

Prof. dr hab. Witold Danikiewicz  
Instytut Chemii Organicznej PAN  
ul. Kasprzaka 44/52  
01-224 Warszawa

Warszawa, 3-12-2017

## RECENZJA

pracy doktorskiej magistra Mateusza Krzysztofa Łąckiego pt. „Computational and Statistical Methods for Mass Spectrometry Data Analysis” wykonanej na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego pod kierunkiem prof. Anny Gambin

Jednym z najszybciej rozwijających się obecnie obszarów badań w dziedzinie biochemii jest chemia białek. W badaniach proteomicznych najważniejszym zagadnieniem jest ustalanie budowy cząsteczek białek, począwszy od ustalenia sekwencji reszt aminokwasowych w łańcuchu polipeptydowym, czyli poznaniu I-rzędowej struktury białka, aż do określenia struktur IV-rzędowych, czyli agregatów złożonych z większej liczby łańcuchów polipeptydowych. Dodatkową trudność sprawiają modyfikacje potranslacyjne, w wyniku których do łańcucha polipeptydowego przyłączane są różne mniejsze cząsteczki. Na każdym etapie ustalania budowy białek wiodącą rolę odgrywa obecnie spektrometria mas. Za jej pomocą można nie tylko określić precyzyjnie masę cząsteczkową białka, ale także ustalić sekwencję łańcucha polipeptydowego i – chociaż nie zawsze – rodzaj i lokalizację modyfikacji potranslacyjnych.

Głównym problemem ograniczającym obecnie wykorzystanie spektrometrii mas do analizy białek jest konieczność przetwarzania ogromnej ilości danych generowanych przez spektrometr mas. Co gorsza, ze względu na samą istotę pomiaru, widma masowe są dość niewdzięcznym obiektem analizy matematycznej. Składają się na to niedoskonałości kalibracji skali  $m/z$ , ograniczona zdolność rozdzielcza spektrometrów, często słaby stosunek sygnału do szumu, słaba powtarzalność względnych intensywności sygnałów (problem statystyki jonów i sprawność detektorów), a także nakładanie się na siebie widm pochodzących od różnych związków, co z kolei wynika z niedoskonałości technik rozdzielania chromatograficznego białek i peptydów. Dlatego też konieczny jest ciągły rozwój technik matematycznych i informatycznych, określanych łącznie mianem bioinformatyki, które ułatwią przetwarzanie ogromnych ilości danych i doprowadzą do możliwie jednoznacznych wyników analitycznych, czyli identyfikacji białek znanych lub ustalenia budowy do tej pory nieznanymi. Tego właśnie obszaru nauki dotyczy rozprawa doktorska mgr. Mateusza Łąckiego. Doktorant podjął się bardzo ambitnego zadania, ponieważ dziedzinę jego badań można określić jako „gorącą”, co wiąże się z dużą konkurencją ze strony innych zespołów badawczych. Wyniki opisane w pracy doktorskiej, a przede wszystkim w publikacjach naukowych, które są wynikiem badań Doktoranta wskazują, że mgr Łącki wyszedł z tej próby zwycięsko.

Recenzowana praca powstała dzięki współpracy naukowej zespołów prof. Gambin i zespołu prof. Sobotta z Uniwersytetu w Antwerpii. Jest to bardzo dobry przykład efektywnej

współpracy grup badaczy z dwóch pozornie bardzo odległych od siebie dziedzin: matematyki i chemii białek. Jednak, jak wspomniałem powyżej, taka współpraca jest obecnie koniecznością, ponieważ z jednej strony specjaliści w dziedzinie proteomiki potrzebują skutecznych narzędzi analizy danych, a z drugiej strony matematycy chcą pokazać, że opracowywane przez nich algorytmy są właśnie takim narzędziem. Wszystkie praktyczne testy procedur opracowanych przez Doktoranta zostały wykonane z wykorzystaniem widm masowych dostarczonych przez zespół prof. Sobotta.

Praca doktorska mgr. Łackiego napisana jest w sposób klasyczny, czyli w formie rozprawy. Liczy ona 121 stron i składa się z sześciu rozdziałów, po których następuje spis literatury liczący ok. 130 pozycji. Piszę „około”, ponieważ odsyłacze nie zostały ponumerowane, a ich przeliczenie to dla chemika już wyższa matematyka... Rozprawa napisana jest w języku angielskim, co niewątpliwie ułatwi jej rozpowszechnienie w międzynarodowym środowisku naukowym. Należy podkreślić już w tym miejscu, że Doktorant bardzo biegle posługuje się tym językiem, będącym *lingua franca* współczesnej nauki. Wyniki opisane w recenzowanej pracy doktorskiej zostały opublikowane w dobrych czasopiśmie międzynarodowych i materiałach konferencyjnych. Doktorant nie skorzystał jednak z możliwości napisania jedynie zwięzłego przewodnika po tych pracach i wybrał formę pełnej rozprawy.

