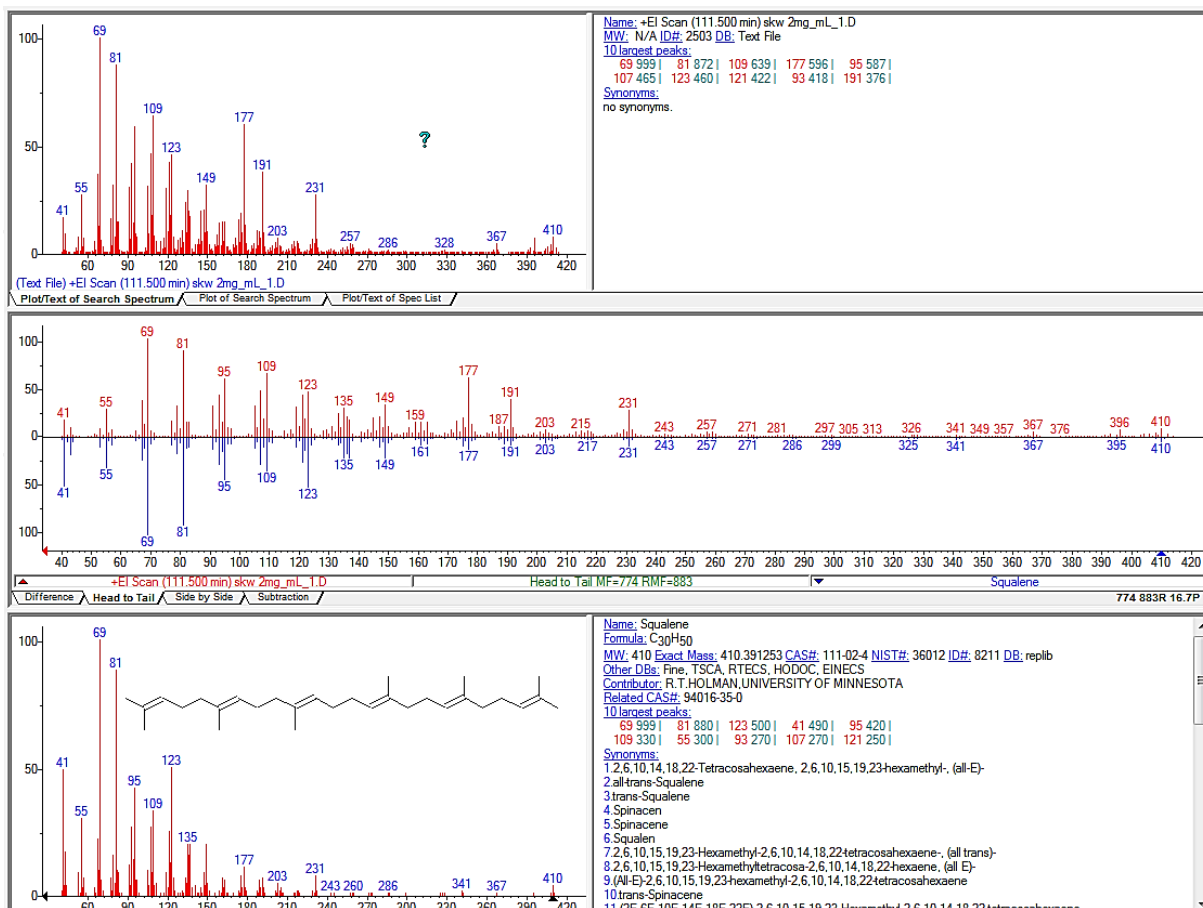


Chromatogram i widmo mas standardu skwalenu (TIC).

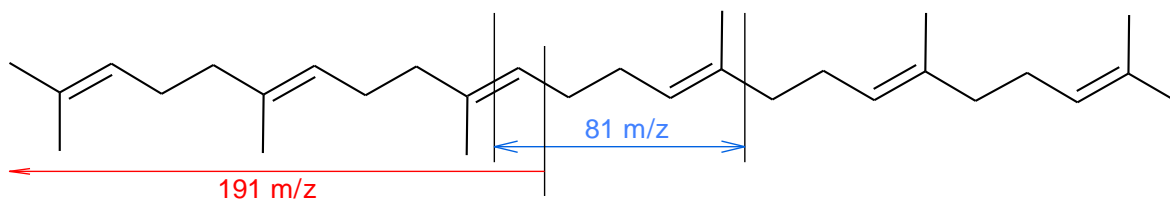
2. Przeszukiwanie biblioteki (NIST MS) dla otrzymanego widma; Skorzystaj z polecenia "Wyszukiwanie przy użyciu programu NIST"

Search Results Table:

#	Lib.	Match	R.Match	Prob. (%)	Name
1	R	774	883	16.7	Squalene
2	M	765	784	12.1	9,19-Cyclohexanediol-2,4-en-3-ol, acetate, (3S)-
3	R	757	763	9.05	Cyclohexane
4	M	750	790	6.93	9,19-Cyclohexanediol-2,4-en-3-ol, 4,14-dimethyl-
5	M	748	758	6.39	2,2,4a,8a,9,10,11,12,12a,12b,12c,12d,12e,12f,12g,12h,12i,12j,12k,12l,12m,12n,12o,12p,12q,12r,12s,12t,12u,12v,12w,12x,12y,12z,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000



3. Wybierz jony do analizy ilościowej (SIM): m/z 191 i 81.



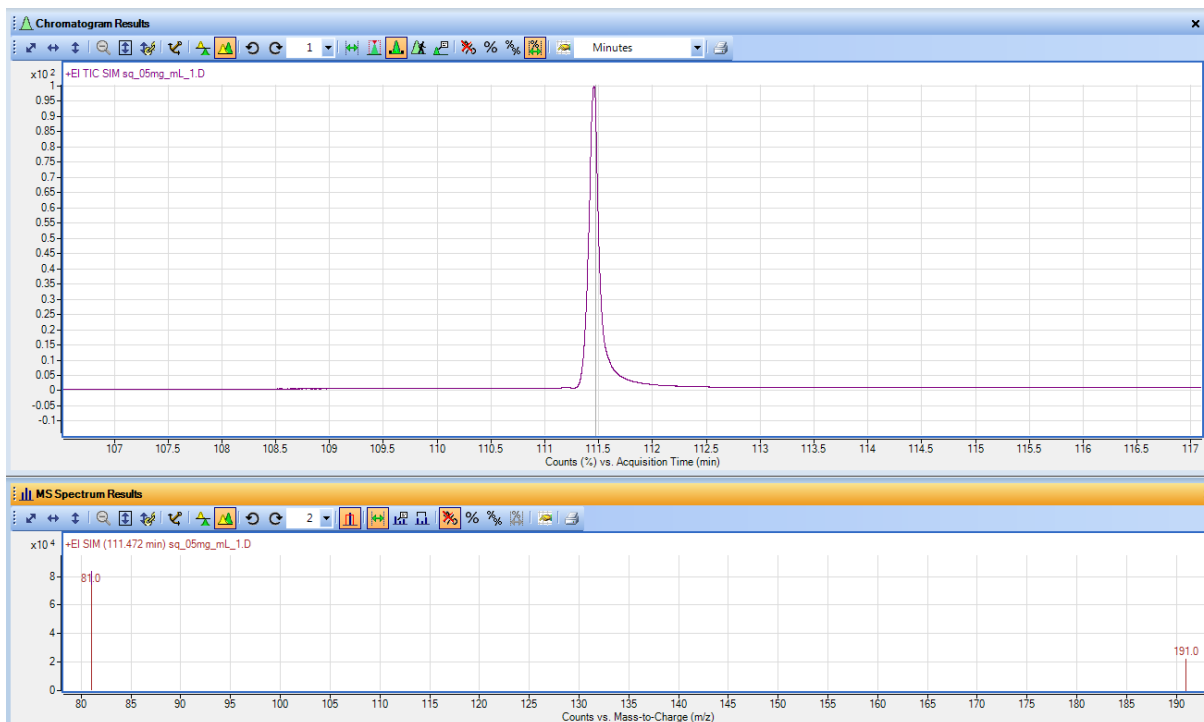
4. Utwórz metodę SIM-MS.

Acquisition					
Instrument		Chromatogram			
Scan segments					
	Compound name	ISTD?	Mass	MS1 resolution	Dwell
▶	squalene	<input type="checkbox"/>	191	Wide	15
	squalene	<input type="checkbox"/>	81	Wide	15

Tryb SIM

W tym trybie pracy spektrometru mas rejestrowane są zmiany intensywności wybranego jonu ("quantifier ion"). Tryb SIM jest najbardziej czułą metodą działania spektrometrów mas.

1. Korzystając z metody SIM zarejestruj chromatogram standardu skwalenu o stężeniu np. 0,5 mg/mL (w heksanie).



Chromatogram i widmo mas wzorca skwalenu otrzymane w trybie pracy SIM.

d) Analiza ilościowa (metoda standardu zewnętrznego - External Standard Method)

1. Korzystając z powyższej metody SIM zarejestruj chromatogramy dla wzorców skwalenu o następujących stężeniach: 0.5 mg/mL, 1.0 mg/mL, 2 mg/mL, 5 mg/mL oraz 10 mg/mL w heksanie.

Jon m/z 191 w metodzie SIM używany jest jako “qualifier ion” natomiast jon m/z 81 jako “quantifier ion” (“qualifier ion” używany jest wyłącznie do kontroli intensywności “quantifier ion”).

Quantifier					
Name	TS	Scan	Type	MZ	Uncertainty
squalene	1	SIM	Target	81.0	Relative

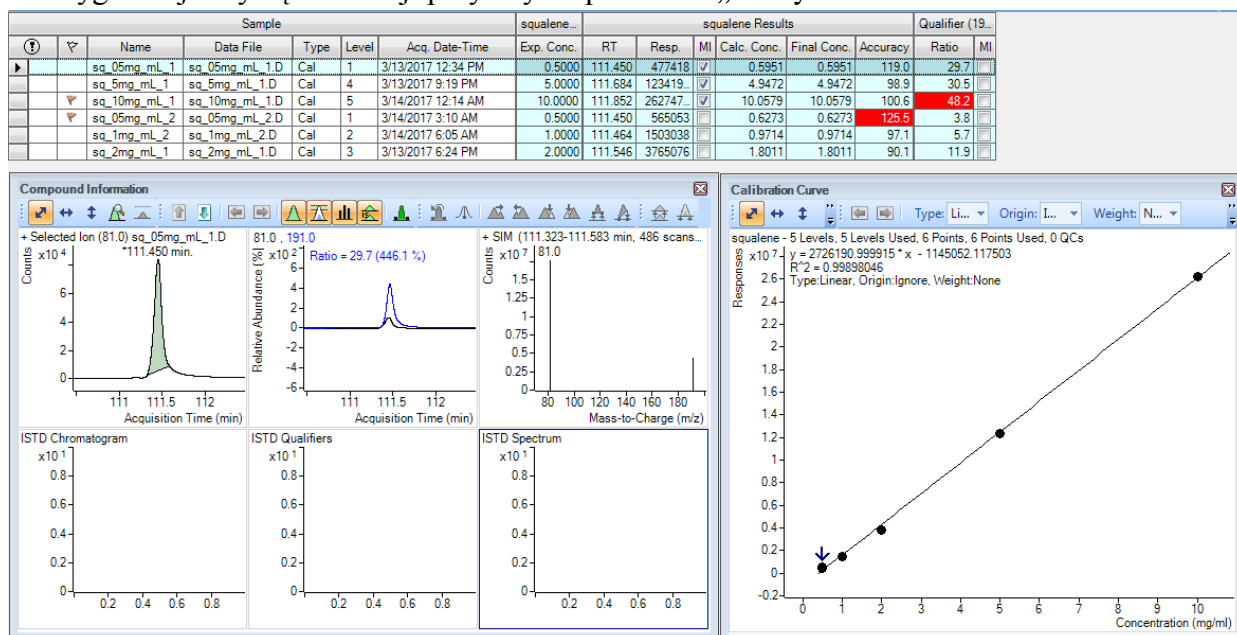
Qualifier			
MZ	Rel. Resp.	Uncertainty	Area Sum
191.0	6.7	20.0	

2. Utwórz poziomy kalibracji.

Quantifier				
Name	TS	Scan	Type	Units
squalene	1	SIM	Target	mg/ml

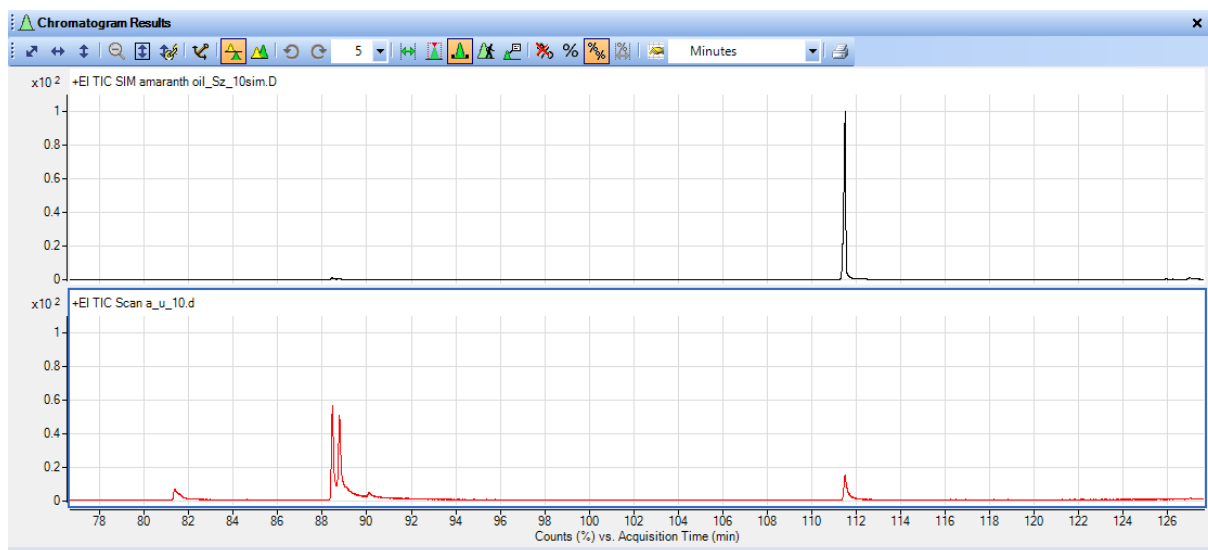
Calibration			
Level	Conc.	Response	Enable
1	0.5000		<input checked="" type="checkbox"/>
2	1.0000		<input checked="" type="checkbox"/>
3	2.0000		<input checked="" type="checkbox"/>
4	5.0000		<input checked="" type="checkbox"/>
5	10.0000		<input checked="" type="checkbox"/>

3. Wygeneruj krzywą kalibracji przy użyciu polecenia „Analyze Batch”.



4. Korzystając z powyższej metody SIM zarejestruj chromatogramy dla siliolowanych próbek olejów jadalnych (olej amarantusowy, olej z wiesiołka, olej rokitnikowy).

Poniżej przedstawiono porównanie chromatogramów zarejestrowanych dla oleju amarantusowego w trybie SIM i w trybie skanowania.



Chromatogramy zarejestrowane w trybie SIM i w trybie skanowania oleju amarantusowego.

Zaznacz wszystkie zarejestrowane chromatogramy próbek oleju, a następnie kliknij przycisk OK, aby dodać je do utworzonego wcześniej pliku do analizy ilościowej. Następnie kliknij ikonę „Analyze Batch”, aby uzyskać tabelę zawierającą wyniki analizy ilościowej.

3.5. Wykoananie analiz przy użyciu metody HPLC-MS oraz konfiguracja parametrów dla trybu data dependent MS/MS

Otrzymane w punkcie 1a ekstrakty lipidów zostaną poddane analizie z użyciem systemu LC-MS-QTOF w dependent MS/MS. Przed przystąpieniem do analizy kluczowe parametry wprowadzone do metody, takie jak poziom sygnału tła matrycy próbki, liczba wybieranych prekursorów czy szybkość skanowania, zostaną dokładnie przedyskutowane. Z kolei w trakcie trwania analiz podstawowe zagadnienia dotyczące budowy hybrydowego spektrometru mas typu QTOF jak również jego tryby pracy ze zwróceniem uwagi na tryb data dependent będą obszernie omówione. Dodatkowo znaczenie wprowadzania do metody list wartości m/z jonów wybieranych jak i wykluczanych z analizy zostanie również wyjaśnione.

1. Przeprowadź analizę ekstraktów osocza w trybie data dependent używając metody "Phospholipid_HILIC_negative.m" po jej wybraniu z głównego menu kontekstowego programu Agilent Mass Hunter Acquisition. Parametry dotyczące modułu chromatografu cieczowego podane są na poniższych rycinach:

Flow: 0.030 ml/min

Solvents

A: 0.00 %
1: 100.0 % (Methanol/ACN 1) ACN:MEOH:
2: 100.0 % Water V.02

B: 100.00 %
1: 100.0 % (Methanol/ACN 1) ACN:MEOH_
2: 100.0 % Acetonitrile V.02

Pressure Limits
Min: 0.00 bar Max: 600.00 bar

Stoptime: 45.00 min As Injector/No Limit

Posttime: Off 1.00 min

+ Timetable (12/100 events)

Time [min]	A [%]	B [%]	Flow [ml/min]	Max. Pressure Limit [bar]
0.00	0.00	100.00	0.030	600.00
8.00	0.00	100.00	0.030	600.00
15.00	40.00	60.00	0.030	600.00
40.00	0.00	100.00	0.030	600.00
45.00	0.00	100.00	0.030	600.00

2. Omów i przedyskutuj wartości parametrów źródła ESI po ich wprowadzeniu do metody (zakładka „source”):

W zakładce “Acquisition” wybierz typ analizy “Auto MS/MS”:

On	Prec. m/z	Delta m/z (ppm)	Z	Prec. Type	Ret. Time	Delta Ret. Time (min)	Iso. Width	Collision Energy
<input checked="" type="checkbox"/>	679.76	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	649.71	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	612.19	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	611.17	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	610.14	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	476.42	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	475.53	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	474.32	100	1	Exclude	0		Narrow (~1.3 m/z)	

3. Przedyskutuj i wyjaśnij znaczenie kluczowych parametrów dla analizy typu data dependent zadeklarowanych w metodzie takich jak:

- liczba wybieranych jonów prekursora

Binary Pump Column Comp. DAD Q-TOF

General Source Acquisition Ref Mass Chromatogram

Spectral Parameters Collision Energy **Precursor Selection I** Precursor Selection II Preferred/Exclude

Mode:
 MS (Seg)
 Auto MS/MS (Seg)
 Targeted MS/MS (Seg)

3 Max Precursor Per Cycle

Precursor Threshold
 Abs. Threshold 1000 counts
 Rel. Threshold (%) 0.01 %

Active Exclusion
 Enabled
 Excluded after 3 Spectra
 Released after 0.7 min

Static Exclusion Range List
 Static Exclusion Range Table

Start m/z	End m/z
50	400

- model izotopowy

Binary Pump Column Comp. DAD Q-TOF

General Source Acquisition Ref Mass Chromatogram

Spectral Parameters Collision Energy **Precursor Selection I** Precursor Selection II Preferred/Exclude

Mode:
 MS (Seg)
 Auto MS/MS (Seg)
 Targeted MS/MS (Seg)

Precursor Charge-State Selection and Preference

Inactive: 2, 3, >3, Unk
 Active: 1
 Buttons: >>, <<, Reset, Up, Down

Abundance Dependent Accumulation
 Scan speed varied based on precursor abundance
 Target 25000 counts/spectrum
 Use MS/MS accumulation time limit
 Reject precursors that cannot reach target TIC within time limit

Purity
 Purity Stringency: 100 %
 Purity Cutoff: 30 %

Isotope Model: **Common Organic Molecules**

Sort Precursors by Charge State then Abundance
 Sort Precursors by Abundance only

- wartości m/z jonów preferencyjnie wybieranych do analizy jak i eliminowanych

Binary Pump Column Comp. DAD Q-TOF

General Source Acquisition Ref Mass Chromatogram

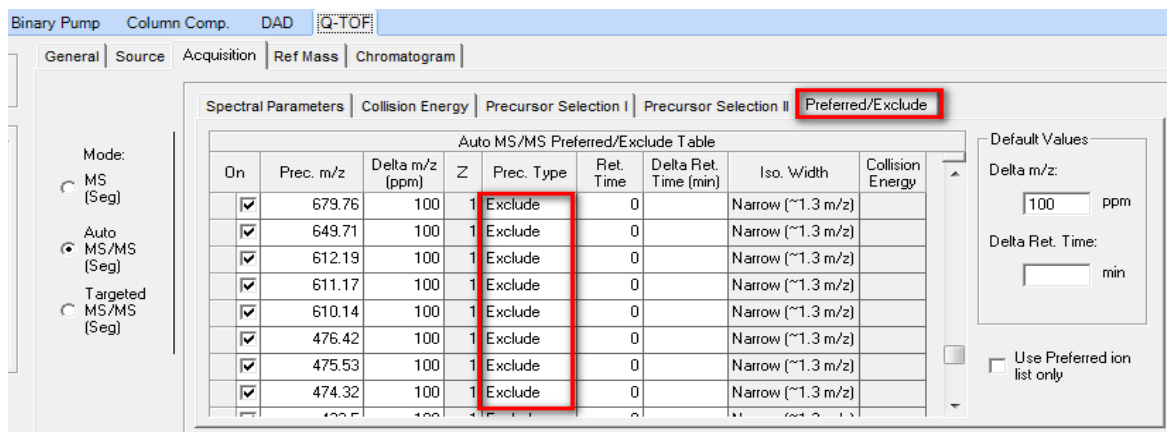
Spectral Parameters Collision Energy Precursor Selection I Precursor Selection II **Preferred/Exclude**

Mode:
 MS (Seg)
 Auto MS/MS (Seg)
 Targeted MS/MS (Seg)

Auto MS/MS Preferred/Exclude Table

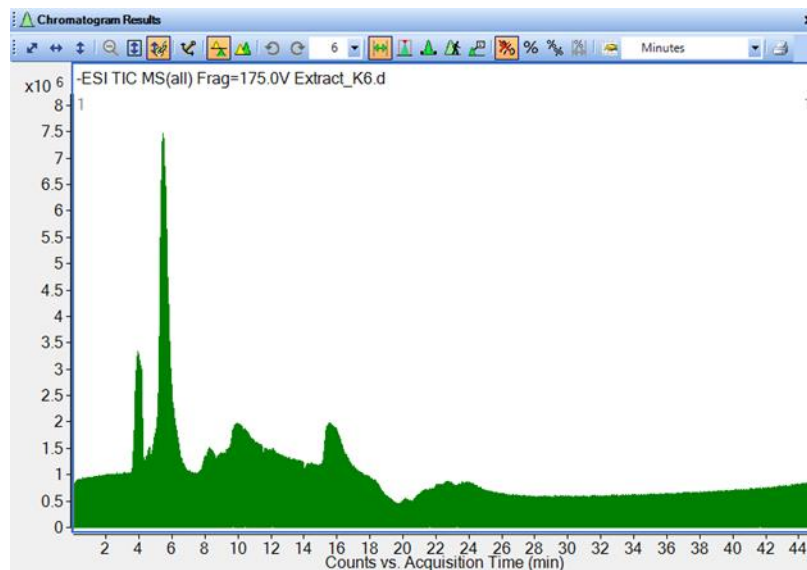
On	Prec. m/z	Delta m/z (ppm)	Z	Prec. Type	Ret. Time	Delta Ret. Time (min)	Iso. Width	Collision Energy
<input checked="" type="checkbox"/>	906.6	100	1	Preferred	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	880.6	100	1	Preferred	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	880.5	100	1	Preferred	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	878.6	100	1	Preferred	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	878.5	100	1	Preferred	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	876.5	100	1	Preferred	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	874.6	100	1	Preferred	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	874.5	100	1	Preferred	0		Narrow (~1.3 m/z)	

Default Values
 Delta m/z: 100 ppm
 Delta Ret. Time: min
 Use Preferred ion list only



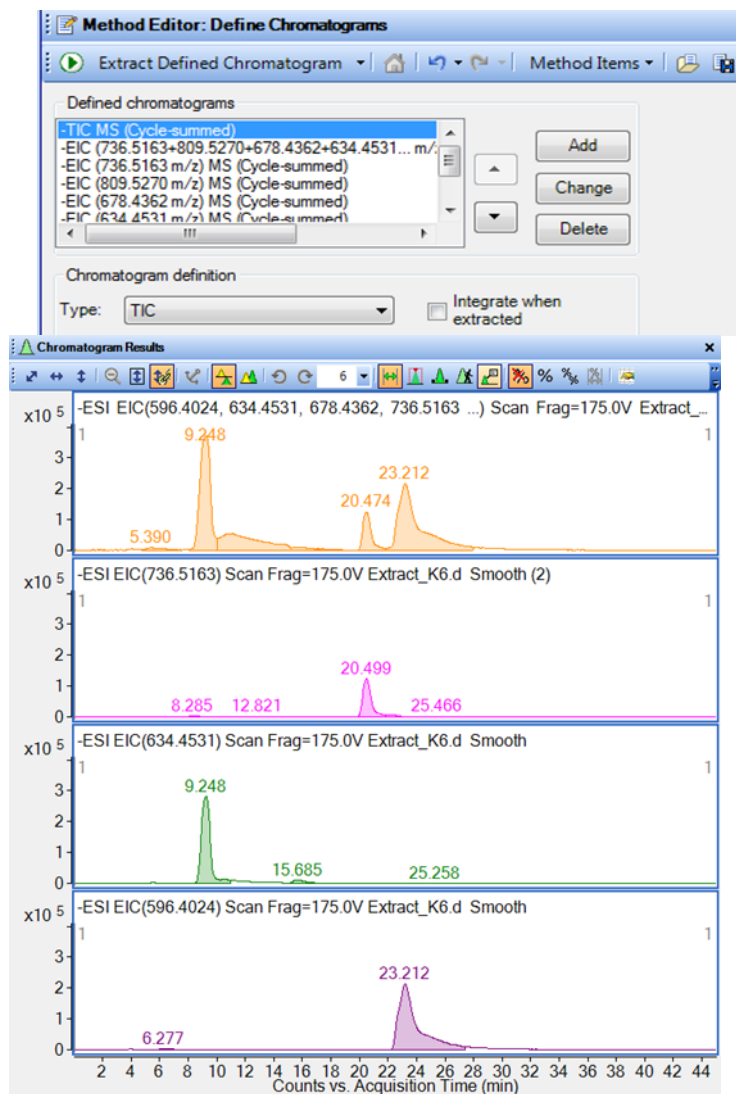
Jaki jest powód i znaczenie tworzenie w metodzie list m/z jonów prekursorów wybieganych i eliminowanych z analiz w trybie data dependent? Jakie parametry metody analitycznej ulegają znacznemu polepszeniu?

- Wygeneruj chromatogram dla wszystkich jonów (TIC) w celu uzyskania profile fosfolipidowego każdej przeanalizowanej próbki osocza.



- Następnie wygeneruj chromatogramy (EIC) dla jonów standardów wewnętrznych poszczególnych klas fosfolipidów w celu ustalenia zakresu czasów retencji odpowiadających fosfolipidom danej klasy.

Użyj w tym celu komendy “Extract defined chromatogram” w edytorze metody program Mass Hunter Qualitative.



6. Wygeneruj listę związków znalezionych przez software dla każdej z analiz. W tym celu użyj funkcji “Find compounds by Auto MS/MS” z menu kontekstowego Mass Huntera Qualitative.

Compound List

Automatically Show Columns

Show/Hide	Cpd	File	RT	m/z
<input checked="" type="checkbox"/>	1	Extract_K6.d	0.062	494.3586
<input checked="" type="checkbox"/>	2	Extract_K6.d	0.811	819.5143
<input checked="" type="checkbox"/>	3	Extract_K6.d	0.822	718.555
<input checked="" type="checkbox"/>	4	Extract_K6.d	0.946	656.8809
<input checked="" type="checkbox"/>	5	Extract_K6.d	1.034	764.5413
<input checked="" type="checkbox"/>	6	Extract_K6.d	1.119	690.5284
<input checked="" type="checkbox"/>	7	Extract_K6.d	1.18	906.5547
<input checked="" type="checkbox"/>	8	Extract_K6.d	1.229	824.4565
<input checked="" type="checkbox"/>	9	Extract_K6.d	1.362	482.9254
<input checked="" type="checkbox"/>	10	Extract_K6.d	1.465	848.5168
<input checked="" type="checkbox"/>	11	Extract_K6.d	1.506	621.4743
<input checked="" type="checkbox"/>	12	Extract_K6.d	1.721	690.5607
<input checked="" type="checkbox"/>	13	Extract_K6.d	1.75	819.5143
<input checked="" type="checkbox"/>	14	Extract_K6.d	1.899	706.4321
<input checked="" type="checkbox"/>	15	Extract_K6.d	2.266	848.4113
<input checked="" type="checkbox"/>	16	Extract_K6.d	2.41	876.4235
<input checked="" type="checkbox"/>	17	Extract_K6.d	2.416	522.4201
<input checked="" type="checkbox"/>	18	Extract_K6.d	2.465	718.5351
<input checked="" type="checkbox"/>	19	Extract_K6.d	2.829	819.5143
<input checked="" type="checkbox"/>	20	Extract_K6.d	2.866	880.5361
<input checked="" type="checkbox"/>	21	Extract_K6.d	3.125	718.555
<input checked="" type="checkbox"/>	22	Extract_K6.d	3.15	611.3993

Context Menu:

- Spectrum
- General
- Reports
- Find Compounds**
 - Find by Auto MS/MS
 - Find by Targeted MS/MS
 - Find by Molecular Feature
 - Find by MRM
- Find Compounds by Formula
- Identify Compounds
- Compound Automation Steps
- Worklist Automation
- Export

7. Na wygenerowanej liście odszukaj związek o znanej wartości m/z np. przy m/z 591.4036.
8. Przypisz wartość czasu retencji wyselekcjonowanemu związkowi.
9. Następnie odszukaj charakterystyczne jony na widmie fragmentacyjnym związku wskazanego w punkcie 7 i przypisać im ich wartości m/z .



Literatura:

- E. de Hoffmann, V. Stroobant. Mass spectrometry: principles and applications. Wiley, 2007.
- H.J. Hübschmann. Handbook of GC-MS: Fundamentals and Applications (3rd Edition). Wiley, 2015.
- J. Kool, W.M. Niessen. Analyzing biomolecular interactions by mass spectrometry. Wiley, 2015.

Moduł 2

Metabolomika

Coral Barbas, Danuta Dudzik, M^a Fernanda Rey-Stolle, Francisco J. Rupérez,
Antonia Garcia

Centre for Metabolomics and Bioanalysis (CEMBIO), Facultad de Farmacia, Universidad San Pablo CEU; Madrid, Spain

I. Wstęp

Celem kursu jest umożliwienie studentom studiów doktoranckich rozwoju wiedzy i umiejętności w nowej dyscyplinie jaką są nauki ometyczne. Metabolomika zajmuje się analizą systematyczną metabolitów wewnątrz- i zewnątrzkomórkowych. Celem metabolomiki jest dostarczenie wiedzy na temat zmian metabolicznych oraz mechanizmu molekularnego leżącego u podstaw rozwoju chorób, a także odkrycie nowych, klinicznych biomarkerów diagnostycznych i prognostycznych. Moduł 2 podzielony jest na cztery sekcje, obejmujące główne aspekty metabolomiki: 1) cele i definicje, 2) stosowane podejścia, w tym planowanie oraz warunki eksperymentu, narzędzia analityczne, kontrola jakości i cykl pracy w metabolomice, 3) wstępna obróbka danych, przetwarzanie danych, w tym analiza statystyczna i identyfikacja biomarkerów; oraz w ostatniej części, 4) interpretacja biochemiczna, analiza szlaków biochemicznych oraz walidacja biomarkerów przy użyciu różnych programów.

Treści kursu będą realizowane w formie wykładów, warsztatów, jak i ćwiczeń praktycznych z danymi, przy wykorzystaniu obecnie dostępnych bezpłatnych zasobów bioinformatycznych. Zastosowane podejście pozwoli na otrzymanie wymiernych rezultatów.

II. Założenia i efekty kształcenia

Założenia

Celem kursu jest zrozumienie przez studentów podstawowych pojęć i metodologii badań metabolomicznych. Studenci zdobędą podstawową wiedzę teoretyczną i praktyczną w zakresie metabolomiki, w tym koncepcji pipeline, wymagań i podejmowania odpowiednich działań, analizy danych metabolicznych, wstępnego przetwarzania danych, statystyki, identyfikacji metabolitów oraz ich znaczenia biologicznego. Poza tym student będzie potrafił przeprowadzić walidację biomarkerów oraz ustalić hipotezę biologiczną.

Specyficzne efekty kształcenia:

Po ukończeniu kursu studenci powinni potrafić:

1. definiować i stosować podstawową terminologię związaną z metabolomiką.
2. omawiać w sposób kompleksowy, różne mechanizmy rozdziału, w połączeniu z techniką MS w zakresie metabolomiki.
3. określać wszystkie etapy badań metabolomicznych.
4. wybierać, pomiędzy różnymi podejściami, w celu rozwiązania konkretnego problemu.
5. znać i rozumieć różne metody przygotowania próbek oraz ograniczenia w badaniach metabolomicznych.
6. znać i rozumieć etapy przetwarzania danych.

7. umiejętnie wykorzystywać darmowe oprogramowanie w celu przetwarzania danych i analizy szlaków biologicznych.
8. znać różne podejścia do analizy statystycznej stosowanej w metabolomice.
9. zdobywać podstawowe umiejętności w zakresie wykorzystania bezpłatnych baz danych.
10. wyjaśniać tym, którzy nie są specjalistami, jak dyscyplina jaką są nauki "omiczne" dostarcza cennych informacji w różnych dziedzinach nauki.
11. wyciągać oraz uzasadniać wnioski, jasno i jednoznacznie zarówno dla specjalistów, jak i nie.
12. kontynuować proces uczenia się, w dużej mierze autonomicznie.

III. Spis treści

1. Wstęp do metabolomiki
2. Podejścia analityczne w metabolomice
 - 2.1. Cykl pracy w badaniach metabolomicznych
 - 2.1.1. Projekt eksperymentu
 - 2.1.2. Pobieranie próbek
 - 2.1.3. Ekstrakcja metabolitów
 - 2.1.4. Rozdział i detekcja
 - 2.2. Kontrola jakości i procedura zapewnienia jakości w metabolomice
3. Przetwarzanie danych i identyfikacja metabolitów
 - 3.1. Przetwarzanie danych koncepcja pipeline
 - 3.1.1. Przetwarzanie surowych danych
 - 3.1.2. Wstępne przetwarzanie danych
 - 3.1.3. Wstępna obróbka danych
 - 3.2. Identyfikacja metabolitów
 - 3.2.1. Identyfikacja za pomocą GC-EI-MS
 - 3.2.2. Identyfikacja metabolitów za pomocą LC lub CE-(ESI)MS: Praca z bazami danych
 - 3.2.3. Bazy danych metabolitów, klasyfikacja w oparciu o wartość masy cząsteczkowej
 - 3.2.4. Bazy danych metabolitów, klasyfikacja w oparciu o widma masowe
 - 3.2.5. Informatyczne bazy danych
 - 3.3. Analiza statystyczna
 - 3.3.1. PCA
 - 3.3.2. PLS-DA
 - 3.3.3. OPLS-DA
 - 3.3.4. Wykres-S, metoda Jack-knife i test projekcyjny VIP
 - 3.3.5. Walidacja statystycznych modeli wielowymiarowych
4. Analiza danych
 - 4.1. Od identyfikacji danych do szlaków biochemicznych: klasyfikacja danych w celu oceny istotności biologicznych.
 - 4.1.1. Bazy danych szlaków biochemicznych
 - 4.1.2. Analiza szlaków biochemicznych
 - 4.1.3. Wzbogacanie metabolitów
 - 4.1.4. Topologiczna analiza szlaków biologicznych

- 4.2. Walidacja biomarkerów: ustalenie istotności metabolitów
- 5. Zajęcia praktyczne z metabolomiki.
 - 5.1. Metabolomika targetowa i nietargetowa
 - 5.2. Wstępna obróbka danych przy użyciu platformy Workflow for Metabolomics (W4M)
 - 5.2.1. Konwersja plików MS
 - 5.2.2. Kontrola surowych danych
 - 5.2.3. Prześlij dane do Galaxy
 - 5.2.4. Wstępne przetwarzanie danych LC-MS za pomocą XCMS.
 - 5.3. Statystyka
 - 5.4. Komentarze

1. Wstęp do metabolomiki

Metabolomika odnosi się do systematycznej analizy całego zestawu metabolitów wewnątrz- i zewnątrzkomórkowych w danym systemie biologicznym (w tym w systemach bakteryjnych, roślinnych i zwierzęcych). Metabolity są cząsteczkami o małych masach molowych (<1500 Da), które pochodzą od wielu składników biologicznych (np. białek, DNA, RNA). Powstają w wyniku różnych przemian biologicznych w odpowiedzi na proces patofizjologiczny. Metabolity uczestniczą w przemianach metabolicznych i są niezbędne do regulacji, wzrostu i prawidłowego funkcjonowania komórki. Biorą także udział w procesach transkrypcyjnych, translacyjnych i posttranslacyjnych. Dlatego też, metabolomika zapewnia całkowity obraz aktualnego stanu i fenotypu organizmu. Zrozumienie metabolomu, zdefiniowanego jako zestaw wszystkich metabolitów w komórce, tkankach lub organizmie jest kluczowe dla uzyskania kompleksowego obrazu funkcjonowania komórek, tkanek i organizmów. W większości przypadków metabolomika polega na badaniu zróżnicowania metabolomicznych odcisków palca (metabolomics fingerprinting), generowanych z podgrup "kontrolnych" i "badanych", w celu znalezienia różnic w ich profilach metabolicznych w odpowiedzi na bodźce zewnętrzne (patologie, działania biochemiczne lub środowiskowe, żywienie itp.). Podsumowanie specyficznych pojęć w dziedzinie metabolomiki, podano w tabeli 2.1.

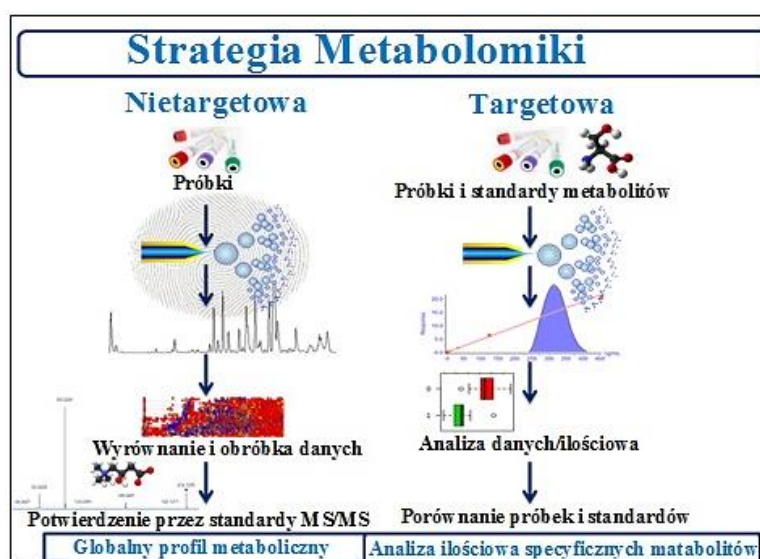
Metabolomika okazała się skuteczną, dobrze ugruntowaną i wartościową metodą badania zdrowia ludzkiego, chorób, procesu starzenia, stylu życia oraz odkrywania i udoskonalenia związków o działaniu terapeutycznym. Dlatego też, ostatecznym celem metabolomiki jest wgląd w zmiany metabolitów i mechanizmy molekularne leżące u podstaw stanu chorobowego oraz odkrycie nowych, klinicznych biomarkerów diagnostycznych oraz prognostycznych. Liczba badań obejmujących analizę metabolomu wzrasta z roku na rok wraz z ulepszeniami w zakresie technologii analitycznych.

Tabela 2.1. Definicje dotyczące metabolomiki.

Pojęcie	Definicja
Metabolomika	Badanie kompletnego zestawu metabolitów różniących się w zależności od fizjologii, rozwoju stanu patologicznego komórki, tkanki, narządu lub organizmu.
Metabolity	Małe cząsteczki, o niskiej masie molowej (<1500 Da), które wchodzą w skład struktury wielu składników biologicznych, niezbędnych do regulacji, wzrostu i normalnego funkcjonowania komórki.
Metabolom	Końcowy produkt genomu, kompletny zestaw wszystkich metabolitów o niskiej masie cząsteczkowej obecnych w komórce lub organizmie. Metabolom dzieli się na egzo-metabolom (metabolity zewnątrzkomórkowe) i endo-metabolom (metabolity wewnątrzkomórkowe).
„Metabotyp”	Fenotyp metaboliczny.
Odcisk palca	Bezstronne, globalne i wysoce wydajne podejście do klasyfikowania próbek opartych na modelu metabolomicznym lub "odciskach palców" zmieniających się w odpowiedzi na choroby, zaburzenia środowiskowe lub genetyczne, których głównym celem jest identyfikacja dyskryminujących metabolitów.
Profilowanie metaboliczne	Identyfikacja i analiza ilościowa lub półilościowa konkretnej grupy metabolitów, które związane są z konkretnym szlakiem metabolicznym.
Odcisk stopy	Analiza (egzo) metabolitów wydzielanych/wydalanych przez organizm w kontrolowanych warunkach.
Metabolomika nietargetowa	Zakres globalny, wolny od hipotez, w celu równoczesnego pomiaru jak największej ilości metabolitów z próbek biologicznych.
Metabolomika targetowa	Wiąże się z użyciem jednej metody analitycznej do oznaczenia ilościowego niewielkiego zestawu znanych metabolitów.
Szlak metaboliczny	Zestaw reakcji chemicznych zachodzących wewnątrz komórki.

2. Podejścia analityczne w metabolomice

W zależności od celu badań metabolomicznych można zastosować dwie strategie analityczne: podejście targetowe i nietargetowe (Rycina 2.1).



Rycina 2.1. Targetowe i nietargetowe podejście analityczne do badań metabolomicznych.

➤ **Podejście targetowe**

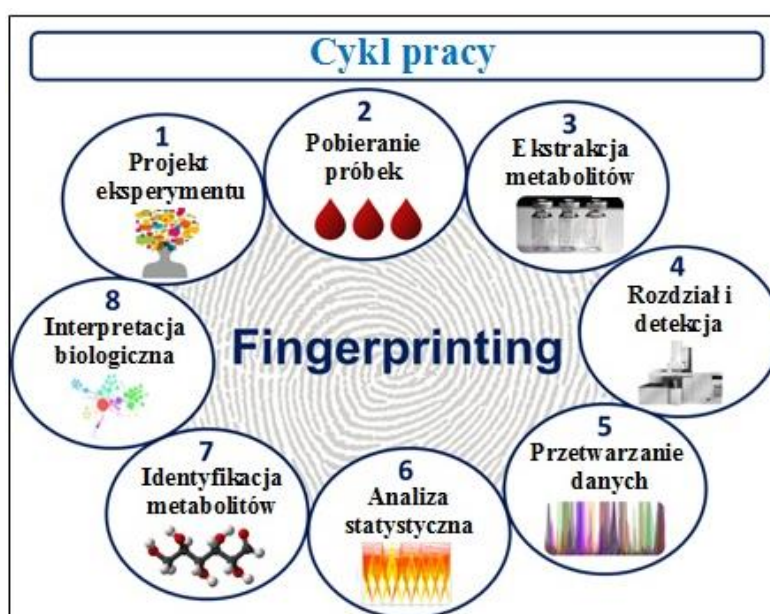
Metabolomika targetowa, odnosi się do badań mających na celu analizę konkretnych, znanych cząsteczek (typowo <20), dotyczących kilku (lub jednego), powiązanych szlaków metabolicznych, które w poprzednich badaniach zostały określone jako biologicznie istotne szlaki metaboliczne. W związku z tym stosuje się metody o wysokim poziomie specyficzności, precyzji i dokładności, w celu określenia ilości każdego metabolitu. Na przykład, ilościowe oznaczanie aminokwasów przy użyciu metody monitorowania reakcji z zastosowaniem spektrometru mas, potrójnego kwadrupolu (QQQ). Analizę ilościową badanych związków, przeprowadza się przy użyciu standardów wewnętrznych oraz porównania uzyskanych danych z krzywymi wzorcowymi.

➤ **Podejście nietargetowe**

Nietargetowe podejście do metabolomiki, odnosi się do koncepcji odcisku palca, posiada globalny zasięg, zaś główny cel stanowi jednoczesne wykrycie jak największej liczby metabolitów w każdej próbce, w ramach jednego nastrzyku próbki. W strategii nietargetowej, metabolity, które zostaną zidentyfikowane, nie są znane przed badaniem. Podejście to stosuje się do badań wolnych od hipotez, zaś celem jest zdefiniowanie nowych i wcześniej nieobserwowanych zmian w metabolomie. Może to, być związane, z biologiczną funkcją organizmu i mechanizmami leżącymi u podłoża zmian metabolicznych, spowodowanych rozwojem choroby. Podejście takie, prowadzi do powstania nowej hipotezy, która powinna zostać zweryfikowana, w późniejszym badaniu walidacyjnym, czyli w podejściu targetowym.

2.1. Cykl pracy w badaniach metabolomicznych

Typowy cykl pracy w nietargetowych badaniach metabolomicznych przedstawiono na rycinie 2.2.



Rycina 2.2 Cykl pracy analizy nietargetowej w badaniach metabolomicznych.

Najważniejsze punkty cyklu pracy w badaniach metabolomicznych to: (1) projekt eksperymentu; (2) pobieranie i przechowywanie próbek; (3) ekstrakcja metabolitów; (4) gromadzenie danych; (5,6) przetwarzanie i analiza danych; oraz (7,8) identyfikacja metabolitów, które umożliwią interpretację biologiczną. Każdy krok, ma duże znaczenie, dlatego też, należy stosować procedury zapewniające jakość analizy w celu zmniejszenia niepożądanych zmian preanalizacyjnych i eksperymentalnych oraz w celu zagwarantowania wiarygodności badań.

2.1.1. Projekt eksperymentu

Projekt eksperymentu opisuje w detalach, jak dane badanie, zostanie przeprowadzone. Powinien obejmować wszystkie istotne aspekty pracy, od pobrania próbek do końcowej analizy danych. Wszystkie próbki do badań wybierane są w sposób rozważny i przemyślany. Jeśli próbki zostały zebrane wcześniej (np. grupa próbek do badań epidemiologicznych) istotne jest, aby wiedzieć (1) jak zostały zebrane oraz jak są przechowywane; (2) jak można dopasować próbki kontrolne i badane; i (3) czy istnieją zależności między grupami kontrolnymi i eksperymentalnymi. Projekt eksperymentu powinien obejmować jasny opis liczby próbek i powtórzeń oraz opis badanych grup, aby zapewnić dane o wystarczającej mocy statystycznej. Badania oparte na modelu eksperymentalnym np. hodowle zwierząt lub hodowle komórkowe, w dobrze kontrolowanym środowisku laboratoryjnym, wymagają zwykle małego rozmiaru próbek (zazwyczaj 6-12), aby zidentyfikować istotne statystycznie zmiany. Badania oparte na populacji ludzkiej są bardziej złożone i obejmują znacznie więcej zmiennych, które mogłyby wpływać na jakość przeprowadzonych badań. Czynniki związane ze stylem życia (np. dieta, aktywność fizyczna), demograficzne (np. etniczność), fizjologiczne (np. płeć, wiek, BMI) oraz dane kliniczne (np. dane medyczne, leki) należy rozważyć i umożliwić ich zidentyfikowanie, unikając tym samym zmiennych zakłócających nasze badania. Skuteczny i starannie dobrany projekt eksperymentu jest pożądanym, aby można było dokładnie zmierzyć, nawet małe, istotne biologicznie zmiany w metabolomie. Zmiany te, będą znacznie łatwiej wykrywalne, niż zmiany w źle zaprojektowanym eksperymencie.

