



UNIwersytet
WARszawski

Wydział Chemii

dr hab. Dominik Gront
Uniwersytet Warszawski
Wydział Chemii



Warszawa, dnia 28 sierpnia 2019r.

Recenzja rozprawy doktorskiej p.t.

*„Przewidywanie struktury białek oraz ich kompleksów
przy użyciu gruboziarnistego pola siłowego UNRES z wykorzystaniem informacji z baz danych”*

opracowanej przez panią magister Agnieszkę Sylwię Karczyńską

Pytanie o strukturę przestrzenną białek - makrocząsteczek będących podstawą wszystkich procesów życiowych - pojawiło się po raz pierwszy w pierwszej połowie ubiegłego stulecia. Na przestrzeni ostatnich kilku dekad odpowiedzi na nie poszukiwano poprzez rozważania teoretyczne (m.in. Linus Pauling), badania eksperymentalne (m.in. John Kendew, Kurt Wüthrich) a także stosując symulacje komputerowe (np. Martin Karplus, Michael Levitt i Arieh Warshell). Dzięki tym badaniom znamy obecnie ponad 150 000 struktur białek i kompleksów białkowych, wyznaczonych różnymi metodami doświadczalnymi. Posiadamy również całkiem dobrą wiedzę podstawową o zasadach, wg których łańcuch złożony z kilkuset reszt aminokwasowych formuje unikalną dla jego sekwencji strukturę trójwymiarową. W oparciu o tę wiedzę jesteśmy w stanie przewidywać struktury małych białek globularnych a także projektować zupełnie nowe, sztuczne białka, które nie istnieją w organizmach żywych. Metody teoretyczne i obliczeniowe są niezmiernie ważnym narzędziem do badania struktur białek. Liczba potencjalnych obiektów tych badań, czyli sekwencji białkowych, przerasta możliwości metod eksperymentalnych o co najmniej trzy rzędy wielkości.

Współczesne metody obliczeniowe stosowane do badania biomakromolekuł to szeroki wachlarz różnych algorytmów, których wykorzystanie zależy dostępnej wiedzy *a priori*, a także możliwych do zagospodarowania zasobów komputerowych. Pierwsze pytanie, które pojawia się przy wyborze odpowiedniego algorytmu, dotyczy dostępności już wyznaczonych struktur białek homologicznych. Modelowanie homologiczne (zwane też *w oparciu o szablon*) umożliwia bowiem uzyskanie dobrych a nawet bardzo dobrych wyników w stosunkowo krótkim czasie. Alternatywny scenariusz, zwany *de novo*, wykorzystywany jest, kiedy struktury takie nie mogą być znalezione. Scenariusz ten sprowadza się do czasochłonnego przeszukiwania przestrzeni konformacyjnej stanów dostępnej dla zadanego polipeptydu i wymaga zaangażowania bardzo poważnych zasobów obliczeniowych.

W praktyce metody te stosują się tylko dla małych białek. Jednym ze sposobów na poszerzenie zakresu ich stosowalności jest zastosowanie reprezentacji gruboziarnistej, w której grupa kilku a nawet kilkunastu atomów zastępowana jest pojedynczym centrum oddziaływań. Przykładem takiej metody jest algorytm UNRES, opracowywany od ponad 25 lat w laboratorium prof. Adama Liwo (Uniwersytet Gdański) a także w grupie prof. Scheragi (Cornell University).

W przedstawionej mi do recenzji rozprawie doktorskiej, pani magister Agnieszka Karczyńska wykorzystwała wspomniany powyżej model UNRES, zarówno w modelowaniu *de novo* jak i w połączeniu z informacją o strukturach białek homologicznych. Celem jej prac było udoskonalenie tej metody obliczeniowej poprzez wprowadzenie wydajniejszego schematu próbkowania (określonego w pracy jako HREMD). Doktorantka wykonała również obszernie i systematyczne obliczenia testowe mające na celu sprawdzenie ergodyczności symulacji w których stosowano więzy strukturalne różnej jakości. Autorka opisała w pracy także wyniki jedenastej oraz w dwunastej edycji międzynarodowego eksperymentu CASP (*Critical Assessment of Protein Structure Prediction Methods*), w których wzięła aktywny udział. W trakcie tych eksperymentów grupy badawcze z całego świata wykorzystują metody obliczeniowe aby wymodelować przestrzenne konformacje białek, dla których struktury wyznaczone doświadczalnie zostaną ogłoszone dopiero po zakończeniu eksperymentu. Udział w CASP11 i CASP12 stanowił zatem obiektywny test metodologii UNRES. Metodologia ta została udostępniona szerokiemu środowisku naukowemu jako zautomatyzowany serwis internetowy, w tworzeniu której Doktorantka miała znaczący wkład.

