

prof. dr hab. Tomasz Puzyn
Kierownik Pracowni Chemoinformatyki Środowiska

RECENZJA

rozprawy doktorskiej mgr Agnieszki Kamedulskiej
pt. "Opracowanie bayesowskich modeli hierarchicznych opisujących retencję
w wysokosprawnej chromatografii cieczowej w odwróconym układzie faz"

1. Wstęp

Przedstawiona mi do recenzji rozprawa doktorska mgr Agnieszki Kamedulskiej poświęcona jest opracowaniu bayesowskich modeli hierarchicznych w celu przewidywania retencji w wysokosprawnej chromatografii cieczowej, w odwróconym układzie faz. Rozprawa wykonana została w Zakładzie Biofarmacji i Farmakokinetyki w strukturze Katedry Biofarmacji i Farmakodynamiki Wydziału Farmaceutycznego Gdańskiego Uniwersytetu Medycznego, pod kierunkiem prof. dr. hab. Pawła Wiczlinga. Warto tu wspomnieć, że jednostka, w której wykonane zostały badania należy do ścisłej czołówki w zakresie modelowania parametrów chromatograficznych w Polsce i na Świecie – tematyki zainicjowanej przez śp. prof. dr. hab. Romana Kaliszana. Tak duża tradycja i wielkie nazwiska naukowe Poprzedników zobowiązują i stanowią dla doktoranta wyzwanie, a pani magister Agnieszka Kamedulska – jak przedstawię w przedmiotowej recenzji – z tym zobowiązaniem i wyzwaniem poradziła sobie wyśmienicie.

2. Ocena formalna

Praca liczy 87 stron i ma formę zbioru opublikowanych i powiązanych tematycznie artykułów naukowych, czyli tzw. „spinki”. Składa się z przedruków trzech oryginalnych prac w języku angielskim. Dwie z tych prac (pierwsza i trzecia) opublikowane zostały w *Analytical Chemistry* (IF_{5-letni} = 7,0), a kolejna (druga) – w *Analytical and Bioanalytical Chemistry* (IF_{5-letni} = 4,0). Są to wiodące czasopisma, zakwalifikowane przez *Web of Science* do pierwszego kwartyła w kategorii tematycznej *analytical chemistry*. Należy więc założyć, że wszystkie manuskrypty przed publikacją przeszły wnikliwy proces oceny z udziałem zagranicznych recenzentów. Znacznie ułatwiło mi to przygotowanie niniejszej recenzji. W dwóch z tych trzech prac mgr Kamedulska jest pierwszą autorką. Oryginalne publikacje zostały opatrzone czterdziestoośmiostronicowym komentarzem autorskim w języku polskim oraz streszczeniami: w języku polskim i angielskim. Formalna strona

rozprawy spełnia zatem wymogi opisane w Art. 187 Ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 r. Dodatkowo spełnione są również wymogi określone w dokumencie wewnętrznym Gdańskiego Uniwersytetu Medycznego pt. „Zasady wyróżniania rozprawy doktorskiej”.

Do rozprawy Doktorantka załączyła (odrębnie) oświadczenia dotyczące wkładu pracy współautorów publikacji, na podstawie których ocenić można było umiejętność samodzielnego prowadzenia pracy naukowej przez mgr Kamedulską.

3. Ocena merytoryczna

Tytuł rozprawy odpowiada zawartym w niej treściom i nie budzi moich uwag.

Lektura pierwszego rozdziału komentarza, zatytułowanego „Wstęp”, już od pierwszych zdań buduje w czytelniku głębokie przekonanie, że Doktorantka posiada rozległą wiedzę teoretyczną w zakresie farmacji, chemii analitycznej (ze szczególnym uwzględnieniem nowoczesnych technik chromatograficznych) oraz metod modelowania matematycznego (zarówno w ujęciu mechanistycznym, jak i statystycznym). Mgr Kamedulska zaprezentowała również rzadko spotykaną umiejętność syntetycznego i precyzyjnego zarazem referowania złożonych treści. Z obowiązku recenzenta wymieniam poniżej niewielkie potknięcia, które nie wpływają na moją wysoką ocenę tej części rozprawy:

