

Gliwice, 14.01.2023

Prof. dr hab. inż. Katarzyna Stapor  
Politechnika Śląska  
Wydział Automatyki, Elektroniki i Informatyki  
Katedra Informatyki Stosowanej

## **R E C E N Z J A**

**pracy doktorskiej mgr Magdaleny Ptak-Kaczor**

*p.t. „Charakter rozkładu oddziaływań niewiązanych (elektrostatycznych, vdW)  
wobec rozkładu oddziaływań hydrofobowych w strukturach białek”*

wykonanej na Wydziale Fizyki Astronomii i Informatyki Stosowanej  
Uniwersytetu Jagiellońskiego  
pod kierunkiem prof. dr hab. Ireny Roterman-Koniecznej

### **1. Obszar problemowy i cel rozprawy**

Tematyka recenzowanej rozprawy doktorskiej dotyczy zagadnień związanych z przewidywaniem struktury przestrzennej białek na podstawie sekwencji. Rozprawa posiada charakter interdyscyplinarny, z pogranicza chemii obliczeniowej, biofizyki oraz informatyki.

Zrozumienie funkcji białek na poziomie molekularnym wymaga poznania ich przestrzennej struktury. Struktura białka jest natomiast zdeterminowana przez jego sekwencję aminokwasową oraz środowisko (otoczenie). Poznanie struktury natywnej białka ma kluczowe znaczenie dla badań nad nowymi metodami terapii różnych chorób. Zrozumienie, jaki jest mechanizm zwijania białek, może doprowadzić do znalezienia przyczyny nieprawidłowego formowania ich struktury jak to ma miejsce w przypadku chorób neurodegeneracyjnych (Alzheimera czy Creutzfeldta-Jakoba). Mechanizm procesu fałdowania białek mimo w długoletnich próbach nie został niestety w pełni rozpoznany.

Metody eksperymentalne (techniki NMR, krystalografia rentgenowska) wyznaczania struktury białek pozwoliły na jej wyznaczenie dla zaledwie małej części sekwencji białkowych dostępnych obecnie w bazie UniProt. Ta olbrzymia i wciąż powiększająca się dysproporcja wskazuje na konieczność rozwijania innych metod służących rozwiązywaniu struktur białek.

Znaczny rozwój metod bioinformatycznych oraz wzrost mocy obliczeniowej komputerów sprawiły, że stało się możliwe również teoretyczne przewidywanie struktur. W niektórych przypadkach dokładność otrzymanego modelu jest porównywalna do struktur białek otrzymanych przy zastosowaniu metod eksperymentalnych. Od 1994 roku, co dwa lata, organizowany jest ogóln światowy konkurs w przewidywaniu struktury trzeciorzędowej białek, CASP (*ang. Critical Assessment of Techniques for Protein Structure Prediction*). Celem konkursu jest ocena skuteczności oraz weryfikacja dostępnych metod i algorytmów.

Dostępne obecnie metody teoretyczne, pozwalające na przewidywanie przestrzennej struktury białek, można podzielić na metody *de novo* (przewidyujące strukturę białka jedynie na podstawie sekwencji aminokwasów) oraz modelowanie porównawcze umożliwiające budowę modelu białka na podstawie białka homologicznego o znanej strukturze przestrzennej. Przewidywanie metodami *de novo* oparte jest o hipotezę Anfinsena stwierdzającą, iż struktura przestrzenna białka jest ściśle zdeterminowana przez jego sekwencję aminokwasów i struktura natywna odpowiada minimum energii swobodnej układu białko-rozpuszczalnik. Do podstawowych problemów tych metod, jak dotąd nie rozwiązanych w pełni od ponad pół wieku należą: stworzenie uproszczonej reprezentacji białka o właściwościach zbliżonych do reprezentacji pełnoatomowej, znalezienie odpowiedniej funkcji dokładnie opisującej minimum energetyczne układu oraz zaprojektowanie algorytmu pozwalającego na jak najlepsze próbkowanie przestrzeni konformacyjnej modelowanego białka. Najważniejszą grupą metod są tu symulacje metodami dynamiki molekularnej w oparciu o modele fizyczne z użyciem różnych empirycznych pól siłowych. Należy też wspomnieć o dużej grupie metod wykorzystujących uczenie maszynowe (*ang. machine learning*) z „rewolucyjnym” systemem AlphaFold2, który mimo najlepszych wyników nie wyjaśnia mechanizmu fałdowania się białek i posiada znaczne ograniczenia, jeśli chodzi o predykcje struktury białek o wielu stanach natywnych.

