

Poznań, 30 grudnia 2024r.

dr hab. Anna Philips, prof. ICHB PAN
Pracownia Bioinformatyki
Instytut Chemii Bioorganicznej PAN

Recenzja
rozprawy doktorskiej mgr Sophii Bałdysz,
pt. „A new identification method of phage cell wall-associated lytic proteins”

Przedłożona do recenzji praca doktorska została wykonana w Zakładzie Wirusologii Molekularnej Instytutu Biologii Eksperymentalnej na Wydziale Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu pod kierunkiem prof. UAM dr hab. Roberta Nawrota oraz dr hab. Jakuba Barylskiego, pełniącego funkcję promotora pomocniczego.

Rozprawa doktorska mgr Sophii Bałdysz została przygotowana w języku angielskim w formie autorskiego manuskryptu liczącego, wraz z załącznikami, 196 stron. Praca została opracowana zgodnie z klasycznym układem monografii naukowej, typowym dla prac doktorskich w dziedzinie nauk ścisłych i przyrodniczych. Rozprawa otwiera się obszernym wprowadzeniem (*Introduction*), które przedstawia problematykę badań. Następny rozdział zawiera jasno określone cele pracy (*Aim of Thesis*), a kolejny szczegółowy opis zastosowanych metod oraz wykorzystanych zestawów danych (*Methods*). W części poświęconej wynikom (*Results*) Autorka precyzyjnie przedstawiła narzędzia, które opracowała na potrzeby realizacji celów badawczych rozprawy a także wyniki przeprowadzonych testów i eksperymentów. Pracę wieńczy sekcja *Dyskusja* oraz krótkie podsumowanie wyników (*Results Summary*). Dodatkowo, rozprawa zawiera rozdział podsumowujący najważniejsze wnioski oraz omawiający kierunki dalszych badań (*Conclusions and Future Directions*). Monografia została uzupełniona o streszczenia w języku polskim i angielskim, wykaz skrótów, listę literatury obejmującą 226 pozycji, a także suplementarne ryciny i tabele. Autorka nie zamieściła natomiast informacji o swoim dorobku publikacyjnym, co w mojej ocenie powinna zrobić, gdzie jest on ściśle związany z tematyką rozprawy i podnosi walory naukowe pracy. W szczególności Pani Sophia jest pierwszą autorką dwóch artykułów przeglądowych związanych z tematyką pracy w *Frontiers in Microbiology* i w *Applied Environmental Microbiology* oraz współautorką pracy w *Cells* dotyczącej identyfikacji nowego enzymu pochodzącego od faga, a celującego w bakterię *Rothia*.

Rozprawa została bardzo dobrze przygotowana pod względem językowym i edytorskim. Poza kilkoma uwagami wymienionymi poniżej nie zgłaszam istotnych zastrzeżeń.

1. Brak numeracji stron.
2. W niektórych tabelach zastosowanie przecinka jako separatora dziesiętnego.
3. Zmienione czcionki i ich rozmiar, szczególnie w przypadku odnośników do tabel i rycin.
4. Tabela 14 – dwie kolumny mają ten sam nagłówek.

Ocena Merytoryczna

Rozprawa dotyczy badań nad litycznymi białkami bakteriofagów, które są kluczowymi czynnikami infekcyjnymi wirusów bakteryjnych. Temat ten jest niezwykle aktualny w kontekście rosnącej liczby infekcji powodowanych przez bakterie odporne na antybiotyki i ma znaczenie zarówno dla biologii podstawowej, jak i zastosowań praktycznych. Lityczne białka bakteriofagów mogą stanowić podstawę dla nowych terapii przeciwbakteryjnych, szczególnie w walce z wielolekoopornymi patogenami, takimi jak *Proteus mirabilis* czy *Enterococcus*. Ponadto, ich zastosowanie w przemyśle kosmetycznym, spożywczym i biotechnologicznym podkreśla ich wszechstronny potencjał użytkowy. Autorka omawia we *Wstępie* te zagadnienia i wprowadza czytelnika w tematykę związaną z bakteriofagowymi białkami litycznymi, podkreślając ich rolę w cyklu infekcyjnym wirusów oraz ich potencjalne zastosowanie w medycynie i przemyśle. Autor podkreśla także trudności związane z tradycyjnymi metodami poszukiwania litycznych enzymów, które są czasochłonne, kosztowne i często mało efektywne. Zastosowane przez nią podejście, oparte na nowoczesnych technikach uczenia maszynowego, jest odpowiedzią na te wyzwania. W rozprawie opracowała algorytm przewyższający istniejące narzędzia pod względem dokładności i wiarygodności przewidywań, co znacząco zwiększa efektywność identyfikacji nowych białek litycznych. Model ten bazuje na analizie właściwości fizykochemicznych białek i domen konserwatywnych, co czyni go innowacyjnym narzędziem, które może zrewolucjonizować poszukiwanie nowych enzymów.

