

## ZÁKLADY INTERPRETACE HMOTNOSTNÍCH SPEKTER

LUDOVÍT ŠKULTÉTY, JIŘÍ NOVÁK a VLADIMÍR HAVLÍČEK

Mikrobiologický ústav AV ČR, Videňská 1083, 142 20  
Praha 4  
vlhavlic@biomed.cas.cz

Došlo 13.11.19, přijato 16.12.19.

Klíčová slova: fragmentace, mechanismus, pravidlo, isotopová obálka, interpretační postup, nomenklatura iontů, jemná isotopová struktura

### Obsah

1. Úvod
2. Elektronová ionizace a některá fragmentační pravidla
3. Stablní isotopy
4. Metoda *per analogiam*
5. *De novo* interpretace spekter biopolymerů s omezeným počtem stavebních kamenů
6. Softwarové nástroje k usnadnění interpretace hmotnostně spektrometrických dat

### 1. Úvod

Hned v úvodu je třeba zdůraznit, že jediný a zaručeně fungující obecný postup pro úspěšnou interpretaci hmotnostních spekter neexistuje. Hmotnostní spektrum není fyzikální konstantou a jeho informační obsah závisí na použité ionizační technice, typu experimentu i použitého hmotnostního analyzátoru. V úvodní části referátu se zaměříme na interpretace spekter jednoduchých organických látek získaných ionizací elektrony, které poskytují ve spektrech kladných iontů částice s lichým počtem elektronů, radikál-kationty. Dále budou zmíněna některá pravidla pro fragmentaci polárních látek poskytujících molekulové ionty se sudým počtem elektronů. Referát bude uzavřen interpretací produktových spekter peptidů jako vybraných představitelů biopolymerů.

Před samotným měřením vzorku a interpretací jeho hmotnostních spekter je třeba vzít do úvahy mj. historii vzorku (zdroj, toxicita, očekávaná molekulová hmotnost), fyzikální vlastnosti vzorku (skupenství, rozpustnost, stabilita) a jeho chemickou povahu (organická nebo anorganická čistá látka nebo směs). Někdy je třeba se smířit s faktem, že hmotnostní spektrometrie se na řešení daného analytického problému vůbec nehodí. Příkladem může být

analýza střevní mikroflóry, která je doménou molekulárních sekvenačních technik. Jinou, pro hmotnostní spektrometrii komplikovanou problematikou, je rozlišení některých typů izomerů či nestabilních látek.

Experimentální strategie se tedy volí podle zadání a povahy vzorku. Zvláště je třeba mít na mysli, co je cílem měření (kvalitativní nebo kvantitativní analýza), jak zní zadání analytického problému (u výstupu z organické syntézy se potvrzují očekávané produkty, jindy se řeší struktura zcela neznámé přírodní látky) a jak detailní je požadovaná strukturní informace (od prokázání přítomnosti funkčních skupin po určení konfigurace molekuly, od sekvence stavebních kamenů v biopolymeru po zjištění jeho konformace). Navržená chemická struktura by se měla potvrdit alternativní analytickou technikou, podpořit kvantově-chemickým výpočtem, případně její spektrální chování porovnat se syntetizovaným standardem.

### 2. Elektronová ionizace a některá fragmentační pravidla

Nepolární až středně polární a tepelně stabilní nízkomolekulové látky lze s výhodou ionizovat elektrony (EI, electron ionization) nebo fotoionizací. Při tvorbě této podkapitoly autoři zvažovali, zdali mají detailněji zpracovávat informace o typických fragmentačních cestách jednotlivých skupin látek, od alkanů přes aminy, alkoholy, atd. Současný profil českých a slovenských uživatelů elektronové ionizace nás nakonec vedl k variantě odkazu na původní monografii, kde zvláště fanoušci spojení hmotnostní spektrometrie s plynovou chromatografií mohou nalézt detailní informace<sup>1–3</sup>.

Vzhled EI spekter závisí na řadě experimentálních parametrů (teplota, tlak, typ analyzátoru). Pro interpretaci spekter jsou významná měření přesné hmotnosti i znalost základních fragmentačních mechanismů. Při interpretaci spekter kladných iontů je hlavním vodítkem molekulový radikál-kation (částice s lichým počtem elektronů, OE, odd electron, odpovídající celé analyzované molekule, přesněji této molekule chybí elektron odtržený při ionizaci). Pomineme-li vyšší stabilní isotopy molekulového klastru, obvykle to bývá ion s nejvyšší hmotností v EI spektru, jehož hodnota  $m/z$  musí korelovat s fragmentovými ionty. Intenzita molekulového iontu v EI spektru má přímý vztah ke struktuře stanovované látky, někdy nemusí být přítomen vůbec. V řadě případů při zvyšování hmotnosti podobné skupiny látek stejné polaritě intenzita jejich molekulových iontů klesá. Obecně ale platí, že s rostoucím počtem násobných vazeb a cyklů v molekule intenzita píku molekulového iontu roste. Ekvivalent počtu dvojných vazeb

a cyklů  $h$  (rdb, rings+double bonds equivalent) ve struktuře organické látky je dán vztahem (1):

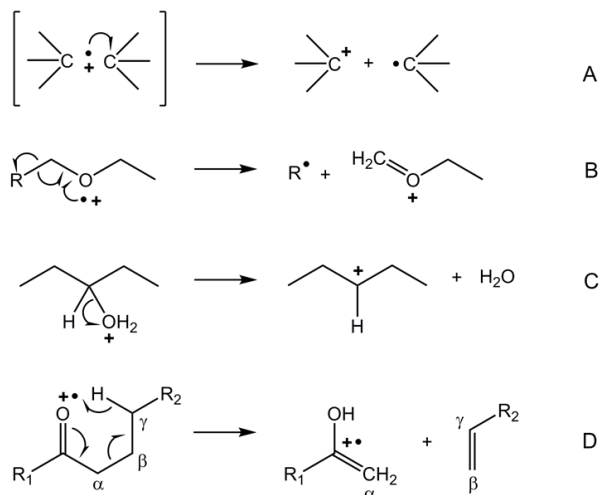
$$h = x - \frac{1}{2}y + \frac{1}{2}z + 1 \quad (1)$$

kde proměnné  $x$ ,  $y$ ,  $z$  představují počty atomů v sumárním vzorci  $C_xH_yN_zO_n$  a platí, že uhlíku dále odpovídá křemík, vodíku odpovídají F, Cl, Br, I, dusíku odpovídá fosfor a kyslíku síra. Např. ion  $C_2H_5OH^+$  má ekvivalent počtu dvojných vazeb  $2 - \frac{1}{2} \cdot 6 + \frac{1}{2} \cdot 0 + 1 = 0$ , zatímco ion  $Ph-CO^+$  má  $7 - \frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 0 + 1 = 5,5$ . Pokud  $h$  je celé číslo, pak příslušný ion má lichý počet elektronů. Jestliže hodnota  $h$  končí jednou polovinou, příslušný ion má sudý počet elektronů. Trojná vazba odpovídá dvěma dvojným. Občas ale platí, že „výjimka potvrzuje pravidlo“, a proto se vždy doporučuje konzultovat studovanou skupinu látek s příslušnou literaturou<sup>4</sup>.