- Pojęcie „pKa” (str. 12) nie zostało zdefiniowane przy pierwszym użyciu w pracy. Zakładam, że chodzi o ujemny logarytm dziesiętny ze stałej kwasowej?
- Na stronie 13. Doktorantka pisze: „Wartości π_i maleją wraz z polaryzacją w systemach z odwróconymi fazami”. W kontekście poprzedniego zdania nie jestem przekonany, czy Autorka miała na myśli właśnie *polaryzację*, czy też może *polarność*? Przy tej okazji proszę o odpowiedź podczas publicznej obrony na pytanie: Na czym polega różnica między *polarnością* a *polaryzowalnością* związku chemicznego?
- Na tej samej stronie (str. 13.) mgr Kamedulska definiuje współczynnik podziału n-oktanol / woda (log P). O ile zastosowanie tej wielkości dla związków niezjonizowanych nie budzi moich zastrzeżeń, to zabrakło mi w pracy odniesienia do współczynnika log D stanowiącego rozszerzenie współczynnika log P właśnie do przypadku związków zjonizowanych. Jaka relacja zachodzi pomiędzy obydwoma współczynnikami?
- Na kolejnej stronie (str. 14.) Autorka wymienia pożądane cechy rozpuszczalników stosowanych w wysokosprawnej chromatografii cieczowej (HPLC). Wśród nich znalazła się „mała toksyczność”. Moim zdaniem zabrakło tutaj wyjaśnienia, w jakim

kontekście jest to ważne. Proszę o krótki komentarz podczas obrony. Czy Doktorantka potrafiłaby wskazać, jakie związki lub grupy związków chemicznych wymieniane są w gronie tzw. „zielonych rozpuszczalników” stanowiących alternatywę dla dotychczas wykorzystywanych w HPLC rozpuszczalników o dużej toksyczności?

- Na str. 20. Doktorantka pisze (o modelach statystycznych): „Mogą one jednak nie uchwycić pełnej złożoności mechanizmów separacji. Są również często nieekstrapolowalne, tj. mają ograniczone zastosowanie poza zakresem związków używanych do opracowania modelu”. Bardzo proszę o głębsze wyjaśnienie obu tych twierdzeń. Jakie metody możemy zastosować w celu określenia tzw. dziedziny modelu?

W drugim rozdziale Autorka krótko opisuje cele rozprawy. Są one przedstawione w sposób zrozumiały i jasno wskazujący na element nowości naukowej (oryginalne rozwiązanie problemu naukowego), którym jest wprowadzenie bayesowskich modeli hierarchicznych w miejsce wykorzystywanego dotychczas podejścia statystycznego (modeli ilościowych zależności pomiędzy strukturą a retencją związku, z ang. *Quantitative Structure-Retention Relationship, QSRR*) w ujęciu niehierarchicznym. Zabrakło mi tutaj zapisania hipotez badawczych, które w rozprawie pozostają domyślne. Proszę Doktorantkę o zapisanie najważniejszej hipotezy (hipotez) i przedstawienie jej w prezentacji pokazywanej podczas obrony.

W rozdziale trzecim mgr Kamedulska krótko definiuje modelowane wielkości oraz przedstawia założenia tworzonych modeli hierarchicznych. Na szczególną pochwałę zasługuje umieszczenie odnośników do repozytoriów zawierających stworzone przez Doktorantkę skrypty w języku R implementujące omawiane w pracy metody.

Rozdział czwarty zawiera syntetyczne podsumowanie wyników prezentowanych w każdej z trzech opublikowanych prac wchodzących w skład rozprawy doktorskiej, a więc:

- A. Wiczling, P., Kamedulska, A. & Kubik, Ł. Application of Bayesian Multilevel Modeling in the Quantitative Structure–Retention Relationship Studies of Heterogeneous Compounds. *Anal. Chem.* 93, 6961–6971 (2021).
- B. Kamedulska, A., Kubik, Ł. & Wiczling, P. Statistical analysis of isocratic chromatographic data using Bayesian modeling. *Anal. Bioanal. Chem.* 414, 3471–3481 (2022).
- C. Kamedulska, A. et al. Toward the General Mechanistic Model of Liquid Chromatographic Retention. *Anal. Chem.* 94, 11070–11080 (2022).



Tę część rozprawy (wraz z oryginalnymi pracami, do których odnosi się Doktorantka) oceniam najwyżej. Najważniejszym oryginalnym osiągnięciem pracy [A] było stworzenie bayesowskiego modelu hierarchicznego oceniającego wpływ masy molowej i rodzaju grup funkcyjnych na retencję. Zaproponowana została również nowa metoda wizualizacji niepewności przewidywania (tzw. chromatogram niepewności). W pracy [B] przedstawiony został bayesowski model hierarchiczny dla retencji w warunkach izokratycznych. Pokazano, że wykorzystanie wartości współczynnika podziału $\log P$ oraz stałej kwasowej pK_a rozdzielanych związków jako predyktorów jest niewystarczające i wymaga uzupełnienia o dodatkowe parametry uzyskiwane eksperymentalnie. Natomiast w pracy [C] przewidywania retencji prowadzone były w warunkach gradientowych przy uwzględnieniu różnych wartości pH, typów modyfikatora organicznego, temperatury oraz czasu trwania elucji gradientowej. Z załączonych oświadczeń współautorów powyższych prac wynika, że wkład Doktorantki w badania udokumentowane pracami był dominujący.