Głównym celem pracy było opracowanie nowej metody identyfikacji białek litycznych związanych ze ścianą komórkową bakteriofagów. Podjęcie tematyki pracy jest uzasadnione, gdyż biologiczne bazy danych zawierają tysiące sekwencji białkowych, w tym potencjalnych litycznych, jednak brak jest wszechstronnego i zweryfikowanego narzędzia umożliwiającego efektywne wyszukiwanie tych enzymów. Dostępne obecnie narzędzia (phiBiScan i Lypred) są przestarzałe, nie są już aktualizowane, a ich silniki wyszukiwania nie uwzględniają nowych domen litycznych.

W ramach realizacji celu pracy Autorka podjęła następujące zadania szczegółowe:

1. Przeprowadzenie przeglądu istniejących narzędzi i baz danych wykorzystywanych w badaniach nad enzymami litycznymi.
2. Identyfikacja domen konserwatywnych typowych dla enzymów litycznych oraz rodzin białkowych grupujących takie enzymy.

3. Analiza cech sekwencyjnych białek litycznych jako podstawy do stworzenia modelu uczenia maszynowego.

4. Opracowanie modelu bioinformatycznego łączącego techniki uczenia maszynowego i ukryte modele Markowa (HMM) w celu odróżniania białek litycznych od nielitycznych oraz wspomagania odkrywania nowych enzymów litycznych.

5. Przeprowadzenie oceny opracowanego modelu oraz porównanie jego skuteczności z obecnie stosowanymi metodami bioinformatycznymi.

6. Eksperymentalna weryfikacja aktywności wybranych białek zidentyfikowanych przez opracowany klasyfikator oraz ocena ich właściwości biochemicznych.

Pierwszym etapem pracy było wybranie reprezentatywnych ukrytych modeli Markowa (pHMMs), które najlepiej odzwierciedlają domeny i rodziny związane z aktywnością lityczną. Proces selekcji zbioru RepLysin66 wyróżniał się precyzją i wieloetapowym podejściem, które połączyło zaawansowane analizy statystyczne, bioinformatyczne i manualną kurację wyników. Ostatecznie wybrano 66 modeli a selekcja obejmowała m. in.: przeszukiwanie baz danych pHMMs, analizę nadreprezentacji domen w stosunku do zbioru kontrolnego, grupowanie i selekcję reprezentatywnych modeli w programie Cytoscape, uwzględniając liczby trafień oraz funkcjonalne adnotacje. Wyjątkowa precyzja metodologii zapewniła eliminację redundancji i wybór najbardziej informacyjnych modeli. A jak wiadomo, jedynie wiarygodny i dobrze przygotowany zbiór uczący jest w stanie doprowadzić do otrzymania wiarygodnych wyników. RepLysin66 stał się podstawą do trenowania modeli uczenia maszynowego oraz identyfikacji nowych białek litycznych, co potwierdzono w analizach metagenomicznych i walidacjach eksperymentalnych w kolejnych etapach pracy. **Nasuwa się jednak pytanie czy kryteria manualnej selekcji reprezentantów w klastrach (np. liczba trafień) mogły wpłynąć na pominięcie nietypowych lub mniej reprezentatywnych domen litycznych? Jak można by dodatkowo uwzględnić takie przypadki?**

Opracowanie zbioru RepLysin66 było nie tylko kluczowym krokiem w rozwoju nowego narzędzia bioinformatycznego, ale jednocześnie przyczyniło się do lepszego zrozumienia strukturalnych i funkcjonalnych cech białek litycznych. W kolejnej części pracy Doktorantka przeanalizowała cechy fizykochemiczne litycznych i losowo wybranych nie litycznych białek (m. in. aromatyczność, hydrofobowość, stabilność struktury, czy budowa struktury drugorzędowej). Ponadto zidentyfikowała domeny białek litycznych często występujących razem lub zastępujących się. **Ciekawi mnie czy funkcje tych domen są znane i czy ich współwystępowanie ma biologiczne uzasadnienie? Np. w omawiamy przez Autorkę przykładzie dlacego domena Amidase_2 często występuje z domeną CHAP i SH3?**

Głównym wynikiem pracy doktorskiej Pan mgr Bałdysz było stworzenie narzędzia bioinformatycznego do identyfikacji białek litycznych. Autorka zastosowała kombinację technik uczenia maszynowego (sieci neuronowe, Random Forest) oraz ukrytych modeli Markowa, optymalizując je pod kątem identyfikacji domen konserwatywnych białek

litycznych. Autorka z powodzeniem zintegrowała w modelu różnorodne cechy białek, takie jak rodzaj reprezentacji sekwencji czy właściwości fizykochemiczne. Optymalizacja modelu została przeprowadzona przy użyciu technik optymalizacji bayesowskiej, co znacząco podniosło jego skuteczność w klasyfikacji. **Czy przeprowadzono szczegółową analizę wpływu poszczególnych cech fizykochemicznych na wydajność klasyfikacji (np. aromatyczność, struktura drugorzędowa)?**

