

ŘEŠENÉ PŘÍKLADY INTERPRETACE PRODUKTOVÝCH SPEKTER PEPTIDŮ

JIŘÍ NOVÁK a VLADIMÍR HAVLÍČEK

Mikrobiologický ústav AV ČR, Vídeňská 1083, Praha 4
vlhavlic@biomed.cas.cz

Došlo 12.1.20, přijato 13.2.20.

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

Obsah

1. Úvod
2. Lineární peptid
3. Cyklický peptid
4. Lineární peptid s postranním řetězcem
5. Cyklický peptid s postranním řetězcem
6. Bioinformatická podpora

1. Úvod

V návaznosti na příspěvek¹ v předchozím čísle *Chemických listů*, uvádíme modelové postupy pro čtyři strukturně rozdílné peptidy. *De novo* přístup je korelován s bioinformatickým řešením poskytnutým volně dostupným nástrojem CycloBranch, který umožňuje *de novo* sekvenování lineárních a cyklických neribozomálních pep-

tidů². Spektrum neznámého lineárního peptidu, jehož sekvence je řešena v kap. 2, bylo naměřeno s nízkým rozlišením. Ostatní příklady pak zahrnují hmotnostní spektra s vysokým rozlišením. Profil kolizního spektra peptidu se odvíjí od procesů zachycených modelem mobilního protonu a bazicity příslušných nábojových center^{3,4}. Relativní intenzitu konkrétních fragmentových iontů ovlivňuje například N-methylace amidového dusíku peptidové vazby. Zatímco vznik N-koncového b_i iontu není u této modifikace zásadně ovlivněn (nukleofilní atak kyslíkového atomu karbonylové skupiny předchází aminokyseliny přes šestičlenný kruh), pravděpodobnost vzniku komplementárního y_i iontu je snížena absencí vodíkového atomu, který by se u demethylovaného analogu jinak přenášet mohl⁵.

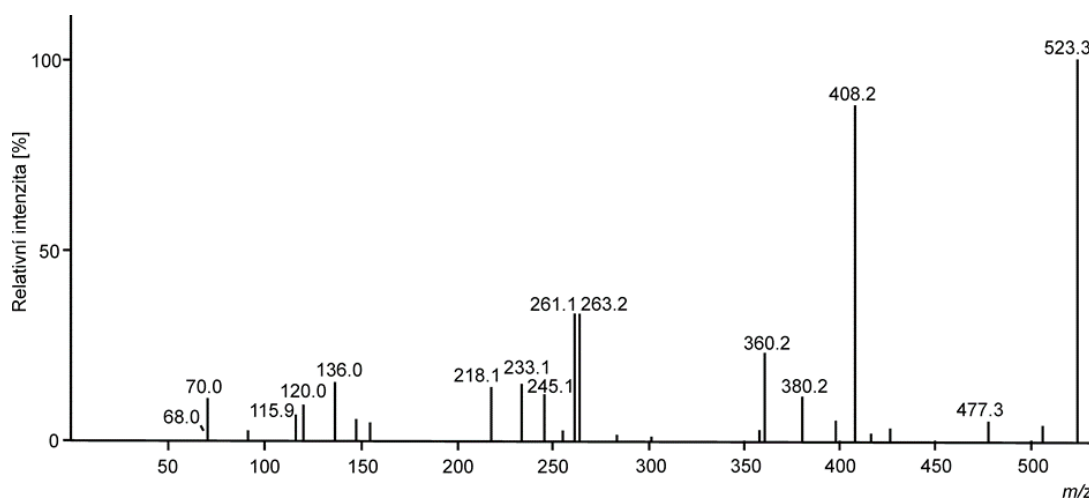
2. Lineární peptid

Zadání:

Identifikujte lineární peptid, jehož hmotnostní spektrum je na obr. 1.

Řešení:

- Podle hodnoty m/z protonované molekuly (523) lze soudit, že peptid obsahuje čtyři nebo maximálně pět aminokyselin.
- Imoniové ionty v oblasti nízkých hodnot m/z ukazují na přítomnost Pro (70), Phe (120) a Tyr (136). Fakt, že jiné imoniové ionty nevidíme, může, ale také nemusí, indikovat vícenásobné zastoupení aminokyseliny jednoho druhu.



Obr. 1. Spektrum lineárního peptidu

- Odečtením hmotností prokazatelně přítomných aminokyselin získáváme chybějící stavební jednotku(ky) peptidu: $523 - (97+147+163+19) = 97$. Zbývající aminokyselinou je tedy prolin (viz Základy interpretace hmotnostních spekter)¹.
- Z dosud identifikovaných aminokyselin víme, jakých teoretických hodnot m/z může nabývat dipeptid b_2 : 195, 245, 261 nebo 311. Správnou hodnotou je m/z 261 (Pro, Tyr), neboť tato je současně podpořena i existencí a_2 iontu (m/z 233). Pokud je naše myšlenková konstrukce správná, komplementárním iontem k b_2 je ion y_{n-2} (n je celkový počet aminokyselin v peptidu), který obsahuje dvojici Pro, Phe. Tento ion je opravdu ve spektru přítomen na m/z 263 (Pro+Phe+19).
- Dosud jsme zjistili, že sekvence má tvar (Pro, Tyr)-(Pro, Phe). Nyní je na řadě určení pořadí jednotlivých aminokyselin, které se získá testováním jednotlivých teoretických variant. Víme, že ion y_2 má hodnotu m/z 263. Další člen řady, ion y_3 , bude buď větší o Tyr (m/z 426) nebo o Pro (360). Pouze druhý z iontů je přítomen, a protože tento je současně o 163 Da menší než protonovaná molekula, musí být N-koncovou aminokyselinou tyrosin. Naše současná sekvence má tedy tvar H-Tyr-Pro-(Pro, Phe).
- Stejným způsobem určíme C-koncovou sekvenci, a to za použití b-iontové série. Víme, že ion b_2 má hodnotu m/z 261. Pokud by byl třetí aminokyselinou od N-konce prolin, b_3 ion by měl hodnotu m/z 358. Pokud by na třetí pozici byl fenylalanin, příslušná hodnota by byla m/z 408. Druhá varianta je správná (vidíme i ion a_3 na m/z 380) a potvrzuje ji i existence iontu y_1 na m/z 116.
- Kompletní sekvence je tedy H-Tyr-Pro-Phe-Pro-OH.