2.1.2. Pobieranie próbek

W czasie pobierania próbek należy stosować się do wymagań zdefiniowanych w projekcie. Należy zebrać wszystkie próbki w takich samych pojemnikach, dokładnie je zamknąć, podpisać i przechowywać w temperaturze -80°C . Rozmrażać próbki wolno, bezpośrednio przed analizą i przechowywać je w lodzie.

2.1.3. Ekstrakcja metabolitów

Aby wyniki badań były wiarygodne ekstrakcja metabolitów musi być kontrolowana. Metoda ekstrakcji nie jest skomplikowana, sposób ekstrakcji zależy od stosowanej matrycy próbki oraz systemu analitycznego.

LC-MS. W celu przeprowadzenia analizy LC-MS, próbki zawierające białka wymagają etapu deproteinizacji, który obejmuje proces usuwania białek. W celu wytrącenia białek dodaje się, zazwyczaj zimny, organiczny rozpuszczalnik (np. metanol, acetonitryl) lub mieszaninę rozpuszczalników (np. metanol/etanol (1:1, obj./obj.)). Tak wytrącone białka, oddziela się poprzez wirowanie. Otrzymany supernatant, zawiera wyekstrahowane metabolity i jest odpowiedni do analizy LC-MS.

GC-MS. W przypadku GC-MS, poza deproteinizacją, konieczne są bardziej skomplikowane procedury, ponieważ metoda GC jest niekompatybilna z nielotnymi i termicznie labilnymi związkami. Procedury te, obejmują ochronę grupy funkcyjnej-ketonowej przez metoksylicję i derywatyzację w celu zwiększenia lotności metabolitów. Proces derywatyzacji obniża temperaturę wrzenia wielu endogennych metabolitów, co czyni je wystarczająco lotnymi, aby umożliwić rozdział chromatograficzny. Związki zawierające grupy funkcyjne z aktywnymi atomami wodoru, takie jak -SH (tiole), -OH (alkohole, poliole, fenole, enole), -NH (aminy, amidy) i -COOH (kwasy karboksylowe) są najczęściej poddawane analizie GC-MS, ze względu na tendencję tych grup funkcyjnych do tworzenia wiązań wodorowych międzycząsteczkowych. Wiązania wodorowe wpływają na: lotność związków, oddziaływania związków z materiałami do pakowania kolumn oraz na stabilność cieplną związków. Po deproteinizacji, próbki należy osuszyć przed dodaniem jakichkolwiek środków derywatyzujących. Metoksylicja wykorzystuje chlorowodorek O-metoksyaminy w pirydynie w celu stabilizacji grup karbonylowych poprzez zahamowanie tautomerizacji ketonolu i utworzenie struktur acetylowych lub ketonowych. Wpływa również na zmniejszenie liczby pochodnych cukrów redukujących i generują dwie postacie pochodnej -N = C <: syn/anty. Kolejnym etapem jest proces silylowania, gdzie następuje zastąpienie aktywnych atomów wodoru grupą alkylsilylową, na przykład trimetylosilylową (-SiMe₃). Modyfikacja grupy funkcyjnej cząsteczki przez chemiczną derywatyzację umożliwia analizę wielu związków polarnych, zwiększając ich lotność, stabilność termiczną i symetrię piku w chromatografii.

CE-MS. Typowe przygotowanie próbek do analizy CE-MS składa się zazwyczaj z dwóch etapów, rozcieńczenia próbki i deproteinizacji. Dla lepszej czułości, siła jonowa w roztworze próbki powinna być niższa niż elektrolitu wypełniającego kapilarę.

2.1.4. Rozdział i detekcja

Wybór platformy analitycznej może znacząco wpłynąć na wyniki w badaniach nietargetowych metabolomu. Spektrometria mas (MS) i jądrowy rezonans magnetyczny (NMR) są najczęściej stosowanymi metodami detekcji w analizie metabolomu. NMR jest bardzo użyteczny w charakterystyce struktury nieznanymi związków, jednakże charakteryzuje się niską czułością oraz koszty zakupu aparatury, są znacznie wyższe niż aparatury do MS. Do najważniejszych zalety techniki MS możemy zaliczyć wysoką czułość i wydajność. Metabolom składa się z bardzo różnorodnych związków chemicznych, które sprawiają, że praktycznie niemożliwe jest jednoczesne oznaczenie całego metabolomu obecnego w próbce za pomocą pojedynczej platformy analitycznej. Dlatego połączenie technik separacyjnych (chromatografii cieczowej, chromatografii gazowej i elektroforezy kapilarnej) z MS znacząco zwiększa zdolność analizy chemicznej bardzo złożonych próbek biologicznych i umożliwia uzyskanie pełniejszego obrazu metabolomu. GC-MS oferuje zdolność do analizowania lotnych metabolitów, tłuszczów i estrów o niskiej polarności oraz metabolitów aminokwasów i kwasów organicznych o wysokiej polarności poddanych derywatyzacji w lotne pochodne, podczas gdy niepolarne i średnio polarne metabolity o wysokiej masie cząsteczkowej (np. sfingolipidy, glicerofosfolipidy) analizuje się stosując platformę LC-MS. Dodatkowo CE-MS pozwala na komplementarne podejście do LC-MS i analizuje aniony, kationy oraz cząsteczki neutralne. Wymienione wyżej metody analityczne zostały opracowane i poddane walidacji w celu zapewnienia powtarzalnych wyników i kompletnych danych. Tabela 2.2 przedstawia

główne cechy technik separacyjnych w połączeniu ze spektrometrią mas w badaniach metabolomicznych.

Tabela 2.2. Techniki separacyjne połączone ze spektrometrią mas w badaniach metabolomicznych.

Technika analityczna	Aplikacyjność	Zalety	Wady
GC-MS	Rozdział, identyfikacja oraz analiza ilościowa lotnych i termicznie stabilnych metabolitów, mniej polarnych.	Wysoka rozdzielczość chromatograficzna, dostępność bibliotek wielkoformatowych do identyfikacji metabolitów.	Niezdolność do analizowania metabolitów termo-labilnych i wysokocząsteczkowych, wymagana derywatywacja, nielotnych metabolitów.
LC-MS	Rozdział, identyfikacja oraz analiza ilościowa szerokiej grupy metabolitów, w zależności od typu kolumny i fazy ruchomej	Wysoka czułość, duża pojemność próbki, derywatywacja nie jest wymagana, zdolność do analizy związków termo-labilnych.	Ograniczona dostępność bibliotek komercyjnych, eluentów LC, efekt matrycy, ograniczony potencjał identyfikacyjny, chyba że, stosowana jest technika MS-MS.
CE-MS	Rozdział, identyfikacja oraz analiza ilościowa polarnych i zjonizowanych metabolitów, przy użyciu zredukowanej objętości próbki.	Wysoka rozdzielczość i szybka analiza, użyteczna metoda dla złożonych próbek biologicznych, nawet dla małych objętości próbek.	Ograniczona dostępność bibliotek komercyjnych, niekompatybilność buforów, granice detekcji, ograniczony potencjał identyfikacyjny, chyba że, stosowana jest technika MS-MS.

Literatura:

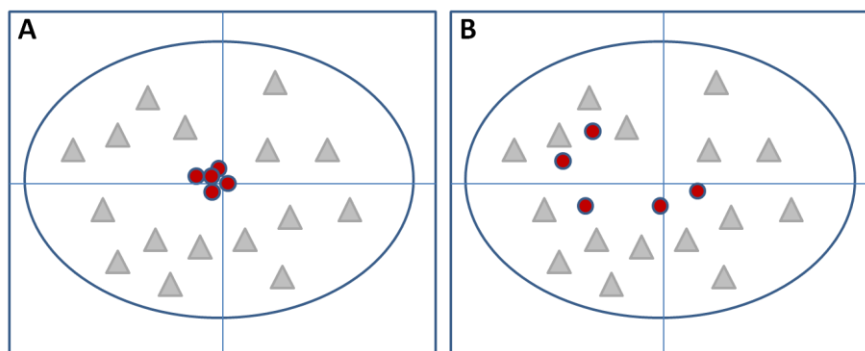
S.G. Villas-Bôas, S. Mas, M. Åkesson, J. Smedsgaard, and J. Nielsen, Mass spectrometry in metabolome analysis. Mass spectrometry reviews 24 (2005) 613-646.

2.2. Kontrola jakości i procedura zapewnienia jakości w metabolomice

Zapewnienie jakości (QA) wymagane jest w przypadku wszystkich strategii analitycznych (targetowa i nietargetowa), ale szczególnie w przypadku badań na dużą skalę. W badaniach targetowych, stosując platformy MS, kontrola jakości próbki (QC) stosowana jest rutynowo do określania dokładności i precyzji metody analitycznej. W kontekście metabolomiki, próbki QC stosowane są do: (1) kontroli działania systemu analitycznego; (2) zrównoważenia systemu w celu uzyskania w pełni powtarzalnych warunków; (3) poprawy małych odchyłeń mierzonego sygnału w analizie (w partiach) oraz pomiędzy partiami analitycznymi; (4) zintegrowania danych z różnych partii analitycznych; (5) obliczenia precyzji pomiarów metabolitów, z powtórzonych nastrzyków QC.

Próbki QC, analizowane są na początku i końcu przebiegu procesu analitycznego, a także w regularnych odstępach czasu w trakcie analizy, w celu monitorowania stabilności i powtarzalności procesu analitycznego. Próbki QC powinny być reprezentatywne dla

jakościowego i ilościowego składu analizowanych próbek. Z założenia, próbka QC powinna zawierać średnie stężenie wszystkich składników, które obecne są w badanych próbkach. Losowe i systematyczne źródła zmian, będą miały wpływ na powtarzalność danych w eksperymencie. Opracowano kilka metod pozyskiwania danych opartych na wydajności kontroli jakości. Ocena jakości danych obejmuje kontrolę precyzji analitycznej w oparciu o (1) badanie uzyskanych, nieobrobionych danych; (2) wykreślenie sumy intensywności cech metabolicznych każdej próbki eksperymentalnej i próbki QC; (3) sprawdzenie dryfowania sygnału, utraty czułości lub zmienności w próbkach QC w modelu wielowymiarowym (analiza głównych składowych, PCA-X); (4) obliczenie współczynnika zmienności CV (% RSD) dla każdej cechy metabolicznej wykrytej w próbkach QC. Obliczenie CV każdego metabolitu (możliwego metabolitu) dla wszystkich próbek QC w jednym przebiegu, zapewnia ilościowy pomiar zmienności. Wartości dopuszczalne CV, niższe niż 20% dla danych LC-MS i CE-MS oraz niższe niż 30% dla GC-MS, umożliwią filtrację złożonej matrycy danych i usunięcie metabolitów wykazujących odchylenia. Na rycinie 2.3 przedstawiono możliwe wyniki przewidywania próbek QC w wielowymiarowym modelu PCA-X.



Rycina 2.3. Schematyczne przedstawienie predykcji próbek QC w wielowymiarowym modelu PCA-X. Panel A. Można zaobserwować szczelne zgrupowanie próbek QC, co wskazuje na dokładne wyniki analityczne. Panel B. Dyspersja próbek QC wskazuje na utratę stabilności i powtarzalności systemu w trakcie analizy.

3. Przetwarzanie danych i identyfikacja metabolitów

Badania metabolomiczne prowadzą do powstania złożonych zbiorów danych, zawierających setki do tysięcy metabolitów. Poprawna obróbka danych jest podstawowym krokiem, który ma olbrzymi wpływ na zakres i jakość wyników metabolomicznych. Jednak przetwarzanie danych metabolomicznych, obróbka danych i identyfikacja metabolitów jest prawdopodobnie jednym z najbardziej wymagających etapów ze względu na pewne wewnętrzne cechy: sygnały analityczne zawierają wiele szumów oraz istnieje wiele skorelowanych zmiennych w porównaniu do liczby osobników. Istnieją jednak zaawansowane narzędzia do wstępnego przetwarzania i analizy danych, zarówno komercyjne (np. Agilent Technologies), jak i bezpłatne platformy online (np. XCMS, MZmine, MetAlign, MetaboAnalyst, MeltDB lub niedawno opracowana W4M).

3.1. Przetwarzanie danych koncepcja pipeline

W badaniach metabolomicznych ważne jest oddzielenie interesujących zmian biologicznych od niepożądanych źródeł zmienności wprowadzonych do eksperymentu. Dlatego wszechstronna analiza danych dotycząca metabolizmu wymaga wyczerpującej analizy danych oraz specjalistycznego oprogramowania, które umożliwia analizę chemometryczną, bioinformatyczną i statystyczną. Ogólny cykl pracy, analizy danych metabolomicznych opiera się na 4 głównych etapach: (1) przetwarzanie surowych danych; (2) wstępne przetwarzanie danych; (3) wstępna obróbka danych i (4) obróbka danych (Rycina 2.4).



Rycina 2.4. Główne etapy analizy danych w cyklu pracy badań metabolomicznych.

Nietargetowe badania metabolomiczne charakteryzują się równoczesnym pomiarem dużej liczby metabolitów w każdej próbce. W rezultacie generowana jest duża ilość danych. Najważniejsze kwestie, które należy wziąć pod uwagę w nietargetowych badaniach metabolomicznych, związane są z dwoma kluczowymi czynnikami wpływającymi w znaczny sposób na nasze dane: biologiczna zmienność i błędy techniczne.

3.1.1. Przetwarzanie surowych danych

Po wygenerowaniu danych wyniki muszą zostać przetworzone tak, aby wyodrębnić informacje z surowych plików danych. Należy przekształcić dostępne dane, aby miały wartość naukową. Surowe dane, konwertowane są zarówno przy użyciu oprogramowania dostarczanego przez sprzedawców, jak i bezpłatnych platform elektronicznych (np. XCMS).

3.1.2. Wstępne przetwarzanie danych

Przetwarzanie wstępne danych dotyczy wszystkich czynności wpływających bezpośrednio na surowe dane przed dalszą analizą danych. Głównym celem, jest poprawa jakości sygnału i zmniejszenie możliwych błędów/zakłóceń obecnych w surowych danych. Ten krok obejmuje: (1) redukcję szumów; (2) detekcję pików i dekonwolucję; (3) wyrównanie; (4) imputację zagubionych danych; oraz (5) filtrację danych.

(1) Redukcja szumów

Korekcja linii bazowej, redukcja szumów, służy do usuwania artefaktów o niskiej częstotliwości i pozornych różnic, które generowane są przez protokół eksperymentalny i wariacje instrumentalne. Krok ten ułatwia detekcję pików i redukuje fałszywie pozytywne wyniki. Niektóre oprogramowania zintegrowały algorytmy filtrowania i detekcji pików w jedną funkcję.

(2) Detekcja pików i dekonwolucja

Celem detekcji pików oraz dekonwolucji jest identyfikacja i oznaczenie ilościowe sygnałów z widm pochodzących od wykrytych związków w próbce.

Każdy pik w chromatogramie lub elektroforegramie przedstawia wykres zależności intensywności (abundance) na osi y od czasu retencji (RT) lub czasu migracji (MT) na osi x. Intensywność mierzona jest w oknie czasu detekcji w każdym skanowaniu jonów, w ustalonym zakresie mas. Poza tym, uzyskane widmo masowe, oferuje podwójną informację, poza czasem retencji. W skomplikowanych próbkach, koelucja różnych związków powoduje iż widzimy tylko pojedynczy pik, a ponadto widmo masowe pików jest sumą wszystkich związków koeluujących. Dekonwolucja to proces oddzielania składników koeluujących i tworzenia czystego widma dla każdego składnika poprzez ekstrakcję jonów z całkowitego prądu jonowego (TIC), nawet jeśli sygnał związku targetowego jest na poziomie śladowym.

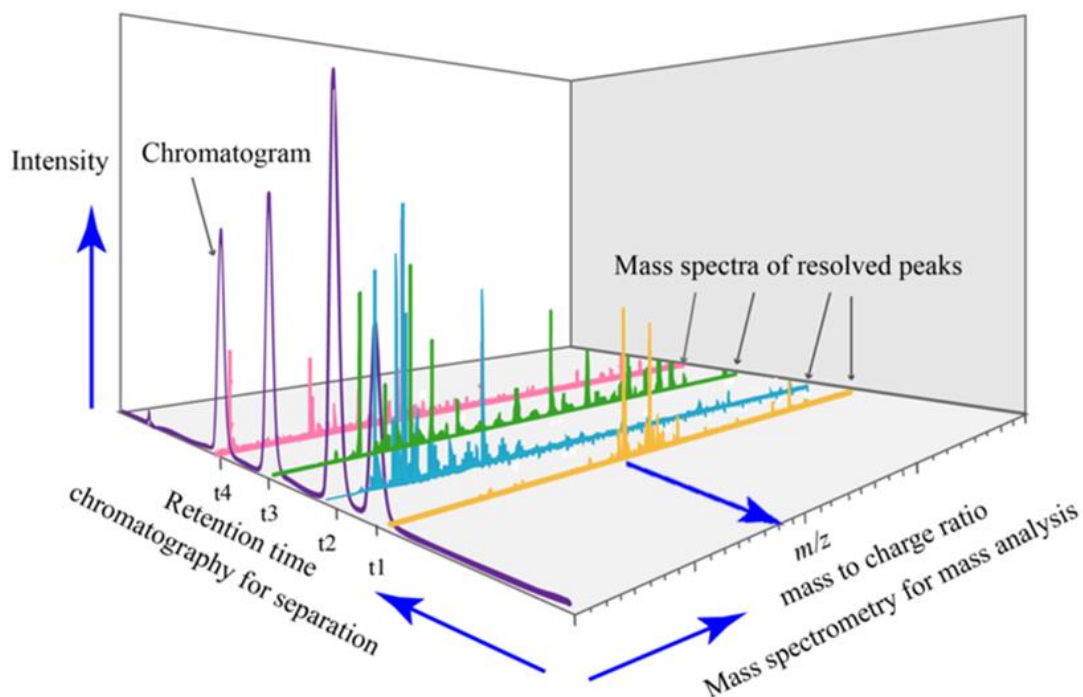
Detekcja pików odbywa się wzdłuż widma, a ich powierzchnie są zintegrowane w celu zapewnienia oznaczenia ilościowego metabolitów. W wyniku tego procesu złożone matryce danych 3D MS przekształcane są w listy cech metabolicznych, odpowiadających RT lub MT w postaci stosunku masy do ładunku (m/z) i względnej intensywności badanych związków.

Detekcja pików powinna opierać się na pomiarze sygnału pochodzącego od związków i unikać fałszywie dodatnich wyników. Metodologia różni się ze względu na źródła jonizacji twardej dla GC-EI-MS lub LC-ESI-MS, oraz miękkiej dla CE-ESI-MS.

GC-EI-MS

Praca ze źródłem EI, powoduje iż otrzymane widmo masowe jest osobliwym odciskiem palca dla cząsteczki, co prowadzi do identyfikacji związków. Ostatecznie powstaje wykres 3D, który składa się zarówno z chromatogramu, jak i widma dla każdego oddzielnego składnika. Pierwszy rozdział pików wynika z rozdzielczości chromatograficznej, przy czym anality w mieszaninie próbek, fizycznie ulegają rozdzielaniu ze względu na różny punkt wrzenia, a także ze względu na ich selektywne oddziaływanie z fazą stacjonarną kolumny analitycznej i eluują w różnych czasach retencji. Istnieją jednak anality eluujące w podobnych czasach retencji. Drugi rozdział, spowodowany jest rozdzielczością widmową, przy czym anality eluujące w podobnych czasach retencji (które nie mają całkowitej rozdzielczości) ulegają rozdzielaniu poprzez ich różne (niepowtarzalne) stosunki masy do ładunku, wykrywane przez analizator masy.

Całkowity prąd jonowy (TIC) przedstawia sumę intensywności w całym zakresie mas, w każdym punkcie analizy. W chromatografii wybranych jonów (EIC) jedna bądź więcej wartości m/z reprezentują jeden lub więcej interesujących nas analitów. Ekstrahujemy analit z całego zestawu danych dla przebiegu chromatograficznego. Na rycinie 2.5 przedstawiono dane 3D uzyskane przy pomocy GC/MS.



Rycina 2.5. Dane 3D uzyskane przy pomocy GC/MS

Narzędzia do badań metabolomicznych w dużej mierze zależą od liczby zidentyfikowanych metabolitów oraz ich interpretacji biologicznej. Fragmentacja metabolitów w źródle EI jest charakterystyczna dla danej struktury chemicznej, umożliwiając wykorzystanie widm do identyfikacji związków z bibliotek widm masowych. Procedura ta, może być zastosowana, jeśli posiadamy widmo czystego związku: tylko dany związek ulega fragmentacji oraz wykryte masy nie mogą należeć do innych związków (koelucja, tło, pozostałości z kolumn). W złożonych próbkach, wiele analitów eluuje jednocześnie, przesłaniając informacje o poszczególnych związkach, dlatego też, ekstrakcja TIC często dostarcza ograniczonych informacji.

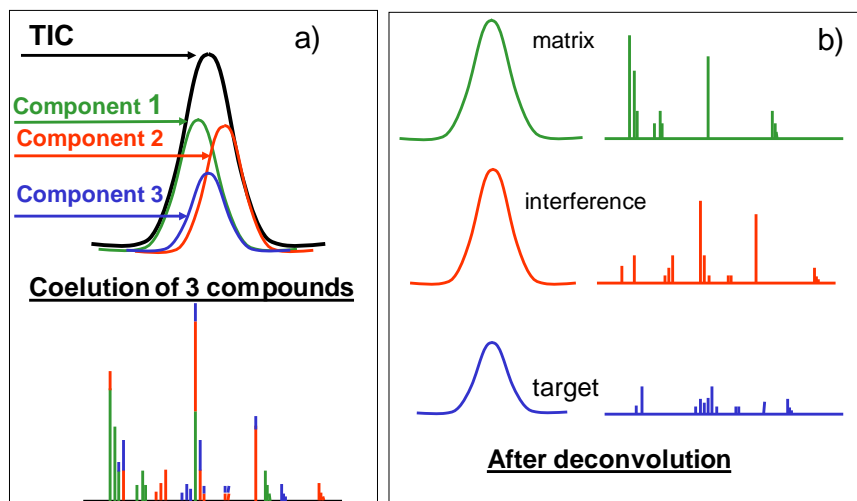
Kiedy ekstrahujemy widma masowe w celu zidentyfikowania nieznanymi składników, trudno sprawdzić, czy istnieje koelucja z innymi składnikami. Koelucja powoduje powstanie mieszanego widma masowego. Porównanie danych z biblioteką widm masowych może doprowadzić do nieprawidłowej identyfikacji związku lub braku potencjalnych dopasowań. Istnieją sposoby sprawdzania czy istnieje koelucja badanych związków. Po pierwsze, ręcznie, klikając w dowolnym punkcie pików i obserwując czy widmo masowe zmienia się wzdłuż pików lub poprzez ekstrakcję kluczowych jonów w celu utworzenia widm EIC. EIC po nałożeniu powinny wykazywać takie same kształty oraz czasy retencji, wskazując, że wszystkie należą do badanego pików.

Dekonwolucja widm w GC-EI-MS

Istnieją sposoby na poprawę jakości widma masowego przed identyfikacją przez biblioteki widm. Najczęstszym sposobem jest odejmowanie tła po obu stronach pików. Jednak w większości programów, ręczne odejmowanie tła, jak i ekstrakcja jonów może być długotrwałym i czasochłonnym procesem, zwłaszcza gdy posiadamy znaczną ilość pików w chromatogramie i/lub gdy badamy wiele próbek. W celu zidentyfikowania i wyekstrahowania ilościowej informacji o odpowiednich metabolitach należy skonstruować widmo dla każdego

pojedynczego metabolitu. Ten etap tworzenia widma nazywamy dekonwolucją. Dekonwolucja jest procesem oddzielania składników koeluuujących i tworzenia czystego widma dla każdego składnika. W szczególności dla każdego widma EIC dla dwóch lub więcej składników, dekonwolucja, oblicza wkład każdego składnika w widmo EIC.

NIST (National Institute of Standards and Technology) opracował technikę AMDIS (Automated Mass Spectrometry Deconvolution and Identification System) służącą do ekstrakcji jonów występujących nawet w śladowych ilościach. Poniższy rysunek (Rycina 2.6) przedstawia wynik procesu dekonwolucji.

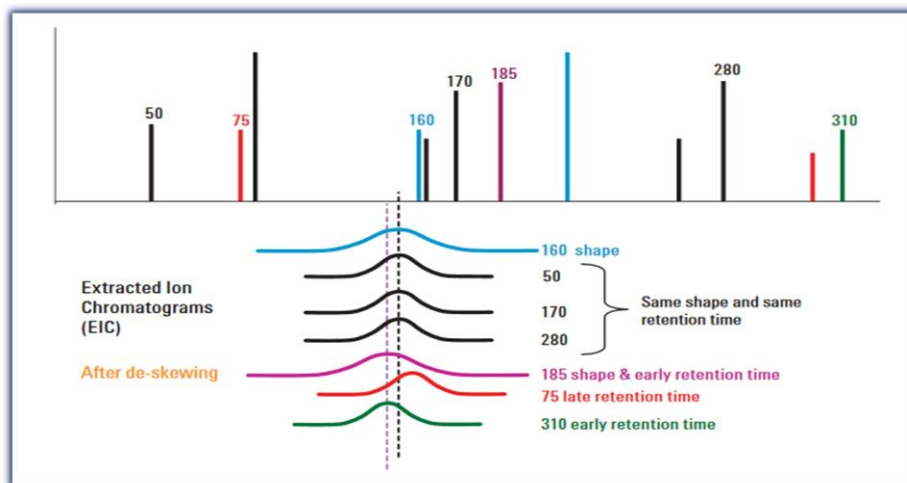


Rycina 2.6. a) przed i b) po procesie dekonwolucji (dostosowany z <https://www.agilent.com/cs/library/Support/Documents/f05017.pdf>).

Ogólny proces dekonwolucji, w technice AMDIS składa się z czterech następujących etapów: 1) analiza szumów, 2) percepcja komponentów, 3) kształt modelu i 4) dekonwolucja widma.

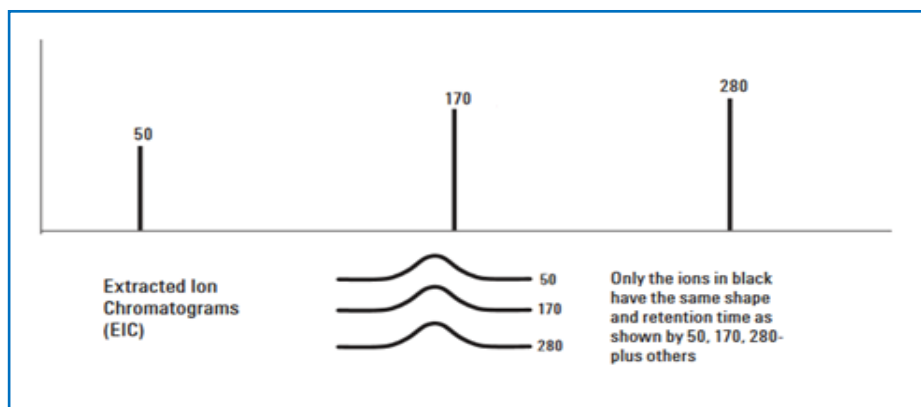
W pierwszym etapie ekstrakcji szumów dla pliku danych GC-MS, obliczamy współczynnik szumu używany do reprezentowania wielkości sygnału w jednostkach szumu. Drugi krok, percepcja komponentów, wychwytuje poszczególne komponenty chromatograficzne. Uzasadnieniem percepcji składnika jest fakt, że komponent może istnieć, gdy intensywność jonów ulega zwiększeniu. Trzeci etap określa model analizy pików, który ma być użyty w następnym kroku jakim jest dekonwolucja. Kształt modelu dla każdego komponentu przyjmuje się jako sumę poszczególnych chromatogramów jonowych, które maksymalizują się razem i których ostrość mieści się w granicy 75% wartości maksymalnej dla danego składnika. Ostatni etap, dekonwolucja czyli "oczyszczenie" widma z poszczególnych chromatogramów jonowych dla każdego składnika obejmuje kształt modelu i metodę najmniejszych kwadratów.

AMDIS rozpatruje kształty pików wszystkich wyekstrahowanych jonów i czasy retencji szczytów pików (RT). Na przykładzie (Rycina 2.7), jedynie niektóre chromatogramy (EIC) pokrywają się z widmem wierzchołkowym. Jon 160 EIC ma taki sam RT jak jony 50, 170 i 280, ale ma inny kształt piku. Jon 185 ma inny kształt piku i wcześniejszy RT. Jony 75 i 310 mają podobny kształt piku, ale mają różne RT.



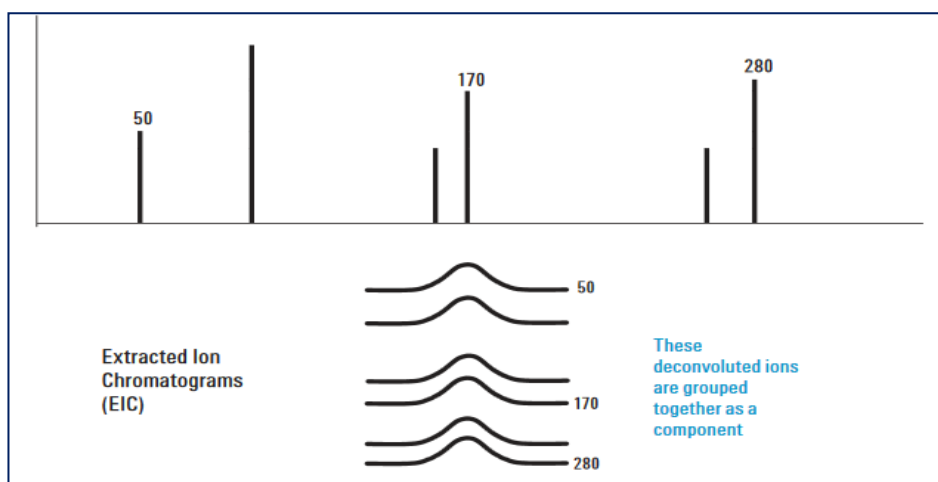
Rycina 2.7. Chromatografia wybranych jonów (obraz dostarczony przez Agilent).

Rycina 2.8 przedstawia EIC po usunięciu innych kształtów pików lub RT z ryciny 2.7. Pozostają jony 50, 170 i 280.



Rycina 2.8. Jony o tym samym kształcie i czasie retencji (obraz dostarczony przez Agilent).

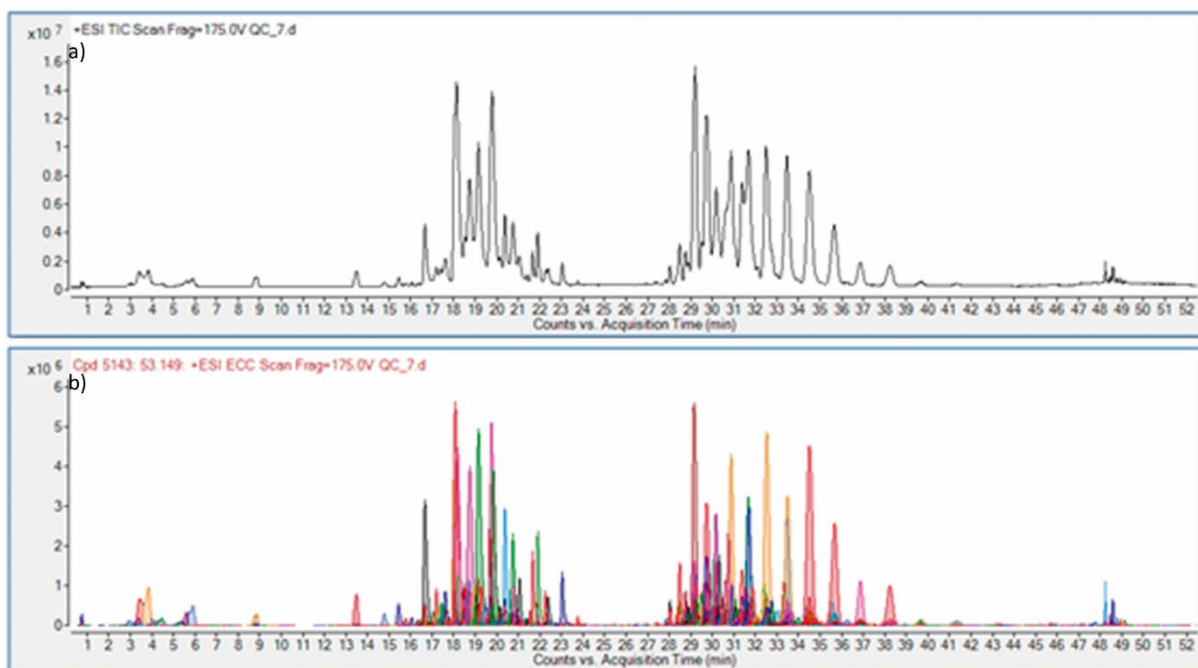
Na rycinie 2.9 przedstawiono wszystkie jony o podobnym kształcie pików i RT w kryteriach ustalonych wcześniej przez analityka. Są one zgrupowane i określane jako komponent przez AMDIS.



Rycina 2.9. Widmo masowe komponentu po dekonwolucji przez program AMDIS (obraz dostarczony przez Agilent).

Dekonwolucja widm w LC-ESI-MS i CE-ESI-MS

Metody oparte na analizie pików, są najczęstszymi algorytmami detekcji cech w badaniach MS. Takie algorytmy jak np. Molecular Feature Extractor (Agilent) oceniają dokładność pomiarów masy w celu grupowania jonów związanych przez stan powłoki ładunku, rozkład izotopowy oraz możliwe relacje chemiczne przy określaniu, czy różne jony mają tę samą cechę metaboliczną. Można rozważyć, czy jony pokrewne tworzą addukty: protonowe, sodowe, potasowe i amonowe w trybie pozytywnej jonizacji lub tracą proton, tworząc addukty mrówczanowe itp. w trybie jonizacji ujemnej. Chromatogram przed i po dekonwolucji przedstawiono na rycinie 2.10.



Rycina 2.10. a) całkowity prąd jonów b) chromatogramy dla każdego pojedynczego związku otrzymanego po dekonwolucji.

(3) Wyrównanie

Wyrównanie widm jest jednym z głównych etapów wstępnego przetwarzania danych w badaniach metabolomicznych, obejmujących analizę wielu próbek. Przy analizie wielu widm, położenie pików odpowiadające tej samej właściwości metabolicznej może mieć wpływ na przesunięcie nieliniowe. W badaniach opartych na analizatorach MS obserwujemy głównie przesunięcia pików wzdłuż osi RT. W związku z tym należy zastosować wyrównanie widm (dopasowanie pików przy pomocy m/z i RT wśród wszystkich próbek) w celu skorygowania zmienności, ponieważ może to mieć istotny wpływ na jakość badania. Algorytmy wyrównywania widm można podzielić na dwie grupy: (1) dane wyrównywane przed detekcją pików; i (2) metoda wyrównywania pików, gdy wykryte widmo naszego pików wyrównywane jest względem innych próbek.

(4) Imputacja zagubionych danych

Niestety, nie wszystkie metody wstępnego przetwarzania danych generują kompletny zestaw danych i zajmują się problemem utraconych danych w macierzy, co stanowi duże wyzwanie w badaniach metabolomicznych. Istnieje, kilka źródeł, które mogą prowadzić do

utruty danych: (1) granica wykrywalności; (2) niska intensywność sygnału; (3) błąd pomiaru; (4) błędy dekonwolucji/wyrównywania; (5) błędy w identyfikacji sygnałów z tła; lub (6) brak metabolitu w próbce. Można zidentyfikować trzy rodzaje utraconych danych: (1) utrata danych w sposób przypadkowy (MCAR), gdy brakujące dane nie są związane z obserwowaną zmienną lub próbką; 2) losowa utrata danych (MAR), gdy brakujące dane dotyczą jednej lub więcej zaobserwowanych zmiennych, ale nie są związane z próbą; (3) nieprzypadkowa utrata danych (MNAR), związanych z próbą. Utrata danych, stanowi poważny problem dla dalszego przetwarzania danych i analizy statystycznej. Spowodowana jest kilkoma czynnikami, w tym stężeniami analitów poniżej granicy wykrywalności systemów MS, ale także błędną identyfikacją przez oprogramowanie. Dlatego, opracowano kilka algorytmów imputacji utraconych danych (MVI). Metody te obejmują: (1) zastąpienie przez 0; (2) zastąpienie przez średnią/medianę, wartość minimalną, połowę wartości minimalnej; (3) metoda k -najbliższy sąsiad (kNN); (4) probabilistyczna PCA (PPCA); (5) Random Forest (RF); (6) metoda Bayesa PCA (BPCA).

(5) Filtracja danych

Matryca danych metabolomicznych zbudowana jest z tysiąca cech zawartych w wielu próbkach. Przed analizą statystyczną, macierz danych musi zostać poddana filtrowaniu w celu zmniejszenia liczby zmiennych. Istnieją różne sposoby filtrowania danych, ale ogólnym celem jest usunięcie właściwości metabolomicznych o słabej powtarzalności. W badaniach metabolomicznych wykorzystuje się głównie, dwie strategie: (1) filtrowanie w oparciu o procent próbek zawierających dane cechy w danej grupie próbek; (2) na podstawie procedur zapewnienia jakości oraz prognozy wartości RSD obliczonego dla każdej właściwości metabolomicznej wykrytej w próbkach QC. Ponadto, dane powinny zostać oczyszczone za pomocą odjęcia tła.

3.1.3. Wstępna obróbka danych

Kilka czynników (np. niepożądana wariancja eksperymentalna i biologiczna oraz błędy techniczne) może powodować trudności z identyfikacją istotnych różnic w profilach metabolicznych. Aby usunąć określone typy niepożądanych wariacji, stosuje się korektę dryfowania sygnałów oraz metodę skalowania.

a). Normalizacja

Normalizację, stosuje się w celu skorygowania niepożądanych różnic w intensywności pików i ustabilizowania wariacji w zbiorze danych. Nieprzewidywalna wariancja eksperymentalna może wynikać z błędu ludzkiego (np. pobieranie próbki i jej przygotowanie), wariacji instrumentu (np. zmiany temperatury, degradacja próbki lub utrata wydajności przyrządu podczas długotrwałej analizy), wariacji pomiędzy przyrządami, różnymi partiami próbek, różnymi laboratoriami i różnymi platformami analitycznymi. Nieznana zmienność biologiczna (np. liczba komórek, stężenie biofluidów) jest również powszechna i może wpływać na pożądane cechy. Normalizację interferencji można przeprowadzić z zastosowaniem lub bez stosowania wzorca wewnętrznego jako odniesienia w celu korekty zaobserwowanych błędów analitycznych. Metody normalizacyjne, które nie opierają się na stosowaniu wzorca wewnętrznego często stosują sumę, średnią lub medianę odpowiedzi wszystkich metabolitów w próbce jako czynnik normalizacyjny. Inne metody obejmują np. probabilistyczną normalizację iloczynową (PQN), normalizację zmiany krotności mediany (FC), normalizację ilościową lub normalizację poprzez krzywą LOESS

(locally weighted scatter plot smoothing). Wybór techniki normalizacji, opiera się na zbiorze założeń dotyczących pomiarów zmienności metabolicznych oraz biologicznych.

b). Transformacja

Ilość metabolitów w macierzy danych zwykle ma prawy rozkład ukośny, dlatego też przed analizą statystyczną często stosuje się transformację logarytmiczną.

c). Skalowanie

Skalowanie przeprowadza się w celu zaobserwowania zmian pomiędzy metabolitami, które mogą wynikać z wariacji odpowiedzi analitycznych. Metody skalowania dzielą każdą zmienną przez współczynnik (współczynnik skalowania), który jest różny dla każdej zmiennej. W metabolomice zastosowano szereg metod skalowania, obejmujących skalowanie automatyczne, skalowanie Pareto, skalowanie zakresu oraz skalowanie VAST.

Literatura:

O. Fiehn. Metabolomics--the link between genotypes and phenotypes. *Plant Mol. Biol.*, 2002, 48(1-2):155-71.

W.B. Dunn, D.I. Broadhurst, H.J. Atherton, et al. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.*, 2011, 40(1):387-426.

W.B. Dunn, I.D. Wilson, A.W. Nicholls, and D. Broadhurst. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*, 2012, 4:2249-2264

A. Alonso, S. Marsal and , A. Julià. Analytical methods in untargeted metabolomics: state of the art in 2015. *Frontiers Bioeng. Biotech.*, 2015, 3:23.

R. Di Guida, J. Engel, J.W. Allwood, et al. Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics*, 2016, 12(5):1-14.

3.2. Identyfikacja metabolitów

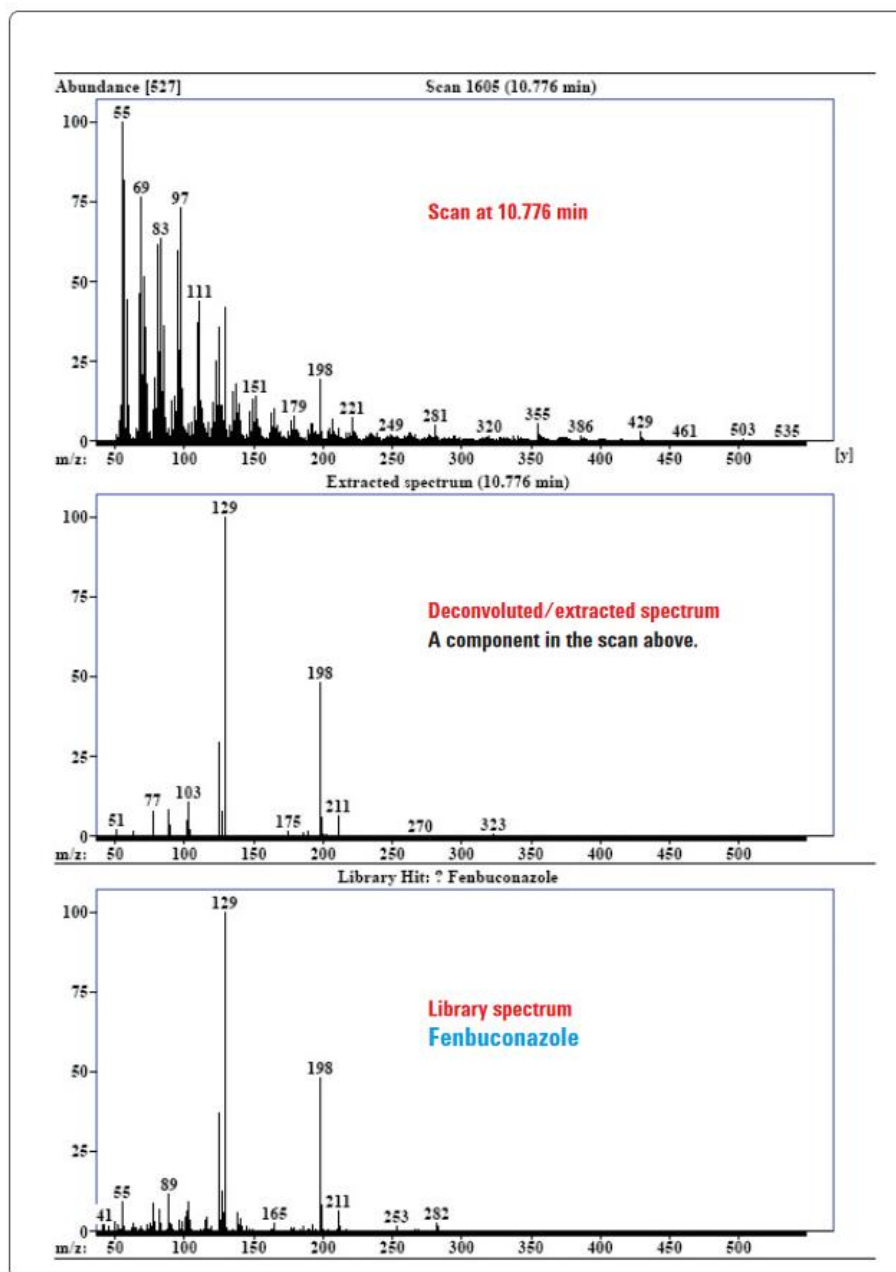
Często jednym z najbardziej wymagających kroków jest identyfikacja metabolitów. Zdaniem Donalda Rumsfelda (sekretarza obrony, USA): "...rzeczy, o których wiemy, że je znamy. Wiemy również, że istnieją znane niewiadome. Innymi słowy, wiemy, że są pewne rzeczy, których nie znamy. Ale są również nieznanne niewiadome – takie, o których nie wiemy, że ich nie znamy”.

3.2.1. Identyfikacja za pomocą GC-EI-MS

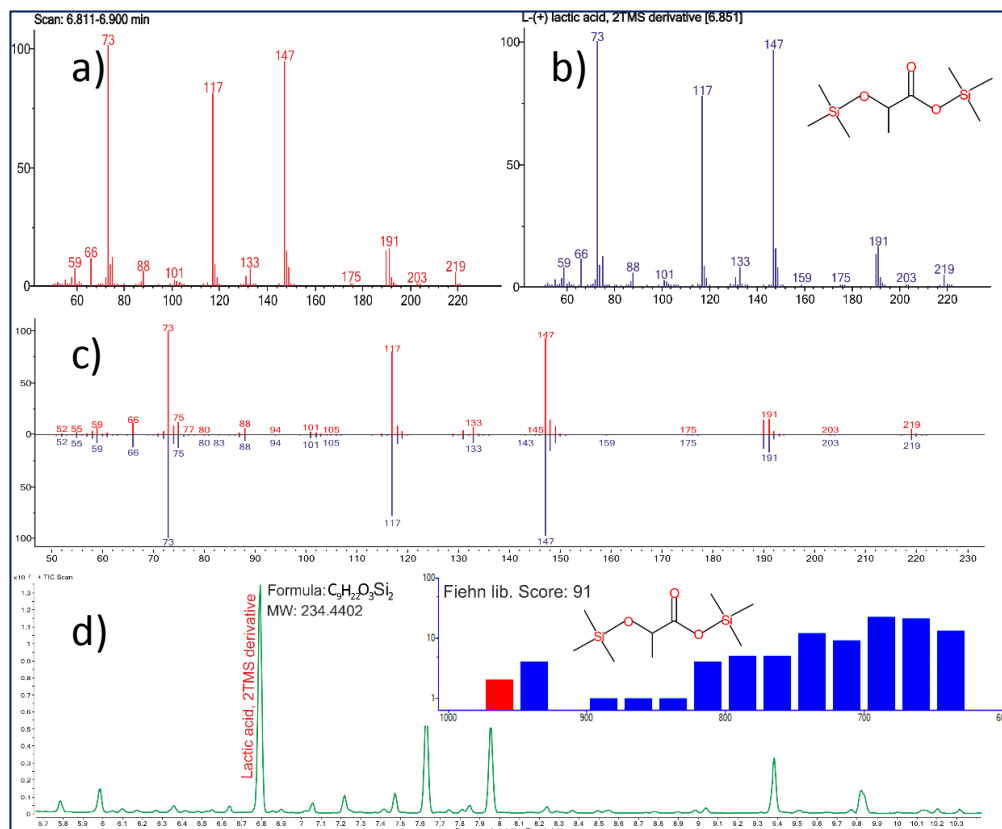
Zdolność dekonwolucji, porównana jest na dwóch górnych widmach na rycinie 2.11. Skanowanie przed dekonwolucją pokazane jest na górnej rycinie. Komponenty po dekonwolucji, widzimy na środkowej rycinie. Skan dolny jest skanem zidentyfikowanego związku z biblioteki widmowej. Bez dekonwolucji, analityk wizualnie porównuje następujące zależności: odcięcie tła przed dekonwolucją oraz skanowanie biblioteczne w celu potwierdzenia badanego związku. Byłoby bardzo trudne, wręcz niemożliwe, stwierdzenie, że fenbuconazol jest związkiem docelowym na tym przykładzie, przy użyciu tego rodzaju porównania.

