

Prof. dr hab. inż. Sebastian Deorowicz,
Instytut Informatyki
Wydział Automatyki, Elektroniki i Informatyki
Politechnika Śląska
44-100 Gliwice, ul. Akademicka 16

Gliwice, 22.01.2019

Recenzja rozprawy doktorskiej

Tytuł rozprawy:

Ocena jakości modeli białek na podstawie parametrów energetycznych i funkcjonalnych modeli przepływowych

Autor:

mgr inż. Monika Kurczyńska

Promotor:

Prof. dr hab. inż. Małgorzata Kotulska

I. Problematyka naukowa oraz przedmiot rozprawy

Poznanie struktur przestrzennych białek ma kluczowe znaczenie m.in. dla zrozumienia tego jak białka funkcjonują, w jakie wchodzi interakcje, jakie są potencjalne zagrożenia wynikające z mutacji w genomie przekładających się na zmianę jednego z aminokwasów tworzących białko. Te i inne czynniki powodują, że od wielu lat toczą się intensywne prace w tym kierunku. Jedną z możliwości jest zastosowanie technik eksperymentalnych, które jednak są kosztowne i nie zawsze do końca skuteczne. Dlatego też wiele wysiłku włożono w opracowanie różnych podejść obliczeniowych. Mimo upływu lat i poczynionego sporego postępu skuteczność modeli komputerowych w dalszym ciągu pozostawia wiele do życzenia. Z tego też powodu problematyka ta ciągle pozostaje otwarta i stanowi interesujące pole do nowych badań. Recenzowana praca dotyczy właśnie tej problematyki.

II. Analiza treści rozprawy oraz uzyskanych wyników

1. Treść rozprawy

Doktorantka postawiła sobie w pracy następujące tezy:

- **Teza 1. Zastosowanie metod przepływowych w ocenie jakości modeli strukturalnych białek transmembranowych podnosi efektywność wyboru optymalnego modelu, charakteryzującego się także dobrą funkcjonalnością.**
- **Teza 2. Istnieją składowe funkcje energii całkowitej, które umożliwiają rozdzielenie modeli strukturalnych białek, uzyskanych z map kontaktów, na zbiory o orientacji prawidłowej i lustrzanej.**

Rozprawa ma charakter obszernego streszczenia oraz cyklu pięciu artykułów opublikowanych w czasopismach indeksowanych w bazie Journal Citation Reports (JCR). W szczególności rozprawa składa się z dziewięciu części.

Pierwsza część zawiera wykaz publikacji będących podstawą nadania stopnia doktora. W drugiej części zawarto wykaz innych publikacji (głównie konferencyjnych), które są związane z rozprawą doktorską. Ponadto w wykazie tym znajduje się także publikacja przeglądowa z czasopisma indeksowanego w bazie JCR. Część trzecia jest krótkim streszczeniem rozprawy. W części czwartej Autorka przedstawia ogólny wstęp do problematyki modelowania struktury białek. Wymienia tu m.in. niektóre z aktualnie stosowanych metod. Część piąta zawiera sformułowanie obu tez. Część szósta zaczyna się od opisu metodologii i infrastruktury badawczej. Następnie Doktorantka omawia uzyskane wyniki dotyczące obu zaproponowanych tez. Część ta kończy się podsumowaniem uzyskanych wyników i odniesieniem ich do tez. Kolejną część stanowi wykaz literatury cytowanej w opracowaniu. W części ósmej znajduje się streszczenie pracy w języku angielskim. Najobszerniejsza, ostatnia, część zawiera pełne teksty artykułów stanowiących podstawę do nadania stopnia doktora. Prace te to:

- A1 Bogumił M. Konopka, Marika Ciombor, Monika Kurczyńska, Małgorzata Kotulska, Automated procedure for contact-map-based protein structure reconstruction. *Journal of Membrane Biology* 2014; 247(5):409–420.
- A2 Witold W. Dyrka, Monika Kurczyńska, Bogumił M. Konopka, Małgorzata Kotulska, Fast assessment of structural models of ion channels based on their predicted current-voltage characteristics. *Proteins: Structure, Function and Bioinformatics* 2016, 84(2):217–231.
- A3 Monika Kurczyńska, Małgorzata Kotulska, Ion Move Brownian Dynamics (IMBD)—simulations of ion transport. *Acta of Bioengineering and Biomechanics* 2014, 16(4): 107–116.
- B1 Monika Kurczyńska, Ewa Kania, Bogumił M. Konopka, Małgorzata Kotulska, Applying PyRosetta molecular energies to separate properly oriented protein models from mirror models, obtained from contact maps. *Journal of Molecular Modeling* 2016; 22(5):111.

B2 Monika Kurczyńska, Małgorzata Kotulska, Automated method to differentiate between native and mirror protein models obtained from contact maps. *PLOS ONE* 2018; 13(5):e0196993.

Praca [A1] została zrealizowana w 4-osobowym zespole. Zaproponowano w niej m.in. metodę rekonstrukcji modeli białek z map kontaktów. Schemat ten był następnie stosowany do budowania modeli analizowanych w przedstawionej rozprawie. Skuteczność tej metody potwierdzono eksperymentalnie w pracy [A1] na dość obszernym zestawie danych testowych obejmującym 50 modeli dla 205 domen SCOP oraz serię modeli kanału potasowego KcsA w konformacji przewodzącej otwartej (3fb8 w bazie PDB). Dane testowe były przygotowane starannie i obejmowały kilka scenariuszy, m.in. takie, w których wprowadzano szum do map kontaktów, bądź usuwano informacje o sporej części kontaktów.

Zgodnie z deklaracją zawartą w rozprawie, Doktorantka była odpowiedzialna przede wszystkim za badania elektrostatyczne KcsA oraz analizę relacji między elektrostatyką i strukturą. Warte podkreślenia jest, że praca ta została do tej pory zacytowana w literaturze światowej dwunastokrotnie.¹

Praca [A2], zrealizowana również w 4-osobowym zespole, dotyczy wykorzystania oceny jakości modeli kanałów na podstawie obliczonej charakterystyki prądowo-napięciowej. Cechy kanału były wyliczone za pomocą modelu 3D PNP a uzyskane wyniki porównano ze znanymi z literatury wynikami eksperymentów typu *patch-clamp*. W celu poszerzenia zakresu analiz badania dotyczyły nie tylko kanału potasowego KcsA, ale także kanału potasowego MthK, kanału jonowego bramkowego ligandem GLIC oraz α -hemolizyny. Eksperymenty były prowadzone na danych takiego samego typu jak opisane w literaturze, aby możliwe było porównanie wyników obliczeń z wynikami literaturowymi. Uzyskane rezultaty potwierdzają przypuszczenie o zasadności wykorzystania oceny funkcjonalnej modeli. W szczególności bardzo korzystna okazała się cecha selektywności.

Wkład Doktorantki w ten artykuł polegał na przeprowadzeniu obliczeń elektrostatycznych, wykonaniu analiz porównawczych między miarami strukturalnymi, elektrostatycznymi i funkcjonalnymi, a także na opracowaniu rankingów modeli. Praca ta w literaturze światowej została zacytowana trzykrotnie.

Ostatnia z prac dotyczących pierwszej tezy została zrealizowana w 2-osobowym zespole, w którym Doktorantka była pierwszą autorką [A3]. Zaproponowano tu narzędzie Ion Move Brownian Dynamics (IMBD), które służy do symulacji transportu jonowego przez nanopory białkowe. Symulacje te prowadzone są przy użyciu modelu dyskretno-ciągłego BD. Zaproponowane i zaimplementowane narzędzie zostało przetestowane na dwóch różnych kanałach: wąskim kanale potasowym KcsA oraz szerokim kanale β -baryłkowym α -hemolizynie. Wyniki eksperymentalne pokazały, że narzędzie to może być z powodzeniem wykorzystywane do prowadzenia takich symulacji.