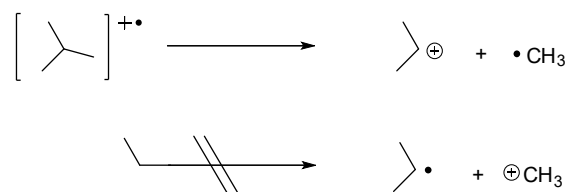
Přehledné práce a monografie zdůrazňují mezi fragmentačními mechanismy zejména sigma, alfa i beta štěpení jedné vazby, štěpení více vazeb, *retro*-Diels-Alderovu reakci<sup>5</sup>, McLaffertyho přesmyky<sup>6</sup>, orthoefekty<sup>7</sup> a celou řadu dalších, které jsou specifické pro konkrétní skupinu organických látek<sup>8</sup>. Obr. 1. uvádí základní fragmentační mechanismy. Stereochemické problémy jsou pak zpracovány v pracích<sup>8,9</sup>.

Při interpretaci spektra lze také použít některá z chemických pravidel. Prvním z nich je *dusíkové pravidlo*, které lze formulovat takto: Pokud studovaná sloučenina obsahuje ve své molekule žádný nebo *sudý* počet dusíkových atomů, pak její molekulový ion musí mít *sudou* hodnotu  $m/z$ , při lichém počtu lichou.

Mezi molekulovým iontem a ionty fragmentovými musí existovat logický vztah podle eliminovaných neutrálních molekul nebo odštěpených radikálů. Hmotnostní rozdíly mezi možným molekulovým iontem a významnými



Obr. 1. Vybrané příklady fragmentačních mechanismů ve spektrech elektronové ionizace. Štěpení  $\sigma$ -vazby mezi dvěma atomy uhlíku (A), fragmentace iniciovaná radikálovým centrem (B), fragmentace iniciovaná nábojovým centrem (C) a přesmyk vodíku z polohy  $\gamma$  (McLaffertyho přesmyk), zde uvedený v ketonu (D)



Obr. 2. Fragmentace 2-methylpropanu. Relativní intenzity fragmentových iontů  $m/z$  43 a  $m/z$  15 jsou 100, resp. cca 10 %

fragmentovými ionty v intervalech  $m/z$  4–14 a 21–25 jsou pro běžné typy organických sloučenin velmi nepravděpodobné (illogical losses). Naopak obvyklými ztrátami jsou odštěpení labilního vodíkového radikálu ( $-1$ ), ztráta alkylových radikálů ( $-15, -29, -43, -57 \dots$ ), eliminace neutrálních molekul  $H_2O$ ,  $HF$ ,  $CO$ ,  $C_2H_4$ ,  $H_2S$ ,  $HCl$  ( $-18, -20, -28, -34, -36$ ), dále jsou běžné současné eliminace vody a ethyleny z alifatických aldehydů ( $-18-28$ ) apod.

Další pravidlo je dle *Stevensona a Audiera*. Podle něj bude kladný náboj lokalizován na fragmentu s nejnižší ionizační energií. Příkladem je přednostní tvorba methylradikálů z radikál-kationtu 2-methylpropanu (obr. 2). Jiným pravidlem je *pravidlo ztráty největšího alkyly*.

Způsob fragmentace molekuly závisí i na lokalizaci náboje, tedy místě ve struktuře měřené látky, kde chybí elektron a fragmentace může mít různé varianty (např. fragmentace řízená nábojovým nebo radikálovým centrem). Velmi často se setkáváme s konkurenčními fragmentačními řadami, a to přímým štěpením jednoduchých vazeb, alfa-štěpením i celou řadou přesmyků. Často stačí malá strukturní změna, která se projeví v dramaticky odlišném profilu spektra.

### 3. Stabilní isotopy

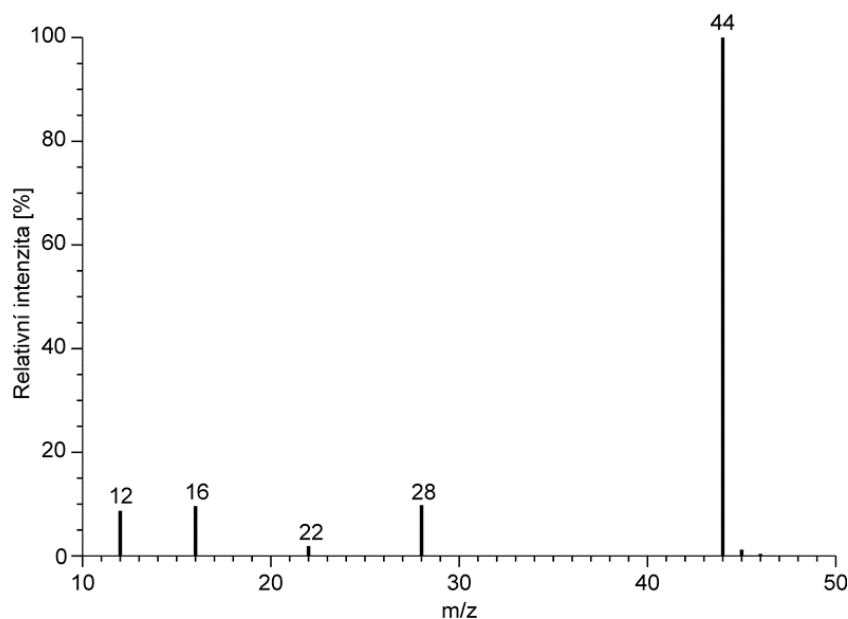
V hmotnostním spektru doplňuje molekulový radikálion ( $M^{\bullet}$ ) jeho isotopová obálka, a to dalšími ionty na hodnotách  $m/z$   $M+1$ ,  $M+2$ ,  $M+3$  apod., které obsahují ostatní stabilní isotopy  $^{13}C$ ,  $^{15}N$ ,  $^{18}O$ ,  $^{34}S$  atd. Tyto se spojí v signálech  $m/z$   $M+1$ ,  $M+2$ ,  $M+3$  v nízkorozlišených spektrech bez ohledu na hmotnostní defekty jednotlivých nuklidů (viz dále). Isotopovou obálku molekulových i fragmentových iontů lze použít k získání informací o jejich elementárním složení.