W GC-EI-MS identyfikacja metabolitów może być przeprowadzona za pomocą dwóch kryteriów: przeszukiwanie czystego widma uzyskanego przez dekonwolucję w komercyjnych bibliotekach widm oraz poprzez dopasowanie RT z biblioteką widm (ten sam czas retencji dla analitu uzyskanego tą samą metodą analityczną). W zależności od współczynnika dopasowania podczas wyszukiwania, docelowe związki mogą zostać zidentyfikowane lub

oznaczone w złożonym TIC. Każdy komponent zostanie przeszukany w oparciu o bibliotekę czasu retencji (zwaną biblioteką targetową). Istnieją targetowe metaboliczne biblioteki widmowe, biblioteki RTL lub specjalnie opracowane, aby ułatwić identyfikację metabolitów. Rycina 2.12 przedstawia ten proces dla potwierdzenia kwasu mlekowego w próbce osocza. Ze względu na wysoką powtarzalność, chromatograficzną rozdzielczość pików i istnienie bibliotek widm masowych, GC-EI-MS jest podstawową techniką wykorzystywaną w badaniach metabolomicznych.



Rycina 2.11. Porównanie widma przed dekonwolucją, po dekonwolucji oraz z biblioteki widm (obraz dostarczony przez Agilent).



Rycina 2.12. Identyfikacja metabolitu za pomocą chromatogramu GC-EI-MS poprzez dopasowanie RT oraz widma w porównaniu do biblioteki targetowej. a) otrzymane widmo, b) widmo biblioteki targetowej, c) wykres po nałożeniu widm, d) chromatogram TIC z wynikami przeszukiwania powyżej 100.

Eksperymenty przeprowadzone na aparaturach z dokładnym pomiarem mas, przy wykorzystaniu precyzyjnych baz danych, zawsze zwiększają liczbę zidentyfikowanych związków. Aparaty te, ułatwiają również identyfikację nieznanymi związków przez tworzenie struktur związków oraz ich fragmentów.

3.2.2. Identyfikacja metabolitów za pomocą LC lub CE-(ESI)MS: Praca z bazami danych

Powszechnie przyjmuje się, że we wspólnym eksperymencie metabolomicznym istnieją trzy poziomy identyfikacji metabolitów od najmniej do najlepiej scharakteryzowanych związków: około 50% związków nie zostanie zidentyfikowanych ("nieznane"), ale wśród pozostałych 50% mogą znajdować się zidentyfikowane przy użyciu różnych baz danych związki. Identyfikacja może zostać przeprowadzona również za pomocą takich parametrów jak czas retencji lub fragmentacja MS/MS. Związki mogą ponadto zostać pozytywnie zidentyfikowane, gdy standard związku analizowany jest w tych samych warunkach i widmo pokrywa się z jeszcze niezidentyfikowanym związkiem.

W ostatnich latach bazy danych dotyczące metabolomiki uległy znacznemu rozwojowi: rozszerzeniu (liczba metabolitów), poszerzeniu informacji o metabolitach oraz ich funkcjonalności. Na przykład, baza danych metabolomu człowieka (HMDB) została wydana po raz pierwszy w 2006 r. wraz z 2180 metabolitami, a obecnie zawiera ponad 100 możliwości wyszukiwania informacji dla każdego związku oraz ponad 42 000 metabolitów.

Niemniej jednak, prace nad bazami danych w dziedzinie metabolomiki są dalekie od ujednolicenia, podobnie jak w innych badaniach "omicznych", przez co naukowcy muszą łączyć dostarczone przez nie informacje. Istnieje ponad 50 różnych zasobów online. Dlatego w ciągu ostatnich kilku lat pojawiło się kilka programów, które zarządzają informacjami z kilku źródeł baz danych.

Bazy danych, które mogą być stosowane w metabolomice, można sklasyfikować ze względu na rodzaj informacji, na te które koncentrują się na szlakach metabolicznych lub na samych metabolitach, tworząc później zarówno spektralne bazy danych jak i bazy danych badanych związków. Klasyfikacja ta nie może być uznana za ostateczną, ponieważ dane ulegają ciągłej modyfikacji, a część informacji znajduje się w różnych bazach danych.

Poniższy tekst podsumowuje niektóre z najważniejszych, najbardziej dostępnych baz danych związanych z metabolomiką. Warto zauważyć iż podstawowym celem części baz danych nie są badania dotyczące metabolomiki np. PubChem lub KEGG.

3.2.3. Bazy danych metabolitów, klasyfikacja w oparciu o wartość masy cząsteczkowej

ChEBI (Chemical Entities of Biological Interest) jest inicjatywą europejskich bioinformatyków z Europejskiego Laboratorium Biologii Molekularnej (EMBL). Obejmuje związki z różnych źródeł w celu dostarczenia wystandaryzowanych opisów istotności molekularnych (blisko 45 000 metabolitów otrzymało tzw. "3 gwiazdki"), które umożliwiają innym bazom przypisywanie adnotacji w spójny sposób. ChEBI koncentruje się na wysokiej jakości ręcznej adnotacji, braku redundancji i udostępnianiu ontologii chemicznej. Można przeszukiwać ją według nazwy, wzoru lub struktury związku.

PubChem jest bazą danych NIH, liczącą wiele milionów (blisko 100) związków (<1000 atomów) i substancji (ponad 200 milionów), oczywiście nie wszystkie są produktami metabolizmu. Dane pochodzą od 80 różnych sprzedawców/deponentów. Substancje mogą być "zanieczyszczone/duplikaty", podczas gdy związki są pojedynczymi jednostkami (Pubchem Compound ID - CID daje unikalny związek). Jednostki zawierają synonimy, właściwości chemiczne, strukturę chemiczną, w tym SMILES i szereg InChI, bioaktywność oraz powiązania strukturalne z innymi bazami danych NCBI, takimi jak PubMed. Związki można wyszukiwać według nazwy, formuły, zakresu MW, struktury, liczby donorów/akceptorów H lub XlogP.

ChemSpider jest bezpłatną bazą danych o strukturze chemicznej, zapewniającą szybki dostęp do ponad 58 milionów struktur, właściwości i powiązanych informacji. Dzięki integracji i połączeniu związków z ~ 500 źródeł danych, ChemSpider w wyniku pojedynczego wyszukiwania umożliwia naukowcom poznanie najbardziej wszechstronnego obrazu badanego związku. Jest własnością Royal Society of Chemistry. Związki można wyszukiwać według nazwy, synonimu, InChi, struktury, rejestru, SMILES, obliczonych właściwości (ale nie według wzoru lub masy). Dane obejmują nazwy, synonimy, artykuły wikipedia, opisy, źródła danych, sprzedawców, patentów, artykułów, właściwości, nagłówków MESH, linków farmakologicznych, widm (UV, IR, NMR, MS) pochodzących z innych witryn.

3.2.4. Bazy danych metabolitów, klasyfikacja w oparciu o widma masowe

Ten typ baz danych wspiera nie tylko przeszukiwanie według MW lub jego zakresu, ale także wyszukuje jony macierzyste (pozytywne, negatywne, neutralne), tworzy listę wyszukiwanych pików (z danych MS lub MS/MS) oraz dopasowuje widma za pomocą MS/MS. Powyższe

bazy danych przeznaczone są w większym stopniu do badania metabolizmu opartego o MS i identyfikacji związku niż proste narzędzia oparte na pomiarze MW.

BioCyc jest zbiorem zaktualizowanych baz danych dla różnych organizmów, który został opracowany przez SRI International (Menlo Park w Kalifornii). Bazy danych budowane są za pomocą ręcznych aktualizacji. Dane poziomu 1, np. EcoCyc (dla *E. coli*) i HumanCycle, są w dużej mierze uaktualniane, a większość baz danych BioCyc, poziomy 2 i 3 zostały stworzone komputerowo. Przydatne są w celu identyfikacji określonych metabolitów organizmu i rekonstrukcji szlaków metabolicznych poprzez przeszukiwanie szlaków biologicznych.

HMDB jest bazą danych poświęconą ludzkiemu metabolizmowi, opracowaną przy wsparciu Canadian Institutes of Health Research, Alberta Innovates - Health Solutions i Metabolomics Innovation Center. Do każdej danej podano informacje na temat właściwości chemicznych, biologicznych i klinicznych danego związku oraz znajdują się także odnośniki literaturowe, zawierające informacje na temat stanów chorobowych oraz związanych z nimi enzymów i transporterów. Istnieją również powiązania z zewnętrznymi bazami danych, takimi jak KEGG.

Komic Market (Kazusa Omics Data Market) to baza danych dotycząca detekcji pików MS w badaniach metabolomicznych. Pochodzi z projektu "Development of Fundamental Technologies for Controlling the Material Production Process of Plants", wspieranego przez Organizację Rozwoju Technologii Przemysłu i Energii w Japonii.

LipidBank jest oficjalną bazą danych Japanese Conference on the Biochemistry of Lipids (JCBL), poświęconą obojętnym lipidom. Wszystkie informacje molekularne zatwierdzane są przez ekspertów w dziedzinie badań nad lipidami. Każda pozycja zawiera nazwę lipidu, strukturę molekularną, informacje spektralne i odnośniki literaturowe.

LipidMaps jest finansowana przez grant badawczy ("Glue Grant") z NIH National Institute of General Medical Sciences. Jej celem jest identyfikacja i ilościowe oznaczenie lipidów ssaków. LipidMaps Proteome Database (LMPD) również uwzględniona jest w tym zasobie.

MassBank przechowuje dane widm masowych, oparte na identyfikacji i strukturalnych wyjaśnieniach związków chemicznych wykrytych za pomocą MS. MassBank jest dostępna poprzez dwie domeny: japońską (<http://massbank.jp>) i europejską (<http://massbank.eu>) (NORMAN MassBank), ale niektóre funkcje dostępne są tylko w języku japońskim.

METLIN jest znakiem towarowym Scripps Research Institute, który opracowuje i stosuje technologie spektrometrii mas w celu zrozumienia metabolizmu. Obejmuje ona, procesy informatyczne przetwarzania danych (XCMS) i nanostrukturalną spektrometrię obrazową (NIMS). Zawiera, prawie milion informacji o związkach i jest jedną z największych dostępnych baz danych (metabolity, lipidy, sterydy, metabolity roślinne i bakteryjne, małe peptydy i egzogenne metabolity leków oraz substancje toksyczne). IsoMETLIN jest dodatkowym modulem dla badań metabolomicznych opartych na izotopach.

MycompoundID jest zasobem sieciowym opracowanym na Uniwersytecie Alberta w celu identyfikacji związków, opartej na właściwościach chemicznych, w tym dokładności mas. Możliwe są różne sposoby przeszukiwania bazy danych, w tym MS, MS2, metoda PEP czyli nieznakowanych i znakowanych dimetylowych peptydów oraz znakowanych izotopów chemicznych za pomocą MS. Biblioteka metabolomów (EML), składa się z 8.021 znanych ludzkich metabolitów endogennych i ich przewidywanych produktów metabolicznych, w tym 375.809 związków wchodzących w jedną reakcję metaboliczną oraz 10.583.901 związków

opartych na dwóch reakcjach. Związki generowane przez badania silico pochodzą z bazy danych HMDB.

3.2.5. Informatyczne bazy danych

CEU MassMediator jest bazą danych powstałą przy współpracy CEMBio i Bioengineering Laboratory of Polytechnic Faculty at Universidad CEU San Pablo, Madrid. Narzędzie to przeprowadza automatyczne przeszukiwanie w zewnętrznych źródłach danych (HMDB, KEGG, LipidMaps, METLIN i MINE) i zapewnia możliwość identyfikacji konkretnego związku na podstawie masy (ujednoczenie trafień z wielu baz danych do jednego trafienia).

CSI: FingerID jest specyficzną bazą danych w celu identyfikacji MSⁿ. Uzupełnia badania nad niezidentyfikowanymi substancjami na poziomie MS. Jest to wspólny projekt pomiędzy Friedrich Schiller University, Germany i Helsinki Institute for Information Technology at Aalto University, Finlandia, łącząc obliczenia dotyczące fragmentacji danej substancji z zaawansowanymi badaniami komputerowymi, w celu poprawy procentu zidentyfikowanych cząsteczek oraz dokładności identyfikacji.

MAGMa jest narzędziem opracowanym w ramach projektu eMetabolomics, finansowanego przez Netherlands eScience Center at Wageningen University we współpracy z Netherlands Metabolomics Centre. Dane MSⁿ mogą być przesłane jako hierarchiczne fragmenty piku, na podstawie stosunku m/z lub wzoru chemicznego. Potencjalni kandydaci (cząsteczki) automatycznie ładowane są z PubChem, KEGG lub HMDB. Odpowiadający wynik, obliczany jest na podstawie fragmentacji pików.

MassTRIX jest internetowym narzędziem służącym do opisu danych spektrometrii mas o dużej precyzji. Wyniki wyświetlane są na specyficznych dla danego organizmu mapach szlaków biochemicznych KEGG, a wszelkie dodatkowe informacje o genomie lub transkryptomie mogą zostać dodane. Narzędzie to zostało opracowane w Helmholtz Zentrum München w kolaboracji pomiędzy Philippe Schmitt-Kopplin i Karsten Suhre.

3.3. Analiza statystyczna

Po otrzymaniu matrycy obrobionych danych zawierających potencjalną listę możliwych związków i ich ilości w każdej pojedynczej próbce, następnym krokiem jest przeprowadzenie analizy statystycznej za pomocą narzędzi chemometrycznych, które dostarczają opisu wariacji biologicznych w badanym systemie. W analizie danych dotyczących metabolomiki nietargetowej wykorzystuje się dwie analizy statystyczne: jednowymiarową oraz wielowymiarową analizę danych (MVDA).

Jednowymiarowa analiza danych zakłada, że na odpowiedź zmiennych wpływ ma tylko jeden czynnik. W chorobach zaangażowane są różne systemy oraz szlaki biologiczne, w których jednocześnie zaangażowanych jest wiele zmiennych. Tradycyjna analiza statystyczna zmierza do przekształcenia wszystkich zmiennych w jednowymiarowe zmienne, nawet tych, które z natury są wielowymiarowe. Z tego powodu MVDA jest właściwym narzędziem statystycznym do interpretacji danych pochodzących z badań metabolomicznych. Podsumowuje tabele danych z wieloma zmiennymi oraz kilkoma obserwacjami, zmniejszając liczbę zmiennych oraz klasyfikuje dane do grup. MVDA dostarcza modeli statystycznych, w szczególności pojedynczych metabolitów będących przedmiotem zainteresowania, które można ostatecznie określić chemicznie lub strukturalnie. Pierwszym krokiem MVDA jest utworzenie macierzy X, gdzie próbki są zawarte w wierszach, a wszystkie zmienne (związki)

w kolumnach. Następnie, konieczne jest wstępne przetwarzanie danych (skalowanie i transformacja). Dane analizowane są za pomocą niezależnej, analizy składowych głównych (PCA) oraz nadzorowanej częściowo, dyskryminacyjnej metody cząstkowych najmniejszych kwadratów (PLS-DA) oraz ortogonalnej metody PLS (OPLS -DA). Obie metody kompresują pierwotną macierz danych w taki sposób, aby wykazać zależności pomiędzy próbkami.

3.3.1. PCA

PCA: jest metodą redukcji wymiarów, która szeroko stosowana jest w poszukiwaniu danych i ich wizualizacji. Została po raz pierwszy zaproponowana przez Pearson w 1901 roku. Celem PCA jest ujawnienie wzorców poprzez kompresję danych, starając się zachować jak najwięcej pierwotnych informacji. PCA stanowi naturalny punkt startu dla każdej wielowymiarowej analizy danych. Przydatne jest dostarczenie, globalnego, graficznego przeglądu wszystkich próbek w postaci macierzy danych, ujawniając odstające próbki, grupy, klastery, podobieństwa/odmienności, zmienne dominujące. Pozycja każdej próbki na wykresie używana jest w celu powiązania prób ze sobą: próbki położone blisko siebie mają podobny profil wielowymiarowy; odległe próbki, które są daleko od siebie mają różne właściwości.

PCA i detekcja „autlajerów”

Odstające od reszty wyniki analiz próbek („autlajery” - outliers) definiowane są jako dane, których wartości oddalone są od wartości danych uzyskanych z analiz pozostałych próbek. Wpływają one na trafność wyników metabolicznych poprzez zmianę wariancji i rozkładu danych, a tym samym zmniejszając statystyczną moc modelu. Próbki te, mogą być, zarówno biologicznymi lub analitycznymi autsajderami, w zależności od źródła zmienności. Biologiczne, są trudne do rozpoznania i spowodowane są przypadkową lub indukowaną zmiennością biologiczną wśród próbek. Często występują w badaniach złożonych chorób z niejednorodną grupą pacjentów. Ponadto w analizie statystycznej należy uwzględnić informacje pochodzące z odstających próbek biologicznych. Przeciwnie, analityczni autsajderzy powodują zwiększenie zniekształcenia informacji biologicznych i należy wykluczyć ich przed analizą statystyczną. Pochodzenie analitycznych autsajderów może być różne: pobieranie próbek, ich przechowywanie, przygotowanie próbek, analiza, znajdowanie cech lub proces dekonwolucji/identyfikacji. Decyzja o pozostawieniu bądź usunięciu autsajderów powinna być zawsze rozważna oraz uzasadniona.

Po uzyskaniu modelu PCA, autsajderów można wyróżnić, wykreślając intensywność wszystkich wykrytych związków w porównaniu z porządkiem analizy, obserwując zakres T2 Hotellinga oraz opcję "Distance to Mode" (dostępny w SIMCA P+). Przeanalizowanie diagramów oraz analiza surowych i przefiltrowanych danych potrzebne są w celu zindywidualizowania i wyeliminowania autsajderów. Na wykresie PCA wyniki są rozproszone, silni autsajderzy stanowią punkty odległe od granicy elipsy, reprezentującej 95% przedziału ufności w modelu wariancji. Wykres T2 Hotellinga przedstawia odległość środka płaszczyzny modelu dla każdej wybranej obserwacji (reprezentowanej jako kolumna). Jeśli wartość znajduje się znacznie powyżej granicy krytycznej, oznacza to, że wybrana seria komponentów nie mieści się w przestrzeni wykresu. Stąd prawdopodobieństwo, że obserwacje należą do tej samej klasy co pozostałe próbki jest niższe niż 5 lub 1%.

3.3.2. PLS-DA

Używana jest do dyskryminacji pomiędzy grupami i aby zmaksymalizować różnice pomiędzy nimi. Modeluje relacje między dwiema lub większą liczbą klas danych, przy użyciu serii lokalnych najmniejszych kwadratów. PLS ocenia związek pomiędzy macierzą deskryptora X a macierzą odpowiedzi Y. Istotną różnicą pomiędzy PLS a PCA jest to, że PLS jest "nadzorowaną" techniką, która wykorzystuje dodatkowe informacje w celu opracowania modelu statystycznego, podczas gdy PCA jest "nienadzorowana", i nie wymaga wprowadzenia danych po raz drugi.

PLS-DA i PCA mogą być stosowane jako metody redukcji wymiarów w modelach predykcyjnych przed analizą dyskryminacji liniowej (PLS-DA, PCA-DA). Dwuliniowe statystyczne podejście do analizy dyskryminacji cząstkowych najmniejszych kwadratów (PLS-DA) jest jednym z najbardziej popularnych metod nadzoru stosowanych w metabolomice. W PLS-DA, macierz X zawiera zmienne dane, natomiast macierz Y zawiera zmienne klas, dla których wartości wybrane są jako deskryptory klas.

3.3.3. OPLS-DA

Jest to ewolucja metody PLS, jest metodą regresji liniowej, która została pomyślnie wykorzystana do predykcyjnego modelowania biologicznych oraz biochemicznych aplikacji. Wśród zalet metody OPLS możemy wyróżnić zdolność do modelowania danych zarówno, zmiennych zakłócających jak i wieloliniowych, takich jak dane widm w metabolomice. W prostych słowach OPLS wykorzystuje informacje macierzy Y do rozkładu macierzy X na bloki zmienności strukturalnej, skorelowane i prostopadłe do macierzy Y. OPLS, analogicznie do PLS-DA, może być używana do dyskryminacji (OPLS-DA). Główną zaletą OPLS-DA jest możliwość oddzielenia predykcyjnych od niepredykcyjnych (ortogonalnych) zmienności. Niemniej jednak, OPLS-DA, przeciwnie do PLS-DA, nie może być używana do analizy więcej niż dwóch klas w tym samym czasie.

Jeśli zbudujemy model statystyczny (PCA, PLS-DA, itp.), istnieją dwa parametry, jakościowe opisujące nasz model (R^2 i Q^2), które mają istotne znaczenie statystyczne i stanowią pierwszy krok w walidacji metody.

- R^2 to procent zmienności zestawu danych w modelu. Określa, czy dany model pasuje do danych. Odpowiedni model ma duży współczynnik R^2 , co oznacza dobrą powtarzalność.
- Q^2 , przewiduje wariancję, informuje nas o zdolności modelu do przewidywania nowych danych. Duża wartość Q^2 wskazuje na dobrą przewidywalność.

W celu zapewnienia mocy modelu w diagnostyce, R^2 i Q^2 powinny być wysokie, ale nie różnić się od siebie o 0.2-0.3 jednostek. Dysproporcja pomiędzy tymi dwiema wartościami stanowi pierwsze ostrzeżenie, że model mógł zostać utworzony przez nadmierne dopasowanie. Przyjmuje się, że dla danych biologicznych dopuszczalne wartości to: $R^2 > 0,7$ i $Q^2 > 0,4$.

3.3.4. Wykres-S, metoda Jack-knife i test projekcyjny VIP

OPLS-DA jest punktem startu do wykonania **wykresu-S**, analizy Jack-Knife i VIP w celu identyfikacji przypuszczalnych biomarkerów. Przedziały ufności wykresu S i Jack-knife są przydatnymi narzędziami do wyboru metabolitów różnicujących dane grupy badawcze. W skrócie, wykres S jest wizualną prezentacją, przedstawioną na wykresie punktowym, zarówno

kowariancji, jak i korelacji pomiędzy metabolitami a oznaczanymi klasami. Wykres S może pokazać udział różnych źródeł w rozkład danych próbki w danej zmiennej.

W celu skonstruowania przedziału ufności dla oszacowania parametrów modelu, wykorzystywana jest technika **Jack-knife**, przedstawiona jako wykres słupkowy. Przedział ufności odzwierciedla stabilność i niepewność zmiennych, które powinny mieć niską wartość. W analizie typu Jack-knife można stworzyć domniemaną listę najistotniejszych metabolitów.

Wynik testu **VIP** wskazują wkład każdej zmiennej w predykcję modelu. W programie SIMCA P+ wykresy VIP sortowane są w oparciu o znaczenie zmiennych, dlatego wartość VIP często wykorzystywana jest do selekcji zmiennych. Aby zbudować dokładny model, metabolity, zwykle wybierane są, z wynikiem powyżej 1, $VIP > 1$.

3.3.5. Walidacja statystycznych modeli wielowymiarowych

Aby uniknąć ryzyka nadmiernego dopasowania, ze względu na to iż wyniki po analizie MVDA wrażliwe są na korelacje prawdopodobieństwa, model statystyczny musi zostać zwalidowany. Narzędzia do walidacji, takie jak a) test permutacji oraz b) walidacja krzyżowa, są najczęściej stosowane w celu obiektywnej oceny wydajności i stabilności modelu.

- (a) *Test permutacji*: używany do oceny znaczenia klasyfikacji. Przypisywanie klasy może być permutowane kilkakrotnie i dla każdej permutacji można zbudować model pomiędzy danymi a permutacją przypisania klas. Dyskryminacja pomiędzy klasami opiera się na porównaniu dyskryminacji modelu, opartego na pierwotnej klasyfikacji. Test permutacji odpowiedni jest dla dużych zbiorów danych, obejmujących ponad 50 próbek.
- b) *Walidacja krzyżowa*: wydaje się, że jest najbardziej odpowiednią metodą dla średnich i małych zestawów danych. Podczas walidacji krzyżowej pierwszym krokiem jest podzielenie próbek na kilka grup. Następnie grupy te ukrywane są sekwencyjnie, a pozostałe próbki wykorzystywane są do budowy modelu walidacji krzyżowej. Ukryte próbki dla każdego modelu są przewidywane, tym samym mierzymy zdolność predykcyjną. Proces jest powtarzany, aż wszystkie próbki zostaną ukryte i przewidziane. Stosujemy podejście ukrycia, 1/3 ilości próbek. Zestaw danych, losowo podzielony jest na trzy części, a 1/3 próbek wykluczona jest w celu wykonania modelu statystycznego z pozostałymi próbkami (2/3). Wykluczone próbki, przewidujemy za pomocą nowego modelu, wykonując analizę wyników T-przewidywania. Procedura ta, powtarzana jest, aż wszystkie próbki zostaną przeanalizowane co najmniej raz. Zależnie od tego, czy wszystkie dane zostaną prawidłowo przypisane do grup, model można uznać za prawidłowy bądź nie.

Literatura:

- E. Saccenti, H.C. Hoefsloot, A.K. Smilde, J.A. Westerhuis, M.M. Hendriks. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*, 2014, 10:361-374.
- K. Varmuza, and P. Filzmoser. Introduction to multivariate statistical analysis in chemometrics. CRC press, 2009.
- J. Godzien, M. Ciborowski, S. Angulo, and C. Barbas. From numbers to a biological sense: How the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC-MS. *Electrophoresis*, 2013, 34:2812-2826.
- J. Trygg, E. Holmes, T. Lundstedt. Chemometrics in metabonomics. *J. Proteome Res.*, 2007, 6:469-479.

4. Analiza danych

4.1. Od identyfikacji danych do szlaków biochemicznych: klasyfikacja danych w celu oceny istotności biologicznych.

Lista lub listy potencjalnych związków (jeszcze nie lista związków) muszą być nadzorowane, starając się skonsolidować nasze ustalenia. W przypadku analiz LC-MS lub CE-MS uwzględnia się addukty i wielokrotność ładunku, fragmenty (utrata neutralnych cząsteczek, produkty rozkładu, przegrupowaniu), izotopy oraz szumy.

Ponadto należy wykonać ostrożną weryfikację zidentyfikowanych związków w celu usunięcia niemożliwych dopasowań. Istnieją dwie możliwości odrzucenia zidentyfikowanych związków: wiedza na temat, powiązania pomiędzy właściwościami identyfikowanego związku (biegunowość, ładunek, addukty ...) a warunkami doświadczalnymi (rozpuszczalnik, faza ruchoma, typ chromatografii, warunki jonizacji ...); oraz wiedza na temat projektu eksperymentu (organizm, leczenie, próbka ...).

Kiedy lista zostanie zaktualizowana, dalsza identyfikacja polega na użyciu zestawu narzędzi bioinformatycznych umożliwiających ocenę biochemicznego znaczenia tych zmian. Narzędzia te, komplementarne są do ogólnych danych biochemicznych zebranych w podręcznikach, w połączeniu z nowymi podejściami do tej wiedzy, takimi jak Roche Biochemical Pathways czy też IUBMB-Sigma-Nicholson Metabolic Pathways Chart oraz dostępnymi mapami metabolicznymi online.

Narzędzia bioinformatyczne, które związane są ze szlakami biochemicznymi można podzielić na dwie grupy: bazy danych szlaków biochemicznych, o podobnej charakterystyce do baz danych związków, ale będące zapisami pełnych ścieżek biochemicznych; oraz narzędzia analizy szlaków biochemicznych, które mogą zastosować algorytmy matematyczne w celu oceny znaczenia zaobserwowanych zmian, w zależności od nasilenia sygnałów, ilości i znaczenia zmieniających się metabolitów.

4.1.1. Bazy danych szlaków biochemicznych

Szlak biochemiczny można łatwo określić jako serię reakcji konwertujących zestaw substratów w zestaw produktów. Definicja ta, jest subiektywna i niestandardowa, ponieważ dobór metabolitów, które wchodzi w jeden konkretny szlak metaboliczny jest arbitralny, ze względu na to, że szlaki metaboliczne mają różny rozmiar (liczba metabolitów i reakcji) i często pokrywają się. Jeden metabolit może być częścią różnych szlaków metabolicznych. Ponadto szlaki metaboliczne utworzone w jednej bazie danych mogą różnić się od innych baz danych.

Bazy danych szlaków metabolicznych stanowią bogate źródło danych biologicznych, które dotyczą metabolitów genów, białek, chorób, sygnalizacji zdarzeń oraz procesów. Dostarczając również narzędzi umożliwiających wizualizację i mapowanie genów/metabolitów, (często) wielu gatunków.

Do najważniejszych baz danych szlaków biochemicznych należą:

KEGG (Encyklopedia Genów i Genomów Kioto) jest bazą danych, służącą do zrozumienia funkcji i narzędzi systemu biologicznego, takich jak komórka, organizm i ekosystem oraz informacji genomowych i molekularnych. Jest to komputerowa prezentacja systemu biologicznego, składająca się z molekularnych bloków strukturalnych genów i białek (informacji genomowej) oraz substancji chemicznych (informacji chemicznych), które zintegrowane są z wiedzą na temat powiązania diagramów interakcji molekularnych,

reakcji i ich wzajemnej relacji (informacje o systemach). Zawiera także informacje o chorobach i lekach (informacje o zdrowiu), w jaki sposób powodują zaburzenia w systemie biologicznym. Obejmuje ona: 508 map szlaków biochemicznych, 17.931 związków i 11.015 glikanów, a także 10.476 reakcji biochemicznych.

Recon 2 zawiera informacje na temat 1.789 enzymów kodujących geny, 7.440 reakcji i 2626 unikalnych metabolitów rozmieszczonych w przeszło ośmiu przedziałach komórkowych. Dlatego możemy zbadać istotę odbudowy komórkowej poprzez przeszukiwanie/przeoglądanie metabolitów i reakcji.

Baza danych **Small Molecule Pathway Database** (SMPDB) składa się z niemal 900 szlaków biochemicznych dla małych cząsteczek, w tym 384 są to szlaki leków, 232- chorób oraz 220 to ogólne szlaki metaboliczne. Przedstawia podziały komórkowe, organelle, lokalizacje białek, struktury czwartorzędowe, konwersje genów, listę białek lub substancji chemicznych w szlakach biochemicznych oraz w diagnostyce chorób.

PathWhiz to serwer zaprojektowany w celu umożliwienia tworzenia biologicznych, dokładnych schematów szlaków biochemicznych, które są czytelne oraz interaktywnie w przeglądarce Google Maps-style viewer.

4.1.2. Analiza szlaków biochemicznych

Dwa typy analiz szlaków biochemicznych zostaną rozważone: analiza przez wzbogacanie oraz topologiczna. Celem wzbogacania metabolitu jest zbadanie, czy istnieją biologicznie znaczące grupy metabolitów, których zmiany są istotne statystycznie w danych. Sensem tej metody jest lokalizacja wykrytych metabolitów oraz w jakich stanach chorobowych są zaangażowane. Analiza topologiczna ma na celu rozszerzenie analizy wzbogacania, poprzez rozważenie struktury i/lub wizualizacji szlaków biochemicznych.

4.1.3. Wzbogacanie metabolitów

IMet-Q (intelligent Metabolomic Quantitation) jest zautomatyzowanym narzędziem z przyjaznym interfejsem dla użytkownika do ilościowego oznaczania metabolitów z danych z pełnego skanowania za pomocą metody LC-MS. IMet-Q posiada procedurę do ilościowego usuwania szumów, wykrywania i wyrównywania pików. Ponadto podaje nam informacje na temat ładunku i stosunku izotopów wykrytych pików w celu ułatwienia identyfikacji metabolitów.

MBRole (Metabolite Biological Role) wykonuje wzbogacaną analizę kategoriycznych adnotacji dla zestawu związków. Adnotacje kategoriyczne odpowiadają informacjom biologicznym i chemicznym dostępnym w wielu publicznych bazach danych oraz oprogramowaniach, aczkolwiek, koncentruje się głównie na adnotacjach jakościowych. MBRole jest wystarczającym narzędziem do przeprowadzenia analizy wzbogacania funkcjonalnego w każdej próbce (zawierając w tym adnotacje biologiczne i chemiczne ludzkich metabolitów).

BiNChE jest narzędziem analizy wzbogacania małych cząsteczek opartym na Cheopsa Ontology. BiNChE wyświetla interaktywny wykres, który można wyeksportować jako obraz o wysokiej rozdzielczości lub w formatach sieciowych. Narzędzie to, zapewnia, jasną, wyważoną i fragmentaryczną analizę opartą na ChEBI Role Ontology lub ChEBI Structural Ontology. BiNChE umożliwia wyszukiwanie dużych zestawów małych cząsteczek w ramach badań metabolomicznych lub innych systemów biologicznych. Narzędzie te zapewnia łatwy i

wysokie interaktywne dostępy przez internet do analizy wzbogacania za pomocą narzędzia ontologicznego ChEBI oraz posiada niezależną bibliotekę danych.

IMPALA jest internetowym narzędziem do analizy danych dotyczących szlaków transkryptomiki lub proteomiki oraz metabolomiki. IMPALA przeprowadza analizę wzbogacania poprzez specyficzne listy metabolitów i genów za pomocą ponad 3000 wstępnie przypisanych szlaków biochemicznych z 11 baz danych. W rezultacie, można zidentyfikować szlaki, które mogą być zaburzone na poziomie transkrypcyjnym, metabolicznym lub obu. Wykrycie zaburzonych szlaków biochemicznych na obu poziomach pozwala na identyfikację szlaków biologicznych o zmienionej aktywności, których nie można uzyskać, gdy analiza stosowana jest tylko na jednym poziomie funkcjonalnym.

MPEA (Metabolite Pathway Enrichment Analysis) jest szybkim narzędziem, stosowanym zarówno do analizy funkcjonalnej jak i biologicznej interpretacji danych profilowania metabolicznego. MPEA stosuje koncepcję analizy wzbogacania genów (GSEA) i sprawdza, jak potencjalne metabolity wpływają na dany szlak biochemiczny (ulegają obniżeniu lub podwyższeniu). W szczególności, MPEA przeznaczony jest do analizy wielu relacji, które mogą występować między metabolitami.

MSEA (Metabolite Set Enrichment Analysis) jest internetowym narzędziem służącym do identyfikacji i interpretacji zmian stężenia metabolitów istotnych biologicznie u ludzi i ssaków. Kluczowym elementem rozwoju MSEA było stworzenie biblioteki około 1000 predefiniowanych zestawów metabolitów, obejmujących różne szlaki metaboliczne, stany chorobowe, biofluidy oraz lokalizacje w tkankach. MSEA wspiera także specjalistyczną analizę niestandardowych, definiowanych przez użytkowników zestawów metabolitów. MSEA oferuje trzy różne analizy wzbogacania dotyczące badań metabolomicznych, w tym analizę nadreprezentatywną (ORA), profilowanie pojedynczych próbek (SSP) oraz ilościową analizę wzbogacania (QEA). ORA wymaga tylko listy nazw związków, podczas gdy SSP i QEA potrzebują zarówno nazw związków jak i ich koncentracji. MSEA generuje wykresy lub tabele zawierające hiperłącza do odpowiednich obrazów szlaków metabolicznych i deskryptorów chorób. W przypadku badań nie na ssakach lub bardziej wyspecjalizowanych, MSEA umożliwia użytkownikom dostarczanie własnych zestawów metabolitów do analizy wzbogacania. Serwer MSEA obsługuje również konwersję między popularnymi nazwami metabolitów, synonimami oraz głównymi identyfikatorami baz danych.

4.1.4. Topologiczna analiza szlaków biologicznych

Struktura szlaków biologicznych reprezentuje naszą wiedzę na temat złożonych korelacji pomiędzy cząsteczkami (aktywacja, hamowanie, reakcja itp.). Jednakże ani analiza nadreprezentatywna ani analiza wzbogacania nie uwzględniają struktury szlaków biologicznych przy określaniu, które szlaki będą bardziej narażone w badanych warunkach. Jest oczywiste, że zmiany w kluczowych pozycjach sieci spowodują silniejsze oddziaływanie na szlak biologiczny niż zmiany w marginalnych lub stosunkowo odizolowanych pozycjach. Istnieją dwie metody pomiaru punktu węzłowego w celu ustalenia jego ważności w szlaku biologicznym- centralność międzyosobowa i stopień centralności. Pierwsza koncentruje się na węźle w stosunku do ogólnej struktury szlaków biologicznych, podczas gdy druga koncentruje się na bezpośrednich powiązaniach lokalnych.

Istnieje wiele komercyjnych narzędzi do analizy szlaków biologicznych, takich jak Pathway Studio, MetaCore, Ingenuity Pathway Analysis itp. W porównaniu z nimi analiza szlaków w programie **MetaboAnalyst** jest bezpłatna, dostępna w sieci, służąca do analizy

danych dotyczących metabolomiki. Wykorzystuje wysokiej jakości szlaki metaboliczne KEGG jako zaplecze wiedzy. Łączy ona, wiele sprawdzonych metod (tj. jednoczynnikowych analiz, nadreprezentatywnych analiz), a także nowych algorytmów/koncepcji (GlobalTest, GlobalAncova, analiza topologii szlaków) w celu analizy szlaków biologicznych. Ponadto MetPA implementuje interaktywny system wizualizacji w stylu Google-Map, aby pomóc użytkownikom zrozumieć wyniki analiz.

4.2. Walidacja biomarkerów: ustalenie istotności metabolitów

Kiedy biomarkery zostaną zaproponowane jako odpowiedni zestaw metabolitów, który mógłby sklasyfikować próbki zgodnie z badanymi warunkami, konieczne jest przeprowadzenie targetowej analizy, ilościowego eksperymentu, zastosowanego do większej grupy próbek (pacjenci, na przykład w badaniach klinicznych), niezależnych od siebie. Wyniki tego eksperymentu, analizowane są w celu oceny ważności takiego biomarkera, a najczęściej stosowana metodologia oparta jest na ocenie krzywej ROC (**receiver operating characteristic**).

Analiza krzywej ROC jest najbardziej obiektywną i statystycznie ważną metodą oceny biomarkerów. Pozwala zbudować model predykcyjny dla jednej lub więcej zmiennych, które można wykorzystać do klasyfikacji nowych, specyficznych cech grup (na przykład zdrowi i chorzy). Krzywa ROC jest wykresem ilustrującym działanie systemu klasyfikacji binarnej, ponieważ jej próg dyskryminacji jest zróżnicowany. Krzywa tworzona jest przez wykreślenie prawdziwie dodatniego wyniku (TPR) względem fałszywego dodatniego wyniku (FPR) przy różnych ustawieniach progowych. Prawdziwie dodatni wynik znany jest również jako czułość, czyli odrzucenie lub prawdopodobieństwo detekcji. Fałszywie dodatni wynik znany jest jako odrzucenie lub prawdopodobieństwo wystąpienia fałszywego wyniku i może być obliczony jako (1 - specyficzność).

Obszar pod krzywą (AUC) mieści się w przedziale od 0,5 do 1,0, a celem badacza jest maksymalizacja AUC pod krzywą ROC przy jednoczesnej minimalizacji liczby metabolitów stosowanych w panelu biomarkerów.

Głównym narzędziem do tworzenia krzywych ROC był Roccet, który obecnie implementowany jest w skrzynce narzędziowej programu MetaboAnalyst.

5. Zajęcia praktyczne z metabolomiki

Praktyczne zajęcia rozpoczną się od wizyty w laboratorium w celu zapoznania się z procesem przygotowania próbek oraz stosowanymi aparaturami analitycznymi.

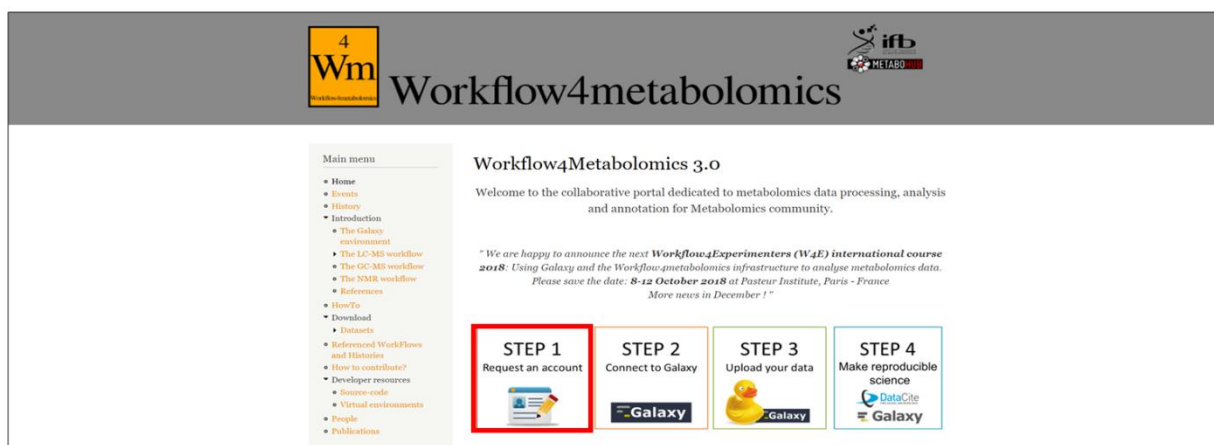
5.1. Metabolomika targetowa i nietargetowa

Protokół przygotowania próbek osocza w celu analizy LC-MS zostanie przedstawiony i omówiony interaktywnie ze studentami w laboratorium. Sesja samouczka pokaże wszystkie kroki wykonane w profesjonalnym środowisku, używając darmowego oprogramowania. Badanie metabolomiczne będzie składało się z dwóch grup: badanej i kontrolnej wraz z próbkami QC analizowanymi za pomocą LC-MS zarówno w jonizacji dodatniej jak i ujemnej.

5.2. Wstępna obróbka danych przy użyciu platformy Workflow for Metabolomics (W4M)

Samouczek ten pokaże Ci, jak przygotować zestaw danych i korzystać z internetowej platformy Workflow4Metabolomics (W4M). W4M jest wspólnym projektem pomiędzy platformami metabolomicznymi (MetaboHUB Francja) oraz bioinformatycznymi (IFB: Instytut Français de Bioinformatique) co doprowadziło do stworzenia platformy typu open-source do badań metabolomicznych.

Przed rozpoczęciem pracy każdy student będzie musiał utworzyć konto w platformie Workflow4Metabolomics. W4M: <http://workflow4metabolomics.org/>. Kliknij "STEP1: „Request an account" i wypełnij formularz. Oświadczenia zostaną dostarczone w ciągu jednego dnia roboczego.

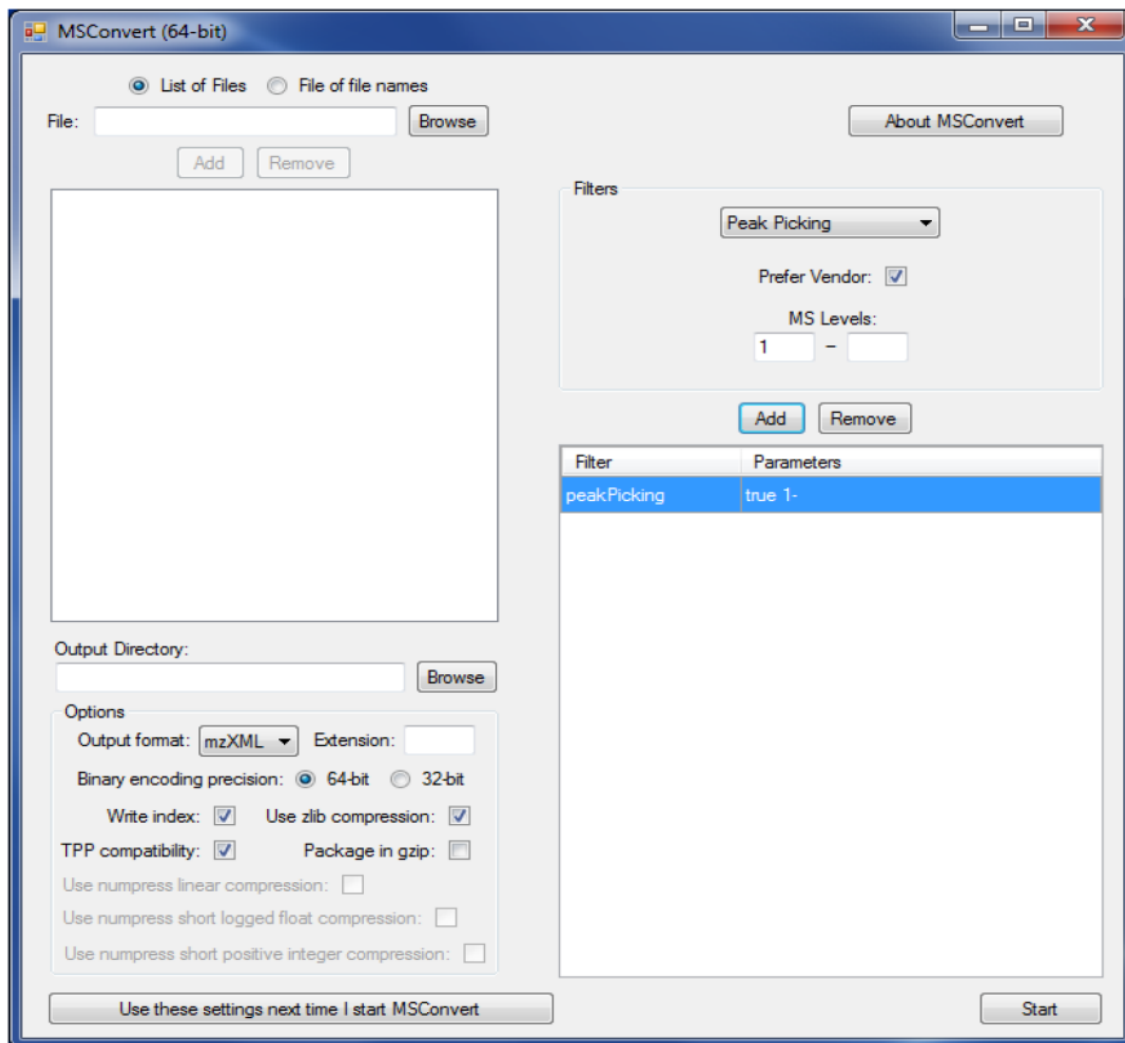


The screenshot shows the homepage of Workflow4Metabolomics 3.0. The header includes the 'Wm' logo and the text 'Workflow4metabolomics'. Below the header, there is a 'Main menu' on the left with a list of links: Home, Events, History, Introduction, The Galaxy environment, The LC-MS workflow, The GC-MS workflow, The NMR workflow, References, HowTo, Download, Datasets, Referenced WorkFlows and Histories, How to contribute?, Developer resources, Source code, Virtual environments, People, and Publications. The main content area features a welcome message: 'Welcome to the collaborative portal dedicated to metabolomics data processing, analysis and annotation for Metabolomics community.' Below this, there is a notice: 'We are happy to announce the next Workflow4Experimenters (W4E) international course 2018: Using Galaxy and the Workflow4metabolomics infrastructure to analyse metabolomics data. Please save the date: 8-12 October 2018 at Pasteur Institute, Paris - France. More news in December!'. At the bottom, there are four steps: STEP 1 (Request an account), STEP 2 (Connect to Galaxy), STEP 3 (Upload your data), and STEP 4 (Make reproducible science). Each step is accompanied by a small icon or logo.

