



Toruń, dnia 12.05.2026

Dr hab. Karolina Mikulska-Rumińska, prof. UMK
Katedra Biofizyki, Instytut Fizyki
Wydział Fizyki, Astronomii i Informatyki Stosowanej
Uniwersytet Mikołaja Kopernika w Toruniu
ul. Grudziądzka 5, Toruń

RECENZJA

Pracy doktorskiej mgr Mateusza Zalewskiego

pt.: „Metody wielkoskalowego modelowania oddziaływań białko-białko i białko-peptyd”

1. Wprowadzenie

Przedstawiona mi do recenzji rozprawa doktorska autorstwa pana mgr Mateusza Zalewskiego powstała na Wydziale Chemii w Centrum Nauk Biologiczno-Chemicznych Uniwersytetu Warszawskiego pod kierunkiem prof. dr hab. Sebastiana Kmiecika oraz prof. dr hab. Dimy Kozakova z Uniwersytetu Stony Brook (USA).

Rozprawa doktorska poświęcona jest badaniu struktur białkowych, modelowaniu ich dynamiki konformacyjnej oraz oddziaływań międzycząsteczkowych. Praca koncentruje się silnie na rozwoju nowatorskich metod obliczeniowych opartych na metodologii CABS i ich wszechstronnej walidacji w kontekście kluczowych problemów biologii strukturalnej i projektowania leków. Rozwijane przez mgr Mateusza Zalewskiego narzędzia (CABS-dock, Aggrescan4D, CABS-flex 3.0) stanowią wszechstronny zestaw metod służących do badania oddziaływań struktur białkowych (białko-białko i białko-peptyd), agregacji białek oraz przewidywania ich dynamiki konformacyjnej. Narzędzia te, oparte na wydajnej metodologii gruboziarnistej CABS, oferują szerokie możliwości modelowania strukturalnego, które są stosowane przez naukowców na całym świecie. Tematyka ta – oddziaływania międzycząsteczkowe oraz identyfikacja konformacji białkowych – jest aktualna i bardzo dynamicznie rozwijana i wykorzystywana m.in. w biofizyce oraz w biologii i bioinformatyce strukturalnej, zwłaszcza po przełomie jaki miał miejsce po wprowadzeniu metod opartych o sztuczną inteligencję (szczególnie po 2021 roku). Badania tego typu mają zastosowanie m.in. przy projektowaniu leków, badaniu mechanizmów chorobotwórczych czy przy tworzeniu wydajnych enzymów biotechnologicznych, gdzie znajomość konformacji i dynamiki białka ma kluczowe znaczenie.

K. Mikulska-Rumińska



2. Zawartość pracy

Przedłożona praca doktorska mgr Mateusza Zalewskiego jest napisana w języku polskim i obejmuje 9 publikacji naukowych w języku angielskim. Ma ona charakter teoretyczno-obliczeniowy, a jej istotnym walorem jest fakt, że cykl publikacyjny zawarty w pracy był realizowany przy współpracy z grupami eksperymentalnymi lub też walidowany w oparciu o niezależne doniesienia empiryczne, co znacząco podnosi wiarygodność uzyskanych wyników i opracowanych narzędzi.

Praca obejmuje 196 stron i składa się z trzech głównych części. Jej układ jest właściwy i odpowiada przyjętym standardom prac doktorskich w formie zbioru artykułów. Ma ona dobrze przemyślaną strukturę, odpowiedni stopień szczegółowości, napisana jest w sposób przystępny dla czytelnika. Zawarte w pracy rysunki są dobrej jakości. Całość prezentuje się też dobrze od strony typograficznej. Czytając rozprawę znalazłam stosunkowo niewiele błędów edytorskich.