Ze względu na skalę czasową i rozmiar układów biomakrocząsteczkowych konieczne staje się projektowanie efektywnych modeli.

Recenzowana rozprawa może być pewną próbą wypełnienia istniejącej luki w tym zakresie.

Konkretnie, **celem pracy doktorskiej** było zweryfikowanie, czy **oddziaływania niewiążące w białkach** (tj. oddziaływania elektrostatyczne i van der Waalsa) **stosują się do modelu centrycznej koncentracji**, jaki ma miejsce w przypadku oddziaływań hydrofobowych.

Doktorantka sformułowała też (w sposób niejawni) tezę rozprawy:

*Wewnętrzne algorytmy optymalizacji energii (w symulacji komputerowej zwiżania białek) w warunkach wstępnych powinny uwzględniać środowisko wodne jedynie dla oddziaływań hydrofobowych.*

Tematykę rozprawy uważam za oryginalną, ważną oraz aktualną dla współczesnych prac z zakresu teoretycznych metod przewidywania struktury białek.

Podjęta przez Doktorantkę tematyka, gdyby została zweryfikowana praktycznie w symulacjach, mogłaby mieć bardzo duże znaczenie dla praktyki.

## **2. Zawartość rozprawy**

Rozprawa składa się z pięciu rozdziałów omawiających uzyskane wyniki, bibliografii (44 strony) oraz 5-ciu załączników – publikacji Doktorantki stanowiących podstawę złożonej rozprawy doktorskiej, jednakże tylko do pierwszej publikacji są odniesienia w rozprawie. Bibliografia rozprawy zawiera 86 pozycji literaturowych, w tym tylko 5 z ostatnich 5 lat. W tekście manuskryptu zamieszczono 11 rysunków (w tym 7 z publikacji 1) oraz dwie tabele. Całość rozprawy została napisana w języku polskim.

W 2-stronicowym Wprowadzeniu został przedstawiony cel pracy, na końcu zaś (w sposób niejawni) teza pracy. Rozdział drugi „Oddziaływania” zawiera opis oddziaływań molekularnych w procesie fałdowania białka: oddziaływania niewiążące (podrozdział 2.1), oddziaływania hydrofobowe (podrozdział 2.2) oraz opis modelu rozmytej kropli oliwy (ang. **Fuzzy Oil Drop**, FOD) użyty w pracy do oceny charakteru rozkładów oddziaływań niewiążących (podrozdział 2.3). Rozdział trzeci “Materiały i metody” opisuje metodykę i protokoły zastosowane w pracy dla wykonania obliczeń w systemie Gromacs posiadającym zaimplementowane różne pola siłowe. Dla obliczenia energii oddziaływań elektrostatycznych i van der Waalsa w białku doktorantka użyła metody MM/PBSA (ang. *Molecular Mechanics/Poisson–Boltzmann surface area*) zaimplementowanej w pakiecie `g_mmpbsa` w Gromacs. W rozdziale czwartym omówione zostały uzyskane wyniki, Podrozdziały tej sekcji poświęcone są konkretnym przykładom białek, dla których przeprowadzono analizę. Podrozdział 4.1 poświęcony jest titinie, która jest reprezentatywnym przykładem idealnej

miceli, podczas gdy białka amyloidogenne (podrozdział 4.2) przeciwnie, wykazują wzór niezgadający się z idealną micelą. Transtyretyna przedstawiona w podrozdziale 4.3 jest przypadkiem pośrednim. Rozdział piąty zawiera Wnioski.