5.2.1. Konwersja plików MS

W platformie W4M pliki MS muszą zostać skonwertowane do formatu otwartego, na przykład .mzML, .mzXML, .mzData i .cdf. Jeśli komercyjne oprogramowanie (zwykle dostarczane z urządzeniem) nie zawiera narzędzia eksportującego, do przetwarzania nieprzetworzonych danych, można użyć innego oprogramowania, jak ProteoWizard (<http://proteowizard.sourceforge.net/>) w celu konwersji plików.

Zaleca się używanie wyłącznie liter (małych i wielkich liter), cyfr i znaków "_" (bez przerw) we wszystkich nazwach (nazwy próbek, nazwy plików, nazwy folderów, zmienne itp.), a także konwertowane pliki raw do formatu mzXML w trybie centroidowym (mniejsze pliki), aby pliki były kompatybilne z metodą xcms centWave.



Rekomendowane parametry dla plików MS podczas konwersji w W4M

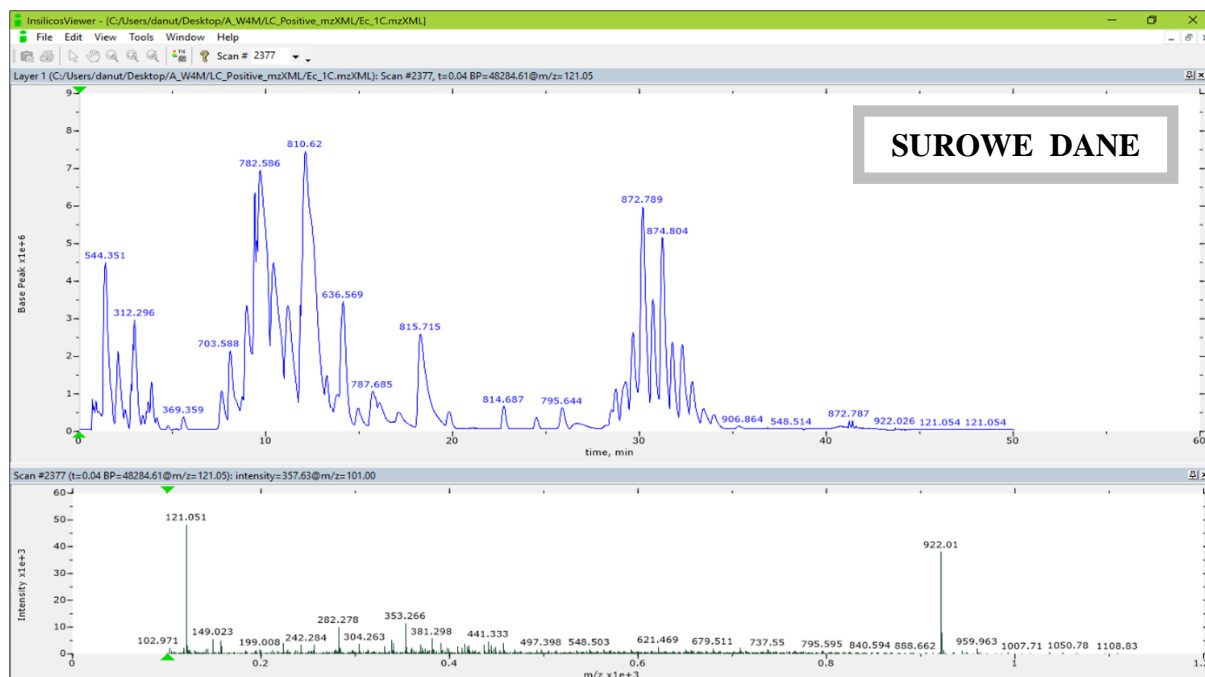
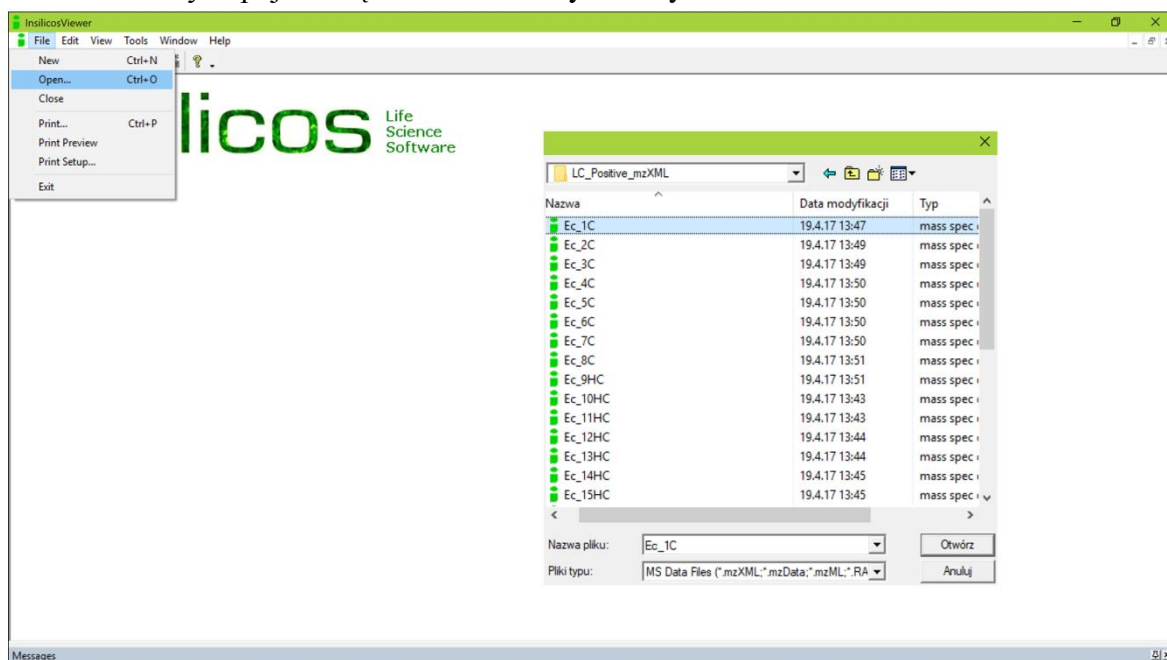
5.2.2. Kontrola surowych danych

Zazwyczaj naukowcy używają oprogramowania dostarczanego przez dostawcę aparatury MS. Jednak, istnieją również alternatywne, otwarte źródła, takie jak InsilicosViewer, darmowa przeglądarka do wyświetlania danych MSZML MS z tą samą funkcjonalnością jaką oferują komercyjne odpowiedniki. Narzędzie te zawiera oprogramowanie do odczytania plików mzXML/mzData opracowane przez Institute of System Biology (<http://www.systemsbio.org>).

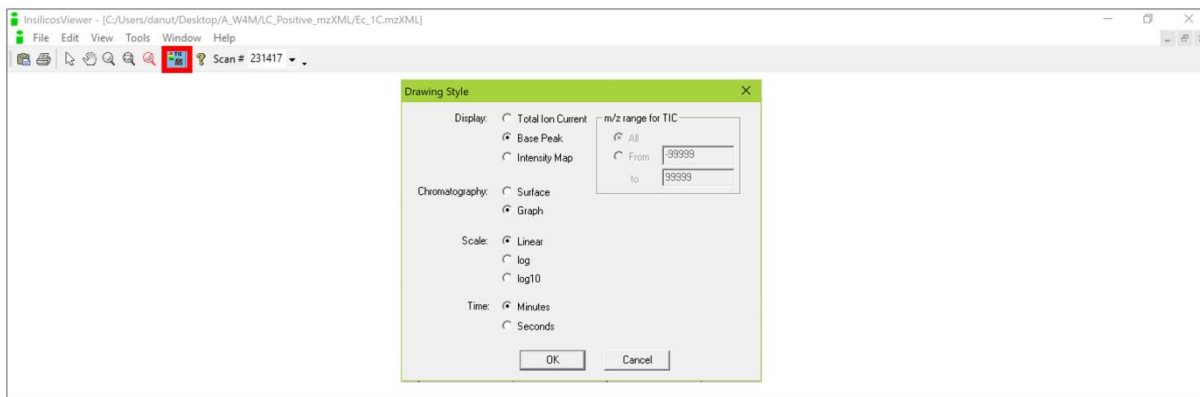
Otwórz surowe dane w formacie mzXML.

1. Wybierz funkcję> Open file.
2. Przejdź do folderu zawierającego dane w formacie mzXML.
3. Wybierz swoje dane.
4. Kliknij Open.

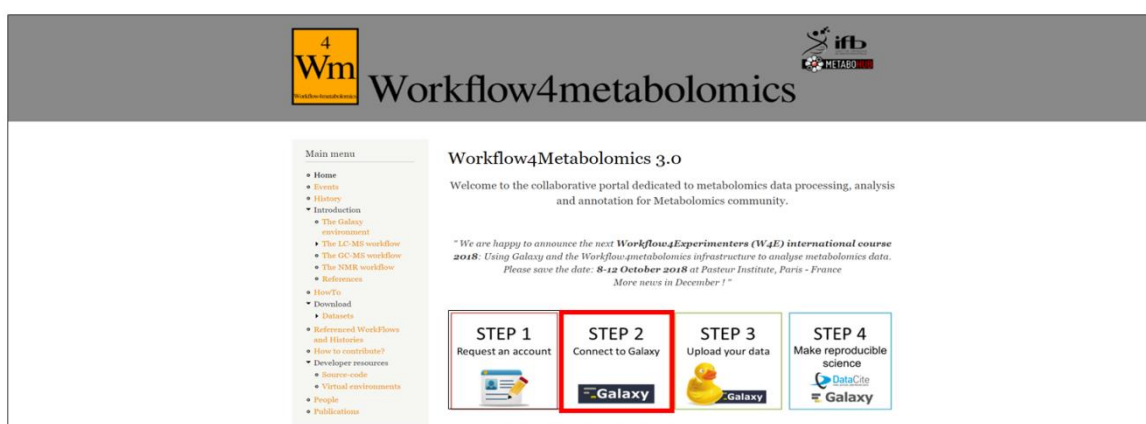
Plik danych pojawi się w oknie surowych danych.



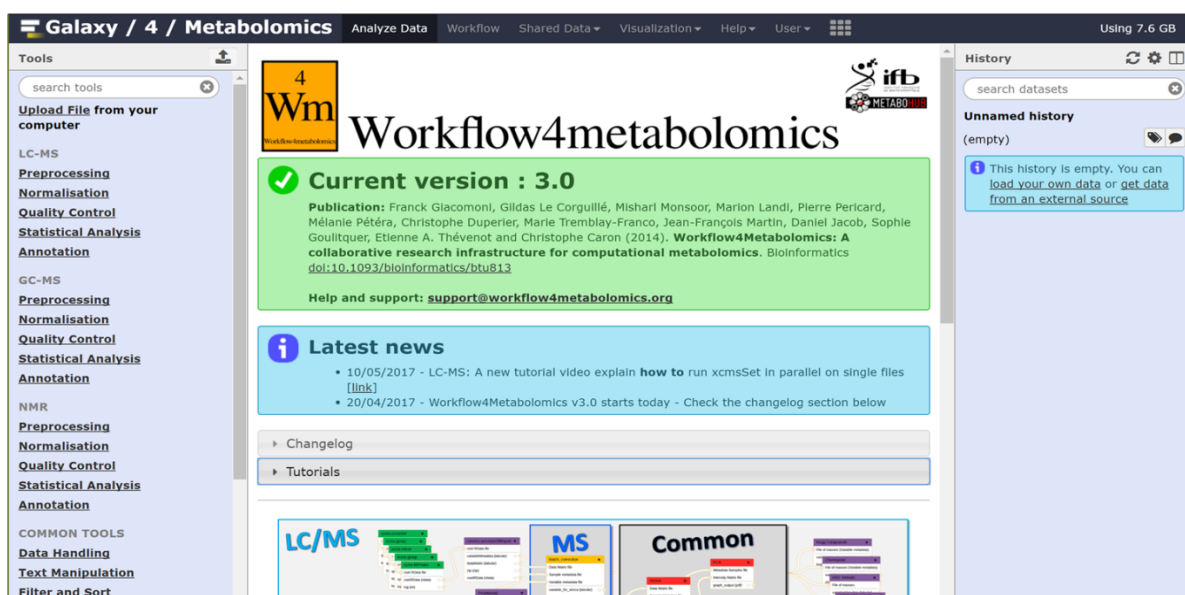
Aby wyświetlić więcej stylów do rysowania kliknij> TIC icon or Tools > Drawing style



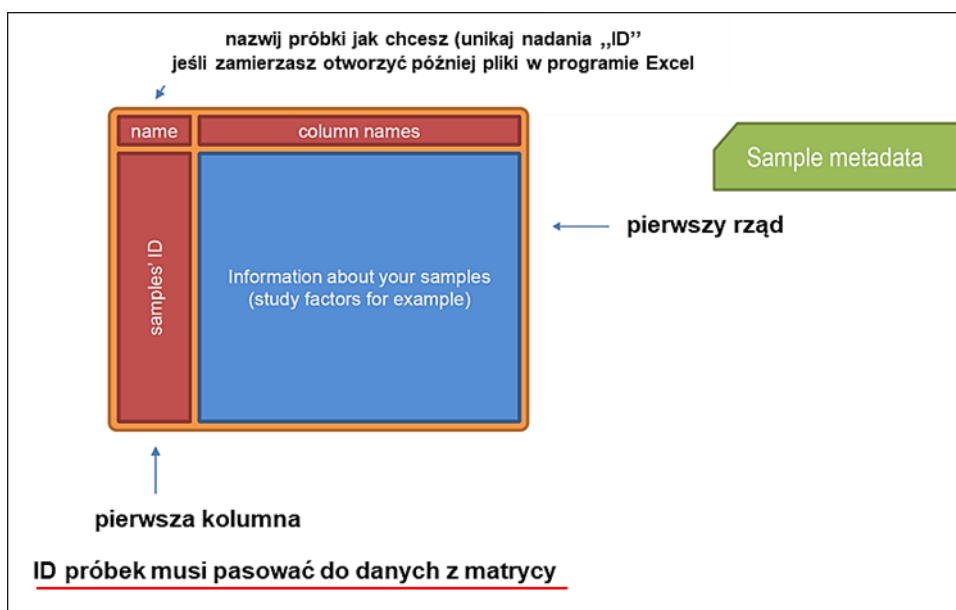
5.2.3. Prześlij dane do Galaxy



Główne okno zostanie wyświetlone.



Przesłanym plikom danych towarzyszą plik meta-danych: informacje dotyczące próbek, informacje te muszą zostać dostarczone do systemu zgodnie z surowymi kryteriami:



Informacje o próbkach w sesji praktycznej muszą być zgodne ze strukturą opisaną w tabeli:

sample name	class	polarity	sample type	batch	injection order	diet
QC	one	positive	pool	B1	1	NA
C1	one	positive	sample	B1	7	C
HC3	one	positive	sample	B1	10	HC
BL	one	positive	blank	B1	12	NA
...

Pliki muszą być zestawione w tabeli: pliki TSV lub TXT z tabulacją jako separatorem.

Istnieją dwie metody przesyłania danych: zestaw pojedynczych plików lub tworzeniem najpierw pliku .zip, z predefiniowaną strukturą podkatalogów. Będziemy ćwiczyć tylko pierwszą opcję .:

- pojedynczy plik (zalecane): Możesz wstawić pojedynczy plik jako dane wejściowe. W ten sposób będziesz mógł uruchomić kilka xcmsSet równolegle i użyć "xcms.xcmsSet Merger" przed "xcms.group"
- plik Zip: możesz umieścić plik zip zawierający dane wejściowe: myinputs.zip (zawierające wszystkie twoje warunki jako podkatalogi).

Pliki Zip: Poszczególne kroki w celu utworzenia plików Zip.

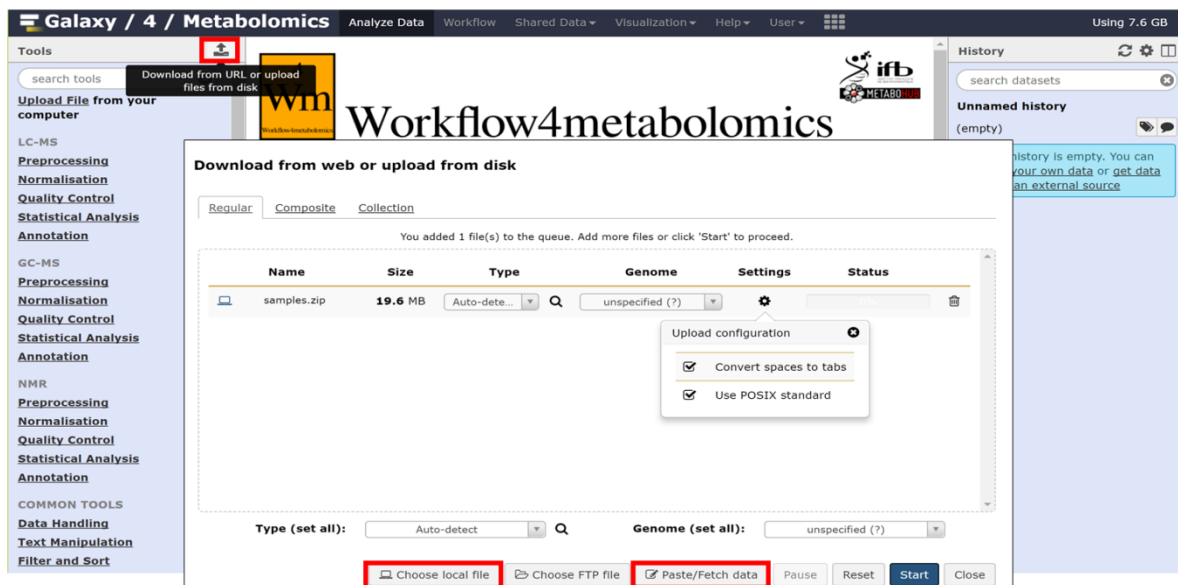
Krok 1: Tworzenie katalogu i hierarchizacja podkatalogów

BARDZO WAŻNE: Jeśli spakujesz pliki w systemie Windows, musisz użyć oprogramowania 7Zip (<http://www.7-zip.org/>), w przeciwnym razie twoje pliki zip nie zostaną rozpakowane na platformie W4M (wystąpi błąd).

Krok 2: Tworzenie plików Zip

Importowanie danych<2Gb

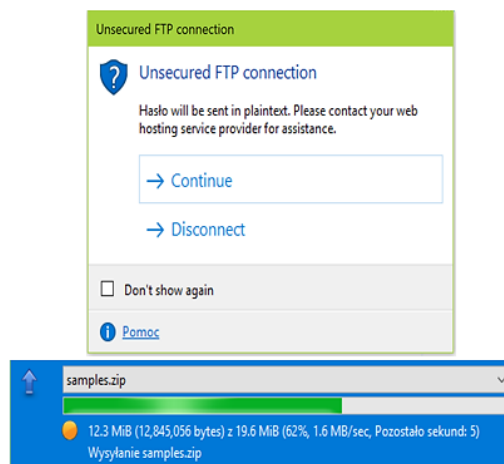
Aby zaimportować dane użyj opcji Paste/Fetch (Kopiuj/Wklej) dane lub wybierz plik lokalny> z opcji Upload configuration kliknij Convert spaces to tabs<Start<Close.



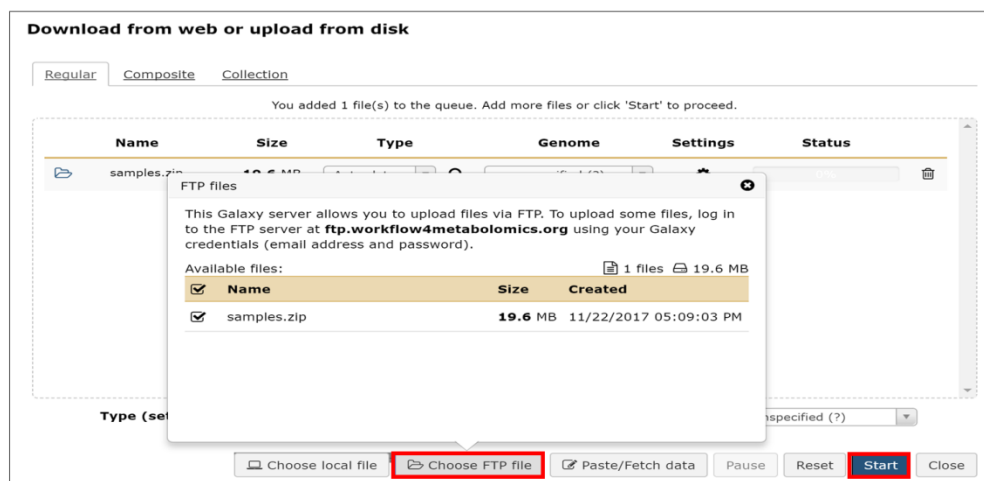
Importowanie danych>2Gb

Krok 1: Wybierz FTP Client<przejdź do Cyberduck<Otwórz Connection<Server: <ftp.workflow4metabolomics.org>> Utwórz login oraz hasło, następnie otwórz Connection i skopiuj.

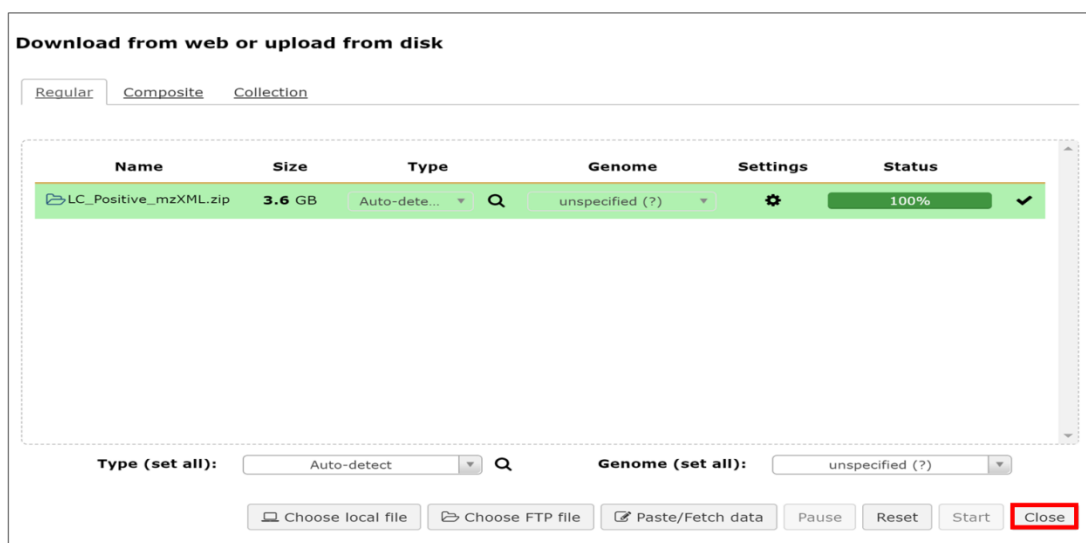




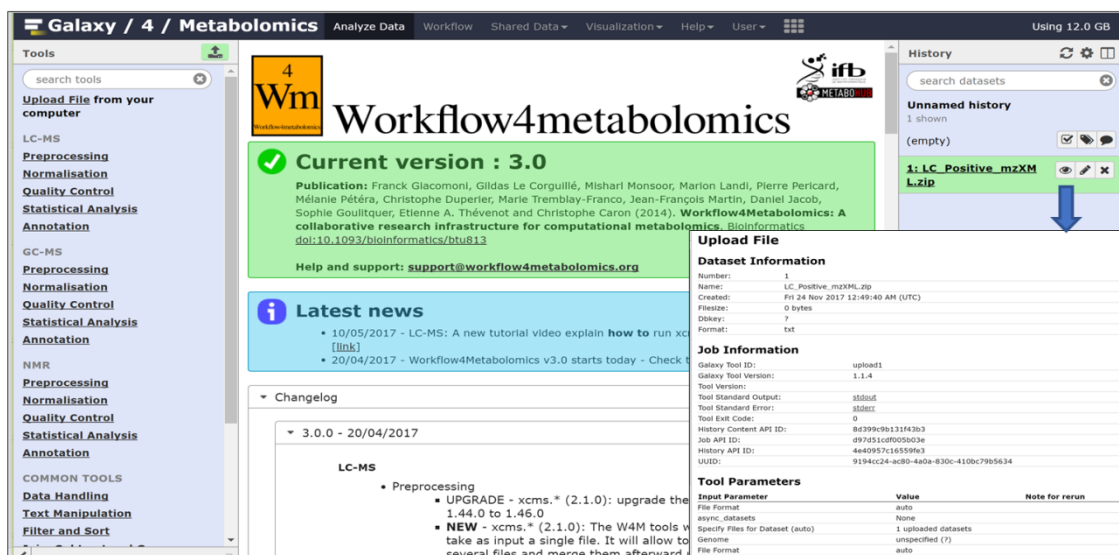
Jeśli przesyłanie danych dobiegnie końca wybierz ponownie plik FTP> wybierz swoje dane> Start.



Twoje dane zostaną przesłane na platformę WM.

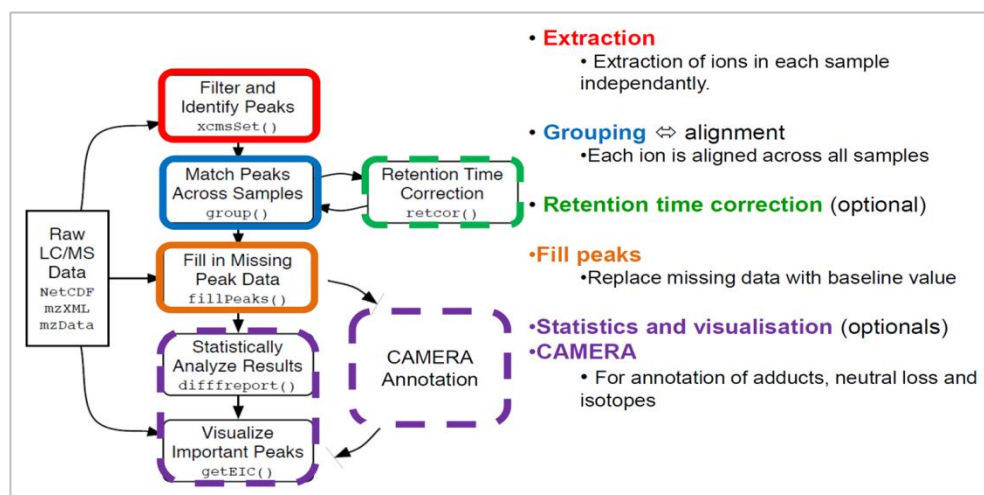


Możesz podejrzeć detale Twoich zaimportowanych danych.



5.2.4. Wstępne przetwarzanie danych LC-MS za pomocą XCMS.

XCMS to pakiet oprogramowania R, przeznaczony do ekstrakcji pików oraz wyrównania czasu retencji plików danych z aparatury GC-MS i LC-MS.



Ekstrakcja pików- xcmsSet

Xcms.xcmsSet Filtracja i identyfikacja pików za pomocą funkcji xcmsSet z pakietu xcms R do procesowania wstępnego danych LC /MS dla względnej analizy ilościowej i statystycznej

Narzędzie te służy do wstępnego przetwarzania danych z plików LC/MS (formaty NetCDF, mzXML i mzData). Ekstrahuje jony z każdej próbki niezależnie i przy użyciu modelu statystycznego, piki są filtrowane oraz integrowane

Pliki wejściowe:

Parameter : num + label	Format
OR : Zip file	zip
OR : Single file	mzXML, mzML, mzData, netCDF

xcms.xcmsSet Filtration and Peak Identification using xcmsSet function Versions Options
from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
(Galaxy Version 2.1.0)

File(s) from your history containing your chromatograms

No mzxml, mzml, mzdata, netcdf, no_unzip.zip or zip dataset available.

Single file mode for the format: mzxml, mzml, mzdata and netcdf. Zip file mode for the format: no_unzip.zip, zip. See the help section below.

Scan range option

hide

Extraction method for peaks detection

centWave

[method] See the help section below

Max tolerated ppm m/z deviation in consecutive scans in ppm

25

[ppm]

Min,Max peak width in seconds

20,50

[peakwidth]

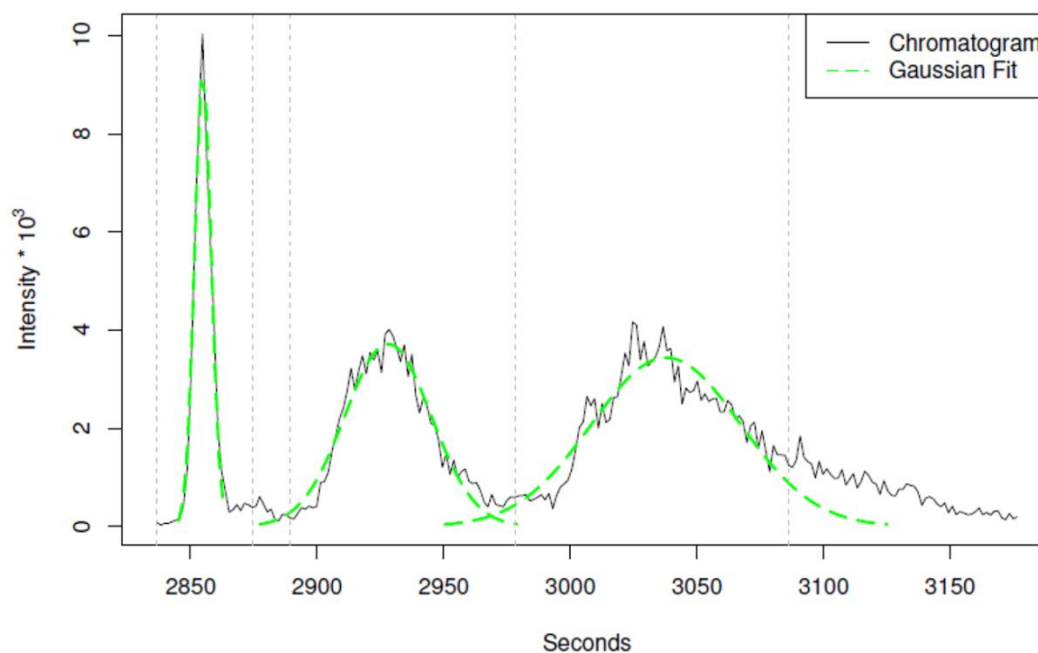
Advanced options

hide

Parametry:

Detekcja pików za pomocą metody ekstrakcji:

- **Dopasowanie filtrów:** dedykowany jest do danych MS centroidowych lub o profilu o niskiej rozdzielczości
- **Centwave:** metoda dydykowana dla danych centroidowych o wysokiej rozdzielczości. Algorytm ma na celu wykrywanie "masowych śladów" lub "obszaru zainteresowania (ROI)", które są zdefiniowane jako regiony z mniej niż zdefiniowanym odchyleniem m/z w kolejnych skanach. Odchylenie te musi być niższe niż wartość parametru "ppm".



- **MSW:** Ciągła transformacja falkowa. Metoda ta opiera się na bezpośredniej infuzji danych, może być użyta w celu lokalizacji pików chromatograficznych w różnych skalach.

Możesz sprawdzić typ swoich danych w pliku surowych danych.

```
<MS1CentroidDataAbsThreshold>200</MS1CentroidDataAbsThreshold>
<MS1CentroidDataRELThreshold>0.010</MS1CentroidDataRELThreshold>
<MS2CentroidDataAbsThreshold>5</MS2CentroidDataAbsThreshold>
<MS2CentroidDataRELThreshold>0.010</MS2CentroidDataRELThreshold>
<TimeSegment>
  <Index>1</Index>
  <StartTime>0.0</StartTime>
  <DiverterValveState>MS</DiverterValveState>
  <StorageMode>Centroid</StorageMode>
  <IonMode>Dual LSI</IonMode>
  <TimeSourceParameter>
    <Id>FGasReater</Id>
    <Value>350</Value>
  </TimeSourceParameter>
  <TimeSourceParameter>
    <Id>FGasFlow</Id>
    <Value>10.0</Value>
  </TimeSourceParameter>
  <TimeSourceParameter>
    <Id>NebulizerPressure</Id>
    <Value>40</Value>
  </TimeSourceParameter>
</TimeSegment>
```

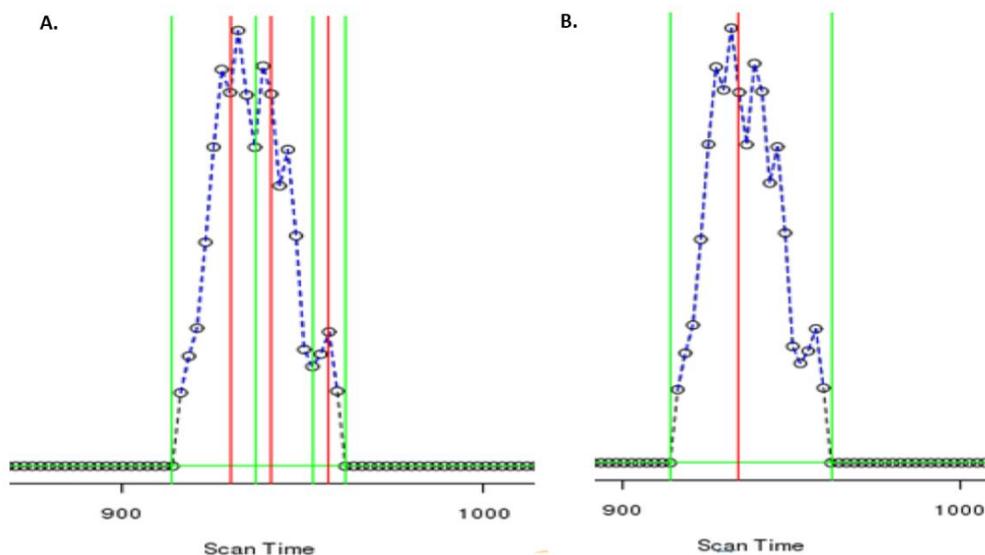
Maksymalna tolerancja odchylenia ppm m/z w kolejnych skanach w ppm (w odniesieniu do m/z)

Fluktuacja m/z wartości (ppm) od skanowania do skanowania. Ten parametr należy ustawić zgodnie z dokładnością spektrometru mas.

Min. Maks. szerokość pików w sekundach: (w zależności od RT)

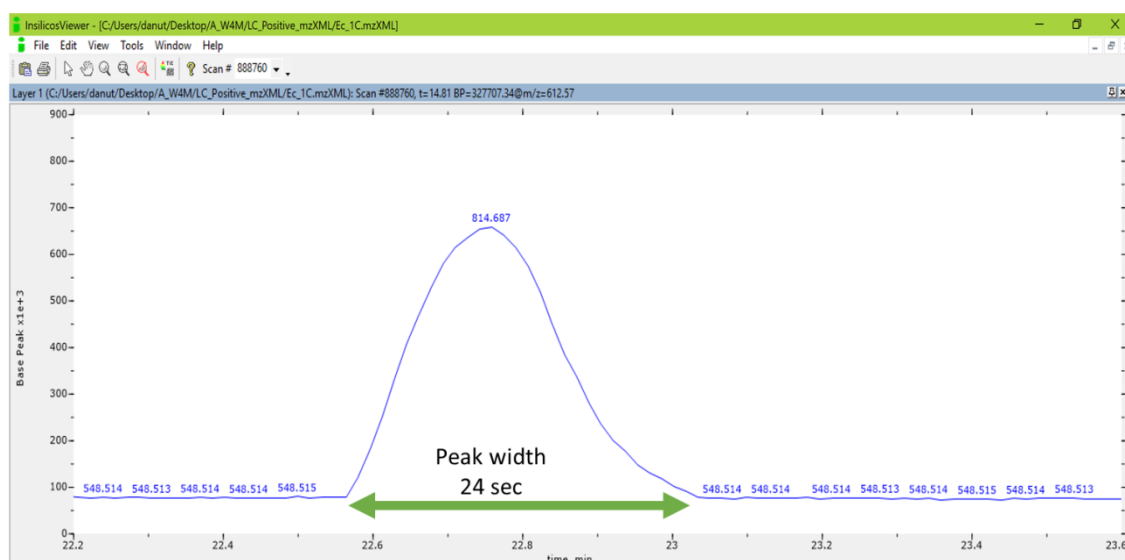
Głównym celem tego parametru jest oszacowanie zakresu szerokości pików, parametr ten nie jest wartością graniczną. Falkowanie wykorzystywane do detekcji pików obliczane są na podstawie tego parametru.

Ważne: Nie wybieraj minimalnego pików za małego, nie zwiększy to czułości, ale spowoduje podział (rozszerzenie) pików.



Przykład: szerokość pików 45s. A. Stosując szerokość pików (20-60) pik będzie podzielony na trzy pików, każdy wykryty jako oddzielny pik o szerokości 10 sekund; B. Stosując szerokość pików (20-120) otrzymamy jeden pik.

Możesz sprawdzić ten parametr w swoich surowych danych. Oblicz szerokość pików dla wąskiego i szerokiego pików.



Opcje zaawansowane:

Advanced options

show

Signal/Noise threshold

10

[snthresh] Signal to noise ratio cutoff

Minimum difference in m/z for peaks with overlapping retention times

-0.0015

[mzdiff] Can be negative to allow overlap

peak limits method

peak limits based on smoothed 2nd derivative (less precise)

[integrate]

Prefilter step for the first phase

3,1000

[prefilter] Separate by coma k,I. Mass traces are only retained if they contain at least 'k' peaks with intensity >= 'I'

Noise filter

0

[noise] optional argument which is useful for data that was centroided without any intensity threshold, centroids with intensity smaller than 'noise' are omitted from ROI detection

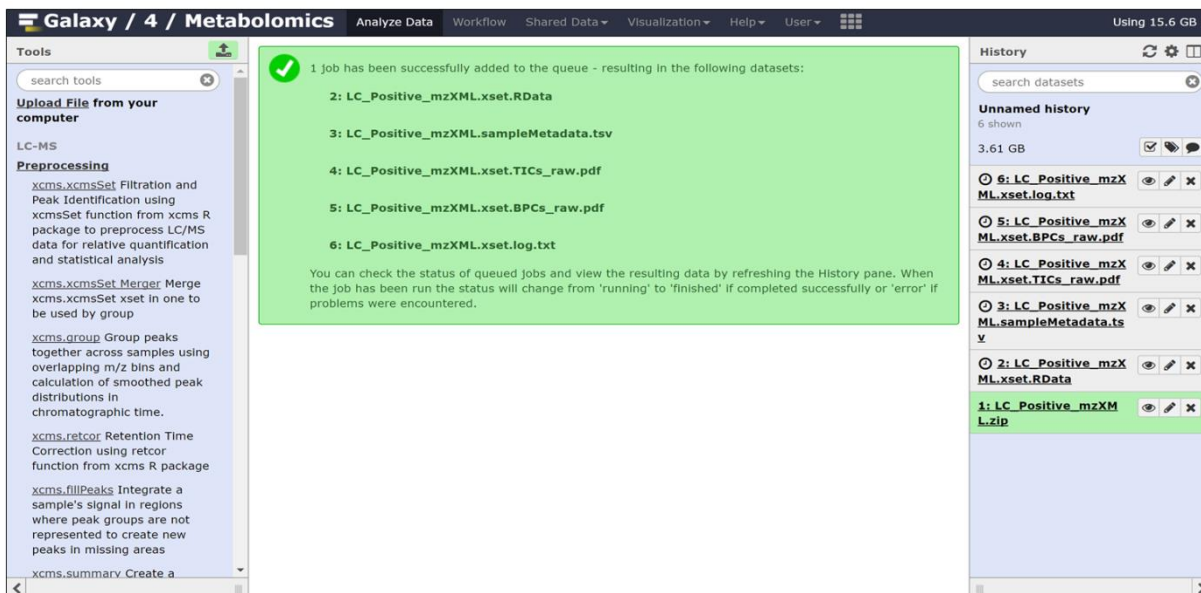
Minimalna różnica m/z dla pików z nakładającym się na siebie czasem RT (w odniesieniu do m/z i RT):

Minimalna różnica m/z dla pików z nakładającym się czasem RT (piki koeluuujące) musi być ujemna, aby umożliwić nakładanie się pików.

Filtr wstępny (w zależności od intensywności):

Pik musi być obecny w n skanach o intensywności większej niż k.

Po ustawieniu wszystkich parametrów naciśnij Execute i poczekaj ... Ten krok wymaga czasu.



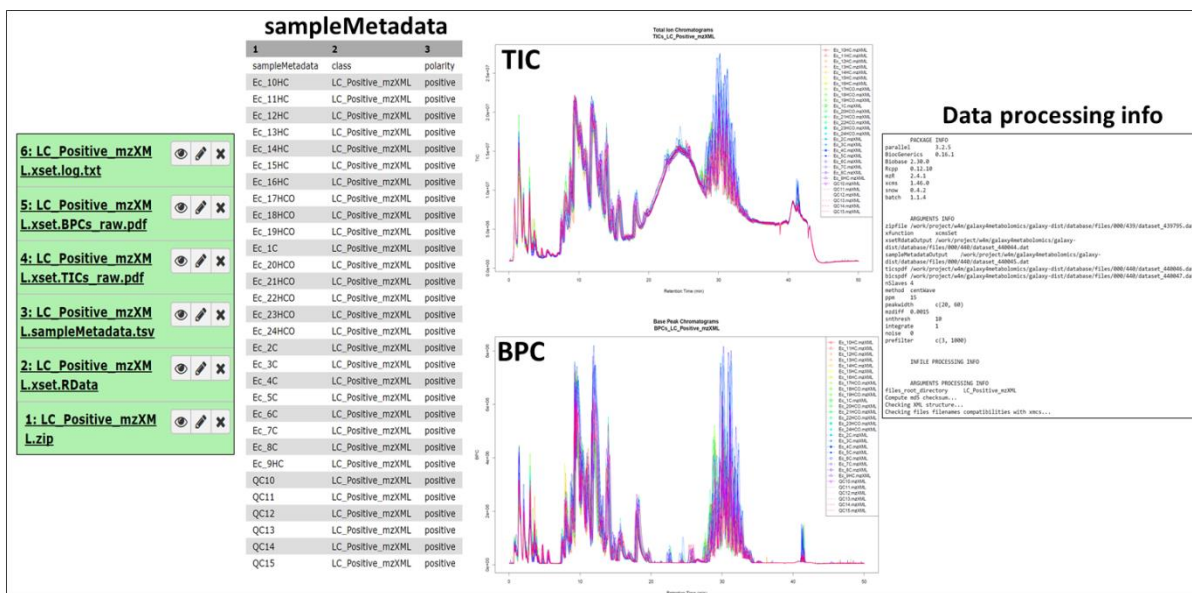
Pliki wyjściowe (historia):

xset.TICs_raw.pdf Wykres "Total Ion Chromatograms" w formacie pdf.

xset.BPCs_raw.pdf Wykres "Base Peak Chromatograms" w formacie pdf z przeciwstawnymi próbkami klas.

sampleMetadata.tsv Plik tabelaryczny, który zawiera dla każdej próbki, podział na klasę oraz polaryzację (dodatnią, ujemną i mieszaną). Ten plik jest niezbędny podczas analizy Anova i PCA.

xset.RData: rdata.xcms.raw format. Plik Rdata, jest niezbędny w drugim etapie pracy "xcms.group".



Plik wyjściowy to plik xset.RData. Możesz kontynuować analizę, używając go w narzędziu xcms.group.

xcms.group Grupuje piki w poprzek próbek przy użyciu nakładających się boksów m / z oraz oblicza wygładzone rozkłady pików w czasie chromatograficznym

Po dopasowaniu pików w grupach, xcms może wykorzystać te grupy do zidentyfikowania i skorygowania skorelowanych dryfów w czasie retencji od startu do startu. Wyrównane piki można następnie wykorzystać do drugiego grupowania pików, które będzie dokładniejsze niż pierwsze.

Cały proces można powtórzyć w sposób interakcyjny (powtarzalny). Nie wszystkie piki w grupach będą pomocne w identyfikacji dryfu czasu retencji. W niektórych grupach w dużej części próbek mogą znajdować się brakujące piki, co powoduje niepełny obraz dryfu w danym punkcie czasowym. Jeszcze inne mogą zawierać wiele pików z tej samej próbki, co jest oznaką niewłaściwego grupowania.

Po identyfikacji pików za pomocą xcmsSet, narzędzie te grupuje piki, które reprezentują ten sam analit w próbkach, z wykorzystaniem nakładania się m/z przedziału oraz obliczanie wygładzonych rozkładów pików w czasie chromatograficznym. Umożliwia to odrzucenie cech, które wykrywane są tylko częściowo w replikatach z przykładowej klasy próbek.

	pool1B1			pool1B2			pool1B3		
	mz	rt	int	mz	rt	int	mz	rt	int
Independent peak lists	196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933325
	158.1180	67.4	71736	342.0310	69.0	74594	158.1173	67.4	82969
	342.0308	67.6	202268	267.0581	65.5	260877	342.0308	21.3	2581
	267.0581	65.5	282039	283.0318	65.2	424631	283.0320	65.3	357448
Group ions by m/z	mz	rt	int	mz	rt	int	mz	rt	int
	196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933325
	158.1180	67.4	71736	342.0310	69.0	74594	158.1173	67.4	82969
	342.0308	67.6	202268	267.0581	65.5	260877	342.0308	21.3	2581
Group ions by RT	mz	rt	int	mz	rt	int	mz	rt	int
	196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933325
	158.1180	67.4	71736				158.1173	67.4	82969
	342.0308	67.6	202268	342.0310	69.0	74594	342.0308	21.3	2581
			267.0581	65.5	260877				
			283.0318	65.2	424631	283.0320	65.3	357448	
Resulting matrix									
	mz	rt	pool1B1	pool1B2	pool1B3				
	196.0905	66.6	7810936	11733921	7933325				
	158.1176	67.4	71736		82969				
	342.0308	21.3			2581				
	342.0309	68.3	202268	74594					
	267.0581	65.5	282039	260877					
	283.0319	65.2		424631	357448				

Pliki wejściowe:

Parameter : num + label	Format
Or : RData file rdata.xcms.raw	
Or : RData file rdata.xcms.retcor	

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time. (Galaxy Version 2.1.0)

versions Options

xset RData file

No rdata.xcems.raw, rdata.xcems.group, rdata.xcems.retcor or rdata dataset available.

output file from another function xcms (xcmsSet, retcor etc.)

Method to use for grouping

density

[method] See the help section below

Bandwidth

30

[bw] bandwidth (standard deviation or half width at half maximum) of gaussian smoothing kernel to apply to the peak density chromatogram

Minimum fraction of samples necessary

0.5

[minfrac] in at least one of the sample groups for it to be a valid group

Width of overlapping m/z slices

0.01

[mzwid] to use for creating peak density chromatograms and grouping peaks across samples

Advanced options

show

Maximum number of groups to identify in a single m/z slice

50

[max]

Get a Peak List

Parametry:

Metoda używana do grupowania:

mzClust. Uruchamia wyrównanie o wysokiej rozdzielczości w próbkach pojedynczych widm przechowywanych w pliku RData, wygenerowanym przez narzędzie **xcmsSet**.