Zanim przystąpię do oceny merytorycznej recenzowanej rozprawy chciałbym wyraźnie podkreślić, że jako chemik specjalizujący się w badaniach związków organicznych i ich reakcji metodami spektrometrii mas nie jestem w stanie ocenić matematycznej poprawności opisanych przez Doktoranta procedur i algorytmów. Z tego obowiązku zwolnili mnie w pewnym stopniu recenzenci uznanych czasopism naukowych, w których zostały opublikowane wyniki opisane w rozprawie. Dlatego też w swojej recenzji skoncentruję się na ocenie tego, w jakim stopniu te wyniki mogą się przyczynić do usprawnienia badań w dziedzinie proteomiki i ew. innych zastosowań spektrometrii mas.

Rozprawę doktorską rozpoczyna rozdział zatytułowany „Introduction”. Mgr Łacki przedstawia w nim w zwięzły sposób tło swojej pracy oraz zawartość kolejnych rozdziałów. Do tej części pracy mam jedynie kilka drobnych zastrzeżeń, o których napiszę w dalszej części recenzji.

Pierwszy merytoryczny rozdział rozprawy nosi tytuł „Isotopic Distribution Calculations” i zgodnie z nim dotyczy metodyki obliczania profili izotopowych jonów, zwłaszcza tych o wysokiej i bardzo wysokiej masie cząsteczkowej. Programy do obliczania profili izotopowych jonów są oczywiście znane od dawna i wchodzą w skład wszystkich pakietów oprogramowania do rejestracji i przetwarzania widm masowych, ale zawodzą w przypadku bardzo dużych cząsteczek. Algorytm zaproponowany przez Doktoranta, ze względu na udowodnioną liniową zależność czasu obliczeń od, mówiąc w uproszczeniu, wielkości cząsteczki ma zastosowanie do praktycznie dowolnie dużych cząsteczek, a na pewno wszystkich, dla których daje się zarejestrować widmo masowe. O wartości tej części pracy doktorskiej mgr. Łackiego świadczy fakt opublikowania opracowanego algorytmu w prestiżowym czasopiśmie *Analytical Chemistry*. Algorytm ten jest wykorzystywany także jako element bardziej złożonych procedur opisanych w dalszej części rozprawy.

W następnym rozdziale rozprawy zatytułowanym „Quantifying Electron Transfer Reactions” Doktorant podejmuje niezwykle ambitne zadanie statystycznej analizy procesów fragmentacji łańcuchów polipeptydowych w wyniku transferu elektronu (metoda ETD). Technika ETD jest obecnie bardzo intensywnie rozwijana, ponieważ, w odróżnieniu od fragmentacji indukowanej kolizyjnie (CID), jest znacznie bardziej selektywna i w jej wyniku tworzą się praktycznie wyłącznie jony fragmentacyjne szeregów *c* i *z*, co bardzo ułatwia interpretację widm i ustalanie sekwencji peptydów. Niemniej jednak widma nadal są dość złożone ze względu na fakt, że przeniesienie elektronu do wielokrotnie naładowanego jonu peptydu nie zawsze skutkuje rozerwaniem łańcucha, a możliwe są także inne procesy. W celu wsparcia interpretacji widm ETD peptydów został opracowany algorytm o nazwie MassTodon. Ani z rozprawy, ani z autoreferatu mgr. Łackiego nie wynika jednoznacznie, jaki był jego udział w powstaniu tego algorytmu - teksty napisane są w liczbie mnogiej, a podsumowująca wyniki publikacja jest wieloautorska. Niemniej jednak należy przyjąć, że ten udział był dominujący lub przynajmniej bardzo znaczący na etapie tworzenia algorytmu. Wyniki przeprowadzonych testów skuteczności algorytmu z wykorzystaniem m. in. widm ETD dwóch peptydów wykazały jego użyteczność i wskazały także drogę dalszego rozwoju.

Kolejny rozdział, zgodnie ze swoim tytułem „Estimating Reaction Kinetics of Electron Transfer Reactions”, dotyczy przewidywania na podstawie opracowanego modelu matematycznego prawdopodobieństwa przebiegu określonych reakcji fragmentacji łańcucha peptydowego w warunkach ETD. Stworzone narzędzie o nazwie ETDetective łącznie z opisanym w poprzednim rozdziale programem MassTodon ma za zadanie zwiększyć prawdopodobieństwo poprawnej interpretacji widm ETD poprzez przewidywanie względnych intensywności sygnałów w widmie na podstawie przyjętej sekwencji peptydu i porównywanie ich z widmem doświadczalnym. Pierwsze próby zastosowania tego narzędzia okazały się obiecujące, ponieważ powala ono na wyjaśnienie obecności do 80% pików w widmie.

W ostatnim rozdziale pt. „Deconvolution of Mass Spectra” mgr Łacki opisuje opracowaną przez siebie metodę dekonwolucji widm masowych, a także próby powiązania sygnału detektora spektrometru mas z rzeczywistą liczbą jonów, które do niego docierają. Dekonwolucja widm jest szczególnie przydatna w często spotykanej sytuacji nakładania się na siebie widm różnych związków. Analiza statystyki jonów pozwala z kolei lepiej oszacować rzeczywiste prawdopodobieństwa przebiegu konkurencyjnych procesów fragmentacji.

