

Poznań, 11.04.2023

Dr hab. inż. Anna Wojakowska
Zakład Proteomiki Biomedycznej
Pracownia Spektrometrii Mas
Instytut Chemii Bioorganicznej PAN
e-mail: astasz@ibch.poznan.pl

RECENZJA

rozprawy doktorskiej mgr inż. Anny Glodek zatytułowanej

„Deisotoping methods in MALDI-ToF Mass Spectrometry Imaging”

wykonanej w Katedrze Inżynierii i Analizy Eksploracyjnej Danych,

Wydział Automatyki, Elektroniki i Informatyki Politechniki Śląskiej,

pod kierunkiem prof. dr hab. inż. Joanny Polańskiej w roli promotora i dr Marty Gawin w roli kopromotora

Podstawa wykonania recenzji

Recenzję wykonano na zlecenie Rady Dyscypliny Inżynieria Biomedyczna Politechniki Śląskiej z dnia 17.02.2022 r.

Uzasadnienie podjęcia tematu rozprawy doktorskiej

Podjęta przez mgr inż. Annę Glodek tematyka pracy doktorskiej dotyczy opracowania metody identyfikacji obwiedni izotopowych dla danych zarejestrowanych podczas prowadzenia eksperymentów obrazowania molekularnego techniką MALDI-TOF. W wyniku eksperymentów obrazowania molekularnego metodami spektrometrii mas (MSI: *ang. - mass spectrometry imaging*) uzyskujemy znaczne ilości danych, zarejestrowane w formie widm masowych, charakteryzujące się dużą wymiarowością. Analiza danych uzyskanych przy wykorzystaniu techniki MSI jest znaczącym wyzwaniem od strony analizy bioinformatycznej zarejestrowanych widm masowych. Jednym z istotnych etapów wstępnego przetwarzania danych z eksperymentów MS, obok m.in. korekcji linii bazowej, eliminacji szumu, detekcji i ekstrakcji cech, dekonwolucji, integracji i wyrównania danych, jest identyfikacja obwiedni izotopowych. Poprawna identyfikacja obwiedni izotopowych w procesie zwanym deizotopowaniem pozwala właściwie wyodrębnić masę monoizotopową, bądź średnią masę identyfikowanych związków (białek, metabolitów czy lipidów), a w konsekwencji prowadzi do poprawnej ich identyfikacji. Istnieje wiele dostępnych skryptów do deizotopingu, stanowiących element oprogramowania do preprocesingu danych zarejestrowanych technikami MS. Jednak, o ile dostępne oprogramowanie identyfikuje w dużej mierze właściwie obwiednie nienakładające się, o tyle większe wyzwaniem stanowią obwiednie nakładające się, w szczególności w kontekście analizy dużych zbiorów danych, jakimi są dane z obrazowania molekularnego MALDI-MSI. Z tego względu, poszukiwanie nowych wydajnych metod deizotopingu stanowi ważny i uzasadniony problem badawczy.

Ocena merytoryczna pracy i jej struktury

Przedłożona do recenzji praca doktorska została napisana w języku angielskim. Ma postać manuskryptu obejmującego łącznie 113 stron i składającego się z następujących części: wprowadzenie wraz z celem pracy, przegląd literatury, materiały i metody, wyniki wraz z dyskusją, podsumowanie i wnioski, wykaz bibliografii liczący 123 pozycje oraz streszczenie w języku angielskim i polskim. Ponadto, praca zawiera 80 rycin, 11 tabel oraz wykaz użytych skrótów. Zważając na fakt, że praca dotyczy opracowania skryptu, który potencjalnie może zostać zaimplementowany do oprogramowania do wstępnego przetwarzania zbiorów widm masowych, brakuje właściwego skryptu, tudzież linku odsyłającego do źródłowej formy skryptu zamieszczonego na serwerze.

We wstępie został uzasadniony istotny problem podjętych badań, jako konieczność opracowania skutecznej metody deizotopingu, właściwej dla dużych i wielowymiarowych zestawów danych zarejestrowanych w wyniku prowadzonych eksperymentów MALDI-MSI. Poprawne zdefiniowanie zakresu obwiedni izotopowej dla danego peptydu jest podstawą właściwej identyfikacji sekwencji aminokwasowej tego peptydu, w konsekwencji prowadzącej do identyfikacji białka.

Celem pracy było stworzenie metody obliczeniowej, umożliwiającej identyfikację obwiedni izotopowych w zbiorach danych uzyskanych z obrazowania molekularnego MALDI-MSI, opartego na systemie wnioskowania rozmytego (preselekcja pików) oraz przestrzennych mapach dystrybucji molekularnej pików wchodzących w skład obwiedni izotopowej. Algorytm testowano na ośmiu zbiorach danych, uzyskanych w wyniku przeprowadzenia eksperymentów obrazowania molekularnego MALDI-MSI na tkankach guzów pochodzących od pacjentów ze zdiagnozowanym nowotworem regionu głowy i szyi, leczonych w Narodowym Instytucie Onkologii w Gliwicach.

