

Wrocław, 14.06.2023 r.

Prof. dr hab. inż. Małgorzata Kotulska
Politechnika Wrocławska
Wydział Podstawowych Problemów Techniki
Katedra Inżynierii Biomedycznej

Recenzja rozprawy doktorskiej: mgr inż. Patryk Jarnot
“Methods for similarity analysis of low complexity regions in protein sequences”

1. Charakterystyka rozprawy

Przedstawiona mi do oceny praca doktorska Patryka Jarnota powstała na Wydziale Automatyki, Elektroniki i Informatyki, Politechniki Śląskiej, pod kierunkiem dr hab. inż. Aleksandry Grucy. Drugim promotorem rozprawy był dr hab. Marcin Grynberg. Rozprawa wpisuje się w obszar badań z dziedziny bioinformatyki, ma charakter teoretyczno-obliczeniowy z uwzględnieniem praktycznych implementacji zaproponowanych rozwiązań. Autor opracował nowe metody i algorytmy do wykrywania i analizy podobieństwa obszarów o niskiej złożoności (LCR, ang. *Low Complexity Regions*) i obszarów z powtórzeniami (STR, ang. *short tandem repeat*).

Praca doktorska została napisana w języku angielskim, liczy 137 stron i podzielona jest na 8 rozdziałów. W rozdziale, który otwiera rozprawę, Autor definiuje cele pracy doktorskiej i krótko przedstawia swój dorobek naukowy, który powstał w wyniku jej realizacji. Rozdziały drugi, trzeci i czwarty to wprowadzenie teoretyczne do podjętej tematyki. Najpierw Autor przedstawia w nich metody dedykowane rozpoznawaniu krótkich powtórzeń i obszarów o niskiej złożoności, w których znalazły się: fLPS, CAST, SEG, LCD-Composer, SIMPLE, T-REKS, XSTREAM, GBA, a także klasyczne metody porównywania sekwencji, takie jak BLAST, HHblits, CD-HIT, MMseqs i MCL. Następnie, w rozdziale 3, omawia problem obszarów o niskiej złożoności i porównuje skuteczność i powtarzalność wyników z różnych metod im dedykowanych. W rozdziale 4 omawia klasyczne metody porównywania sekwencji, stosowane do wyszukiwania białek homologicznych i problemy z ich zastosowaniem w przypadku obszarów o niskiej złożoności. W rozdziale 5, Autor przedstawia rozwiązanie, które mogłoby wspomóc zastosowanie BLASTa w takim przypadku, wprowadzając nową metodę nazwaną LCR-BLAST. W rozdziale 6, Autor prezentuje kolejną, całkowicie oryginalną metodę przeznaczoną do wykrywania krótkich powtórzeń, opartą na grafowym podejściu do klastrowania sekwencji. Jej zadaniem jest umożliwienie bardziej precyzyjnego rozpoznawania obszarów sekwencji białkowych z powtórzeniami, w tym również takich o niskiej złożoności. W rozdziale 7 zostały zaprezentowane przykłady zastosowania nowych metod opracowanych przez doktoranta oraz wyniki skuteczności ich działania

wyników, i opublikowane w bardzo dobrym czasopiśmie. Ta część pracy doprowadziła do istotnego rezultatu, ułatwiając świadomy wybór narzędzia optymalnego w konkretnej analizie i najbardziej adekwatnego wyniku.

W kolejnej części pracy, Doktorant przeanalizował działanie klasycznych narzędzi do porównywania sekwencji i poszukiwania białek homologicznych. Zauważył, że w domyślnych konfiguracjach świadomie pomijają lub zaniedbują w swojej analizie obszary o niskiej złożoności. Przeprowadził analizę wyników z ich działania z uwzględnieniem rozmaitych konfiguracji parametrów. Ostatecznie wykazał, że są one bardzo mocno oparte na obszarach o wysokiej złożoności (HCR, ang. *High Complexity Regions*) i segmenty typu LCR mają niewielki wpływ na wyniki ich pracy. Zaproponował więc zestaw parametrów i modyfikację macierzy oceny, które podnoszą wrażliwość BLASTa na obszary inne niż HCR. W analizach porównawczych umożliwi to wykorzystanie fragmentów LCR, zwłaszcza krótkich, zwykle eliminowanych ze względu na podwyższenie wartości *e-value*. Uważam, że jest to bardzo potrzebne rozwiązanie i warto byłoby je włączyć również do klasycznego BLASTa w formie narzędzia internetowego.

Ostatni projekt, zawarty w pracy, dotyczy opracowania metody GBSC (*Graph Based on Sequence Clustering*). Metoda ma na celu identyfikację STR w sekwencjach białkowych i grupuje je według podobnych wzorców w sekwencji, wykorzystując podejście oparte na grafach. Dodatkowo może identyfikować i grupować powtórzenia za pomocą zredukowanych alfabetów. Pozornie metoda ta przypomina podejście zastosowane w klasycznej metodzie CD-HIT, jednak grupowanie przebiega inaczej, co daje możliwość innych analiz, pokazujących inne zależności. To bardzo przydatna nowa metoda, która pozwoli wykorzystać specyficzne alfabety zredukowane, które nie muszą się opierać na klasycznych własnościach fizykochemicznych, ale będą mogły odnosić się do specyficznych wzorców odpowiadających różnym szczegółowym problemom.

Ostatecznie, Autor testuje dwie własne metody na białkach z bazy UniProtKB/Swiss-Prot. Testy przeprowadza najpierw na wybranym motywie aminokwasowym, a następnie na sekwencjach białkowych z proteomu wirusa SARS-CoV-2. Analiza ta pokazuje istnienie w proteomach drobnoustrojów obszarów o niskiej złożoności. Autor wskazuje jak brzemienne w skutki może być ich przeoczenie, co łatwo może się zdarzyć, jeżeli korzysta się wyłącznie ze standardowych narzędzi do porównywania sekwencji, na przykład przy opracowaniu szczepionki. To istotna część pracy, która dowodzi poprawności i stosowalności wyników uzyskanych przez Doktoranta.

Część badawcza rozprawy w spójny pokazuje systematyczną analizę działania istniejących metod oraz wkład Autora w ich rozwinięcie oraz tworzenie nowych metod, jak również ich implementacja w postaci gotowych narzędzi. Zaproponowane metody znalazły zastosowanie w opracowanych programach, z których część jest już udostępniona na portalach internetowych, takich

podkreślić, że praca doktorska jest przygotowana bardzo starannie, zarówno edycyjnie, jak i językowo. Bardzo dobrze się ją czyta – jest przejrzysta i interesująco napisana.

Biorąc pod uwagę omówione powyżej elementy oceny, stwierdzam, że oceniana rozprawa doktorska spełnia wymagania określone w Ustawie z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (tekst jednolity: Dz.U. 2022 r. poz. 574 z późn. zm.) w sprawie szczegółowego trybu i warunków przeprowadzenia czynności w przewodach doktorskich, postępowaniu habilitacyjnym oraz w postępowaniu o nadanie tytułu profesora.

Niniejszym, wnioskuję o dopuszczenie autora do kolejnych etapów przewodu doktorskiego oraz do publicznej obrony przedstawionej rozprawy.

Ponadto, biorąc pod uwagę bardzo wysoki poziom merytoryczny rozprawy, dodatkowo potwierdzony publikacjami w najlepszych czasopismach z dziedziny oraz praktyczną implementacją opracowanych metod, wnioskuję o wyróżnienie rozprawy.

Małgorzata Kotulska