

# Identyfikacja białek z wykorzystaniem techniki Peptide Mass Fingerprinting (PMF)

## Część I – charakterystyka eksperymentu identyfikacji

### The identification of proteins by Peptide Mass Fingerprinting (PMF)

#### Part I – properties of the identification experiment

Hanna Kamińska<sup>1,2</sup>, Halina Podbielska<sup>1</sup>

<sup>1</sup> Instytut Inżynierii Biomedycznej i Pomiarowej, Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, tel. +48 71 320 28 25, e-mail: hanna.kaminska@pwr.wroc.pl

<sup>2</sup> MedicWave AB, Stålvärksgatan 1, SE-302 45 Halmstad, Sweden, tel. +46 35 133 600, e-mail: hanna.kaminska@medicwave.com

#### Streszczenie

Wprowadzenie w spektrometrach jonizacji typu MALDI zrewolucjonizowało proces identyfikacji białek. Automatyzacja procesu identyfikacji oraz bezpośrednie połączenie analizy spektrometrem masowym z separacją białek dwuwymiarową elektroforezą żelową (2D-GE) pociągnęły za sobą znaczny rozwój proteomiki. Późniejszy wzrost proteomicznych baz danych pozwolił na zwiększenie dokładności identyfikacji, z wykorzystaniem pierwszej w historii techniki wydajnej identyfikacji białek – *peptide mass fingerprinting*, w skrócie: PMF. Metoda *peptide mass fingerprinting* pozwala identyfikować białka z widm masowych uzyskanych w wyniku analizy próbki spektrometrem masowym. Przez wzgląd na powszechność stosowania metody, jak i ciągle obserwowane jej ulepszenia, autorzy postanowili podsumować obecny stan wiedzy w tym zakresie. Praca została podzielona na dwie części: w pierwszej znajduje się opis historii powstania metody PMF wraz z charakterystyką części eksperymentalnej i opisem najpopularniejszych baz danych stosowanych przy identyfikacji, natomiast druga część pracy jest poświęcona zagadnieniu algorytmicznemu, związanym z wyszukiwaniem w bazie danych protein najlepiej odzwierciedlających białko analizowane w próbce. Specyfikacja eksperymentu w pierwszej części pracy uwzględniła zarówno opis metody separacji, trawienia białek w próbce, jak i późniejszej ich analizy z wykorzystaniem spektrometru masowego. Eksperymentalne fazy metody PMF są opisane z uwzględnieniem ich cech biochemicznych, mających wpływ na dalsze etapy schematu identyfikacji.

**Słowa kluczowe:** proteomika, identyfikacja protein, spektrometria masowa, peptide mass fingerprinting, schematy scoringu

#### Abstract

The development of MALDI ionization method in mass spectrometers, had revolutionized the protein identification procedure. The automation of an identification procedure and the mass spectrometry direct connection to the protein separation with the two-dimensional gel electrophoresis (2D-GE) implicated the significant proteomics development. The later growth of the proteomics databases contributed to the enhancement of the identification accuracy, by using the first method of effective protein identification in the history: the peptide mass fingerprinting (PMF). The peptide mass fingerprinting enabled the protein identification from the mass spectra acquired by

the mass spectrometry sample analysis. Due to the common use of method and its continuous improvements, the authors decided to summarize the current state of the knowledge in this field of science. The publication is divided into two parts. The first one is devoted to the origins of PMF scheme, the characteristics of its experimental part and a description of the most popular databases used in the identification procedure. The second part relates to the algorithmic issues of searching the database protein, which reflects the sample content in the best way. The experiment specification in the first part takes into the consideration the description of separation and sample digestion methods, as well as the later protein sample analysis by the mass spectrometer. The experimental steps of the PMF method are described according to their biochemical properties, having an impact for the later stages of the identification procedure.

**Keywords:** proteomics, identification of proteins, mass spectrometry, peptide mass fingerprinting, scoring schemes

#### Wstęp

Przez identyfikację białek rozumiemy eksperymentalny pomiar ich właściwości i odnalezienie sekwencji aminokwasów o identycznej lub zbliżonej charakterystyce spośród już udokumentowanych i opisanych białek. Do informacji o zidentyfikowanych białkach i ich potwierdzonych właściwościach mamy dostęp poprzez otwarte bazy danych. Dane o zgromadzonych tam opisach sekwencji aminokwasów możemy przeszukiwać, z uwzględnieniem przynależności gatunkowej poszukiwanych białek, ich całkowitej masy czy też znanych nam fragmentów sekwencji. Przy porównywaniu eksperymentalnych właściwości białka z właściwościami białek opisanymi w bazie danych, obliczamy prawdopodobieństwo określające, w jakim stopniu białko eksperymentalne może być utożsamiane z białkiem z bazy.

Z analitycznego punktu widzenia białka mają wiele ważnych właściwości. Wszystkie właściwości białek determinowane są przez sekwencję ich aminokwasów, która również jednoznacznie decyduje o ich rozróżnieniu. Niektóre białka pełnią podobne funkcje, niejednokrotnie są homologiczne pod względem budowy, jednak sekwencja aminokwasów wchodzących w ich skład jednoznacznie określa typ białka. Ponieważ ustalenie pełnej sekwencji białka jest czasochłonne, niezbędne jest wykorzystanie jego innego mierzalnego atrybutu, pozwalającego na ta-

ką identyfikację, którą z bardzo dużym prawdopodobieństwem można by uznać za poprawną. Wybrany atrybut białka, aby nadawał się do wykorzystania w identyfikacji, powinien spełniać poniższe założenia [1]:

- Wartość atrybutu powinna być łatwa do zmierzenia.
- Atrybut powinien mieć dużą czułość, tj. różne białka powinny mieć istotnie różną wartość atrybutu.
- Wartość atrybutu powinna być niezmienna w czasie i jak najmniej zależna od czynników zewnętrznych.
- Jeżeli sekwencja białka jest znana, to wartość wybranego atrybutu powinna być prosta do określenia.

Nie istnieje atrybut białka, który idealnie spełniałby wszystkie powyższe warunki, jednak są takie atrybuty, które spełniają część z powyższych założeń. Atrybutem białek, który może być w sposób efektywny zmierzony z dużą dokładnością, jest np. ich masa molekularna. Masa całkowita białka jako atrybut ma jednak słabą czułość, ponieważ bardzo wiele białek może mieć podobną całkowitą masę molekularną. Rozwiązaniem tego problemu miała być próba pomiaru mas molekularnych fragmentów białka, a nie całej jego sekwencji. Gdyby dało się podzielić łańcuch aminokwasów w pewien z góry określony sposób i zmierzyć masy peptydów go tworzących, to taki zbiór mas identyfikowałby białko w sposób znacznie bardziej czuły niż masa całej molekuly. Zbiór mas peptydów wchodzących w skład białka tworzy jego unikalny identyfikator, masowy odcisk palca (*fingerprint*), który może posłużyć do jego identyfikacji. Ta hipoteza stała się motywem powstania techniki identyfikacji białek znanej obecnie pod angielską nazwą *peptide mass fingerprinting* (PMF) [2].