Wyniki testów wskazują, że stworzony model przewyższa istniejące narzędzia (phiBiScan, Lypred) pod względem kluczowych metryk, takich jak: F1 (balans pomiędzy rozpoznawaniem prawdziwych pozytywów, a identyfikacją fałszywych pozytywów) który średnio wyniósł ponad 0.8 dla klasyfikatora opracowanego w niniejszej pracy. **Kolejnym pytaniem jaki mi się nasuwa to czy opracowane narzędzie może być łatwo zaadaptowane do identyfikacji innych typów białek wirusowych lub enzymów bakteryjnych?**

Warto podkreślić, cały kod źródłowy programu został udostępniony na platformie GitHub. Umieszczenie kodu na GitHub zapewnia transparentność i dostępność projektu, umożliwiając śledzenie historii zmian, co dodatkowo potwierdza samodzielności pracy Doktorantki. Wszystkie zarejestrowane zmiany w kodzie (commity) zostały wprowadzone przez Panią Bałdysz.

Stworzone narzędzie zostało następnie zastosowane do identyfikacji nowych białek litycznych, a wybrane wyniki zostały przetestowane eksperymentalnie. Model został zastosowany do analizy różnych zestawów danych, w tym: środowiskowych danych metagenomicznych – dzięki temu zidentyfikowano szereg białek litycznych z próbek zwierzęcych oraz środowiskowych, w tym unikalne białka skierowane przeciwko *Enterococcus*. Dane pochodzące od pacjentów – w tym zestawie odkryto nowy enzym lityczny skierowany przeciwko *Rothia*, bakterii będącej częstym składnikiem mikrobiomu ludzkiego, ale mogącej wywoływać infekcje oportunistyczne.

Walidacja eksperymentalna potwierdziła skuteczność opracowanego modelu bioinformatycznego. Weryfikacja przeprowadzona w laboratorium pozwoliła na ocenę aktywności enzymatycznej przewidzianych przez klasyfikator białek. Wybrane białka lityczne zidentyfikowane za pomocą modelu zostały wyprodukowane w systemach ekspresyjnych, a następnie aktywność wyprodukowanych białek oceniono za pomocą techniki umożliwiającej wykrywanie zdolności enzymów do trawienia peptydoglikanu, głównego składnika ścian komórkowych bakterii. Wyniki potwierdziły aktywność kilku z przewidzianych białek, w tym nowo odkrytego białka litycznego skierowanego przeciwko: *Proteus mirabilis* – pierwszy przypadek takiego enzymu związanego z fagiem infekującym tego patogena, oraz enzymów aktywnych przeciwko *Rothia* i *Enterococcus*.

Znaczenie pracy wykracza poza granice nauki akademickiej, oferując realne narzędzia do zastosowania w klinikach, laboratoriach badawczych i przemyśle. Praca wpisuje się w globalne wysiłki zmierzające do rozwiązania problemu antybiotykooporności, oferując alternatywne rozwiązania w postaci białek o właściwościach bakteriobójczych. Rozprawa wyróżnia się interdyscyplinarnym podejściem, które łączy bioinformatykę, biologię molekularną oraz walidację eksperymentalną. Imponujące są szerokie zdolności Doktorantki, która z powodzeniem zrealizowała wszystkie etapy projektu – od sformułowania pomysłu, przez dobór danych i stworzenie narzędzia bioinformatycznego, aż po jego praktyczne zastosowanie i eksperymentalne potwierdzenie wyników. Praca stanowi zamkniętą i spójną całość, co jest jej dużym atutem, świadczącym o przemyślanej konstrukcji i umiejętności kompleksowego podejścia do realizacji projektu badawczego. Jej wyniki mogą stanowić solidną podstawę dla przyszłych badań w tej dziedzinie.

Podsumowując, wyniki przedstawione w rozprawie jednoznacznie dowodzą, że Pani mgr Sophia Bałdysz w pełni zrealizowała wszystkie założone cele badawcze. Praca doktorska została przygotowana na wysokim poziomie merytorycznym, a zawarte w niej wyniki wnoszą oryginalny i istotny wkład do nauki. Praca wyróżnia się aktualnością i znaczeniem tematu, oraz interdyscyplinarnością. Przedstawione rozwiązanie naukowego problemu świadczy o solidnej wiedzy teoretycznej Kandydatki w dziedzinie nauk biologicznych oraz o jego zdolności do samodzielnego prowadzenia badań. Uwagi sformułowane powyżej mają charakter dyskusyjny i nie wpływają na wartość naukową rozprawy.

Stwierdzam, że przedstawiona mi do oceny rozprawa doktorska spełnia warunki określone w art. 187 ust. 1-2 i art. 190 ust. 3 Ustawy z dn. 20.07.2018 r. Prawo o Szkolnictwie wyższym i nauce (Dz. U. 2024 poz. 1571) i wnoszę do Rady Dyscypliny Nauki Biologiczne Uniwersytetu im. Adama Mickiewicza w Poznaniu o dopuszczenie Pani mgr Sophii Bałdysz do dalszych etapów przewodu doktorskiego.