Gęstość. Grupowanie pików razem wzdłuż próbek przy użyciu nakładania się m/z przedziału oraz obliczanie wygładzonych rozkładów pików w czasie chromatograficznym.

Najbliższy. Grupowanie pików razem wzdłuż próbek, tworząc listę pików wzorcowych i przypisując odpowiednie piki ze wszystkich próbek. Używany jest algorytm dopasowywania mzMine.

Pasma (przepustowość) [bw] (związane z RT)

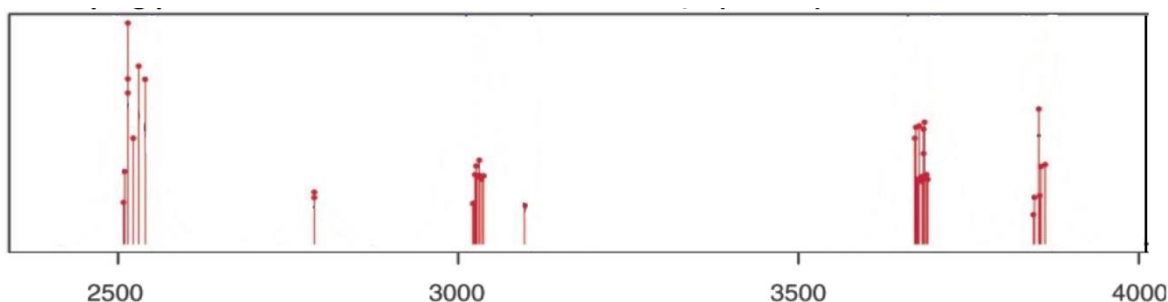
Standardowe odchylenie metapiku gaussowskiego, które grupuje razem piki.

Szerokość nakładających się na siebie przedziału m/z [mzwid] (związanych z m/z)

Rozmiar zakresu m/z (przedziału). Zakres m/z powinien być włączony do grupy. Zależy od dokładności spektrometru mas.

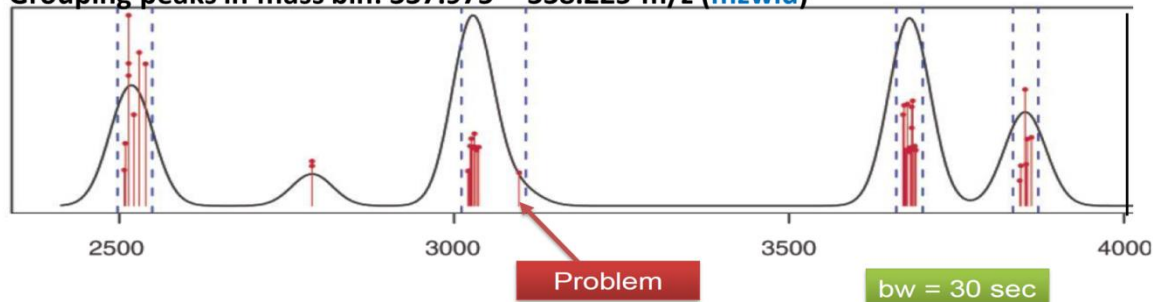
Dokonuje się przedziału domeny masowej. mzwid określił rozmiar przedziału mas. Następnie dla każdego przedziału m/z, wszystkie jony wszystkich próbek są brane pod uwagę dla wszystkich czasów retencji. Metoda estymatora gęstości Kernela służy do wykrywania regionu czasu retencji przy dużej gęstości jonów.

Grupowanie pików w przedziale mas: 337,975 - 338,225 m / z (mzwid)

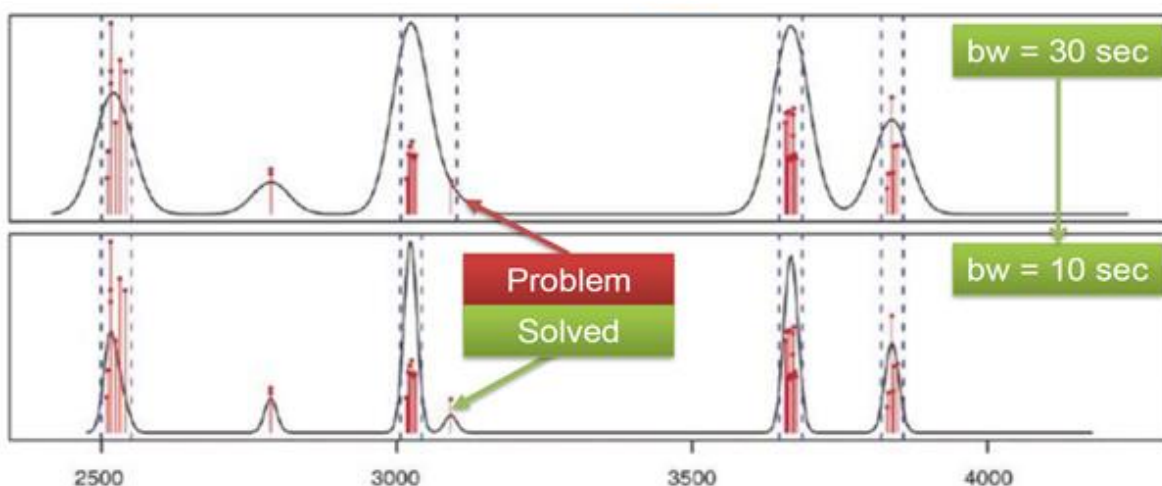


- Model gaussowski grupuje piki razem o zbliżonych RT.
- Inkluzywność jonów w grupie jest określona przez standardowe odchylenie modelu gaussowskiego (pasmo) odpowiadające parametrowi bw (może być interpretowane jako okno RT).
- Pionowe linie przerywane wskazują, że funkcja jest poprawna i zostanie zachowana w macierzy danych.
- Liczba pików w grupie musi być większa niż procent całkowitej liczby próbek. Ten próg jest zdefiniowany przez parametr **minfrac**.

Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)



Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)



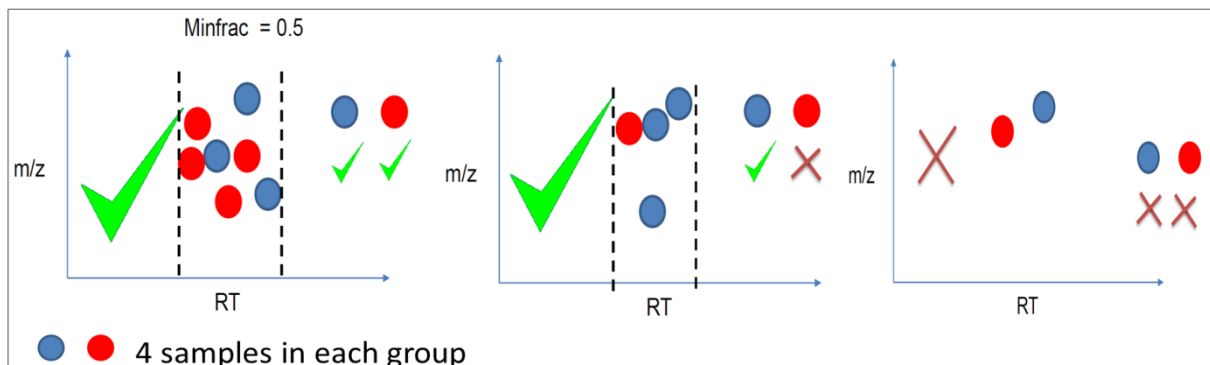
- Zmniejszenie bw pozwala rozdzielić te 2 grupy.
- Wyniki cech m/z i RT odpowiadają medianie m/z i RT wszystkich jonów zgrupowanych razem jako jedna cecha.

Minimalna potrzeba frakcjonowania próbki [minfrac] (w odniesieniu do próbek)

Minimalna część całkowitej liczby próbek dla grupy, która ma być uznana za ważną (minimalna próbka wykryta w co najmniej jednej klasie, która ma być uznana za grupę). Minfrac = 0.5 odpowiada 50%. (n = 10 minfrac = 0.5 znalezione w co najmniej 10 z 50).

Max (w zależności od liczby jonów)

Maksymalna liczba grup wykrytych w jednym zakresie m/z.



Pliki wyjściowe (historia)

xset.group.RData: rdata.xcms.group format

xset.group.Rplots.pdf

Rdata file będzie to konieczne na trzecim i czwartym etapie pracy (xcms.retcor i xcms.fillpeaks).

Galaxy / 4 / Metabolomics

Analyze Data Workflow Shared Data Visualization Help User

Using 15.7 GB

Tools

search tools

Upload File from your computer

LC-MS

Preprocessing

xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis

xcms.xcmsSet Merger Merge xcms.xcmsSet xset in one to be used by group

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.

xcms.retcor Retention Time Correction using retcor function from xcms R package

xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas

xcms.summary Create a summary of XCMS analysis

1 Job has been successfully added to the queue - resulting in the following datasets:

7: LC_Positive_mzXML.xset.group.RData

8: LC_Positive_mzXML.xset.group.Rplots.pdf

9: LC_Positive_mzXML.xset.group.variableMetadata.tsv

10: LC_Positive_mzXML.xset.group.dataMatrix.tsv

11: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from "running" to "finished" if completed successfully or "error" if problems were encountered.

History

search datasets

Unnamed history

10 shown

3.62 GB

10: LC_Positive_mzXML.xset.group.dataMatrix.tsv

9: LC_Positive_mzXML.xset.group.variableMetadata.tsv

8: LC_Positive_mzXML.xset.group.Rplots.pdf

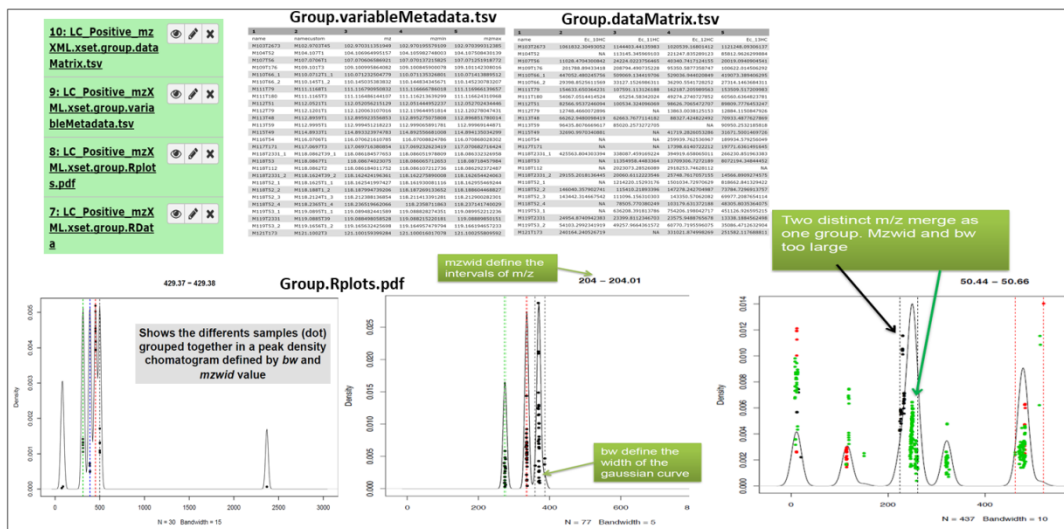
7: LC_Positive_mzXML.xset.group.RData

6: LC_Positive_mzXML.xset.log.txt

5: LC_Positive_mzXML.xset.BPCs_raw.pdf

4: LC_Positive_mzXML.xset.TICs_raw.pdf

3: LC_Positive_mzXML.sampleMetadata.tsv



1. xcms.retcor (wyrównanie)

xcms.retcor Korekta czasu retencji za pomocą funkcji rektora z pakietu xcms R.

Po dopasowaniu pików do grup, xcms może wykorzystać te grupy do zidentyfikowania i skorygowania skorelowanych dryfów w czasie retencji od analizy do analizy. Wyrównane piki można następnie wykorzystać do drugiego przebiegu grupowania pików, które będzie dokładniejsze niż pierwsze. Cały proces można powtórzyć. Nie wszystkie piki w grupie będą pomocne w identyfikacji dryfu czasu retencji. Niektóre grupy mogą zawierać brakujące piki w dużej części próbek, co powoduje niepełny obraz dryfu w tym punkcie czasowym. Jeszcze inne mogą zawierać wiele pików z tej samej próbki, co jest oznaką niewłaściwego grupowania.

Pliki wejściowe

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

xcms.retcor Retention Time Correction using retcor function from xcms R package (Galaxy Version 2.1.0) Versions Options

xset RData file

No rdata.xcms.raw, rdata.xcms.group or rdata dataset available. ▼
 output file from another function xcms (xcmsSet, retcor etc.)

Method to use for retention time correction

peakgroups ▼
 [method] See the help section below

Smooth method

loess ▼
 [smooth] either 'loess' for non-linear alignment or 'linear' for linear alignment

Number of extra peaks to allow in retention time correction correction groups

1
 [extra]

Number of missing samples to allow in retention time correction groups

1 Number of admitted missing well behaved peak in a group.
 [missing]

Advanced options

hide ▼

Resubmit your raw dataset or your zip file 🔗

Parametry

Metoda

➤ grupowanie pików

xcms bierze pod uwagę tylko te grupy pików, które nie mają co najwyżej jednej próbki i mają co najwyżej jeden dodatkowy pik. (Wartości te można zmienić za pomocą funkcji **missing** i **extra**).

Dla każdej z tych grup algorytm oblicza medianę RT i dla każdej próbki odchylenie od tej mediany. W obrębie próbki zaobserwowane odchylenie na ogół zmienia się w czasie w sposób nieliniowy. Zmiany te są przybliżane za pomocą techniki regresji wielomianowej, zaimplementowanej w funkcji loess. Domyślne dopasowywanie krzywej odbywa się za pomocą metody najmniejszych kwadratów we wszystkich punktach danych. Jednakże, możliwe jest jednak wykrywanie outlajerów i usuwania wartości odstających poprzez ustawienie argumentu family na symetryczny.

➤ metoda obiwarp

Obliczenie odchylenia RT dla każdej próbki za pomocą kodu obiwarp na at "<http://obiwarp.sourceforge.net/>". Funkcja ta jest w stanie wyrównać wiele próbek za pomocą strategii gwiazdy centralnej. Ordered Bijective Interpolated Warping (OBI-Warp) wyrównuje macierze wzdłuż pojedynczej osi za pomocą dynamicznego dopasowania czasowego (DTW) i interpunowanej interpunkcji jeden-do-jednego (bijective). OBI-Warp wykorzystuje nieliniową, funkcję DTW i opiera się na dyskretnym, nie-bijektywnym wyjściu DTW, aby uzyskać naturalne interpolanty, które można wykorzystać w wielu zestawach danych.

Advanced options

show

Degree of smoothing for local polynomial regression fitting

0.2

[span]

Family

gaussian

[family] if gaussian fitting is by least-squares with no outlier removal, and if symmetric a re descending M estimator is used with Tukey's biweight function, allowing outlier removal

plotype

deviation **Plot to visualize the result of the retention time correction.**

[plotype] if deviation plot retention time deviation points and regression fit, and if mdevden also plot peak overall peak density and retention time correction peak density

Resubmit your raw dataset or your zip file

Execute

Pliki wyjściowe (historia)

xset.group.retcor.TICs_corrected.pdf Wykres TIC w formacie pdf, poprawiony po kroku retcor.

xset.group.retcor.BPCs_corrected.pdf Wykres BPC w formacie pdf, poprawiony po kroku retcor.

xset.group.retcor.RData: rdata.xcms.retcor Plik Rdata, które będą niezbędne w kroku xcms.group.

Plik wyjściowy to plik xset.retcor.RData. Możesz kontynuować analizę, używając narzędzia **xcms.fillPeaks**.

Po kroku retcor należy wykonać krok grupowania, w przeciwnym razie następne etapy nie będą działać z plikiem RData. (początkowe zgrupowanie pików staje się nieważne i zostaje odrzucone).

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Help User Using 15.8 GB

Tools

search tools

Upload File from your computer

LC-MS

Preprocessing

xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis

xcms.xcmsSet Merger Merge xcms.xcmsSet xset in one to be used by group

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.

xcms.retcor Retention Time Correction using retcor function from xcms R package

xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas

xcms.summary Create a summary of XCMS analysis

CAMERA.annotate CAMERA

1 job has been successfully added to the queue - resulting in the following datasets:

- 12: LC_Positive_mzXML.xset.group.retcor.RData
- 13: LC_Positive_mzXML.xset.group.retcor.Rplots.pdf
- 14: LC_Positive_mzXML.xset.group.retcor.TICs_corrected.pdf
- 15: LC_Positive_mzXML.xset.group.retcor.BPCs_corrected.pdf
- 16: xset.log.txt

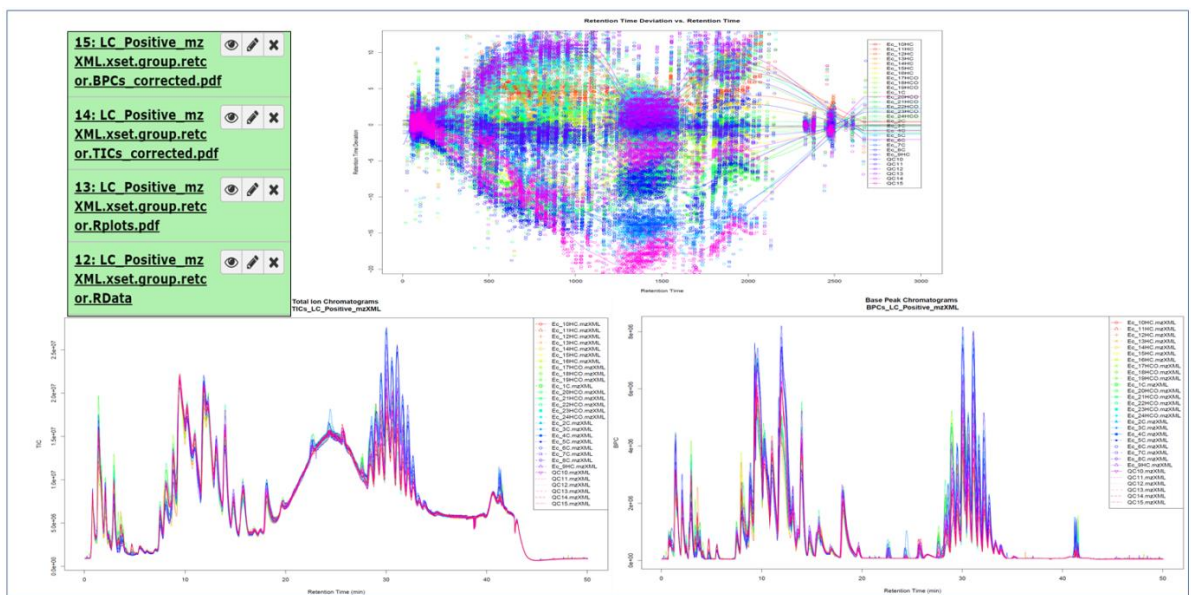
You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Unnamed history
14 shown, 1 hidden
3.71 GB

- 15: LC_Positive_mzXML.xset.group.retcor.BPCs_corrected.pdf
- 14: LC_Positive_mzXML.xset.group.retcor.TICs_corrected.pdf
- 13: LC_Positive_mzXML.xset.group.retcor.Rplots.pdf
- 12: LC_Positive_mzXML.xset.group.retcor.RData
- 10: LC_Positive_mzXML.xset.group.dataMatrix.tsv
- 9: LC_Positive_mzXML.xset.group.variableMetadata.tsv
- 8: LC_Positive_mzXML.xset.group.Rplots.pdf
- 7: LC_Positive_mzXML.xset.group.RData



Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Help User Using 15.8 GB

Tools

search tools

xcms.xcmsSet Merger Merge xcms.xcmsSet xset in one to be used by group

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.

xcms.retcor Retention Time Correction using retcor function from xcms R package

xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas

xcms.summary Create a summary of XCMS analysis

CAMERA.annotate CAMERA

CAMERA.combineXcnos Wrapper function for the

1 job has been successfully added to the queue - resulting in the following datasets:

- 17: LC_Positive_mzXML.xset.group.retcor.group.RData
- 18: LC_Positive_mzXML.xset.group.retcor.group.Rplots.pdf
- 19: LC_Positive_mzXML.xset.group.retcor.group.variableMetadata.tsv
- 20: LC_Positive_mzXML.xset.group.retcor.group.dataMatrix.tsv
- 21: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Unnamed history
18 shown, 2 hidden
3.73 GB

- 20: LC_Positive_mzXML.xset.group.retcor.group.dataMatrix.tsv
- 19: LC_Positive_mzXML.xset.group.retcor.group.variableMetadata.tsv
- 18: LC_Positive_mzXML.xset.group.retcor.group.Rplots.pdf
- 17: LC_Positive_mzXML.xset.group.retcor.group.RData
- 15: LC_Positive_mzXML.xset.group.retcor.BPCs_corrected.pdf
- 14: LC_Positive_mzXML.xset.group.retcor.TICs_corrected.pdf
- 13: LC_Positive_mzXML.xset.group.retcor.Rplots.pdf
- 12: LC_Positive_mzXML.xset.group.retcor.RData

A group step

2. xcms.fillpeaks

xcms.fillPeaks Integruje sygnał próbek w regionach, w których grupy pików nie są reprezentowane, aby utworzyć nowe piki w brakujących obszarach

Integruje powierzchnie brakujących pików. Dla każdej próbki należy zidentyfikować grupy pików, w których próbka nie jest reprezentowana. Dla każdej z tych grup zintegruj sygnał w obszarze tej grupy pików i utwórz nowy pik.

W zależności od rodzaju surowych danych dostępne są 2 różne metody do uzupełnienia danych GCMS/LCMS: metoda "chrom" integruje surowe dane w domenie chromatograficznej, podczas gdy "MSW" jest używana dla listy pików bez informacji o RT, takich jak z widm bezpośrednich infuzji.

Pliki wejściowe

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas (Galaxy Version 2.1.0) Versions Options

xset RData file

No rdata.xcms.group or rdata dataset available. ▼
output file from another xcms function (group)

Filling method

chrom ▼
[method] See the help section below

Get a Peak List

[Resubmit your raw dataset or your zip file](#) ↻

Parametry

Metoda filling

➤ chrom

Metoda ta wytwarza wartości intensywności dla brakujących próbek poprzez integrację surowych danych w regionie grupy piku. W danej grupie początkowe i końcowe punkty RT dla integracji są określane przez medianę punktów początkowych i końcowych innych wykrytych pików. Początkowe i końcowe wartości m/z są podobnie określane. Intensywności mogą nadal wynosić zero, co jest dość nietypową intensywnością dla piku np. dane pierwotne zostały poddane progowaniu, a obszar integracji nie zawiera rzeczywistych intensywności,

lub jeśli jedna próbka jest błędnie skalibrowana, takie surowe punkty danych znajdują się (tuż) poza obszarem integracji.

Co ważne, jeśli korekcja danych RT jest dostępna, informacje o wyrównaniu są wykorzystywane do dokładniejszej integracji właściwego obszaru surowych danych. Jeśli poprawiony RT jest poza końcem danych pierwotnych, wartość będzie nie-numerowana (NaN).

➤ MSW

"MSW" jest używany do list pików bez informacji o RT, takich jak te z widm bezpośrednich infuzji.

Zdobądź listę pików

Jeśli "true", moduł generuje dwa dodatkowe pliki odpowiadające liście pików: - zmienny plik metadanych (odpowiadający informacjom o wyekstrahowanych jonach, takich jak masa lub RT) - macierz danych (odpowiadająca powiązanym intensywnościom).

Miejsca dziesiętne dla wartości [masy lub czasu retencji] w zidentyfikowanych jonach.

Zidentyfikowane jony są skonstruowane jako MxxxTyyy, gdzie "xxx" to mediana masy jonów, a "yyy" mediana RT jonów. Dwa parametry służą do dostosowania liczby miejsc dziesiętnych poszukiwanych w zidentyfikowanych jonach odpowiednio dla masy i RT. Parametry te nie wpływają na miejsca dziesiętne w kolumnach innych niż zidentyfikowany jon.

Zgłoszone wartości intensywności

Parametry te określają, wartości, które należy zgłosić jako intensywności w tabeli dataMatrix; odpowiada to parametrowi "intval" xcms:

-**into**: zintegrowane powierzchnie surowego pików

-**maxo**: maksymalna intensywność surowego pików

-**intb**: korekcja linii podstawowej zintegrowanego obszaru pików (dostępna tylko w przypadku, gdy wykrywanie pików zostało wykonane przez 'findPeaks.centWave')

The screenshot shows the Galaxy Metabolomics interface. A green notification box in the center states: "1 job has been successfully added to the queue - resulting in the following datasets: 22: LC_Positive_mzXML.xset.group.retcor.group.fillPeaks.RData, 23: LC_Positive_mzXML.xset.group.retcor.group.fillpeaks.variableMetadata.tsv, 24: LC_Positive_mzXML.xset.group.retcor.group.fillpeaks.dataMatrix.tsv, 25: xset.log.txt". Below this, instructions are provided: "You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." The History pane on the right shows a list of datasets, including the ones mentioned in the notification, such as "24: LC_Positive_mzXML.xset.group.retcor.group.fillpeaks.dataMatrix.tsv", "23: LC_Positive_mzXML.xset.group.retcor.group.fillpeaks.variableMetadata.tsv", "22: LC_Positive_mzXML.xset.group.retcor.group.fillPeaks.RData", "20: LC_Positive_mzXML.xset.group.retcor.group.dataMatrix.tsv", "19: LC_Positive_mzXML.xset.group.retcor.group.variableMetadata.tsv", "18: LC_Positive_mzXML.xset.group.retcor.group.Rplots.pdf", "17: LC_Positive_mzXML.xset.group.retcor.group.RData", and "15: LC_Positive_mzXML.xset.group.retcor.group.RData".

3. xcms. Podsumowanie

[xcms.summary](#) Create a summary of XCMS analysis

Narzędzie te udostępnia plik HTML podsumowując analizę przy użyciu narzędzi [W4M] XCMS i CAMERA.

xcms.summary Create a summary of XCMS analysis (Galaxy Version 1.0.3) Versions Options

xset RData file

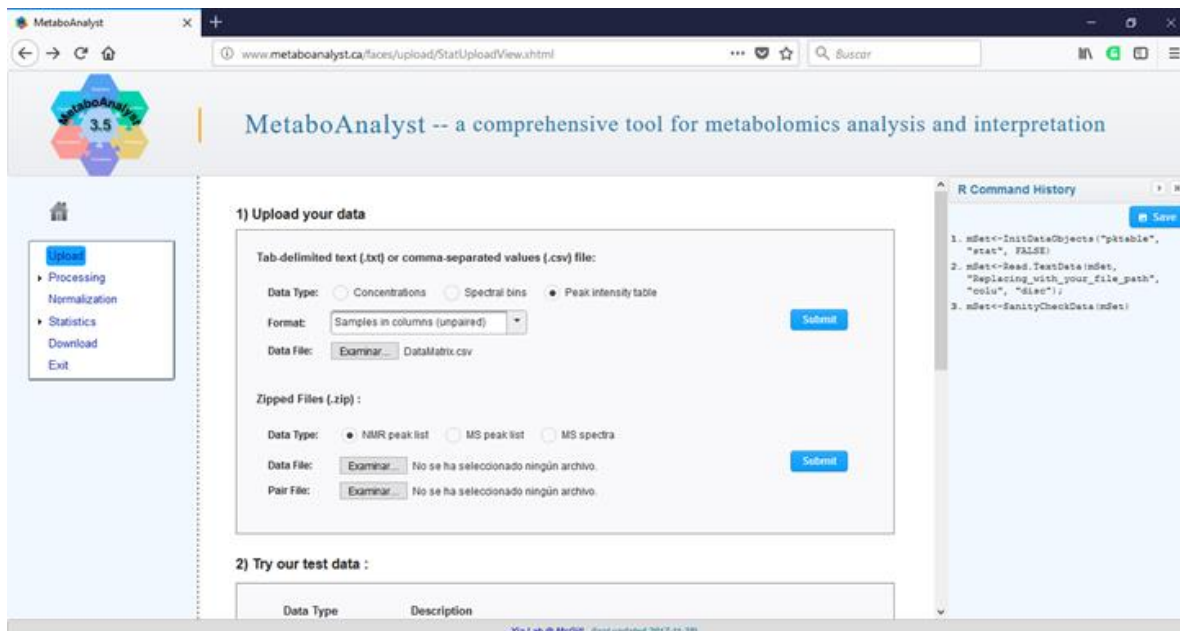
No rdata.xcms.raw, rdata.xcms.group, rdata.xcms.retcor, rdata.xcms.fillpeaks, rdata...
output file from another function xcms (xcmsSet, group, retcor, fillpeaks etc.)

Execute

5.3. Statystyka

Analiza statystyczna jednoczynnikowa i wieloczynnikowa: PCA, PLS-DA zostanie wykonana przy użyciu Metaboanalyst (<https://metaboanalyst.ca>). Celem MetaboAnalyst jest zapewnienie bezpłatnego, przyjaznego dla użytkownika i łatwo dostępnego narzędzia do analizy danych metabolomicznych. Został zaprojektowany w celu rozwiązania dwóch problemów: 1) w celu zidentyfikowania cech znacząco różniących się między dwoma warunkami (wykrywanie biomarkerów); 2) wykorzystanie danych metabolomicznych do przewidywania badanych warunków (klasyfikacja). Ponadto, MetaboAnalyst zapewnia również narzędzia do identyfikacji związków i mapowania ścieżek biochemicznych do opisywania istotnych cech.

Przejdź do strony głównej i kliknij >>click here to start<<



Wybierz odpowiedni typ danych i format i prześlij dane. Ponieważ nie ma brakujących wartości w dostarczonych danych, następny krok można pominąć.

Na każdym etapie dostępnych jest wiele opcji do analizy statystycznej. Uczniowie są zachęceni do wypróbowania alternatywnych dróg analiz, które zostaną zaproponowane w ramach tego kursu.

Ponieważ materiał badawczy to osocze, nie jest wymagana zewnętrzna normalizacja.

Praca będzie odbywać się po transformacji log i skalowaniu Pareto.

Studenci będą starali się znaleźć istotne związki po analizie jednowymiarowej ANOVA oraz teście post hoc, a także po analizie PLS-DA.

Opcje klastrowania zostaną ocenione.

Wyniki zostaną połączone w arkuszu kalkulacyjnym MS Excel i uzyskana zostanie lista znaczących wartości m/z.

5.4. Komentarze

- Identyfikacja za pomocą CEU Mass Mediator, <http://ceumass.eps.uspceu.es/>
- Analiza szlaków biochemicznych: integracja analizy wzbogacania i analiza topologii szlaków biochemicznych.
- Wizualizacja uzyskanych wyników poprzez modelowanie za pomocą Metaboanalyst.

Literatura:

E.G. Armitage, F. J. Rupérez, C. Barbas. Metabolomics of diet-related diseases using mass spectrometry. *Trends Anal. Chem.*, 2013, 52:61-73.

A. Mastrangelo, A. Ferrarini, F. Rey-Stolle, A. García, C. Barbas. From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. *Anal. Chim. Acta*, 2015, 900:21-35.

A. García, S. Naz, C. Barbas. Metabolite fingerprinting by capillary electrophoresis-mass spectrometry. *Methods Mol Biol*, 2014, 1198:107-223.

S. Naz, A. García, M. Rusak, C. Barbas. Method development and validation for rat serum fingerprinting with CE-MS: application to ventilator-induced-lung-injury study. *Analytical Bioanal Chemistry*, 2013, 405(14):4849-4858.

W4M Workflow4Metabolomics. <http://workflow4metabolomics.org/>

Moduł 3

Lipidomika

Elisabete Maciel¹, Eliana Alves², Pedro Domingues², Rosário Domingues^{1,2}

Mass Spectrometry Center, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal

1 CESAM (Centro de Estudos do Ambiente e do Mar)

2 QOPNA (Química Orgânica de Produtos Naturais e Agroalimentares)

I. Wprowadzenie

Lipidomikę można zdefiniować jako wielkoskalową analizę mającą na celu ocenę składu lipidów w próbkach biologicznych. Na przestrzeni ostatnich dziesięciu lat zainteresowanie lipidomiką znacząco wzrosło co związane było z rozwojem technik analitycznych, zwłaszcza spektrometrii mas, jak również poznaniem kluczowej roli lipidów w utrzymaniu równowagi w komórce, a także w rozwoju różnych stanów patologicznych. Analizy lipidomiczne mogą stanowić zatem nowatorskie podejście pozwalające na zidentyfikowanie lipidowych biomarkerów choroby, jak również określać rodzaj modyfikacji szlaków metabolicznych umożliwiając w ten sposób opracowanie innowacyjnych strategii terapeutycznych.

Pierwsza część modułu będzie skupiała się wokół zagadnień związanych z głównymi analitycznymi strategiami wykorzystywanymi w celu uzyskania profilu lipidowego próbek biologicznych. Przedstawione zostaną najważniejsze technologie stosowane w tym celu z uwzględnieniem protokołów ekstrakcji lipidów, metod ich rozdzielania na frakcje, sposobów wykorzystania spektrometrii mas do ich analizy, jak również platform bioinformatycznych stosowanych do obróbki otrzymanych danych.

Druga część modułu stanowi studium przypadku i ma na celu zaznajomienie studentów z analizami lipidomicznymi. Szkolenie będzie oparte na prezentacjach/interaktywnych wykładach dla studentów, pokazach/warsztatach praktycznych oraz analizie konkretnego przykładu analizy, co pozwoli na całkowite zrealizowanie założonych celów i osiągnięcie specyficznych efektów kształcenia.

II. Założenia i efekty kształcenia

Założenia

Celem tego kursu jest przedstawienie studentom zaawansowanego i zintegrowanego obrazu strategii analitycznych, które można wykorzystać do badania lipidomu. Studenci zdobędą teoretyczną i praktyczną wiedzę z zakresu lipidomiki w kontekście nauk bioanalitycznych. Uwzględnione zostaną zarówno podstawowe jak i zaawansowane zagadnienia dotyczące struktury lipidów i ich funkcji. Szczególna uwaga zostanie poświęcona aspektom związanym z metodologią lipidomiczną, w tym różnym rodzajom ekstrakcji, frakcjonowania lipidów oraz analizie ich struktury, przede wszystkim w oparciu o metody wykorzystujące spektrometrię mas. Kurs ma również na celu pokazanie potencjalnych możliwości wykorzystania zmian w lipidomie, będących skutkiem różnych stanów

patologicznych, takich jak zaburzenia w układzie sercowo-naczyniowym, choroby neurodegeneracyjne, nowotwór, cukrzyca czy też stan zapalny, w celu zidentyfikowania biomarkerów tych schorzeń oraz opracowania nowych strategii ich leczenia.

Specyficzne efekty kształcenia:

Po ukończeniu kursu studenci powinni potrafić:

1. definiować i posługiwać się powszechnie stosowaną terminologią lipidomiczną;
2. opisać zasady technik najczęściej stosowanych w lipidomice;
3. rozpoznać różnice wynikające z zastosowania różnych typów analiz lipidomicznych z użyciem różnych spektrometrów mas;
4. analizować dane lipidomiczne;
5. wyrażać i uzasadniać wnioski, w sposób zrozumiały i jednoznaczny, zarówno odbiorcom specjalistycznym, jak i osobom niebędącym specjalistami;
6. samodzielnie kontynuować naukę.

III. Spis treści

1. Wstęp do lipidomiki
 - 1.1. Podejścia analityczne stosowane w lipidomice
 - 1.1.1. Przygotowanie próbek: ekstrakcja i rozdział lipidów
 - 1.1.2. Identyfikacja lipidów oparta na analizach MS: koncepcje
 - 1.1.2.1. Nietargetowe podejścia lipidomiczne oparte na LC-MS
 - 1.1.2.2. Targetowe podejścia lipidomiczne “Shotgun”
 - 1.2. Ilościowa analiza lipidomiczna
 - 1.3. Przetwarzanie danych w celu identyfikacji/ilościowej oceny lipidów
 - 1.4. Część praktyczna
 - 1.4.1. Analiza lipidomiczna profilu fosfolipidów monocytów linii THP1 z wykorzystaniem HILIC-LC-MS
 - 1.5. Dodatkowe odczyty zasobów

1. Wstęp do lipidomiki

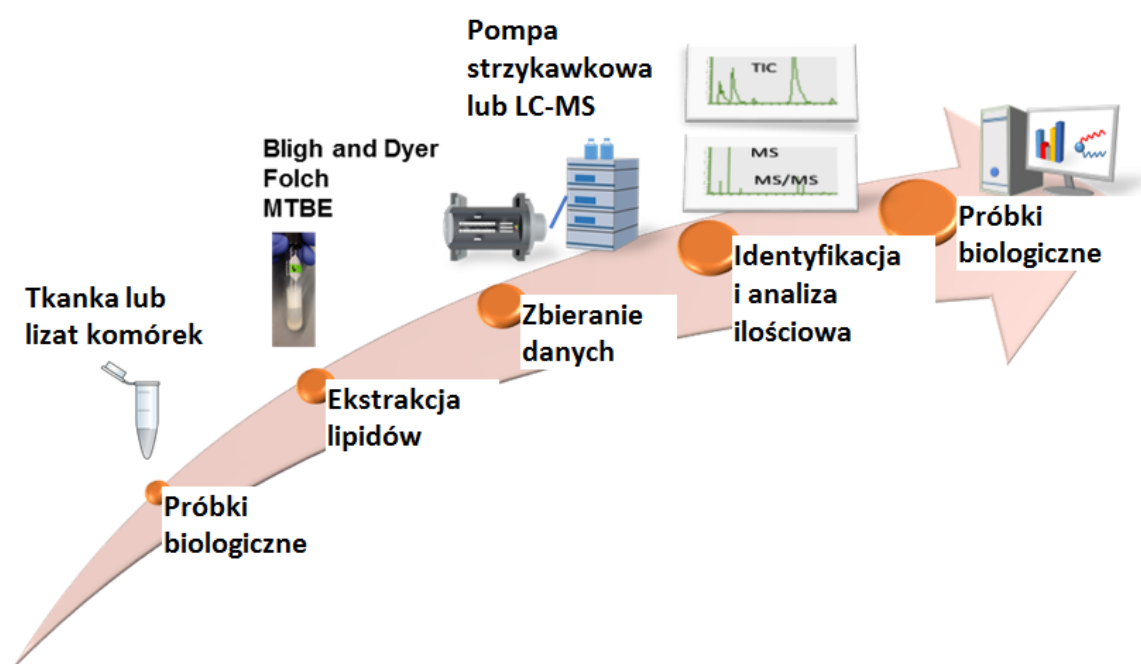
Lipidomika rozumiana jest jako prowadzona na szeroką skalę analiza lipidów zawartych w próbkach biologicznych. Ma ona na celu dokładną charakterystykę lipidów, jako grupy związków, ich biologicznej roli w powiązaniu z ekspresją białek zaangażowanych w regulację genów, oraz funkcje i metabolizm lipidów (<http://lipidlibrary.aocs.org/Analysis/content.cfm?ItemNumber=39284>, April 2018).

Lipidy stanowią ogromną grupę naturalnych związków występujących w roślinach, organizmach zwierząt oraz mikroorganizmach, które z uwagi na różnice strukturalne podzielono na klasy. Różnorodność związków lipidowych, które ta grupa obejmuje znacznie wzrasta po uwzględnieniu strukturalnych modyfikacji lipidów będących wynikiem ich utleniania bądź glikacji. Scharakteryzowanie całego lipidomu stanowi ogromne wyzwanie, nie tylko ze względu na strukturalną różnorodność lipidów, lecz również z uwagi na ogromną liczbę związków, występujących na bardzo niskich stężeniach, co z kolei determinuje potrzebę zastosowanie odpowiednio wysokoczułych technik analitycznych.

Lipidom obejmuje obszerną grupę związków różniących się zarówno pod względem struktury jak również pełnionych funkcji. Lipidy pełnią rolę elementów strukturalnych błony komórkowej, gdzie pełnią ważną funkcję modulatorów aktywności białek, jak również rozpoznawalnych i sygnałowych cząsteczek. Dzięki temu lipidy biorą udział w regulowaniu ważnych procesów takich jak proliferacja komórek, apoptoza, metabolizm czy migracja.

Biosynteza i metabolizm lipidów są regulowane przez grupę enzymów umożliwiających zachowanie strukturalnej różnorodności lipidów należących do różnych klas. Biosynteza lipidów zachodzi przede wszystkim w retikulum endoplazmatycznym oraz mitochondriach, skąd mogą być one przenoszone do innych błon oraz organelli w celu zapewnienia różnorodności lipidów i ich obecności w specyficznych miejscach komórki. W związku z tym zaburzenia w profilu lipidów komórki są często związane z rozwojem wielu chorób. Tak więc lipidomika, jako nowe narzędzie posiadające ogromny potencjał w określaniu zmian w lipidomie, pozwala zrozumieć patologiczne procesy w przebiegu choroby oraz umożliwia identyfikację nowych biomarkerów diagnostycznych i pozwala na opracowanie nowych strategii leczenia.

Określenie profilu lipidowego próbki biologicznej wymaga zastosowania szeregu procedur przygotowania próbek oraz technik analitycznych, które muszą być wcześniej ustalone zależnie od informacji jakie należy uzyskać jako wynik końcowy. Z reguły procedurę przygotowania próbek do analizy lipidomicznej rozpoczyna ekstrakcja całkowitej zawartości lipidów, które w kolejnym etapie rozdzielane są na frakcje, następnie analizowane przy wykorzystaniu spektrometrii mas, a uzyskane dane są odpowiednio przetwarzane (Rycina 3.1).



Rycina 3.1. Podstawowe etapy procedury przygotowania próbek do analizy lipidomicznej.

Pierwszy etap przygotowania próbek do analizy lipidomicznej polega na wyekstrahowaniu wszystkich lipidów zawartych w materiale. Ekstrakcję można przeprowadzić z różnych rodzajów materiału biologicznego takich jak lizat komórek, tkanka czy płyny ustrojowe. Ekstrakcja powinna zapewniać odzysk większości lipidów lub grup związków lipidowych, których dotyczy analiza. Z uwagi na to, że ekstrakt lipidowy zawiera

ogromną liczbę różnych związków, zalecane jest w niektórych specyficznych przypadkach zastosowanie frakcjonowania umożliwiające wydzielenie poszczególnych frakcji związków. Frakcjonowanie przeprowadza się używając różnych metod separacyjnych takich jak cienkowarstwowa chromatografia cieczowa, wysokosprawna chromatografia cieczowa czy też przy zastosowaniu procedury wykorzystującej ekstrakcję do fazy stałej. Identyfikacja poszczególnych związków zawartych w danej klasie lipidów lub całym ekstrakcie może być przeprowadzona przy użyciu takich technik jak spektrometria mas. Z kolei kwasy tłuszczowe zestryfikowane w obrębie struktury danego lipidu mogą być zidentyfikowane przy użyciu połączenia chromatografii gazowej ze spektrometrią mas, po przeprowadzeniu hydrolizy tych kwasów oraz procedury otrzymywania ich lotnych pochodnych. Dzięki rozwojowi spektrometrii mas uzyskano znaczny postęp jeśli chodzi o identyfikację oraz charakterystykę związków lipidowych, zwłaszcza pod względem czułości i szybkości analiz. Począwszy od lat dziewięćdziesiątych dzięki wprowadzeniu nowego typu jonizacji związków typu elektrosprej możliwe było połączenie techniki spektrometrii mas z wysokociśnieniową chromatografią cieczową (HPLC-MS), co stanowiło ogromny postęp w dziedzinie analizowania próbek o złożonej matrycy.

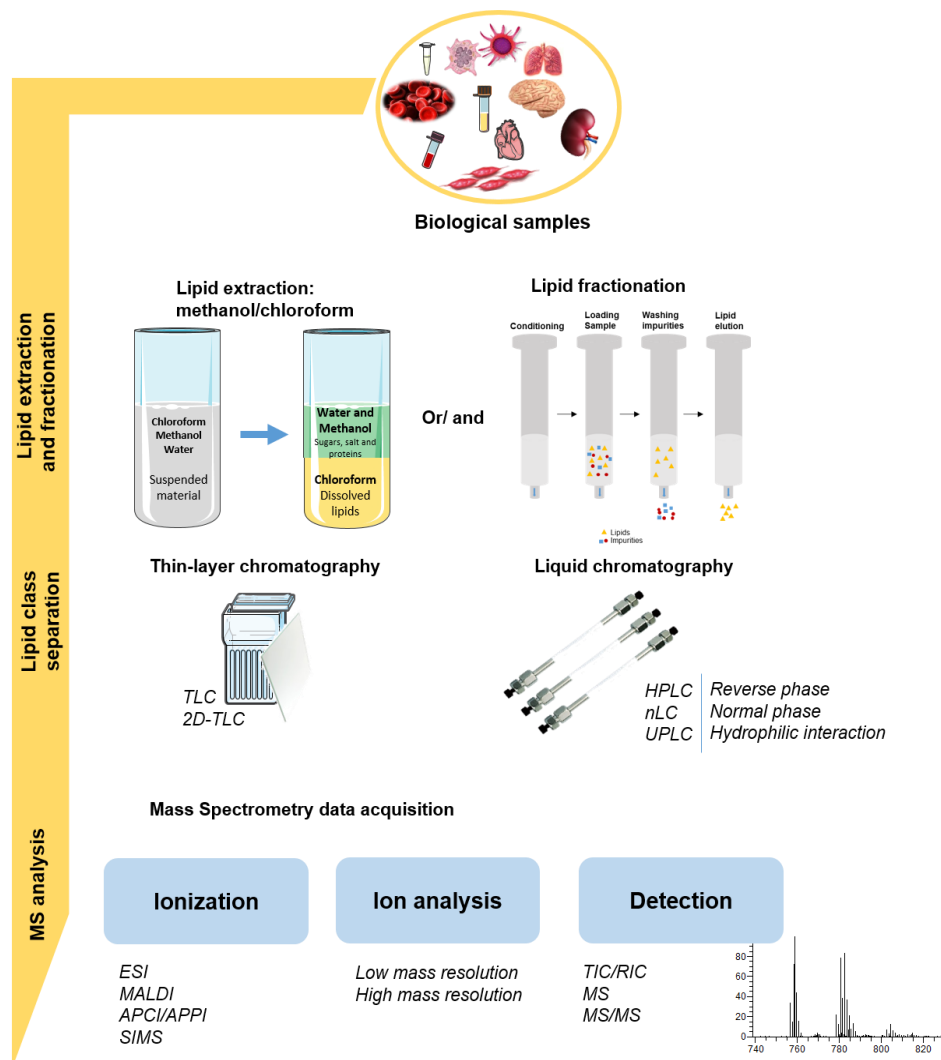
W ostatnich dziesięcioleciach dzięki komercjalizacji wysokorozdzielczych spektrometrów mas typu Q-TOF czy Orbitrap, pojawiła się możliwość zbierania danych przy wykorzystaniu wysokiej czułości. Zastosowanie techniki LC-MS umożliwia w krótkim czasie identyfikację i ilościową ocenę tysiąca związków lipidowych, co jest bardzo korzystne zwłaszcza w przypadku analizy lipidów zawartych w próbkach biologicznych. Uzyskanie informacji ze zgromadzonej ogromnej ilości danych uzyskanych z wielkoskalowych analiz lipidomicznych wymaga zastosowania odpowiedniej procedury obróbki i analizy tych danych.

