

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta



**Nástroj pro generování elementárního
složení z dat hmotnostních spektrometrů s
vysokým rozlišením**

Bakalářská práce

Autor: Marie Doležalová

Vedoucí práce: Ing. František Drdák, CSc.

České Budějovice 2015

Jihočeské univerza v Českých Budějovicích
Přírodovědecká fakulta

ZADÁVACÍ PROTOKOL BAKALÁŘSKÉ PRÁCE

Student: Marie Doležalová
(jméno, příjmení, tituly)

Obor - zaměření studia: Aplikovaná Informatika

Katedra: Ústav aplikované informatiky

Školitel: Ing. František Drdák, CSc.
(jméno, příjmení, tituly, u externího š. název a adresa pracoviště, telefon, fax, email)

Garant z Přírodovědecké fakulty:
(jméno, příjmení, tituly, katedra - jen v případě externího školitele)

Školitel - specialista, konzultant: Ing. Iva Opekarová, LAMCH, BC AV ČR
(jméno, příjmení, tituly, u externího š. název a adresa pracoviště, telefon, fax, email)

Téma bakalářské práce:

Nástroj pro generování elementárního složení z dat hmotnostních spektrometrů s vysokým rozlišením

Popis práce:

Vytvořte nástroj na zjišťování elementárního složení neznámých látek z dat naměřených hmotnostním spektrometrem firmy Thermo Orbitrap Exactive Plus. Důkladně prostudujte současné vědecké metodiky pro stanovení elementárního složení včetně heuristických či experimentálních. Celou problematiku důkladně popište v úvodní části práce. Implementaci nástroje proveďte v programovacím jazyku C++. Na základě údajů známých látek z volně dostupných databází na internetu se pokuste provést optimalizaci parametrů programu s ohledem na rychlost či efektivitu běhu a proveďte její zhodnocení.

Cíl práce:

Software pro zpracování chemometrických dat z HRMS s možností predikce elementárního složení neznámých látek

Základní doporučená literatura:




[1] Kind, Tobias; Fiehn, Oliver, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC BIOINFORMATICS, Volume: 8, 2007

[2] Abate, Salvatore; Ahn, Yun Gyong; Kind, Tobias; et al.

Determination of elemental compositions by gas chromatography/time-of-flight mass spectrometry using chemical and electron ionization, RAPID COMMUNICATIONS IN MASS SPECTROMETRY, Volume: 24,2010

[3] Pluskal, Tomas; Uehara, Taisuke; Yanagida, Mitsuhiro, Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching, ANALYTICAL CHEMISTRY, 2012

[4] Virius M, Od C k C++, Koop, 2003

Financování práce:
Vedoucí práce: Ing. František Drdák, CSc. podpis: 
U externích vedoucích fakultní garant práce: podpis:
Garant oboru bak. studia (nepožaduje se u zaměření, příprava na mag. stud. biologie):
..... podpis:
Vedoucí katedry: Libor Dostálek (vz. Ing. Jan Fesl) podpis:
Odborný konzultant (chemická část): Ing. Iva Opekarová podpis: 
Případný souhlas vedoucí ústavu AV: podpis:
V Českých Budějovicích dne: ..19.12.2014.....
Převzal/a dne: ..19.12.2014..... podpis: 

Bibliografické údaje

Doležalová, Marie, 2015: Nástroj pro generování elementárního složení z dat hmotnostních spektrometrů s vysokým rozlišením. [Tool for generating elemental composition of the mass spectrometer data with high resolution, Bc. Thesis, in Czech] - 43 p., Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic.

Anotace: Bakalářská práce se zabývá predikcí elementárního složení chemických látek na základě znalosti hmotnosti stanovené s vysokou přesností. Predikce představuje nalezení validní kombinace počtu a druhu atomů, tj. elementárního složení molekuly, která svou hmotností odpovídá přesně naměřené hodnotě.

Anotation: This bachelor thesis deals about elementary element composition of chemical compounds, which is based on the knowledge of accurate measured mass. The prediction introduces finding of the valid combination counts of element numbers, that means elementary composition of some molecule, which mass is equal to the measured value.

Prohlašuji, že svoji bakalářskou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 15. 4. 2015

Poděkování

Chtěla bych poděkovat svému školiteli Františku Drdákovi za pomoc a přínosné rady během tvorby mé bakalářské práce a své konzultantce Ivě Opekarové za vedení mé práce po chemické stránce. Nemalý dík patří zejména mému příteli za velkou podporu a trpělivosti v průběhu přípravy na bakalářskou práci a při samotné její tvorbě.

1. OBSAH

1.	OBSAH	7
2.	ÚVOD.....	8
3.	TEORETICKÁ ČÁST	8
3.1.	Obecný princip hmotnostní spektrometrie (MS)	8
3.1.1.	Hmotnostní spektrometr.....	9
3.1.2.	Hmotnostní spektrum	13
3.2.	Spojení hmotnostní spektrometrie a chromatografie	14
3.3.	Orbitrap Q Exactive Plus.....	16
3.4.	Izotopové složení prvků a molekul	17
3.4.1.	Izotopická obálka.....	18
3.4.2.	Relativní atomová hmotnost	20
3.5.	Predikce elementárního složení.....	21
3.5.1.	Ověření predikované molekulové hmotnosti	23
3.5.2.	Heuristická pravidla	25
4.	IMPLEMENTACE	27
4.1.	Návrhová a implementační část programu	27
4.2.	Algoritmy generování sumárního vzorce	27
4.2.1.	Parametry algoritmu generování sumárního vzorce	28
4.2.2.	Algoritmus generování sumárního vzorce	29
4.2.3.	Algoritmus generování izotopické obálky (AGIO)	30
4.2.4.	Algoritmus výpočtu shody teoretické a naměřené izotopické obálky.....	30
4.2.5.	Základní důležité datové typy a funkce	31
4.2.6.	Implementace dynamicky linkované knihovny.....	32
4.3.	Optimalizace paralelního běhu programu	33
4.4.	Grafické prostředí.....	33
4.5.	Testování programu.....	36
4.6.	Optimalizace heuristických pravidel	38
5.	ZHODNOCENÍ VÝSLEDKŮ A ZÁVĚR	41
6.	SEZNAM CITOVANÉ LITERATURY	42

2. ÚVOD

Cílem této multidisciplinární bakalářské práce je vytvoření softwarového modulu pro interpretaci elementárního složení dat z hmotnostních spektrometrů. Vzhledem k tomu, že se ideově jedná o práci, která spojuje znalosti různých oborů, je nutné v úvodní části práce vysvětlit základní pojmy a principy hmotnostní spektrometrie. Na tuto část bude navazovat kapitola, ve které budou detailně zmíněna heuristická optimalizační pravidla pro generování výstupu. Na základě zkušeností z předchozích kapitol by navržen software, který rutinní práci zautomatizuje. Srovnání teoretického přístupu řešení a reálných dosažených výsledků bude provedeno v závěrečné fázi. Vývoj daného softwarového modulu bude hrát klíčovou roli při interpretaci dat naměřených v Laboratoři analytické chemie v Biologickém Centrum Akademie věd ČR, kde bude posléze i reálně nasazen.

3. TEORETICKÁ ČÁST

3.1. Obecný princip hmotnostní spektrometrie (MS)

Hmotnostní spektrometrie je fyzikálně-chemická metoda sloužící k identifikaci a kvantifikaci jak známých, tak zcela neznámých látek. Hmotnostní spektrometrie je velmi citlivá analytická technika. Je schopná s velkou rychlostí stanovit složení i velmi komplexních směsí molekul. Nepoužívá se pouze k prostému stanovení molekulové hmotnosti analyzované molekuly, ale slouží díky možnosti fragmentace molekul i k objasňování její vnitřní struktury a chemických vlastností.

Zpracování vzorků hmotnostními spektrometry se děje pomocí elektrických a elektromagnetických polí. (9) Elektricky neutrální částice nejsou těmito poli ovlivněny a hmotnostní spektrometrie je tak závislá na převedení látek obsažených ve vzorku na nabitě částice (ionty). Aby se ionty mohly pohybovat v elektrostatických polích, převádí hmotnostní spektrometrie analyzované vzorky do plynné fáze. (2)

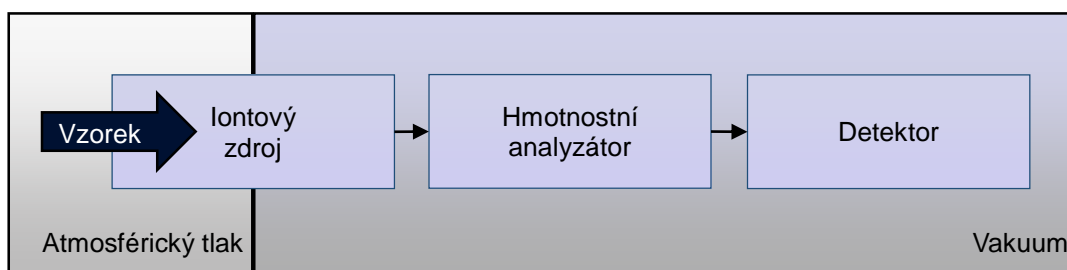
3.1.1. Hmotnostní spektrometr

Hmotnostní spektrometr je iontově-optický přístroj umožňující měřit hmotnosti a relativní četnosti iontů ve vzorku. Hmotnostní spektrometr měří hmotnost jednotlivých iontů nepřímo, tj. na základě poměru hmotnosti iontu m a počtu jeho elementárních nábojů z (m/z). Většina iontů vytvořených v hmotnostním spektrometru má náboj rovný jedné. Poměr m/z iontu tak odpovídá molekulové hmotnosti. (1)

3.1.1.1. Základní princip hmotnostního spektrometru

Hmotnostní spektrometr je složen ze tří základních částí:

- 1) Iontový zdroj
- 2) Hmotnostní analyzátor
- 3) Detektor



Obrázek č. 1 – Obecné schéma hmotnostního spektrometru

Vzorek je vstříknut do prostoru iontového zdroje. Zde dochází k ionizaci molekul vzorku na ionty. Vytvořené ionty jsou následně elektrostaticky směřovány do hmotnostního analyzátoru. Vzhledem k tomu, že ionty jsou velmi reaktivní a existují jen po velmi krátkou dobu, musí být jejich vznik a manipulace prováděna ve vakuu. Vysokým vakuem je zamezeno kolizím s molekulami vzduchu. (1)

V hmotnostním analyzátoru dochází k oddělení iontů na základě poměru jejich hmotnosti a náboje (m/z) a v detektoru jsou ionty zachytávány. Detektor převádí energii jednotlivých iontů na elektrické signály, a ty jsou posléze přenášeny do počítače jak tzv. hmotnostní spektrum. (2)

3.1.1.2. Rozlišovací schopnost hmotnostního analyzátoru

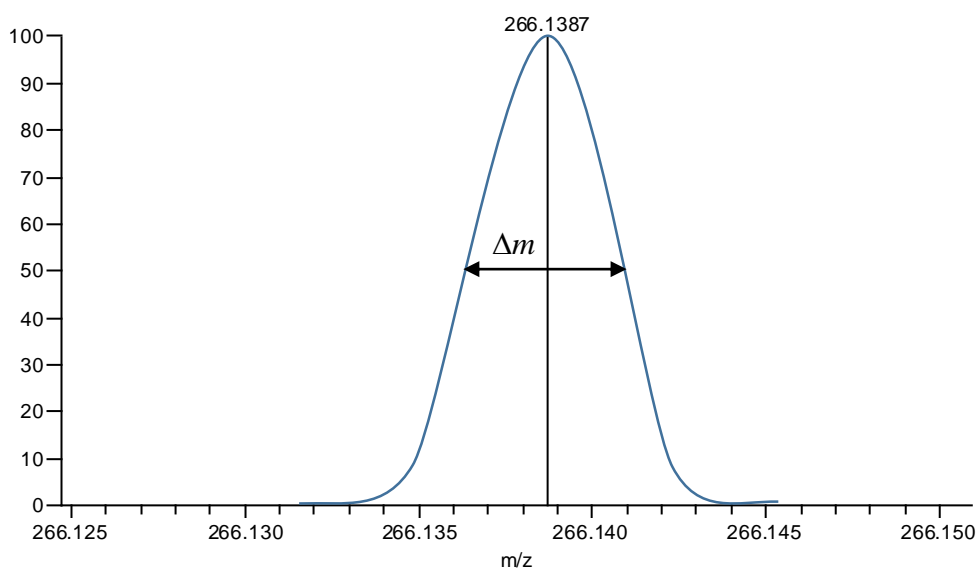
Ionty se stejným poměrem m/z by měly dosahovat detektoru ve stejný okamžik, nicméně tomu tak není a výstup z detektoru je časově proměnný. Příčiny mohou být

od různé počáteční energie u iontů se stejnou hmotností až po nízkou rozlišovací schopnost analyzátoru. V takovém případě se stane, že pokud se ve vzorku nacházejí ionty mající velmi podobné elementární složení, tudíž i velmi podobný poměr hmotnosti a náboje, přístroj nemusí být schopen tyto sloučeniny od sebe odlišit. (3)

V ideálních podmínkách a za stálé přesnosti hmotnostního analyzátoru by bylo spektrum tvořeno čarami, které by odpovídaly příslušným poměrům m/z . Z ideálního čárového hmotnostního spektra se díky nestabilitám stává graf tvořený sadou čar, které se shlukují do tvarů křivky Gaussova rozložení, jejichž vrcholy odpovídají statisticky nejpřesnějším hodnotám m/z analyzovaných iontů.

