

## Recenzja pracy doktorskiej mgr inż. Moniki Kurczyńskiej

dr hab. Witold Rudnicki, prof. UwB  
Instytut Informatyki, Uniwersytet w Białymstoku,  
ul. Konstantego Ciołkowskiego 1M, 15-245 Białystok,  
tel. +48 85 738 8276  
e-mail: [W.Rudnicki@uwb.edu.pl](mailto:W.Rudnicki@uwb.edu.pl)

Białystok, 31-01-2019

Ocena pracy doktorskiej pt. „*Ocena jakości modeli białek na podstawie parametrów energetycznych i funkcjonalnych modeli przepływowych*” przygotowanej przez mgr inż. Monikę Kurczyńską, pod opieką dr hab. inż. Małgorzaty Kotulskiej.

### **Omówienie rozprawy**

Rozprawa doktorska została przygotowana na podstawie materiału opisanego w pięciu publikacjach, w tym trzech doktorantka była pierwszą autorką. Wszystkie artykuły ukazały się w międzynarodowych recenzowanych czasopismach, jeden z nich w czasopiśmie o wysokim impact factorze PLoS One. Praca składa się z dziewięciu rozdziałów, zawiera również spis skrótów oraz zestawienie dorobku naukowego doktorantki. Spośród 9 rozdziałów rozdział I, II i VII są w istocie spisami publikacji autorki (I i II) oraz spisem literatury (VII), część merytoryczna pracy to rozdziały III (Streszczenie), IV (Wstęp ogólny), V (Tezy pracy doktorskiej), VI (Przebieg i wyniki badań będących podstawą nadania stopnia doktora). Rozdział VIII to streszczenie w języku angielskim a rozdział IX zawiera włączone w całości oryginalne publikacje stanowiące podstawę pracy.

Jak widać z powyższego zestawienia treść merytoryczna pracy jest skoncentrowana w rozdziałach III do VI oraz przede wszystkim w oryginalnych pracach zawartych w rozdziale IX. Reszta pracy ma charakter pomocniczy, dlatego w omówieniu rozprawy skoncentruję się na tych rozdziałach. Uwagi co do struktury pracy umieszczam w sekcji Ocena rozprawy.

### **Ocena rozprawy**

Rozprawa doktorska pani Moniki Kurczyńskiej zawiera opis rzetelnych i innowacyjnych badań mających na celu rozwiązanie ważnych problemów w modelowaniu białek – szerokiej dziedzinie uprawianej przez duże środowisko naukowe. Praca jest generalnie dobrze napisana, zrozumiała i merytorycznie poprawna. Autorka w zasadzie uniknęła błędów językowych, literówek, itp.

Największe zastrzeżenia mam do struktury pracy, która choć zawiera niezbędne elementy jest moim zdaniem niewłaściwie wyważona i prezentowana w dyskusyjnej kolejności. Nie widzę uzasadnienia do przedstawiania spisu prac doktorantki jako dwóch wydzielonych rozdziałów – zajmujących razem nieco ponad stronę – zamiast umieszczenia ich jako pierwszych dwóch sekcji w spisie literatury. Podobnie Tezy Pracy Doktorskiej moim zdaniem powinny zostać umieszczone jako sekcja we wstępie, a nie wydzielane jako oddzielny rozdział zajmujący w sumie 7 linijek. Również umieszczenie anglojęzycznego streszczenia pracy jako rozdziału

przedostatniego, zamiast bezpośrednio po (lub przed) streszczeniem polskim uważam za wybór niefortunny. Zamykając tę część uwag rozdział IX powinien być moim zdaniem umieszczony w formie 5 oddzielnych, numerowanych załączników, a nie zgrupowany w jednym nieustrukturyzowanym rozdziale.

Należy przy tym podkreślić, że ta dość niefortunna organizacja pracy nie była dużą przeszkodą w czytaniu, stanowiła raczej drobną uciążliwość.

Pozostałe aspekty edycyjne pracy są bez zarzutu, autorka ustrzegła się literówek. Jedyne zidentyfikowane przez mnie poważniejsze błędy gramatyczne to niewłaściwy rodzaj w zdaniu „[...]”, że białka z gęstością kontaktów większym niż 10 [...]” (linia 3 na stronie 22). W kilku miejscach autorka zastosowała szyk wyrazów przeniesiony z języka angielskiego, który brzmi źle po polsku, przykładowo: „bioinformatycznego modelowania” zamiast „modelowania bioinformatycznego” (linia 6 od dołu na stronie 15), „są związane z typowymi dla danej metody problemami” zamiast „są związane z problemami typowymi dla danej metody” (pierwsza linia na stronie 16), „Nie zawsze jednak przeniesienie skali chemii kwantowej do makromolekuł wielkości białek jest możliwe” zamiast „Nie zawsze jest jednak możliwe przeniesienie skali chemii kwantowej do makromolekuł wielkości białek” (ostatnie pełne zdanie na stronie 16). Nie są to wielkie uchybienia i rozumiem doskonale, że trudno się ich ustrzec gdy podstawowym językiem oryginalnej literatury naukowej jest język angielski.