3. Cyklický peptid

Zadání:

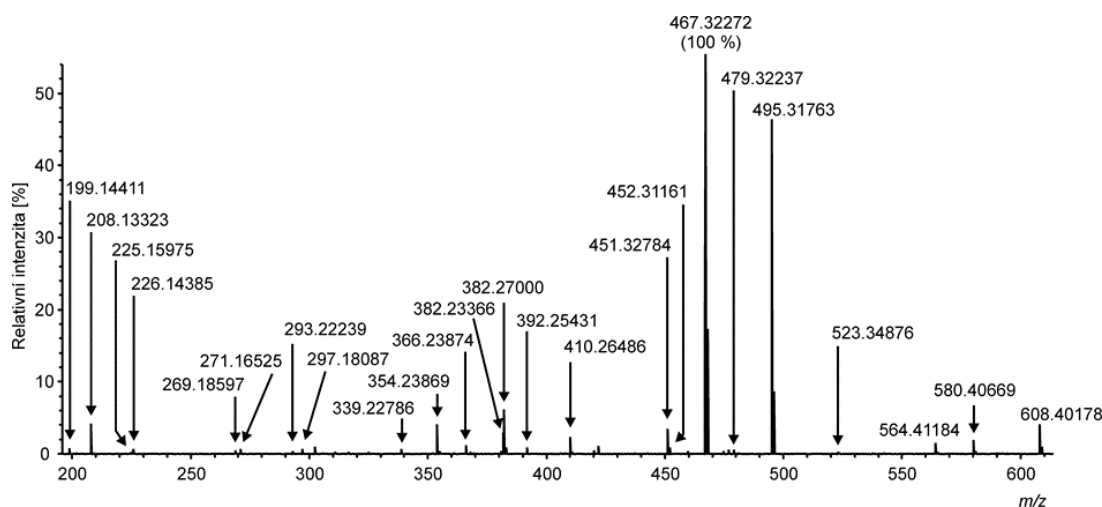
Určete strukturu cyklického neribozomálního peptidu, jehož hmotnostní spektrum je na obr. 2. Hodnota m/z prekursoru 608,40178 odpovídá jednou protonované molekule $[M+H]^+$.

Nápověda:

U neribozomálních peptidů se často definuje nový stavební blok jako kombinace základní aminokyseliny a určité modifikace. Mezi běžné modifikace patří např. methylace (Me), kdy dochází k náhradě amidového vodíku za skupinu $-CH_3$ a hmotnost aminokyseliny se tak zvýší o 14 Da. Tato modifikace existuje např. u alaninu (MeAla), prolinu (MePro) nebo valinu (MeVal). Obvyklou modifikací je rovněž deamidace (1 Da), při které se skupina $-NH_2$ nahradí skupinou $-OH$. Například z leucinu o nominální hmotnosti 113 tak získáváme stavební blok o hmotnosti 114 Da (leucic acid, kyselina 2-hydroxy-4-methylbutanová; La). Hmotnostní rozdíly, které odpovídají jiným než základním aminokyselinám, lze dohledat např. v databázi neribozomálních peptidů Norine⁶.

Řešení:

- Při interpretaci spektra cyklického peptidu si musíme uvědomit, že se jedná o superpozici spekter tolika lineárních peptidů, kolik má daný cyklický peptid aminokyselin. Při fragmentaci se totiž nejprve otevře kruh mezi dvěma po sobě následujícími aminokyselinami a následný „linearizovaný“ peptid se dále fragmentuje. V ideálním případě se tak stane pro všechny dvojice po sobě následujících aminokyselin. Výhodou je fakt, že spektrum poskytuje více informací vzhledem k tomu, že sekvence jednotlivých linearizovaných peptidů se překrývají. Nevýhodou je, že interpretace spektra je komplikovanější než u lineárního peptidu.

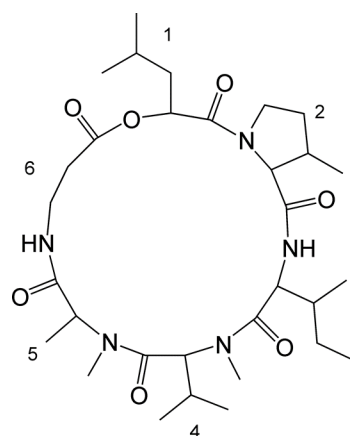


Obr. 2. Spektrum cyklického neribozomálního peptidu

- Protože cyklický peptid z principu nemůže mít –OH skupinu navázanou na C-konci, ve spektru nepozorujeme píky odpovídající y-iontům (+19 Da). Hodnoty m/z , které ve spektru pozorujeme, běžně odpovídají b-iontům nebo a-iontům. Abychom od sebe odlišili píky odpovídající jednotlivým linearizovaným peptidům, aminokyseliny v cyklickém peptidu obvykle číslováme. Ionty pak označujeme jako $x^y\text{b}$, kde x odpovídá číslu N-koncové aminokyseliny a y pozici C-koncové aminokyseliny linearizovaného peptidu, který vznikne po rozdělení sekvence mezi aminokyselinami očíslovanými jako x a y .
- Výjimečně můžeme pozorovat hodnoty m/z , které odpovídají součtu hmotností aminokyselin, které v původní sekvenci cyklického peptidu nenavazují hned za sebou. Tento jev označujeme jako „scrambling“⁷. Při fragmentaci dojde nejprve k otevření kruhu a odštěpení aminokyseliny. Následně se kruh opět uzavře a znovu otevře na jiném místě, kde fragmentace dále pokračuje.
- S využitím přesné hodnoty m/z prekurzoru 608,40178 nejprve určíme sumární vzorec iontu hledané látky $[\text{C}_{31}\text{H}_{53}\text{N}_5\text{O}_7+\text{H}]^+$. Povšimneme si píků, které odpovídají neutrálním ztrátám prekurzoru. V našem případě m/z 580 odpovídá ztrátě CO (a-ion) a hodnota 564 ztrátě CO_2 .
- V dalším kroku anotujeme b-ionty. Hledáme tedy takové hodnoty m/z , které odpovídají součtu hmotností zbytků (residues) jednotlivých aminokyselin zvýšenému o 1 Da. Timto způsobem přiřadíme následující hodnoty m/z : 199 b_2 (Lxx, MeAla), 225 b_2 (Lxx, MePro), 226 b_2 (La, MePro), 271 b_3 (Ala, MeAla, La), 297 b_3 (Ala, MePro, La), 339 b_3 (Lxx, MePro, La), 382.23366 b_4 (Ala, MeAla, MePro, La), 410 b_4 (Lxx, Ala, MePro, La), 452 b_4 (Lxx, Lxx, MePro, La), 495 b_5 (Ala, MeAla, Lxx, MePro, La), 523 b_5 (Ala, Lxx, Lxx, MePro, La). Lxx znamená, že daná aminokyselina odpovídá leucinu, *N*-methylvalinu, allo-isoleucinu, isoleucinu, atd. (isobarické aminokyseliny).
- Z hodnot b-iontů 226, 297, 382 a 495 vypočítáme sekvenci (La, MePro)-Ala-MeAla-Lxx. Z rozdílu m/z hodnoty prekurzoru a iontu 495 vidíme, že sekvence musí obsahovat ještě další stavební blok Lxx (608–495=113). Celkem tedy obsahuje 6 stavebních bloků. S využitím hodnot 452 a 523 dále určíme, že sekvence peptidu odpovídá cyklo((La, MePro)-Ala-MeAla-Lxx-Lxx), protože na koncovou část sekvence Lxx-Lxx navazuje její začátek (La, MePro) resp. (La, MePro)-Ala. Výsledek potvrzují i hodnoty 199, 339 a 410, protože odpovídají b-iontům, které mohou vzniknout fragmentací doposud určeného peptidu. Zbývá ještě stanovit pořadí bloků (La, MePro), k čemuž nám pomohou ionty 225 a 271. Vidíme totiž, že na blok Lxx navazuje MePro, a že na La navazuje posloupnost bloků Ala-MeAla. Dosavadní sekvence je tedy cyklo(MePro-La-Ala-MeAla-Lxx-Lxx). Jelikož se jedná o cyklický peptid, nemůžeme na základě pozorovaných hodnot b-iontů rozhodnout o orientaci N-konce a C-konce v molekule. Musíme tedy ještě

uvažovat opačnou sekvenci peptidu cyklo(Lxx-Lxx-MeAla-Ala-La-MePro).