Do jaké míry je hmotnostní analyzátor schopen s jistotou rozlišovat hmotnostně blízké ionty je závislé na jeho typu a konstrukci. Tzv. rozlišovací schopnost (Resolving power, RP) hmotnostního analyzátoru je jedním z hlavních parametrů i celého hmotnostního spektrometru. Definice stanovení rozlišovací schopnosti hmotnostního analyzátoru, není jednotná a liší se podle typu analyzátoru. Obecně se rozlišovací schopností rozumí schopnost rozlišit ionty s blízkou hodnotou poměru m/z . (3) Rozlišovací schopnost lze definovat dvojím způsobem. (1)

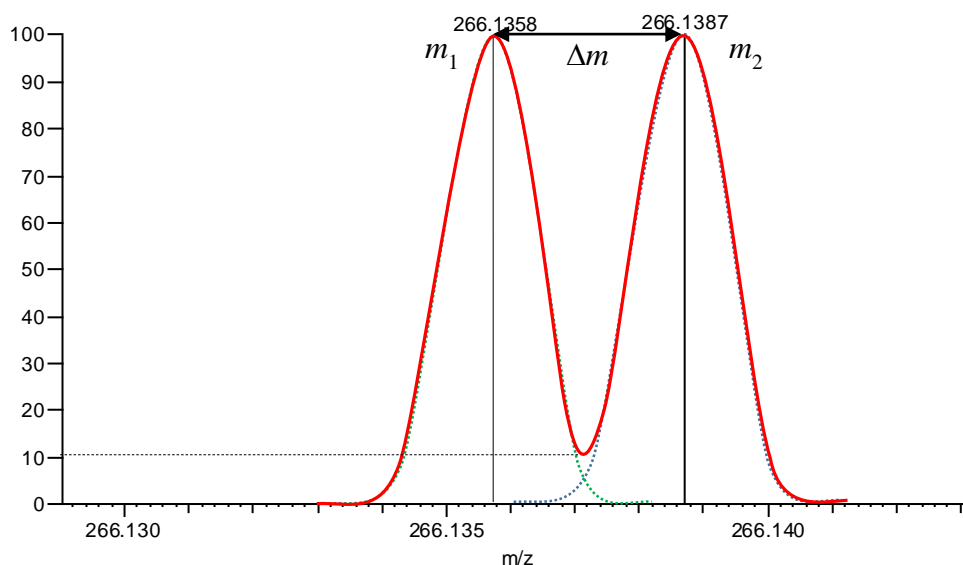
- 1) Definice na základě izolovaného píku. Rozlišovací schopnost je stanovena na základě podílu poměru m/z a šířky píku v určité výšce píku. Nejčastěji se používá šířka píku v polovině výšky (FWHM, Full Width at Half Maximum, Δm). (1)



Obrázek č. 2 – Určení rozlišovací schopnosti na základě jednoho hmotnostního píku

$$RP = \frac{\frac{m}{z}}{\Delta m}$$

- 2) Definice na základě překryvu dvou stejně vysokých sousedních píků a s jednotkovým nábojem. Rozlišovací schopnost je pak stanovena při jejich 10% překryvu, tzn. „údolí“ mezi nimi dosahuje k 10 %. (1)



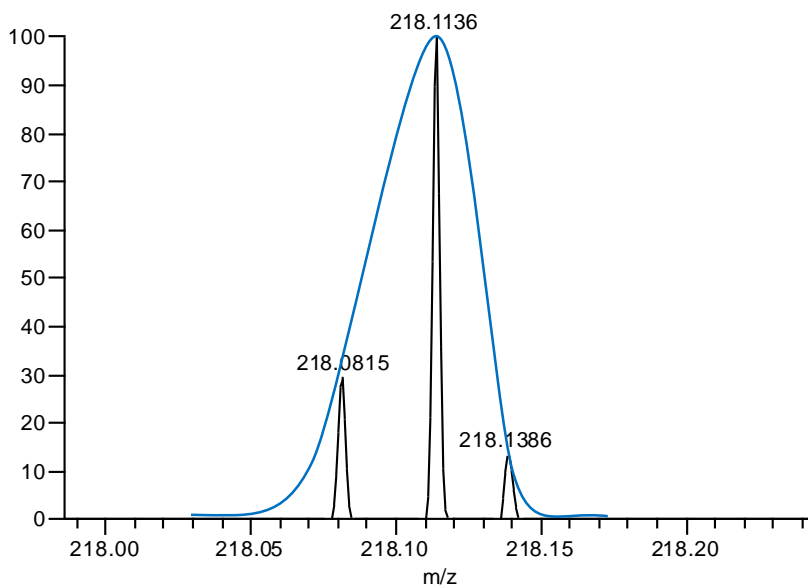
Obrázek č. 3 – Určení rozlišovací schopnosti na základě překryvu hmotnostních píků

$$RP = \frac{\frac{m_1}{z}}{\Delta m}$$

Podle rozlišovací schopnosti definované na základě překryvu dvou sousedních píků dělíme hmotnostní spektrometry do dvou tříd. Hmotnostní spektrometry s vysokou rozlišovací schopností a spektrometry s nízkou rozlišovací schopností. Mezi spektrometry s nízkou rozlišovací schopností patří přístroje s tzv. jednotkovým rozlišením, kdy můžeme rozlišit hmotnost iontů pouze s přesností na celá čísla. (5) Hmotnostními analyzátory s nízkou rozlišovací schopností jsou například kvadrupóly a iontové pasti. Rozlišovací schopnost těchto spektrometrů se pohybuje mezi 10² až 5x10³. Do skupiny hmotnostních spektrometrů s vysokou rozlišovací schopností řadíme ty, které mají rozlišovací schopnost mezi 10⁴ až 10⁶. (4)

Rozlišovací schopnost hmotnostního analyzátoru, kterou udávají výrobci, se vztahuje vždy k určitému hmotovému rozsahu. Může být pro celý rozsah analyzátoru,

tj. hmotnosti iontů, které je schopen spektrometr změřit, anebo může být stanovena pro jednotlivé intervaly hmot různě.



Obrázek č. 4 – Porovnání možnosti zobrazení blízkých hodnot m/z

Na obrázku je znázorněno srovnání rozlišení blízkých hmotností, resp. m/z oběma třídami hmotnostních spektroskopů. Přístroj s nízkou rozlišovací schopností by rozlišil pouze jeden pík, zakreslený modrou barvou, $RP(FWHM_{50\%}) = 218.1136/0,07 = 3116$. Spektroskop s vysokou rozlišovací schopností by rozlišil již tři odlišné píky, které jsou zachyceny černou barvou, $RP(FWHM_{50\%}) = 218.1136/0.0028 = 77\ 898$.

3.1.1.3. Správnost hmoty naměřené hmotnostním spektrometrem

Dalším důležitým parametrem hmotnostního analyzátoru je správnost, se kterou je zjištěná hmotnost iontu měřena. Důležitý je zejména u hmotnostních spektrometrů s vysokou rozlišovací schopností. Vysoká správnost naměřené hmoty umožňuje stanovení elementárního složení příslušného iontu. Hmotnostní spektroskop musí poskytnout hmotnost s daným rozlišením a s garantovanou správností.

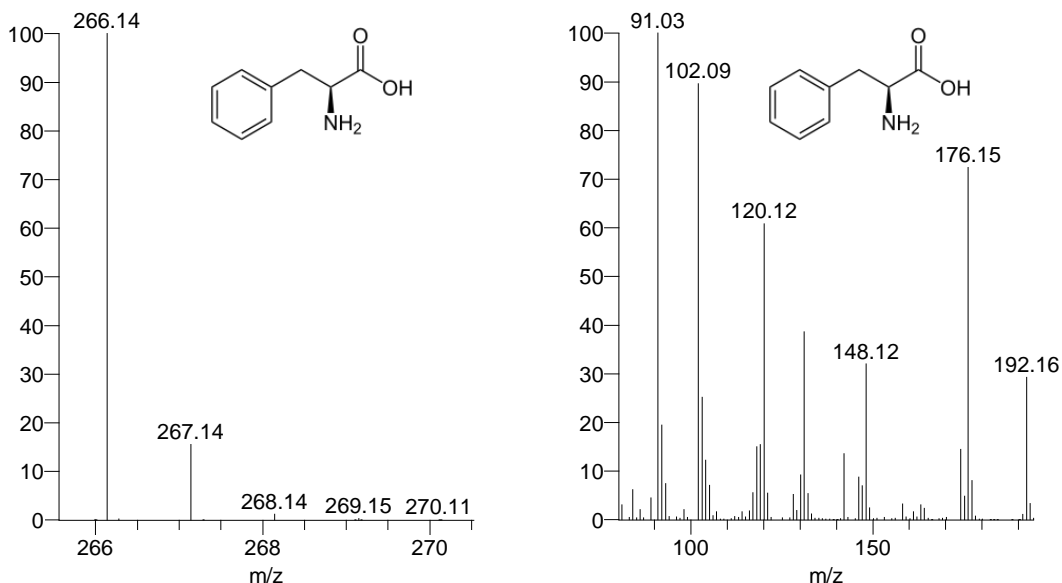
Správnost naměřené hmoty udává rozdíl mezi hmotností, kterou očekáváme, tj. hmotností vypočtenou, a hmotností naměřenou přístrojem. Správnost se nejčastěji vyjadřuje jako bezrozměrná veličina ppm (miliontina z celku, parts per million).

$$ppm = \frac{\frac{m_{(\text{naměřená})}}{Z} - \frac{m_{(\text{vypočtená})}}{Z}}{\frac{m_{(\text{vypočtená})}}{Z}} \times 10^6$$

3.1.2. Hmotnostní spektrum

Hmotnostní spektrum je záznam závislosti četnosti iontů a poměru jejich hmotnosti a náboje. Spektrum je obvykle zjednodušeně zobrazováno jako vertikální čárový graf, v němž čára představuje iont odpovídající poměru m/z a délka čáry udává poměrnou četnost příslušného iontu. Každé spektrum je normalizováno vzhledem k tzv. základnímu iontu, což je iont s nejvyšší četností. Četnosti ostatních iontů spektra jsou pak vyjádřeny v procentech vzhledem k základnímu iontu. (3)

Vzhled hmotnostního spektra je závislý na mechanismu ionizace vzorku. (1) Nejčastěji používaným mechanismem u plynové chromatografie je elektronová ionizace (EI) a u kapalinové chromatografie ionizace elektrospřejem (ESI). Vzhled spekter je ilustrován na obrázku 5. Na obou hmotnostních spektrech je zachycena aminokyselina fenylalanin. V případě ESI se jedná o tzv. měkkou ionizační techniku, kdy nedochází k destrukci primárního analyzovaného iontu. Ve spektru tak můžeme najít molekulový ion, který odpovídá analyzované sloučenině. V levém hmotnostním spektru se jedná o $m/z = 266.14$. Oproti tomu elektronová ionizace je tzv. tvrdou ionizační technikou, kdy dochází k destrukci původního iontu a ke vzniku několika fragmentů, které jsou pro původní iont charakteristické. Molekulový ion mívá při této ionizaci velmi malou četnost a ve spektru nemusí být patrný.



Obrázek č. 5 – Porovnání hmotnostních spekter vytvořených odlišnou ionizací, na ose x je uvedena hodnota m/z iontů, na ose y je relativní četnost iontů.