Pro ilustraci je uvedeno EI spektrum oxidu uhličitého (obr. 3). V jeho spektru dominuje molekulový radikálkation ( $m/z$  44) a dále jsou zřejmé i signály jednotlivých atomů ( $m/z$  12,  $m/z$  16). Kromě oxidu uhelnatého ( $m/z$  28) je patrná i tvorba dvakrát nabitého molekulového iontu ( $m/z$  22).

Ionty uvedené v tomto spektru mají nominální hmotnosti, které byly získány součtem nominálních hmotností v nich zastoupených prvků. Nominální hmotnosti získáme zaokrouhlením monoisotopových hmotností v přírodě nejvíce zastoupených stabilních nuklidů (tab. I). Tyto no-

Tabulka I  
Monoisotopové a průměrné hmotnosti vybraných biogenních prvků

Prvek	Hlavní stabilní isotop	Monoisotopová hmotnost	Nominální hmotnost	Průměrná hmotnost
H	$^1\text{H}$	1,007825	1	1,00794
C	$^{12}\text{C}$	12,00000	12	12,011
N	$^{14}\text{N}$	14,003074	14	14,00674
O	$^{16}\text{O}$	15,994915	16	15,9994
S	$^{32}\text{S}$	31,972072	32	32,066
Cl	$^{35}\text{Cl}$	34,9689	35	35,4527



Obr. 3. EI hmotnostní spektrum oxidu uhličitého

minální hodnoty nezohledňují tzv. hmotnostní defekt, viz dále.

V této tabulce mají všechny uvedené prvky průměrnou hmotnost větší než je hmotnost hlavního isotopu. Jinak je tomu např. u isotopů lithia. Zatímco  $^7\text{Li}$  atom obsahuje 3 protony, 4 neutrony a 3 elektrony, jeho nuklid  $^6\text{Li}$  má neutrony jen tři, čímž snižuje průměrnou atomovou hmotnost na 6,941. Přirozeně se vyskytující isotopy sice na jednu stranu komplikují hmotnostní spektra, na druhou stranu ovšem zpřesňují informace o elementárním složení iontů. Zvláště důležité je zastoupení uhlíku, jehož méně běžný isotop  $^{13}\text{C}$  je zastoupen v průměru 1,1 % všech

nuklidů uhlíku. Zastoupení isotopů v příslušném iontu lze spočítat podle koeficientů binomických rozvoje. V případě uhlíku se bude jednat o rozvoje  $(^{12}\text{C}+^{13}\text{C})^x$ , čili  $(99a+1b)^1$  u jednoho atomu uhlíku,  $(^{12}\text{C}+^{13}\text{C})^2$  u dvou atomů uhlíku,  $(^{12}\text{C}+^{13}\text{C})^3$  u tří atomů atd. Tab. II uvádí zaokrouhlené intenzity píků v isotopové obálce příslušného iontu v závislosti na počtu přítomných atomů uhlíku.

U malých počtů atomů uhlíku není přírůstek vyššího isotopu tak zásadní, jako u některých jiných prvků (tab. III). Chlor se v přírodě vyskytuje ve formě dvou hlavních isotopů  $^{35}\text{Cl}/^{37}\text{Cl}$  se vzájemným poměrem přibližně 3:1, zatímco  $^{79}\text{Br}/^{81}\text{Br}$  dosahuje poměru téměř 1:1. Přítom-

Tabulka II  
Isotopová obálka v závislosti na počtu atomů uhlíku

$\text{C}_1$ :	99 % $^{12}\text{C}$	1 % $^{13}\text{C}$		
$\text{C}_2$ :	98 % $^{12}\text{C}^{12}\text{C}$	2 % $^{12}\text{C}^{13}\text{C}$	0,01 % $^{13}\text{C}^{13}\text{C}$	
$\text{C}_3$ :	97 % $^{12}\text{C}^{12}\text{C}^{12}\text{C}$	3 % $^{12}\text{C}^{12}\text{C}^{13}\text{C}$	0,04 % $^{12}\text{C}^{13}\text{C}^{13}\text{C}$	$10^{-4}$ % $^{13}\text{C}^{13}\text{C}^{13}\text{C}$

## Tabulka III

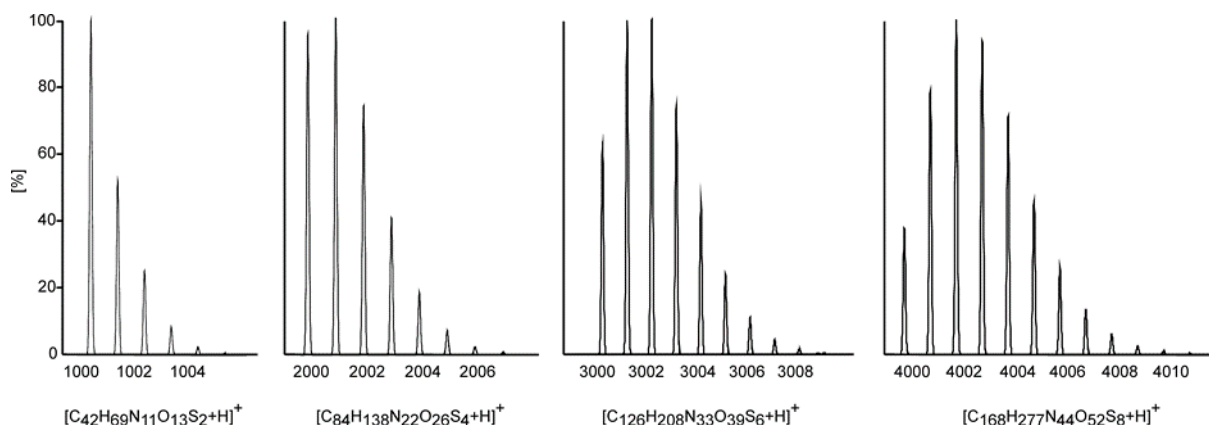
Příspěvky atomů Cl a Br do isotopových profilů. Poměr isotopů se projeví v koeficientech před sčítacími obsahujícími proměnné  $a/b$ , exponent pak odpovídá počtu atomů

Br <sub>2</sub> :	$(a+b)^2$	$a^2+2ab+b^2$	(1:2:1)
Br <sub>3</sub> :	$(a+b)^3$	$a^3+3a^2b+3ab^2+b^3$	(1:3:3:1)
Cl <sub>2</sub> :	$(3a+b)^2$	$9a^2+6ab+b^2$	(9:6:1)

nost být jediného takového prvku v nízkomolekulové struktuře je z hmotnostního spektra zřejmá na první pohled a může být použita ke správnému odhadu sumárního vzorce příslušného iontu.