Wkład Doktorantki polegał na zaprojektowaniu i zaimplementowaniu narzędzia IMBD oraz na przeprowadzeniu badań. Warte podkreślenia jest, że praca ta została do

¹Dane bibliometryczne podano w recenzji według bazy Web of Science; , na dzień 18.01.2019.

tej pory dwukrotnie zacytowana w literaturze światowej.

Praca [B1], opracowana w zespole 4-osobowym, przy czym Doktorantka jest pierwszą autorką, dotyczy znalezienia statystycznie różnych składowych dla modeli lustrzanych i natywnie zorientowanych białek, których struktura drugorzędowa to głównie α -helisy. Problem, którym zajmują się autorzy dotyczy tego, że modele rekonstrukcji struktur białkowych z map kontaktów generują wiele różnych modeli, z których część ma odwrotną orientację przestrzenną. Automatyzacja poszukiwania takich modeli jest o tyle istotna, że czasami takie modele powinny być wykluczone z dalszych analiz, a czasami, wręcz przeciwnie, możemy być zainteresowani właśnie takimi wariantami. Zadanie to nie jest proste, ponieważ często modele lustrzane i odpowiadające im modele zorientowane prawidłowo różnią się bardziej niż by tego można było oczekiwać. Głównym celem Doktorantki było opracowanie metodologii, która będzie niezależna od rodzaju struktury i nie będzie wymagała wiedzy a priori. Z uwagi na skomplikowanie tematyki, praca [B1] dotyczy jednak tylko białek α -helikalnych a ich klasyfikację oceniano wizualnie. Idea podejścia polega na poszukiwaniu składowych ET, które są statystycznie różne dla modeli lustrzanych i zorientowanych natywnie. Badania eksperymentalne prowadzono na losowo wybranych 55 modelach domek białek α -helikalnych z klasy A bazy SCOP.

Praca ta w literaturze światowej była cytowana jednokrotnie. Głównym wkładem Doktorantki było opracowanie metodologii automatyzującej proces oceny, implementacja części algorytmów, przeprowadzenie części obliczeń a także udział w analizie wyników.

Praca [B2], została opracowana w zespole 2-osobowym, w którym Doktorantka była pierwszą autorką. Celem prowadzonych tutaj badań była weryfikacja, które składowe ET pozwalają na rozróżnienie modeli natywnie zorientowanych od zorientowanych lustrzanie. Jednym z wyników była też automatyzacja procesu klasyfikacji modeli. Badania przeprowadzono na bardzo obszernej bazie ponad 130 tys. modeli strukturalnych, co pozwoliło dość dobrze oszacować skuteczność metody. Głównym wynikiem pracy jest wskazanie, które ET są statystycznie różne dla modeli natywnych i lustrzanych.

Godne podkreślenia jest to, że praca ta ukazała się w renomowanym czasopiśmie PLOS ONE o stosunkowo dużym współczynniku wpływu (Impact Factor). Praca ta do tej pory nie była cytowana, ale może na to mieć wpływ niedługi okres (kilka miesięcy) od jej ukazania się. Wkład Doktorantki w ten artykuł polegał na opracowaniu metodologii, implementacji narzędzia oraz przeprowadzeniu obliczeń. Ponadto uczestniczyła Ona w analizie danych.

2. Najważniejsze wyniki przedstawione w rozprawie

Wyniki przedstawione w rozprawie można podzielić na dwie zasadnicze części. Pierwsza z nich obejmuje prace [A1,A2,A3] i dotyczy pierwszej z postawionych tez. Druga część, zawierająca prace [B1,B2] dotyczy drugiej tezy.

W artykułach [A1,A2,A3] Doktorantka pokazała, że zastosowanie dodatkowo ocenionej profili elektrostatycznych wewnątrz kanału wpływało na lepsze rozróżnienie analizowanych modeli. Zbadane zostały też zależności między strukturą, elektrostatyką

i funkcją modeli. Opracowano wreszcie program pozwalający wykorzystać dynamikę brownowską w modelowaniu struktury białek. Osiągnięcia te uzasadniają pierwszą z postawionych tez. W niektórych z tych artykułów Doktorantka była odpowiedzialna głównie za przeprowadzanie badań i analizę wyników, ale w ostatnim z nich opracowała samodzielnie interesujące narzędzie do symulacji transportu jonów, co według mnie jest bardzo wartościowym wynikiem.