W pracy, w sekcji 2 zostanie omówiona historia powstania techniki identyfikacji PMF, ze szczególnym uwzględnieniem rozwoju spektrometrii masowej, jako głównego czynnika, który doprowadził do powstania metody. W sekcji 3. zostaną omówione główne założenia techniki PMF związane z eksperymentem spektrometrii masowej: przygotowanie próbki do analizy (tj. separacja i trawienie białek) oraz techniczne aspekty spektrometrii (tj. analiza masy cząstek). Dodatkowo w sekcji 3. zostaną omówione podstawowe bazy danych, na podstawie których identyfikuje się białka. Bazy danych zostaną opisane pod kątem zawartości proteomicznej, związanej z biochemicznymi właściwościami eksperymentu.

## Historia

### Pierwsze próby identyfikacji białek

Pierwszą techniką identyfikacji białek w historii była metoda degradacji Edmana, opracowana w 1950 r. przez szwedzkiego naukowca Pehra Edmana [3]. W każdym kroku degradacji Edmana koniec aminowy łańcucha aminokwasów jest odrywany od białka bez uszkodzenia reszty sekwencji aminokwasów, a następnie określany jest typ oderwanego aminokwasu. Krok odseparowania N-końca łańcucha polipeptydowego nazywany jest pojedynczym cyklem degradacji Edmana. Cykle są powtarzane, aż do momentu odczytania całej sekwencji białka i dotarcia do C-końca łańcucha. Wadą metody jest niemożność kontynuowania sekwencjonowania w przypadku modyfikacji chemicznej końca aminowego analizowanego polipeptydu. Pomimo, iż w ciągu 60 lat istnienia metoda została zautomatyzowana i udoskonalona, to wciąż nie pozwalała na analizę łańcucha składającego się z więcej niż 60 aminokwasów, a identyfikacja pojedynczego aminokwasu zajmuje około godziny [4]. Metoda jest wciąż powszechnie stosowana, ponieważ umożliwia potwierdzanie sekwencji białek gromadzonych w bazach danych, jednak czasochłonność i duży nakład pracy niezbędny przy sekwencjonowaniu stanął u podstaw poszukiwań bardziej efektywnej metody identyfikacji białek [5].

Pierwsze próby analizy mas białek i peptydów zostały podjęte w latach 80. ubiegłego stulecia i były one wynikiem gwałtownego rozwoju technologii spektrometrów masowych. W porównaniu z cyklem eksperymentu degradacji Edmana, pozyskanie spektrum mas komponentów obecnych w próbce trwało kilka sekund. Wprowadzenie w 1981 r. metody jonizacji cząstek w spektrometrach za pomocą bombardowania atomami o dużej energii (*Fast-Atom Bombardment* – FAB) [6], pozwoliło po raz pierwszy w historii na analizę mas tak dużych cząstek, jakimi są białka. Wykorzystując FAB, prekursorami techniki identyfikacji PMF stali się Wiliam J. Henzel, Colin Watanabe i John T. Stults, którzy w 1989 r. jako pierwsi zastosowali pomiar mas peptydów tworzących białko do jego identyfikacji w bazie danych, a swoje rezultaty potwierdzili eksperymentem degradacji Edmana [7]. Byli także autorami pierwszego programu, generującego na podstawie sekwencji opisanych w bazie danych, teoretyczne masy peptydów, które były następnie porównywane z masami uzyskanymi w wyniku eksperymentu spektrometrem masowym typu FAB [2].

Pomimo, iż analiza białek spektrometrem FAB była możliwa, nie została wykorzystana masowo w identyfikacji, ponieważ spektrometry do działania wymagały zbyt dużej ilości materiału badawczego (rzędu nanomoli) oraz nie były w stanie wyprodukować jonów cząsteczek większych niż 20 kDa [2]. Dopiero pojawienie się dwóch nowych technik jonizacji: elektrozpylenia (*Electrospray Ionization* – ESI) [8] i desorpcji laserowej z udziałem matrycy (*Matrix Assisted Laser Desorption Ionisation* – MALDI) [9] umożliwiło analizę białek na szeroką skalę. Dzięki masowej produkcji tych urządzeń, ograniczenia związane z wielkością próbki, jak i zakres mas obserwowanych na podstawie spektrum (w przypadku ESI i MALDI zakres przekraczający 100 kDa) nie były już dłużej przeszkodą w identyfikacji. Z tego powodu w roku 1993 spektrometria masowa MALDI została po raz pierwszy wykorzystana do identyfikacji złożonego zbioru białek z lizatu komórek *E. coli*, odseparowanych za pomocą dwuwymiarowej elektroforezy żelowej (2D-GE) [10]. Rok później, również w przypadku zbioru białek *E. coli*, przeprowadzono z sukcesem identyfikację za pomocą chromatografii cieczowej połączonej ze spektrometrem masowym wyposażonym w ESI [11].

### Rozwój techniki PMF po roku 1993

Do roku 1993 technika identyfikacji białek za pomocą pomiaru mas peptydów zobrazowanych na spektrum nie miała swojej oficjalnej nazwy. Masy peptydów, będące rezultatem eksperymentu spektrometrem masowym, były często określane mianem częściowej mapy peptydowej białka [12, 13]. Kilka grup naukowców na świecie prowadziło w tym samym czasie badania w zakresie identyfikacji białek za pomocą spektrometrii mas. Pierwszymi badaczami, którzy użyli pojęcia *peptide mass fingerprint*, byli naukowcy z grupy badawczej Darryla Pappina, którzy określili tym pojęciem masy peptydów powstałych w wyniku trawienia białka [14]. Od tej pory nazwa *peptide mass fingerprinting* (PMF) na stałe przyjęła się jako definicja techniki identyfikacji białek na podstawie widm otrzymanych z analizy spektrometrem masowym. Oczywiście metoda składa się z kilku złożonych etapów, które szczegółowo zostaną przedstawione w sekcji 3.

Potrzeba analizy i identyfikacji białek na szeroką skalę wynikała z faktu pomyślnego sekwencjonowania pełnych genomów organizmów. Kolejny krok w zdobywaniu wiedzy na temat funkcjonowania organizmów wiązał się właśnie z identyfikacją i charakteryzacją licznych białek dekodowanych przez poznane geny. Wtedy też jako uzupełnienie pojęcia genomu, powstało pojęcie proteomu jako zbioru wszystkich białek istniejących w badanej jednostce, w pewnym określonym momencie czasu i w pewnych warunkach biologicznych [15]. Coraz bardziej zaawansowane rozwiąza-

nia techniczne w spektrometrii, możliwość analizy tysięcy białek w bardzo krótkim czasie, uczyniły ją głównym narzędziem identyfikacji i charakteryzacji, będącym w stanie sprostać ogromowi informacji niesionych przez proteom. Bardzo istotne dla poprawy efektywności identyfikacji było znaczące zwiększenie dokładności pomiaru masy przez spektrometrię, jak i zastosowanie bardziej zaawansowanych analizatorów masy, np. analizatora cyklotronowego rezonansu jonów z fourierowską transformacją wyników (*Fourier Transform Ion Cyclotron Resonance* – FT-ICR) [16, 17]. Niemniej jednak wciąż najczęściej wykorzystywanym w schemacie PMF spektrometrem jest spektrometr ze źródłem jonizacji typu MALDI [2]. Spektrometry typu MALDI są proste w obsłudze, pozwalają na automatyzację procesu identyfikacji oraz obecność w próbce detergentów, soli czy roztworów buforowych [4]. Dodatkowo, zastosowanie w MALDI po roku 1995 analizatora czasu przelotu (*Time Of Flight* – TOF), a następnie analizatora czasu przelotu ze zwierciadłem elektrostatycznym, umożliwiło znaczne zwiększenie zakresu obserwowanych mas oraz rozdzielczości i czułości tych spektrometrów [18, 19]. Obecnie spektrometry masowe typu MALDI mogą badać masy molekuł od ok. 100 Da do nawet do 500 kDa, z dokładnością maksymalnie kilku jednostek ppm, osiągając przy tym rozdzielczość rzędu 5000 FWHM w liniowej konfiguracji i nawet 15000 FWHM w konfiguracji ze zwierciadłem elektrostatycznym.