Wspomniane 5 publikacji zostały opublikowane w czasopiśmie notowanych w Journal Citation Reports, zaś ich sumaryczny współczynnik oddziaływania wynosi 23.769. Jedna publikacja (omówiona w pracy) – „Internal force fields in selected proteins” opublikowana w *Acta Biochimica Polonica* (IF 2.15), dwie publikacje w *International Journal of Molecular Sciences* (IF 5.54), jedna w *Biomolecules* (IF 4.57) oraz jedna w *Symmetry* (IF 3.11). W trzech pierwszych publikacjach Doktorantka jest pierwszym autorem, co sugeruje jej dominujący wkład w ich utworzenie.

### **3. Uzyskane wyniki**

Najważniejszym osiągnięciem pracy doktorskiej jest identyfikacja typu rozkładów oddziaływań w białkach. Oddziaływania w białkach stabilizujące strukturę III-rzędową to głównie oddziaływania elektrostatyczne i typu van der Waalsa. Dodatkowo w stabilizacji biorą udział oddziaływania hydrofobowe. W procedurach poszukiwania struktury białek metodami *in silico* wykorzystuje się procedury optymalizacji w poszukiwaniu minimum energetycznego dla stabilnej struktury. Optymalizacja ta – jak wykazała praca doktorska pani mgr Magdaleny Ptak-Kaczor – powinna mieć odmienny charakter. Oddziaływania elektrostatyczne i oddziaływania typu van der Waalsa wykazują charakter rozkładu jednostajnego, podczas gdy rozkład hydrofobowości wykazuje rozkład zbliżony do rozkładu normalnego (Gaussa). Oznacza to odmienną funkcję celu dla procedury optymalizacji oddziaływań prowadzących do stabilizacji struktury III-rzędowej. Wnioski te wynikają z analizowanych przykładów wybranych białek – titiny, gdzie obserwowany rozkład hydrofobowości wykazuje wysoką zgodność z rozkładem Gaussa. Ze względu na aktualny problem zjawiska transformacji amyloidowej, analizę przeprowadzono też dla białek ulegających tej transformacji oraz dla białek wykazujących już strukturę fibryla amyloidowego. Analiza ta wykazała, że relacje rozkładów elektrostatycznych, van der Waalsa oraz hydrofobowości nie wyróżniają struktur amyloidowych, które w przeprowadzonej analizie dostarczają charakterystyki podobnej dla analizowanych białek. Dla wszystkich przeanalizowanych białek niezależnie od ich hydrofobowego wzorca centrycznej koncentracji nie znaleziono podobnych wzorców w zakresie oddziaływań elektrostatycznych czy van der Waalsa.

Należy również podkreślić, że doktorantka wykazała się znakomitym opanowaniem i posługiwaniem się systemem Gromacs oraz dostępnymi tu pakietami i polem siłowym OPLS, co świadczy o jej dużych umiejętnościach programistycznych i uzdolnieniach w dziedzinie informatyki. Wykonanie obliczeń w systemie Gromacs wymagało od doktorantki prócz dużego, czasowego nakładu – cierpliwości i kreatywności przy dostosowywaniu dostępnych bibliotek procedur dla wykonania swojego zadania.

#### **4. Uwagi o charakterze dyskusyjnym**

Podczas czytania pracy nasuwają mi się pewne uwagi.