Część pierwsza pracy zawiera *Podstawy teoretyczne i metodyczne* pozwalające na lepsze wdrożenie się w tematykę pracy. W tej części Autor opisał cel i zakres pracy oraz podstawy teoretyczne używanych (tj. dynamika molekularna) i rozwijanych przez niego metod obliczeniowych (bazujących na modelu CABS), które zostały wykorzystane do uzyskania wyników zawartych w kolejnej części pracy. Pierwsza część pracy jest dobrze opisana, jest rzeczowa, i zgrabnie przybliży aktualny stan wiedzy w zgłębianej tematyce, jej rozwoju i ograniczeniach. Zawiera ona nieliczne błędy interpunkcyjne (np. str. 19, rozdział 2.2.1, drugi paragraf) oraz pewne żargonowe sformułowania (np. oddziaływania π -stackingowe (str. 13), trajektoria asocjacji (str. 13)), które można by było ująć w nieco inny sposób. W niektórych miejscach – jak np. Rozdział 1.1., ostatni akapit – Autor mógłby uwzględnić odnośniki literaturowe odnoszące się do wymienianych przez niego przykładów. Tak samo w przypadku opisu metod – tam też w powodzeniu można by dodać więcej odnośników literaturowych. To jednak nie zmienia mojego pozytywnego odbioru tej części pracy. Referencje do tej części znajdują się na końcu rozprawy i jest ich łącznie 145. Są one aktualne, odpowiednio przypisane i zdecydowana większość to odnośniki do artykułów naukowych z ostatniej dekady. Widać, że Autor ma dobre rozeznanie w aktualnym stanie wiedzy dotyczącej jego tematyki pracy.

Część druga pracy zawiera 9 publikacji Autora (nazwanych A1-A9), które są podstawą rozprawy doktorskiej. Każda z publikacji jest poprzedzona wstępem do publikacji, gdzie Autor opisuje co zostało zrobione w ramach publikacji, jakie to miało znaczenie dla dziedziny oraz jaki był jego wkład w jej powstanie. Publikacje są opublikowane w dobrych i bardzo dobrych czasopiśmiech z listy filadelfijskiej – *Impact Factor* czasopiśmie plasuje się między 4-15 – a pan Mateusz Zalewski w pięciu z nich jest pierwszym lub pierwszym równorzędnym autorem. Podkreśla to jego znaczny wkład i dominującą rolę w sukcesie opublikowanych doniesień publikacyjnych. Dodatkowo warto podkreślić, że publikacje te są cytowane przez innych



badaczy. Dla przykładu publikacje [A4] i [A6] są cytowane odpowiednio 20 i ponad 30 razy (wg Google Scholar), a zostały opublikowane w zeszłym roku. Pokazuje to użyteczność opracowanych metod i narzędzi oraz chęć używania ich przez innych badaczy.