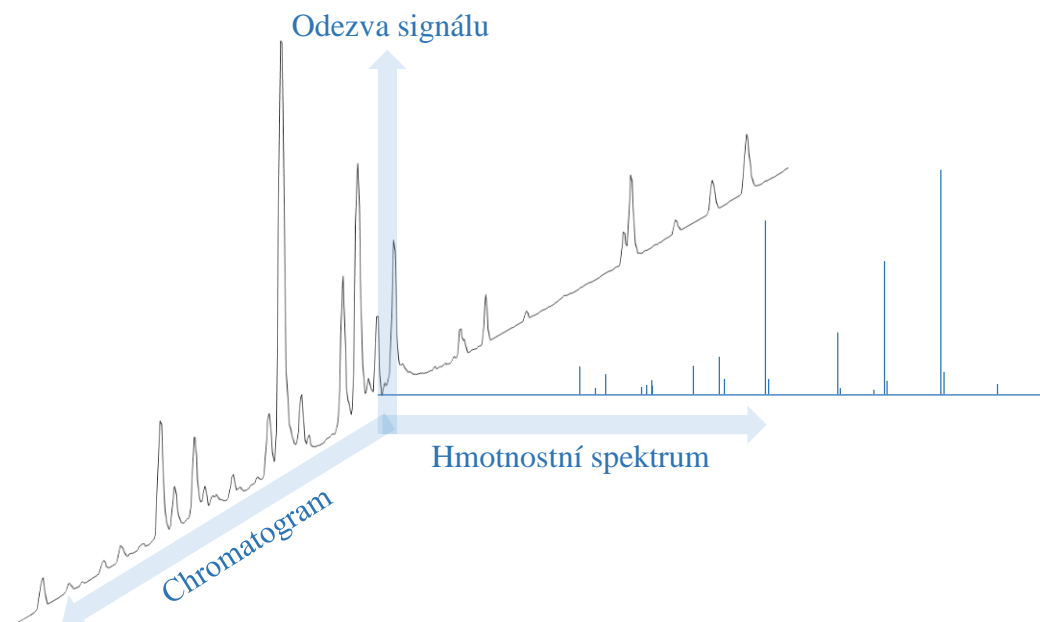
3.2. Spojení hmotnostní spektrometrie a chromatografie

Hmotnostní spektrometrie je schopná oddělit molekuly ve vzorku na základě jejich molekulové hmotnosti a umožňuje jejich detekci a kvantifikaci s extrémně vysokou citlivostí. Nicméně u velmi složitých vzorků obsahujících řádově stovky a tisíce sloučenin nastává problém, protože hmotnostní spektrometr je schopen oddělit sloučeniny pouze na základě podílu m/z . Například hmotu 250Da má v přírodě více než 1500 různých sloučenin. Spojením se separačními technikami, které provedou velmi rychlé, kvantitativní oddělení sloučenin v čase, se docílí jedinečného, rychlého, nákladově efektivního a kvantitativního měření celkové směsi. (2)

Chromatografie odliší sloučeniny na základě jejich fyzikálně-chemických vlastností a hmotnostní spektrometrie rozliší sloučeniny podle jejich hmotnosti, resp. poměru m/z . Je to právě tato dvojí selektivita, která dělá z tandemových spojení tak silný analytický nástroj. Síla této techniky je znázorněna na obrázku 6. (2)

Směs sloučenin je nanášena na chromatografickou kolonu a ta následně podle vlastností jednotlivých sloučenin tyto sloučeniny různě dlouho v koloně zadržuje. Dojde tak k separaci jednotlivých sloučenin v čase. Hmotnostní spektrometr v daných časových intervalech provádí měření, tzv. provádí skeny aktuálního složení výstupu

z chromatografické kolony a vytváří hmotnostní spektrum. Během měření dochází k simultánnímu záznamu průběhu chromatografie do tzv. chromatogramu. Existuje více druhů chromatogramů, nicméně se vždy jedná o závislost relativní četnosti iontů skenu a tzv. retenčního času, tj. času, ve kterém byl sken proveden.



Obrázek č. 6 – Propojení měření chromatografem a hmotnostním spektrometrem

Spojení separačních technik a hmotnostní spektrometrie se věnuje řada firem. Je možné vybrat si na trhu téměř jakoukoliv kombinaci. Možnost volby začíná samotným spektrometrem, kdy každá ze tří základních součástí hmotnostního spektrometru může být jinak konstrukčně řešena. První důležitou součástí je iontový zdroj. Na typu iontového zdroje a mechanismu vzniku iontů je závislý tvar hmotnostního spektra. Druhou důležitou součástí je hmotnostní analyzátor, který zejména zodpovídá za to, s jakou přesností budou hmotnosti iontů měřeny. Co se týká separačních technik, hmotnostní spektrometrie se používá ve spojení jak s plynovou chromatografií (Gas chromatography mass spectrometry, GC/MS), tak kapalinovou chromatografií (Liquid chromatography mass spectrometry, LC/MS).

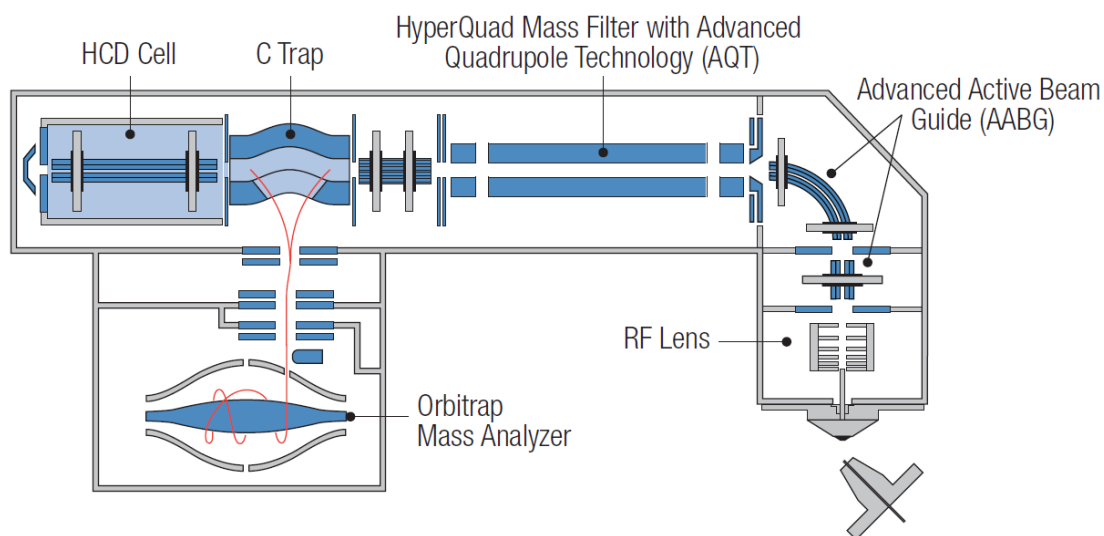
3.3. Orbitrap Q Exactive Plus



Obrázek č. 7 – Orbitrap Q Exactive Plus

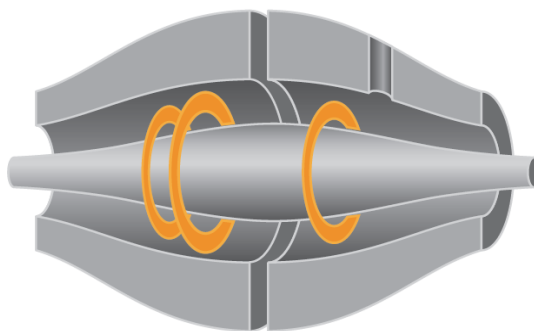
Jedná se o hmotnostní spektrometr s vysokou rozlišovací schopností. Vysoká rozlišovací schopnost je dána konstrukčním řešením hmotnostního analyzátoru. Analyzátozem, jak již název napovídá, je Orbitrap, elektrostatická orbitální past, který navrhl ruský fyzik Alexander Makarov (13).

Orbitrap Q Exactive Plus je svými částmi klasickým hmotnostním spektrometrem.



Obrázek č. 8 – Vnitřní organizace spektrometru Orbitrap Q Exactive Plus

V iontovém zdroji dojde ke vzniku iontů, které jsou následně zaostřovány a směřovány do kvadrupólového hmotového filtru, kde dojde odchylení iontů, které nepatří do zvoleného hmotového rozsahu. Svazek iontu pak směřuje do zakřivené lineární pasti, C-trapu (Curved Linear Trap). Zde ionty ztrácejí svou energii kolizemi s molekulami dusíku, Poté, co jsou dostatečně zbrzděny, jsou stlačeny v centru lineární pasti a vstřeleny do hmotnostního analyzátoru, Orbitrapu. (14)



Obrázek č. 9 – Schématické zobrazení Orbitrapu

Orbitrap je tvořen dvěma elektrodami. Uvnitř je centrální elektroda vřetenovitého tvaru, kolem které během měření krouží ionty, a vnější elektrody, podél které se kroužící ionty pohybují v závislosti na elektrostatickém poli. (13)

Na centrální elektrodě Orbitrapu se postupně zvyšuje napětí a ionty se začínají spirálově pohybovat. Ionty se stejnou hodnotou m/z vytvářejí kolem centrální elektrody prstence. Během pohybu generují proudový signál, který je pomocí Fourierovy transformace převeden na hmotnostní spektrum. (14)

Pokud je prováděna fragmentace iontu, jsou ionty s danou hodnotou m/z z lineární pasti směřovány do kolizní cely (HCD Cell, Higher Energy Collisional Dissociation Cell), kde dojde k rozbití molekul na dílčí fragmenty, které jsou přivedeny zpět do lineární pasti a odtud stejným principem uvedeným výše do Orbitrapu. (13)

3.4. Izotopové složení prvků a molekul

Každý chemický prvek, který najdeme v periodické tabulce je charakterizován několika základními čísly odrážejícími elementární atomové složení daného prvku. Každý atom prvku je složen z jádra, kde se nacházejí protony a neutrony, a obalu, kde se nacházejí elektrony. Z hlediska příslušnosti atomu k danému prvku je rozhodující

počet protonů. Počet protonů v jádře se pak na základě této skutečnosti označuje jako tzv. atomové číslo. Aby byl atom považován za daný prvek, musí mít stejné atomové číslo, nicméně počet elektronů a neutronů může být různý.

Pokud má atom stejný počet protonů v jádře a elektronů v obalu, má celkový náboj nulový. Jestliže atomu odebereme nebo naopak přidáme elektrony, pak se stává nabitou částicí, která je označována jako *iont*. Jestliže počet neutronů v jádře neodpovídá atomovému číslu, resp. počtu protonů, pak se jedná o tzv. *izotop*. Součet protonů a neutronů v jádře se označuje jako tzv. nukleonové číslo.

Izotopy daného prvku se liší svými fyzikálními vlastnostmi, mohou mít v přírodě různou stabilitu a četnost, nicméně svými chemickými vlastnostmi se nijak vzájemně neliší a podléhají stejným chemickým reakcím. (9) Důležitým dopadem tvorby izotopů je, že díky nim dochází ke změně atomové hmotnosti.

Většina chemických prvků vyskytujících se v přírodě je tvořena více než jedním izotopem. Zastoupení jednotlivých izotopů stanovuje a kontroluje mezinárodní unie pro čistou a užitou chemii, která každé dva roky tabulku izotopového rozložení prvků aktualizuje. (7)

Tabulka č.1 – Zastoupení izotopů šesti nejvíce zastoupených prvků v přírodě

Prvek	Izotop	Četnost	Hmotnost [Da]	Izotop	Četnost	Hmotnost [Da]
Vodík	¹ H	0.999 885	1.007 825 0322	² H	0.000 115	2.014 101 7781
Uhlík	¹² C	0.989 3	12 (přesně)	¹³ C	0.010 7	13.003 354 835
Dusík	¹⁴ N	0.996 36	14.003 074 004	¹⁵ N	0.003 64	15.000 108 899
Kyslík	¹⁶ O	0.997 57	15.994 914 620	¹⁷ O	0.000 38	16.999 131 757
				¹⁸ O	0.002 05	17.999 159 613
Fosfor	³¹ P	1	30.973 761 998			
Síra	³² S	0.949 9	31.972 071 174	³³ S	0.007 5	32.971 458 910
				³⁴ S	0.042 5	33.967 867 0
				³⁶ S	0.000 1	35.967 081

3.4.1. Izotopická obálka

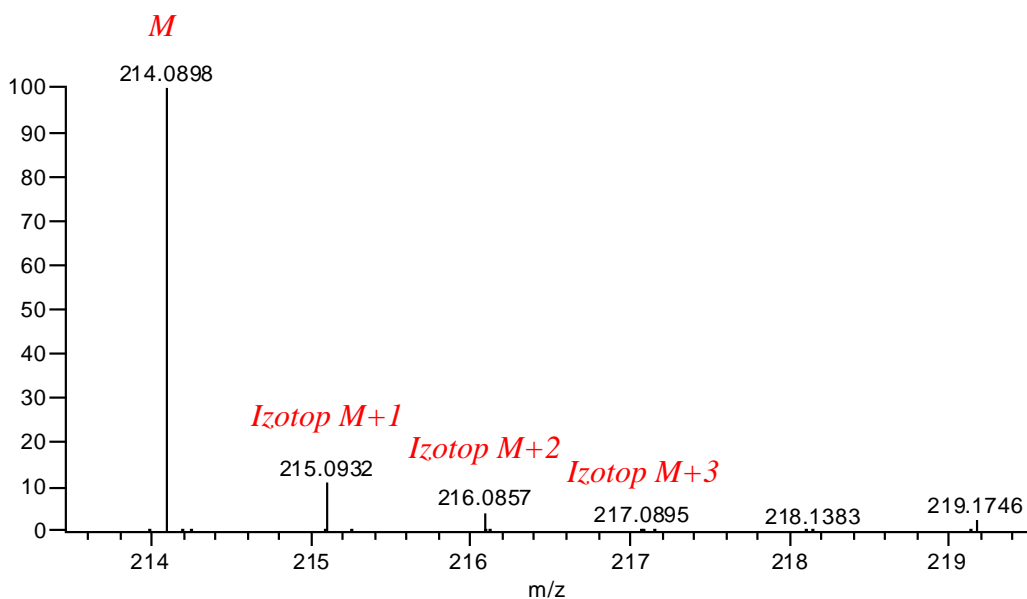
Díky vysoké rozlišovací schopnosti hmotnostních spektrometrů, a to již u tzv. jednotkových hmotnostních spektrometrů, hmotnostní spektrometrie neidentifikuje molekuly pouze jediným údajem, a to údajem o její molekulové hmotnosti, ale je schopna zachytit i hmotnosti jednotlivých izotopů.