- Naše pozorování můžeme ještě doplnit tím, že anotujeme a-ionty a píky odpovídající ztrátě vody: 208 $\text{b}_2\text{-H}_2\text{O}$ (La, MePro), 269 a_3 (Ala, MePro, La), 293 $\text{a}_3\text{-H}_2\text{O}$ (Lxx, MePro, La), 354 a_4 (Ala, MeAla, MePro, La), 366 $\text{b}_4\text{-H}_2\text{O}$ (Ala, MeAla, Lxx, La), 382.27000 a_4 (Ala, Lxx, MePro, La), 392 $\text{b}_4\text{-H}_2\text{O}$ (Ala, Lxx, MePro, La), 451 $\text{a}_5\text{-H}_2\text{O}$ (Ala, MeAla, Lxx, Lxx, La), 467 a_5 (Ala, MeAla, Lxx, MePro, La), 479 $\text{b}_5\text{-H}_2\text{O}$ (Ala, MeAla, Lxx, Lxx, La). V našem příkladu pro hodnotu m/z 366 $\text{b}_4\text{-H}_2\text{O}$ nepozorujeme ve spektru příslušný pik 384 b_4 . Pro hodnotu 451 $\text{a}_5\text{-H}_2\text{O}$ pozorujeme 479 $\text{b}_5\text{-H}_2\text{O}$, nevidíme však odpovídající píky 469 a_5 a 497 b_5 . To může nastat v případě, že intenzita chybějících píků je příliš nízká a píky tedy nelze odlišit od šumu. Někdy může nejprve dojít k dehydrataci iontu a jeho následné fragmentaci, což rovněž způsobí, že odpovídající píky nevidíme. Zatímco u lineárního peptidu by nám a-ionty pomohly určit pozici N-konce a C-konce v peptidové sekvenci, v případě cyklického peptidu nám a-ionty k jejich stanovení nepomohou.
- Struktura přírodního cyklického peptidu (roseotoxin A⁸) je zobrazena na obr. 3. Povšimneme si však, že daný peptid ve skutečnosti obsahuje β -alanin místo alaninu, a že jedna aminokyselina Lxx sice odpovídá isoleucinu, druhá však *N*-methylvalinu, který je isomerem Lxx. Struktura rovněž obsahuje 3-methylprolin. Místo, na kterém je připojena methylová skupina v prolinu, však z kolizního spektra rovněž nerozpoznáme. Správná sekvence peptidu je tedy cyklo(La-MePro-Ile-MeVal-MeAla- β Ala). Odlišení isobarických aminokyselinových zbytků je možno dosáhnout například charakteristickou fragmentací příslušných imoniových iontů⁹ nebo z retenčního času v analýze těkavého derivátu příslušné aminokyseliny po hydrolyze peptidu. Technikou plynové chromatografie a hmotnostní spektrometrie je možno případně určit i chiralitu příslušné aminokyseliny¹⁰.



Obr. 3. Řešení struktury cyklického peptidu

- Pro úplnost uvedme ještě značení jednotlivých b-iontů, pokud stavební bloky očíslováme podle obr. 3: 199 ⁴⁻³b₂, 225 ²⁻¹b₂, 226 ¹⁻⁶b₂, 271 ⁵⁻⁴b₃, 297 ⁶⁻⁵b₃, 339 ¹⁻⁶b₃, 382.23366 ⁵⁻⁴b₄, 410 ⁶⁻⁵b₄, 452 ¹⁻⁶b₄, 495 ⁵⁻⁴b₅ nebo ⁴⁻³b₅, 523 ⁶⁻⁵b₅.

4. Lineární peptid s postranním řetězcem

Zadání:

Identifikujte lineární peptid s postranním řetězcem (m/z 758), jehož hmotnostní spektrum jedenkrát protonované molekuly $[M+H]^+$ je na obr. 4.

Nápověda:

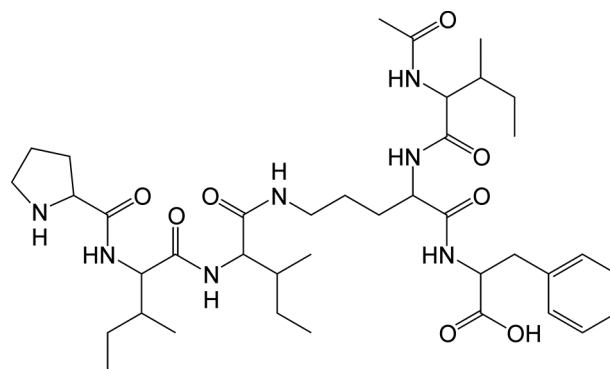
Ornithin (Orn) je aminokyselina, jejíž stavební blok má sumární vzorec $C_5H_{10}N_2O$ a nominální hmotnost 114. Acetylace (Ac) je běžná N-koncová modifikace, která má sumární vzorec C_2H_4O a hmotnost 42. Například aminokyseliny AcLeu nebo AcIle tak odpovídají vzorcům $C_8H_{13}NO_2$ s nominální hmotností $113+42=155$.

Řešení:

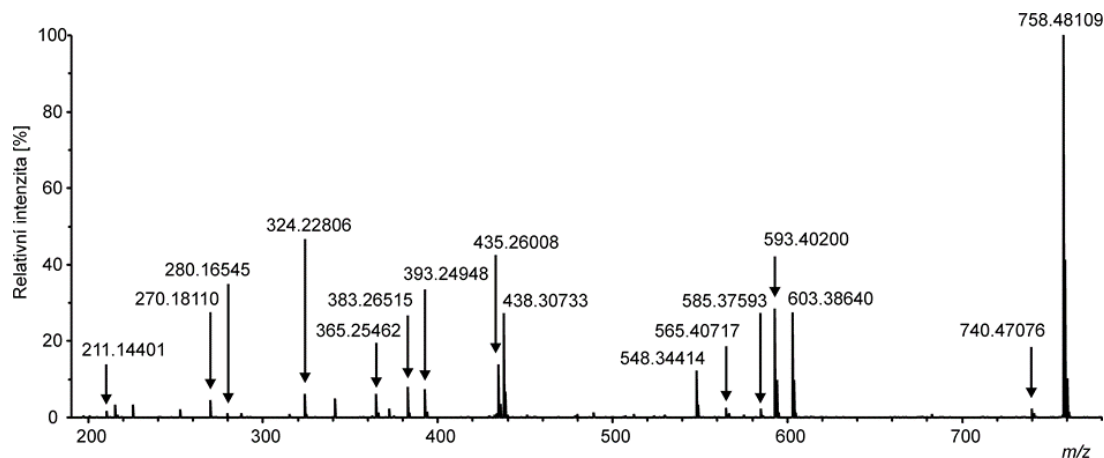
- Ve spektru nejprve přiřadíme píky odpovídající kombinacím základních aminokyselin. Vidíme, že hodnota m/z 211 odpovídá dvojici (Pro, Lxx), tedy iontu b_2 . Dále pozorujeme ion b_3 na m/z 324, čímž získáme částečnou sekvenci (Pro, Lxx)-Lxx, protože rozdíl $324 - 211 = 113$.
- V dalším kroku se pokusíme přiřadit rozdíly hmotností mezi m/z píků nestandardním aminokyselinám. Z hmotností 324 a 438 určíme, že další aminokyselinou je Orn. Známe tedy ion b_4 a sekvenci (Pro, Lxx)-Lxx-Orn. Dále si všimneme, že rozdíl hodnot 585 a 438 odpovídá Phe, a že rozdíl mezi 593 a 438 odpovídá AcLxx. Pozorujeme tedy dva různé ionty b_5 a b_5' odpovídající sekvencím aminokyselin (Pro, Lxx)-Lxx-Orn-Phe a (Pro, Lxx)-Lxx-Orn-AcLxx. Rovněž pozorujeme ion a_5' na m/z 565. Z rozdílu mezi m/z prekurzoru 758 ($[C_{39}H_{63}N_7O_8+H]^+$) a m/z iontů b_5 a b_5' určí-

me dvě sekvence (Pro, Lxx)-Lxx-Orn-Phe-AcLxx a (Pro, Lxx)-Lxx-Orn-AcLxx-Phe (m/z prekurzoru vypočteme v tomto případě stejně jako u lineárního peptidu tj. součet hmotností zbytků aminokyselin zvýšený o 19 Da). Z přítomnosti dvou různých iontů typu b_5 však vidíme, že sekvence peptidu nemůže být lineární a dochází tedy k rozvětvení na Orn, ke kterému jsou připojeny obě dvě aminokyseliny Phe i AcLxx.

- Naše tvrzení podpoříme ještě analýzou y -iontové série. Pozorujeme ion y_2 (Phe, Orn) na m/z 280, y_3 na m/z 435 obsahující aminokyseliny (Phe, Orn, Ac-Lxx) a rovněž ztrátu acetylu z iontu y_3 na m/z 393. Dále vidíme ion y_4 (Phe, Orn, AcLxx)-Lxx na m/z 548 a ion y_5 (Phe, Orn, Lxx, Lxx, Pro) na m/z 603.
- Dále můžeme pozorovat píky odpovídající interním fragmentům: (AcLxx, Orn) na m/z 270, (AcLxx, Orn, Lxx) na m/z 383 a po následné ztrátě vody rovněž na m/z 365.
- Pořadí aminokyselin (Pro, Lxx) už dále nelze jednoznačně určit, protože spektrum neobsahuje ion b_1 na m/z 98 ani ion y_5' (Phe, Orn, Ac-Lxx)-Lxx-Lxx na m/z 661. Rovněž nelze jednoznačně určit, zda-li peptid obsahuje Leu nebo Ile, viz předchozí příklad. Struktu-



Obr. 5. Řešení struktury lineárního peptidu s postranním řetězcem



Obr. 4. Spektrum lineárního peptidu s postranním řetězcem

ra odpovídající spektru (linearizovaná forma pseudacyclinu A^2) je zobrazena na obr. 5. Povšimněme si, že daný peptid má dva N-konce a jeden C-konec.

5. Cyklický peptid s postranním řetězcem

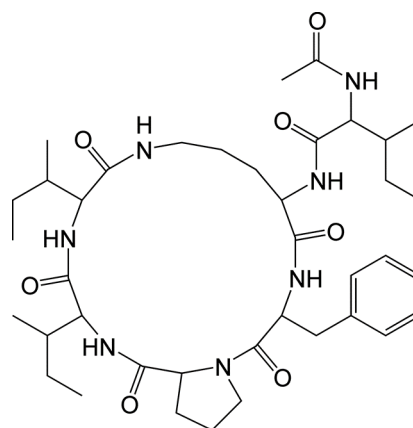
Zadání:

Určete strukturu cyklického peptidu s postranním řetězcem, jehož hmotnostní spektrum jednou protonované molekuly $[M+H]^+$ je na obr. 6.

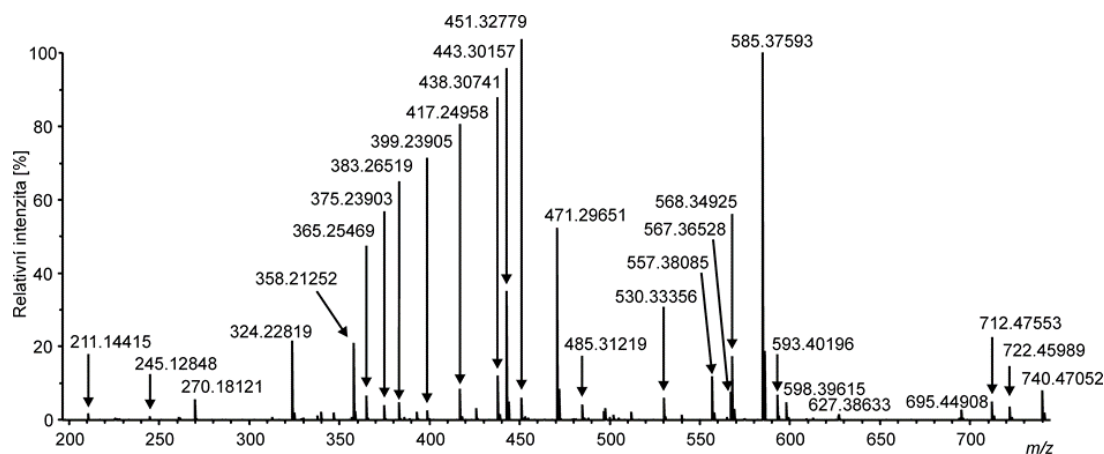
Řešení:

- Vzhledem k tomu, že známe přesnou hodnotu m/z prekursoru, můžeme určit sumární vzorec hledané látky $[C_{39}H_{61}N_7O_7+H]^+$. Rovněž určíme píky s hodnotou m/z 722 (ztráta H_2O), 712 (ztráta CO , a-ion) a 695 (ztráta CO a NH_3).
- Ve spektru nejprve anotujeme b-iony: 211 b_2 (Pro, Lxx), 245 b_2 (Phe, Pro), 270 b_2 (AcLxx, Orn), 324 b_3 (Lxx, Lxx, Pro), 358 b_3 (Phe, Pro, Lxx), 375 b_3 (Phe, Orn, Lxx), 383 b_3 (AcLxx, Orn, Lxx), 417 b_3 (AcLxx, Orn, Phe), 438 b_4 (Orn, Lxx, Lxx, Pro), 471 b_4 (Lxx, Lxx, Pro, Phe), 530 b_4 (AcLxx, Orn, Phe, Lxx), 585 b_5 (Lxx, Lxx, Pro, Phe, Orn), 593 b_5 (Lxx, Lxx, Pro, Orn, AcLxx), 627 b_5 (Lxx, Pro, Phe, Orn, AcLxx). Dále můžeme pozorovat hodnoty 451 a_4-NH_3 (Lxx, Lxx, Orn, AcLxx) a 598 a_5-NH_3 (Lxx, Lxx, Orn, AcLxx, Phe). Podobně jako v oddílu 3 však nevidíme odpovídající píky 468 a_4 , 479 b_4-NH_3 , 496 b_4 , ani 615 a_5 , 626 b_5-NH_3 a 643 b_5 , protože jejich intenzita je zřejmě nižší než hladina šumu. Zbývající píky ve spektru odpovídají ztrátám CO , H_2O , NH_3 a jejich kombinacím z výše uvedených b-ionů.
- Z m/z hodnot iontů typu b_5 a prekursoru vypočteme, že se struktura skládá ze 6 stavebních bloků (Orn, Phe, Pro, Lxx, Lxx, AcLxx). Protože pozorujeme více překrývajících se sérií b-ionů a zároveň nepozorujeme žádné y-iony, předpokládáme, že analyzovaná struktura je cyklická.

- Nyní se pokusíme sestavit sekvenci peptidu. Z m/z hodnot 211, 324, 438 a 585 určíme sekvenci (Pro, Lxx)-Lxx-Orn-Phe. Mohli bychom se tedy domnívat, že sekvence peptidu odpovídá cyklo((Pro, Lxx)-Lxx-Orn-Phe-AcLxx). Tato úvaha je však v rozporu s hodnotou m/z 245, která nám říká, že Phe a Pro musí být v sekvenci za sebou, což v tomto případě není splněno kvůli přítomnosti AcLxx. Dojdeme tedy k závěru, že struktura obsahuje cyklo(Pro-Lxx-Lxx-Orn-Phe), kde bloky Phe a Pro navazují.
- Podobně s pomocí hodnot 245, 358, 471 a 585 určíme sekvenci (Phe, Pro)-Lxx-Lxx-Orn. Zde bychom se mohli domnívat, že sekvence peptidu odpovídá cyklo((Phe, Pro)-Lxx-Lxx-Orn-AcLxx). Dostaneme se však do rozporu s hodnotou 375, která definuje sousedící bloky Phe, Orn a Lxx. Analogicky nám tedy po vyjmutí bloku AcLxx vychází sekvence cyklo(Phe-Pro-Lxx-Lxx-Orn), kde již bloky Phe, Orn a Lxx navazují, a která po rotaci bloků odpovídá předchozímu výsledku cyklo(Pro-Lxx-Lxx-Orn-Phe).



Obr. 7. Řešení struktury cyklického peptidu s postranním řetězcem



Obr. 6. Spektrum cyklického peptidu s postranním řetězcem

- Oba rozpory se nám podaří úspěšně vyřešit, pokud předpokládáme, že struktura peptidu odpovídá cyklo (Pro-Lxx-Lxx-Orn-Phe), kde chybějící stavební blok AcLxx navážeme na Orn. Naše tvrzení podporuje pik s hodnotou m/z 585 a relativní intenzitou 100 %, který je charakteristický při oddělení postranního řetězce od cyklické části peptidu. Při hlubším studiu literatury můžeme rovněž zjistit, že aminokyseliny jako ornithin nebo lysin se vzhledem ke své struktuře často vyskytují jako stavební bloky, které umožňují větvení peptidů. V hmotnostním spektru můžeme pozorovat ještě b-ionty 270, 383, 417, 530 a 627, které potvrzují sousednost jim odpovídajících bloků v navržené struktuře. Výsledná struktura peptidu (pseudacyclin A¹¹) je zobrazena na obr. 7.

6. Bioinformatická podpora

Protože manuální anotace fragmentů neribozomálních peptidů je poměrně náročná – zejména pak v případě cyklických struktur – a navíc často nemáme k dispozici informaci o jednotlivých stavebních blocích, ze kterých může být daný peptid složen, můžeme pro výpočet použít volně dostupnou aplikaci CycloBranch². CycloBranch využívá pro *de novo* sekvenování databázi stavebních bloků např. 20 základních aminokyselin nebo 287 bloků (521 včetně

isomerů) získaných z databáze Norine⁶. Databázi stavebních kamenů si lze však nadefinovat libovolně. Algoritmus pro *de novo* sekvenování peptidů nejprve mapuje hmotnosti jednotlivých stavebních bloků z databáze na rozdíly hodnot m/z mezi libovolnou dvojicí piků ve spektru. Vznikne tak graf, jehož uzly odpovídají hodnotám m/z a hrany rozdílům mezi těmito hodnotami. Algoritmus poté hledá všechny cesty z hypotetického startovního bodu (ion b_0 s hodnotou m/z 1) do cílového bodu (m/z prekurzoru). V případě lineárního peptidu lze rovněž využít i startovní bod y_0 s hodnotou m/z 19 Da.

Nalezené cesty odpovídají sekvencím peptidů (kandidátům), ze kterých je potřeba vybrat správnou sekvenci, a proto CycloBranch pro každého kandidáta vygeneruje jeho teoretické hmotnostní spektrum. U lineárních peptidů se standardně vygenerují píky požadovaných sérií iontů (např. y, b, a). V případě cyklických peptidů CycloBranch pro každého kandidáta vygeneruje teoretické hmotnostní spektrum, které obsahuje píky všech jeho linearizovaných peptidů. Teoretická spektra jednotlivých kandidátů jsou následně porovnávána s experimentálním spektrem a seřazena podle skóre (počet anotovaných piků, součet intenzit anotovaných piků, apod.).