Drugą część pracy Autor podzielił dodatkowo na tematycznie spójne podrozdziały. Pierwszy z nich, *Rozwój nowych metod dokowania białek i peptydów*, obejmuje cztery pierwsze publikacje [A1]-[A4] z prezentowanego cyklu. Prace te tworzą spójny blok tematyczny dotyczący rozwoju, testowania i porównywania metod obliczeniowych do modelowania kompleksów typu białko-białko oraz białko-peptyd. W pracy [A1] przedstawiono podejście, oparte na modelu gruboziarnistym CABS i symulacjach Monte Carlo. Autor odszedł tu od klasycznego traktowania struktur jako obiektów sztywnych na rzecz uwzględnienia elastyczności szkieletu aminokwasowego jednego z partnerów białkowych. Praca ta ma charakter metodologiczny i pokazuje potencjał podejścia gruboziarnistego w przypadkach, w których klasyczne dokowanie sztywnych struktur jest niewystarczające. Artykuł [A2] rozszerza zastosowanie CABS-dock na peptydy cykliczne. Unikalnym i nowym osiągnięciem Autora jest tutaj opracowanie protokołu do globalnego dokowania peptydów cyklicznych bez uprzedniej wiedzy o miejscu wiązania, co zrealizowano dzięki systemowi więzów dystansowych zachowujących topologię pierścienia podczas próbkowania. Opublikowane wyniki pokazały, że rozwijana przez Autora metoda może efektywnie generować modele bliskie natywnym, choć jej istotnym ograniczeniem pozostaje kwestia selekcji najlepszych modeli przez funkcję oceniającą. Ten problem został podjęty w kolejnej pracy, [A3], w której wykorzystano krótkie symulacje pełno-atomowej dynamiki molekularnej do ponownej oceny modeli CABS, co pozwoliło na identyfikację struktur bliższych natywnym. Praca ta zatem zidentyfikowała jedno z kluczowych ograniczeń rozwijanej przez Autora metodologii czyli rozróżnienie modeli natywnych zgodnych z obserwacjami krystalograficznymi od modeli ocenionych korzystnie przez funkcję oceniającą. Ostatnia praca, [A4], wpisuje się w aktualny trend wykorzystania metod sztucznej inteligencji (AI) w modelowaniu strukturalnym. Autor dokonał walidacji modelu językowego AI ESMFold w dokowaniu typu białko-peptyd, porównując jego skuteczność z metodami opartymi na AlphaFold oraz z klasycznymi narzędziami dokowania, w tym CABS-dock. Wykazał, że ESMFold, choć ogólnie mniej dokładny niż najlepsze metody bazujące na AlphaFold, oferuje istotną przewagę pod względem szybkości i niskiego kosztu obliczeniowego. W pojedynczych przypadkach metoda ta pozwalała również uzyskać wyniki lepsze niż podejścia oparte na AlphaFold. Jednak główna wartość tej metody polega na możliwości szybkiego generowania modeli w analizach przesiewowych oraz jako narzędzie pomocnicze w strategiach opartych na porównaniu wyników kilku niezależnych metod, niż na samodzielnym dostarczaniu najbardziej wiarygodnych predykcji.

Drugi rozdział, *Modelowanie elastyczności struktur białkowych: CABS-flex*, zawiera publikacje [A5] i [A6]. Dotyczy on ewolucji narzędzia CABS-flex, które zostało wykorzystane



do stworzenia platformy integrującej dane eksperymentalne i sztuczną inteligencję, oferującej wydajność o 3-4 rzędy wielkości większą niż pełno-atomowe symulacje dynamiki molekularnej. Przełomową kwestią w wersji 3.0 tego serwera jest bezpośrednia integracja ze wskaźnikami ufności pLDDT z modeli AlphaFold, co pozwala na inteligentne sterowanie więzami w symulacji – regiony, które AI uznaje za niepewne, są traktowane przez algorytm CABS jako bardziej elastyczne, co znacząco urealnia uzyskiwane zespoły konformacyjne. Kolejną nowością podnoszącą jakość modeli było wdrożenie nowoczesnej metody rekonstrukcji *cg2all* opartej na grafowych sieciach neuronowych, która zastąpiła tradycyjne podejście oparte na programie Modeller. To pozwoliło na uzyskiwanie znacznie lepszej geometrii struktur białkowych po ich odtworzeniu do struktur pełno-atomowych. Autor wspominał we wcześniejszych publikacjach, że niedoskonałości rekonstrukcji były jednym z czynników ograniczających dokładność funkcji oceniającej (*scoringu*).