W kolejnym – piątym – rozdziale Doktorantka przeprowadza dyskusję przedstawiając koncepcję hierarchicznych modeli bayesowskich niejako w kontrze do powszechnie wykorzystywanych w celu przewidywania parametrów retencji modeli QSRR. Odniosłem jednak wrażenie, że ta dyskusja nie jest do końca *fair* w odniesieniu do metodyki QSRR (jednoznacznie wynika z niej znacznie większa sympatia Autorki dla modeli Bayesa). Broniąc klasycznych modeli QSRR chciałbym poprosić mgr Kamedulską o przedyskutowanie podczas publicznej obrony następujących kwestii:

1. Autorka powołuje się na prace wykorzystujące QSRR, w których deskryptory (predyktory) wybierane są do modelu przy wykorzystaniu algorytmów selekcji zmiennych (odrębnych lub „zaszytych” w metodzie modelowania). Takie podejście – zdaniem Doktorantki (z którym się zgadzam) – prowadzić może do prognoz „zaszumionych” oraz w wielu przypadkach do ograniczonej możliwości interpretacji mechanizmu retencji na podstawie deskryptorów wybranych przez algorytm do modelu. W przeciwieństwie do metod QSRR, w podejściu bayesowskim deskryptory wybierane są *a priori*, bazując na założeniach dotyczących mechanizmu. Powstaje w tym miejscu pytanie: Czy nie byłoby możliwe (i rozsądne) wybranie odpowiedniej kombinacji deskryptorów do modelu QSRR *a priori* w celu zweryfikowania hipotezy dotyczącej wpływu wybranych predyktorów na modelowane parametry retencji?
2. Doktorantka twierdzi ponadto, że: „Równania QSRR wyjaśniają pewną część zmienności między analitami (parametry na poziomie populacji). Jednak dokładność prognozy opartej tylko na ich podstawie jest raczej ograniczona”. Powstaje pytanie, czy podejścia hierarchicznego nie można byłoby zastosować również w przypadku metod QSRR wychodząc od modelu „globalnego” (mającego zastosowanie dla całej populacji), a następnie serii modeli „lokalnych” (opracowanych dla poszczególnych grup związków o zbliżonym mechanizmie retencji)?

3. Czy dałoby się przeprowadzić porównanie efektywności metod klasycznych (QSRR) z metodami bayesowskimi przy wykorzystaniu tego samego zbioru deskryptorów (predyktorów)? Jest to o tyle istotne, że właściwy wybór deskryptorów o wiele bardziej wpływać powinien na jakość przewidywania niż sama metoda.

Komentarz autorski zamyka poprawnie przygotowany spis bibliografii zawierający 75 aktualnych pozycji.

Po lekturze rozdziałów od 2 do 6 i zapoznaniu się z oświadczeniami współautorów nie mam najmniejszej wątpliwości, że mgr Kamedulska posiadała umiejętność samodzielnego prowadzenia pracy naukowej, a przedmiotem przedstawionej do oceny rozprawy jest oryginalne rozwiązanie problemu naukowego.

4. Ocena strony edytorskiej

Obowiązkiem recenzenta jest również ocena edytorskiej strony dysertacji. Rozprawa jest napisana przejrzysto, przy użyciu poprawnego języka. Tekst jest sformatowany poprawnie przy użyciu pakietu LaTeX, co czyni pracę bardzo czytelną. Zdziwiło mnie użycie w rozprawie archaicznej już w pracach naukowych formy bezosobowej (np. „prowadzono”), zamiast formy osobowej (np. „prowadziłam”), wskazującej jednoznacznie, kto wykonał dane badania.

5. Wnioski końcowe

W mojej ocenie rozprawa mgr Agnieszki Kamedulskiej spełnia wymagania formalne i merytoryczne określone w Ustawie Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 r. stawiane rozprawom doktorskim. Wnioskuje zatem o dopuszczenie Kandydatki do dalszych etapów przewodu doktorskiego.

Dodatkowo, biorąc pod uwagę dużą wartość merytoryczną pracy wnioskuje o przyznanie mgr Agnieszce Kamedulskiej wyróżnienia zgodnie z procedurą przyjętą w Gdańskim Uniwersytecie Medycznym.

Gdańsk, 4 września 2023 r.