S rostoucím počtem atomů se nejvyšší pík isotopové obálky posouvá k vyšším hmotnostem. Simulaci tohoto stavu pro hypotetické peptidy ilustruje obr. 4. Mějme na paměti, že některé další prvky mají vyšší počet stabilních isotopů (Sn, Pt apod.) a hmotnostní spektra odpovídajících organometalických sloučenin pak mohou být velmi komplexní.

menšími než je hmotnost elektronu ( $9,109 \cdot 10^{-31}$  kg). Uvědomíme-li si, že hmotnosti neutronu ( $1,675 \cdot 10^{-27}$  kg) a protonu ( $1,673 \cdot 10^{-27}$  kg) jsou odlišné a současně jsou v jádře vázány energií, které odpovídá určitý úbytek hmotnosti (celková hmotnost jádra se nerovná prostému součtu hmotností přítomných nukleonů), nabízí se využití hmotnostních defektů jednotlivých nuklidů (tab. I, obr. 6) k určení sumárního vzorce iontu. Analyzátoři s extrémně vysokou rozlišovací schopností navíc přidávají informaci o jemné isotopové struktuře, jejíž profil příslušný sumární vzorec potvrzuje (obr. 7).

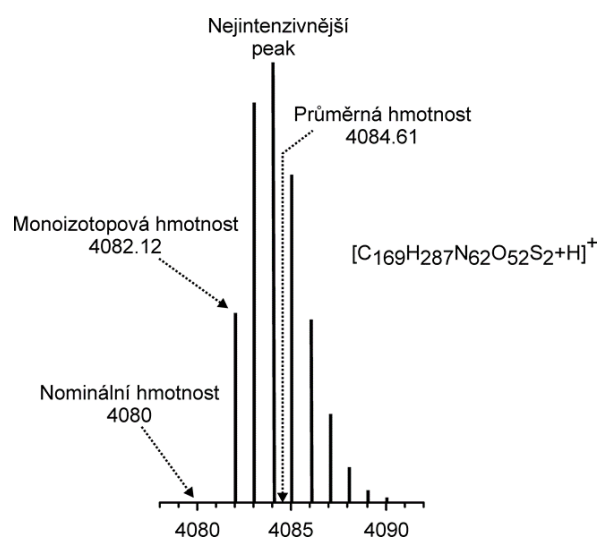


Obr. 4. Simulovaná isotopová obálka peptidu v závislosti na zvyšující se molekulové hmotnosti

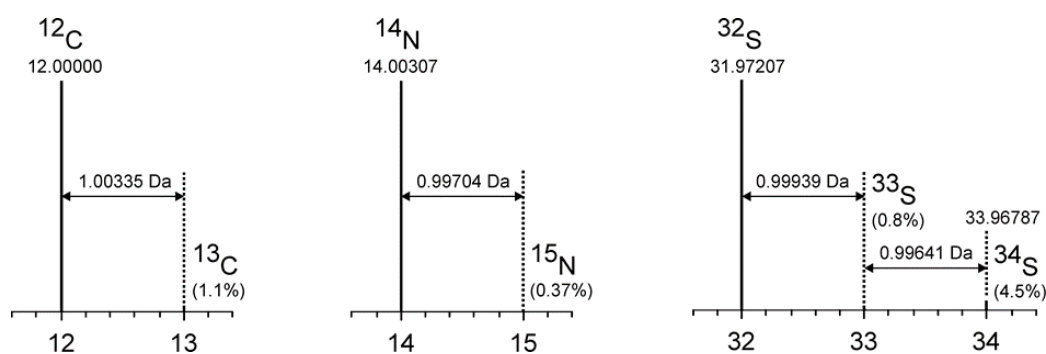
U velkých biomolekul v řádu desítek kilodaltonů nemusí být molekulový monoisotopový signál patrný vůbec, jindy signál nejintenzivnějšího píku nemusí představovat průměrnou hmotnost iontového druhu (obr. 5).

Rozlišení isotopových obálek biomolekul je možné pouze vysoce rozlišujícími hmotnostními spektrometry. Podle definice je rozlišovací schopnost dána poměrem  $m/\Delta m = 1000$ . Např. k rozlišení běžného dubletu glutamin/lysin u peptidových iontů, kde  $\Delta m = 0,0364$  u, je potřeba 36 ppm přesnost na hmotnosti 1000. U iontů s hmotností 10 000 stoupá potřebná přesnost na 3,6 ppm. Spektrometry s nízkým hmotnostním rozlišením pak u velkých biomolekul mohou poskytovat pouze průměrné hmotnosti. U nich je navíc třeba počítat s inherentní chybou měření v řádu až 20 ppm. Ta je dána odlišným přírodním zastoupením nuklidů v závislosti na původu vzorku, např. různým poměrem isotopů  $^{12}\text{C}/^{13}\text{C}$  v organických nebo anorganických materiálech<sup>10</sup>.

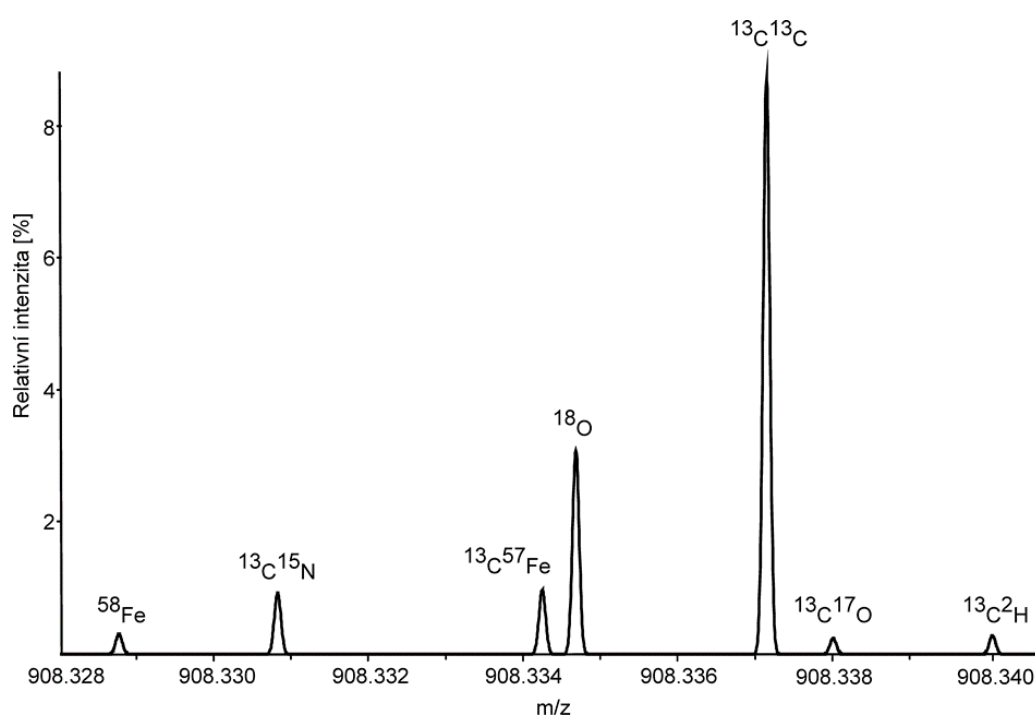
Současné možnosti vysoce rozlišující hmotnostní spektrometrie umožňují měřit rozdíly mezi hmotnostmi