1. Wątpliwość budzi sposób obliczania energii dla oddziaływań elektrostatycznych i van der Waalsa pojedynczego białka za pomocą metody MM/PBSA. Metoda mechaniki molekularnej/pola powierzchni Poissona Boltzmanna jest używana do obliczania energii swobodnej wiązania kompleksów białko-ligand. Zastosowanie w roli liganda/inhibitora – konglomeratu cząsteczek wody, jak to zrobiła Doktorantka wydaje się dyskusyjne. Prosiłabym o uzasadnienie.
2. W protokole obliczeń opisanym w podrozdziale 3.2 brak specyfikacji użycia (i parametrów) do pakietu `g_mmpbsa` użytego do obliczeń energii. Parametry te są kluczowe dla obliczenia składowych energii. Nie jest też jasne, które "brakujące elementy" są wymienione w kroku 1 protokołu.
3. Czy Doktorantka podejmowała próby zaimplementowania uzyskanych wyników do sterowania optymalizacją w trakcie symulacji zwijania jakiegoś białka w polu siłowym systemu Gromacs? Innymi słowy, czy postawiona w sposób niejawnym we Wprowadzeniu teza o optymalnym obliczeniowo sposobie prowadzenia symulacji została zweryfikowana praktycznie? Jeśli nie, to co stanęło na przeszkodzie?
4. Dlaczego Doktorantka odwołuje się w doktoracie tylko do publikacji pierwszej spośród 5-ciu załączonych?
5. Pierwsze dwa akapity w podrozdziale 2.1, a także pewne fragmenty przedstawione w rozdziale 2.2 dotyczące systemu Gromacs powinny należeć do rozdziału "Materiały i metody", ponieważ zawierają szczegóły techniczne, a nie odpowiadają tematowi podrozdziału. Powoduje to, że tekst jest nieuporządkowany.
6. "Mostki disulfidowe" nie są oddziaływaniami "niewiązującymi" i dlatego nieodpowiednio zostały włączone do tego podrozdziału 2.1.

7. W rozdziale 2-gim nie podano także źródła rysunków 1 i 2, które można znaleźć w Internecie.
8. W punkcie 4 opisującym uzyskane wyniki Doktorantka poświęciła moim zdaniem zbyt dużo miejsca na dane literaturowe dotyczące funkcji biologicznych badanych białek, natomiast przedstawiła stosunkowo mało szczegółów obliczeniowych analizy.
9. Jako sugestię do weryfikacji tezy o możliwości sterowania optymalizacją w symulacji zwijania białka poprzez odpowiednie „włączanie” sił hydrofobowych – podałabym zrealizowaną pracę, w której podjęto pewne wstępne próby optymalizacji naprzemiennej z wykorzystaniem użytego przez Doktorantkę modelu FOD w polu siłowym UNRES (ang. UNited RESidue):

Roterman I, Sieradzan A, Stapor K, Fabian P, Wesolowski P, Konieczny L. On the need to introduce environmental characteristics in ab initio protein structure prediction using a coarse-grained UNRES force field. *J Mol Graph Model*. 2022 Jul;114:108166. doi: 10.1016/j.jmgm.2022.108166. Epub 2022 Mar 18. PMID: 35325843.

Myszę, że wnioski z w/w pracy mogłyby być przydatne w przyszłościowej implementacji wyników uzyskanych w pracy doktorskiej.

## 5. Ocena końcowa rozprawy

Wymienione powyżej uwagi mają charakter dyskusyjny i nie podważają merytorycznej wartości pracy.

Reasumując stwierdzam, że postawiony w rozprawie **cel pracy doktorskiej** dotyczący identyfikacji typów rozkładów oddziaływań w białkach **został osiągnięty**. Praca ta stanowi wartościowy wkład w poznanie ich natury. Uzyskane wyniki, jeśli zostaną zastosowane w praktycznych polach siłowych mogą znacznie zmniejszyć złożoność obliczeniową i potrzebne zasoby dla realizacji symulacji zwijania białek metodami dynamiki molekularnej.

W konkluzji uważam, że recenzowana **praca spełnia wszystkie wymagania stawiane pracom doktorskim przez odpowiednią Ustawę i wnioskuję o jej przyjęcie i dopuszczenie mgr Magdaleny Ptak-Kaczor do dalszych etapów przewodu doktorskiego.**