Rozprawę otwierają podziękowania i wykazy skrótów oraz wykorzystywanego oprogramowania. Następnie Autorka charakteryzuje cele swojej pracy. Kolejny, niemalże sześćdziesięciostronicowy rozdział to wstęp teoretyczny, mający za zadanie przybliżyć czytelnikowi podstawy dziedziny swoich badań, w którym zawarto aż 194 odsyłacze do pozycji literaturowych. Główną część rozprawy stanowi 5 wieloautorskich prac oryginalnych, publikowanych w czasopiśmie z listy filadelfijskiej o wysokim współczynniku *Impact Factor*, które zawierają wyniki badań prowadzonych przez doktorantkę. Prace te zostały już zrecenzowane przez niezależnych ekspertów. Pozwolę sobie jednak poniżej podsumować ich główne wnioski, które jednocześnie stanowią meritum osiągnięć Autorki.

Gruboziarnisty model UNRES, umożliwiający modelowanie struktury, dynamiki i termodynamiki zarówno białek jak i kwasów nukleinowych, rozwijany jest nieprzerwanie od ponad 25 lat. Badania pani magister Karczyńskiej stanowią tutaj istotny wkład. W szczególności jej dwie pierwsze prace prezentują dogłębną ocenę możliwości modelu. Publikacja 1 opisuje wyniki jakie osiągnięto w XI konkursie CASP, podczas którego grupa

Cornell-Gdańsk stosowała podejście oparte na fizyce (modelowanie *de novo*). Pomimo, że grupa nie korzystała z informacji ewolucyjnej, zanotowała kilka bardzo udanych predykcji. Wykorzystanie nowych elementów pola siłowego pozwoliło poprawić wyniki w porównaniu do poprzedniej, dziesiątej edycji konkursu. Publikacja 2 z kolei opisuje wyniki modelowania porównawczego, uzyskane w trakcie dwunastej edycji eksperymentu CASP przez grupę KIAS-Gdańsk. Wykorzystano w tym celu opracowane uprzednio podejście hybrydowe, w którym model UNRES połączono z więzami odległości. Więzy te, zaczerpnięte z modeli generowanych przez cztery najlepsze zautomatyzowane serwery, kierowały symulację w stronę najbardziej prawdopodobnej topologii. Jednocześnie oparte na fizyce pole siłowe UNRES pozwalało uniknąć konformacji niemożliwych do zrealizowania w przyrodzie oraz pomagało wygładzić lokalną geometrię generowanych modeli. Uzyskane w ten sposób struktury nierzadko były istotnie lepsze od najlepszego modelu wykorzystanego w procesie modelowania.

Metodyka uwzględniania więzów strukturalnych w procesie modelowania została poprawiona i gruntownie przetestowana w Publikacji 3. Zaproponowano w niej zmianę funkcji oceniającej zgodność modelu z więzami: logarytm z funkcji Gaussa zastąpiono funkcją Lorentza, której wartości są ograniczone z góry i dość szybko zbliżają się do wartości asymptotycznej. Modyfikacja ta spowodowała obniżenie barier energetycznych a tym samym istotnie poprawiła wydajność schematu próbkowania. W pracy dodatkowo zaproponowano wymianę replik w przestrzeni Hamiltonianu. Wykazano, że nowy schemat wymiany replik (*Hamiltonian Replica Exchange Molecular Dynamics*, HREMD) próbkuje wydajniej, niż schemat MREMD, z którego korzystano w pracach (1) i (2).

W Publikacji 4 zaprezentowano metodę, która pozwala na uwzględnienie pomiarów niskokątowego rozpraszania promieni X (SAXS) w symulacjach modelem UNRES. Dane te włączano w postaci rozkładów odległości międzyatomowych $P(r)$, porównywanych z rozkładem teoretycznym, obliczanym na podstawie modeli gruboziarnistych w reprezentacji UNRES. Przyjęto tu przybliżenie, w którym zjednoczone atomy traktowano jako chmury prawdopodobieństwa a odległości międzyatomowe - zmienne o rozkładzie normalnym. Uwzględnienie danych eksperymentalnych spowodowało istotną poprawę modeli uzyskanych dla dwóch z jedenastu celów. W pracy tej zabrakło mi szczegółów opisujących w jaki sposób modelowana jest czwartorzędowa struktura białek. Czy układem modelowanym metodą UNRES jest pojedynczy łańcuch, czy cały oligomer? Jak ustalana jest wzajemna orientacja cząsteczek białka w kompleksie? Dodatkowo pozwolę sobie wdać się w polemikę o roli rozpuszczalnika w widmie SAXS. Rola ta jest marginalizowana przez autorów w Publikacji 4, którzy stwierdzają, iż SAXS rejestruje jedynie różnicę w gęstości elektronowej pomiędzy otoczką solwatacyjną a rozpuszczalnikiem. Stwierdzenie to jest oczywiście prawdziwe, niemniej jednak przy obliczaniu wkładu cząsteczek białka

do krzywej rozpraszania również należy uwzględnić wyparty rozpuszczalnik. Jedną ze stosowanych metod jest wprowadzenie skorygowanych czynników strukturalnych.