W zależności od głównego celu prowadzonych badań może być zastosowana odmienna strategia badań oparta o targetowe bądź nietargetowe analizy. Lipidomika nietargetowa ma na celu zidentyfikowanie i ilościową ocenę możliwie jak największej ilości związków lipidowych uzyskując w ten sposób specyficzną sygnaturę składu lipidowego określonej matrycy w postaci tkanki, lizatu komórkowego, organelli bądź płynu ustrojowego. Strategia ta może być również stosowana w celu poszukiwania biomarkerów chorób poprzez porównanie danych lipidomicznych uzyskanych dla osób zdrowych jak i cierpiących na daną jednostkę chorobową. Lipidomika nietargetowa wykorzystuje platformę LC-MS do analizy ekstraktów zawierających wszystkie lipidy w jednym przebiegu generując w ten sposób ogromną ilość danych wymagających przeprowadzenia analizy bioinformatycznej jak również dalszej walidacji. Informacje pozyskane z nietargetowych analiz mogą być wykorzystane w celu zaplanowania i przeprowadzenia odpowiednich celowanych analiz lipidomicznych. Lipidomika targetowa przeprowadzana jest w celu identyfikacji panelu specyficznych związków lipidowych. Strategia ta może być wykorzystana do określania specyficznych zmian w obrębie danej klasy lipidów. Realizując to założenie stosowane są analizy typu "shotgun" w trybie "precursor ion scan" (PIS) lub neutral loss scan (NLS). Analiza ilościowa specyficznych związków lipidowych przeprowadzana jest w trybie MRM przy zastosowaniu spektrometru mas typu potrójnego kwadrupola.

1.1. Podejścia analityczne stosowane w lipidomice

1.1.1. Przygotowanie próbek: ekstrakcja i rozdział lipidów

Procedurę przygotowania próbek do analizy lipidomicznej rozpoczyna ekstrakcja całkowitej zawartości lipidów przy użyciu rozpuszczalników organicznych z matrycy biologicznej takiej jak płyny ustrojowe, komórki, tkanki lub inne (Rycina 3.2).



Rycina 3.2. Przygotowanie próbek do analizy lipidomicznej: ekstrakcja lipidów i ich frakcjonowanie

Powszechnie stosuje się kilka metod ekstrakcji opartych o rozpuszczalniki organiczne różniące się polarnością w zależności od rodzaju materiału biologicznego stanowiącego próbkę poddawaną ekstrakcji. Jeśli celem ekstrakcji jest uzyskanie ekstraktów zawierających konkretną klasę związków lipidowych, proporcje zawartości rozpuszczalników organicznych w użytych mieszaninach ekstrakcyjnych mogą się różnić zmieniając w ten sposób ich polarność.

Najbardziej rozpowszechnionymi metodami ekstrakcji lipidów są: metoda Folcha (wykorzystująca mieszaninę chloroform : metanol, 2 : 1, w stosunku objętościowym) oraz

metoda Bligh-Dyera (wykorzystująca mieszaninę chloroform : metanol, 1 : 2, w stosunku objętościowym). Pierwsza z nich jest częściej stosowana w celu ekstrakcji lipidów z płynów ustrojowych (osocze, surowica krwi oraz mocz), podczas gdy druga jest wykorzystywana przede wszystkim do ekstrakcji lipidów z tkanek oraz liszt komórkowego. Wprowadzono także kilka innych metod ekstrakcji w celu zaprzestania stosowania rozpuszczalników chlorowanych takich jak chloroform. Tak jest w przypadku ostatnio zaproponowanej metody ekstrakcji opartej o mieszaninę eter metylo-tert-butyłowy (MTBE)/metanol/woda (w stosunku objętościowym 10:3:2.5). Dodatkową zaletą takiej ekstrakcji jest to, że faza organiczna, odpowiedzialna za odzysk lipidów z próbki, stanowi górną fazę mieszaniny co znacznie ułatwia jej zebranie i oddzielenie od fazy wodnej. Dzięki tej właściwości ekstrakty są czystsze, uzyskuje się je szybciej, a co istotne stanowią niemal gotową próbkę do profilowania lipidów metodą shotgun, która wykorzystuje mieszaninę eter metylo-tert-butyłowy (MTBE)/metanol (w stosunku objętościowym 1:1). Natomiast w celu otrzymania ekstraktów bogatych w neutralne lipidy takie jak triglicerydy (TG), jest wskazane stosowanie mniej polarnych rozpuszczalników takich jak heksan.

Ekstrakcja do fazy stałej SPE (solid phase extraction) jest kolejną techniką wykorzystywaną do izolacji frakcji fosfolipidów (PL) z osocza. Procedura tego typu ekstrakcji polega na strąceniu białek poprzez dodanie do osocza acetonitrylu z dodatkiem 1% kwasu mrówkowego, odwirowaniu, zebraniu górnej warstwy, przeniesieniu jej na jonowe wypełnienie kolumny hybrid SPE-PL cartridge (Thermo Scientific) i wymyciu PL mieszaniną acetonitryl/amoniak (roztwór 5%). Całkowity ekstrakt zawierający wszystkie związki lipidowe może być bezpośrednio analizowany przy użyciu nowoczesnych metod lipidomicznych: targetowych jak i nietargetowych wykorzystujących spektrometrię mas lub może być poddany frakcjonowaniu, a uzyskane frakcje zawierające specyficzne klasy związków są następnie analizowane przy wykorzystaniu odpowiednich metod lipidomicznych. Do frakcjonowania całkowitego ekstraktu lipidów wykorzystywana jest SPE lub cienkowarstwowa chromatografia cieczowa (TLC). W przypadku SPE w celu wyizolowania cholesterolu, TG i PL wykorzystywane są kolumny z wypełnieniem aminopropylowym, natomiast zastosowanie kolumn zawierających krzemionkę zalecane jest do rozdzielania lipidów neutralnych (wolnych kwasów tłuszczowych, cholesterolu i TG) oraz frakcji PL.

Rozdział lipidów na klasy można również uzyskać stosując TLC. W badaniach lipidomicznych stosowana jest zarówno jednowymiarowa jak i dwuwymiarowa TLC. Wybór układu rozpuszczalników stosowanych w TLC zależy do tego które klasy lipidów będą poddawane rozdziałowi. Dotyczy to zwłaszcza rozdziału fosfolipidów na klasy. Zazwyczaj klasy PL rozdzielane są przy użyciu jednokierunkowej TLC i zastosowaniu fazy ruchomej w postaci mieszaniny chloroform/etanol/woda/trietylamina (w stosunku objętościowym 35:30:7:35). Po rozwinięciu, wizualizacja płytki w świetle UV, dokonywana jest po jej wcześniejszym spryskaniu acetonowym roztworem primuliny. Identyfikacja poszczególnych klas PL dokonywana jest na podstawie porównania ich współczynników retencji (retention factors R_f) ze współczynnikami dla standardów nanoszonych na tę samą płytkę TLC przed jej rozwinięciem. W każdej zdrapanej z płytki frakcji PL oznaczana jest zawartość fosforu, na podstawie której dokonywana jest ilościowa ocena danej klasy PL względem całkowitej zawartości wszystkich PL zawartych w analizowanym ekstrakcie. Ponadto fosfolipidy zawarte w zdrapanej z płytki TLC krzemionce mogą być wyekstrahowane ponownie przy użyciu odpowiedniego rozpuszczalnika organicznego i poddane dalszej analizie. Do

rozdzielania lipidów używana jest także dwuwymiarowa TLC, a dobór rozpuszczalników stosowanych do separacji, podobnie jak w jednowymiarowej TLC, jest determinowany charakterem rozdzielanych klas związków. Pomimo tego, że rozdział lipidów przy użyciu 2-D TLC ulega znacznej poprawie, technika ta ma wadę w porównaniu z jednowymiarową TLC, z uwagi na brak możliwości stosowania standardów ponieważ tylko jedna próbka jest наносzona na pojedynczą płytkę. Ponadto rozdział z użyciem 2-D TLC jest bardziej czasochłonny.

Otrzymane metodą SPE lub TLC ekstrakty zawierające poszczególne frakcje związków lipidowych są następnie analizowane przy zastosowaniu targetowych bądź nientargetowych podejść lipidomicznych. W praktyce najczęściej stosowane są dwa podejścia wykorzystujące do analizy ekstraktów/frakcji lipidowych spektrometrię mas: bezpośrednie wstrzyknięcie (direct infusion) ekstraktu przy użyciu metody shotgun oraz podejście wykorzystujące platformę LC-MS. Oba podejścia analityczne opracowano dla spektrometrów mas wyposażonych w źródło jonizacji typu ESI. Pomimo różnych wariantów w chwili obecnej najbardziej rozpowszechnionym podejściem lipidomicznym jest analiza całych ekstraktów lipidowych przy użyciu LC-MS. Oba wspomniane podejścia wymagają umiejętności obsługi aparatury MS, oraz analizy i interpretacji uzyskiwanych danych w celu identyfikacji poszczególnych klas związków lipidowych.

1.1.2. Identyfikacja lipidów oparta na analizach MS: koncepcje

Analiza lipidów z użyciem spektrometrii mas wymaga informacji uzyskanych na podstawie interpretacji widm MS i MS/MS. Całkowite ekstrakty lipidowe wyizolowane z komórek bądź tkanek zawierają przede wszystkim fosfolipidy. Dlatego biorąc pod uwagę ten fakt widma masowe mogą być rejestrowane zarówno w negatywnym jak i pozytywnym trybie jonizacji, w zależności od specyficznej budowy strukturalnej fosfolipidów należących do poszczególnych klas. W przypadku fosfolipidów należących do: fosfatydylocholin (PC), fosfatydyloetanolamin (PE) i sfingomielin (SM) identyfikacja opiera się o widma masowe rejestrowane w trybie pozytywnej jonizacji, na których główny sygnał reprezentuje protonowany jon molekularny tych fosfolipidów $[M+H]^+$, bądź jon adduktu sodowego $[M+Na]^+$, którego intensywność jest znacznie mniejsza. Fosfatydyloinozytole (PI), fosfatydyloseryny (PS), fosfatydyloglicerole (PG) oraz kardiolipiny (CL) są analizowane preferencyjnie w trybie negatywnej jonizacji dając na widmie masowym deprotonowane jony $[M-H]^-$. PE mogą być również analizowane w trybie jonizacji negatywnej tworząc deprotonowany jon $[M-H]^-$. W przypadku CL ma również miejsce tworzenie jonów dwuujemnych ($[M-2H]^{2-}$). W przypadku analiz LC-MS w trybie negatywnym, przy zastosowaniu eluentów zawierających dodatek soli octanu amonu bądź mrówczanu amonu, PC i SM tworzą odpowiednie addukty obserwowane na widmie masowym jako jony, $[M+CH_3COO]^-$ i $[M+COO]^-$ (Tabela 1).

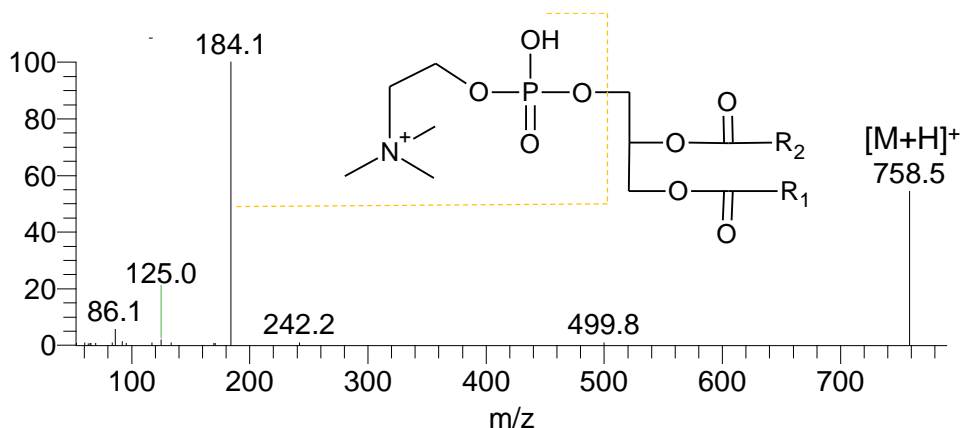
Tabela 1. Rodzaje jonów oraz adduktów uzyskiwanych w wyniku jonizacji polarnych lipidów w źródle ESI.

Klasa PL	Tworzone jony	
	Pozytywny tryb jonizacji	Negatywny tryb jonizacji
PC	$[M+H]^+$, $[M+Na]^+$,	$[M+Ac-H]^-$
PE	$[M+H]^+$, $[M+Na]^+$	$[M-H]^-$
PG	$[M+NH_4]^+$, $[M+Na]^+$	$[M-H]^-$
PI	$[M+NH_4]^+$	$[M-H]^-$
PS	$[M+H]^+$	$[M-H]^-$
CL		$[M-H]^-$, $[M-2H]^{2-}$

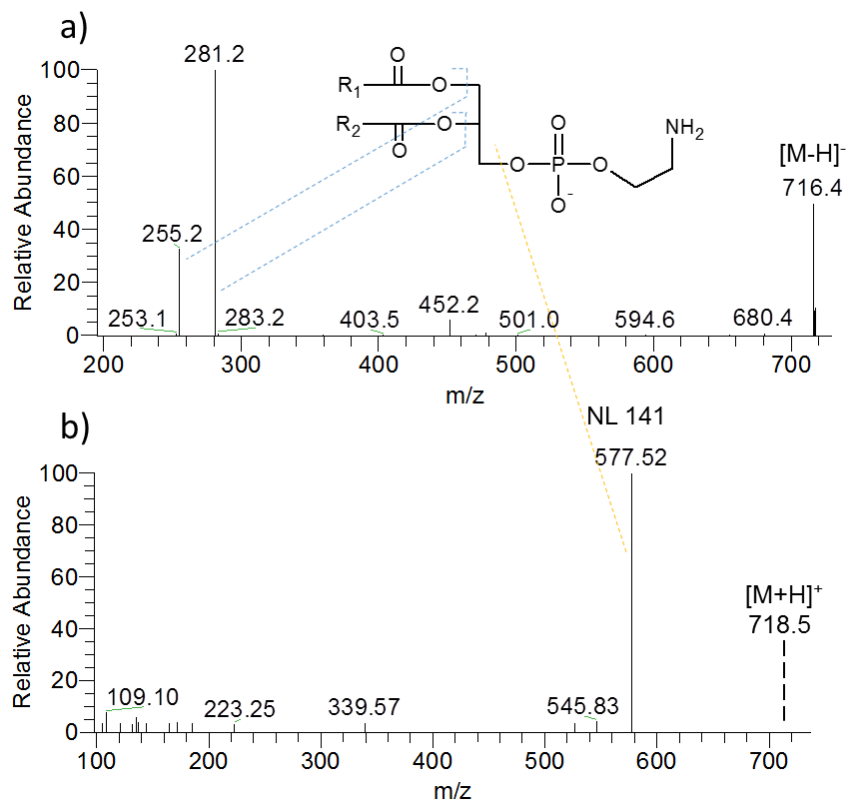
Wyłuszczone główne jony na widmie ESI-MS. PC (fosfatydylocholina), PE (fosfatydyloetanoloamin), PG (fosfatydyloglicerole), PI (fosfatydyloinozytyle), PS (fosfatydyloseryny), CL (kardiolipiny). PIS- Precursor ion scan, NLS- Neutral loss scan.

PI mogą także tworzyć jony $[M+NH_4]^+$. Prawidłowa identyfikacja odpowiedniego typu jonów obserwowanych na widmie masowym pozwala na obliczenie masy molekularnej fosfolipidów, którym one odpowiadają. W chwili obecnej dzięki zastosowaniu spektrometru mas typu orbitrap, pozwalającego na uzyskiwanie widm masowych o dużej rozdzielczości od 70 000 do 500 000 możliwe jest uzyskanie informacji na temat dokładnej wartości masy związku z dokładnością poniżej 2 ppm.

Do ostatecznego potwierdzenia tożsamości poszczególnych związków niezbędne są widma uzyskiwane przy użyciu tandemowej spektrometrii mas (widma MS/MS). Na widmach MS/MS obecne są charakterystyczne jony odpowiadające „polarnym głowom” poszczególnych fosfolipidów, co umożliwi ich przypisanie do odpowiedniej klasy. Natomiast utrata cząsteczki kwasu tłuszczowego w trybie neutral loss i jonizacji pozytywnej w postaci ketenu lub wolnego kwasu oraz utworzenie anionu reszty kwasowej $RCOO^-$ w trybie jonizacji negatywnej pozwala zidentyfikować zestryfikowane kwasy tłuszczowe w obrębie struktury poszczególnych fosfolipidów (Rycina 3.3 i 3.4).



Rycina 3.3. Widmo MS/MS, protonowanego jonu $[M+H]^+$ fosfolipidu należącego do PC, jon fragmentaryczny o m/z 184 jest charakterystyczny dla PC.



Rycina 3.4. Widmo MS/MS fosfolipidu należącego do PE w trybie jonizacji negatywnej (a) zawierające sygnały pochodzące od jonów RCOO^- pozwalające zidentyfikować kwasy tłuszczowe obecne w cząsteczce PE, i pozytywnej (b) ilustrujące mechanizm utraty neutralnej cząsteczki o masie 141 Da, charakterystycznej dla klasy PE, co w przedstawionym przypadku skutkuje wytworzeniem intensywnego jonu fragmentarycznego o m/z 577.5.

Główne typy jonów, charakterystycznych dla poszczególnych klas PL, charakteryzujące się utratą neutralnej cząsteczki NLS oraz pozwalające zidentyfikować kwasy tłuszczowe zebrane zostały w tabeli 2.

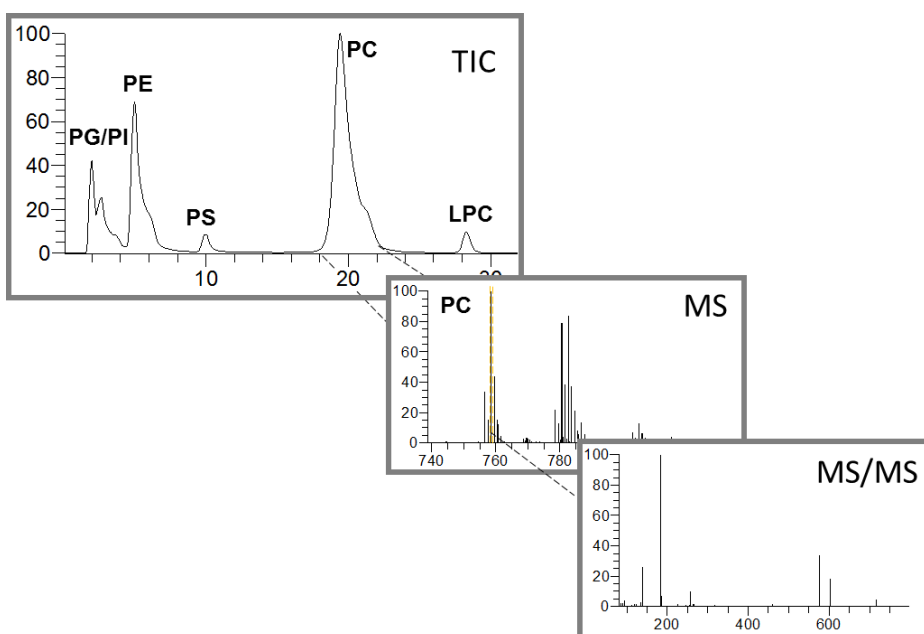
Tabela 2. Charakterystyczne jony fragmentaryczne obserwowane dla fosfolipidów na widmach MS/MS.

Klasa fosfolipidów	Wykrywane jony	Tryb pozytywny	Tryb negatywny
Fosfatydylocholino (PC)	$[\text{M}+\text{H}]^+$	PIS m/z 184	-
Fosfatydyloetanolaminy (PE)	$[\text{M}+\text{H}]^+$ $[\text{M}-\text{H}]^-$	NLS 141 Da	-
Fosfatydyloseryny (PS)	$[\text{M}+\text{H}]^+$ $[\text{M}-\text{H}]^-$	NLS 185 Da	NLS 87 Da
Fosfatydyloglicerole (PG)	$[\text{M}-\text{H}]^-$	-	NLS 74 Da
Fosfatydyloinosytole (PI)	$[\text{M}-\text{H}]^-$	-	PIS m/z 241
Kardiolipiny (CL)	$[\text{M}-\text{H}]^-$ $[\text{M}-2\text{H}]^{2-}$		
Sfingomieliny (SM)	$[\text{M}+\text{H}]^+$	PIS m/z 184	-

1.1.2.1. Nietargetowe podejścia lipidomiczne oparte na LC-MS

Lipidomika nietargetowa wykorzystuje połączenie LC-MS, którego ogromną zaletą jest możliwość rozdzielenia i identyfikacji związków zawartych w złożonym ekstrakcie lipidowym. Jest to bardzo istotne z uwagi na strukturalną różnorodność polarnych lipidów. Określanie związków należących do danej klasy lipidów przy zastosowaniu połączenia spektrometrii mas oraz chromatografii cieczowej, daje możliwość wyizolowania i zateżenia różnych grup związków zależnie od ich właściwości fizykochemicznych. Stosowanie różnych platform LC-MS polepsza rozdzielczość i czułość analiz lipidomicznych.

Identyfikacja związków lipidowych przy zastosowaniu platform LC-MS przeprowadzana jest na podstawie czasu retencji (RT) oraz interpretacji widm MS i MS/MS (Rycina 3.5). Rozdział chromatograficzny prowadzony jest zarówno w odwróconym układzie faz (RP), normalnym (NP) lub przy wykorzystaniu chromatografii oddziaływań hydrofilowych (HILIC). Czas retencji zależy od warunków analizy między innymi rodzaju kolumny LC, prędkości przepływu, czy rodzaju eluentów. Pojemność oraz selektywność kolumny są zróżnicowane i zależą w dużej mierze od producenta, podczas gdy sprawność i rozdzielczość w pewnym stopniu można modyfikować zmieniając parametry elucji.



Rycina 3.5. Dane LC-MS uzyskane w wyniku analizy ekstraktu lipidów z wykorzystaniem platformy HILIC-MS. Przykład chromatogramu dla wszystkich jonów (TIC), oraz widmo MS dla PC ilustrujące protonowane jony $[M+H]^+$ związków należących do tej klasy. Widmo MS/MS jednego wyselekcjonowanego z widma MS jonu prekursora przedstawia jony fragmentaryczne, na podstawie których ustalana jest struktura związku.

W lipidomice najczęściej stosowane są kolumny NP lub HILIC umożliwiające rozdzielenie wszystkich klas związków lipidowych w trakcie pojedynczej analizy. Analizy tego typu wykorzystywane są w celu znalezienia biomarkerów danej choroby, jak również wykrycia zaburzeń szlaków metabolicznych, występujących w jej przebiegu. Lipidy wymywane są z tego typu kolumn w zależności od ich charakteru hydrofilowego, który warunkują właściwości ugrupowania polarnego ich cząsteczek tzw. „głowy hydrofilowej”. Z kolei rozdział lipidów na kolumnach C18 w odwróconym układzie faz determinowany jest ich właściwościami hydrofobowymi, warunkowanymi przede wszystkim ilością atomów węgla i

stopnia nasycenia w łańcuchach acylowych kwasów tłuszczowych. Dlatego lipidy zawierające długołańcuchowe i nasycone kwasy tłuszczowe wymywają się z kolumny później niż lipidy zawierające wielonienasycone kwasy tłuszczowe o krótszych łańcuchach acylowych. Podejścia lipidomiczne wykorzystujące kolumny C18 zazwyczaj dotyczą analiz konkretnej klasy lipidów po wcześniejszym jej wyizolowaniu z ekstraktu, lub typowych analizach targetowych.

Podejścia nietargetowe umożliwiają identyfikację oraz ilościową analizę ogromnej liczby związków lipidowych jednocześnie. Dlatego identyfikacja kilkuset związków oraz ich ilościowa ocena stanowi duże wyzwanie, lecz w celu analizy, procesowania, organizacji jak również wizualizacji danych z analiz nietargetowych wciąż opracowywane są nowe narzędzia bioinformatyczne.

1.1.2.2. Targetowe podejścia lipidomiczne „Shotgun”

Wspomniane już wcześniej analizy targetowe wykorzystują wiedzę na temat charakterystycznego sposobu fragmentacji lipidów należących do konkretnej klasy w celu zdefiniowania parametrów monitorowanych w trybie skanowania jonów prekursora (PIS) lub utraty cząsteczki obojętnej (NLS). Przykładowo monitorowanie jonów prekursora o m/z 184 wykorzystywane jest przy identyfikacji lipidów należących do PC, natomiast monitorowanie utraty masy związku o odpowiadającej 141 Da jest powszechnie stosowane w analizach lipidomicznych typu „shotgun”. Selektywność tego typu podejścia analitycznego jest podstawową jego zaletą, gdyż aby ją uzyskać wykorzystywana jest specyfika sposobu fragmentacji związków należących do konkretnej klasy fosfolipidów, posiadających charakterystyczne polarne ugrupowanie (Tabela 1). Analizy tego typu mogą być przeprowadzane bezpośrednio podając roztwór próbki na źródło jonizacji spektrometru mas typu ESI-QqQ lub Q-trap lub przy wykorzystaniu systemu LC-MS. Podejście polegające na monitorowaniu wybranych par jonów (MRM) zazwyczaj realizowane jest przy zastosowaniu spektrometrów mas typu potrójnego kwadrupola i stanowi z kolei typ analizy targetowej, polegającej na screeningu specyficznych par jon macierzysty/jon potomny. Ten tryb tandemowej spektrometrii mas ma swoje podstawowe zastosowanie do analizy ilościowej konkretnych związków, w oparciu o odpowiednie krzywe kalibracyjne.

Najważniejszą zaletą lipidomiki „shotgun” jest fakt, że rozdział chromatograficzny nie jest wymagany, aby uzyskać widmo masowe dla wszystkich związków lipidowych należących do tej samej klasy. Dodatkową zaletą tego podejścia jest szybkość, pozwalająca na analizę kilkuset próbek w krótkim czasie. Z kolei podstawową wadą jest dramatycznie niska czułość spowodowana zjawiskiem supresji jonów w źródle. Dodatkowo określenie parametrów do analizy typu NLS lub PIS dla niektórych związków może nie być możliwe, lub parametry takie mogą nie być wystarczająco specyficzne dla tylko jednej grupy związków (np. monitorowanie m/z 184 wykorzystywane jest zarówno przy identyfikacji PC jak i SM, Tabela 2).

Literatura:

- U. Loizides-Mangold. On the future of mass-spectrometry-based lipidomics. *FEBS J.*, 2013, 280(12):2817-29.
- X. Han. Lipidomics for studying metabolism. *Nature Rev. Endocrin.*, 2016, 12(11):668-679.
- L. Yang et al. Recent advances in lipidomics for disease research. *J. Separation Sci.*, 2016, 39:38-50.
- M. Holčapek, G. Liebisch, K. Ekroos. Lipidomic Analysis. *Analytical Chemistry*, 2018, 2090:4249-4257.

1.2. Ilościowa analiza lipidomiczna

W lipidomice analizę ilościową można przeprowadzić na różnym poziomie metodą relatywną jak i absolutną. Relatywna analiza ilościowa (relative quantification) fosfolipidów dokonywana jest po wcześniejszym ich rozdzielaniu na klasy z użyciem TLC, identyfikacji poszczególnych klas przy użyciu odpowiednich standardów PL i ilościowym oznaczeniu fosforu w poszczególnych plamkach ujawnionych na płycie odpowiadających konkretnym klasom.

Na poziomie molekularnym analizę ilościową metodą relatywną jak i absolutną można przeprowadzić stosując podejście wykorzystujące połączenie LC-MS, jak również podejście typu "shotgun". Metoda relatywna ma na celu ocenę profilu danej klasy związków, jak również znalezienie różnic tego profilu w wyniku porównania odpowiednich grup badanych. W tym celu po zidentyfikowaniu poszczególnych lipidów, należących do danej klasy, konieczne jest obliczenie powierzchni pod pikiem chromatograficznym odpowiadającym każdemu z nich. Następnie obliczone pole powierzchni normalizowane jest do sumy pól powierzchni pików wszystkich związków zidentyfikowanych w danej klasie, uzyskując w ten sposób relatywną zawartość konkretnego fosfolipidu. Stosuje się także dodatek odpowiedniej ilości standardu wewnątrzny do każdej z próbek (liaztu komórek, homogenatu tkanki czy płynu ustrojowego) przed dokonaniem ekstrakcji lipidów, bądź do ekstraktu lipidów bezpośrednio przed analizą LC-MS. W takim przypadku normalizacji powierzchni pików poszczególnych fosfolipidów dokonuje się w oparciu o powierzchnię pików standardu wewnętrznego. Oba opisane sposoby pozwalają na określenie tzw. relatywnej zawartości fosfolipidu w obrębie danej klasy.

W celu przeprowadzenia analizy ilościowej z użyciem LC-MS konkretnego lipidu metodą absolutną, przeprowadza się analizy z użyciem spektrometru mas typu potrójnego kwadrupola w trybie MRM. Do przeprowadzenia obliczeń konieczne jest wygenerowanie krzywej kalibracyjnej dla związku, którego zawartość w próbce jest określana.

Wadą zarówno ilościowych jak i jakościowych analiz wykorzystujących chromatografię w układzie NP lub HILIC jest brak możliwości rozróżnienia fosfolipidów należących do tej samej klasy posiadających tę samą masę molekularną. Przykładowo, fosfolipid zidentyfikowany jako PC(36:4), może zawierać różne reszty kwasów tłuszczowych w obrębie swej struktury dając kilka opcji PC(16:0/20:4), PC(18:1/ 18:1), lub PC(18:0/ 18:2). W tym przypadku powinna być zastosowana alternatywna metoda uwzględniająca etap wstępnego oczyszczania próbki jak również rozdziału z użyciem kolumny C18.

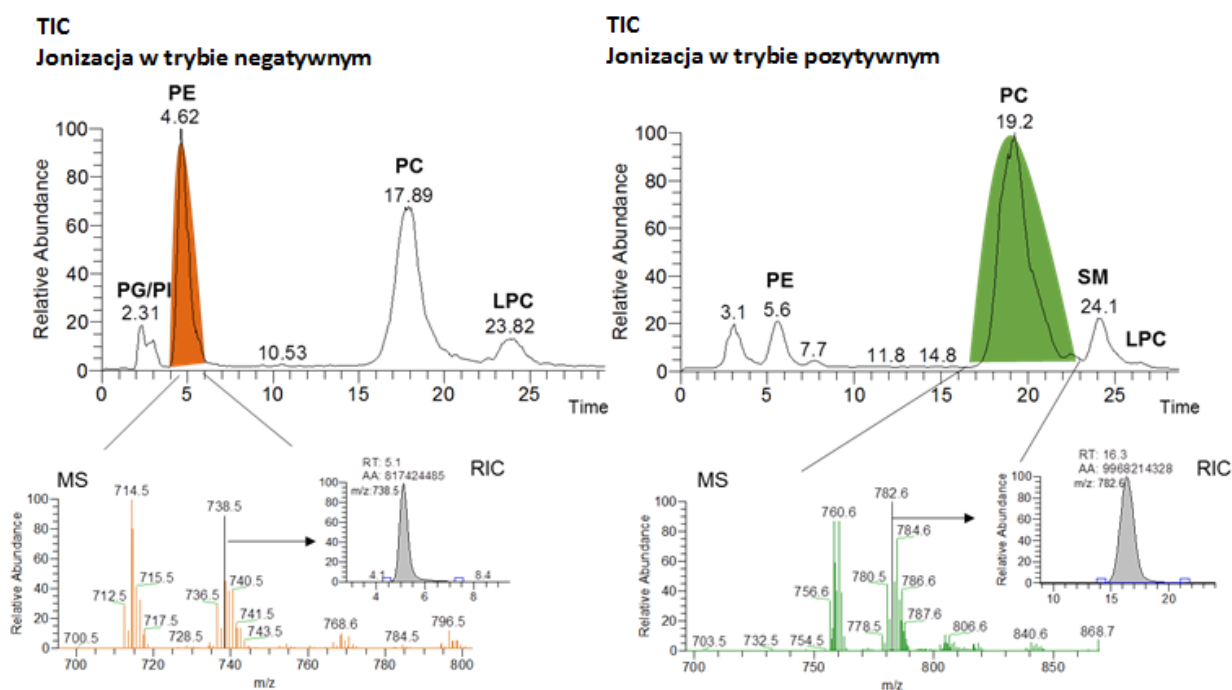
Literatura:

X. Han, K. Yang, R.W. Gross. Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrom. Rev.*, 2012, 31:134-78.

1.3. Przetwarzanie danych w celu identyfikacji/ilościowej oceny lipidów

W wyniku przeprowadzenia wielkoskalowych analiz lipidomicznych z użyciem LC-MS generowana jest ogromna ilość danych, których analiza wymaga wykorzystania odpowiednio wyspecjalizowanych narzędzi bioinformatycznych. Lipidomika wymaga także użycia obszernych baz danych w celu dokonania identyfikacji. Z uwagi na to, że nie istnieją uniwersalne narzędzia bioinformatyczne pozwalające na przeprowadzenie zautomatyzowanej procedury obróbki danych LC-MS, ich interpretacja stanowi nadal duże wyzwanie.

Pierwszy etap analizy danych lipidomicznych powinna stanowić identyfikacja związków w oparciu o RT oraz interpretację widma MS w celu znalezienia wartości m/z jonu molekularnego, a następnie interpretację widm MSMS dla tego jonu (Rycina 3.6).



Rycina 3.6. Chromatogram LC uzyskany z analizy lipidomicznej całkowitego ekstraktu lipidów keratynocytów. Na rycinie wyróżniono kolorami piki odpowiadające klasie fosfolipidów PE i PC oraz ich widma MS. Poprzez wyselekcjonowanie odpowiedniego sygnału na widmie MS istnieje możliwość wygenerowania chromatogramu (reconstructed ion chromatogram RIC) dla związku, reprezentowanego przez dany sygnał. Powierzchnia pod każdym z pików może posłużyć za podstawowy parametr do analizy ilościowej.

Identyfikacja lipidów przeprowadzana jest zazwyczaj „ręcznie” w oparciu o wnikliwą analizę danych, dokładną wartość masy oraz bazy danych zawierające widma masowe sfragmentowanych związków. Obecnie nie istnieją uniwersalne bazy widm fragmentacyjnych związków lipidowych pozwalające odpowiednio je sklasyfikować oraz zidentyfikować, aczkolwiek dostępnych jest kilka baz stanowiących pewne wsparcie takie jak: LipidBank (<http://lipidbank.jp/>), LIPIDAT (<http://lipidata.in/>) and LIPID MAPS (<http://www.lipidmaps.org/>). Również prawie każdy producent spektrometrów mas opracował swoje autorskie oprogramowanie do obróbki danych uzyskiwanych z analiz lipidomicznych. Opracowane pakiety programów umożliwiają obróbkę danych lipidomicznych uzyskiwanych przy zastosowaniu specyficznego trybu zbierania, danych jak również konkretnej aplikacji. W chwili obecnej komercyjnie dostępne są oprogramowania: LipidView™ (zaproponowane przez firmę Sciex), dedykowane dla analiz w trybie monitorowania jonów prekursora jak również w trybie utraty cząsteczki neutralnej, LipidSearch™ (firmy Thermo Scientific) dedykowane do obróbki danych lipidomicznych LC/MS, w oparciu widma masowe wysokiej rozdzielczości otrzymywane przy zastosowaniu spektrometrów mas typu Orbitrap™. Dostępne są również bezpłatne narzędzia informatyczne oraz biblioteki widm masowych analiz lipidomicznych.

Do najpopularniejszych zaliczany jest

- LipidXplorer (https://wiki.mpi-cbg.de/wiki/lipidx/index.php/Main_Page.),
- Alex software (http://mslipidomics.info/contents/?page_id=133),
- Lipid Blast (<http://fiehnlab.ucdavis.edu/projects/LipidBlast>)
- MS-DIAL (http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/).

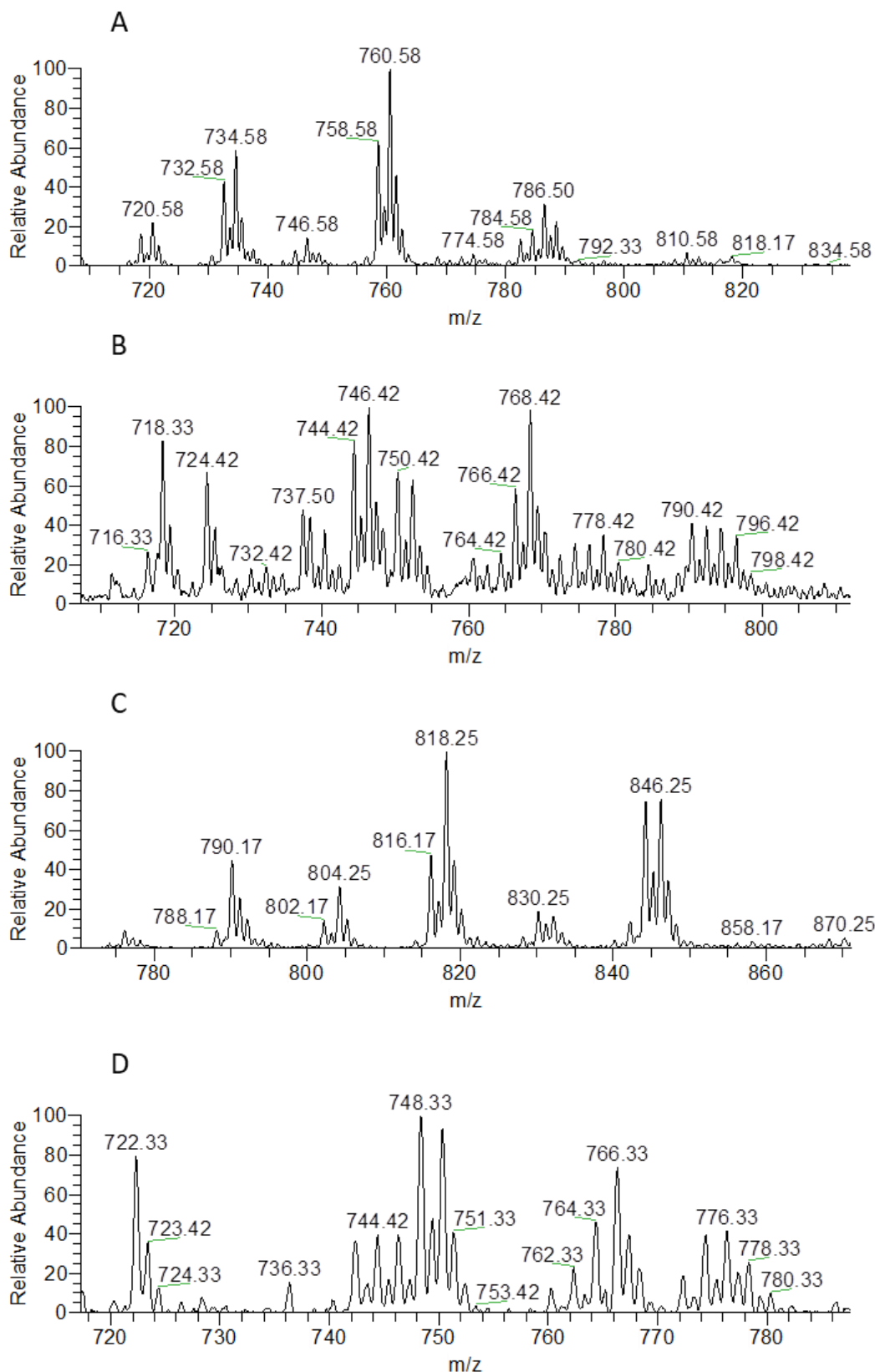
Narzędzia te pozwalają na przeprowadzenie zarówno jakościowej jak i ilościowej analizy danych lipidomicznych uzyskiwanych z zastosowaniem różnych podejść, jak również spektrometrów mas od różnych producentów. LipidXplorer oraz ALEX zostały zaprojektowane z myślą o obróbce danych z analiz typu “shotgun” uzyskiwanych przy zastosowaniu spektrometrów mas wysokiej rozdzielczości. Z kolei LipidBlast umożliwia obróbkę danych uzyskiwanych z analiz MSMS polarnych lipidów przeprowadzanych z użyciem spektrometrów mas wysokiej jak i niskiej rozdzielczości. MS-DIAL przystosowany jest do obróbki danych analiz zarówno typu “data dependent” jak i „independent MS/MS”.

Po przeprowadzeniu analizy ilościowej dalsza obróbka danych zazwyczaj obejmuje ich normalizację w celu przeprowadzenia analizy statystycznej przy zastosowaniu podejść opisanych w dziale niniejszego manuskryptu dotyczącym analiz metabolomicznych.

1.4. Część praktyczna

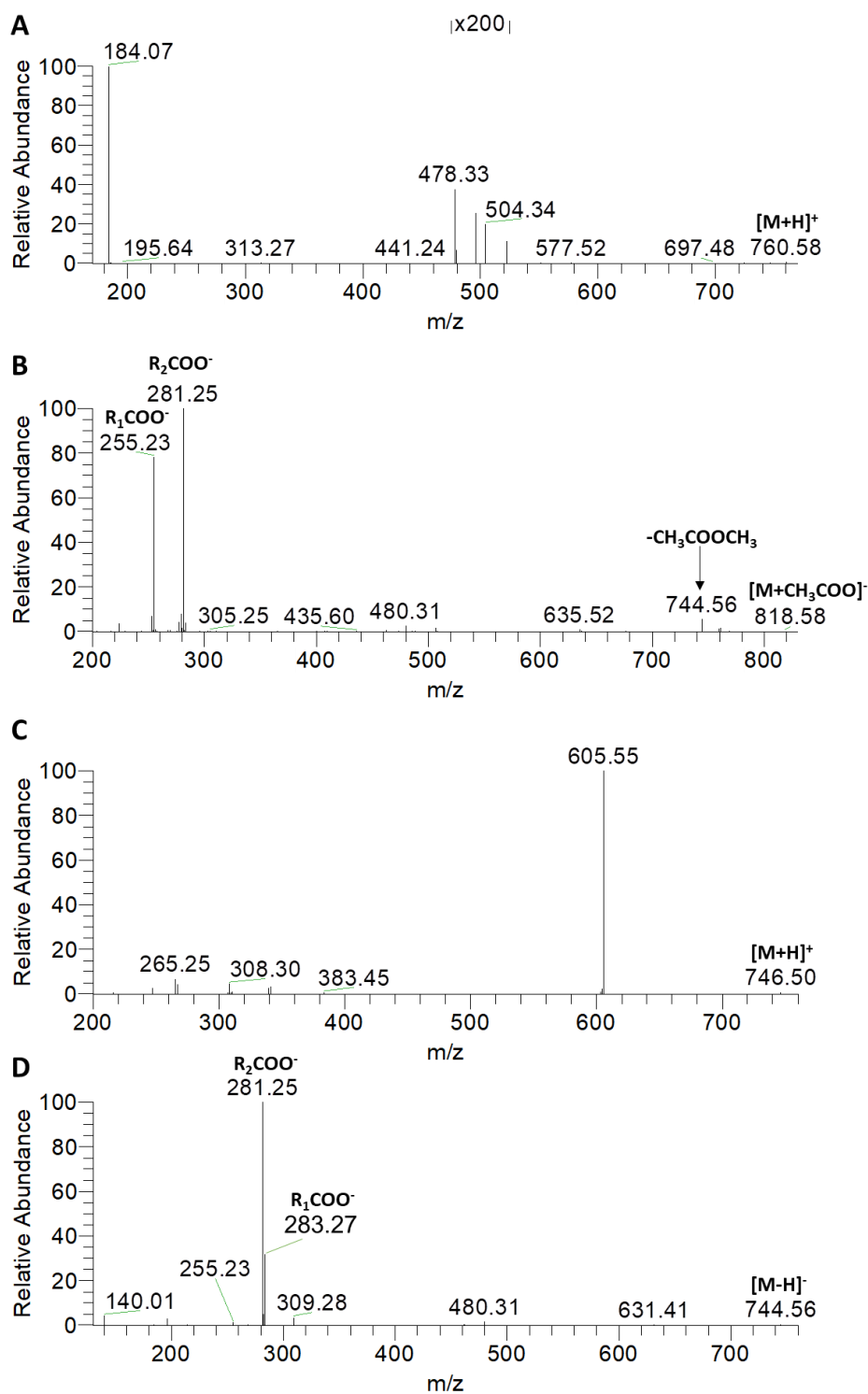
1.4.1. Analiza lipidomiczna profilu fosfolipidów monocytów linii THP1 z wykorzystaniem HILIC-LC-MS.

Ekstrakty lipidów uzyskane z monocytów linii THP1 przeanalizowano w trybie pozytywnej jak i negatywnej jonizacji metodą LC-ESI-MS i LC-ESI-MS/MS. Rycina 3.7 przedstawia widmo masowe LC-ESI-MS otrzymane w trybie pozytywnej jonizacji, dla czasie retencji (RT) 22 min (A) i 7 min (B) and oraz w trybie jonizacji negatywnej, przy RT 22 min (C) oraz RT 7 min (D).



Rycina 3.7. Widma masowe LC-ESI-MS uzyskane w wyniku analizy całkowitego ekstraktu lipidów w trybie pozytywnej jonizacji, dla czasu retencji (RT) 22 min (A) i 7 min (B) and oraz w trybie jonizacji negatywnej, przy RT 22 min (C) oraz RT 7 min (D).

W celu uzyskania informacji na temat klasy fosfolipidów konieczne jest przeanalizowanie sposobu fragmentacji wybranych jonów molekularnych. Rycina 3.8 przedstawia widma masowe ESI-MS/MS uzyskane dla jonów molekularnych o m/z 760.58 (Rycina 3.8A), m/z 818.58 (Rycina 3.8B), m/z 746.5 (Rycina 3.8C) oraz m/z 744.56 (Rycina 3.8D).



Rycina 3.8. Widma masowe ESI-MS/MS dla jonów molekularnych o m/z 760.58 (A), m/z 818.58 (B), w trybie jonizacji pozytywnej, oraz jonów o m/z 746.50 (C) and m/z 744.56 (D), w trybie jonizacji negatywnej.

Zadania

1. Na podstawie charakterystycznych sposobów fragmentacji fosfolipidów, rozstrzygnij do jakiej klasy fosfolipidów należą fosfolipidy, dla których piki na chromatogramie LC-MS zaobserwowano przy czasie retencji: 7 min oraz 22 min?
2. Na podstawie widm masowych ESI-MS/MS uzyskanych dla związku o czasie retencji 7 min w trybie jonizacji negatywnej i pozytywnej, zaproponuj budowę strukturalną tego fosfolipidu.
3. Obecności jakich innych klas fosfolipidów można się spodziewać w przypadku lipidomicznej LC-ESI-MS ekstraktu lipidów monocytów linii THP1s?
4. Opisz inne podejście pozwalające ocenić zmiany w profile fosfolipidów monocytów linii THP1.

1.5. Dodatkowe źródła wiedzy

J.D. Martins, E.A. Maciel, A. Silva, I. Ferreira, P. Domingues, B.M. Neves, M.T. Cruz, M.R.M. Domingues. Phospholipidomic profile variation on THP-1 cells exposed to skin or respiratory sensitizers and respiratory irritant. *Journal of Cell Physiology*, 2016, 231(12):2639-51.

M. Pulfer, R.C. Murphy. Electrospray mass spectrometry of phospholipids. *Mass Spectrometry Reviews*, 2003, 22(5):332-64.

Y.H. Rustam, E. Gavin. Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics. *Anal. Chem.*, 2018, 90:374-397.