Od strony merytorycznej praca jest bardzo dobrze napisana, zawiera dokładne omówienie wyjaśnienie wyników uzyskanych w pracach stanowiących podstawę rozprawy. Zdarzające się drobne uchybienia można pogrupować w dwie klasy.

Pierwsza to niedopowiedzenie pewnych myśli, które dla autorki są tak oczywiste, że nie warto ich pisać. Jednak dla klarowności tekstu warto było je wyartykułować wprost. Przykładowo na stronie 14 w końcowym zdaniu sekcji pierwszej brakuje stwierdzenia, że problem przewidywania struktury jest szczególnie ważny dla białek membranowych, dla których stosunek znanych struktur do sekwencji jest dużo mniejszy niż dla białek cytoplazmatycznych. W pierwszym akapicie sekcji 3 rozdziału IV, autorka podzieliła metody oceny jakości białek na dwie klasy, wymieniła najważniejszych reprezentantów tych klas, ale nie wyjaśniła jakie znaczenie ma to dla jej własnej pracy. Podobnie w sekcji 3.1 tego rozdziału autorka wspomina niewłaźconą do rozprawy pracę Kurczyńska i in. 2017, poświęconą analizie różnych pól siłowych, ale nie wskazuje w jaki sposób wiąże się ona z rozprawą. Na stronie 18, ostatnie zdanie sekcji „Należy jednak zaznaczyć, że model PNP ze względu na swoje ograniczenia nie umożliwia dokładnego odtworzenia eksperymentalnych charakterystyk prądowo-napięciowych modelowanych białek.” pozostawia czytelnika nie będącego specjalistą w dziedzinie w niepewności jakie ten fakt ma znaczenie w kontekście pracy.

Na stronie 24 pożyteczne byłoby podanie wzoru na korelację Kendalla.

Na stronie 25, pożyteczne byłoby wyjaśnienie dlaczego uzyskanie AUROC w przedziałach (0.59,0.78) dla przewodności i (0.70 do 0.78) wskazuje na dobrą jakość cech funkcjonalnych – w oderwaniu od problemu te wartości AUC nie są bardzo imponujące. Na stronie 28, na koniec sekcji 3.3 przydałoby się jedno zdanie podsumowania, które by wyjaśniało/podsumowywało, dlaczego uzyskanie opisanych wyników dla jonów chlorkowych było ważnym wynikiem.

Niżej na tej stronie doktorantka opisuje wynik, uzyskany dla kryterium selektywności, które pozwoliło na znalezienie 82% modeli o  $RMSD < 1\text{\AA}$  i niemal 100% modeli o  $RMSD < 3\text{\AA}$ . Jest to wynik paradoksalny, zważywszy na to, że modele o  $RMSD < 1\text{\AA}$  są podzbiorem modeli o  $RMSD < 3\text{\AA}$ . Niżej na tej samej stronie doktorantka pisze, TOP100 selektywności to zakres od 62:1 do 518:1. To stwierdzenie nie jest jasne dla niespecjalisty. Definicja selektywności jest



podana wcześniej w rozprawie w postaci wzoru, ale przydałoby się nieco dokładniejsze, słowne wyjaśnienie tego pojęcia i jego znaczenia, co ułatwiłoby odbiór rozprawy.

Druga klasa to nieściśności i nietrafione uogólnienia. Na stronie 13 doktorantka pisze „Oczywiście, jako ludzie mamy skłonność do komplikowania prostoty i już od dziesięcioleci próbujemy wprowadzić nowe elementy do tej białkowej układanki. W efekcie na przestrzeni 55 lat dodano do proteogenicznego alfabetu dwa aminokwasy: selenocysteinę i pirolizynę”. Jest to efektowna figura retoryczna, która jednak nie jest prawdziwa, ponieważ jest dokładnie odwrotnie. To Natura nie daje się wtłoczyć do prostych schematów i generuje na każdym kroku wyjątki.

Na stronie 14 autorka rozprawy pisze „Jednak jednym z pierwszych kroków do większości badań bioinformatycznych jest pozyskanie struktury przestrzennej białka”, co jest oczywistą nieprawdą, jako że modelowanie białek jest tylko jedną z wielu obszarów badawczych bioinformatyki. Te zdania nie mają wielkiego znaczenia w kontekście całej rozprawy, ale bez nich byłaby ona na pewno lepsza.

Kwestia doboru cytowanej literatury jest oczywiście bardzo indywidualna. Nie jest możliwe wymienienie całej istotnej literatury przedmiotu, zważywszy, że modelowanie białek jest obszarem badawczym uprawianym przez bardzo wiele zespołów badawczych. Szkoda jednak, że autorka nie zauważyła w swoich cytowaniach znaczącego w skali światowej dorobku polskich zespołów badawczych kierowanych przez Andrzeja Kolińskiego, Adama Liwo, Janusza Bujnickiego czy Krzysztofa Ginalskiego.