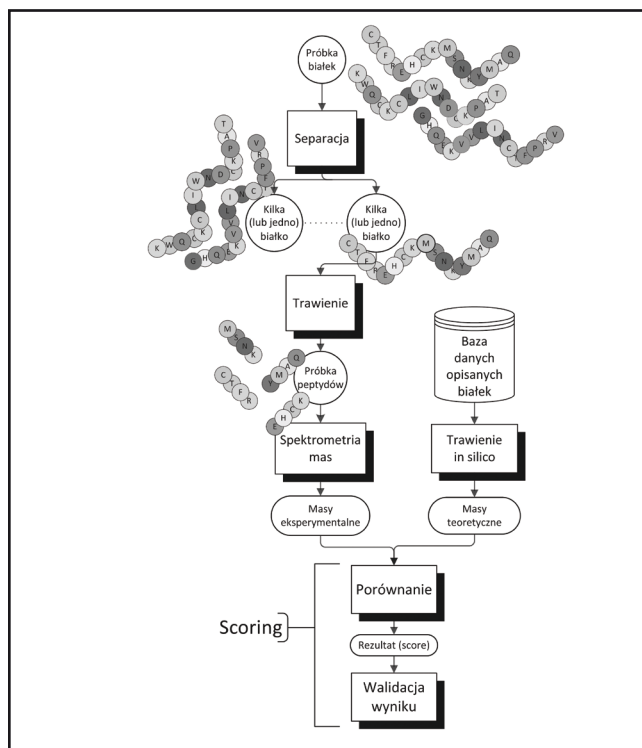
W ciągu ostatnich 10 lat naukowcy skupili się na udoskonalaniu każdego z etapów identyfikacji [20]: od separacji [21] i trawienia białek [22], przez jonizację [23, 24] i analizę ich masy [16, 25], po przeszukiwanie baz danych [2]. Postęp w dziedzinie komputerów i rozwój internetu umożliwił jeszcze bardziej efektywną identyfikację oraz zagwarantował korzystanie z ogólnodostępnego oprogramowania do identyfikacji białek za pomocą schematu PMF [26].

## Metoda

W dzisiejszej proteomice możemy wyróżnić dwa główne paradygmaty analizy i charakteryzacji białek: bottom-up i top-down [27]. Oba podejścia bazują na innych założeniach odnośnie do właściwości badanych białek, metody ich separacji oraz spektrometrów masowych wykorzystywanych w analizie. Głównym założeniem podejścia top-down jest analiza pełnych sekwencji białek, z użyciem spektrometrów masowych umożliwiających ich cykliczną fragmentację na peptydy, a nawet pojedyncze aminokwasy. W odróżnieniu od techniki top-down, podejście bottom-up, które jest wykorzystywane w technice PMF, zakłada trawienie białek, w celu otrzymania peptydów, których masy są następnie analizowane za pomocą spektrometru masowego.

Wybór paradygmatu identyfikacji bottom-up determinuje podział techniki PMF na określone kroki. Główne etapy metody zostały przedstawione na rys. 1. Pierwszym krokiem jest separacja białek, służąca głównie ich wyborowi z pewnego określonego przedziału masowego. Następnie przeprowadzane jest ich trawienie, w celu uzyskania podziału białka na peptydy. Próbka zawierająca peptydy trafia potem do spektrometru masowego. Zarejestrowane w spektrometrze widmo podlega złożonej obróbce, dając w efekcie masy peptydów zaobserwowane w próbce. Kolejnym krokiem w schemacie PMF jest porównanie mas zaobserwowanych na widmie z masami peptydów powstałych w wyniku teoretycznego trawienia białek opisanych w bazie danych.

W praktyce masy peptydów otrzymanych dzięki spektrum należą do molekuł, które tylko w 20-40% pokrywają sekwencję całego białka [28]. Zatem technika PMF w swojej ostatecznej fazie, musi także umożliwiać za pomocą analizy statystycznej (głównie metod estymacji przedziałowej), określenie prawdo-



Rys. 1. Schemat identyfikacji białek w metodzie peptide mass fingerprinting

podobieństwa dopasowania zbioru mas eksperymentalnych do teoretycznych mas białka z bazy danych. Ten niezwykle istotny dla otrzymania wiarygodnych wyników identyfikacji etap jest nazywany scoringiem, od angielskiego słowa *score* (wynik punktowy), ponieważ każdemu białku z bazy danych jest przypisywany pewien score, świadczący o tym, w jakim stopniu może ono być utożsamiane z wynikiem eksperymentu. W kolejnych sekcjach pracy bardziej szczegółowo przyjrzymy się poszczególnym etapom schematu PMF związanym z jego częścią eksperymentalną, natomiast opis drugiej, algorytmicznej części schematu, w ujęciu bioinformatycznym, zostanie podany w kolejnej części publikacji.

## Separacja

Jak możemy zauważyć na rys. 1, pierwszym krokiem w technice PMF jest separacja białek. Etap separacji jest niezbędny, ponieważ PMF z dużą dokładnością identyfikuje tylko białka, które nie wchodzi w skład złożonych mieszanin. Idealnie, w efekcie separacji powinniśmy uzyskać próbkę zawierającą dostateczną ilość jednego rodzaju białka. Przez wzgląd na zbliżone masy molekularne różnych białek, ograniczenia techniczne oraz modyfikacje posttranslacyjne, uzyskanie tak czystej próbki jest niejednokrotnie niemożliwe. Niemniej jednak, gdy w próbce znajdują się maksymalnie trzy różne rodzaje białek, identyfikacja PMF wciąż może być przeprowadzona z dużą dokładnością [1].

Metodą, która jest najczęściej kojarzona z separacją białek w schemacie PMF, jest dwuwymiarowa elektroforeza żelowa (2D Gel Electrophoresis, 2D-GE) [28, 29]. Jednowymiarowa elektroforeza żelowa umożliwia separację białek tylko względem ich masy (*Sodium Dodecyl Sulfate* – *Polyacrylamide Gel Electrophoresis* – SDS-PAGE) [30] lub punktu izoelektrycznego (*Isoelectric Focusing* – IEF) [31]. Nie pozwala to najczęściej w wystarczającym stopniu na rozróżnienie analizowanych białek i uzyskanie w próbce pobranej z żelu minimalnej ilości różnych protein. W dwuwymiarowej elektroforezie żelowej białka w pierwszym wymiarze separuje się względem ich



punktu izoelektrycznego, a w następnym kroku analizowany jest ruch białek w żelu zależny od ich masy (SDS-PAGE) [31]. Przemieszczenie się białek w żelu względem ich masy następuje w kierunku różniącym się o 90 stopni od poprzedniego. Masa i ładunek białka są właściwościami na tyle niezależnymi od siebie, że separacja białek jednocześnie względem masy i punktu izoelektrycznego (2D SDS-PAGE) pozwala na ich znacznie lepsze rozróżnienie, niż ma to miejsce w przypadku metod jednowymiarowych. Dodatkowo, estymacja masy i pI białka na bazie jego końcowego położenia w żelu jest dodatkową informacją, niezwykle przydatną przy końcowej identyfikacji. Dzięki znajomości pI i masy białka możemy się skupić na przeszukiwaniu bazy danych białek tylko pod kątem tych określonych parametrów, co pozwala na zwiększenie prawdopodobieństwa poprawności identyfikacji, oraz na znaczne przyspieszenie obliczeń.