Je však potřeba mít na paměti, že neribozomální peptidy mohou být složeny z různých bloků, z nichž mnoho tvoří isomery. Navíc fragmentace často není dokonalá a v naměřeném spektru tak nemusí být přítomny všechny

*	Result ID	Group ID	Peptide Sequence	Summary Formula	Monoisotopic Mass	Number of Bricks	Matched Peaks	Ratio of Matched Peaks
1	1	1	[NMe-Ala]-[Ala]-[C5:0-Me(4)-OH(2)]-[3Me-Pro]-[Leu]-[Leu]	C31H53N5O7	607.394499	6	27	39.705882
2	2	2	[3Me-Pro]-[C5:0-Me(4)-OH(2)]-[Ala]-[NMe-Ala]-[Leu]-[Leu]	C31H53N5O7	607.394499	6	27	39.705882
3	3	3	[Leu]-[3Me-Pro]-[C5:0-Me(4)-OH(2)]-[Ala]-[NMe-Ala]-[Leu]	C31H53N5O7	607.394499	6	27	39.705882
4	4	2	[C5:0-Me(4)-OH(2)]-[3Me-Pro]-[Ala]-[NMe-Ala]-[Leu]-[Leu]	C31H53N5O7	607.394499	6	26	38.235294
5	5	4	[3Me-Pro]-[C5:0-Me(4)-OH(2)]-[Leu]-[Ala]-[NMe-Ala]-[Leu]	C31H53N5O7	607.394499	6	24	35.294118
6	6	5	[3Me-Pro]-[C5:0-Me(4)-OH(2)]-[Ala]-[Leu]-[NMe-Ala]-[Leu]	C31H53N5O7	607.394499	6	24	35.294118
7	7	6	[Leu]-[3Me-Pro]-[C5:0-Me(4)-OH(2)]-[Ala]-[Leu]-[NMe-Ala]	C31H53N5O7	607.394499	6	24	35.294118
8	8	7	[Leu]-[3Me-Pro]-[C5:0-Me(4)-OH(2)]-[Leu]-[Ala]-[NMe-Ala]	C31H53N5O7	607.394499	6	24	35.294118
9	9	1	[Ala]-[NMe-Ala]-[C5:0-Me(4)-OH(2)]-[3Me-Pro]-[Leu]-[Leu]	C31H53N5O7	607.394499	6	22	32.352941
10	10	3	[3Me-Pro]-[Leu]-[C5:0-Me(4)-OH(2)]-[Ala]-[NMe-Ala]-[Leu]	C31H53N5O7	607.394499	6	22	32.352941
11	11	8	[3Me-Pro]-[Ala]-[C8:0-Me(4)-OH(3)]-[Ala]-[NMe-Ala]-[Leu]	C31H53N5O7	607.394499	6	22	32.352941
12	12	9	[3Me-Pro]-[NMe-Leu]-[Hiv]-[Ala]-[NMe-Ala]-[Leu]	C31H53N5O7	607.394499	6	22	32.352941
13	13	10	[C5:0-Me(4)-OH(2)]-[3Me-Pro]-[Leu]-[Ala]-[Leu]-[NMe-Ala]	C31H53N5O7	607.394499	6	22	32.352941
14	14	11	[C5:0-Me(4)-OH(2)]-[3Me-Pro]-[Leu]-[Ala]-[Leu]-[NMe-Ala]	C31H53N5O7	607.394499	6	22	32.352941

Calculating the number of paths forming peptide sequence candidates... ok
Number of paths found: 71
Comparing theoretical spectra of candidates with the peak list... ok
Preparing the report... ok
CycloBranch successfully finished at 19:12:59 (time elapsed: 0 hrs, 0 min, 1 sec).

Obr. 8. Seznam kandidátních sekvencí peptidů

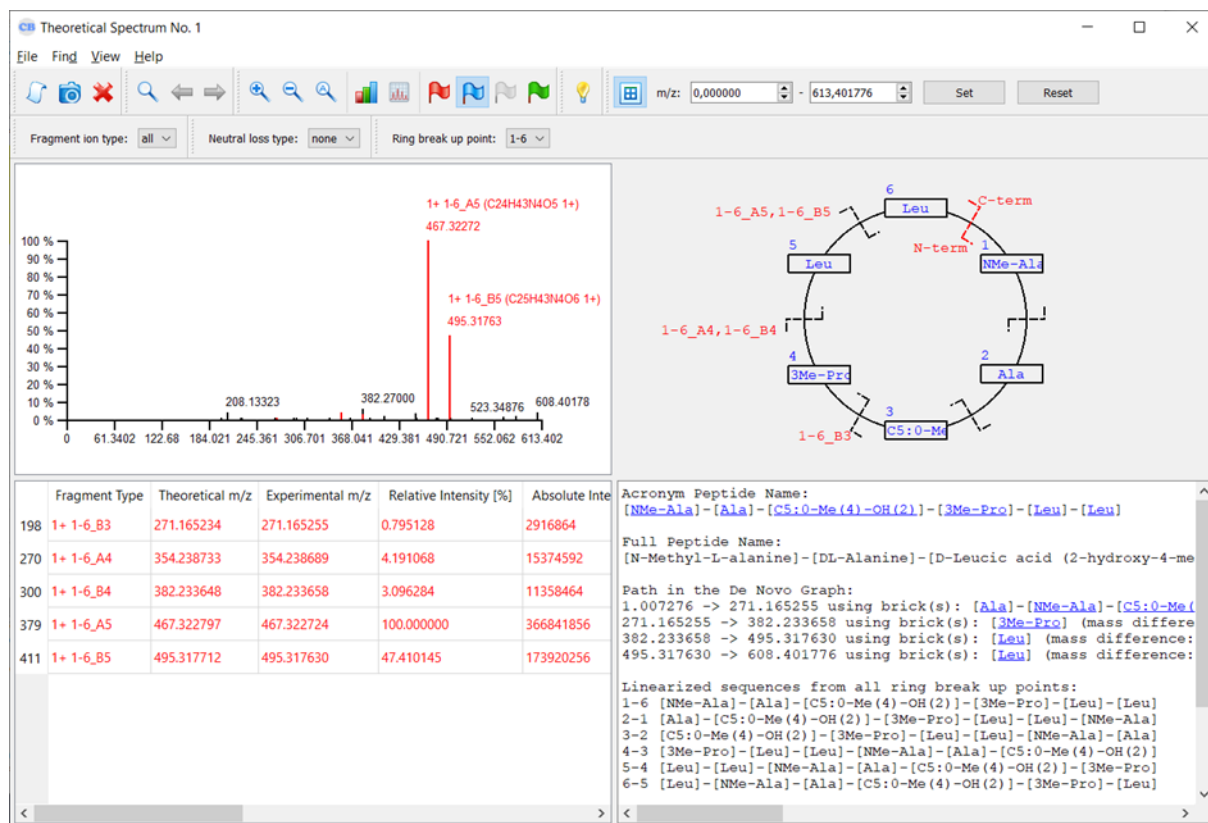
teoretické píky. Běžně je proto potřeba mapovat na rozdíly hodnot m/z nejen hmotnosti jednotlivých stavebních kamenů, ale i jejich kombinace – dvojice, trojice, čtveřice, atd. CycloBranch umožňuje podrobné nastavení maximálního počtu bloků v jedné kombinaci dokonce s využitím tří parametrů $p_1/p_2/p_3$ (první parametr p_1 udává maximální počet bloků mezi m/z b_0 (resp. y_0) a jinou hodnotou ve spektru, poslední parametr p_3 mezi m/z prekursoru a jinou hodnotou ve spektru a prostřední parametr p_2 mezi libovolnou jinou dvojicí hodnot m/z). Motivací pro tento způsob nastavování je fakt, že série (např. b-iontů) bývají obvykle více kompletní „uprostřed“ spektra než na jeho „začátku“ a „konci“. Např. konfigurací 3/1/3 povolíme kombinace maximálně tří bloků na „začátku“ a „konci“ spektra, zatímco „uprostřed“ spektra je zakážeme. Kombinace stavebních bloků však přesto mezi sebou tvoří mnoho isomerů a nepříjemným důsledkem je často fakt, že pro jedno experimentální hmotnostní spektrum můžeme získat mnoho kandidátních sekvencí, které mají shodné nejlepší nebo velmi dobré skóre, a přesto takový výsledek nemusí být nutně správný. Řešením může být syntéza a měření peptidového standardu, či využití alternativní analytické techniky (jaderná magnetická rezonance nebo rentgenová difrakce)¹².