Trzeci rozdział, *Agregacja białek zależna od pH: Aggrescan4D*, obejmuje publikacje [A7] i [A8]. Koncentruje się on na problemie agregacji białek, wprowadzając pH środowiska jako kluczowy, „czwarty wymiar” analizy w narzędziu Aggrescan4D (A4D). System ten wyróżnia się wykorzystaniem uczenia głębokiego (algorytm *pKa-ANI*) do przewidywania specyficznych dla danej struktury przestrzennej wartości pKa, co pozwala na precyzyjne modelowanie zmian ładunku i lipofilowości aminokwasów w zależności od środowiska. Autor opracował tu również unikalny protokół automatycznego projektowania mutacji zwiększających rozpuszczalność, który sugeruje zmiany neutralne ewolucyjnie, chroniąc tym samym stabilność i funkcję biologiczną leku białkowego. Za szczególnie użyteczne w A4D uważam wprowadzenie dwóch funkcjonalności: analizy pH-zależnej skłonności białek do agregacji oraz trybu dynamicznego, pozwalającego uwzględnić fluktuacje strukturalne modelowane z wykorzystaniem CABS-flex. Obie funkcjonalności znacząco zwiększają użyteczność biologiczną narzędzia, ponieważ agregacja białek może zależeć zarówno od warunków środowiskowych, jak i od chwilowego odsłaniania się regionów skłonnych do agregacji. W obecnej implementacji funkcjonalności te są jednak rozdzielone i nie mogą być stosowane jednocześnie. Połączenie analizy pH z trybem dynamicznym jest, w mojej opinii, naturalnym i wartościowym kierunkiem dalszego rozwoju A4D. Z tego względu interesujące byłoby zatem wyjaśnienie, czy ograniczenie to wynika z przyczyn metodologicznych, braku odpowiedniej walidacji, kosztu obliczeniowego czy raczej z obecnych ograniczeń implementacyjnych narzędzia.

Czwarty i ostatni rozdział tej części pracy to *Dynamika molekularna (MD) w biologii strukturalnej*. Obejmuje on jedną publikację [A9], w której Autor wykonał symulacje dynamiki molekularnej dla białka P-stalk, aby wyjaśnić, jak fosforylacja steruje funkcją rybosomu. Dzięki symulacjom pełno-atomowym wykonanym przez pana Mateusza Zalewskiego udało się pokazać unikalny motyw strukturalny stabilizowany przez mostki solne między fosfoserynami a resztami lizyny („U-bent”), który nadaje białku swobodę konformacyjną niezbędną do



interakcji z czynnikami pomocniczymi translacji. W moim odczuciu rozdział ten nieco odbiega od pozostałych, głównie dlatego, że nie bazuje on na rozwijanej przez Autora metodologii CABS, a na powszechnie stosowanych symulacjach pełno-atomowych dynamiki molekularnej. Nie mniej jednak tematycznie jest on powiązany z tematem pracy doktorskiej. Warto w tym momencie podkreślić, że nawet bez publikacji [A9] doktorant mgr Mateusza Zalewskiego prezentowałby się bardzo dobrze.

Ostatnia, trzecia część rozprawy to pięciostronicowe podsumowanie i wnioski dotyczące badań zawartych w pracy doktorskiej. Część ta dobrze domyka rozprawę i stanowi czytelne zestawienie najważniejszych wniosków płynących z przeprowadzonych badań. Została ona przygotowana starannie porządkując najważniejsze rezultaty pracy Autora. Autor jasno wskazuje znaczenie rozwijanych metod, ich zastosowania oraz ograniczenia, co ułatwia ocenę spójności całego osiągnięcia. Praca zawiera również rozwinięcie wykorzystywanych skrótów, które znajduje się na początku rozprawy.

3. Wartość naukowa prezentowanych rezultatów

Przedstawiona praca doktorska ma istotną wartość naukową, szczególnie w obszarze bioinformatyki strukturalnej i modelowania molekularnego. Głównym osiągnięciem Autora jest jego znaczny wkład w rozwój i testowanie metod obliczeniowych służących do analizy elastyczności białek, dokowania białko-białko oraz białko-peptyd, a także przewidywania właściwości agregacyjnych struktur białkowych. Prace te wpisują się w aktualne potrzeby biologii strukturalnej, w której coraz większe znaczenie ma nie tylko przewidywanie statycznych struktur biomolekuł, lecz także opis ich dynamiki i zmienności konformacyjnej.