Artykuły [B1,B2] dotyczą sposobów klasyfikacji modeli według tego czy mają one orientację prawidłową czy też lustrzaną. Doktorantka zaproponowała systematyczną analizę ponad 1000 modeli białkowych. Znalazła także najbardziej różnicujące składowe energii. Kolejnym osiągnięciem było zautomatyzowanie procesu oceny co pozwoliło na przebadanie ponad 130 tys. modeli i wybraniu najbardziej różnicujących składowych energetycznych. Tym samym wykazano drugą z postawionych tez. W obu tych artykułach wkład Doktorantki był dominujący a uzyskane wyniki bardzo interesujące.

3. Uwagi merytoryczne

Zasadniczą część rozprawy stanowią publikacje, które podlegały już procesowi recenzji w uznanych czasopismach naukowych. Stąd też trudno o sformułowanie jakichś zasadniczych uwag krytycznych. Tym niemniej mam zastrzeżenia do sformułowania pierwszej z tez. W szczególności chodzi o użycie w niej przymiotnika „optymalny”. Zgodnie ze słownikową definicją oznacza on „najlepszy z możliwych”. W teorii optymalizacji, którą tu należy przywołać, ponieważ mowa jest o optymalizacji parametrów, przez optymalne rozumie się takie rozwiązanie, którego już w żaden sposób poprawić się nie da, gdyż każde inne rozwiązanie jest gorsze, bądź równoważne. Tymczasem z samego sformułowania tezy zdaje się wynikać, że chodzi tylko o rozwiązanie (model), które jest najlepsze spośród badanych i które wcale nie musi posiadać cech optymalności. W moim odczuciu zamienienie w tezie słowa „optymalny” na sformułowanie „jak najlepszego” (bądź podobne) lepiej oddawałoby to co faktycznie zostało osiągnięte.

W rozprawie nie znalazłem wyjaśnienia, dlaczego w eksperymentach opisanych w pracy [A2] dla białka GLIC przygotowano inną liczbę modeli z użyciem zestawu kontaktów o innej charakterystyce. Uwaga ta nie ma charakteru fundamentalnego, ale interesujące byłoby wyjaśnienie tej różnicy.

W rozprawie poświęcono wiele uwagi wynikom jakościowym. Jest to jak najbardziej uzasadnione, ponieważ dążenie do poprawy jakości predykcji struktury białek było osią pracy. Tym niemniej, jak zresztą Doktorantka sama zauważa, dążenie to musi brać pod uwagę także aspekt czasowy, albo inaczej budżet obliczeniowy, który zwykle nie jest nieograniczony. Stąd też chętnie dowiedziałbym się też czegoś odnośnie czasów realizacji poszczególnych analiz. Mimo iż te informacje nie są kluczowe, to jednak byłbym zobowiązany, gdyby Doktorantka odniosła się do tego w trakcie publicznej obrony.

Z powyższą uwagą wiąże się także moje ostatnie pytanie w tej części. W ostatnim akapicie podsumowania pracy [A3] znajdujemy stwierdzenie: „In our next steps we will focus on improvement of the efficiency of the algorithm to shorten the CPU ti-

me and reduce memory usage, by adding the precalculated Lennard–Jones potential grids, and we are preparing a similar procedure for Coulomb interactions.” Ponieważ praca ta pochodzi z roku 2014, a nie znalazłem w streszczeniu informacji o tym czy te usprawnienia zostały poczynione, to prosiłbym o informację czy udało się coś tutaj uzyskać. Może to mieć o tyle znaczenie, że można się domyślać, iż czas obliczeń był stosunkowo długi. Tymczasem samo narzędzie mogłoby się cieszyć popularnością i ew. jego rozwój mógłby być interesujący.