We wstępie literaturowym przedstawiono podstawowe informacje związane z aspektami metodycznymi prowadzenia eksperymentów proteomicznych na tkankach świeżo mrożonych (FF) oraz utrwalonych w formalinie i zatopionych w parafinie (FFPE). O ile dobrze przybliżono temat obwiedni izotopowej i problemów związanych z właściwą klasyfikacją poszczególnych pików do obwiedni i stosowanych do tego celu metod obliczeniowych (algorytmów), o tyle zabrakło wyjaśnienia w jaki sposób powstają jony w przypadku stosowanej jonizacji typu MALDI. Proszę o ustosunkowanie się do tej kwestii w odpowiedzi na recenzję. Doktorantka przedstawia typy najczęściej stosowanych technik jonizacji oraz analizatorów mających zastosowanie w układach MS w badaniach proteomicznych. W mojej opinii, bardziej właściwym z punktu widzenia wprowadzenia do tematu analizy MS, byłoby bliższe skoncentrowanie się na technice, za pomocą której zostały zarejestrowane dane będące przedmiotem pracy doktorskiej, czyli technice MALDI-TOF. Zabrakło również opisu w jaki sposób identyfikowane są białka analizowane techniką MALDI-TOF, w przeciwieństwie do szerzej opisanej techniki ESI. Proszę w tym przypadku również o uzupełnienie brakujących informacji w odpowiedzi na recenzję. Ponadto, w opisie metod i podejść opartych o spektrometrię mas występują pewne nieścisłości i drobne błędy merytoryczne np. w technice LC/MS fazę mobilną nie stanowi gaz, a jak nazwa techniki wskazuje ciecz; w wyniku jonizacji „miękkiej” uzyskujemy jon pseudomolekularny (wg. starej nomenklatury) lub protonowany/deprotonowany nie zaś jon molekularny. Zdaję sobie sprawę z faktu, że Doktorantka jest przede wszystkim bioinformatykiem a nie spektrometrystą mas, jednak tworzy narzędzie mające służyć analizie danych MS, stąd używanie właściwych pojęć i znajomość podstaw dotyczących powstawania jonów pochodzących od peptydów tryptycznych rejestrowanych techniką MALDI-TOF jest w tym

przypadku konieczne. Stosunkowo niewielki fragment wstępu teoretycznego stanowi opis techniki obrazowania molekularnego (MSI), co również z punktu widzenia późniejszej analizy obrazu zasługiwałoby na szerszy opis, np. poruszenie kwestii dotyczących wpływu szerokości rastra na rozdzielczość otrzymanego obrazu molekularnego, czy wyzwania obliczeniowych związanych z analizą tak dużych i złożonych zbiorów danych. W rozdziale dotyczącym obwiedni izotopowych, Doktorantka wspomina o peptydach wielokrotnie naładowanych, w odniesieniu do problemu dechargingu i deizotopingu. Czy takie peptydy mogą powstawać w jonizacji typu MALDI?

Pod koniec wstępu literaturowego Doktorantka przedstawia istniejące algorytmy służące identyfikacji obwiedni izotopowych, w tym oparte o teorię grafów (tworzenia klastra pików izotopowych), oraz o teoretyczną obwiednię izotopową (porównanie teoretycznej obwiedni izotopowej z eksperymentalną). Głównie programy do deizotopingu stworzone są do analizy danych pochodzących z eksperymentów LC/MS, rzadziej MALDI-TOF. Stosunkowo niewiele jest programów dedykowanych analizie tak dużym zbiorom danych jakimi są zbiory widm rejestrowane techniką MALDI-MSI. Ponadto, niewiele algorytmów radzi sobie z deizotopowaniem w obrębie obwiedni nakładających się. Czy Doktorantka jest w stanie procentowo określić jaką część spośród analizowanych obwiedni stanowiły obwiednie nakładające się?

Doktorantka analizowała dwa zestawy danych proteomicznych zarejestrowanych techniką MALDI-MSI dla preparatów (FF i FFPE) tkanek guzów pochodzących od pacjentów cierpiących na nowotwory regionu głowy i szyi. W opisie materiałów i metod zabrakło podania nr zgody komisji bioetycznej dotyczącej analizowanych próbek. Uzyskane dane w postaci widm masowych zostały poddane redukcji wymiarowości. W konsekwencji po wstępnym przetworzeniu (m.in. usunięcie szumu i korekcja linii bazowej, wyrównanie i normalizacja do TIC) i wyborze cech, w przypadku preparatów FF z zarejestrowanych 45 738 surowych widm uzyskano 3 714 komponent gaussowskich. W przypadku preparatów FFPE podano wartości nieuśrednione, osobno dla wszystkich 4 preparatów. Proszę doprecyzować dla preparatów FFPE stopień redukcji wymiarowości danych. Ponadto, widma były rejestrowane w różnych zakresach m/z dla obu typów preparatów. Czy to miało wpływ na uzyskane wyniki procesu deizotopingu?