Na szczególne podkreślenie zasługuje fakt, że Autor dostrzega zarówno atuty, jak i ograniczenia rozwijanych metod, co świadczy o dużej dojrzałości naukowej. Z jednej strony zauważa rosnące znaczenie narzędzi opartych na sztucznej inteligencji, które często pozwalają uzyskać bardzo dokładne modele kompleksów białkowych. Z drugiej strony jasno wskazuje miejsce metod CABS jako podejść umożliwiających szybkie i tanie próbkowanie przestrzeni konformacyjnej, stanowiących alternatywę lub uzupełnienie zbyt kosztownych symulacji pełno-atomowych. Jest to szczególnie ważne w analizie układów charakteryzujących się dużą elastycznością, które mogą przyjmować różne stany konformacyjne lub przejściowe, trudne do uchwycenia na podstawie pojedynczej predykcji strukturalnej uzyskiwanej metodami AI.

Istotną wartością cyklu jest również krytyczne podejście do ograniczeń rozwijanych metod i konsekwentne dążenie do ich ulepszenia m.in. poprzez włączanie nowoczesnych rozwiązań. Jest to widoczne m.in. w integracji symulacji CABS z metrykami ufności modeli AI oraz w wykorzystaniu metod głębokiego uczenia do rekonstrukcji struktur atomowych. Ponadto Autor nie ogranicza się tylko do prezentowania nowych narzędzi, ale również porównuje metody klasyczne (grubozłaziste, pełno-atomowe) z narzędziami opartymi na sztucznej



inteligencji. Dzięki temu przedstawione prace mają nie tylko charakter aplikacyjny, ale również metodologiczny, wskazując kierunki dalszego rozwoju narzędzi do modelowania biomolekuł.

Całość cyklu pokazuje konsekwentny rozwój tematyki badawczej, obejmującej zarówno tworzenie i walidację metod, jak i ich zastosowanie do istotnych problemów biologii molekularnej. W mojej ocenie prezentowane rezultaty stanowią bardzo wartościowy wkład w rozwój współczesnych metod modelowania strukturalnego białek.

4. Uwagi do pracy

Praca doktorska mgr Mateusza Zalewskiego jest starannie zredagowana. Od strony merytorycznej jest bardzo dobra i trudno w niej znaleźć słabe punkty. Niemniej jednak w czasie czytania lektury pojawiło się kilka kwestii, które wzbudziły moje zainteresowanie, do których chciałabym, aby Doktorant się odniósł. Oto one:

- Narzędzie i samo podejście CABS jest niezwykle użyteczne m.in. dla osób chcących generować różne konformacje białka do dokowania ligandów/partnerów białkowych, jako struktury startowe do symulacji MD, czy po prostu do badania ich dynamiki. Biorąc pod uwagę jak wiele białek zawiera jednak nie tylko naturalne/standardowe aminokwasy oraz niestandardowe elementy strukturalne (np. modyfikacje postranslacyjne (PTMs), kofaktory), to bez wątpienia porządnym kierunkiem było by uwzględnienie w rozwijanych narzędziach chociaż części z nich. Zwłaszcza, że dla wielu układów białkowych nie są one „dodatkiem”, tylko elementem determinującym geometrię miejsca aktywnego, elementem stabilizującym bądź inicjującym zdolność do wiązania partnera. Spektrum użytkowników i zastosowań mogłyby się wtedy znacząco zwiększyć dając możliwość do eksplorowania coraz to nowszych kierunków badawczych. Stąd też moje pytania: (i) Czy w rozwijanych metodach/narzędziach planowane jest uwzględnienie chociażby najczęściej występujących PTMs? (ii) Czy rozwijane przez Autora narzędzia w wersji stand-alone dają możliwość załączenia samodzielnie sparametryzowanych niestandardowych elementów strukturalnych symulowanego układu? Jeśli nie to czy w przyszłości jest to planowane? (iii) Czy jest jakiś sposób by uwzględnić w symulacjach jony? Są białka, które bez jonów nie mogą przybrać odpowiedniej konformacji do oddziaływania z partnerem np. kalmodulina.
- Ciekawi mnie, dlaczego narzędzia typu CABS-dock i CABS-flex nie uwzględniają protonacji aminokwasów. Czy na etapie rozwoju metodologii CABS rozważano traktowanie takich stanów jako odrębnych typów reszt albo jako modyfikacji potencjałów kontaktowych lub elektrostatycznych? Czy metodologia CABS była testowana pod kątem wrażliwości na różne stany protonacyjne aminokwasów?
- Autorzy w swoich publikacjach dot. narzędzi opartych o metodologię CABS, w tym sam Autor rozprawy, rekomendują by w przypadku fosforylowanej seryny zastąpić ją



kwasem asparaginowym. Ciekawa jestem dlaczego tego typu podejście nie zostało zastosowane w publikacji [A9].