Výpis kandidátních sekvencí peptidů pro spektrum roseotoxinu A je na obr. 8. Po rozkliknutí řádku daného kandidáta je možné zobrazit detail porovnání jeho teoretic-

kého spektra s experimentálním spektrem (obr. 9). Aplikace umožňuje kromě *de novo* sekvenování i přímé porovnání experimentálního spektra s teoretickým spektrem vygenerovaným ze sekvence peptidu, a rovněž porovnání experimentálního spektra s databází peptidových sekvencí.

Ve všech případech je potřeba mít k dispozici databázi stavebních bloků (obr. 10). Pro každý stavební blok definujeme jeho název, unikátní zkratku, sumární vzorec residua. Volitelně pak můžeme definovat seznam neutrálních ztrát oddělených středníkem a referenci do externí databáze např. ChemSpider (<https://www.chemspider.com/>) nebo PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). Například reference „CSID: 969“ (ChemSpider ID) slouží k linkování zkratky „Phe“ na stránku obsahující informace o fenylalaninu. Alternativně bychom mohli napsat „CID: 6140“ (Compound ID) pro odkaz do databáze PubChem. Monoisotopická hmotnost je vypočtena vždy automaticky ze sumárního vzorce. Pokud má nějaký stavební blok isomery, je vhodné je seskupit do jednoho řádku. Ve sloupcích název, zkratka a reference je pak nutné pro oddělení isomerů nutně použít symbol „/“.

Příklad databáze sekvencí neribozomálních peptidů je na obr. 11. Pro každý peptid definujeme jeho typ: lineární (linear), lineární s jednou větví (branched), cyklický (cyclic) a cyklický s jednou větví (branch-cyclic). Dále definujeme název peptidu, sumární vzorec neutrální mole-



Obr. 9. Detail porovnání teoretického a experimentálního spektra

	Name(s)	Acronym(s)	Residue Summary	Monoisotopic Residue Mass	Neutral Loss(es)	Reference(s)	Preview
1	<input type="checkbox"/> 3-hydroxy-4-methylotanoic acid	C8:0-Me(4)-OH(3)	C9H16O2	156.115030		CSID: 17247549	C8:0-Me(4)-OH(3)
2	<input type="checkbox"/> DL-Phenylalanine	Phe	C9H9NO	147.068414		CSID: 969	Phe
3	<input type="checkbox"/> DL-Alanine/beta-Alanine/Sarcosine	Ala/bAla/Sar	C3H5NO	71.037114		CSID: 582/CSID: 234/CSID: 1057	Ala bAla Sar
4	<input type="checkbox"/> Leucine/Isoleucine/N-Methyl-L-valine	Leu/Ile/NMe-Val	C6H11NO	113.084064		CSID: 834/CSID: 769/CSID: 392093	Leu Ile NMe-Val
5	<input type="checkbox"/> D-Leucic acid (2-hydroxy-4-methylpentanoic acid)	C5:0-Me(4)-OH(2)	C6H10O2	114.068080		CSID: 83753	C5:0-Me(4)-OH(2)
6	<input type="checkbox"/> 3-Methylproline/Pipecolic acid	Me-Pro/Pip	C6H9NO	111.068414		CSID: 10610754/CSID: 826	Me-Pro Pip
7	<input type="checkbox"/> N-Methyl-L-alanine/ABU/2-Methylalanine	NMe-Ala/Abu/Aib	C4H7NO	85.052764		CSID: 4450824/CSID: 6405/CSID: 5891	NMe-Ala Abu Aib
8	<input type="checkbox"/> DL-Ornithine	Orn	C5H10N2O	114.079313		CSID: 380	Orn
9	<input type="checkbox"/> DL-Proline	Pro	C5H7NO	97.052764		CSID: 594	Pro
10	<input type="checkbox"/> N-Acetylsoleucine	N-Ac-Ile	C8H13NO2	155.094629		CSID: 270615	N-Ac-Ile
11	<input type="checkbox"/> MeBmt	MeBmt	C10H17NO2	183.125929		CSID: 9944222	MeBmt
12	<input type="checkbox"/> N-methylleucine	Me-Leu	C7H13NO	127.099714		CSID: 493595	Me-Leu
13	<input type="checkbox"/> DL-Valine/Isovaline	Val/Ival	C5H9NO	99.068414		CSID: 1148/CSID: 85483	Val Ival
14	<input type="checkbox"/> DL-Arginine/L-(+)-Arginine	Arg/L-Arg	C6H12N4O	156.101111		CSID: 227/CSID: 6082	Arg L-Arg
15	<input type="checkbox"/> DL-Asparagine	Asn	C4H6N2O2	114.042927		CSID: 231	Asn
16	<input type="checkbox"/> DL-Aspartic acid	Asp	C4H5NO3	115.026943		CSID: 411	Asp
17	<input type="checkbox"/> DL-Cysteine	Cys	C3H5NOS	103.009184		CSID: 574	Cys
18	<input type="checkbox"/> DL-Glutamine	Gln	C5H8N2O2	128.058578		CSID: 718	Gln
19	<input type="checkbox"/> DL-Glutamic acid	Glu	C5H7NO3	129.042593		CSID: 591	Glu
20	<input type="checkbox"/> Glycine	Gly	C2H3NO	57.021464		CSID: 730	Gly

