



Uniwersytet
Gdański



Związek
Uczelni
Fahrenheita

Wydział Chemii
ul. Wita Stwosza 63
80-308 Gdańsk

Gdańsk, dnia 28.08.2023

**Recenzja rozprawy doktorskiej Pana mgra Marcina Sobieraja pt.
„Nowe metody analizy mechanizmów funkcjonowania układów
biomolekularnych w oparciu o symulacje dynamiki molekularnej”**

Analiza wyników symulacji dynamiki molekularnej białek oraz innych makromolekuł biologicznych i ich układów jest wyzwaniem z uwagi na mnogość zmiennych, która często skutecznie ukrywa istotne informacje o zmianach zachodzących w układzie. W ciągu długiego rozwoju metody dynamiki molekularnej opracowano również mniej lub bardziej zaawansowane metody analizy wyników, z których najbardziej znane są te oparte na analizie głównych składowych, analizie skupień czy metodzie map dyfuzyjnych. Każda z tych metod wymaga wyboru zmiennych do analizy, które mogą być współrzednymi kartezjańskimi, kątami dwuściennymi, kontaktami międzyresztowymi lub charakterystykami globalnymi takimi, jak odchylenie od struktury natywnej, ułamek natywnych kontaktów międzyresztowych, promień żyracji, czy też eliptyczność. Taka analiza umożliwia konstrukcję krajobrazów energii swobodnej ale jest mniej użyteczna do opisu ścieżek zwijania białek czy też przemian konformacyjnych. Rozprawa doktorska Pana mgra Marcina Sobieraja, wykonana pod kierunkiem znanego specjalisty w dziedzinie biofizyki obliczeniowej, dra hab. Piotra Setnego, jest poświęcona temu ostatniemu zagadnieniu.

Rozprawa została napisana w formacie klasycznym, bez publikacji dołączonych jako jej integralna część. Tym niemniej ze spisu literatury wynika, że jej materiał w pewnej części znalazł się w publikacjach Kandydata, co jest oczywiście wielką zaletą. Preambulę rozprawy stanowią streszczenie (w języku polskim i angielskim) oraz podziękowania. Rozprawa jest podzielona na 5 rozdziałów, z którym pierwszy jest podzielony na wstęp, cel i strukturę pracy oraz znaczenie wyników badań, trzy kolejne opisują poszczególne części badań własnych kandydata a ostatni jest poświęcony perspektywom dalszego rozwoju badań. Każdy z wymienionych trzech rozdziałów jest zamkniętą całością, zawierającą przedstawienie zagadnienia, podstawy teoretyczne (teoria jest w bardzo dużej części oryginalnym dorobkiem Kandydata), metody, wyniki, podsumowanie zagadnienia oraz podsumowanie rozdziału. Po zakończeniu opisu badań własnych

następują cztery dodatki, zawierające szczegóły rozważań teoretycznych lub obliczeń. Ostatnimi częściami są spis cytowanej literatury, zawierający 172 pozycje oraz spis rysunków. Łącznie rozprawa zawiera 182 strony. Oprócz anglojęzycznej wersji streszczenia rozprawa jest napisana w języku polskim. Organizacja rozprawy jest bardzo systematyczna i przejrzysta.

Celem pracy było opracowanie, wraz z aparatem matematycznym, metod selekcjonowania istotnej informacji z trajektorii dynamiki molekularnej, włączając w to badanie zależności między procesami zachodzącymi w różnych miejscach i czasie. Takie metody docelowo umożliwiają poznanie mechanizmów zwijania białek, dużych przemian konformacyjnych w białkach (np. przejść między zamkniętymi i otwartymi stanami chaperonów molekularnych), a przez to mechanizmów funkcjonowania maszyneryi komórkowej. Opracowana metodologia została przetestowana poprzez zastosowanie jej do analizy trajektorii zwijania minibiałka chignoliny oraz modelowego układu ładunków we wnęce kulistej.