- Czy tryb globalny i lokalny CABS-dock różnią się jedynie zakresem przeszukiwanej przestrzeni, czy również efektywną rozdzielczością próbkowania?
- Ciekawią mnie również techniczne aspekty przygotowania struktur wejściowych do metod opartych na metodologii CABS. W szczególności, w jaki sposób narzędzia te trakturą struktury odbiegające od „czystych” plików PDB np. zawierające alternatywne konformacje grup bocznych aminokwasów (np. ALYS, BLYS) czy niekompletne łańcuchy boczne.
- W odniesieniu do publikacji [A7]: W jaki sposób w metodzie A4D definiowane są regiony skłonne do agregacji oraz czy algorytm pozwala odróżnić je od funkcjonalnych powierzchni wiążących partnerów białkowych lub peptydowych?
- Na czym polegała strategia losowego maskowania zastosowana w pracy [A4] i jak dobrano poziom maskowania 0.25?

Kilka drobnych spostrzeżeń (poza wcześniej wspomnianymi):

- W języku polskim zwykle nowy akapit tekstu powinien zaczynać się od wcięcia akapitowego.
- W pracy [A1] autorzy pominęli jednostki przy wykresach i tabelach (m.in. Å przy RMSF i RMSD).
- Przy wprowadzeniach do poszczególnych artykułów dobrze by było by Autor wyjaśnił znaczenie poszczególnych parametrów (np. LRMSD, Top 10k, Fnat) dla osób, które mogą nie mieć doświadczenia z tego typu badaniami.

5. Podsumowanie

Podsumowując stwierdzam, że rozprawa doktorska mgr Mateusza Zalewskiego pt. *Metody wielkoskalowego modelowania oddziaływań białko-białko i białko-peptyd* spełnia wymogi prawne stawiane pracom doktorskim zgodnie z Art. 187 ustawy z dnia 20 lipca 2018 r. (z późn. zm.) *Prawo o szkolnictwie wyższym i nauce* i wnioskuję do Rady Dyscypliny Nauki Fizyczne o dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego. Wskazane w recenzji uwagi, poczynione z obowiązku recenzenta, w najmniejszym stopniu nie umniejszają wartości rozprawy, którą oceniam bardzo pozytywnie. Mogę śmiało stwierdzić, że założone cele pracy zostały zrealizowane, a stwierdzenia i wnioski sformułowane prawidłowo i w pełni wynikają z uzyskanych wyników badań. Praca stanowi oryginalne rozwiązanie problemu naukowego, dowodzi gruntownej wiedzy Autora w zakresie Dyscypliny Nauki Fizyczne, a także umiejętności samodzielnego prowadzenia pracy naukowej. Uzyskane wyniki badań wnoszą istotny wkład w rozwój nowoczesnej biofizyki i bioinformatyki strukturalnej, dostarczając innowacyjnych narzędzi do modelowania dynamiki oraz oddziaływań białek i peptydów (w tym również peptydów cyklicznych) przy uwzględnieniu elastyczności szkieletu białkowego.



Opracowane metodologie stanowią cenne wsparcie w obszarze inżynierii białek i projektowania bioterapeutów, co potwierdzają udane walidacje przy użyciu danych eksperymentalnych. Z tego względu wnoszę o wyróżnienie rozprawy doktorskiej mgr Mateusza Zalewskiego.

K. Mikulska-Rumińska

Dr hab. Karolina Mikulska-Rumińska, prof. UMK