Obr. 10. Příklad databáze stavebních bloků neribozomálních peptidů

	Type	Name	Summary Formula	Monoisotopic Mass	Sequence
1	<input type="checkbox"/> cyclic	beauverolide I	C27H41N3O5	487.304621	[C8:0-Me(4)-OH(3)]-[Phe]-[Ala]-[Leu]
2	<input type="checkbox"/> cyclic	cyclosporin A	C62H111N11O12	1201.841368	[NMe-Bmt]-[Abu]-[NMe-Gly]-[NMe-Leu]-[Val]-[NMe-Leu]-[Ala]-[Ala]-[NMe-Leu]-[NMe-Leu]-[NMe-V
3	<input type="checkbox"/> branched	dau 677 T2	C61H91N15O20	1353.656480	[Glu]-[Ser]-[Leu]([Lys]-[Asn]-[Phe]-[Ile])-[Asp]-[Gln]-[Tyr]-[Gly]
4	<input type="checkbox"/> linear	gramicidin C	C97H139N19O18	1858.054549	[Val]-[Gly]-[Ala]-[Leu]-[Ala]-[Val]-[Val]-[Val]-[Trp]-[Leu]-[Tyr]-[Leu]-[Trp]-[Leu]-[Trp]
5	<input type="checkbox"/> branched	linearized pseudocyclin A	C39H63N7O8	757.473812	[Pro]-[Ile]-[Ile]([Orn]-[NAC-Ile])-[Phe]
6	<input type="checkbox"/> cyclic	pseudocyclin A	C39H61N7O7	739.463247	[Phe]-[Pro]-[Ile]-[Ile]-[Orn]-[NAC-Ile]
7	<input type="checkbox"/> branch-cyclic	pseudocyclin A	C39H61N7O7	739.463247	[Phe]-[Pro]-[Ile]-[Ile]([Orn]-[NAC-Ile])
8	<input type="checkbox"/> branch-cyclic	pyoverdin Pa A (pyoverdin D)	C55H83N17O22	1333.589857	[Fo-OH-Orn]-[Thr]-[Thr]([Lys]-[Fo-OH-Orn]-[Ser]-[Arg]-[Ser]-[ChrPaA]-[Suc])
9	<input type="checkbox"/> cyclic	roseotoxin A	C31H53N5O7	607.394499	[C5:0-Me(4)-OH(2)]-[3Me-Pro]-[Ile]-[NMe-Val]-[NMe-Ala]-[bAla]
10	<input type="checkbox"/> linear	substance P	C63H98N18O13S	1346.728146	[L-Arg]-[Pro]-[Lys]-[Pro]-[Gln]-[Gln]-[Phe]-[Phe]-[Gly]-[Leu]-[Met]
11	<input type="checkbox"/> cyclic	surfactin C	C53H93N7O13	1035.683136	[C14:0-Me(13)-OH(3)]-[Glu]-[Leu]-[Leu]-[Val]-[Asp]-[Leu]-[Leu]
12	<input type="checkbox"/> cyclic	valinomycin	C54H90N6O18	1110.631160	[Val]-[Lac]-[Val]-[Hiv]-[Val]-[Lac]-[Val]-[Hiv]-[Val]-[Lac]-[Val]-[Hiv]

Obr. 11. Příklad databáze sekvencí neribozomálních peptidů

kuly a sekvenci stavebních bloků. Monoisotopická hmotnost se opět vypočítává automaticky ze sumárního vzorce. Podobně jako v předchozím případě je možné vkládat reference do externích databází ChemSpider, PubChem, apod. Sekvence lineárního a cyklického peptidu se zapisují ve

formátu „[A]-[B]-[C]-[D]-[E]“, kde A až E označují zkratky stavebních bloků. V případě cyklického peptidu se automaticky předpokládá, že na stavební blok E navazuje blok A. Sekvence lineárního i cyklického peptidu s větvi se zapisují ve formátu „[A]([B]-[C])-[D]-[E]“, kde staveb-

The screenshot shows the 'Modifications Editor' window with a table of defined modifications. The table has columns for Name, Summary Formula, Monoisotopic Mass, N-terminal, and C-terminal. There are four rows of modifications listed.

	Name	Summary Formula	Monoisotopic Mass	N-terminal	C-terminal
<input type="checkbox"/>	Acetyl	H ₂ C ₂ O	42.010565	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Amidated	HNO-1	-0.984016	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Ethanolamine	H ₅ C ₂ N	43.042199	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Formyl	CO	27.994915	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Obr. 12. Příklad definice koncových modifikací

ní blok B způsobuje rozvětvení (např. ornithin) a blok C označuje konec větve. U cyklického peptidu s větví opět předpokládáme, že na blok E navazuje blok A. Pro definici sekvencí peptidů je možné použít i vestavěný grafický editor. Pro lineární peptid můžeme dále definovat N-koncovou a C-koncovou modifikaci, pro lineární peptid s větví pak můžeme navíc definovat i koncovou modifikaci větve (příslušné sloupce nejsou na obr. 11 zobrazeny). Pro cyklický peptid se koncové modifikace nedefinují. Pro cyklický peptid s větví ale můžeme definovat koncovou modifikaci větve. K tomuto účelu se analogicky využívá editor modifikací (obr. 12). Ukázkové příklady konfigurace programu pro různé režimy vyhledávání a různé typy peptidů jsou součástí instalačního balíčku aplikace CycloBranch, která je volně k dispozici na <https://ms.biomed.cas.cz/cyclobranch>.

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

LITERATURA

- Škultéty L., Novák J., Havlíček V.: Chem. Listy 114, 145 (2020).
- Novák J., Lemr K., Schug K.A., Havlíček V.: J. Am. Soc. Mass Spectrom. 26, 1780 (2015).
- Wright P., Alex A., Pullen F.: Rapid Commun. Mass Spectrom. 30, 1163 (2016).
- Cautereels J., Blockhuys F.: J. Am. Soc. Mass Spectrom. 28, 1227 (2017).
- Vaisar T., Urban J.: J. Mass Spectrom. 33, 505 (1998).
- Caboche S., Pupin M., Leclère V., Fontaine A., Jacques P., Kucherov G.: Nucleic Acids Res. 36, D326 (2008).
- Bleholder C., Osburn S., Williams T. D., Suhai S., Van Stipdonk M., Harrison A. G., Paizs B.: J. Am. Chem. Soc. 130, 17774 (2008).

- Jegorov A., Paizs B., Žabka M., Kuzma M., Havlíček V., Giannakopoulos A. E., Derrick P. J.: Eur. J. Mass Spectrom. 9, 105 (2003).
- Havlíček V., Jegorov A., Sedmera P., Wagnerrederer W., Ryska M.: J. Mass Spectrom. 30, 940 (1995).
- Zahradníčková H., Hartvich P., Šimek P., Hušek P.: Amino Acids 35, 445 (2008).
- Pavlásková K. a 13 spoluautorů: J. Nat. Prod. 73, 1027 (2010).
- Přichystal J., Schug K. A., Lemr K., Novák J., Havlíček V.: Anal. Chem. 88, 10338 (2016).

J. Novák and V. Havlíček (*Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic*): **Interpretation of Peptide Product Ion Mass Spectra: Examples**

The fragmentation of four peptide types, i.e. linear, cyclic, branched and branched-cyclic peptides, is described in four tutorial examples. The possible *de novo* sequencing approaches are supported by an open tool called CycloBranch. This tutorial represents a practical extension of our preceding article Fundamentals of Mass Spectra Interpretation published in the previous Chemické listy issue (2/2020) and can serve the newcomers for the basic orientation in the peptidomics field.

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

Acknowledgements

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (LO1509) and Czech Science Foundation (19-10907S).