Rozmieszczenie białek na żelu może być zaobserwowane dzięki przyłączeniu do nich barwników lub metali [32]. W przypadku elektroforezy żelowej poprzedzającej analizę spektrometrem masowym najczęściej białka barwi się barwnikiem Coomassie Brilliant Blue G-250 (CBB G-250) [33] lub Coomassie Brilliant Blue R-250 (CBB R-250) [34]. Przyłącza się on do podstawowych aminokwasów, a jego ilość jest proporcjonalna do dodatniego ładunku białka. Białka mogą być także barwione srebrem [21], które przyłącza się do ich grup cysteinowych. Badania wykazały, iż modyfikacje, jakich dokonuje srebro przy przyłączeniu się do grup cysteinowych białka, skutkują najłagodniejszym odzwierciedleniem jego sekwencji analizowanej za pomocą spektrometru MALDI [35]. W przypadku analizy spektrometrem typu MALDI dowiedziono również, że w przypadku uwidocznienia białek, których zawartość w żelu jest znikoma, lepszą czułość ma barwnik CBB G-250, natomiast w przypadku białek, których jest dużo, lepiej sprawdza się CBB R-250 [35].

Niezwykle istotne jest, że w trakcie przygotowania żelu, a także w późniejszych fazach analizy należy zachować ostrożność, aby próbka nie została zanieczyszczona keratyną [28]. Keratyna, jako składnik ludzkiego nabłonka, w łatwy sposób może się przedostać do próbki i pojawić się jako dodatkowa, niepożądana masa wykryta podczas analizy spektrometrem masowym. W gotowym żelu plamy reprezentujące poszczególne białka mogą być wycinane manualnie za pomocą skalpela lub z zastosowaniem robota wycinającego plamy (*spot picker*), który pozwala na automatyzację pracy [36]. Robot może być wyposażony w system umożliwiający trawienie białek poprzedzające naniesienie próbki na matrycę MALDI, a w przypadku trawienia białek następującego bezpośrednio w żelu [37], próbki są od razu przenoszone na matrycę.

### Trawienie

Niezwykle istotnym krokiem w schemacie PMF jest trawienie oddzielonych za pomocą separacji białek (patrz rys. 1), ponieważ to właśnie trawienie pozwala na wytworzenie unikatowego identyfikatora masowego białka (*fingerprint*). W metodzie PMF takie trawienie będziemy określać mianem eksperymentalnego trawienia, ponieważ w odróżnieniu od trawienia teoretycznego (białek z bazy danych) ma ono miejsce w próbce analizowanej w eksperymencie laboratoryjnym. Podstawowym enzymem, wykorzystywanym w trawieniu białek poprzedzającym analizę spektrometrem masowym, jest trypsyna [4]. Trypsyna trawi łańcuch aminokwasów za arginina lub lizyna, o ile nie następuje po nich prolina. Przez wzgląd na swoje specyficzne właściwości trypsyna jest enzymem najlepiej nadającym się do trawienia białek, które mają być następnie analizowane spektrometrem masowym. Niżej wymienione właściwości czynią trypsynę enzymem preferowanym w metodzie PMF [1]:

1. Jest enzymem, który jest pozyskiwany w prosty i tani sposób.
2. Jest swoista, tzn. bardzo rzadko opuszcza w łańcuchu aminokwasów miejsca, w których powinna go trawić, oraz praktycznie nigdy nie trawi łańcucha na nieoczekiwanych pozycjach.
3. Arginina i lizyna występują w łańcuchu aminokwasów średnio co 11 pozycji, a prawdopodobieństwo wystąpienia po nich prolina jest niskie. Trawienie łańcucha aminokwasów w takich odległościach pozwala na tworzenie peptydów o przybliżonej masie najlepiej rozpoznawalnej przez spektrometrię typu MALDI.
4. Może być wykorzystana w większości procedur laboratoryjnych, bardzo dobrze sprawdza się w trawieniu białek zarówno w roztworach, jak i w żelach.
5. W przypadku spektrometrii masowej wykorzystującej ładunki dodatnie, masa peptydu może być odzwierciedlona na spektrum, tylko wtedy, gdy przyłączony jest do niego proton. Ponieważ arginina i lizyna mają odczyn zasadowy, trawienie za tymi aminokwasami gwarantuje peptydowi posiadanie miejsca do przyłączenia protonu.

Trypsyna nie jest jednak jedynym enzymem wykorzystywanym w analizie MALDI. Białka zasadowe z dużym prawdopodobieństwem zawierają w swoich łańcuchach wiele argininy i lizyny, a trypsyna trawi takie białka na krótkie łańcuchy peptydowe. Proteiny o kwaśnym odczynie pH z dużym prawdopodobieństwem zawierają w swoich łańcuchach wiele kwasów asparaginowych i glutaminowych, a znacznie mniej argininy i lizyny. Dlatego też tego typu białka mogą być trawione przez trypsynę na kilka bardzo długich peptydów. W takich przypadkach bardzo często stosuje się alternatywne enzymy trawienne [1, 38], tj.:

- Lys-C – rozcinająca wiązania peptydowe przed lizyną,
- Asp-N – rozcinająca wiązania peptydowe za kwasem asparaginowym,
- Glu-C – rozcinająca wiązania peptydowe za kwasem asparaginowym lub glutaminowym, o ile nie występuje po nich prolina.

Powyższe proteazy są najczęściej stosowane w przypadkach, gdy pH analizowanych białek jest znane i zależy nam na większym pokryciu sekwencji białka na podstawie mas peptydów analizowanych na spektrum. W przypadku, gdy pH analizowanej próbki nie jest znane, można zastosować dodatkowe endopeptydazy w eksperymencie weryfikującym rezultat identyfikacji z próbki, która została wcześniej poddana działaniu trypsyny. Powtórzenie eksperymentu z wykorzystaniem innych enzymów pozwala na potwierdzenie identyfikacji uznanych za poprawne w pierwszym eksperymencie oraz uwidocznienie poprawnych identyfikacji, które w istocie są fałszywe [38]. Takie podejście pozwala znacznie zwiększyć dokładność identyfikacji, jednak przez wzgląd na dodatkowy czas, potrzebny do powtórzenia analizy oraz wykorzystanie znacznego droższego w produkcji proteaz jest znacznie rzadziej stosowane [28].