4. Uwagi redakcyjne

Rozprawa została opracowana dość starannie. Niemniej występują w niej nieliczne błędy gramatyczne lub stylistyczne, np.

- [str. 11] „Niekonwencjonalnym podejściem do problemu jest spojrzenie na proces modelowania białek od końca i zweryfikowanie, czy pełni **on** poprawnie swoją funkcję„ — raczej „ono”,
- [str. 13] „kilku **rodzajów pierwiastków**: węgla, tlenu, azotu, wodoru, fosforu i siarki” — raczej „rodzajów atomów” albo „pierwiastków”,
- [str. 17] „które są odpowiednio **sparametryzowanych**”,
- [str. 20] „W konsekwencji prace pokazują, że zastosowanie elektrostatyki i komputerowo obliczonej funkcjonalności modeli może być wykorzystane **jako** ocena jakości” — raczej „do oceny”,
- [str. 22] „białka z gęstością kontaktów **większym** niż 10”,
- [str. 23] „od 75%-82%” — raczej „od 75% do 82%”; podobnie w innych miejscach pracy.

Czasami pojawia się niekonsekwencja w stosowanych oznaczeniach, np. w niektórych miejscach pracy występuje oznaczenie C_α a w innych znajdujemy $C\alpha$. Niekonsekwentnie stosowane są także odwołania do literatury: czasami odwołanie jest ujęte w nawiasy a czasami nie. W przypadku równań matematycznych także chętnie widziałbym bardziej dopracowany skład. Przykładowe równanie (4) (str. 21) powinno być (zgodnie z polskimi zasadami złożone następująco):

$$\Delta F_{max} = \frac{F_{max_{model}} - F_{max_{3fb8}}}{F_{max_{3fb8}}},$$

gdzie:

– ΔF_{max} — odchyłka ...

5. Podsumowanie

Przedstawiona do oceny rozprawa składa się w przeważającej części z artykułów opublikowanych w prestiżowych czasopismach indeksowanych w bazie JCR. Prace te były opracowywane w zespołach składających się z od 2 do 4 osób, przy czym w 3 z tych artykułów Doktorantka jest pierwszą autorką. Tematyka pracy dotyczy problematyki modelowania struktury białek a także oceny czy model ma orientację prawidłową czy też lustrzaną. Tezy pracy uważam za ambitne. Badania eksperymentalne były wszechstronne i z pewnością czasochłonne. Pozytywnie oceniam opracowane narzędzia.

Rozprawa zawiera pewne niedociągnięcia, co wskazałem w niniejszej recenzji. Tym niemniej nie umniejszają one mojej wysokiej oceny. Na zakończenie dodam, że choć nie uważam, aby wskaźniki bibliometryczne były istotne na tak wczesnym etapie kariery naukowej, to jednak wobec faktu, że są one wyróżniające, to je przytoczę. Baza Web of Science zawiera 7 prac, których współautorką jest Doktorantka. Prace te były cytowane 18 razy (11 bez autocytowań) a indeks Hirscha Doktorantki wynosi 2.

III. Konkluzja

Rozprawa doktorska mgr inż. Moniki Kurczyńskiej zawiera oryginalne i interesujące wyniki naukowe dotyczące modelowania struktury przestrzennej białek. Wykazuje również dużą ogólną wiedzę Doktorantki w dyscyplinie biocybernetyka i inżynieria biomedyczna (nauki techniczne) i świadczy o umiejętności samodzielnego prowadzenia pracy naukowej. Zawarte w recenzji uwagi krytyczne nie wpływają na moją ogólną bardzo dobrą ocenę rozprawy. Uważam, że zostały spełnione wymagania stawiane rozprawom doktorskim przez *Ustawę o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 roku*. Wnoszę zatem o dopuszczenie wspomnianej rozprawy do publicznej obrony.

Ponadto, biorąc pod uwagę wysoką jakość uzyskanych wyników oraz fakt, iż rezultaty prac zostały już opublikowane w kilku artykułach w renomowanych czasopismach naukowych wnoszę o wyróżnienie pracy.