Obr. 5. Simulovaný isotopový profil peptidu



Obr. 6. Přesné hmotnosti a hmotnostní odstupy mezi nuklidy prvků C, N, S

Obr. 7. Detail  $M+2$  píku jednou protonovaného triacetylfulvarininu C ( $\text{C}_{39}\text{H}_{57}\text{N}_6\text{O}_{15}\text{Fe}$ ). Teoretické spektrum bylo generováno pro rozlišení  $R = 900\,000$ 

#### 4. Metoda *per analogiam*

Struktury složitějších látek s molekulovou hmotností zhruba do 2000 Da se někdy řeší metodou *per analogiam*, která je založena na znalosti fragmentačního chování standardu látky blízké. Pomocí porovnání elementárních složení iontových sérií v charakterizovaném analogu se standardem je někdy možno určit místo a povahu jednoduché substituce ve struktuře nové látky.

Zjištění molekulové hmotnosti, sumárního vzorce a měření produktových, byť vysoce hmotnostně rozlišených, spekter často k určení struktur přírodních látek nestačí. Přestože některá obecná fragmentační pravidla a postupy pro interpretaci spekter získaných měkkými

ionizačními technikami byly publikovány<sup>11</sup>, právě použití alternativních strukturálních metod (zejména nukleární magnetické rezonance a rentgenové krystalografie) je nezbytné při charakterizaci složitějších organických látek, o jejichž povaze je málo známo.

#### 5. *De novo* interpretace spekter biopolymerů s omezeným počtem stavebních kamenů

Při interpretaci hmotnostních spekter biopolymerů, jako jsou peptidy, oligonukleotidy, oligosacharidy, glykopeptidy apod., lze využít skutečnosti, že jejich molekuly jsou vystavěny z omezeného počtu stavebních bloků. V ribozomálně syntetizovaných peptidech se setkáváme

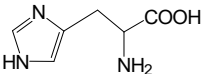
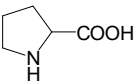
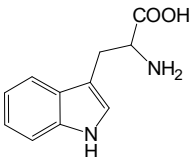
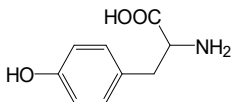
s 20 kódovanými aminokyselinami (tab. IV), u nukleotidů/nukleosidů s příslušnými čtyřmi bázemi apod. Pokud není přítomna neočekávaná kovalentní modifikace, lze pro interpretaci spekter biopolymerů využít řadu interpretačních *de novo* postupů, a to i ve spektrech s nízkým hmotnostním rozlišením. Biopolymery jsou látky teplotně málo stálé a pro jejich ionizaci je třeba použít některou z měkkých ionizačních technik. Základem pro charakterizaci biopolymerů jsou interpretace jejich různých produktových hmotnostních spekter. Například u FT-ICR přístrojů se používá kolizí indukovaná disociace nerezonanční iontovou excitací (SORI-CID, Sustained Off-resonance Irradiation, Collisionally Induced Dissociation), multifotonová disociace infračerveným laserem (IRMPD, Infrared Multi-

Photon Dissociation), disociace elektronovým záchytem (ECD, Electron Capture Dissociation) či disociace přenosem elektronů (ETD, Electron transfer dissociation). Kolizní spektra získaná těmito technikami se mohou významně lišit a dávat tak odlišnou, ale nikoli rozdílnou strukturní informaci.

Jeden z možných obecných postupů při *de novo* interpretaci kolizního spektra lineárního peptidu se může skládat z následujících bodů.

1. Identifikace molekulového iontu a vydělení jeho *m/z* hodnoty číslem 110 (průměrnou molekulovou hmotností všech základních aminokyselin). Takto získáme orientační počet aminokyselin přítomných v daném peptidu.

Tabulka IV  
Základní aminokyseliny

Aminokyselina	Akronym	Kód	Sumární vzorec	Strukturní vzorec
Alanin	Ala	A	$C_3H_7NO_2$	$CH_3-CH(NH_2)-COOH$
Arginin	Arg	R	$C_6H_{14}N_4O_2$	$H_2N-C(=NH)-NH-[CH_2]_3-CH(NH_2)-COOH$
Asparagin	Asn	N	$C_4H_8N_2O_3$	$H_2N-CO-CH_2-CH(NH_2)-COOH$
Kyselina asparagová	Asp	D	$C_4H_7NO_4$	$HOOC-CH_2-CH(NH_2)-COOH$
Cystein	Cys	C	$C_3H_7NO_2S$	$HS-CH_2-CH(NH_2)-COOH$
Glutamin	Gln	Q	$C_5H_{10}N_2O_3$	$H_2N-CO-[CH_2]_2-CH(NH_2)-COOH$
Kyselina glutamová	Glu	E	$C_5H_9NO_4$	$HOOC-[CH_2]_2-CH(NH_2)-COOH$
Glycin	Gly	G	$C_2H_5NO_2$	$CH_2(NH_2)-COOH$
Histidin	His	H	$C_6H_9N_3O_2$	
Isoleucin	Ile	I	$C_6H_{13}NO_2$	$C_2H_5-CH(CH_3)-CH(NH_2)-COOH$
Leucin	Leu	L	$C_6H_{13}NO_2$	$(CH_3)_2CH-CH_2-CH(NH_2)-COOH$
Lysin	Lys	K	$C_6H_{14}N_2O_2$	$H_2N-[CH_2]_4-CH(NH_2)-COOH$
Methionin	Met	M	$C_5H_{11}NO_2S$	$CH_3-S-[CH_2]_2-CH(NH_2)-COOH$
Fenylalanin	Phe	F	$C_9H_{11}NO_2$	$C_6H_5-CH_2-CH(NH_2)-COOH$
Prolin	Pro	P	$C_5H_9NO_2$	
Serin	Ser	S	$C_3H_7NO_3$	$HO-CH_2-CH(NH_2)-COOH$
Threonin	Thr	T	$C_4H_9NO_3$	$CH_3-CH(OH)-CH(NH_2)-COOH$
Tryptofan	Trp	W	$C_{11}H_{12}N_2O_2$	
Tyrosin	Tyr	Y	$C_9H_{11}NO_3$	
Valin	Val	V	$C_5H_{11}NO_2$	$(CH_3)_2CH-CH(NH_2)-COOH$