Swoistym podsumowaniem tych prac jest Publikacja 5, w której opisano konstrukcję zautomatyzowanego serwera obliczeniowego. Serwer ten udostępnia oprogramowanie UNRES szerokiemu środowisku naukowemu, bez konieczności instalacji i uaktualizacji u użytkowników końcowych. Z założenia serwer udostępniać będzie bowiem zawsze najnowszą wersję oprogramowania i pola siłowego. Serwer UNRES jest zatem kamieniem milowym w rozwoju tego oprogramowania.

Prace te w sumie stanowią ogromną wartość naukową i w mojej opinii bez wątpienia kwalifikują Kandydatkę do dalszych etapów przewodu doktorskiego. Z obowiązku recenzenta pozwolę sobie jednak przedstawić kilka uwag krytycznych, które nasunęły mi się w trakcie lektury rozprawy oraz załączonych publikacji.

- Informacja o udziale Doktorantki w uzyskanych wynikach pojawia się dopiero na samym końcu rozprawy; załączone oświadczenia pani magister Karczyńskiej oraz pozostałych współautorów stanowią ostatni rozdział. W moim przekonaniu informacja powinna się pojawić znacznie wcześniej, na przykład zostać uwzględniona w polskojęzycznym podsumowaniu każdej z prac (podrozdziały 6.1 - 6.5)
- Opis do Rysunku 3 niefortunnie przemieścił się do środka akapitu.
- W Publikacji 3 wspomniano ilustracje 7g oraz 7h, których niestety nie udało mi się odnaleźć zarówno w głównym tekście ani w załączonym do publikacji suplemencie.
- Pojęcie ergodyczności jest kluczowe dla zrozumienia i poprawnej interpretacji wyników przedstawionych w Publikacji 3, nie jest ono niestety wyraźnie zdefiniowane. W mojej ocenie Autorka powinna poświęcić mu co najmniej kilka akapitów w „*Części Teoretycznej*” pracy.
- Tytuł rozdziału 6 „*Część Wynikowa*” jest w moim odczuciu sformułowany niezręcznie.
- Autorka nie stosuje zwyczajowo przyjętej metody cytowania wielu prac przez podanie zakresu, np [24-37], lecz wymienia wszystkie kolejne numery, które nierzadko zajmują prawie cały wiersz (*vide* str. 22)

Uwagi powyższe mają głównie charakter edytorski i w żadnym stopniu nie umniejszają wartości merytorycznej rozprawy. Kandydatka wykonała ogromną pracę, ulepszając i testując metodę badawczą UNRES. Jest współautorką co najmniej ośmiu publikacji w renomowanych czasopismach, z których pięć stanowi treść rozprawy. Co więcej, Kandydatka jest pierwszym autorem aż dwóch publikacji. Dorobek naukowy pani magister

Karczyńskiej jest zatem imponujący. Rozprawa spełnia zatem z nawiązką wymagania stawiane rozprawom doktorskim przez **Ustawę o z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki** (Dz. U. nr 65 poz 595 z późn. zm.) jak również zwyczajowe standardy stawiane rozprawom doktorskim w dziedzinie nauk przyrodniczych i ścisłych. Dlatego z pełnym przekonaniem wnoszę do Rady Wydziału Chemii Uniwersytetu Gdańskiego o dopuszczenie pani magister Agnieszki Sylwii Karczyńskiej do dalszych etapów przewodu doktorskiego. Biorąc pod uwagę wysoką jakość wyników oraz ogromny wkład Doktorantki w ich uzyskanie wnoszę o wyróżnienie rozprawy. Szczegółowe uzasadnienie wniosku o wyróżnienie stanowi załącznik do niniejszej recenzji.



Dominik Gront



UNIwersytet
Warszawski

Wydział Chemii

dr hab. Dominik Gront
Uniwersytet Warszawski
Wydział Chemii



Warszawa, dnia 28 sierpnia 2019r.

Uzasadnienie wniosku o wyróżnienie rozprawy doktorskiej
pani magister Agnieszki Sylwii Karczyńskiej pt.

*„Przewidywanie struktury białek oraz ich kompleksów
przy użyciu gruboziarnistego pola siłowego UNRES z wykorzystaniem informacji z baz danych”*

W ramach realizacji swojej rozprawy doktorskiej pani magister Karczyńska wykonała ogromną pracę projektując, optymalizując i testując nowe komponenty modelu UNRES. Dodatkowo wzięła udział w aż trzech kolejnych edycjach eksperymentu CASP. Praca ta na pewno zaowocowała ogromnym doświadczeniem oraz warsztatem badawczym. Wyniki tych badań opisano w pięciu publikacjach stanowiących integralną część rozprawy. Pani magister Karczyńska jest jednak współautorką co najmniej ośmiu artykułów w czasopiśmie z listy filadelfijskiej, a w dwóch z nich jest pierwszym autorem. Kandydatka miała ogromny wkład w opracowanie serwera obliczeniowego UNRES, który ma duże znaczenie praktyczne; publikacja ta będzie zapewne szeroko cytowana.

Uwzględniając powyższe argumenty, stawiam formalny wniosek o wyróżnienie niniejszej rozprawy zgodnie z panującymi na Wydziale Chemii UG zasadami.

Dominik Gront