K. Yang, X. Han. Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biomedical Sciences*, 2018, 41:954-969.

H.C. Lee, T Yokomizo. Applications of mass spectrometry-based targeted and nontargeted lipidomics. *Biochem Biophys Res Commun*, 2018, doi: 10.1016/j.bbrc.2018.03.081.

S. Sethi, E. Brietzke. Recent advances in lipidomics: Analytical and clinical perspectives. *Prostaglandins Other Lipid Mediat.*, 2017, 128-129:8-16.

Thanks are due to Fundação para a Ciência e a Tecnologia (FCT, Portugal), the European Union, the Quadro de Referência Estratégica Nacional (QREN), the Programa Operacional Potencial Humano (POPH), the Fundo Europeu de Desenvolvimento Regional (FEDER) and Programa Operacional Factores de Competitividade (COMPETE) for funding the QOPNA (Química Orgânica de Produtos Naturais e Agroalimentares) research unit and to RNEM (REDE/ 1504/REM/2005) for the Portuguese Mass Spectrometry Network.

Moduł 4

Proteomika

Tânia Melo¹, Rita Ferreira¹, Rosário Domingues^{1,2}, Pedro Domingues¹,

Mass Spectrometry Center, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal

1 QOPNA (Química Orgânica de Produtos Naturais e Agroalimentares)

2 CESAM (Centro de Estudos do Ambiente e do Mar)

I. Wstęp

Tradycyjne metody stosowane dotychczas w analizie białek dostarczają jedynie cząstkowych informacji na temat zmian zachodzących w organizmie zarówno podczas przemian fizjologicznych, jak również podczas rozwoju chorób. Proteomika jest nauką zajmująca się wielkoskalową analizą białek, która daje nam narzędzia do poznania i lepszego zrozumienia mechanizmów molekularnych, a także interakcji białko-białko oraz roli białek w sygnalizacji w warunkach fizjologicznych i patologicznych, co ma na celu identyfikację biomarkerów białkowych oraz poprawę terapii. Pierwsza część tego modułu skupia się na aktualnej metodologii stosowanej w analizie i identyfikacji białek. Obejmuje to opis najważniejszych technik, w tym metod rozdzielania białek stosowanych przed analizą z użyciem spektrometrii mas. Druga część modułu ma na celu zapoznanie studentów z dostępnym oprogramowaniem do analizy białek. Zajęcia prowadzone są w formie wykładów interaktywnych z udziałem studentów, sesji demonstracyjnych/praktycznych oraz bioinformatycznej analizy otrzymanych widm, które pozwolą na analizę konkretnych wyników.

II. Założenia i cele przedmiotu

Założenia

Głównym celem kursu jest zapoznanie studentów z podstawowymi pojęciami dotyczącymi analizy proteomicznej. Podczas kursu studenci zdobędą zarówno podstawową wiedzę teoretyczną, jak i umiejętności praktyczne z zakresu przeprowadzania analiz proteomicznych i zastosowania proteomiki w badaniach biomedycznych, w tym metod przygotowania próbek biologicznych, oczyszczania i ekstrakcji, identyfikacji i oznaczania ilościowego białek, a także wykorzystania oprogramowania bioinformatycznego do opracowania uzyskanych wyników proteomicznych.

Specyficzne efekty kształcenia:

Po ukończeniu kursu studenci będą potrafili:

1. definiować i posługiwać się powszechnie stosowaną terminologią proteomiczną;
2. opisać zasady technik najczęściej stosowanych w proteomice;
3. rozpoznać różnice w przeprowadzaniu analiz proteomicznych z użyciem różnych spektrometrów mas;
4. wybrać odpowiednie podejście do analizy proteomicznej w celu rozwiązania konkretnego problemu;

5. wykorzystywać oprogramowanie bioinformatyczne do opracowania uzyskanych wyników proteomicznych;
6. interpretować dane wynikające z identyfikacji białek;
7. w sposób naukowy oraz popularnonaukowy zaprezentować wyniki i jednoznacznie uzasadnić wnioski;
8. samodzielnie kontynuować naukę i rozwijać swoje umiejętności.

III. Spis treści

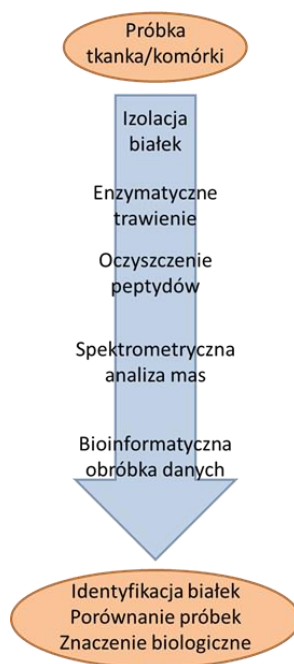
1. Wstęp do proteomiki
2. Podejścia analityczne stosowane w proteomice
 - 2.1. Przygotowanie próbek
 - 2.2. Identyfikacja białek oparta na analizach MS: koncepcje
 - 2.3. Metody spektrometrii mas stosowane do identyfikacji białek
 - 2.4. Ilościowa analiza proteomiczna
 - 2.5. Identyfikacja PTM
 - 2.6. Identyfikacja wewnątrzkomórkowej lokalizacji białek
 - 2.7. Identyfikacja interakcji między białkami
3. Przetwarzanie danych w celu identyfikacji białek
 - 3.1. Bazy danych zawierające sekwencje białkowe
 - 3.2. Przeszukiwanie białkowych baz danych
 - 3.3. Analiza danych
4. Część praktyczna
 - 4.1. Identyfikacja białek w komórkach z użyciem podejścia PMF i PFF za pomocą platformy MASCOT
 - 4.2. Identyfikacja białek z eksperymentu LC-MS
 - 4.2.1. Zamiana otrzymanych danych na pliki mzML z użyciem MSConverter (ProteoWizard)
 - 4.2.2. Utworzenie bazy danych FASTA używając serwera Uniprot (SwissProt)
 - 4.2.3. Wyszukiwarka: SearchGUI
 - 4.2.4. Generowanie i ocena wyników: użycie PeptideShaker do wizualizacji peptydów i białek oraz do walidacji wyników. Analiza PTM
 - 4.2.5. Analiza danych: informacja o białkach, analiza szlaków i ontologia genów

1. Wstęp do proteomiki

Proteomika jest nauką zajmującą się badaniem białek w oparciu o szereg najnowszych technologii umożliwiających szybkie i precyzyjne określenie całego proteomu. Jednakże podczas realizacji tego zadania pojawiają się problemy dotyczące zarówno ograniczeń technicznych, jak i biologicznych. Na przykład, ogromne zróżnicowanie w zakresie stężeń białek w próbkach biologicznych powoduje wiele trudności technicznych, w tym konieczność frakcjonowania i wzbogacania próbek o niższym poziomie białek oraz rozwijania szybkich i czułych metod przetwarzania dużych ilości zróżnicowanych próbek. Ponadto, proteom zdefiniowany jako zestaw białek zależny od genomu specyficznej komórki, tkanki lub organizmu w określonym czasie, jest bardzo dynamiczny i może zmieniać się z wiekiem, w warunkach stresu, przy zróżnicowanym odżywianiu, itp., co dodatkowo utrudnia dokładne określenie profilu białkowego. Również potranslacyjne modyfikacje (post-translational modifications – PTM) w dużej mierze przyczyniają się do zwiększenia złożoności problemu.

Obecnie wyróżnia się kilka różnych podejść do analiz proteomicznych, które w zależności od sytuacji dostarczają danych pozwalających na rozwiązanie różnych kwestii biologicznych. Różny sposób prowadzenia badań proteomicznych może prowadzić do analizy ekspresji genomu w określonych warunkach fizjologicznych, jak również pomiaru stężeń zidentyfikowanych białek prowadzącego do porównania grup eksperymentalnych i próbek kontrolnych. Dzięki analizom proteomicznym można także identyfikować post-transkrypcyjne modyfikacje białek (PTM) oraz bezpośrednie oddziaływania pomiędzy białkami. Jednakże do podstawowych zadań badań proteomicznych zalicza się wykrywanie i ocenę: poziomu białek, ich modyfikacji, lokalizacji w obrębie komórki oraz interakcji pomiędzy białkami.

Do każdej z powyższych analiz są stosowane różne metody proteomiczne, jednakże każda z nich powinna zawierać podstawowe etapy, takie jak: izolacja białek z tkanki/komórki, enzymatyczna degradacja, oczyszczenie peptydów, spektrometryczną analizę mas i bioinformatyczną obróbkę danych (rycina 4.1.)



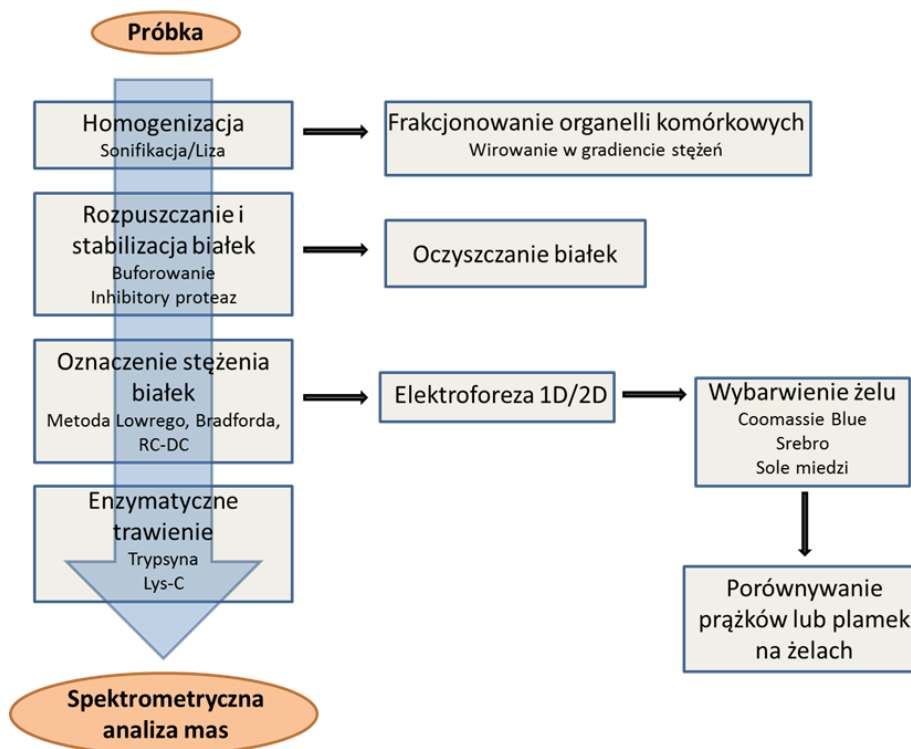
Rycina 4.1. Podstawowe etapy analizy proteomicznej

W zależności od otrzymanych wyników wyróżnia się różne podejścia do analiz proteomicznych, które można zakwalifikować jako proteomikę poznawczą (discovery proteomics) lub proteomikę ukierunkowaną (targeted proteomics). Standardowo w proteomicie poznawczej stosuje się ograniczoną liczbę próbek w celu zoptymalizowania identyfikacji białka. Natomiast w ukierunkowanych analizach proteomicznych ilość monitorowanych białek jest ograniczana ze względu na konieczność uzyskania najwyższej czułości. Niemniej jednak im większa liczba analizowanych białek, tym trudniej jest je zidentyfikować i oznaczyć ilościowo.

2. Podejścia analityczne stosowane w proteomicie

2.1. Przygotowanie próbek

Kluczowym elementem w otrzymaniu rzetelnych wyników jest prawidłowe przygotowanie próbki. Na jakość próbek przygotowywanych do analizy proteomicznej wpływa przede wszystkim powtarzalność zastosowanych metod ekstrakcji i oczyszczania białek. Są to zazwyczaj bardzo złożone procesy, ze względu na fakt, że białka są wysoce heterogenicznymi grupami związków różniących się wielkością, ładunkiem, rozpuszczalnością oraz stężeniem (np. zakres stężeń białek w surowicy ma rozpiętość większą niż 10^8 mg/mL). W związku z tym nie ma standardowej metody przygotowywania próbek, co dodatkowo utrudnia porównanie wyników badań. Jednakże pomimo różnic w protokołach przygotowania próbki wszystkie metody dążą do określonego celu – otrzymania czystej mieszaniny peptydów przeznaczanej do analizy z użyciem spektrometru mas. Na rycinie 4.2. przedstawiono ogólną procedurę przygotowania próbki, podkreślając różne możliwości obróbki próbki.



Rycina 4.2. Ogólna procedura przygotowywania próbek do analizy proteomicznej

Klasycznym podejściem do proteomiki jest użycie dwuwymiarowej elektroforezy żelowej (2-DE) do analizy spektrometrii mas (MS) i identyfikacja białek. Inną powszechnie stosowaną metodą jest zastosowanie jednowymiarowej elektroforezy SDS-PAGE w podejściu nazwanym GeLC-MS/MS. Jednak z powodu problemów związanych z technikami rozdzielania w żelu (takimi jak: wytrącanie białek, trudności związane z rozdzieleniem białek, niska odtwarzalność), w ostatnich latach opracowano podejścia omijające rozdział elektroforetyczny (tzw. analizy typu "no-gel"). Zwykle obejmują one zastosowanie wielokierunkowej chromatografii ciekłowej o wysokiej rozdzielczości, która umożliwia wstępne frakcjonowanie próbek i zapewnia wysoką czułość prowadzonych analiz.

Literatura:

J. Lovric. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*. ISBN: 978-0-470-03524-5, Chapter 2.1, 2011.

2.2. Identyfikacja białek oparta na analizach MS: koncepcje

Istnieje kilka strategii identyfikacji białek z zastosowaniem spektrometrii mas. Do głównych zalicza się tzw. podejścia „bottom-up” oraz „top-down” (rycina 4.3.). Najbardziej tradycyjnym jest podejście „bottom-up”, które opiera się na analizie mas peptydów powstałych w wyniku enzymatycznego trawienia białka. Identyfikacja poszczególnych peptydów z użyciem spektrometrii mas lub tandemowej spektrometrii mas pozwala następnie na określenie sekwencji cząsteczki całego białka.

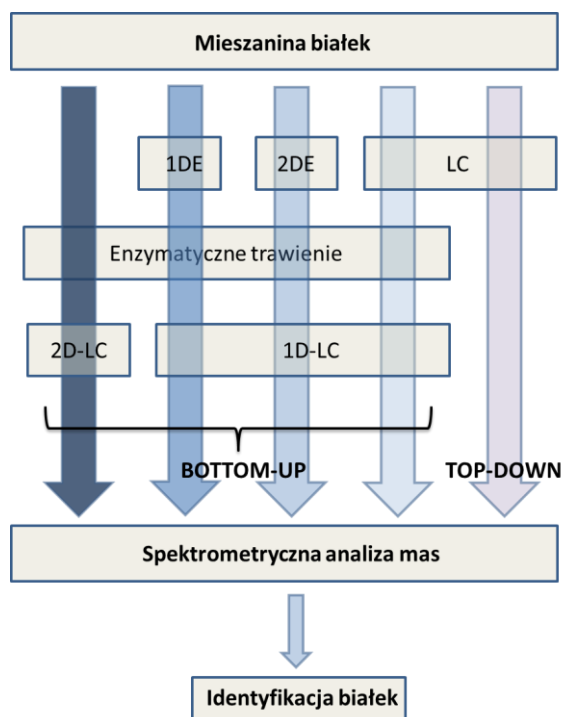
Do enzymatycznego trawienia białek można stosować różne enzymy lub kombinację enzymów, chociaż najczęściej stosowana jest trypsyna. Trawienie białek rozdzielonych elektroforetycznie jest zazwyczaj wykonywane bezpośrednio w żelu, po wycięciu odpowiednich fragmentów zawierających interesujące nas białka. Następnie strawione peptydy są ekstrahowane z żelu. Przy stosowaniu metod typu „no-gel” białka są trawione bezpośrednio w roztworze, a następnie otrzymane peptydy są oczyszczane przed analizą za pomocą MS.

W celu zwiększenia liczby oznaczonych białek w próbce zwykle wymagane jest zastosowanie chromatograficznej separacji peptydów. Jest to powiązane z tzw. „efektem supresji jonów” wynikającym z używania technik jonizacji takich jak MALDI czy ESI, które tłumią sygnały peptydów hydrofilowych o niższym powinowactwie protonowym. Rozdział chromatograficzny prowadzi się wówczas za pomocą nanoLC-MS na kolumnie C18 w odwróconym układzie faz. Peptydy są wówczas wymywane z kolumny z gradientem rozpuszczalnika, który stanowi woda/acetonitryl z dodatkiem środka sprzęgającego jon (na przykład kwasu mrówkowego). W odniesieniu do złożonych próbek stosuje się dłuższe gradienty lub wykorzystuje się rozdział z użyciem dwóch typów chromatografii, np. chromatografii jonowymiennej oraz chromatografii w układzie odwróconych faz (2DLC, MudPIT). Przy zastosowaniu metod, które nie zakładają przeprowadzania żadnego wcześniejszego rozdzielania białek przed trawieniem („shotgun”), zwykle konieczne jest oddzielenie peptydów z użyciem 2DLC w celu uzyskania lepszych wyników.

Po rozdzieleniu peptydów dochodzi do ich jonizacji z zastosowaniem ESI lub MALDI jako źródła jonizacji, a następnie analizy za pomocą spektrometrii mas (MS) lub tandemowej

spektrometrii mas (MS/MS). Istniejąca dyskryminacja różnych peptydów podczas jonizacji MALDI i ESI przyczynia się do różnych wskaźników pokrycia sekwencji analizowanych białek w przypadku stosowania różnych sposobów jonizacji. Przyjmuje się jednak, że ESI umożliwia nieco lepszy zasięg w przypadku większości peptydów, chociaż MALDI jest mniej wrażliwy na zanieczyszczenie i efekt supresji jonów.

W podejściu „top-down”, oczyszczone białko jest jonizowane, zwykle z użyciem źródła ESI i fragmentowane z wykorzystaniem techniki ECD (Electron Capture Dissociation) lub ETD (Electron Transfer Dissociation). Następnie fragmenty jonów są analizowane za pomocą analizatorów o wysokiej rozdzielczości, takich jak FT-ICR lub orbitrap. Podejście to zazwyczaj wymaga wyższego stopnia oczyszczania białek, zazwyczaj przy użyciu wielowymiarowej chromatografii ciekowej. Ponadto podejście takie zakłada większe stężenia białek w analizowanych próbkach i wymaga bardzo drogiego oprzyrządowania.



Rycina 4.3. Podejścia do identyfikacji białka opartej na MS

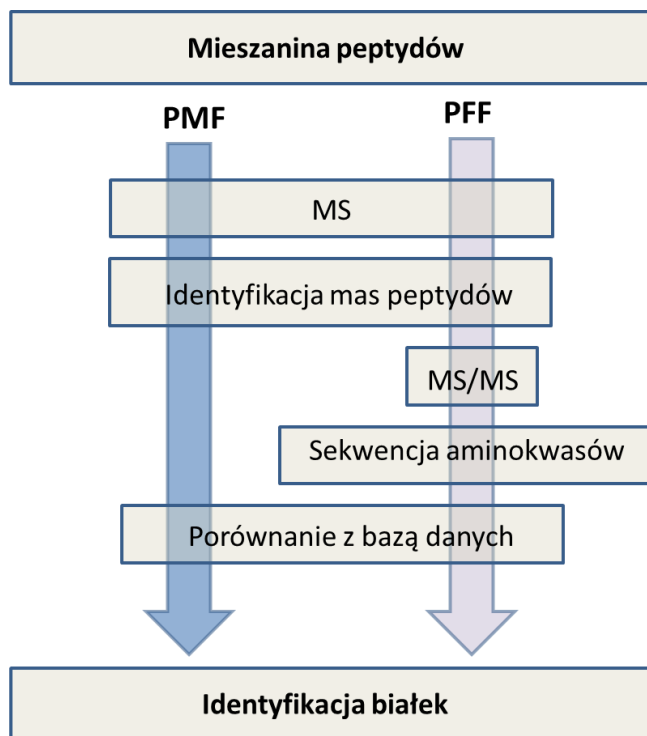
Literatura:

Z. Szabo, T. Janaky. Challenges and developments in protein identification using mass spectrometry. *Trends Anal. Chem.*, 2015, 69:76-87.

2.3. Metody spektrometrii mas stosowane do identyfikacji białek

W przypadku analiz typu „bottom-up” do identyfikacji białek stosuje się dwie główne techniki, tzw.: technikę peptydowych odcisków palców (PMF, peptide mass fingerprinting) lub technikę specyficznego fragmentacji peptydów (PFF, peptide fragmentation fingerprinting) (rycina 4.4). Technika PMF identyfikuje białka, dopasowując masę peptydów pochodzących z enzymatycznego trawienia z teoretycznymi masami peptydów generowanych z białkowych baz danych. Technika ta może być stosowana w próbkach zawierających pojedyncze białka, dlatego pozwala na szybką analizę, pod warunkiem, że obserwuje się kilka specyficznych

peptydów (najlepiej powyżej pięciu). Identyfikacja białka z zastosowaniem tej metody odbywa się na podstawie różnych algorytmów, jednakże zazwyczaj są one oparte na porównaniu mas powstałych peptydów z teoretycznymi masami peptydów powstałymi w wyniku trawienia białka.



Rycina 4.4. Główne techniki stosowane do identyfikacji białek przy użyciu podejścia “bottom-up”.

Technika PFF identyfikuje białka dopasowując wzorec fragmentacji MS/MS otrzymanych peptydów z teoretycznym fragmentem widma z bazy danych sekwencji białek. Ta technika bazuje na przewidywaniu i ze względu na wysokie zautomatyzowanie oraz dużą przepustowość próbek jest bardzo przydatna do identyfikacji białek w złożonych matrycach, umożliwiając identyfikację wielu zmiennych modyfikacji. Używając tego podejścia, z zastosowaniem LC-MS/MS otrzymuje się duże ilości danych MS/MS o najwyższej czułości.

Literatura:

J. Lovric. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*. ISBN: 978-0-470-03524-5, Chapters 4.2 and 4.3, 2011.

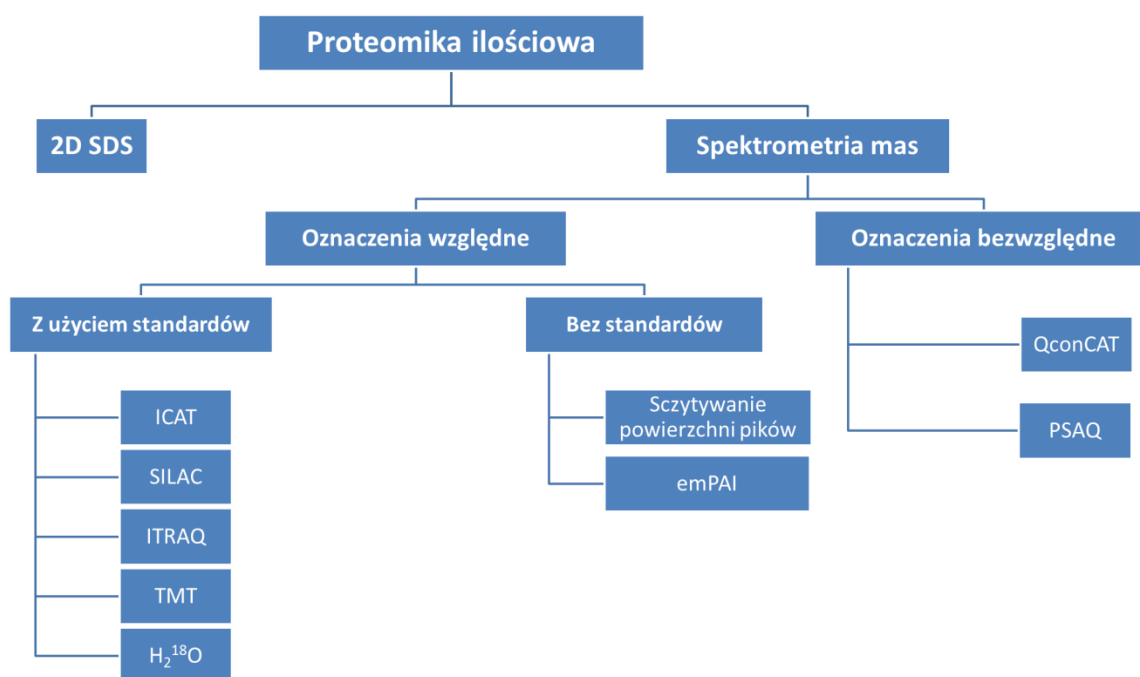
2.4. Ilościowa analiza proteomiczna

Celem niektórych eksperymentów proteomicznych jest jedynie zidentyfikowanie białek obecnych w próbce, jednakże obecnie większość eksperymentów dotyczy oznaczeń ilościowych zmian zachodzących w poziomie ekspresji białek w celu porównania próbek pochodzących od pacjentów w różnych warunkach, na przykład proteomu osób zdrowych i chorych. W tych zastosowaniach konieczne jest dokładne oznaczanie ilościowe białek co stanowi centralny aspekt proteomiki ilościowej. Obecnie istnieją różne metody spektrofotometryczne do wykrywania i pomiaru ilości pojedynczego białka w roztworze. Metody te jednak zawsze wymagają uprzedniego oczyszczenia białka będącego przedmiotem

zainteresowania. Istnieją również metody izolacji poszczególnych białek, zarówno z roztworu, jak i przy użyciu testu w fazie stałej, wykorzystujące specyficzne przeciwciała. Jednakże, jakość takich oznaczeń opiera się na specyficzności przeciwciała i wytrzymałości wiązania antygen-przeciwciała, dlatego uważa się je za metody o małej przepustowości. Niemniej jednak, pogrupowanie takich przeciwciał w mikromacierze białkowe znacząco podwyższa przepustowość analizy ilościowej. Niestety, najbardziej selektywne i czułe mikromacierze zawierają ograniczoną liczbę dobrze scharakteryzowanych przeciwciał.

Wielkoskalowa analiza ilościowa w proteomice jest obecnie realizowana w dwóch głównych podejściach: na podstawie rozdziału białek o wysokiej rozdzielczości przez dwuwymiarową elektroforezę żelową (2D) lub na podstawie względnej ilości jonów peptydów w widmie masowym (rysunek 4.5.). Chociaż umożliwia to jednoczesne oznaczanie stężenia tysięcy białek, zastosowanie elektroforezy w żelu 2D stanowi wąskie gardło w szeroko zakrojonej analizie białek ze względu na trudności w jej przeprowadzaniu i niskiej powtarzalności. Może to być częściowo wykluczone poprzez fluorescencyjne znakowanie próbek poprzedzające rozdział elektroforetyczny (2D DIGE). W 2D DIGE białka z różnych próbek są oznaczane fluorescencyjnie (Cy2, Cy3 i Cy5), a następnie połączone przed rozdzielaniem i ilościową analizą za pomocą dwuwymiarowej elektroforezy żelowej. Takie podejście ma tę zaletę, że minimalizuje przesunięcia pomiędzy oznaczeniami w różnych żelach oraz zmniejsza liczbę żeli w eksperymencie.

Analiza peptydów metodą spektrometrii mas nie jest zasadniczo analizą ilościową. Wynika to z wielu czynników, w tym z różnych skuteczności jonizacji peptydów, efektu tłumienia przy użyciu powszechnych technik jonizacji i ograniczonego zakresu dynamicznego niektórych analizatorów. W związku z tym opracowano różne podejścia do ilościowego oznaczania białek przy użyciu spektrometrii mas, które można zaklasyfikować do metod względnych i bezwzględnych (rycina 4.5.). Względne podejścia ograniczają się do porównania względnego poziomu białek w próbce badanej w stosunku do innych próbek, podczas gdy metody bezwzględne określają dokładną ilość lub stężenie białka.



Rycina 4.5. Proteomiczne metody ilościowej analizy białek.

Podejście do oznaczeń względnie ilościowych przy użyciu LC-MS i LC-MS/MS może być przeprowadzone zarówno przy użyciu standardów wewnętrznych, jak również bez nich. Metody pozbawione standardów wewnętrznych opierają się na zliczaniu otrzymanych widm lub szczytowaniu intensywności poszczególnych pików. Podejście oparte na zliczaniu widm oparte jest o liczbę peptydów zidentyfikowanych z danego białka w celu określenia ich względnej ilości przy użyciu różnych algorytmów, np. emPAI, które obliczają przewidywany poziom białek, biorąc pod uwagę ich sekwencję oraz długość. Oznaczenia względne przy użyciu oceny intensywności pików przeprowadza się porównując względne natężenie jonów peptydowych pochodzących z identycznego białka w różnych próbkach. Podejścia oparte na stosowaniu znakowanych standardów polegają na wprowadzeniu różnych znaczników masowych, które zmieniają masę białka lub peptydu. Następnie próbki łączy się w jedną, a porównanie ilościowe przeprowadza się porównując względną intensywność jonów peptydowych (z analizy MS) lub jonów produktu peptydowego (z eksperymentów MS/MS). Znaczniki te można wprowadzać metabolicznie, chemicznie lub enzymatycznie na poziomie białka lub peptydu podczas przygotowywania próbki. Znakowanie metaboliczne obejmuje trwałe przyłączenie izotopów do aminokwasów w hodowli komórkowej (SILAC) lub organizmów w pożywkach ze znakowanymi atomami: ^2H , ^{15}N , ^{13}C i ^{18}O . Próbkę z różnych warunków hodowlanych można analizować jednocześnie oszczędzając czas i odczynniki. W znakowaniu chemicznym, izotopy są wprowadzane do białek lub strawionych peptydów w wyniku reakcji chemicznej. Na przykład przy użyciu znaczników różnicujących próbki, poprzez przyłączenie izotopów do różnych fragmentów białka (isotope-coded affinity tags, ICAT), znacznik wiąże się z resztami cysteiny zawierającej w łańcuchu atomy węgla "ciężkiego" (^{13}C) lub "lekkiego" (^{12}C). Izobaryczne znaczniki masy, takie jak TMT (tandem mass tag) czy ITRAQ (isobaric tags for relative and absolute quantification) nie wpływają na masy powstałych po trypsynizacji peptydów, a do ich identyfikacji dochodzi wyłącznie za pomocą tandemowej spektrometrii mas, co pozwala na ilościowe określenie względnej zawartości białek z maksymalnie ośmiu próbek. Natomiast znakowanie enzymatyczne polega wykorzystaniu przyłączenia ^{18}O przy karboksylowym końcu peptydów podczas proteolitycznej degradacji. Taka modyfikacja w C-końcowej grupie karboksylowej fragmentów proteolitycznych prowadzi do zastąpienia dwóch atomów ^{16}O dwoma atomami ^{18}O podczas enzymatycznej degradacji w obecności H_2^{18}O . Peptydy z oznakowanych próbek można wówczas łatwo zidentyfikować w widmach MS przez obecność zmiany masy o 4Da w stosunku do peptydów trawionych w obecności H_2^{16}O .

Proteomiczne oznaczenia ilościowe w ujęciu bezwzględny zazwyczaj polegają na stosowaniu peptydów o stabilnych izotopach jako standardów wewnętrznych. Ocenę ilościową uzyskuje się przez dodanie znanej ilości takiego standardu do próby przed eksperymentem LC-MS. Peptydy stosowane jako standardy mają taki sam czas retencji jak peptydy uzyskane po trawieniu trypsyną, dlatego możemy pozwolić sobie na odczytanie ich poziomu stosując krzywą standardową. Alternatywnym podejściem jest użycie tzw. sztucznych białek wykonanych z syntetycznych peptydów (QconCAT) lub stosowanie standardów białkowych do bezwzględnej ilościowej oceny (PSAQ). W przypadku QconCAT każdy peptyd będący standardem wewnętrznym reprezentuje konkretne białko, podczas gdy w przypadku PSAQ stosuje się białko, które jest znakowane izotopowo i odpowiada rekombinowanemu analogowi białka, które ma być oznaczone ilościowo. Jednak obie te

metody są bardzo kosztowne, a ich wykorzystanie ogranicza się do niewielkiej liczby eksperymentów.

Literatura:

J. Lovric. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*. ISBN: 978-0-470-03524-5, Chapters 4.4 and 5.3, 2011.

2.5. Identyfikacja PTM

Post-translacyjne modyfikacje białek (PTM) to najczęściej modyfikacje chemiczne łańcucha białkowego odpowiedzialne za zmianę wielkości, składu, funkcji i lokalizacji białek. Jak dotąd zidentyfikowano ponad 900 różnych PTM i włączono je do bazy danych UNIMOD (www.unimod.org). Do najczęstszych modyfikacji post-transakcyjnych zalicza się fosforylację, acetylację, utlenianie, alkilowanie, metylowanie oraz tworzenie mostków disiarczkowych (tabela 4.1.). Spektrometria mas jest obecnie najlepszym podejściem do sekwencjonowania peptydów i identyfikowania PTM, umożliwiając przy tym analizę dużej liczby próbek.

Tabela 4.1. Lista najczęstszych modyfikacji post-translacyjnych

PTM	Modyfikowana reszta aminokwasowa
Glikozylacja atomu azotu	Asparagina i lizyna
Glikozylacja atomu tlenu	Lizyna, prolina, seryna, treonina i tyrozyna
Glikozylacja atomu węgla	Tryptofan
Fosforylacja	Seryna, treonina, tyrozyna, asparaginian, histydyna i cysteina
Acetylacja	N-końcowy fragment peptydów oraz lizyna i cysteina
Amidacja	C-końcowy fragment dojrzałego peptydu po usunięciu glicyny
Hydroksylacja	Asparagina, asparaginian, prolina i lizyna
Metylacja	N-końcowy fragment fenyloalaniny, C-końcowy fragment cysteiny, łańcuch boczny lizyny, argininy, histydyny i asparaginy

PTM zazwyczaj identyfikuje się za pomocą spektrometrii mas, obserwując zmianę mas peptydów powstałych po trypsynizacji (tabela 4.2.). Najczęściej zmiana masy jest wynikiem modyfikacji chemicznej zachodzącej w łańcuchu bocznym reszt aminokwasowych. Jednak dla jednoznacznej identyfikacji miejsc PTM konieczne są eksperymenty tandemowej spektrometrii mas (MS/MS). Jednakże zmodyfikowane peptydy mogą fragmentować w inny sposób niż niezmodyfikowane cząsteczki, dlatego identyfikacja miejsca modyfikacji jest utrudniona bądź niemożliwa z powodu utraty danej modyfikacji podczas fragmentacji. Większość post-translacyjnych modyfikacji występuje na bardzo niskich poziomach w stosunku do ilości białek w próbce, a dodatkowo podczas analizy MS/MS są one bardzo nietrwałe. Ponadto, wiele z tych modyfikacji ma charakter hydrofilowy, co może powodować problemy z oczyszczaniem białek. Ich obecność może także wpływać na efektywność proteolizy, generując niespecyficzne lub zbyt długie łańcuchy peptydowe. Niektóre PTM zmniejszają skuteczność jonizacji, przez co trudno je zarejestrować podczas analizy MS. Z tych względów warto rozważyć frakcjonowanie białek w próbkę. Biochemiczna separacja

organeli komórkowych, kompleksów białkowych lub poszczególnych białek jest użytecznym podejściem do analizy modyfikacji białek, ponieważ zmniejsza złożoność próbki białkowej.

Podczas analiz PTM istotnym jest także izolowanie zmodyfikowanych białek i/lub peptydów w celu ograniczenia złożoności próbki, a tym samym zwiększenia liczby zidentyfikowanych PTM. W tym celu zaleca się stosowanie np. immunoprecypitacji białek zawierających interesujące nas PTM. Zatężanie PTM może odbywać się na poziomie białka (np. przy użyciu przeciwciał, czy chromatografii powinowactwa [IMAC w stosunku do fosfopeptydów] lub immobilizowanych lektyn) lub na poziomie peptydu (np. przy użyciu przeciwciał [AcR, MeK, pTyr], chromatografii IMAC, immobilizowanych lektyn lub TiO₂ w stosunku do fosfopeptydów).

Tabela 4.2. Zmiany w masach peptydów wywołane poprzez PTM

PTM	Δ Monoizotopowej masy (Da)
Hydroksylacja	15.9949
Fosforylacja	
pTyr	79.9663
pSer,	79.9663
pThr	79.9663
Acetylacja	42.0105
Metylacja	14.0156
Sulfonylacja (sTyr)	79.9568
Deamidacja	0.9840
Acylacja	
Farnesyl	204.1878
Myristoyl	210.1983
Palmitoyl	238.2296
Glikozylacja	
Atom azotu	>800
Atom tlenu	203, >800
Mostki disiarczkowe	-2.0157
Nitrowanie tyrozyny	44.9850

Literatura:

A.M.N. Silva, R. Vitorino, M.R.M. Domingues, C.M. Spickett, P. Domingues. Post-translational Modifications and Mass Spectrometry Detection. *Free Radic. Biol. Med.*, 2013, 65:925-941.

J. Lovric. Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis. ISBN: 978-0-470-03524-5, Chapter 5.4, 2011.

2.6. Identyfikacja wewnątrzkomórkowej lokalizacji białek

Pojęcie proteomiki organelli komórkowych powstało w celu stworzenia metod analizy funkcji białek w różnych organellach. Jednakże wyjaśnienie wewnątrzkomórkowej lokalizacji białek w różnych warunkach jest dużym wyzwaniem ze względu na dynamiczny charakter lokalizacji białka. Dotychczas rozmieszczenia białek w obrębie komórki było mapowane przy użyciu tradycyjnych metod, takich jak mikroskopia immunofluorescencyjna, czy frakcjonowanie biochemiczne. Niemniej jednak przystosowano narzędzia proteomiki wielkoskalowej do identyfikowania białek z całych organelli, a także do przyporządkowania określonego białka do lokalizacji wewnątrzkomórkowej. Standardowe podejście do proteomiki organelli opiera się na frakcjonowaniu i zateżaniu otrzymanych białek zazwyczaj przy użyciu wirowania w gradiencie gęstości lub z użyciem zestawów komercyjnych. Organelle komórkowe mogą być również izolowane poprzez immunoprecypitację specyficznymi przeciwciałami przeciwko epitopom prezentowanym na ich powierzchni. Białka z uzyskanych różnych frakcji organelli są następnie identyfikowane i oznaczone ilościowo przy użyciu omówionych wcześniej metod proteomicznych. Chociaż proteomika organelli może dostarczyć cennych informacji na temat profilu białek w specyficznych organellach komórki, równocześnie istotnym jest również badanie lokalizacji białek w kontekście całej komórki, aby uzyskać pełny obraz profilu proteomicznego.

Literatura:

R. Drissi, M.-L. Dubois, F.-M. Boisvert, Proteomics methods for subcellular proteome analysis. *FEBS J.*, 2013, 280(22):5626-5634.

2.7. Identyfikacja interakcji między białkami

Rola pojedynczego białka w komórce zależy głównie od jego interakcji z innymi cząsteczkami i tworzenia wielobiałkowych kompleksów. Takie kompleksy są bardzo dynamiczne, a ich skład zmienia się wraz z upływem czasu i stanu komórek. Do badania interakcji białek wykorzystywane są trzy główne podejścia, tzw.: „affinity pull-down”, „proximity labeling” oraz „protein correlation profiling”. Affinity pull-down wykorzystuje specyficzne przeciwciała do izolowania wybranego białka wraz z towarzyszącymi mu białkami, które są następnie identyfikowane i ilościowo oznaczone za pomocą spektrometrii mas. Proximity labeling to technologia oparta na kowalencyjnym przyłączeniu biotyny do interesującego nas białka, a następnie samoistnego przeniesienia takiego znakowania na cząsteczki będące w bezpośrednim kontakcie z tym białkiem. Przyłączenie biotyny do białek identyfikuje się stosując tandemową spektrometrię mas. Natomiast protein correlation profiling wykorzystuje różne metody separacji (np. wirowanie w gradiencie gęstości lub chromatografia natywna) i zakłada, że współdziałające białka zostaną wyizolowane w tej samej frakcji.

Interakcje białko-białko w sygnalizacji komórkowej są często zależne od specyficznych oddziałujących ze sobą sekwencji aminokwasowych. Jednakże, pomimo istotnego znaczenia tych interakcji w sygnalizacji komórkowej identyfikacja wiążących się ze sobą peptydów jest nadal bardzo złożonym zadaniem z wykorzystaniem narzędzi bioinformatycznych, takich jak np. STRING (<http://string-db.org>).

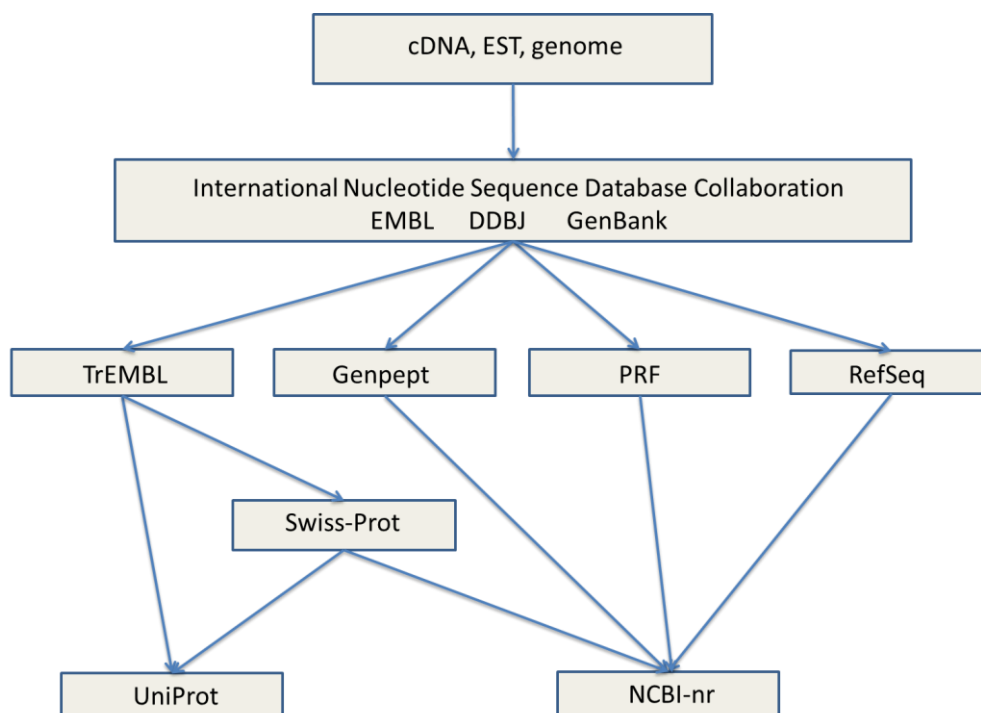
Literatura:

Syafrizayanti, C. Betzen, J.D. Hoheisel, D. Kastelic. Methods for analyzing and quantifying protein-protein interaction. *Expert. Rev. Proteomics*, 2014(2), 11(1):107-20.

3. Przetwarzanie danych w celu identyfikacji białek

3.1. Bazy danych zawierające sekwencje białkowe

Jak wspomniano wcześniej, bazy danych sekwencji białkowych są kluczowe w badaniach profili proteomicznych z zastosowaniem spektrometrii mas. Obecnie głównym źródłem danych o sekwencji białkowej są przepisane z sekwencji nukleotydowych dane zebrane w GenBank (<https://www.ncbi.nlm.nih.gov/genbank>), EMBL (<http://www.ensembl.org/index.html>) lub bazach danych DDBJ (<http://www.ddbj.nig.ac.jp>). Wymienione bazy danych są zintegrowane z International Nucleotide Sequence Database Collaboration (<http://www.insdc.org>), gdzie gromadzone są wszystkie znane sekwencje nukleotydów. Również baza danych NCBI (Entrez Proteins, <http://www.ncbi.nlm.nih.gov/entrez>) zawiera szeroki zestaw białek. Baza ta wykorzystuje informacje z Protein Research Foundation (PRF), Genpept (zawierającej sekwencje nie poddane recenzji), RefSeq (zawierającej opisane i sprawdzone sekwencje białkowe) oraz ze Swiss-Prot (zawierającej opisane i sprawdzone sekwencje białkowe) (Rycina 4.6.). Jednakże, ponieważ baza ta przypisuje każdą sekwencję białka z unikalną liczbą identyfikującą gen (gi), jest bardzo duża i trudna w przeszukiwaniu. Baza danych UniProt Knowledgebase (UniProtKB) zawiera obszerne informacje o zaktualizowanych białkach (w tym funkcję, klasyfikację oraz literaturę) i podzielona jest na dwie różne bazy danych. UniProtKB/Swiss-Prot, która jest dokładnie udokumentowana i sprawdzona oraz UniProtKB/TrEMBL, która zawiera niezrecenzowane sekwencje białkowe.



Rycina 4.6. Najczęściej stosowane białkowe bazy danych oraz ich wzajemne zależności.

Literatura:

C. Brooksbank, M.T. Bergman, R. Apweiler, E. Birney, J. Thornton. The European Bioinformatics Institute's data resources 2014. *Nucleic Acids Res.*, 2014, 42(D1):D18-D25.

3.2. Przeszukiwanie białkowych baz danych

Do przeszukiwania wspomnianych powyżej baz danych konieczne jest odpowiednie oprogramowanie opierające się na dopasowywaniu do istniejących w bazie sekwencji peptydowych otrzymanych podczas analizy mas peptydów (podejście PMF) lub sekwencji peptydowych (podejście PFF). Istnieje kilka narzędzi bioinformatycznych, które zostały do tego zaprojektowane – mogą to być wyszukiwarki komercyjne takie jak Mascot (Matrix Science), Proteome Discoverer (Thermo), ProteinPilot (Sciex), ProteinLynx (Waters), czy ProteinScape (Bruker), a także wyszukiwarki z wolnym dostępem: ProteinProspector, Andromeda, OMSSA, SEQUEST, X! Tandem, Amanda, czy Mascot (wersja ograniczona). Prawidłowa identyfikacja białka zależy od wielu czynników, w tym algorytmu wykorzystywanego przez wyszukiwarki i czynników eksperymentalnych mających wpływ na zawartość informacji w danych MS. Zatem różne wyszukiwarki niekoniecznie identyfikują te same białka w próbce. Niemniej jednak istnieje kilka programów, na przykład SearchGUI, które pozwalają na łączenie wyników z wielu wyszukiwarek. W proteomice uzyskuje się ogromne ilości danych, a zatem takie przeszukiwanie może trwać nawet kilka dni w odniesieniu do jednego eksperymentu, zwłaszcza gdy wykonywane jest równoległe wyszukiwanie PTM.

Literatura:

Z.F. Yuan, S. Lin, R.C. Molden, B.A. García. Evaluation of proteomic search engines for the analysis of histone modifications. *J. Proteome Res.*, 2014, 13:4470-4478.

3.3. Analiza danych

Głównym wyzwaniem dla analiz proteomicznych z wykorzystaniem spektrometrii mas jest interpretacja otrzymanych danych. Obejmuje to integrację informacji taksonomicznych i funkcjonalnych z próbek zawierających kilkaset białek w sposób umożliwiający interpretowanie eksperymentów. W tym celu, gdy identyfikacja i oznaczenie ilościowe białek zostaną zakończone, konieczne jest przejście do analizy funkcjonalnej odpowiednich białek różnicujących badane próbki. Istnieje kilka bezpłatnych narzędzi bioinformatycznych, które mogą pomóc w tym zadaniu, takie jak OpenMS Proteomics Pipeline, Trans-proteomic pipeline, czy PeptideShaker. Analiza takich danych obejmuje zazwyczaj identyfikację PTM, porównanie i biochemiczną interpretację danych ilościowych, identyfikację procesów biologicznych i oddziaływań białkowych. Pierwszym krokiem do funkcjonalnej interpretacji powstałej listy białek jest powiązanie białek ze związanymi z nimi pojęciami ontologicznymi genów. System klasyfikacji ontologicznej genów (<http://www.geneontology.org>) definiuje koncepcje lub klasy używane do opisu funkcji genu i klasyfikuje te funkcje w trzech różnych perspektywach: molekularną funkcję produktu genowego, organella, w których produkty genowe są aktywne i proces biologiczny, w którym produkty genowe są aktywne. System klasyfikacji PANTHER (<http://www.pantherdb.org/about.jsp>), będący częścią projektu referencyjnego GO Genome, jest ważnym narzędziem do klasyfikacji białek i ich genów, który dodatkowo ułatwia i przyspiesza wykonanie analizy danych.