Niezależnie od wymienionych powyżej drobnych uchybień i nieściśności, przedstawiona przez doktorantkę rozprawa mi się podobała. Jest dobrze napisana, zawiera wszystkie treści potrzebne do zrozumienia istoty problemu, trafny opis przedstawionych badań i interesującą dyskusję wyników. Autorka wykazała się szeroką, interdyscyplinarną wiedzą o przedmiocie badań i potrafi ją przekazać w sposób jasny, logiczny i poprawnym językiem. Wstęp w przejrzysty sposób definiuje problem i odpowiada na pytanie czemu warto podjąć badania opisane w pracy.

Prace stanowiące podstawę rozprawy zostały opublikowane w dobrych i bardzo dobrych czasopismach międzynarodowych. Rola doktorantki w pracach, w których nie była pierwszą autorką jest precyzyjnie określona a przedmiot badań poprawnie wyodrębniony, układający się w logiczny i spójny obszar tematyczny. Zaproponowana przez doktorantkę metodologia, zrealizowana w pięciu zawartych w pracy publikacjach, doprowadziła do odpowiedzi na pytania zawarte w tezach doktorskiej.

W trakcie lektury pracy moją uwagę zwróciły dwa wybory metodologiczne dokonane przez doktorantkę, których uzasadnienie chętnie usłyszałbym w czasie obrony.

W pracy **Kurczyńska i in., 2016** w analizie statystycznej użyto trzech bardzo podobnych testów statystycznych. Moim zdaniem wystarczyłoby użyć jedynie testu rang Wilcoxona, który jest najbardziej ogólnym z trzech użytych, odpornym na obserwacje nietypowe i niezależnym od rozkładu populacji. Czy użycie pozostałych testów wniosło coś do rezultatu? Jeżeli wyniki testów były różne, to w jaki sposób były uzyskane ostateczne rezultaty?

Z kolei w pracy **Kurczyńska i Kotulska, 2018** zaproponowano metodę rozdzielania modeli prawidłowych od lustrzanych wykorzystującą uczenie maszynowe. Użyto uczenia nienadzorowanego – analiza skupień, popularnie zwana klastrowaniem. To jest dobre podejście jeżeli nie znamy klasy decyzyjnej. Jednak w tym wypadku, na podstawie lektury rozprawy i artykułu wydaje się, że klasa decyzyjna dla modeli znanych białek może być określona z dość dobrą dokładnością na podstawie odległości modelu od struktury prawdziwej i od struktury

lustrzanej. Z jakich powodów nie użyto tej informacji do budowy modeli uczenia nadzorowanego?

Powyższe pytania nie są zarzutami. Wybór metod badawczych jest często sprawą otwartą, a odmienne rozwiązania problemów narzucają się często dopiero po uzyskaniu wyników. Jednak ciekaw jestem motywów podjętych wyborów.

## **Podsumowanie**

Doktorantka jest współautorką pięciu artykułów opublikowanych w międzynarodowych czasopismach. W trzech przypadkach doktorantka była pierwszą autorką.

Poza tym doktorantka aktywnie uczestniczyła w studenckim życiu naukowym. Była współautorką pięciu komunikatów na konferencjach studenckich, dwa razy jako pierwsza autorka.

Doktorantka wykazała w pracy bardzo dobre opanowanie warsztatu naukowego, a w szczególności metod modelowania, sprawność warsztatową, umiejętność stawiania interesujących celów badawczych, formułowania hipotez i stosowania właściwych metod do ich weryfikacji.

Doktorantka wykazała bardzo dobrą znajomość literatury przedmiotu, udokumentowany poprawnym wyborem cytowanych źródeł.

Przeprowadzone przez autorkę badania są wartościowym wkładem w problem odróżniania struktur poprawnych od ich odbić lustrzanych, który jest ważny w modelowaniu białek.

Praca jest przygotowana bardzo dobrze, przedstawia problematykę i wyniki własnych badań w jasny, przejrzysty i zwięzły sposób, wnioski dobrze udokumentowane i uzasadnione.

Rozprawa doktorska mgr inż. Moniki Kurczyńskiej „*Ocena jakości modeli białek na podstawie parametrów energetycznych i funkcjonalnych modeli przepływowo-tych*” stanowi oryginalne rozwiązanie problemu naukowego i wykazuje ogólną wiedzę teoretyczną doktorantki w dziedzinie biocybernetyki i inżynierii biomedycznej. Jest ona również świadectwem umiejętności samodzielnego prowadzenia pracy naukowej. W mojej ocenie rozprawa wypełnia w bardzo dobrym stopniu wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim. W związku z powyższym wnioskuję o dopuszczenie pani Moniki Kurczyńskiej do dalszych etapów przewodu doktorskiego.

Mimo pewnych uchybień w warstwie edytorskiej i drobnych błędów w części teoretycznej pracy, ze względu na bardzo wysoką ocenę jakości własnych osiągnięć naukowych doktorantki zawartych w pracy, wnioskuję o wyróżnienie rozprawy.

dr hab. Witold Rudnicki, prof. UwB