Hmotnostní spektrometr by ukázal jeden hmotnostní údaj (m/z) jedině tehdy, pokud by se daná molekula skládala pouze z nejčtenějších izotopů všech obsažených atomů. Taková pravděpodobnost se u molekul pocházející z přírodního zdroje téměř blíží nule. Pokud bychom brali v úvahu jednoduchou pravděpodobnost a molekula by byla tvořena jen jedním atomem uhlíku, mělo by se na základě známého poměrného zastoupení jeho výskytu v přírodě v 98,93 % jednat o uhlík $^{12}_6\text{C}$. Pokud by molekula obsahovala 10 atomů uhlíku, pak by měla mít 9,893 uhlíků $^{12}_6\text{C}$ a pouze 0,107 uhlíků $^{13}_6\text{C}$. Pokud by obsahovala 100 atomů uhlíku, pak by již měla obsahovat 98,93 uhlíků $^{12}_6\text{C}$ a 1,07 uhlíků $^{13}_6\text{C}$. Až v této chvíli by se měla na hmotnostním spektru objevit kromě čáry odpovídající hmotě (m/z) uhlíku $^{12}_6\text{C}$ i čára odpovídající hmotě (m/z) uhlíku $^{13}_6\text{C}$. Ve skutečnosti by se na hmotnostním spektru čáry obou izotopů ukázaly již při měření molekuly tvořené jedním uhlíkem. Je to dáno tím, že hmotnostní spektrometr neměří jedinou molekulu, ale řádově 100ky tisíc, milióny, molekul té samé látky.

Zatím bylo uvažováno, že se molekula skládala pouze z uhlíků, nicméně molekuly obsahují více druhů prvků. Většina prvků, stejně jako uhlík, tvoří své izotopy a svým poměrným zastoupením přispívají pak k četnosti příslušných izotopů. Výsledné hmotnostní spektrum jakékoliv měřené molekuly tak vždy koresponduje s izotopovým rozložením prvků molekuly v přírodě. (9)

Izotopickou obálku si lze představit jako hmotnostní spektrum v omezeném rozsahu. Pokud se podíváme na hmotnostní spektrum molekuly na následujícím obrázku, najdeme v něm nejintenzivnější hmotnost (m/z) odpovídající molekulové hmotnosti látky a pak další méně intenzivní hmotnosti (m/z), které odpovídají izotopům tvořených prvky, ze kterých se molekula látky skládá. Takovéto spektrum pro určitou látku označujeme jako izotopická obálka, pattern. (6).

Zastoupení izotopů se normalizuje vzhledem k nejčtenějšímu izotopu, jak je znázorněno na obrázku 10, nebo se sčítají v procentech do sumy 100 %. Jednotlivé hmotnosti izotopů se značí jako $M + \text{posun}$ od molekulového iontu M . Písmeno M je zástupné pro obecný molekulový iont, neboť, jak je níže uvedeno, molekulový iont může být různého typu.



Obrázek č. 10 – Izotopická obálka sloučeniny

Aby bylo možné takovéto hmotnostní spektrum správně interpretovat, musíme znát informace jak o hmotnostech izotopů, tak o jejich vztahu k příslušným prvkům.

3.4.2. Relativní atomová hmotnost

Jak je patrné již z názvu technologie hmotnostní spektrometrie, hraje hmotnost, ať již atomová nebo molekulová, a zejména její přesnost v hmotnostní spektrometrii nedůležitější roli. Je řada možností, jak hmotnost vyjadřovat. Podle národního ústavu pro standardy a technologie jsou hmotnosti elementárních částic v tabulce 2(12).

Tabulka č.2 – Přesná hmotnost elementárních částic

Částice	Náboj	Hmotnost [kg]	Hmotnost [Da]
Proton	+1	1.672 621 777 e-27	1.007 276 466 812
Neutron	0	1.674 927 351 e-27	1.008 664 916 000
Elektron	-1	9.109 382 91 e-31	5.485 799 0946 e-4

Hmotnosti jednotlivých částic, potažmo celých atomů, vyjádřené v kilogramech jsou velmi malá čísla. Byla proto pro ně zavedena nová míra, tzv. relativní atomová hmotnost. Relativní atomová hmotnost je skutečnou hmotností atomu vztaženou k tzv. atomové hmotnostní konstantě, která odpovídá 1/12 hmotnosti atomu $^{12}_6\text{C}$:

$$1 u = 1.660\,538\,921 \times 10^{-27} \text{ kg}$$

Jednotkou relativní atomové hmotnosti se stala unifikovaná atomová hmotnostní

jednotka u , v anglické verzi amu . V hmotnostní spektrometrii se nicméně častěji používá synonymní jednotka Da (dalton), $1 Da = 1 u$. (1)

Relativní hmotnost atomu udává, kolikrát je skutečná hmotnost atomu větší než atomová hmotnostní konstanta. Relativní molekulová hmotnost je pak součtem relativních atomových hmotností prvků, z nichž je molekula složená. (6)

3.4.2.1. Střední relativní hmotnost

Tato hmotnost se používá v případě existence izotopů, kdy je skutečná hmotnost stanovena jako vážený průměr hmotností všech přirozeně se vyskytujících stabilních izotopů daného prvku. Tuto hmotnost najdeme uvedenou v periodické tabulce. A_i je četnost i -tého izotopu a m_i je hmotnost i -tého izotopu v Da . (6)

$$m_{Ar} = \frac{\sum_{i=1}^i A_i \cdot m_i}{\sum_{i=+}^i A_i}$$

3.4.2.2. Monoizotopická hmotnost

Skutečná hmotnost se počítá oproti střední relativní hmotnosti jako suma hmotnosti atomů vyjádřených pouze hmotností nejčtenějšího, tj. většinou nejlehčího izotopu daného prvku uvedeného rovněž v Da . (5)

$$m_{Ar} = \sum_{i=1}^i m_i$$

3.5. Predikce elementárního složení

Každá molekula má stejně jako chemický prvek několik základních vlastností, kterými se liší od ostatních molekul. Jednou z nejzákladnějších vlastností je hmotnost. Hmotnostní spektrometrie s vysokou rozlišovací schopností jde dále. Molekulová hmotnost odráží tzv. elementární složení, sumární vzorec, což je počet jednotlivých atomů prvků, ze kterých se molekula skládá. Pokud je použit hmotnostní analyzátor, který umožňuje určit m/z hodnoty s přesností na minimálně čtyři desetinná místa, je možné rozlišit i přesné izotopové složení.

Proč na čtyři desetinná místa? Vše to souvisí s jádrem atomu. Jádro atomu se skládá z protonů a neutronů, a ty jsou vzájemně přitahovány jadernými silami. Tyto

síly mají sice velmi krátký dosah, ale jsou asi tisíckrát silnější než elektromagnetické síly, kterými se stejně nabitě částice odpuzují a různě nabitě částice přitahují. Tyto síly začínají působit ve chvíli, kdy se protony a neutrony téměř dotýkají. Při vytvoření jádra, tj. při spojení protonů a neutronů, dochází k uvolnění energie, která dle Einsteinovy teorie relativity vede i k odpovídajícímu úbytku hmotnosti. Rozdíl mezi těmito hmotnostmi označujeme jako hmotnostní schodek jádra, resp. vazebná energie jádra. Součet hmotností samostatných částic jádra je proto vždy větší než hmotnost jádra. Izotopy jsou částice, které se liší počtem neutronů v jádře. Váha samotného neutronu je vždy stejná, nicméně pokud dojde k jeho připojení do jádra, pak rozdíl hmotnosti původního atomu a tohoto atomu izotopu závisí na množství energie, která se uvolní. Velikost uvolněné energie, a tedy i hmotnostní úbytek závisí na atomovém čísle prvku. Při vytváření molekuly tak každý prvek ovlivňuje výslednou molekulovou hmotnost podle svého izotopového složení. (7)

Tabulka č.3 – Zastoupení izotopů prvků C, H, N, O, P, S

Prvek	Izotop	Hmotnost [Da]	Izotop	Hmotnost [Da]	Rozdíl hmotnosti [Da]
Vodík	¹ H	1.007 825 032 2	² H	2.014 101 7781	1.006 276 745 9
Uhlík	¹² C	12 (přesně)	¹³ C	13.003 354 835	1.003 354 835
Dusík	¹⁴ N	14.003 074 004	¹⁵ N	15.000 108 899	0.997 034 895
Kyslík	¹⁶ O	15.994 914 620	¹⁷ O	16.999 131 757	1.004 217 137
			¹⁸ O	17.999 159 613	2.004 244 993
Fosfor	³¹ P	30.973 761 998			
Síra	³² S	31.972 071 174	³³ S	32.971 458 910	0.999 387 736
			³⁴ S	33.967 867 0	1.995 795 826
			³⁶ S	35.967 081	3.995 009 826

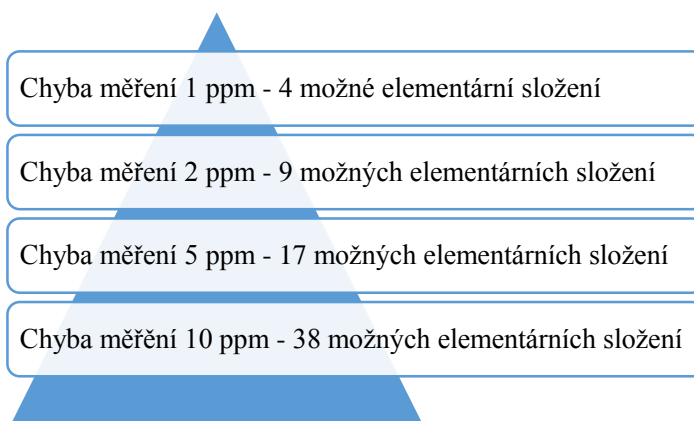
Kdybychom měřili každý prvek zvlášť, na hmotnostním spektru by se zobrazilo tolik čar, kolik by odpovídalo počtu izotopů, které prvek v přírodě tvoří a velikost čar by odpovídala jejich poměru. Při měření molekul se měří několik prvků najednou a výsledný tvar izotopické obálky je tak dán superpozicí izotopických příspěvků jednotlivých atomů. Každá molekula má tuto izotopovou obálku jako jakýsi otisk prstu a podle něj lze potvrdit či vyvrátit navržené elementární složení.

Predikce elementárního složení je počítáním celkové hmotnosti kombinací atomů, ze kterých se může sloučenina skládat. Při predikci elementárního složení jsou generovány všechny možné kombinace z různých prvků a hledá se ta správná. Správná je taková kombinace atomů, která svou vypočtenou hmotností odpovídá přesně nebo

s co nejmenší odchylkou od hmotnosti naměřené.

Základním problémem predikce elementárního složení je exponenciální nárůst počtu validních elementárních složení na několika úrovních. Základem všeho je mít hmotu s dostatečnou přesností a správností.

Množství možných kombinací narůstá s klesající správností měření. Pro predikci elementárního složení iontu je limitní hodnota 5 ppm což odpovídá odchylce 0,005 Da při m/z 1000 či 0,001 Da při m/z 200. (4)



Obrázek č. 11 – Nárůst množství možných elementárních složení

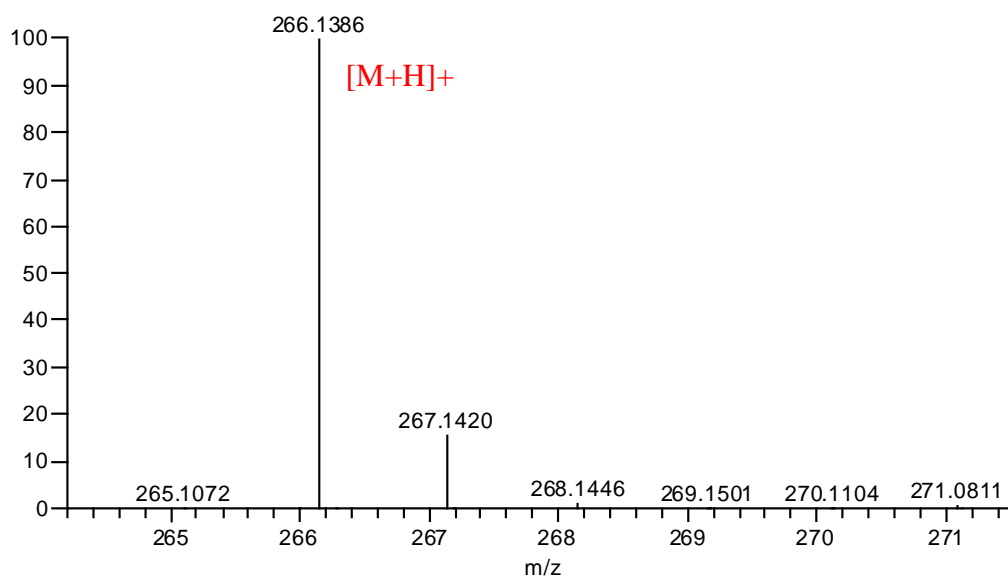
Množství možných kombinací roste a roste se vzrůstající hmotností sloučeniny. S každým dalším atomem se razantně zvyšuje počet možných kombinací prvků, ze kterých se může analyzovaná sloučenina skládat. Důležité je proto omezit počet možných kombinací. Již v prvním kroku lze jejich počet omezit volbou atomů, ze kterých by se měla sloučenina skládat. V přírodě je nejvíce zastoupeno pouze šest prvků, tj. vodík, uhlík, dusík, kyslík, fosfor a síra. (9)

Ne každá matematicky generovaná kombinace atomů je z hlediska chemického, biologického a funkčního správná. Základem ověření správnosti navrhovaných elementárních složení se stala práce Dr. Tobiaše Kinda a Dr. Olivera Fiehna (8), ve které pomocí analýz velkých databází stanovili 7 heuristických pravidel.