- Přřazení imoniových iontů, které obsahují pouze jedinou aminokyselinu, v oblasti nízkých hodnot  $m/z$  ve spektru jednotlivým aminokyselinám (tab. V). Pozor na charakteristická hmotnostní omezení u jednotlivých druhů analyzátorů. Případná nepřítomnost daného iontu ještě nemusí znamenat, že odpovídající aminokyselina v peptidu není. Ionty vzniklé během vysokoenergetických procesů ztrátami bočních řetězců z  $[M+H]^+$  iontů mohou poskytnout další informace.
- S pomocí tab. VI určení iontu typu  $b_2$  na hodnotě  $m/z$  ( $AK^1+AK^2+1$ ), který je obvykle doprovázen příslušným iontem typu  $a_2$  na hodnotě  $m/z$  o 28 hmotnostních jednotek nižší. Hodnota  $b_2$  nás současně navádí na ionty  $y_{n-2}$  příp.  $y_{n-1}$ . Tabulka protonovaných aminokyselinových párů pokrývá interval hodnot  $m/z$  od 115 (Gly-Gly+H) do 373 (Trp-Trp+H). S pomocí imoniových iontů, přiřazeného dipeptidu a iontu  $y_{n-1}$  získáme počátek N-koncové sekvence. Nomenklaturu iontových druhů uvádí obr. 8.
- U tryptických peptidů můžeme nalézt ion  $y_1$  na  $m/z$  147, pokud byl C-koncovou aminokyselinou lysin (odpovídá Lys-OH). Druhou možnou C-koncovou variantou je arginin ( $m/z$  175 odpovídá Arg-OH). Je-li nalezen  $y_1$  ion, lze následně přiřadit komplementární ion typu  $b_{n-1}$ . V případě, že je přítomen i ion  $b_n$ , může se hmotnostní rozdíl mezi protonovanou molekulou a iontem  $b_n$  využít i na potvrzení C-koncové skupiny (18 Da ztráta  $H_2O$  odpovídá karboxylové skupině nebo 17 Da ztráta  $NH_3$  odpovídá amidové skupině, atd.).
- Pokud je neznámý vzorek tryptickým peptidem, neměl by Arg nebo Lys být jinde než na C-konci peptidu (platí pro dokonalé štěpení a když nenásleduje prolin). Pokud je nalezena aminokyselina o velikosti 128 uvnitř sekvence, mělo by se tedy jednat o Gln.
- Identifikace zbývajících iontů typu  $b_i$  a  $y_i$  a použití jejich hmotnostních ( $m/z$ ) rozdílů ke konstrukci částečných sekvencí. Pokud se dvě neúplné sekvence nepřekrývají, lze tyto použít k sestavení sekvence delší. V případě překryvových sekvencí získáváme kontrolu správnosti dříve navržené krátké sekvence. Pokud některé fragmentové ionty významně intenzivně převyšují ionty ostatní, může být v sekvenci přítomen Pro, His nebo jiná bazická aminokyselina zodpovědná za preferenční štěpení.
- Součet hmotností všech již určených aminokyselin (informace z imoniových iontů, dipeptidu, z rozdílu hmotností mezi sekvencními ionty) a hodnoty 19 (hmotnost  $H_3O^+$ ) můžeme odečíst od hodnoty  $m/z$  protonované molekuly, abychom získali celkovou hmotnost chybějících aminokyselin.
- Použití dalších pravidel k určení úplné sekvence: ionty typu  $a_i$  jsou na hodnotách  $m/z$  o 28 Da nižších než ionty typu  $b_i$ . Dále některé  $b_i$  ionty eliminují neutrální částice, např. Ser, Thr, Asp a Glu ztrácejí vodu ze svých C-konců (pak pozorujeme ionty  $b_i-18$ ). Asn, Gln, Lys, Arg uvolňují amoniak z N-konců příslušných fragmentových iontů (pozorujeme ionty  $b_i-17$ ). Obě výše uvedené eliminace lze pozorovat i u  $a_i$  iontů. Cys často poskytuje  $b_i-34$  pík (ztráta  $H_2S$ ). Met poskytuje  $b_i-46$  pík ( $-CH_2S$ ), oxidovaný Met dává  $b_i-64$  píky, atd.

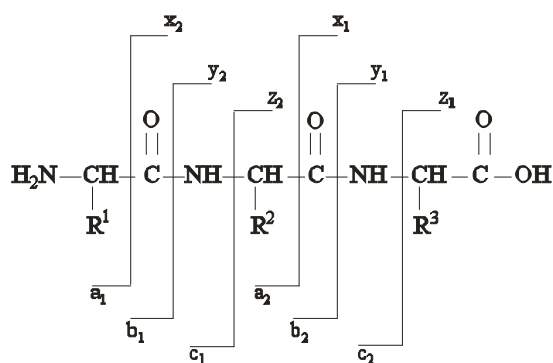
Tabulka V  
Imoniové ionty a další nízkomolekulární fragmenty aminokyselin

Aminokyselina	Hmotnost	Boční řetězec	Imoniové ionty a další fragmenty	
G	Gly	57,02	1	30
A	Ala	71,08	15	44
S	Ser	87,03	31	60
P	Pro	97,05	41	70
V	Val	99,07	43	72
T	Thr	101,05	45	74
C	Cys	103,01	47	76
L	Leu	113,08	57	86 (72)
I	Ile	113,08	57	86 (72)
N	Asn	114,04	58	87 (70)
D	Asp	115,03	59	88
Q	Gln	128,06	72	101 (84, 129)
K	Lys	128,09	72	101 (129, 112, 84, 70)
E	Glu	129,04	73	102
M	Met	131,04	75	104 (61)
H	His	137,06	81	110 (166, 138, 123, 121, 82)
F	Phe	147,07	91	120 (91)
R	Arg	156,10	100	129 (112, 100, 87, 73, 70, 59)
Y	Tyr	163,06	107	136
W	Trp	186,08	130	159

## Tabulka VI

Hodnoty  $m/z$  iontů typu  $b_2$  vzniklých kombinací dvou aminokyselin. Mo je oxidovaný methionin, C\* je karbamidomethylcystein, C<sup>a</sup> je akrylocystein, dvojice L/I, Q/K a Mo/F jsou isobarické