STRING (<http://string-db.org>) i Cytoscape (<http://www.cytoscape.org>) są dwoma ważnymi narzędziami bioinformatyki, które także można wykorzystać do analizy interakcji białko-białko i projektowania sieci białkowych. Dane te można skategoryzować, co pomaga w interpretowaniu wyników eksperymentu. Także, baza danych Uniprot

(<http://www.uniprot.org>) może być używana jako źródło wysokiej jakości informacji o funkcjonowaniu wybranego białka.

Literatura:

S.W. Haga, H.F. Wu. Overview of software options for processing, analysis and interpretation of mass spectrometric proteomic data. *J. Mass Spectrom.*, 2014, 49:959-969.

4. Część praktyczna

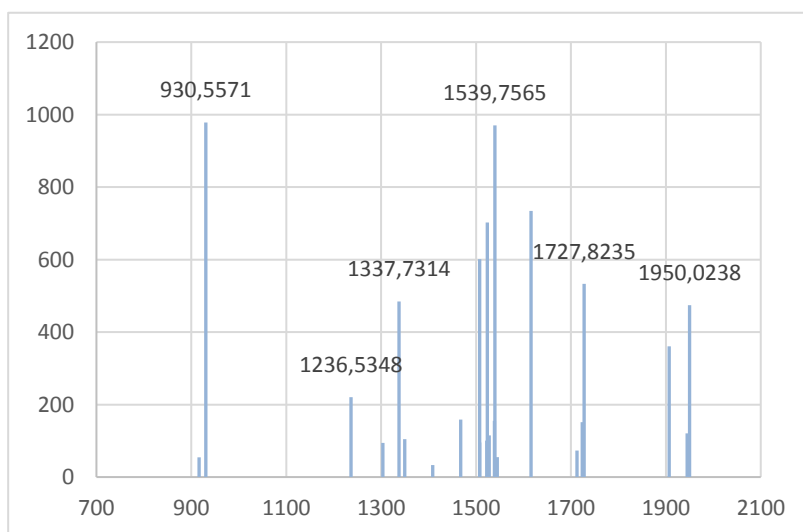
4.1 Identyfikacja białek w komórkach z użyciem podejścia PMF i PFF za pomocą platformy MASCOT

Eksperyment

- 1) próbkę z linii komórkowej HEK 293 (Human Embryo Kidney) zawierającą 120 µg białka nałożono na paski IPG i przeprowadzono analizę 2-DE;
- 2) Po wybarwieniu i wycięciu z żelu fragmentów zawierających wybrane białko, próbki poddano redukcji i alkilacji z użyciem ditiotretolu (DTT) oraz jodoacetamidu (IAA), a następnie przeprowadzono całkowitą trypsynizację;
- 3) Otrzymane peptydy liofilizowano, a następnie zawieszono w 10 µl 50% roztworu acetonitrylu/0,1% kwasu mrówkowego. Próbki wymieszano z nasyconym roztworem matrycy (kwas α -cyjano-4-hydroksycynamonowy) w stosunku 1: 1 I 0,35 µl nakropiono na płytkę do testów MALDI;
- 4) Przeprowadzono analizę z użyciem spektrometru mas MALDI-TOF/TOF (4700 Proteomics Analyzer, Applied Biosystems, Foster City, CA, USA) w trybie odczytywania jonów dodatnich.
- 5) Dla każdej próbki wybrano dwa najintensywniejsze piki, pomijając sygnały od matrycy, trypsyny, czy żelu poliakrylamidowego, w celu przeprowadzenia eksperymentu MS/MS.

Przykładowe spektrum MS:

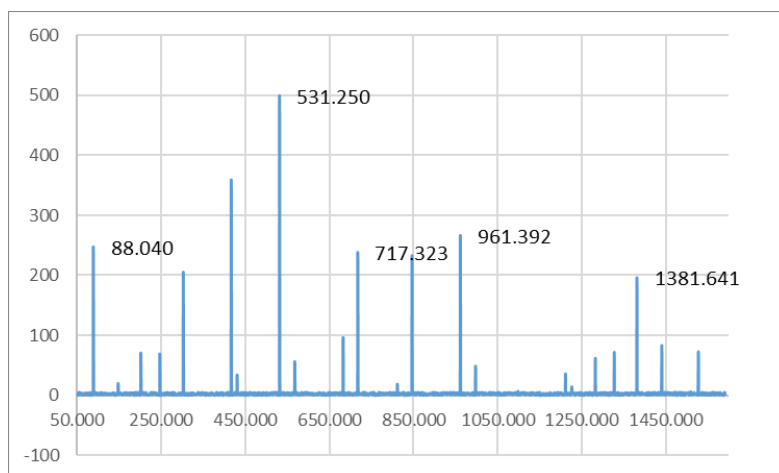
m/z
916.4055
30.5571
1236.5348
1303.6696
1337.7314
1349.7409
1408.7048
1467.6749
1507.6632
1507.6686
1522.6397
1523.7924
1527.7532
1538.7780



1539.7565
1544.7386
1615.9085
1712.7815
1723.8684
1727.8235
1907.0442
1944.9352
1950.0238

MS/MS jonu o $m/z = 1527.7532$

MS/MS	AR%
88.0399	246.9417
201.1301	70.34612
302.1732	204.9268
416.2143	358.6303
531.2498	498.7431
717.3232	237.9741
846.3714	232.3693
961.3924	266.1303
1098.4522	6.643215
1211.5352	35.65641
1282.5703	61.62781
1381.6409	195.4986
1509.7398	5.945193
1527.7472	72.32519
1440.7182	82.95763
1327.6311	71.58412
1226.5858	14.23603
531.2480	7.539087
997.5173	48.5566
811.4397	18.36804
682.3962	96.0226
567.3680	56.00374
430.3032	33.83856
317.2258	1.29883
246.1897	68.81332
147.1151	19.68073



a) Identyfikacja białka z użyciem podejścia PMF

1) Przejdź do strony internetowej MASCOT

(http://www.matrixscience.com/search_form_select.html)

i wybierz opcję wyszukiwania peptydowych odcisków palców „Peptide Mass Fingerprint”

(http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF).

2) Uzupełnij poszczególne parametry wyszukiwania. Do wprowadzania danych można kopiować i wklejać wartości m/z lub załadować plik danych Mascot MS.txt.

3) Powinieneś zwrócić uwagę na następujące kwestie:

- dlaczego zaznaczono Swissprot i powiązane bazy danych?
- dlaczego zaznaczono dwie „missed cleavages”?
- dlaczego dozwolone są dwie zmienne modyfikacje?
- dlaczego tolerancja wynosi 25 ppm?
- co znaczy wartość „decoy”?

MASCOT Peptide Mass Fingerprint

Your name Pedro **Email** p.domingues@ua.pt

Search title

Database(s) Vertebrates_EST
contaminants
cRAP
NCBIprot
SwissProt

Enzyme Trypsin

Allow up to 2 missed cleavages

Taxonomy Homo sapiens (human)

Fixed modifications --- none selected ---

Display all modifications

Variable modifications Oxidation (M)
Carbamidomethyl (C)

Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamyl (K)
Carbamyl (N-term)
Carboxymethyl (C)

Protein mass kDa **Peptide tol. ±** 25 ppm

Mass values MH⁺ M_r M-H⁻

Monoisotopic **Average**

Data file No file chosen

Query

Data input 916.4055
930.5571
1236.5348
1303.6696
1337.7314
1349.7409

Decoy **Report top** AUTO hits

4) Po zakończeniu wyszukiwania pojawiają się poniższe wyniki:

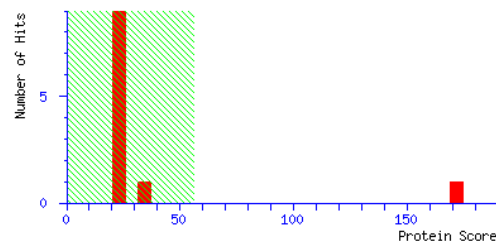
MATRIX SCIENCE Mascot Search Results

User : Pedro
Email : p.domingues@ua.pt
Search title :
Database 1 : contaminants 20090624 (262 sequences; 133770 residues)
Database 2 : SwissProt 2017_01 (553474 sequences; 198069095 residues)
Taxonomy 2 : Homo sapiens (human) (20172 sequences)
Timestamp : 16 Feb 2017 at 17:10:23 GMT
Top Score : 172 for 2::HS90A_HUMAN, Heat shock protein HSP 90-alpha 05=Homo sapiens GN=HSP90AA1 PE=1 SV=!

	Target	Decoy
Protein hits above identity threshold	1	0
Highest scoring protein hit	172	29

Mascot Score Histogram

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than 56 are significant ($p < 0.05$).



Concise Protein Summary Report

Format As [Help](#)

Significance threshold p< Max. number of hits

Preferred taxonomy

1. [2::HS90A_HUMAN](#) Mass: 84607 Score: 172 Expect: 1.3e-13 Matches: 18
Heat shock protein HSP 90-alpha 05=Homo sapiens GN=HSP90AA1 PE=1 SV=5
[2::HS90B_HUMAN](#) Mass: 83212 Score: 48 Expect: 0.36 Matches: 8
Heat shock protein HSP 90-beta 05=Homo sapiens GN=HSP90AB1 PE=1 SV=4

- 5) Przyjrzyj się otrzymanym wynikom i wybierz białko o najwyższym wyniku prawdopodobieństwa. Spójrz na dane „Protein View: HS90A_HUMAN”

b) Identyfikacja białka z użyciem podejścia PFF

- 1) Przejdź do strony internetowej MASCOT (http://www.matrixscience.com/search_form_select.html) i wybierz opcję wyszukiwania sekwencji peptydowych „MS/MS Ions Search”.
- 2) Wypełnij poszczególne parametry wyszukiwania. Do wprowadzania danych można załadować plik danych Mascot MSMS.txt. Uruchom wyszukiwanie (“start search”)

MASCOT MS/MS Ions Search

Your name	<input type="text" value="Pedro"/>	Email	<input type="text" value="p.domingues@ua.pt"/>
Search title	<input type="text"/>		
Database(s)	<input type="text" value="contaminants (AA)"/> <input type="text" value="SwissProt (AA)"/>	<input type="button" value=">"/> <input type="button" value="<"/>	<input type="text" value="Amino acid (AA)"/> <input type="text" value="cRAP"/> <input type="text" value="NCBIprot"/> <i>Nucleic acid (NA)</i> <input type="text" value="Environmental_EST"/> <input type="text" value="Fungi_EST"/> <input type="text" value="Human_EST"/> <input type="text" value="Invertebrates_EST"/> <input type="text" value="Mammals_EST"/> <input type="text" value="Mus_EST"/>
Taxonomy	<input type="text" value="..... Homo sapiens (human)"/>		
Enzyme	<input type="text" value="Trypsin"/>	Allow up to	<input type="text" value="2"/> missed cleavages
Quantitation	<input type="text" value="None"/>		
Fixed modifications	<input type="text" value="-- none selected --"/>	<input type="button" value=">"/> <input type="button" value="<"/>	<input type="text" value="Acetyl (K)"/> <input type="text" value="Acetyl (N-term)"/> <input type="text" value="Acetyl (Protein N-term)"/> <input type="text" value="Amidated (C-term)"/> <input type="text" value="Amidated (Protein C-term)"/> <input type="text" value="Ammonia-loss (N-term C)"/> <input type="text" value="Biotin (K)"/> <input type="text" value="Biotin (N-term)"/> <input type="text" value="Carbamidomethyl (C)"/> <input type="text" value="Carbamyl (K)"/> <input type="text" value="Carbamyl (N-term)"/>
	<input type="checkbox"/> Display all modifications		
Variable modifications	<input type="text" value="-- none selected --"/>		
Peptide tol. ±	<input type="text" value="25"/> ppm	# ¹³C	<input type="text" value="0"/>
MS/MS tol. ±	<input type="text" value="50"/> mmu		
Peptide charge	<input type="text" value="1+"/>		
Monoisotopic	<input checked="" type="radio"/> Average <input type="radio"/>		
Data file	<input type="button" value="Choose File"/> Mascot MSMS.txt		
Data format	<input type="text" value="Mascot generic"/>		
Precursor	<input type="text"/> m/z		
Instrument	<input type="text" value="Default"/>		
Error tolerant	<input type="checkbox"/>		
Decoy	<input checked="" type="checkbox"/>		
Report top	<input type="text" value="AUTO"/> hits		
<input type="button" value="Start Search ..."/>		<input type="button" value="Reset Form"/>	

- 3) Przyjrzyj się otrzymanym wynikom. Co oznaczają? Wybierz białko o najwyższym wyniku prawdopodobieństwa. Spójrz na dane „Protein View: HS90A_HUMAN”

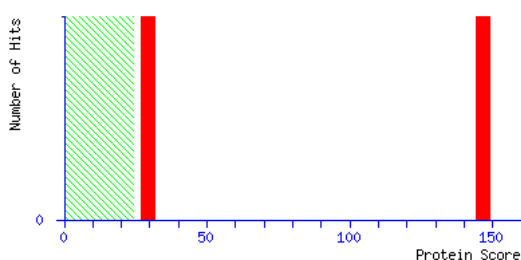
```

User          : test
Email         : p.domingues@ua.pt
Search title  : 10 pmol digest of Sample HS90A_HUMAN
MS data file  : Mascot MSMS.txt
Database 1    : contaminants 20090624 (262 sequences; 133770 residues)
Database 2    : SwissProt 2017_01 (553474 sequences; 198069095 residues)
Taxonomy 2    : Homo sapiens (human) (20172 sequences)
Timestamp     : 16 Feb 2017 at 17:36:09 GMT
Protein hits  : 2::HS90A\_HUMAN Heat shock protein HSP 90-alpha OS=Homo sapiens GN=HSP90AA1 PE=1 SV=5
               : 2::PTPRQ\_HUMAN Phosphatidylinositol phosphatase PTPRQ OS=Homo sapiens GN=PTPRQ PE=1 SV=2
    
```

	Target	Decoy	False discovery rate
Peptide matches above identity threshold	1	0	0.00 %
Peptide matches above homology or identity threshold	1	0	0.00 %

Mascot Score Histogram

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 24 indicate identity or extensive homology ($p < 0.05$). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As	Peptide Summary	Help	Help
Significance threshold p<	0.05	Max. number of hits	AUTO
Standard scoring	<input checked="" type="radio"/> MudPIT scoring <input type="radio"/>	Display non-significant matches	<input type="checkbox"/>
Show pop-ups	<input checked="" type="radio"/> Suppress pop-ups <input type="radio"/>	Sort unassigned	Decreasing Score
Preferred taxonomy	All entries	Require bold red	<input type="checkbox"/>

Select All Select None Search Selected Error tolerant

1. [2::HS90A_HUMAN](#) Mass: 84607 Score: 147 Matches: 1(1) Sequences: 1(1)
Heat shock protein HSP 90-alpha OS=Homo sapiens GN=HSP90AA1 PE=1 SV=5

4) Wybierz pozycję 1 w otrzymanych wynikach i spójrz na widok peptydu „Peptide View”. Spójrz na sposób fragmentacji peptydu. Dlaczego jony typu b są bardziej intensywne niż jony typu y?

4.2. Identyfikacja białek z eksperymentu LC-MS

Do tej części ćwiczeń należy pobrać następujące pliki:

- Dane z eksperymentu lub wstępnie przekształcone pliki
- ProteoWizard – ogólnodostępny program do konwersji oprogramowania w formacie MS. Jest niezbędny przy plikach typu RAW. Dodatkowo SearchGUI obejmuje tę aplikację. (<http://proteowizard.sourceforge.net/>)
- mMass (opcjonalnie) - Jest to program, który można wykorzystać do wizualizacji widm. Zawiera również kilka innych narzędzi, w tym narzędzia do proteomiki (<http://www.mmass.org/>)
- Dbtoolkit - jest to edytor danych w formacie FASTA. (<https://github.com/compomics/dbtoolkit>)

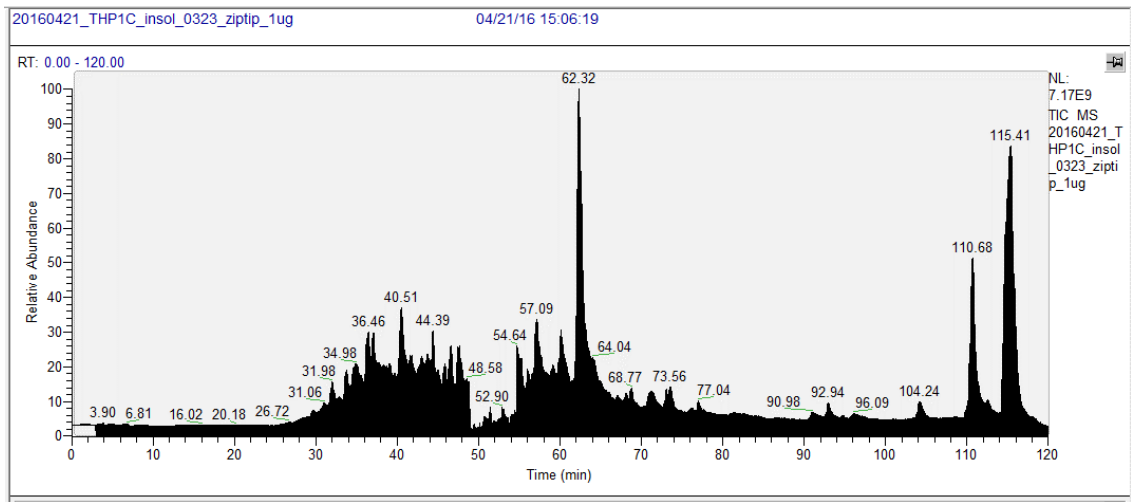
- e) SearchGUI - program do konfigurowania i uruchamiania wyszukiwarek identyfikacji białek. (<http://compomics.github.io/projects/searchgui.html>)
- f) Peptide shaker - jest to niezależny program do interpretacji wyników proteomicznych. (<http://compomics.github.io/projects/peptide-shaker.html>). Bardzo ważne jest również dokładne przeczytanie pełnego tutoriala dotyczącego do programu (<https://compomics.com/bioinformatics-for-proteomics/>).

Eksperyment:

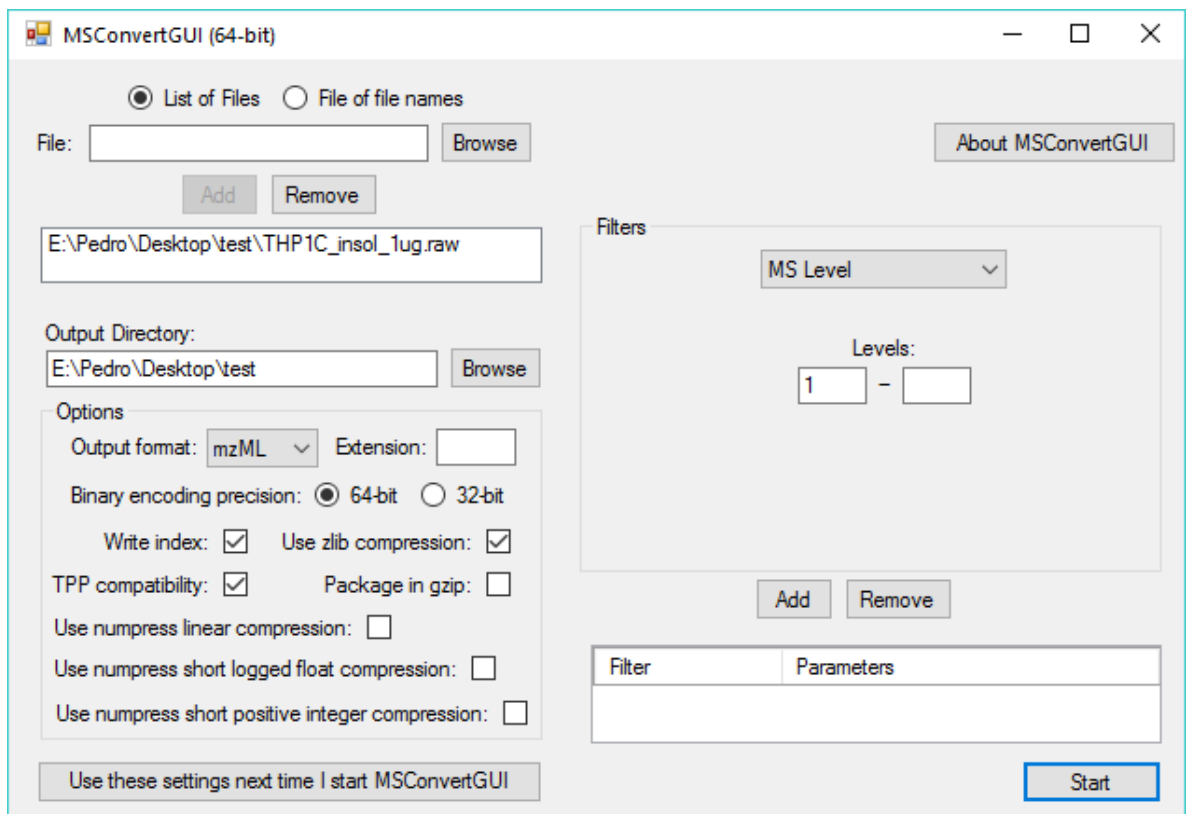
- 1) W celu szybkiej identyfikacji białek występujących w wysokich stężeniach w próbkach pochodzących z linii komórkowej THP 1, komórki te zawieszono w 500 μ l PBS i sonikowano przy 53 kHz w 37 ° C przez 30 minut. Następnie białka wytrącono dodając dwie objętości acetonu do zawiesiny lizatu komórkowego, inkubowano w -20°C przez 16 godzin i odwirowano. Wysuszone pelety zawierające wytrącone białka zawieszono ponownie w 30 μ l buforu do przechowywania próbki. Poziom wyekstrahowanych białek zmierzono za pomocą metody Lowry'ego.
- 2) Cysteinowe reszty aminokwasowe zostały zredukowane za pomocą DTT i alkirowane jodo-acetamidem, a próbki poddano następnie trawieniu „in-solution” z użyciem trypsyny. Otrzymane peptydy liofilizowano i ponownie zawieszano w roztworze 5% ACN/0,1% FA.
- 3) Próbkę zawierającą 250ng ekstraktu białkowego analizowano za pomocą QExactive Orbitrap, który połączono do chromatografu Ultimate 3000 z przepływem typu nano. Zastosowano kolumnę analityczną C18 (150 mm x 75 μ m I.D.). Spektrometr mas pracował w trybie „data dependent”, z wybraniem do eksperymentu MS/MS 10 najbardziej intensywnych jonów.

4.2.1. Zamiana otrzymanych danych na pliki mzML z użyciem MSConverter (ProteoWizard)

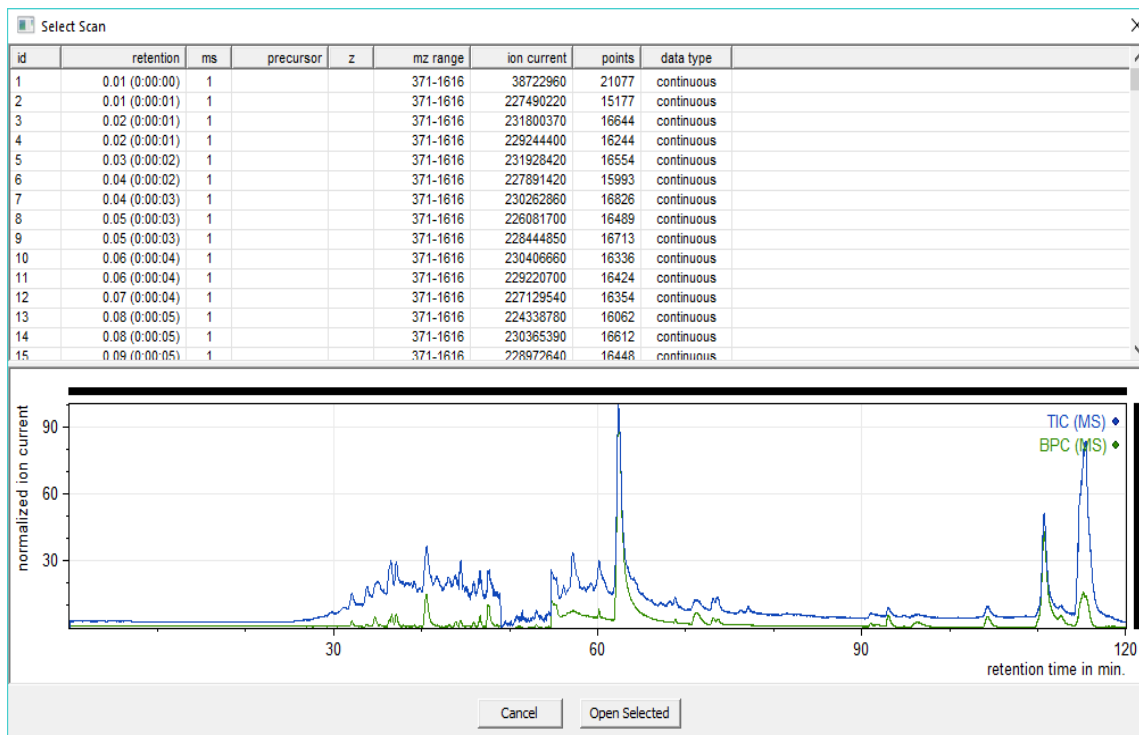
Użyj programu MSConverter w ProteoWizard, aby przekształcić dane HPC-MS otrzymane w wyniku analizy spektrometrem mas (dane w formacie RAW) do formatu, który może być odczytany przez SearchGUI (dane w formacie MGF). Jeśli chcesz obejrzeć dane w aplikacji mMass, możesz odczytać format pliku MGF, który jest obsługiwany przez SearchGUI, jednak najlepiej, jeśli konwertujesz dane do formatu mzML.



chromatogram w formacie RAW



Oprogramowanie MSConverter



Przykład danych w formacie mMass

4.2.2. Utworzenie bazy danych FASTA używając serwera Uniprot (SwissProt)

Aby utworzyć plik FASTA proteomu Homo sapiens, należy na stronie Uniprot wyszukać gatunek Homo Sapiens i pobrać recenzjonowaną wersję (Swiss-prot) pliku.

The screenshot shows the UniProt website interface. At the top, there is a search bar with 'Taxonomy' selected and 'human' entered. Below the search bar, there are navigation links: 'BLAST', 'Align', 'Retrieve/ID mapping', 'Peptide search', 'Help', and 'Contact'. The main heading is 'Taxonomy results' with a link to 'About Taxonomy'. Below this, there is a 'Filter by' section with a 'Download' button and a pagination control showing '1 to 25 of 8,220' items. A 'Repeat search in UniProtKB (2,393,722)' button is also present. The main content area shows a list of taxons, with 'Homo sapiens (Human)' selected. Below the taxon name, there is a taxonomic hierarchy: 'Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorhini > Catarrhini > Hominoidea > Homo'. At the bottom, there is a summary of proteomes: 'Proteomes (1) UniProtKB (156,851) - Reviewed (Swiss-Prot) (20,168) Unreviewed (TrEMBL) (136,683)'.

UniProtKB results

Filter by: Reviewed (20,168) Swiss-Prot

Popular organisms: Human (20,168)

Proteomes: UP000005640 (20,162)

View by: Results table

Entry	Entry	Gene names	Organism	Length
P20848	A1ATR_HUMAN	SERPINA2 ARG5, ATR, PIL, SERPINA2P	Homo sapiens (Human)	420
Q9UNA3	A4GCT_HUMAN	alpha-1,4-N-acetylglucosaminyltrans...	Homo sapiens (Human)	340
Q96GX2	A7L3B_HUMAN	Putative ataxin-7-like protein 3B	Homo sapiens (Human)	97
Q8N5Z0	AADAT_HUMAN	Kynurenine/alpha-aminoadipate amino...	Homo sapiens	425

Można oglądać i zmieniać pliki typu FASTA używając programu Dbtoolkit (data Base Processing Tool).

Database Processing Tool

File selection

Database file: E:\Pedro\Desktop\uniprot-organism%3A9606+reviewed%3Ayes.fasta

Preview pane (2 FASTA entries)

```
>sp|P20848|A1ATR_HUMAN Putative alpha-1-antitrypsin-related protein OS=Homo sapiens GN=SERPINA2 PE=1 SV=1
MPFSVSWGVLVLLAGLCCLVPSLVEDPQGDAQKTDTSHHDQGDWEDLACQKISYNVTDLAFDLYKSWLIYHNQHVLTPTSVAMAFRMLSLGKADTRTEIILEGLNVNLTETPEAKIHECF
QQVLQALSRPDTRLQLTTGSSLFVNKSMKLVDTFLEDTKKLYHSEASSINFRDTEEAKEQINNVYEKRTGRKVVLDLKVHLKDDTSLALVDYISFHGKWKDKFKAERIMVEGFHVDDKTIIRV
PMINHLGRFDIHRDRELSWVLAQHYVGNATAFFILPDPKMMQWLEEKLTYSHLENIQRAFDIRSINLHFKLSISGTYKLRVPRNLGITKIFSNADLSGVSQEAPLKLKAVHVAVLTI
DEKGTATGAPHLEEKAWSKYQTMFNRPFLVIEKEYIINFPLFIGKVVNPTQK
```

```
>sp|Q9UNA3|A4GCT_HUMAN Alpha-1,4-N-acetylglucosaminyltransferase OS=Homo sapiens GN=A4GNT PE=2 SV=1
MRKELQLSLSVTLVLCGFLYQFTLKSSCLFCLPSFKSHQGLEALLSHRRGIVFLETSERMEPPHLVSCSVEAAKIYPEWVVFVFMKGLTDTSPMPSNSTYPAFSFLSAIDNVLFPLDMK
RLEEDTPLFSWYNQINASAERNWLHISDASRLAIWKYGGIYMDTDVVISIRPIPEENFLAAQASRYSSNGIFGFLPHHPFLWECMENFVEHYNSAIWGNQGPMLMTRMLRVVWCKLEDPFQEV
SDLRCLNISFLHPQRFPYPISYREWRRYYEVDTEPFSFNVSVALHLWHRMNIQEGRAVIRGNSNLTVENLYRKHCPRTRYRDLIKGPEGSVTGELGPGNK
```

Status panel

Status:

21-02-2017 (16:51:23): No file loaded.

21-02-2017 (16:51:35): Processing incoming file (E:\Pedro\Desktop\uniprot-organism%3A9606+reviewed%3Ayes.fasta)

21-02-2017 (16:51:35): FASTA formatted file 'E:\Pedro\Desktop\uniprot-organism%3A9606+reviewed%3Ayes.fasta'

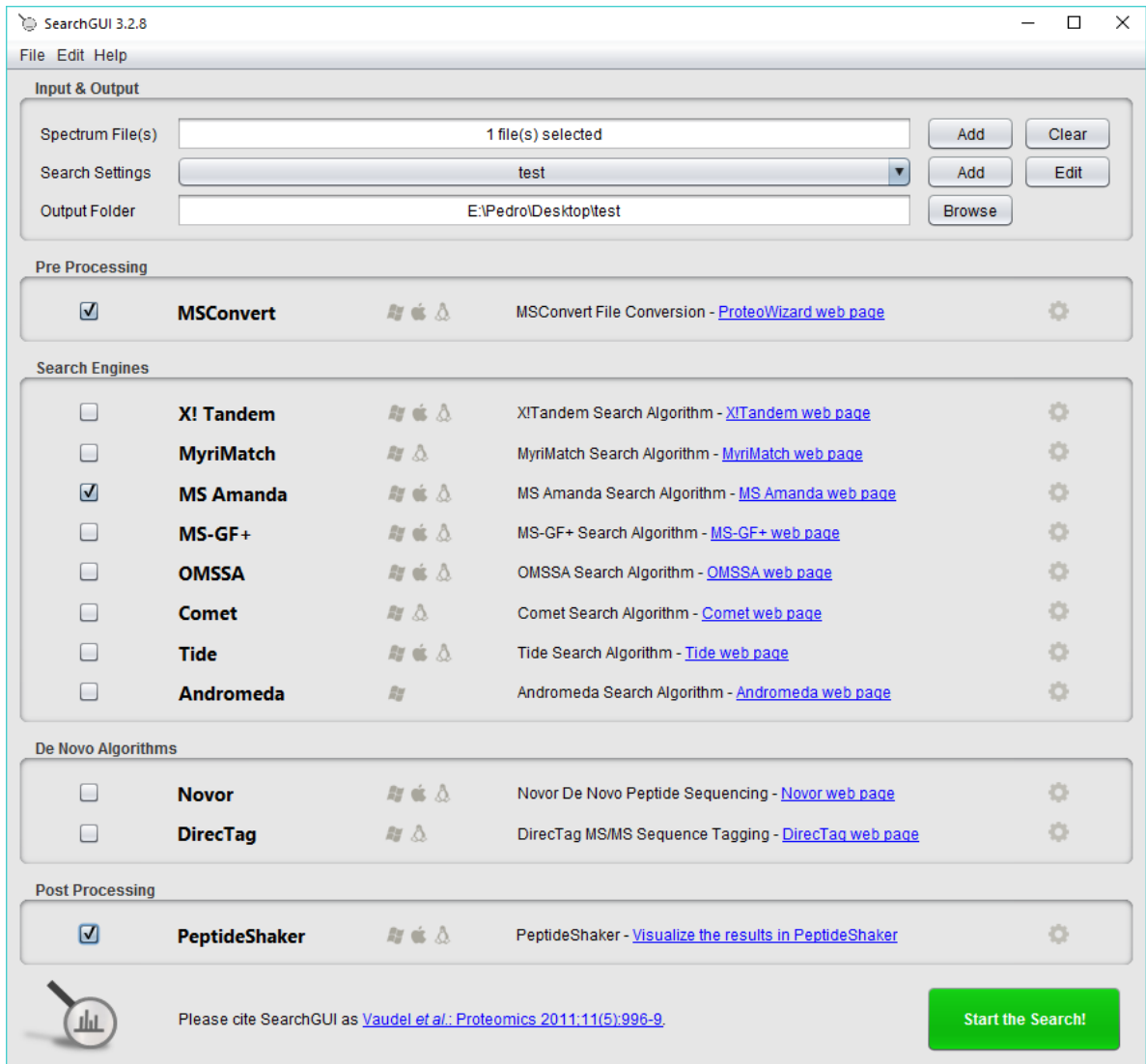
Errors:

21-02-2017 (16:51:23): None.

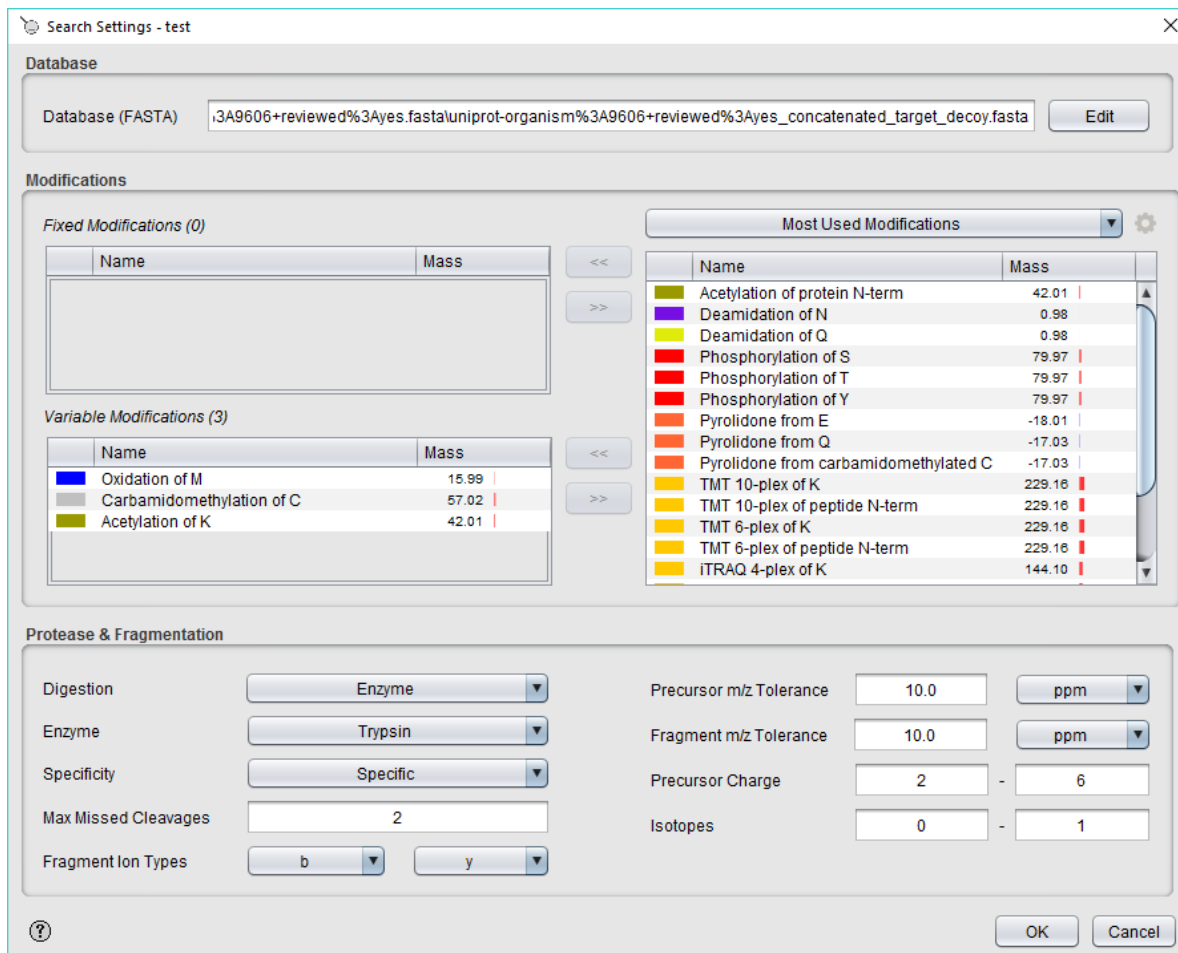
21-02-2017 (16:51:35): None

4.2.3. Wyszukiwarka: SearchGUI

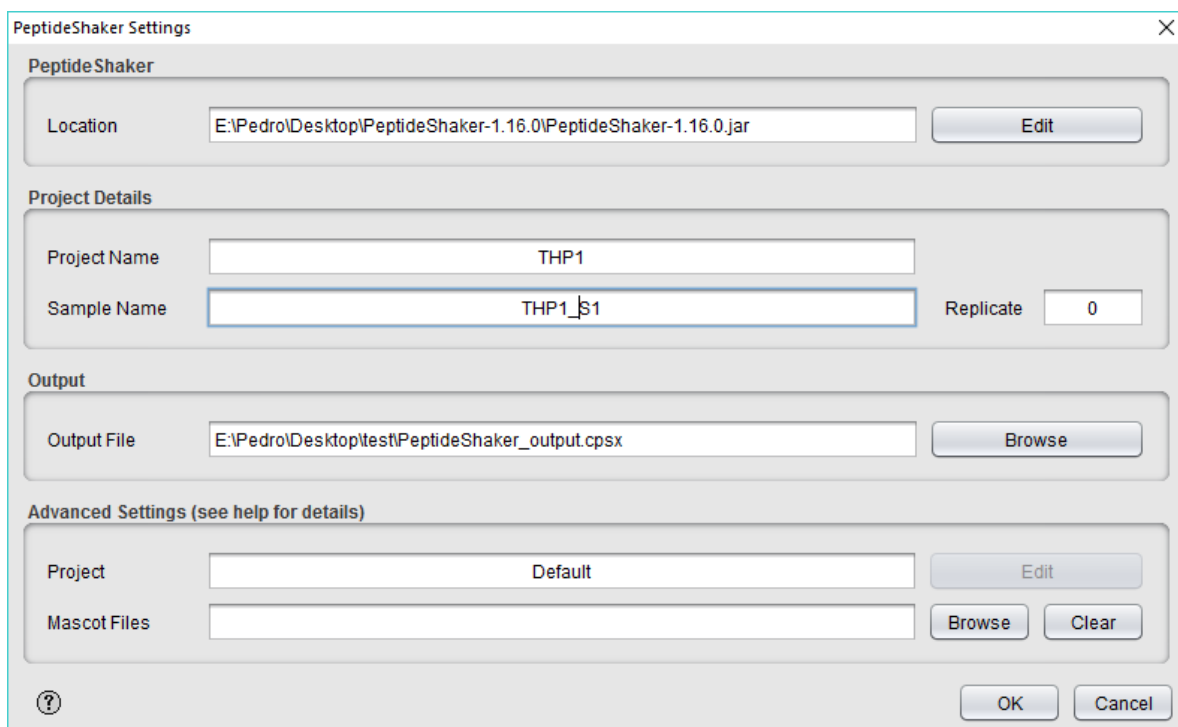
Uruchom program searchGUI. W powyższym przykładzie wybrany plik był konwertowany, więc tutaj również wybrano konwersję MS. W ustawieniach wyszukiwania wybierz edycję i wypełnij formularz, jak pokazano poniżej. Zastanów się dlaczego wybrano „variable modification”.



W opcjach wyszukiwania wybierz bazę danych FASTA.

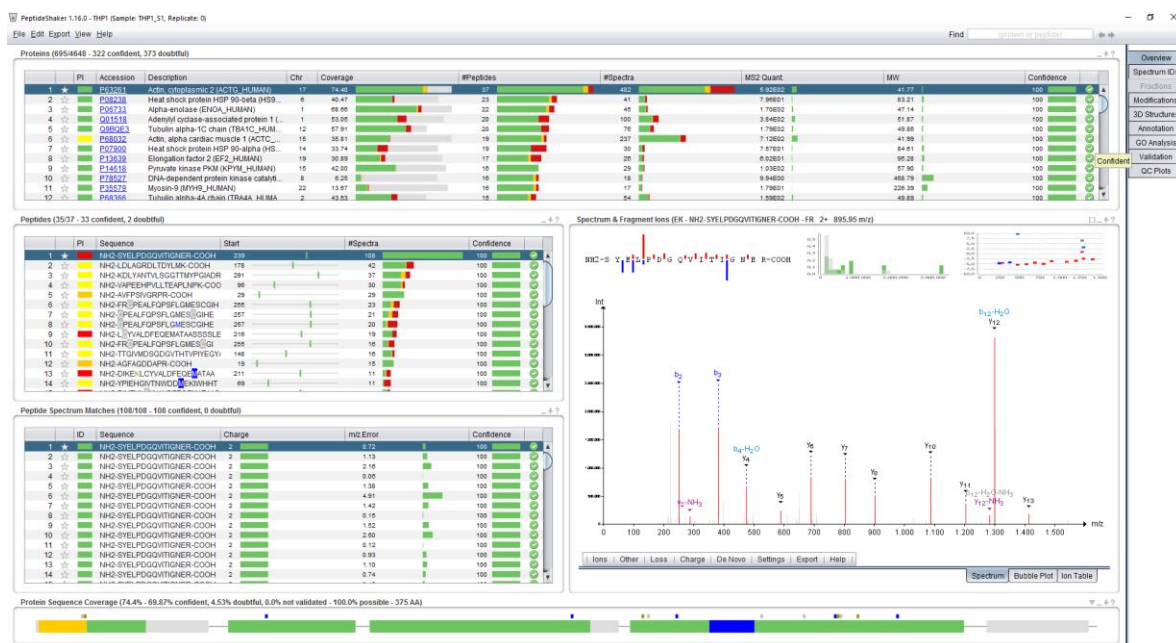


W opcjach peptideshaker ustaw wybrane wartości.



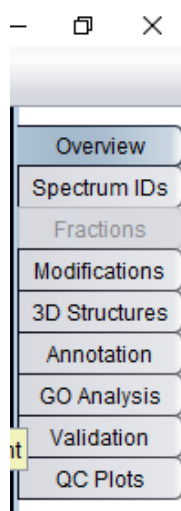
4.2.4. Generowanie i ocena wyników: użycie PeptideShaker do wizualizacji peptydów i białek oraz do walidacji wyników. Analiza PTM

Po przeprowadzeniu wyszukiwania przez searchGUI (~ 3 minut w przypadku Intel I7-6700K z 16 MB RAM), wyniki zostaną otworzone przez platformę Peptideshaker. Można tu zobaczyć ilość zidentyfikowanych białek oraz jaka część z nich została sklasyfikowana jako „doubtful”. Wyjaśnij dlaczego oraz co to oznacza. Spróbuj także zobaczyć informacje o peptydach zidentyfikowanych dla każdego białka oraz widma z opisanym schematem fragmentacji.



4.2.5. Analiza danych: informacja o białkach, analiza szlaków i ontologia genów

Zapoznaj się z zaawansowanymi opcjami analizy danych platformy peptideshaker otwierając kartę modyfikacji (modifications tab) i kartę analizy (GO analysis).



W zakładce adnotacji możesz dodawać komentarze do każdego białka. Jeśli jednak chcesz opisać wiele białek, musisz wyeksportować wyniki (default protein report).

PeptideShaker 1.16.0 - THP1 (Sample: THP1_S1, Replicate: 0)

File Edit Export View Help Find (protein or peptide)

Basic Protein Annotation

Accession: P63261

Description: Actin, cytoplasmic 2 (ACTG_HUMAN)

Gene Name: ACTG1 Chromosome: 17

Taxonomy: Homo sapiens

Database: UniProt

Protein Annotation - Help

Single Protein
To access the annotations for the currently selected protein, simply click the button corresponding to the wanted resource.

Multiple Proteins
To get the list of all validated proteins in your project click [here](#).
Advanced export options: *Export > Identification Features*.

To query using multiple proteins, click the [web](#) link next to the resource and follow the instructions provided at the resource web page.

UniProt - protein knowledgebase
[Search UniProt](#) High-quality protein sequence and functional information. [web](#)

Reactome - pathway database
[Search Reactome](#) Manually curated and peer-reviewed pathway database. [web](#)

neXtProt - human protein knowledgebase
[Search neXtProt](#) High-quality human protein information and annotation. [web](#)

DAVID - functional annotation
[Search DAVID](#) Database for Annotation, Visualization and Integrated Discovery. [web](#)

STRING - protein interaction
[Search STRING](#) Known and Predicted Protein-Protein Interactions. [web](#)

IntAct - protein interaction
[Search IntAct](#) Analysis tools for protein interaction data. [web](#)

QuickGO - gene ontology terms and annotations
[Search QuickGO](#) Web-based browser for Gene Ontology terms. [web](#)

InterPro - predictive protein signatures
[Search InterPro](#) Integrated database of predictive protein signatures. [web](#)

DASTy - protein sequence features
[Search DASTy](#) Web client for visualizing protein sequence feature information. [web](#)

PDB - protein data bank
[Search PDB](#) Biological macromolecular resource. [web](#)

[PICR - Protein Identifier Cross-Reference Service](#)

Overview
Spectrum IDs
Fractions
Modifications
3D Structures
Annotation
GO Analysis
Validation
QC Plots