3.5.1. Ověření predikované molekulové hmotnosti

Je důležité si uvědomit, že hmotnostní analyzátor rozlišuje hmotnosti na úrovni jednotlivých izotopů, proto jsou všechny údaje v hmotnostním spektru tedy vždy tzv. monoizotopickými hmotnostmi. Dalším důležitým faktem je, že v iontovém zdroji dochází k přeměně neutrální molekuly na nabitě ionty, a tak i v hmotnostním spektru

najdeme hmotu iontu, nikoliv výchozí molekuly. Pokud se jedná o kladnou ionizaci elektrosprejem (ESI), pak vzniká kladný ion označovaný $[M+H]^+$, pokud o zápornou ionizaci ESI, vzniká záporný ion označovaný $[M-H]^-$. Na obrázku 12 je pozitivní ESI hmotnostní spektrum. Nejintenzivnější pík odpovídá iontu $[M+H]^+$.



Obrázek č. 12 – Spektrum izotopického složení látky s $[M+H]^+$ 266.1386 Da

Monoizotopická hmotnost $[M+H]^+$ ze spektra: 266,1386 Da

Predikované elementární složení, sumární vzorec $[M]$: $C_{14}H_{19}O_4N$

Monoizotopická hmotnost vypočtená pro $[M]$: 265.1314081035 Da

Monoizotopická hmotnost H: 1,0078250321 Da

Monoizotopická hmotnost e^- : 0,00054857990946 Da

Výpočet monoizotopické hmotnosti $[M+H]^+$:

$$M+H = 265,1314081035 + 1,0078250321 = 266,1392331356$$

$$M+H-e^- = 266,1392331356 - 0,00054857990946 = 266,13868455569054$$

K ověření správnosti predikované molekulové hmotnosti, se používá stejného postupu jako při stanovení přesnosti měření hmotnostního analyzátoru, a to pomocí *ppm*, což představuje odchylku predikované hmotnosti od hmotnosti naměřené přístrojem. V konečném důsledku by se mělo při stanovení odchylky od predikované

hmotnosti brát i v úvahu, že naměřená hodnota hmotnosti je měřena jen s určitou mírou spolehlivosti, která je daná měřicím přístrojem.

$$ppm = \frac{\frac{m_{(\text{naměřená})}}{Z} - \frac{m_{(\text{vypočtená})}}{Z}}{\frac{m_{(\text{vypočtená})}}{Z}} \times 10^6$$

$$ppm = \frac{266,1386 - 266,138685}{266,138685} \times 10^6$$

$$ppm = 0,319$$

3.5.2. Heuristická pravidla

S rozvojem rozlišovací schopnosti hmotnostních spektrometrů se otevřela možnost predikce elementárního složení analyzovaných látek na základě přesné hmotnosti molekuly a přesných hmotností jejích jednotlivých částí, tj. izotopické obálky. Heuristická pravidla usnadňující nalezení správného sumárního vzorce pro analyzovanou přesnou molekulovou hmotnost.

3.5.2.1. Dusíkové pravidlo

Dusíkové pravidlo říká, že pokud má molekula lichou molekulovou hmotnost, pak obsahuje lichý počet atomů dusíků. Nicméně když toto pravidlo bylo tvořeno, molekulová hmotnost se počítala z celočíselných hodnot atomových hmotností, což z dnešního pohledu vneslo do hodnot značné nepřesnosti. Pokud použijeme výpočty s přesnými monoizotopickými hmotnostmi, pravidlo při vyšších hmotnostech neplatí. Heuristicky bylo zjištěno, že pravidlo je nespolehlivé od hmoty 500 Da. (8).

3.5.2.2. Počty prvků v molekule

Toto pravidlo říká, že pokud známe hmotnost molekuly, není nutné do predikce jejího elementárního složení zahrnovat neadekvátně vysoké počty jednotlivých atomů. Pro 4 intervaly molekulových hmotností byly stanoveny maximální počty jednotlivých prvků, které je do predikce elementárního složení smysluplné zahrnout. (8)

3.5.2.3. Lewisovo a Seniorovo pravidlo

Hmotnostní spektrometrie měří nabitě ionty. Některé sloučeniny sice vytvářející měřitelné ionty ale nejsou schopny v přírodě jako neutrální molekuly existovat. Tato

dvě pravidla zkoumají, zda se z navrženého elementárního složení dá sestavit stabilní neutrální molekula. Pravidla se vztahují k tzv. oktetovému pravidlu. (8)

Oktetové pravidlo říká, že pokud jsou zaplněny valenční orbitály 8 elektrony (tvorba oktetu), pak můžeme považovat sloučeninu za stabilní. Pravidlo ale neplatí vždy, striktně platí jen u prvků druhé periody. Prvky vyšších period jej často porušují. Prvky třetí periody mohou dosáhnout oktetu i v nestabilním stavu.

Pravidla vycházejí z tzv. RDB (Rings + Double Bonds) rovnice (11):

$$RDB = 1 + \frac{1}{2} \sum_i n_i (v_i - 2)$$

n_i = počet atomů prvku i , v_i = valence prvku i

Vztah platí jen tehdy, pokud každý prvek molekuly zaujímá nejnižší valenční stav a pokud hodnota RDB nepřesahuje 40. (8)

Lewisovo pravidlo říká, že pokud je hodnota RDB neceločíselná, pak se nejedná o stabilní molekulu, molekula má lichý počet elektronů. Toto pravidlo neplatí pro nitrososloučeniny, tj. sloučeniny s $-N=O$ skupinou. (8)

Seniorovo pravidlo říká, že pokud se nejedná o hypervalentní molekulu, což je molekula, která obsahuje alespoň jeden prvek nesoucí ve své valenční sféře více než 8 elektronů, pak by hodnota RDB neměla být záporná. (8)

3.5.2.4. Izotopická obálka, pattern

Izotopová obálka je tzv. otiskem prstu molekuly a patří při predikci elementárního složení mezi jedno z nejdůležitějších porovnávacích kritérií. Izotopová obálka navrženého elementárního složení musí odpovídat izotopové obálce naměřené látky.

Hmotnostní spektrometrie s vysokou rozlišovací schopností umožňuje rozlišit izotopy několika stupňů a poskytuje tak přesné informace o elementárního složení. (10) Při porovnání se zahrnují do izotopické obálky izotopy M , $M+1$ a $M+2$. Izotop $M+4$ mívá již tak malou intenzitu, že je téměř neměřitelná.

3.5.2.5. Poměr počtu atomů k počtu atomů uhlíku

Bylo zjištěno, že v molekulách platí zákonitosti v poměrech počtů atomů, ze kterých se sloučenina skládá, k počtu atomů uhlíku. Poměry, které stanovili Tobias Kind a Oliver Fiehn ve své publikaci (8), budou uvedeny v části testování.

3.5.2.6. Kontrola pravděpodobnosti výskytu prvků

V předcházejícím pravidle došlo k omezení výskytu prvků pouze na základě poměru počtu jednotlivých prvků k počtu uhlíku. K dalšímu omezení slouží pravidla pro vzájemný výskyt prvků a omezení maximálního rozsahu počtu jednotlivých atomů. Byly stanoveny prahové hodnoty pro kombinaci čtveřice čtyř základních prvků N, O, P, S a všechny kombinace trojic těchto prvků. Pravidla budou uvedena níže, v oddílu zabývajícím se testováním heuristických pravidel.

3.5.2.7. Tandemové, MS/MS pravidlo

Hmotnostní spektrometrie umožňuje tzv. MS/MS fragmentaci. V přístroji lze zvolit hmoty (m/z), které jsou po detekci rozštěpeny na fragmenty a následně znovu změřeny. Tyto fragmenty se pak promítnou i do naměřeného hmotnostního spektra. Pokud bylo predikované elementární složení iontu, pak fragmenty, které z něj vznikají, musí mít složení sestavitelné z prvků elementárního složení výchozího iontu. (10)

4. IMPLEMENTACE

4.1. Návrhová a implementační část programu

Reálné nasazení programu pro interpretaci elementárního složení představovalo jeho začlenění do systému Metabolite Mapper, který sehraává klíčovou roli ve zpracování dat naměřených v Laboratoři analytické chemie BC AV ČR. Jelikož systém Metabolite Mapper je poměrně rozsáhlý a není efektivní jej kvůli nové verzi modulu pro interpretaci elementárního složení neustále rekompilovat, byla vlastní implementace modulu umístěna do dynamicky linkované knihovny, jejíž časově invariantní rozhraní Metabolite Mapper pro její volání používá. Implementace optimalizačních metod pomocí heuristických pravidel, zmíněných v předchozích kapitolách, byla sumarizována do jediné třídy s názvem formulaGenerator.

4.2. Algoritmy generování sumárního vzorce

Generování jednotlivých validních sumárních vzorců je ovlivněno nastavením parametrů, které předpokládají určité vlastnosti (např. i empiricky zjištěné) látky.

Primárním problémem při generování teoretického sumárního vzorce je rapidní nárůst kombinací validních sumárních vzorců v závislosti na vzrůstající hmotě molekuly. Studie Tobiase Kinda a Olivera Fiehna (8) prokázaly, že se v biogenních látkách v závislosti na hmotě vyskytují jednotlivé chemické prvky pouze do určitého počtu a že existují určité zákonitosti v poměru počtu atomu uhlíku a jiných prvků. Tyto předpoklady lze velmi efektivně zahrnout do rozhodovací podmínky rekurzivní části generování sumárního vzorce a zabránit tak generování množství podstromů s chybnými vstupy již na začátku. Ostatní pravidla, která byla uvedena výše, mohou být aplikována až na navržený hmotnostně validní sumární vzorec, protože počty všech atomů obsažených v sumárním vzorci jsou podstatné pro správnost výpočtu.

4.2.1. Parametry algoritmu generování sumárního vzorce

Některá pravidla lze využít již v průběhu generování sumárního vzorce. Tato pravidla jsou brána jako parametry funkce generování sumárního vzorce:

- a. Rozdíl teoretické sumární hmotnosti vygenerovaného elementárního složení a hmotnosti naměřené. Validní rozdíl je omezený hmotnostním rozlišením přístroje.
- b. Minimální a maximální počty všech použitelných chemických prvků.
- c. Hodnoty poměrů počtů použitelných chemických prvků k počtu atomů uhlíku.

Další pravidla lze aplikovat až na finální vygenerovaný sumární vzorec:

- d. Procentuální shoda izotopické obálky naměřeného spektra s teoretickou.
- e. Hodnota z rovnice pro výpočet RDB.
- f. Kontrola sumárního vzorce pomocí Lewisova, Seniorova dusíkového pravidla.
- g. Je-li pro analyzovanou hmotu k dispozici MS/MS fragmentační spektrum, následuje ověřování, zda je možné sestavit, interpretovat, všechny vzniklé fragmenty MS/MS spektra pomocí prvků, které obsahuje vygenerovaný sumární vzorec. Interpretovatelnost spektra znamená, kolik % fragmentů z naměřeného MS/MS spektra lze interpretovat pomocí prvků navrženého elementárního složení.

4.2.2. Algoritmus generování sumárního vzorce

Vstupy:	Výstup:
Naměřená hmota	Seznam vygenerovaných validních elementárních složení
Izotopická obálka	
MS/MS spektrum	
Hmotnostní rozlišení přístroje	
Parametry generování	

Privátní proměnné předávající se jako parametry funkce generateFormula:

Množina chemických prvků P

Množina vygenerovaných sumárních vzorců F

Externí proměnná:

Množina validních sumárních vzorců FG

Začátek algoritmu: Množina P obsahuje všechny prvky získané z parametrů
Množiny F a FG jsou prázdné

Rekurzivní fáze algoritmu:

1. Vyber první prvek z množiny P, existuje-li, pokud ne, ukonči algoritmus.
2. Pro aktuálně předanou množinu vygenerovaných sumárních vzorců F proved' její duplikaci a spoj ji s dalším volným prvkem množiny P do všech možných kombinací pro vytvoření dalších částí generovaných vzorců.
3. Pro nově vytvořené kombinace duplikované množiny F aplikuj pravidla a až c a vyřaď nevalidní prvkové kombinace.
4. Pro kombinace, jejichž sumární hmotnost je menší než naměřená, duplikuj množinu P, odeber z ní aktuální prvek a pokračuj rekurzivně bodem 1.
5. Pro kombinace, jejichž hmota je přibližně rovna naměřené hmotě, aplikuj pravidla c až f a sumární vzorce, které jsou validní, ulož do množiny FG, ostatní sumární vzorce vymaž.

Iterativní fáze algoritmu:

Pro všechny validní formule spočítej celkové skóre validity formule.

$$Validita = \frac{RP}{|m_{\text{naměřená}} - m_{\text{vypočítaná}}|} \cdot simIP \cdot iMSMS$$

RP – rozlišovací schopnost přístroje, *simIP* – shoda izotopové obálky,
iMSMS – interpretovatelnost MS/MS spektra

V poslední fázi algoritmu je množina FG je seříděna vzestupně dle validity.