	G	A	S	P	V	T	C	L/I	N	D	Q/K	E	M	H	Mo/F	R	C*	Y	C <sup>a</sup>	W
G	115																			
A	129	143																		
S	145	159	175																	
P	155	169	185	195																
V	157	171	187	197	199															
T	159	173	189	199	201	203														
C	161	175	191	201	203	205	207													
L/I	171	185	201	211	213	215	217	227												
N	172	186	202	212	214	216	218	228	229											
D	173	187	203	213	215	217	219	229	230	231										
Q/K	186	200	216	226	228	230	232	242	243	244	257									
E	187	201	217	227	229	231	233	243	244	245	258	259								
M	189	203	219	229	231	233	235	245	246	247	260	261	263							
H	195	209	225	235	237	239	241	251	252	253	266	267	269	275						
Mo/F	205	219	235	245	247	249	251	261	262	263	276	277	279	285	295					
R	214	228	244	254	256	258	260	270	271	272	285	286	288	294	304	313				
C*	218	232	248	258	260	262	264	274	275	276	289	290	292	298	308	317	321			
Y	221	235	251	261	263	265	267	277	278	279	292	293	295	301	311	320	324	327		
C <sup>a</sup>	232	246	262	272	274	276	278	288	289	290	303	304	306	312	322	331	335	338	349	
W	244	258	274	284	286	288	290	300	301	302	315	316	318	324	334	343	347	350	361	373



Obr. 8. Nomenklatura fragmentových druhů v produktovém spektru tripeptidu

Ionty typu  $a_i$ ,  $b_i$ ,  $c_i$  jsou N-koncové a dolní index označuje počet aminokyselinových zbytků v příslušném iontu. Iontové druhy  $x_i$ ,  $y_i$ ,  $z_i$  jsou C-koncové. Fragmentace za dusíkem peptidové vazby, tj. ionty  $c$  a  $z$  vznikají především pomocí technik ECD a ETC (cit.<sup>12</sup>).

V komplikovaných případech lze dále použít řadu derivatizačních reakcí k získání přesnější sekvenční informace. Lze provést například:

1. Acetylaci N-konce nebo jakéhokoli lysinu (42 Da přírůstek na jednu skupinu). To současně přináší zlepšení N-koncové fragmentace tryptických peptidů obsahujících Arg na C-konci.
2. Methylaci C-konce nebo jakékoli aminokyseliny nesoucí karboxy skupinu (Glu, Asp). Následně sledujeme

přírůstek 14 Da na jednu skupinu a lepší tvorbu  $y_i$  iontové série.

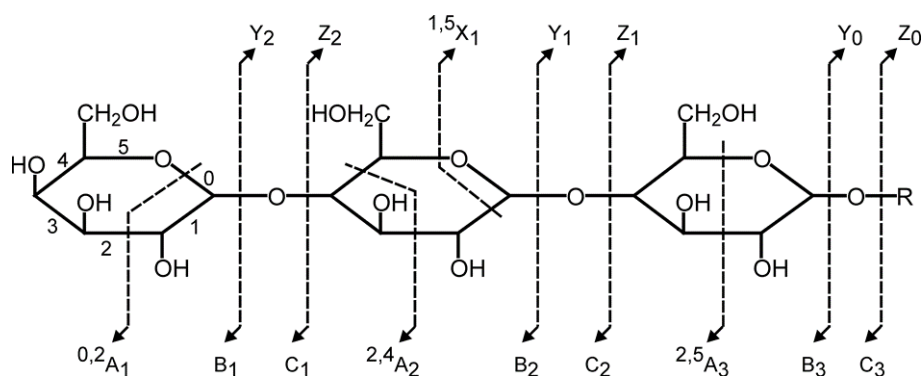
3. N-pyridylacetylaci, která zlepšuje nejen tvorbu obou  $b_i$  i  $y_i$  sérií, ale podpoří i vznik původně nepřítomného iontu  $b_1$ .
4. Při interpretacích *de novo* lze dále využít řadu internetových nástrojů, viz <http://ms-utils.org>, či molekulárního modelování<sup>13</sup>.

Podobně jako u peptidů je zavedena nomenklatura i pro další typy biopolymerů. Obr. 9 uvádí hmotnostně spektrometrickou nomenklaturu pro sekvenování oligosacharidů. Podobně lze sekvenovat oligonukleotidy nebo jiné biopolymery, které mají předem definovaný omezený počet stavebních bloků. Je třeba zdůraznit, že ve všech případech se získává pouze hrubá sekvenční informace, a to bez zřetel na izobarické stavební bloky. Navíc mezi aminokyselinami, sacharidy, lipidy apod. lze nalézt nejen mnoho izobarických stavebních bloků, ale i celou řadu posttranslačních modifikací. Je pak namístě hovořit pouze o charakterizaci struktury neznámého biopolymeru než o určování struktury.

## 6. Softwarové nástroje k usnadnění interpretace hmotnostně spektrometrických dat

K charakterizaci struktur přírodních látek významně přispívá bioinformatika. Při určování sumárního vzorce na základě přesně změřené hmotnosti většina algoritmů pracuje za použití sedmi heuristických pravidel zahrnujících: (1) omezení počtu nuklidů, (2) použití chemických pravidel omezujících maximální valenční stav každého prvku,



Obr. 9. Nomenklatura oligosacharidů (přepřacováno podle cit.<sup>14</sup>)

(3) isotopové profily, (4) poměr počtu atomů vodíku a uhlíku, (5) poměr počtu atomů dusíku, kyslíku, fosforu a síry vůči počtu atomů uhlíku, (6) pravděpodobnost poměrů prvků a konečně (7) přítomnost trimetylsilylovaných látek<sup>15</sup>. Určení sumárního vzorce neznámé látky je i prvním krokem tzv. dereplikace. Termín byl zaveden pro opětné určení již dříve popsanych a strukturně charakterizovaných přírodních látek. Dereplikace je založena na vyhledání sumárního vzorce v lokálně integrované nebo vzdálené (on-line) databázi struktur látek. Programy obvykle vyžadují přesná hmotnostní spektra, někdy zpracovávají i struktury stavebních bloků biopolymerů a fragmentační cesty. Významné veřejně přístupné databáze jsou japonská MassBank<sup>16</sup>, fytochemická databáze ReSpect<sup>17</sup>, Metlin, GNPS (<http://gnps.ucsd.edu/>) a NIST (<http://webbook.nist.gov/chemistry/form-ser.html>).