W rozdziale 2 Kandydat omawia opracowaną przez siebie metodę znajdowania optymalnej odległości obcięcia dla kontaktów międzyresztowych. Jak wspomniałem na wstępie mojej recenzji, kontakty międzyresztowe należą do zestawu miar struktury białka i dlatego określenie kiedy kontakt występuje jest ważne i cały czas otwarte. W pierwszych pracach Tanaki i Scheragi arbitralnie przyjmowano odległość między atomami węgla α mniejszą niż 7 Å jako kryterium utworzenia kontaktu, obecnie zwykle przyjmuje się, że kontakt powstaje, gdy najmniejsza odległość między atomami dwóch reszt jest mniejszą niż 5 Å, co jest odległością van der Waalsa większości atomów drugiego okresu. Kandydat podszedł do zagadnienia znalezienia optymalnej odległości obcięcia na gruncie teorii informacji a konkretnie maksymalizacji informacji, do którego to zagadnienia zastosował przybliżenie Gaussa ze średnią korelacją, aby uniknąć liczenia całek wielowymiarowych (co z góry jest skazane na niepowodzenie nie tylko z uwagi na wysiłek obliczeniowy ale również niewystarczającą ilość danych) i uzyskać rozsądną statystykę. Wyszedł z założenia, że optymalna odległość to taka, przy której występuje lokalne minimum średniej korelacji między kontaktami aktywnymi czyli takimi, które tworzą się i znikają w toku symulacji dynamiki molekularnej. Lokalne minimum korelacji oznacza, że poszczególne kontakty niosą wiele informacji o zmianach strukturalnych. W celu znalezienia korelacji Kandydat opracował oryginalny aparat matematyczny oparty na teorii informacji. Następnie wykonał i przeanalizował przy pomocy opracowanej metody symulacje kanonicznej dynamiki molekularnej pięciu struktur białek: PKA (w stanie aktywnym i nieaktywnym) LAO (dwie struktury), represora λ , trypsyny oraz domeny SH3. Znaleziona wartość odległości obcięcia wynosiła 5 Å dla wszystkich białek oprócz represora λ i 5.1 Å dla represora λ , czyli mniej więcej tyle, ile suma promieni van der Waalsa atomów drugiego okresu.

Do rozdziału drugiego Kandydat dołączył dodatki A i B. W dodatku A przedstawił bardzo interesujący fragment konwersacji między Shannonem i von Neumannem, w którym ten ostatni doradza nazwanie nowo wprowadzonej wielkości entropią głównie dlatego, że nikt dokładnie nie wie co to jest. Faktycznie, zapewne już wtedy nie było a na pewno teraz niewielu jest ludzi władających językiem niemieckim a jednocześnie jako tako znających grekę klasyczną, żeby połączyć $\epsilon\upsilon\tau\rho\omicron\pi\acute{\eta}$ z *Verwandlunginhalt* (zawartością przekształcenia), wprowadzoną przez Clausiusa. Jakkolwiek używanie słów i sformułowań nieznanymi i górnolotnie brzmiącymi było przyczyną niechęci Richarda Feynmana do filozofów (niesłusznie, chyba nie czytał *Obrony Sokratesa*, w której tenże tłumaczy się, że będzie mówił w sądzie używając słów nieupiększonych i przypadkowych tak, jak na agorze albo na ulicy) a znane jest też powiedzenie księdza Tischnera, że jeżeli coś nie da się powiedzieć po „gorolsku” to jest nic nie warte. W każdym razie bardzo dobrze o Kandydacie świadczy, że dociekał również pochodzenia używanych pojęć

– jeszcze jeden argument za przyznaniem stopnia doktora, bądź co bądź filozofii. W dodatku B rozważa równoważność transferu entropii i analizy Grangera dla sygnałów gaussowskich.

Do tej części rozprawy mam następującą uwagę. O ile rysunek 2.1 pokazuje wyróżnione minima lokalne dla korelacji w funkcji odległości obciążenia dla represora λ i aktywnej formy PKA, o tyle dla domeny SH3 wyraźne minima występują dla ok. 5.5 oraz 6 Å, jakkolwiek najwięcej minimów jest dla 5 Å. Czy kryterium wyboru była również dyspersja rozkładu minimów dla różnych progów aktywności i nieaktywności kontaktów? Jeżeli tak, to dlaczego?

W rozdziale trzecim Kandydat zaadaptował analizę Grangera, w szczególności poprzez oryginalną implementację algorytmu autoregresji, do badania trajektorii dynamiki molekularnej na przykładzie chignoliny, której długoczasową trajektorię uzyskał od grupy D.E. Shawa. Chignolina jest dziesięcioreztowym minibiałkiem o strukturze pojedynczej spinki β . Jest dobrym modelem do testowania metod badania procesu zwijania/rozwijania z dwóch względów: po pierwsze jest układem małym, a przez to względnie łatwym w opisie, a po drugie struktura drugo- i trzeciorzędowa są tutaj tożsame. W toku analizowanej symulacji białko zwijało się i rozwijało wielokrotnie.