4.2.3. Algoritmus generování izotopické obálky (AGIO)

Vstupy:	Výstup:
Sumární vzorec	Izotopická obálka
Izotopy jednotlivých prvků	

1. Vyber aktuální prvek ze vzorce, existuje-li, jinak ukonči algoritmus,
2. Všechny jeho izotopy přidej do izotopické obálky následujícím způsobem: pokud již existuje ion spektra, který je hmotnostně shodný (dle rozlišení přístroje) s přidávaným, sečti jejich intenzity, jinak vlož do spektra nový ion.
3. Pokračuj bodem 1.

Výsledný tvar izotopické obálky je silně závislý na hmotnostním rozlišení přístroje, které se může měnit v závislosti na měřené hmotě.

4.2.4. Algoritmus výpočtu shody teoretické a naměřené izotopické obálky

Vstupy:	Výstup:
Naměřená izotopická obálka	Shoda teoretické izotopické obálky vycházející ze sumárního vzorce a naměřené izotopické obálky
Sumární vzorec	

1. Pomocí AGIO vygeneruj teoretickou izotopickou obálku pro sumární vzorec.
2. Porovnej teoretickou izotopickou obálku s naměřenou následujícím způsobem:
 - Proveď spárování jednotlivých iontů dle hmotnostního rozlišení přístroje.
 - Vypočti shodu mezi izotopickými obálkami dle následujícího vzorce.

$$Shoda = 100 \cdot \frac{CMinP}{CMaxP + CN}$$

CMinP - součet minimálních intenzit všech dvojic spárovaných iontů, *CMaxP* - součet maximálních intenzit všech spárovaných iontů, *CN* - součet všech intenzit nespárovaných iontů

4.2.5. Základní důležité datové typy a funkce

Primárním problémem implementace predikce elementárního složení byla definice vhodných datových struktur a návrh základních třídních metod. Detailní implementace v C++ je uvedena v příloze, následující odstavce obsahují výčet nejdůležitějších datových struktur a třídních metod v pseudokódu.

4.2.5.1. Datové struktury

Izotop (isotope), Prvek (element), Prvek (seznam<Izotop>, parametry), PouzityPrvek (*Prvek, parametry), SumarniVzorec (seznam<PouzityPrvek>)

4.2.5.2. Funkce

- countFormulaMass(&Formula) – vypočítá hmotnost sumárního vzorce
- countSpectralLikeness(&Spectrum1, &Spectrum2) – vypočítá podobnost dvou spekter – např. reálné a teoretické izotopické obálky

4.2.5.3. Specifikace třídy formulaGenerator

Použité třídní metody:

1. Nastavení a načtení parametrů generování sumárního vzorce (např. setRdbeValue, setMassResolution)
2. Otestování generovaného vzorce (checkFormula, generateFormulas)
3. Vygenerování izotopické obálky (generateIsotopePatternFromFormula)
4. Rekurzivní generování sumárního vzorce funkcí generate (vnější funkce) a generateFormulas (vnitřní funkce)

Veškeré metody jsou včetně okomentování umístěny ve zdrojovém kódu, který je součástí cd přiložených k bakalářské práci.

4.2.6. Implementace dynamicky linkované knihovny

Implementovanou třídu `formulaGenerator` bylo nutné načítat z dynamicky linkované knihovny (Win32 DLL) `massInterpreterLib.dll`. Vzhledem k tomu, že architektura knihoven DLL není objektová, ale pouze procedurální, bylo nutné implementaci třídy mírně poupravit. Využitím schopností pozdní vazby u abstraktních tříd lze dosáhnout toho, že metody třídy lze volat stejným způsobem, jako v případě dynamicky vytvořené její instance s tím, že tyto metody musí být v předkovi uvedeny jako abstraktní. Knihovna tudíž reálně exportuje pouze jedinou metodu, pomocí které se vytvoří instance třídy `formulaGenerator`.

Zdrojový kód knihovny je následující:

```
#include <vcl.h>
#include <windows.h>
#include "formulaGenerator.h"

extern "C"
{
    formulaGenerator* __stdcall __declspec(dllexport) createFormulaGenerator();
}

#if __WIN32
int WINAPI DllEntryPoint(HINSTANCE hinst, unsigned long reason, void*
    lpReserved)
{
    return 1;
}
#endif

#include "massInterpreterLibDll.h"

formulaGenerator* __stdcall __declspec(dllexport) createFormulaGenerator()
{
    return new _formulaGenerator();
}
```

Načtení knihovny lze provést v programu `Metabolite Mapper` pomocí `LoadLibrary`, získání přístupu k metodě pro vytvoření instance třídy `formulaGenerator` lze provést pomocí funkce `GetProcAddress`.

Příložený zdrojový kód knihovny je kompilovatelný jak pro 32 tak i 64 bitovou verzi OS Windows.

4.3. Optimalizace paralelního běhu programu

S vzrůstající hmotností molekuly exponenciálně narůstá počet hmotnostně korektních sumárních vzorců. Jedním ze způsobů, jak je možné tento fakt částečně kompenzovat, je vyšší výpočetní výkon. Hlavní idea celého řešení je založena na rozložení výpočetní práce na větší počet procesorů prostřednictvím procesů či vláken. Protože vlákna na rozdíl od separátního procesu sdílejí s mateřským procesem paměť programu, je v tomto případě vhodnější použít právě je.

Základní myšlenka byla tedy taková, že do určité hloubky od kořene stromu všech sumárních vzorců, bude pro kombinaci dané skupiny vzorců vytvořeno rekurzivně volané separátní vlákno, které otestuje určitý společný podstrom sumárních vzorců. Tato myšlenka byla opuštěna, neboť se ukázalo, že režie tvorby paralelních vláken v návaznosti na rekurzivní volání funkce pro generování kombinací sumárního vzorce má takovou režii, že je paralelní nárůst výpočetních prostředků degradován režii operačního systému po jejich správě. Bylo tedy nutné sáhnout k jiné variantě řešení.

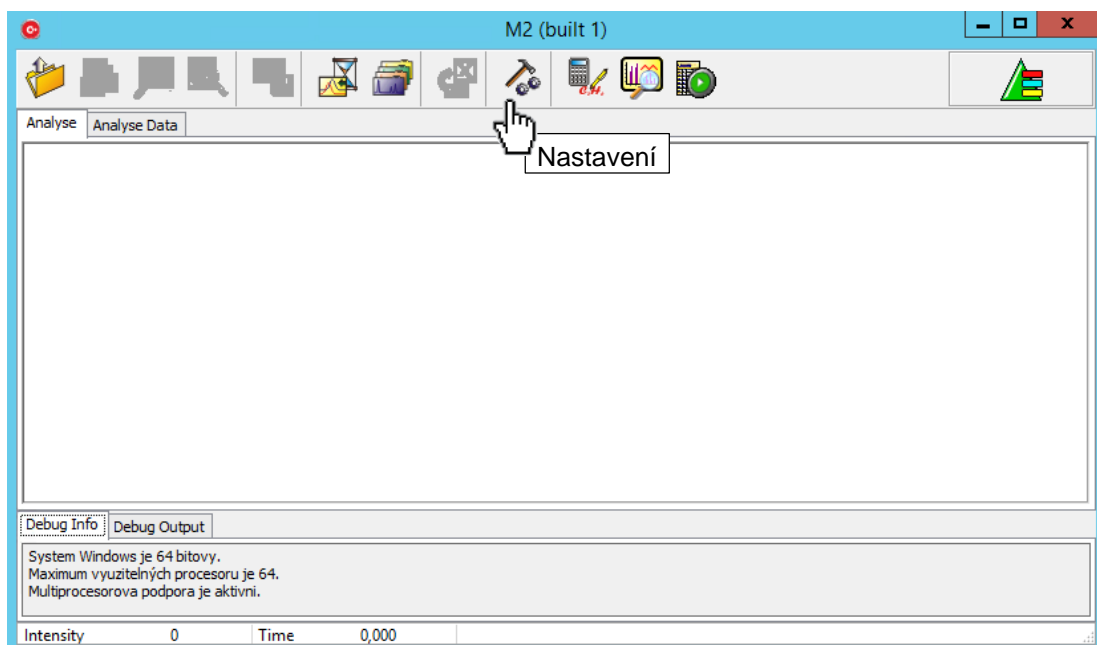
Bylo realizováno sériově paralelnímu řešení na bázi tzv. thread-poolu, což je skupina vláken, která se prostřednictvím sdílené instance určité datové struktury periodicky dotazuje o přidělení práce. Prvotní sériová část toho řešení spočívá v předgenerování počátků chemických vzorců (1-3 znaků, dle hmotnosti), které jsou pro další generování již rozděleny mezi vlákna thread-poolu.

Pro synchronizaci přístupu do společné datové struktury - zejména kvůli odstraňování již prošlých podstromů, bylo nutné použít vhodný mechanismus, což byl jednoduchý binární semafor – mutex. Pro tvorbu, synchronizaci a správu paralelních vláken byla použita multiplatformní knihovna s názvem Boost. Algoritmicky na úrovni jednotlivých vláken se jedná pouze o mírnou modifikace řešení z přechozí části.

4.4. Grafické prostředí

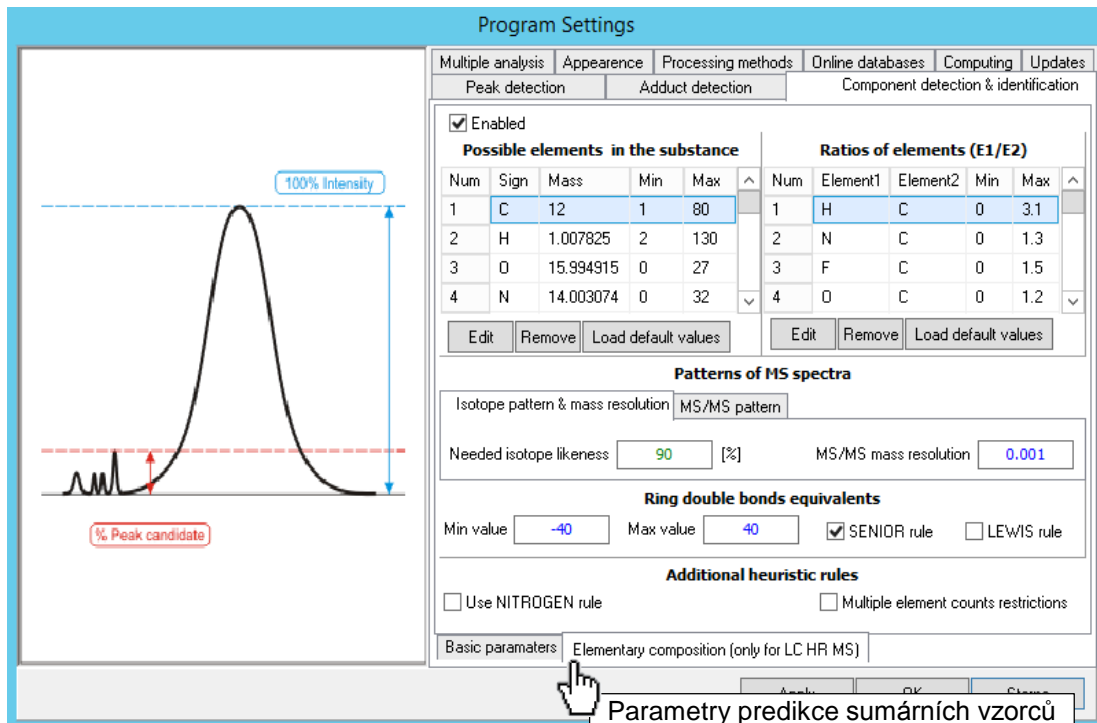
Grafické prostředí bylo vytvořeno ve vývojovém prostředí Embarcadero C++ Builder. Grafické rozhraní systému je velmi intuitivní a jednoduché pro práci.

Nastavení parametrů generování sumárních vzorců se provádí v hlavním okně mateřského programu Metabolite Mapper ikonou *Nastavení*.