Rozprawę zamyka zwięzłe podsumowanie osiągniętych rezultatów oraz przedstawienie planów na przyszłość. Opracowane przez mgr. Łackiego algorytmy mogą stać się podstawą rozbudowanego programu wspomagającego interpretację złożonych widm masowych. Biorąc pod uwagę wspomniany już na początku niniejszej recenzji stały rozwój technik spektrometrii mas i lawinowy wzrost ilości danych dostarczanych przez spektrometry, takie programy staną się wręcz niezbędne.

Podsumowując merytoryczną część recenzji pragnę stwierdzić, że recenzowana praca reprezentuje niewątpliwie bardzo wysoki poziom merytoryczny. Nie potrafię ocenić jej z punktu widzenia wkładu Doktoranta w rozwój metod statystyki matematycznej, ale dla naukowców zajmujących się spektrometrią mas, a w szczególności spektrometrią peptydów i

białek, praca ma duże znaczenie. Spodziewany dalszy rozwój opisanych w pracy programów wspomagających interpretację widm masowych peptydów i białek doprowadzi do stworzenia narzędzi użytecznych w badaniach proteomicznych. Nie mam wątpliwości, że zarówno ilość materiału przedstawionego w pracy, jak i jego poziom merytoryczny przekracza wymagania stawiane przez odnośną ustawę rozprawom doktorskim. Co więcej, o wartości pracy świadczy także jej interdyscyplinarność – Doktorant wykazał się wiedzą nie tylko z obszaru nauk matematycznych, ale także spektrometrii mas i proteomiki. Należy także podkreślić, że praca jest napisana bardzo poprawnym językiem angielskim i jest bardzo starannie opracowana edytorsko.

Z obowiązku recenzenta muszę zwrócić uwagę na pewne błędy i nieścisłości w pracy, a ściślej w jej fragmentach dotyczących spektrometrii mas. Kilka błędów znajduje się w akapicie rozpoczynającym podrozdział „Understanding Reaction Pathways” na str. 14. Oczywisty błąd zawiera zdanie: „*Ions can be fragmented either outside the instrument, via proteolytic digestion, or inside the instrument*”. Nie ma czegoś takiego jak fragmentacja jonów poza spektrometrem mas – trawienie proteolityczne dotyczy oczywiście białek. Także zdanie „*ETD is much more subtle technique: it consists of an ion-ion reaction between the sample cations and anions, each carrying a radicals – an electron in a higher energy state*” nie jest ściśle. Technika ETD polega na transferze elektronu od anionorodnika do wielokrotnie naładowanego kationu. W wierszu 8 od dołu na str. 46 użyto błędnie sformułowania „...*the radical passes...*” zamiast: „...*the electron passes...*”. Poważniejszy błąd znajduje się na rysunku 4.3 (str. 74). Zaznaczono ma nim, że dwa dwukrotnie naładowane jony po jednej reakcji typu PTR lub ETnoD trafiają „na cmentarz”, czyli całkowicie tracą ładunek, co oczywiście nie jest prawdą, ponieważ wymienione reakcje powodują utratę tylko jednego ładunku. Wymienione błędy i nieścisłości nie obniżają wartości merytorycznej rozprawy, a wynikają z faktu, że trudno jest być specjalistą w dwóch bardzo odległych od siebie dziedzinach.

W konkluzji niniejszej recenzji stwierdzam, że rozprawa doktorska mgr. Mateusza Łackiego pt. „Computational and Statistical Methods for Mass Spectrometry Data Analysis” spełnia z dużym nadstatkiem wymagania określone w ustawie o stopniach i tytule naukowym oraz stopniach i tytule naukowym w dziedzinie sztuki, w związku z czym zwracam się do Rady Wydziału Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego o dopuszczenie mgr. Mateusza Łackiego do dalszych etapów przewodu doktorskiego.

Równocześnie, biorąc pod uwagę bardzo wysoki poziom merytoryczny recenzowanej pracy, potwierdzony publikacją wyników w uznanych czasopismach o zasięgu międzynarodowym, a także interdyscyplinarny charakter pracy, w której jej Autor połączył wiedzę z dziedziny statystyki matematycznej z wiedzą z dziedziny spektrometrii mas i proteomiki, wnoszę o wyróżnienie rozprawy. Nie będąc matematykiem nie potrafię ocenić znaczenia tej pracy z punktu widzenia nauk matematycznych, ale jako praktyk w dziedzinie spektrometrii mas doceniam jej znaczenie dla rozwoju mojej dyscypliny naukowej. Należy też podkreślić, że wysokiemu poziomowi merytorycznemu rozprawy – co nie zawsze ma miejsce – towarzyszy bardzo staranne opracowanie edytorskie. Pisząc pracę w języku angielskim mgr Łacki wykazał także, że problemy językowe na pewno nie będą przeszkodą w jego dalszej karierze naukowej.