### Analiza spektrometrem masowym

Jak zostało już powiedziane w sekcji 2., spektrometry masowe typu MALDI są podstawowymi instrumentami stosowanymi w identyfikacji białek metodą PMF. Najpierw próbka składająca się ze skoncentrowanego roztworu peptydów jest łączona z materiałem mającym silne własności absorpcji światła, czyli z tak zwaną matrycą. Za matrycę służą najczęściej: kwas  $\alpha$ -cyjano-4-hydroksycynamonowy lub kwas gentyzynowy, czyli 2,5-dihydroksybenzoesowy [39]. Probki są łączone z matrycą z wykorzystaniem różnych rodzajów rozpuszczalników, np. acetonu lub kwasu trifluorooctowego. Odpowiedni dobór rozpuszczalnika pozwala zwiększyć dokładność analizy [23]. Po wysuszeniu i całkowitym połączeniu z matrycą próbka jest umieszczana w źródle jonizacji spektrometru, skąd cykliczne impulsy lasera wybijają cząst-

ki matrycy [4]. Energia wiązki laserowej jest dobrana w taki sposób, żeby nie doprowadzać do fragmentacji molekuł z matrycy.

Praktycznie wszystkie jony wybijane z matrycy mają dodatni ładunek (jeden przyłączony proton), co w efekcie czyni spektra MALDI bardzo łatwymi do interpretacji. Najczęściej jonizacja w MALDI następuje poprzez przyłączenie protonu. Badania wskazują jednak, że w pewnych przypadkach jonizacja ujemna, czyli przyłączanie do cząstek elektronu, pozwala poprawnie zidentyfikować białka, których nie udało się określić w standardowym trybie [24]. Czas lotu molekuł w tubie próżniowej świadczy o ich masie, ponieważ lepsze cząsteczki szybciej dołączają do detektora masy. W celu zwiększenia dokładności pomiaru masy cząsteczek, do urządzenia MALDI-TOF dołączono tzw. zwierciadło elektrostatyczne, które, po pierwsze, miało na celu zwiększenie dystansu przelotu molekuł, a po drugie, wyrównanie prędkości molekuł mających jednakową masę, jednak dolatujących do detektora w nieco innym czasie [1].

Efektom uderzenia molekuł w detektor masy jest powstające widmo, na którym oś pozioma określa stosunek masy do ładunku molekuły ( $m/z$ ), natomiast oś pionowa natężenie sygnału poszczególnych cząstek. Widmo masowe składa się jednak nie tylko z sygnałów odpowiadających peptydom zaobserwowanym w próbce, ale również jest tworzone przez różnego rodzaju szumy. Szumy mogą być skutkiem elektronicznych zakłóceń wynikających z zanieczyszczenia próbki niepożądanymi substancjami chemicznymi. Dlatego też uzyskanie ostatecznego wyniku eksperymentu, czyli listy mas peptydów, wiąże się z przeprowadzeniem obróbki wstępnej danych ze spektrometru. Przetworzenie danych jest procesem złożonym i nie będziemy w szczegółach opisywać kolejnych etapów. Warto jednak nadmienić, że do podstawowych kroków, mających na celu pozyskanie ze spektrum interesujących nas informacji, należą [1]:

1. Korekcja linii bazowej.
2. Wygładzanie i redukcja szumów.
3. Detekcja pików.
4. Normalizacja intensywności sygnału.
5. Kalibracja.
6. Redukcja obwiedni izotopowej.
7. Usunięcie niepożądanych pików.

Po przeprowadzeniu wyżej wymienionych operacji jako efekt końcowy otrzymujemy tak zwaną listę pików, składającą się jedynie z mas molekuł zaobserwowanych w próbce oraz odpowiadającemu im natężeniu sygnału. Oczywiście końcowy rezultat identyfikacji w dużym stopniu zależy od dokładności algorytmów zastosowanych w wyżej wymienionych krokach. Usprawnień powyższych etapów często dokonuje się w odniesieniu do techniki MALDI i samego PMF [40].

W większości przypadków natężenia sygnału peptydów nie mają wpływu na dalsze kroki metody PMF, ponieważ same masy peptydów, będące efektem obróbki wstępnej danych, stanowią na tym etapie unikatowy, masowy identyfikator białka obecnego w próbce, czyli tzw. *peptide mass fingerprint*. Należy przy tym pamiętać, że zbiór mas otrzymany dla danego spektrum nie musi pochodzić od jednego białka, ale może być skutkiem trawienia dwóch lub trzech różnych białek [2]. Mieszanina peptydów pochodzących z więcej niż jednego białka komplikuje zadanie identyfikacji, jednak nie staje się ono niewykonalne. Metoda identyfikacji w schemacie PMF, nakierowana na identyfikację bardziej złożonego zestawu peptydów w próbce, została zaprezentowana w [41]. W sytuacji złożonej zawartości próbki, na zasadzie włączeń i wyłączeń, brane są pod uwagę możliwości takiego doboru mas w niej zawartych, aby odpowiadały one pewnemu istniejącemu białku. Pozostałe masy zaś są traktowane jak zanieczyszczenie i nie są brane pod uwagę przy obliczaniu prawdopodobieństwa dopasowania mas eksperymentalnych do mas proteiny z bazy danych.

Istotny jest również fakt, iż masy nie wszystkich peptydów będących efektem trawienia białka zostaną odzwierciedlone w eksperymencie. Przeważnie nie udaje się pozyskać mas z więcej niż 20-50% peptydów budujących sekwencję poszukiwanego białka [4]. Przyczyną takiego stanu rzeczy są po części możliwości spektrometru i jego zakres mas, uniemożliwiający zaobserwowanie mas peptydów znajdujących się powyżej, jak i poniżej pewnego progu. Takich mas jest jednak relatywnie mało, a główną przyczyną niepełnego pokrycia sekwencji białka stanowi fakt, iż niektóre peptydy nie ulegają jonizacji. Ponieważ na spektrum mogą zostać zobrazowane tylko masy peptydów, które przyjęły ładunek protonu, wszystkie pozostałe peptydy, nawet jeżeli stanowią sporą część próbki, nie zostaną wykryte. Badania pokazują, że zdolność do jonizacji w peptydach jest warunkowana przez obecność wybranych aminokwasów w ich łańcuchach [1].

### Przeszukiwanie baz danych

Podstawowe dane uzyskane po eksperymencie spektrometrem masowym to lista mas peptydów zaobserwowanych w próbce. Dodatkowo, wiedząc, z jakiego proteomu białka badano, można również określić przynależność gatunkową białek. Jeżeli białka były separowane za pomocą elektroforezy żelowej, to możemy także znać ich przybliżoną masę molekularną i dodatkowo punkt izoelektryczny. Obszar poszukiwania białek w bazie danych pasujących do białka z próbki eksperymentalnej może zatem zostać ograniczony na podstawie wyżej wymienionych parametrów [42].

Istnieje wiele baz danych poświęconych proteomice: przechowujących trójwymiarową strukturę białek, ich szlaki metaboliczne, gromadzące dane o enzymach, substancjach toksycznych czy białkach związanych z chorobami. Bazy danych wykorzystywane w schemacie PMF to bazy przechowujące podstawowe, niezmodyfikowane sekwencje białek prostych, czyli protein. Pojęcie identyfikacji białek możemy zatem zamiennie używać z pojęciem identyfikacji protein, ponieważ masy eksperymentalne dopasowujemy do sekwencji białek prostych opisanych w bazie danych. Wpisy w bazach danych dotyczące protein możemy pobrać najczęściej jako zwykły tekst lub w popularnych formatach: XML lub FASTA. Każdy wpis zawiera wiele podstawowych informacji dotyczących wybranego białka (np. 25 różnych kategorii informacji zawartych w pojedynczym wpisie bazy UniProt [43]). Z punktu widzenia identyfikacji protein najważniejsze w tym wpisie są:

- Unikatowy numer dostępowy proteiny, na podstawie którego można się odwoływać do wpisów pomiędzy różnymi bazami danych.
- Sekwencja aminokwasów proteiny.
- Masa molekularna proteiny.
- Punkt izoelektryczny proteiny.
- Przynależność gatunkowa proteiny.
- Możliwe modyfikacje posttranslacyjne.