Obrázek č. 13 – Vnitřní prostředí programu Metabolite Mapper

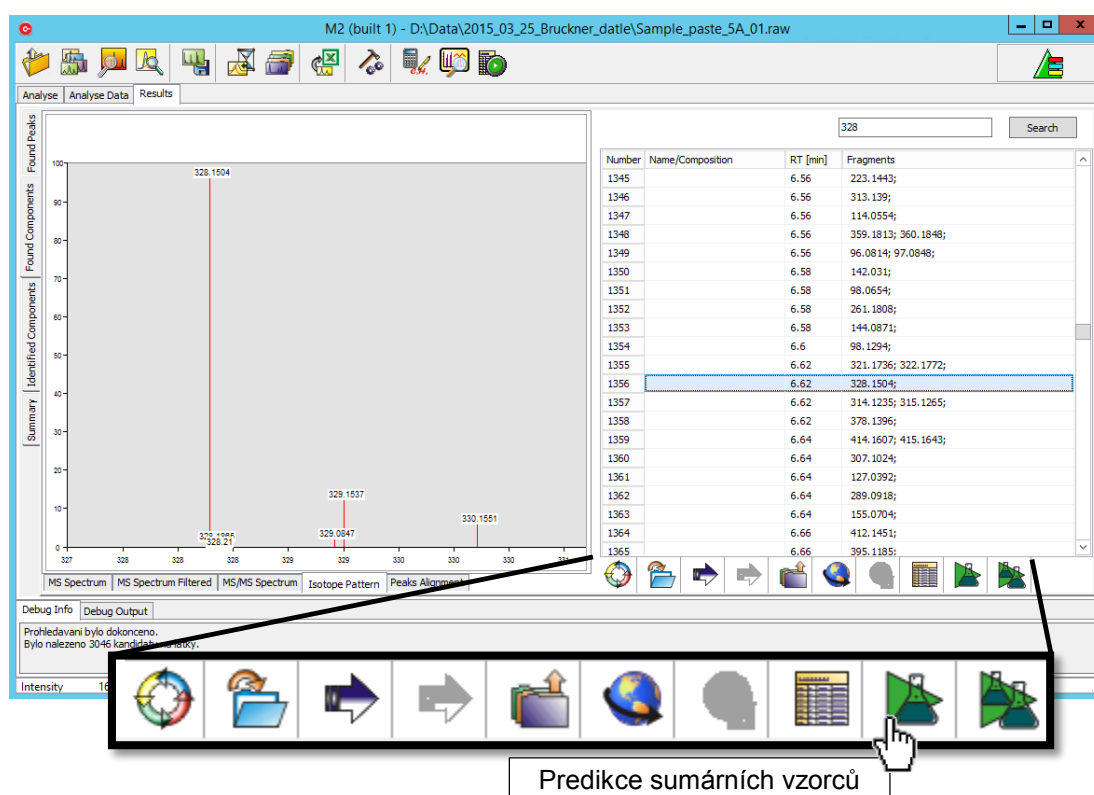
Pak se otevře dialogové okno, kde je možné na záložce *Detekce komponent a identifikace* v listu *Predikce elementárního složení* možné měnit parametry odpovídající implementovaným chemickým a heuristickým pravidlům.



Obrázek č. 14 – Dialogové okno parametrů programu Metabolite Mapper

Predikce sumárního vzorce je možná buď v rámci Metabolite Mapperu, poté co zpracuje analýzu z přístroje nebo je možné spustit samostatný modul massInterpreter.

Jak spustit predikci v rámci Metabolite Mapperu je zachyceno na obrázku 15. Jedná se o vnitřní prostředí Metabolite Mapperu po zpracování analýzy směsi sloučenin. V pravé části okna je výpis nalezených sloučenin ve směsi seřazených dle jejich retenčního času. Po výběru jedné látky ze seznamu, je možné si v levé části okna vybírat zobrazení podrobností o zvolené látce. Je zde i volba pro zobrazení izotopické obálky zvolné sloučeniny. V pravé části pod seznamem látek se nachází ikona pro spuštění predikce elementárního složení, které je předmětem mé bakalářské práce.



Obrázek č. 15 – Vnitřní prostředí Metabolite Mapperu po zpracování analýzy

Po kliknutí se otevře dialogové okno, v němž se po sléze vypíše seznam nalezených validních sumárních vzorců. Vzorce jsou řazeny dle celkového skóre.

Elementary composition prediction

Start generating Stop generating Thread count: 21

Number	Formula	Difference [M/z]	Rdbe	I-likeness [%]	MS/MS score [%]	Score
1	C ₂₀ H ₂₅ P ₂	2E-5 (-0.1ppm)	9.5	92.2	0	26.6
2	C ₁₀ H ₁₇ O _N 7F ₄	6E-5 (0.2ppm)	4	91	0	7.1
3	C ₁₄ H ₂₁ O ₆ N ₃	0.0001 (0.3ppm)	6	95.5	0	4.9
4	C ₁₃ H ₁₅ O _N 10	0.0001 (0.3ppm)	11.5	95.3	0	4.6
5	C ₁₇ H ₂₄ O ₃ F ₅	0.00011 (0.3ppm)	5.5	90.3	0	4
6	C ₁₁ H ₂₇ N ₄ F ₃	0.00013 (-0.4ppm)	1.5	97.6	0	3.8
7	C ₁₅ H ₁₈ O _N 4F ₃	0.00014 (-0.4ppm)	7.5	90.7	0	3.3
8	C ₁₂ H ₂₀ O _N F ₇	0.00018 (-0.5ppm)	0	94.1	0	2.7
9	C ₁₅ H ₂₄ N ₃ F ₂	0.00018 (0.6ppm)	6	95.7	0	2.6
10	C ₁₀ H ₂₄ O ₆ N ₄ P	0.00022 (-0.7ppm)	1.5	92.4	0	2.1
11	C ₉ H ₂₀ O ₆ N ₆ F	0.0003 (0.9ppm)	2.5	94	0	1.6
12	C ₁₂ H ₂₃ O ₃ N ₃ F ₂ S	0.00031 (1ppm)	2	93.6	0	1.5
13	C ₁₆ H ₂₈ N _P 3	0.00033 (-1ppm)	5	95.3	0	1.4

Select formula

Obrázek č. 16 – Dialogové okno s predikovanými sumárními vzorci

Druhý způsob, jak spustit predikci sumárních vzorců, je samostatný modul massInterpreter, který již není závislý na zpracování analýzy Metabolite Mapperem. Hlavní dialogové okno je víceméně společné pro okno s oknem nastavení Metabolite Mapperu s malými rozdíly. Sloučenina, kterou chceme analyzovat, resp. její spektrum, se načte po stisknutí tlačítka *Set*. Spektrum je možné načíst z *.xlsx* souboru nebo z proprietárního *raw* souboru výrobce přístroje.

4.5. Testování programu

K testování programu byla použita směs látek naměřených Orbitrapem Q Exactive Plus. Směs byla naměřena v deseti opakování a obsahovala 28 látek.

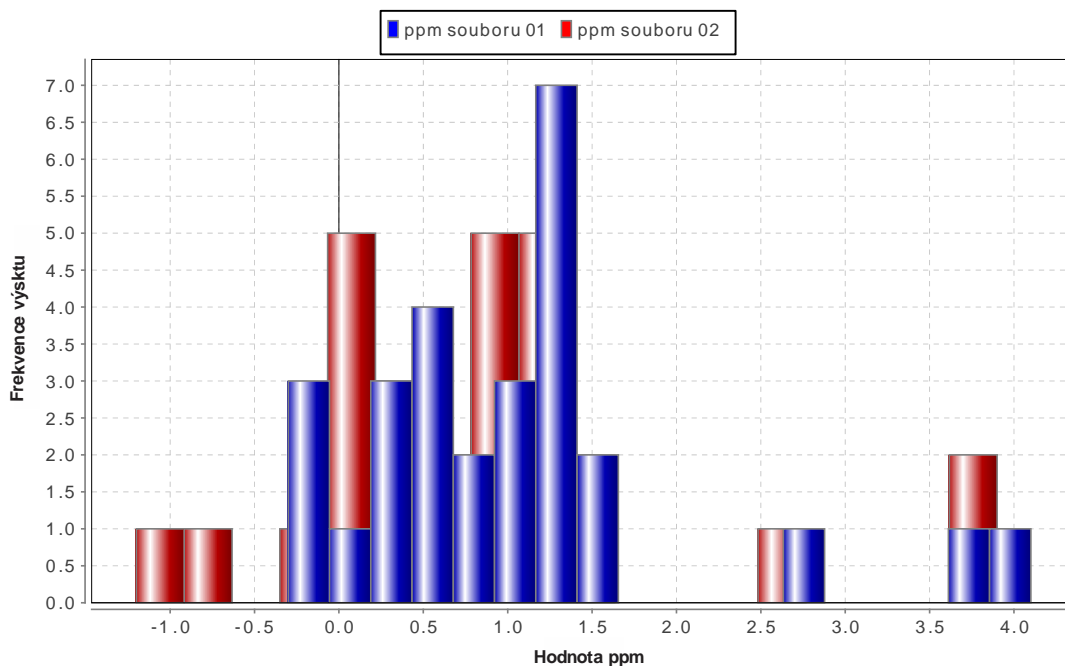
Predikcí navrženým programem byly stanoveny sumární vzorce pro monoizotopické hmotnosti látek směsi. V tabulce 5 je zachyceno srovnání 2 opakovaných měření. Kromě sumárních vzorců je v tabulce zachycena i správnost naměřené monoizotopické hmoty. V posledním sloupci je uveden hmotnostní rozdíl naměřený v prvním a druhém opakovaném měření.

Tabulka č.5 – Příklad vyhodnocované tabulky pro predikci elementárního složení

<i>Soubor 01</i>			<i>Soubor 02</i>			
Navržený sumární vzorec	ppm	<i>Monoizotopická hmotnost [Da]</i>	Navržený sumární vzorec	ppm	<i>Monoizotopická hmotnost [Da]</i>	Rozdíl hmot
C13H19O5N	0.3	269.1264	C13H19O5N	0	269.1263	0.0001
C10H11O3N	1.3	193.0741	C10H11O3N	1.1	193.0741	0
C4H3O3N	3.8	113.0117	C4H3O3N	3.8	113.0117	0
C9H13O5N	1.2	215.0796	C9H13O5N	0.9	215.0796	0
C8H13O4N	1.1	187.0847	C8H13O4N	0.8	187.0846	0.0001
C13H20O5N2	0.5	284.1374	C13H20O5N2	0.2	284.1373	0.0001
C10H17O6N3	0.2	275.1118	C10H17O6N3	0.4	275.1118	0
C5H5N5	1.5	135.0547	C5H5N5	1.3	135.0547	0
C7H8O2N2	0.9	152.0587	C7H8O2N2	1.4	152.0588	-0.0001
C11H15O5N	0.3	241.0951	C11H15O5N	0	241.095	0.0001
C8H12O4	1.3	172.0738	C8H12O4	1.4	172.0738	0
C12H14O4N2	0.6	250.0952	C12H14O4N2	-1.2	250.0951	0.0001
C6H11N4S	0	171.0704	C6H11N4S	0.2	171.0705	-0.0001
C7H15ON4S	0.5	203.0968	C7H15ON4S	0.8	203.0968	0
C9H16O4N2	1.5	216.1113	C9H16O4N2	0.5	216.1111	0.0002
C10H18O5N2	0.5	246.1217	C10H18O5N2	0.4	246.1217	0
C8H10O3	1.1	154.0632	C8H10O3	0.9	154.0631	0.0001
C12H17O4N	-0	239.1157	C12H17O4N	0.5	239.1159	-0.0002
C8H10O5	1.1	186.053	C8H10O5	0.9	186.053	0
C7H15N	4.1	113.1209	C7H15N	3.9	113.1209	0
C11H18O5N2	0.9	258.1218	C11H18O5N2	0.4	258.1217	0.0001
C13H17O5N3	1.2	295.1169	C13H17O5N3	0	295.1168	0.0001
C6H8O2N2	1.2	140.0587	C6H8O2N2	1.2	140.0587	0
C9H14O4N2	1.2	214.0956	C9H14O4N2	1.1	214.0956	0
C6H10O2N2	1.3	142.0744	C6H10O2N2	1.1	142.0744	0
C9H14O4N3S	-0	260.0704	C9H14O4N3S	-0.2	260.0705	-0.0001
C7H8O2	2.7	124.0528	C7H8O2	2.6	124.0528	0
C16H9ON3	-0	259.0745	C16H9ON3	-0.9	259.0743	0.0002

Během testování byla predikce sumárních vzorců téměř bezchybná.

Ke špatné predikci došlo pouze v jednom případě. Po prozkoumání příčiny bylo zjištěno, že v daném případě došlo ke stanovení nesprávné monoizotopické hmotnosti díky fluktuaci signálu z přístroje. Chyba přístroje se pohybovala v rozmezí -1 až 4 ppm, což bylo dostačující k správné predikci sumárních vzorců. Na obrázku 17 je zachycen histogram četností jednotlivých ppm pro dva srovnávané soubory.