Pokud se týče algoritmů samotných, významná je technika molekulového síťování, která vzájemně porovnává větší počty MS/MS spekter a hledá strukturní podobnosti<sup>18</sup>. Peptidogenomické nástroje pracují s virtuálními databázemi, které byly vytvořeny *in silico* podle sekvencí předpovězených na základě mechanismů známých z ribozomální nebo neribozomální biosyntézy přírodních látek. K popisu metabolických dat lze s výhodou použít prostředí MyCompoundID (MCID), ve kterém je uloženo přes 8 tisíc lidských endogenních metabolitů a dalších více než 380 tisíc předpovězených metabolických produktů. Podobně jako ostatní bioinformatické nástroje MCID v prvním kroku porovnává změřené molekulové hmotnosti s databázovými. Relevantní navržené struktury s vysokým skóre jsou následně podrobeny disociaci *in silico*<sup>19</sup>. Fragmentace je iniciována na vazbách s heteroatomem, které byly odvozeny z procesů známých z databáze Dictionary of Natural Products. Tato obsahuje přes 220 tisíc látek, ale je komerční. K dispozici je i volná spektrální kolekce s názvem „A database of In-Silico predicted MS/MS spectrum of Natural Products“, která obsahuje přes 170 tisíc látek (<http://oolonek.github.io/ISDB/>)<sup>20</sup>. Menší sbírka mzCloud (<http://www.mzcloud.org>), která používá spektrální korelační algoritmy, je ze slovenské dílny Roberta Mítrika, který je znám díky svému produktu MassFrontier (distributorem je Thermo Scientific).

Metody pro identifikaci biopolymerů je možné rozdělit do dvou základních skupin: vyhledávání v databázích a *de novo* sekvenování. V případě analýzy proteinů se první metoda nejčastěji provádí porovnáváním změřených hmotnostních spekter s databází obsahující teoretická spektra peptidových sekvencí. Peptidové sekvence se získávají *in silico* štěpením proteinových sekvencí, o nichž se předpokládá, že jsou přítomny v analyzovaném vzorku. Druhá metoda získává sekvence přímo z experimentálních spekter bez znalosti genomických dat. Obě metody implementuje celá řada různých komerčních i volně dostupných softwarů ([http://en.wikipedia.org/wiki/List\\_of\\_mass\\_spectrometry\\_software](http://en.wikipedia.org/wiki/List_of_mass_spectrometry_software)). Za zmínku stojí i interpretace hmotnostních spekter s použitím metod molekulární dynamiky<sup>13</sup>, které dokáží prozkoumat energeticky dostupné oblasti fázového prostoru, a tak vypočítat hmotnostní spektra neznámých organických sloučenin.

Pro *de novo* sekvenování lineárních, cyklických, větvených nebo cyklických peptidů s větvelemi lze využít i volně dostupný nástroj CycloBranch<sup>21</sup>. Produkt je určen výhradně pro přesná hmotnostní spektra. Využívá databázi několika stovek stavebních kamenů z ribozomální i neribozomální syntézy peptidů nebo ketidů a dokáže pracovat s celou řadou biogenních prvků. Má integrované knihovny lipidů (<http://www.lipidmaps.org>) a mikrobiálních metabolitů. Kromě *de novo* sekvenovacího modulu lze CycloBranch použít i k dereplikaci organických látek z datových souborů získaných technikami kapalinové chromatografie/hmotnostní spektrometrie nebo hmotnostně spektrometrického zobrazování.

*Práce byla podpořena Ministerstvem školství, mládeže a tělovýchovy České republiky (LO1509) a Grantovou agenturou České republiky (19-10907S).*

## LITERATURA

1. Hamming M. C., Foster N. G., v knize: *Interpretation of Mass Spectra of Organic Compounds*, str. 217. Academic Press, New York 1972.
2. Zinkevich I. G., Ioffe B. V., v knize: *Interpretation of Mass Spectra of Organic Compounds*, str. 174. Khi-

- miya, Leningradskoe Otdelenie, Leningrad 1986.
- McLafferty F.W., Turecek F., v knize: *Interpretation von Massenspektren*; Springer Spektrum, Berlin 2013; anglická verze: *Interpretation of Mass Spectra*, 4. vyd., University Science Books, Mill Valley 1993.
  - McLafferty F. W.: *Org. Mass Spectrom.* 15, 114 (1980).
  - Tureček F., Hanus V.: *Mass Spectrometry Reviews* 3, 85 (1984).
  - Nibbering N. M. M.: *J. Am. Soc. Mass Spectrom.* 15, 956 (2004).
  - Schwarz H., v knize: *Topics in Current Chemistry*, Vol. 73, str. 231. Springer, Berlin 1978.
  - Splitter J.S., Turecek F., v knize: *Applications of Mass Spectrometry to Organic Stereochemistry*; Wiley, NJ, USA 1995.
  - Tureček F.: *Collect. Czech. Chem. Commun.* 52, 1928 (1987).
  - Beavis R. C.: *Anal. Chem.* 65, 496 (1993).
  - Holčapek M., Jirásko R., Lísa M.: *J. Chromatogr. A* 1217, 3908 (2010).
  - Chu I. K., Siu C. K., Lau J. K. C., Tang W. K., Mu X. Y., Lai C. K., Guo X. H., Wang X., Li N., Xia Y., Kong X. L., Oh H. B., Ryzhov V., Turecek F., Hopkinson A. C., Siu K. W. M.: *Int. J. Mass Spectrom.* 390, 24 (2015).
  - Grimme S.: *J. Chem. Theory Comp.* 15, 2847 (2019).
  - Domon B., Costello C. E.: *Glycoconj. J.* 5, 397 (1988).
  - Kind T., Fiehn O.: *BMC Bioinf.* 8, 20 (2007).
  - Horai H. a 34 spoluautorů: *J. Mass Spectrom.* 45, 703 (2010).
  - Sawada Y. a 11 spoluautorů: *Phytochemistry* 82, 38 (2012).
  - Bouslimani A., Sanchez L. M., Garg N., Dorrestein P. C.: *Nat. Prod. Rep.* 31, 718 (2014).
  - Huan T., Tang C., Li R., Shi Y., Lin G., Li L.: *Anal. Chem.* 87, 10619 (2015).
  - Allard P.-M., Peresse T., Bisson J., Gindro K., Marcourt L., Van Cuong P., Roussi F., Litaudon M., Wolfender J.-L.: *Anal. Chem.* 88, 3317 (2016).
  - Novák J., Lemr K., Schug K. A., Havlíček V.: *J. Am. Soc. Mass Spectrom.* 26, 1780 (2015).

**L. Škultéty, J. Novák, and V. Havlíček** (*Institute of Microbiology of the Czech Academy of Sciences, Prague*):  
**Fundamentals of the Mass Spectra Interpretation**

In this paper the authors present their entirely subjective approaches to identify mass spectra of small molecules generated by electron ionization or product ion mass spectra of biopolymers generated by collision-induced dissociation. A rich set of interpretation approaches is supported by knowledge of isotopic profiles including fine isotope structures, exact mass measurements, and application of common fragmentation rules known from gas-phase ion chemistry as well as bioinformatics.

**Keywords:** fragmentation, mechanism, rule, isotope cluster, interpretation approach, ion nomenclature, fine isotope structure

*Acknowledgements*

*Ministry of Education, Youth and Sports of the Czech Republic (LO1509) and Czech Science Foundation (19-10907S).*