Prawdopodobnie najbardziej znane źródło informacji o proteinach i ich funkcjach stanowi baza danych UniProt (Universal Protein Resource) [44]. Utworzenie bazy danych było efektem współpracy trzech instytucji: Europejskiego Instytutu Bioinformatyki (European Bioinformatics Institute, EBI) [45], Szwajcarskiego Instytutu Bioinformatyki (Swiss Institute of Bioinformatics, SIB) [46] oraz instytucji Protein Information Resource (PIR) [47, 48]. Uni-Prot składa się z dwóch głównych baz danych: Swiss-Prot i TrEMBL.

Pierwsza z nich (Swiss-Prot) jest uzupełniana i weryfikowana manualnie, w ramach rygorystycznych zasad zapewniających jasność, poprawność i autentyczność wpisów dodawanych do bazy. Swiss-Prot jest zatem najmniej redundantną ze wszystkich baz, składającą się tylko wpisów podlegających złożonemu procesowi weryfikacji. Przez wzgląd na



czasochłonna weryfikację jest również bazą danych, składającą się z najmniejszej ilości wpisów, co jest jednak w pełni rekompensowane przez ich jakość i wiarygodność [49]. 5 kwietnia 2011 r. baza składała się z 526 969 wpisów dotyczących sekwencji protein [50].

Dруга z baz wchodzących w skład Uni-Prot to TrEMBL. W odróżnieniu od Swiss-Prot nie podlega ona manualnej korekcie, a wszystkie jej wpisy są dodawane automatycznie. W konsekwencji TrEMBL zawiera wiele redundantnych informacji o proteinach, co jest jednak uzupełnione przez ich bardzo bogatą charakterystykę oraz uwzględnienie w spisie protein dopiero co odkrytych. Baza danych TrEMBL zawiera zarówno dane, które po weryfikacji zostaną przeniesione do bazy Swiss-Prot, jak i np. informacje o sekwencjach, co do których eksperymentalnie potwierdzono, iż nie kodują one żadnych białek [49]. Na 5 kwietnia 2011 r. baza składała się z 14 555 721 wpisów dotyczących sekwencji protein [51].

Inną, bardzo bogatą bazę stanowi baza danych protein tworzona przez Narodowe Centrum Informacji Biotechnologicznej (National Center for Biotechnology Information, NCBI) [52, 53]. Zrzesza ona w sobie dane proteomiczne z baz tj.: PIR, Swiss-Prot, Protein Research Foundation (PRF) [54], Protein Data Bank (PDB) [55], a także uwzględnia transkrypcje z sekwencji kodujących opisanych w GenBank [56], RefSeq [57] oraz Third Party Annotation (TPA) Sequence [58], Europejskiego Laboratorium Biologii Molekularnej (European Molecular Biology Laboratory, EMBL) [59] oraz Banku Danych DNA Japonii (DNA Data Bank of Japan, DDBJ) [60]. Baza gwarantuje odrębność wszystkich wpisów na poziomie sekwencji białek. Przeszukiwanie bazy jest możliwe dzięki specjalnie zaprogramowanemu interfejsowi *Entrez* [61] do przeszukiwania połączonych zasobów odrębnych baz danych.

Trzecią, bardzo znaną bazą danych jest International Protein Index (IPI) [62]. IPI opisuje proteomy organizmów eukariotycznych wyższego rzędu przy wykorzystaniu zasobów takich baz, jak: Swiss-Prot, TrEMBL, RefSeq, Ensembl [63], The Arabidopsis Information Resource (TAIR) [64], H-Invitational Database (H-InvDB) [65] oraz The Vertebrate Genome Annotation (VEGA) [66]. Baza IPI jest nakierowana na dostarczanie maksymalnie kompletnych zbiorów protein dla kluczowych organizmów, przy zapewnieniu minimalnej redundancji wpisów [67].

Dobór odpowiedniej bazy danych zależy przede wszystkim od informacji, jaką chcemy na temat danej zidentyfikowanej proteiny uzyskać, oraz od tego, jak bardzo wiarygodnych informacji poszukujemy [68]. Jeżeli żadne z białek zawartych w próbce nie zostało zidentyfikowane z dostatecznym prawdopodobieństwem przy wykorzystaniu najmniej redundantnej bazy danych, to warto rozszerzyć zakres poszukiwań na bazy danych zawierające redundantne, ale najnowsze informacje. Jeżeli nawet na podstawie adnotacji o ostatnio opisanych białkach wiarygodność identyfikacji pozostawia wiele do życzenia, to może się okazać, iż badane białko nie zostało jeszcze w żadnej bazie opisane. Ponieważ schemat identyfikacji PMF pozwala identyfikować tylko białka o opisanych sekwencjach, w takim wypadku należy zastosować inną metodę, która pozwoli na poznanie sekwencji takiego białka od podstaw. Można do takiego celu zastosować degradację Edmana lub przeprowadzić eksperyment identyfikacji sekwencji białka spektrometrem tandemowym z wykorzystaniem algorytmu *de novo* [69].

## Podsumowanie

W dzisiejszych czasach wprowadzane są coraz nowsze i bardziej dokładne metody identyfikacji, wykorzystujące spektrometry tandemowe i skomplikowane algorytmy budowania sekwencji aminokwasów. Choć widma masowe tandemowych spektrometrów są niezwykle dokładne i dają informację bezpośrednio o fragmentach sekwencji zidentyfikowanego biał-

ka, to wciąż powszechnie wykorzystuje się identyfikację metodą PMF. Spektrometry tandemowe przez wzgląd na swoją cenę oraz koszt przeprowadzania pojedynczej analizy nie wyparły z rynku wciąż powszechnie używanych spektrometrów typu MALDI. Przeprowadzenie analizy MALDI, na której opiera się protokół PMF, wymaga znacznie mniejszego nakładu pracy, kosztów i czasu niż przeprowadzenie analizy spektrometrem tandemowym. Identyfikacja opiera się zatem najczęściej na ponawianiu eksperymentu z wykorzystaniem bardziej dokładnych metod (spektrometry tandemowe i sekwencjonowanie) głównie wtedy, gdy metoda PMF dała wątpliwy rezultat o małej istotności lub gdy chcemy ten rezultat potwierdzić [70].

Znajomość właściwości kolejnych etapów schematu PMF oraz wprowadzenie w pewnych prostych modyfikacji, umożliwi w łatwy sposób nakierowanie techniki na identyfikację tylko wybranych typów protein. Przykładem mogą być fosfoproteidy, proteiny będące efektem alternatywnego splicingu [71] czy też mające dużą masę molekularną [28]. Metoda PMF jest używana nie tylko do zidentyfikowania kilku wybranych białek, ale za pomocą rozbudowanego oprogramowania można wykorzystać ją do identyfikacji pełnych kompleksów białek separowanych za pomocą żeli i chromatografii cieczowej [72, 73].