Zaproponowany przez Doktorantkę skrypt do identyfikacji obwiedni izotopowych jest dwuetapowy. W pierwszej kolejności identyfikowane są pary pików należących do obwiedni za pomocą systemu wnioskowania rozmytego Mamdaniego-Assilana. Zastosowane podejście poddające analizie pary pików umożliwia wykrywanie obwiedni nakładających się. W drugim etapie następuje weryfikacja poprawności zakwalifikowania par pików do obwiedni izotopowej za pomocą przestrzennych map dystrybucji, obrazujących intensywności pików zarejestrowanych przy danym m/z dla całego przekroju preparatu w oryginalnym układzie współrzędnych. Co warto podkreślić, system wnioskowania rozmytego pozwala na znaczną redukcję liczby analizowanych pików, które będą poddawane dalszej analizie w drugim etapie metody. Dzięki takiemu podejściu metoda z powodzeniem może zostać wykorzystana do deizotopingu w przypadku dużych zbiorów danych.

Doktorantka dysponowała 8 zbiorami danych, z czego jeden został wykorzystany do stworzenia modelu (służąc jako zbiór uczący - 115 par pików i zbiór testowy - 40 par pików) a 7 pozostałych zbiorów stanowiły dane walidacyjne. W konsekwencji bez względu na rodzaj badanego zbioru danych, tj. sposób zabezpieczenia materiału tkankowego, preparatykę czy ilość zarejestrowanych widm, dokładność i czułość procesu identyfikacji pików zakwalifikowanych do obwiedni izotopowej jest zbliżony. Ponadto, mgr inż. Anna Glodek porównała zaproponowaną

przez siebie metodę z istniejącymi i ogólnie dostępnymi skryptami do deizotopingu, a uzyskane wyniki weryfikowała przy wykorzystaniu wiedzy eksperckiej doświadczonego spektrometrysty mas. W efekcie zaproponowany przez Doktorantkę skrypt klasyfikował piki dokładniej w porównaniu z badanymi algorytmami.

Doktorantka w pracy nie ustrzegła się błędów edytorskich, w szczególności dotyczy to spisu literatury, w przypadku wielu pozycji brakuje podania nazwy czasopisma. Na uwagę zasługują trzy doniesienia literaturowe, dotyczące metody do deizotopingu z wykorzystaniem systemów rozmytych, których doktorantka jest pierwszym autorem. Praca opatrzona jest znaczącą liczbą rycin, które ułatwiają zrozumienie wyników przedstawionych badań.

Pozostałe uwagi:

- Proszę o wyjaśnienie definicji kanałów masowych?
- Jaka dokładność masy była brana pod uwagę podczas obliczeń?
- Czy stosunek intensywności pików wchodzących w skład obwiedni lub porównanie z teoretyczną obwiednią izotopową były brane pod uwagę, podczas tworzenia klasyfikatora?
- Czy zaproponowany przez Doktorantkę skrypt będzie zaimplementowany do programu do analizy danych MS, dostępnego w dostępie otwartym/komercyjnym?
- Czy zaproponowany algorytm może zostać wykorzystany do deizotopingu dla danych zarejestrowanych innymi technikami MS?

Podsumowanie:

Zaproponowana przez Doktorantkę metoda identyfikacji obwiedni izotopowych, do analizy dużych zbiorów danych, takich jak te otrzymane za pomocą technik obrazowania molekularnego, stanowi istotny wkład w rozwój metod wstępnej analizy i przetwarzania danych MS. Metoda uwzględniająca analizę par pików, skutkująca wykrywaniem obwiedni nakładających się, stanowi oryginalne narzędzie analityczne dla danych uzyskanych techniką MALDI-MSI. Pomimo przedstawionych powyżej drobnych uwag krytycznych, które nie umniejszają wagi zrealizowanego celu naukowego, rozprawę doktorską oceniam pozytywnie. Doktorantka wykazała się wiedzą z zakresu charakteru analizowanych danych i właściwym warsztatem obliczeniowym, które pozwoliły na stworzenie oryginalnego rozwiązania problemu badawczego.

Wnioski

W mojej opinii przedstawiona praca spełnia kryteria stawiane rozprawie doktorskiej określone w aktualnie obowiązujących regulacjach prawnych. Wnoszę do Rady Naukowej Inżynieria Biomedyczna Politechniki Śląskiej o dopuszczenie mgr inż. Anny Glodek do dalszych etapów postępowania o nadanie stopnia doktora.

Dr hab. inż. Anna Wojakowska