Obrázek č. 17 – Histogram rozložení přesnosti měření

4.6. Optimalizace heuristických pravidel

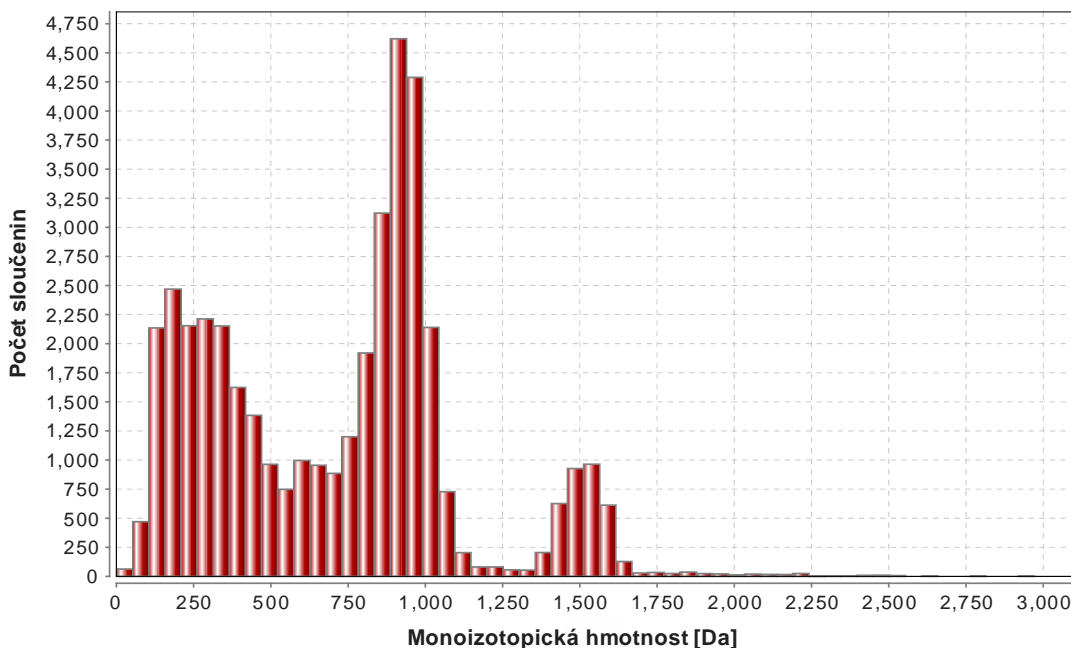
Jako testovací databázi jsem zvolila databázi *Human Metabolome Database*, *HMDB*. Database je volně stažitelná na stránkách <http://www.hmdb.ca/> ve formě datového balíčku obsahujícího samostatné html soubory pro každou látku. Databáze obsahuje zejména informace o nízkomolekulárních metabolitech, které se nacházejí v lidském těle. Jedná se o 41 471 záznamů, tzv. MetaboKaret. Každá MetaboKarta obsahuje více než 110 datových polí. Vzhledem k zaměření mé bakalářské práce jsem většinu polí nepotřebovala, využila jsem pouze údaje o sumárním vzorci, monoizotopické hmotnosti a zařazení do chemické taxonomie.

Pro získání těchto položek jsem naprogramovala jednoduchý syntaktický HTML analyzátor. Jako výstup jsem zvolila výstup do souboru formátu csv odděleného středníky. Tento soubor sloužil jako vstup pro program RapidMiner 5.3.013, ve kterém jsem posléze pracovala na vlastním testování databáze. Program RapidMiner 5.3.013 je volně stažitelný z internetových stránek <https://rapidminer.com/>.

Tobias Kind a Oliver Fiehn ve své publikaci (8) uvedli, že jejich trénovací databáze vykazovaly nerovnoměrné zastoupení sloučenin, a tím jsou jejich heuristická pravidla ovlivněna. Nicméně v diskuzi uvádí, že validační databází byla databáze

PubChem. Databáze kromě přírodních sloučenin obsahuje i sloučeniny připravené synteticky a v té době čítala 5 miliónu sloučenin a dle grafu v publikaci vykazovala normální rozložení s maximem na hmotnosti 435 Da.

Jako první jsem zpracovala rozložení sloučenin HMDB v závislosti na jejich monoizotopické hmotnosti databázi. HMDB vykazuje rovněž nerovnoměrné rozložení, má 3 lokální maxima, a to na hmotnosti 1 525 Da, 925 Da a 240 Da.



Obrázek č. 18 – Histogram rozložení četnosti sloučenin HMDB v závislosti na hmotě

Druhým krokem bylo zkontrolovat, zda platí pravidlo o maximálních počtech jednotlivých prvků zastoupených v molekule, zda souhlasí i jejich poměry k uhlíku a zda platí pravidlo o výskytu jednotlivých prvkových kombinací.

V tabulce 6 je uvedeno, jakých maximálních hodnot dosahují prvky v HMDB ve srovnání s maximy uvedenými v publikaci Tobiasse Kinda a Olivera Fiehna (8).

Tabulka č.6 – Srovnání maximálních hodnot zastoupení prvků v molekule

Interval [Da]	Databáze	C	H	N	O	P	S	F	Cl	Br	Si
1-500	DNP	29	72	10	18	4	7	15	8	5	
	Wiley	39	72	20	20	9	10	16	10	4	8
	HMDB	34	70	10	18	4	7	17	10	4	2
501-1000	DNP	66	126	25	27	6	8	16	11	8	
	Wiley	78	126	20	27	9	14	34	12	8	14
	HMDB	70	128	15	31	8	7	17	12	7	2
1001-2000	DNP	115	236	32	63	6	8	16	11	8	
	Wiley	156	180	20	40	9	14	48	12	10	15
	HMDB	108	188	28	61	8	8	17	12	7	2
2001-3000	DNP	162	208	48	78	6	9	16	11	8	
	HMDB	144	240	28	91	8	8	17	12	7	2

Rozložení sloučenin HMDB je nerovnoměrné, ale zastoupení prvků v molekulách v jednotlivých intervalech molekulových hmotností víceméně odpovídá databázím Wiley (John Wiley & Sons database) a DNP (Dictionary of Natural Products).

Tabulka 7 ukazuje srovnání poměrů prvků vzhledem k uhlíku. Jedná se o vyhodnocení na základě zúženého intervalu hmotností látek databáze na 30 – 1500 Da. Jsou zde uvedeny hodnoty pro 2 rozsahy, které zahrnují buď 99,7 % všech sloučenin databáze, nebo 99,99 %.

Tabulka č.7 – Srovnání intervalů hodnot poměrů počtu prvků k počtu uhlíků

Poměr	Databáze	99.7 %	99.99 %
H/C	Wiley	0,2 – 3,1	0,1 – 6,0
	HMDB	0,1 – 4,0	0,1 – 5,0
F/C	Wiley	0,0 – 1,5	0,1 – 6,0
	HMDB	0,0 – 0,7	0,0 – 1,9
Cl/C	Wiley	0,0 – 0,8	0,0 – 2,0
	HMDB	0,0 – 1,0	0,0 – 1,5
Br/C	Wiley	0,0 – 0,8	0,0 – 2,0
	HMDB	0,0 – 0,5	0,0 – 0,7
N/C	Wiley	0,0 – 1,3	0,0 – 4,0
	HMDB	0,0 – 1,0	0,0 – 2,0
O/C	Wiley	0,0 – 1,2	0,0 – 3,0
	HMDB	0,0 – 3,3	0,0 – 4,6
P/C	Wiley	0,0 – 0,3	0,0 – 2,0
	HMDB	0,0 – 0,7	0,0 – 1,0
S/C	Wiley	0,0 – 0,8	0,0 – 3,0
	HMDB	0,0 – 1,3	0,0 – 2,0

Hodnoty jednotlivých poměrů počtu prvků k počtu uhlíku ve sloučenině z HMDB se nijak výrazně neliší od hodnot z databází použitých v publikaci.

Posledním testem databáze bylo otestování pravidel vzájemného výskytu prvků a poměru počtu jejich atomů prvků ve sloučenině.

Tabulka č.8 – Tabulka heuristických pravidel

Heuristické pravidlo	Vstupní podmínky				Výstupní podmínky			
	N	O	P	S	N	O	P	S
1.	> 1	> 1	> 1	> 1	< 10	< 20	< 4	< 3
2.	> 3	> 3	> 3		< 11	< 22	< 6	
3.		> 1	> 1	> 1		< 14	< 3	< 3
4.	> 1		> 1	> 1	< 4		< 3	< 3
5.	> 6	> 6		> 6	< 19	< 14		< 8

Tato heuristická pravidla při testování nebyla potvrzena z důvodu nesplnění vstupních podmínek heuristických pravidel. Z testovaných 41 471 sloučenin by tato pravidla označila jako vyhovující sloučeniny řádově pouze jednotky. Podmínce prvního pravidla, tj. že pokud je ve sloučenině alespoň jeden N, jeden, O, jeden fosfor a jedna S, vyhovovalo pouze 12 sloučenin a žádná z nich nesplňovala parametry druhé poloviny pravidla. Vstupním podmínkám druhého pravidla vyhovovalo pouze 20 sloučenin a výstup splnilo 9. U třetího pravidla splnily obě podmínky 2 sloučeniny z 12. U čtvrtého pravidla splnilo vstupní podmínku 12 sloučenin, ale výstupní podmínce již nevyhovovala žádná. U pátého pravidla nebyla v celé databázi nalezena sloučenina, která by vyhovovala vstupní podmínce pravidla.

To, že tato pravidla jsou porušována sloučeninami existujícími běžně v přírodě, zmínil ve své publikaci i Tomáš Pluskal (10). Poukázal na to, že pravidlo vyžaduje poměr kyslíku a uhlíku ve sloučenině menší než 1,20, což porušují např. běžně známé ATP (adenosin trifosfát) či fruktóza-1,6-bisfosfát. I v HMDB je obecně poměr kyslíku a uhlíku o něco vyšší, viz tabulka 7. Poměr O/C > 1,20 má ale jen 264 sloučenin.

5. Zhodnocení výsledků a závěr

Základním cílem mé bakalářské práce bylo vytvořit aplikaci pro predikci elementárního složení z dat z hmotnostních spektrometrů s vysokým rozlišením.

Po prostudování literatury týkající se predikce elementárního složení jsem

vyvinula program pro predikci sumárních vzorců, do kterého jsem implementovala heuristická pravidla uvedená v sekci výše. Analýzou sloučenin databáze *Human Metabolome* jsem odhalila několik rozdílů oproti stanoveným pravidlům. Souhrn je uveden rovněž v sekci výše. Tyto změny jsem promítla do implementace programu a následně otestovala, zda došlo v predikci vzorců ke zlepšení, tj. zda byl program schopen stanovit validní elementární složení v rychlejším čase. Přestože se nejednalo o příliš razantní odchylky od hodnot, které uvedl ve své publikaci Tobias Kind a Oliver Fiehn (8), na výsledném čase zrychlení predikce se to projevilo. Průměrný čas predikce všech validních sumárních vzorců pro testovací směsi 28 látek se z původních 34ms vylepšil na 24ms. Vzhledem k výpočetní náročnosti stanovení sumárních vzorců jsem implementovala sériově paralelní zpracování pomocí thread-poolu. Jako další možný vývoj programu spatřuji právě v možnostech optimalizace programu pro pokročilé paralelní architektury. Nemalý vliv by mohla mít i hlubší analýza heuristických pravidel. Plné pochopení a následné vypracování vylepšených pravidel vyžaduje nicméně nastudování další odborné chemické literatury.

6. SEZNAM CITOVANÉ LITERATURY

- (1) HOFFMANN, Edmond De., STROOBANT Vincent, *Mass Spectrometry: Principles and Applications*, 3rd ed., 2007, ISBN 978-0-470-03310-4.
- (2) SARGENT, Mike (Ed.), *Guide to achieving reliable quantitative LC-MS measurements*, RSC Analytical Methods Committee, 2013, ISBN 978-0-948926-27-3.
- (3) SMITH, R. Martin, *Understanding mass spectra: a basic approach*, 2nd ed., 2004, ISBN 047142949x.
- (4) FRIEDECKÝ, David, Lemr, Karel, *Hmotnostní spektrometrie - zdroj analytických informací*, Klinická biochemie a metabolismus, 2012, roč. 20, č. 4, s. 210-215. ISSN: 1210-7921.
- (5) HERBERT, Christopher G., JOHNSTONE Robert Alexander Walker, *Mass spectrometry basics*. Boca Raton: CRC Press, 2003, ISBN 0849313546.
- (6) GROSS, Jürgen H, *Mass spectrometry: a textbook*, Berlin: Springer, 2004, ISBN 3540407391.
- (7) COMMISSION ON ISOTOPIC ABUNDANCES AND ATOMIC WEIGHTS.

- Isotopic Abundances [online]. [cit. 2015-04-22]. Dostupné z: <http://www.ciaaw.org/>
- (8) KIND, Tobias, FIEHN Oliver, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spektrometry, *BMC Bioinformatics*, 2007, vol. 8, issue 1, DOI: 10.1186/1471-2105-8-105.
- (9) EIDHAMMER, Ingvar, BARSNES, Harald, EIDE, Geir Egil, MARTENS, Lennart, *Computational and statistical methods for protein quantification by mass spectrometry*, 2013, ISBN 1119964008.
- (10) PLUSKAL, Tomáš, Taisuke UEHARA a Mitsuhiro YANAGIDA. Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Analytical Chemistry*. 2012, vol. 84, issue 10, s. 4396-4403. DOI: 10.1021/ac3000418.
- (11) MALLET, A a S DOWN. *Dictionary of mass spectrometry*. Hoboken, NJ: Wiley, 2009, vii, 174 p. ISBN 9780470027615
- (12) The NIST Reference on Constants, Units and Uncertainty. NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. *Fundamental Physical Constants --- Complete Listing* [online]. [cit. 2015-04-1]. Dostupné z <http://physics.nist.gov/cuu/Constants/Table/allascii.txt>
- (13) MAKAROV, Alexander. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry*, 2000, vol. 72, issue 6, s. 1156-1162, DOI: 10.1021/ac991131p.
- (14) HU, Qizhi, Robert J. NOLL, Hongyan LI, Alexander MAKAROV, Mark HARDMAN a R. Graham COOKS. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*. 2005, vol. 40, issue 4, s. 430-443. DOI: 10.1002/jms.856.