Przez wzgląd na powszechność identyfikacji metodą PMF w przyszłości możemy spodziewać się dalszych ulepszeń techniki, zarówno na etapach przygotowania próbki, jak i rozwijania metod matematycznych, związanych z poszukiwaniem białka z bazy danych odpowiadającemu wynikowi eksperymentu. Technika PMF sama w sobie może także z powodzeniem stanowić wsparcie czy też ulepszenie bardziej zaawansowanych metod identyfikacji. ■

## Literatura

1. I. Eidhammer, K. Flikka, L. Martens, S. O. Mikalsen: *Computational methods for mass spectrometry proteomics*, ch. 1.3, 1st ed., Chichester: Wiley-Interscience 2008, s. 6-118.
2. W.J. Henzel, C. Watanabe, J.T. Stults: *Protein identification: The origins of peptide mass fingerprinting*, Journal of the American Society for Mass Spectrometry, vol. 14, 2003, s. 931-942.
3. P. Edman: *Method for determination of the amino acid sequence in peptides*, Acta chem. scand., vol. 4, 1950, s. 283-293.
4. A.E. Ashcroft: *Protein and peptide identification: the role of mass spectrometry in proteomics*, Natural Product Reports, vol. 20, 2003, s. 202-215.
5. A. Gooley, K. Ou, J. Russell, M. Wilkins, J. Sanchez, D. Hochstrasser, K. Williams: *A role for Edman degradation in proteome studies*, Electrophoresis, vol. 18(7), 1997, s. 1068-1072.
6. M. Barber, R. Bordoli, R. Sedgwick, A. Tyler: *Fast atom bombardment of solids as an ion source in mass spectrometry*, Nature, vol. 293(5830), 1981, s. 270-275.
7. W. J. Henzel, J. T. Stults, C. Watanabe: *Proceedings of the third symposium of the protein society*, Seattle 1989.
8. J. Fenn, M. Mann, C. Meng, S. Wong: *Electrospray ionization for mass spectrometry of large biomolecules*, Science, 1989.
9. M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp: *Matrix-assisted ultraviolet laser desorption of non-volatile compounds*, International Journal of Mass Spectrometry and Ion Processes, vol. 78, 1987, s. 53-68.
10. W. J. Henzel, T. M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe: *Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases*, Proceedings of the National Academy of Sciences of the United States of America, vol. 90, 1993, s. 5011-5015.
11. W. J. Henzel: *Analysis of two-dimensional gel proteins by mass spectrometry and microsequencing*, Methods, vol. 6(3), 1994, s. 239-247.

12. M. Mann, P. Hojrup, P. Roepstorff: *Use of mass spectrometric molecular weight information to identify proteins in sequence databases*, Biological Mass Spectrometry, vol. 22(6), 1993, s. 338-345.
13. O.N. Jensen, A.V. Podtelejnikov, M. Mann: *Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching*, Analytical chemistry, vol. 69, 1997, s. 4741-4750.
14. D.J. Pappin, P. Hojrup, A. Bleasby: *Rapid identification of proteins by peptide-mass fingerprinting*, Current biology, vol. 3(6), 1993, s. 327-332.
15. M. Wilkins, J. Sanchez, A. Gooley, R. Appel, I. Humphrey-Smith, D. Hochstrasser, K. Williams: *Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it*, Biotechnology & Genetic Engineering Reviews, vol. 13, 1995, s. 19-50.
16. T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolić, R.D. Smith: *Utility of accurate mass tags for proteome-wide protein identification*, Analytical Chemistry, vol. 72, 2000, s. 3349-3354.
17. M. Witt, J. Fuchser, G. Baykut: *Fourier transform ion cyclotron resonance mass spectrometry with NanoLC/microelectrospray ionization and matrix-assisted laser desorption/ionization: Analytical performance in peptide mass fingerprint analysis*, Journal of the American Society for Mass Spectrometry, vol. 14, 2003, s. 553-561.
18. R.S. Brown, J.J. Lennon: *Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer*, Analytical Chemistry, vol. 67, 1995, s. 1998-2003.
19. E.J. Takach, W.M. Hines, D.H. Patterson, P. Juhasz, A.M. Falick, M.L. Vestal, S.A. Martin: *Accurate mass measurements using MALDI-TOF with delayed extraction*, Journal of Protein Chemistry, vol. 16, 1997, s. 363-369.
20. O. Jahn, D. Hesse, M. Reinelt, H.D. Kratzin: *Technical innovations for the automated identification of gel-separated proteins by MALDI-TOF mass spectrometry*, Analytical and Bioanalytical Chemistry, vol. 386, 2006, s. 92-103.
21. A. Shevchenko, M. Wilm, O. Vorm, M. Mann: *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels*, Analytical Chemistry, vol. 68, 1996, s. 850-858.
22. A. Otto, B. Thiede, E. Müller, C. Scheler, B. Wittmann-Liebold, P. Jungblut: *Identification of human myocardial proteins separated by two-dimensional electrophoresis using an effective sample preparation for mass spectrometry*, Electrophoresis, vol. 17(10), 1996, s. 1643-1650.
23. N.D. Padliya, T.D. Wood: *A strategy to improve peptide mass fingerprinting matches through the optimization of matrix-assisted laser desorption/ionization matrix selection and formulation*, Proteomics, vol. 4, 2004, s. 466-473.
24. T. Sanaki, M. Suzuki, S.H. Lee, T. Goto, T. Oe: *A simple and efficient approach to improve protein identification by the peptide mass fingerprinting method: concomitant use of negative ionization*, Analytical Methods, vol. 2(8), 2010, s. 1144.
25. N.D. Padliya, T.D. Wood: *Improved peptide mass fingerprinting matches via optimized sample preparation in MALDI mass spectrometry*, Analytica Chimica Acta, vol. 627, 2008, s. 162-168.
26. W.A. Joo, J.B. Lee, M. Park, J.W. Lee, H.J. Kim, C.W. Kim: *Comparison of search engine contributions in protein mass fingerprinting for protein identification*, Biotechnology and Bioprocess Engineering, vol. 12, 2007, s. 125-130.
27. B.T. Chait: *Chemistry. Mass spectrometry: bottom-up or top-down?*, Science, vol. 314, 2006, s. 65-66.
28. B. Thiede, W. Höhenwarter, A. Krah, J. Mattow, M. Schmid, F. Schmidt, P.R. Jungblut: *Peptide mass fingerprinting*, Methods, vol. 35, 2005, s. 237-247.
29. F. Magni, C. Sarto, C. Valsecchi, S. Casellato, S.F. Bogetto, S. Bosari, A. Di Fonzo, R.A. Perego, M. Corizzato, G. Doro, C. Galbusera, F. Rocco, P. Mocarelli, M. Galli Kienle: *Expanding the proteome two-dimensional gel electrophoresis reference map of human renal cortex by peptide mass fingerprinting*, Proteomics, vol. 5, 2005, s. 816-825.
30. A. Shapiro, J. Maizel: *Molecular weight estimation of polypeptides by SDS-polyacrylamide gel electrophoresis: further data concerning resolving power and general considerations*, Analytical Biochemistry, vol. 29(3), 1969, s. 505-514.
31. D.B. Friedman, S. Hoving, R. Westermeier: *Chapter 30 isoelectric focusing and two-dimensional gel electrophoresis*, [w:] R.R. Burgess, M.P. Deutscher red.: *Guide to protein purification*, 2nd Edition, vol. 463, 2009, s. 515-540.
32. B. Lanne, O. Panfilov: *Protein staining influences the quality of mass spectra obtained by peptide mass fingerprinting after separation on 2-d gels. A comparison of staining with coomassie brilliant blue and sypro ruby*, Journal of Proteome Research, vol. 4(1), 2005, s. 175-179.
33. V. Neuhoff, R. Stamm, H. Eibl: *Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: a systematic analysis*, Electrophoresis, vol. 6(9), 1985, s. 427-448.
34. T. Meyer, B. Lamberts: *Use of coomassie brilliant blue R250 for the electrophoresis of microgram quantities of parotid saliva proteins on acrylamide-gel strips*, Biochimica et Biophysica Acta, vol. 107(1), 1965, s. 144.
35. C. Scheler, S. Lamer, Z. Pan, X. Li, J. Salnikow, P. Jungblut: *Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS)*, Electrophoresis, vol. 19(6), 1998, s. 918-927.
36. L. Canelle, C. Pionneau, A. Marie, J. Bousquet, J. Bigeard, D. Lutowski, T. Kadri, M. Caron, R. Joubert-Caron: *Automating proteome analysis: improvements in throughput, quality and accuracy of protein identification by peptide mass fingerprinting*, Rapid Communications in Mass Spectrometry: RCM, vol. 18, 2004, s. 2785-2794.
37. H. Katayama, T. Nagasu, Y. Oda: *Improvement of in-gel digestion protocol for peptide mass fingerprinting by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*, Rapid Communications in Mass Spectrometry: RCM, vol. 15, 2001, s. 1416-1421.
38. F. Levander, T. Rögngvaldsson, J. Samuelsson, P. James: *Automated methods for improved protein identification by peptide mass fingerprinting*, Proteomics, vol. 4, 2004, s. 2594-2601.
39. C. Zhang, H. Zhang, D. Litchfield, K. Ken: *CHCA or DHB? Systematic comparison of the two most commonly used matrices for peptide mass fingerprint analysis with MALDI MS*, Spectroscopy, vol. 25(2), 2010.
40. E.D. Dodds, H.J. An, P.J. Hagerman, C.B. Lebrilla: *Enhanced peptide mass fingerprinting through high mass accuracy: Exclusion of non-peptide signals based on residual mass*, Journal of proteome research, vol. 5, 2006, s. 1195-1203.
41. Z. He, C. Yang, W. Yu: *A partial set covering model for protein mixture identification using mass spectrometry data*, IEEE/ACM Transactions on Computational Biology and Bioinformatics/IEEE, ACM, vol. 8(2), 2011, s. 368-380.
42. P. James, M. Quadroni, E. Carafoli, G. Gonnet: *Protein identification in DNA databases by peptide mass fingerprinting*, Protein Science: A Publication of the Protein Society, vol. 3, 1994, s. 1347.
43. B. Boeckmann, M. Blatter, L. Famiglietti, U. Hinz, L. Lane, B. Roechert, A. Bairoch: *Protein variety and functional diversity: Swiss-Prot annotation in its biological context*, Comptes Rendus Biologies, vol. 328(10-11), 2005, s. 882-899.
44. <http://www.uniprot.org>
45. <http://www.ebi.ac.uk/>
46. <http://www.isb-sib.ch>
47. <http://pir.georgetown.edu/>
48. T.U. Consortium: *Ongoing and future developments at the Universal Protein Resource*, Nucleic Acids Research, vol. 39, 2010, s. 214-219.
49. C. O'Donovan, M. Martin, A. Gattiker, E. Gasteiger, A. Bairoch, R. Apweiler: *High-quality protein knowledge resource: SWISS-PROT and TrEMBL*, Briefings in Bioinformatics, vol. 3, 2002, s. 275.

50. Swiss-Prot, statystyka: <http://www.expasy.org/sprot/relnotes/relstat.html>
51. TrEMBL, statystyka: <http://www.ebi.ac.uk/uniprot/TrEMBLstats/>
52. <http://www.ncbi.nlm.nih.gov/>
53. L. Geer, A. Marchler-Bauer, R. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, S. Bryant: *The NCBI BioSystems database*, Nucleic Acids Research, vol. 38, 2010, s. D492.
54. <http://www.proteinresearch.net>
55. <http://www.pdb.org>
56. <http://www.ncbi.nlm.nih.gov/genbank/index.html>
57. <http://www.ncbi.nlm.nih.gov/RefSeq/>
58. <http://www.ncbi.nlm.nih.gov/genbank/TPA.html>
59. <http://www.embl.org>
60. <http://www.ddbj.nig.ac.jp/>
61. <http://www.ncbi.nlm.nih.gov/sites/gquery>
62. <http://www.ebi.ac.uk/IPI/IPIhelp.html>
63. <http://www.ensembl.org/index.html>
64. <http://www.arabidopsis.org/>
65. <http://jbirc.jbic.or.jp/hinv/ahg-db/index.jsp>
66. <http://vega.sanger.ac.uk/index.html>
67. P.J. Kersey, J. Duarte, A. Williams, Y. Karavidopoulou, E. Birney, R. Apweiler: *The international orotein index: an integrated database for proteomics experiments*, Proteomics, vol. 4, 2004, s. 1985-1988.
68. R. Apweiler, A. Bairoch, C. Wu: *Protein sequence databases*, *Current opinion in chemical biology*, vol. 8, 2004, s. 76-80.
69. V. Dancik, T. Addona, K. Clauser, J. Vath, P. Pevzner: *De novo peptide sequencing via tandem mass spectrometry*, Journal of Computational Biology, vol. 6(3-4), 1999, s. 327-342.
70. C. Borchers, J.F. Peter, M.C. Hall, T.A. Kunkel, K.B. Tomer: *Identification of in-gel digested proteins by complementary peptide mass fingerprinting and tandem mass spectrometry data obtained on an electrospray ionization quadrupole time-of-flight mass spectrometer*, Analytical Chemistry, vol. 72, 2000, s. 1163-1168.
71. S.W. Lee, J.P. Choi, H.J. Kim, J.M. Hong, C. G. Hur: *ASPMF: a new approach for identifying alternative splicing isoforms using peptide mass fingerprinting*, Biochemical and Biophysical Research Communications, vol. 377, 2008, s. 253-256.
72. T. Lee, J. Horng, H. Juan, H. Huang, L. Wu, M. Tsai, H. Huang: *An agent-based system to discover protein-protein interactions, identify protein complexes and proteins with multiple peptide mass fingerprints*, Journal of computational chemistry, vol. 27(9), 2006, s. 1020-1032.
73. K. Lekpor, M.J. Benoit, H. Butler, M. Schirm, D. Vasilescu, K. Bonter, D. Chelsky, P. Hugo, J. Hunter, G. Opitck, E. Paramithiotis, P. Kearney: *An evaluation of multidimensional fingerprinting in the context of clinical proteomics*, Proteomics. Clinical applications, vol. 1, 2007, s. 457-466.

otrzymano / received: 01.02.2011  
zaakceptowano / accepted: 15.03.2011