Analiza Grangera jest ukierunkowana na następstwa poszczególnych zdarzeń co, jak słusznie zauważa Kandydat, jest często mylone z analizą przyczynowo-skutkową. Podobnie jak w poprzednim rozdziale, Kandydat opracował oryginalny aparat matematyczny, umożliwiający zastosowanie analizy Grangera do badania trajektorii dynamiki molekularnej. Do analizy użył kontaktów natywnych, średniego odchylenia względem struktury odniesienia (RMSD; ang. *Root Mean Square Deviation*) oraz kątów łańcucha głównego ϕ i ψ . W celu znalezienia krajobrazu energii swobodnej, a dalej współrzędnej reakcji, przeprowadził analizę TICA (*Time-lagged Independent Component Analysis*). Przy pomocy opracowanego aparatu matematycznego zmodyfikowanej analizy Grangera zidentyfikował hierarchię tworzenie kontaktów stwierdzając, że najważniejsze w procesie zwijania są te, w które zaangażowane są reszty zgięcia a następnie (w sensie negatywnym) kontakty na N-końcu łańcucha, które muszą być rozerwane aby mogły się utworzyć kontakty dalekozasięgowe i struktura natywna stała się kompletna. Podsumowując stwierdził, że mechanizm zwijania chignoliny można porównać do zapinania zamka błyskawicznego, natomiast nie ma on a charakteru zapadnięcia (kolapsu) hydrofobowego. Co ciekawe, do takiego samego wniosku (dla innego układu) doszedł parę lat temu Ołdziej i współpracownicy, którzy badali eksperymentalnie termiczne rozwijanie C-końcowej spinki β z białka G (*Biophys. Chem.*, **2010**, 151, 1).

Kandydat określił również współrzędną reakcji zwijania. Stwierdził, że bariera potencjału średniej siły przejścia jest większa, niż oszacowana w pracy grupy D.E. Shawa, gdzie barierę oszacowano na podstawie arbitralnie wybranej współrzędnej reakcji, którą był ułamek kontaktów natywnych. Znalezione poprzez analizę Grangera „kamienie milowe” zwijania leżą na ścieżce reakcji, którą można poprowadzić tylko na podstawie mapy potencjału średniej siły. W szczególności punkt odpowiadający strukturze przejściowej znajduje się dokładnie w punkcie siodłowym mapy potencjału średniej siły.

Uzupełnieniem informacji o analizie Grangera oraz analizie trajektorii jest dodatek C.

Mam jedną uwagę merytoryczną do trzeciego rozdziału. Jak zrozumiałem, przeprowadzona analiza Grangera była ukierunkowana raczej na proces zwijania a nie rozwijania. Również przykładowe trajektorie pokazane na rysunku 3.4 są trajektoriami zwijania. Na podstawie przewidywalności tworzenia innych kontaktów z analizy Grangera Kandydat słusznie zauważa,

że utworzenie kontaktów bliskich zgięciu niesie największy ładunek przewidywania utworzenia kolejnych. Jeżeli dobrze zrozumiałem, to na tej podstawie Kandydat sugeruje, że również rozwijanie zaczyna się od rozerwania tych kontaktów. Wydaje się jednak, że dla poparcia takiego stwierdzenia należałoby przeanalizować któreś ze zdarzeń rozwijania podobnie, jak zrobiono to dla zwijania (rysunek 3.4). Intuicyjnie należałoby się spodziewać, że kontakty w okolicy zgięcia pękają jako ostatnie.

W rozdziale czwartym Kandydat rozszerza analizę Grangera na zmienne wektorowe. Wprowadza tutaj algebrę kwaternionów, co docelowo może umożliwić zastosowanie tej analizy do opisu tworzenia kompleksów białkowych. Analizę stosuje do układu trzech ładunków dodatnich we wnęce kulistej w dwóch trybach: ładunków poruszających się swobodnie w obrębie wnęki oraz dwóch ładunków poruszających się swobodnie i jednego orbitującego po znanej trajektorii. Do opisu ruchu stosuje dynamikę Langevina z różnymi wartościami współczynnika lepkości; również różne są wielkości ładunków. Podobnie jak chignolina, układ jest prosty ale nie nadmiernie ponieważ wiadomo, że rozwiązania równań ruchu trzech ciał oddziałujących siłami centralnym w przestrzeni trójwymiarowej wykazują niestabilność Lyapunova, a zatem są „numerycznie” indeterministyczne, nawet w nieobecności sił stochastycznych. Kandydat wykazał, że wzajemny wpływ ładunków na ruchy zmniejsza się ze zwiększeniem współczynnika lepkości i że największy jest wpływ ładunku orbitującego (w przypadku drugiego trybu symulacji). Ta część rozprawy wygląda raczej na badanie wstępne ponieważ, aby w pełni wykorzystać zalety użycia algebry kwaternionów, należałoby go zastosować do układów o niesferycznej symetrii oddziaływań, np. dipoli albo kwadrupoli, jeżeli ograniczyć się do oddziaływań elektrostatycznych. Tym niemniej, w tej części rozprawy opracowano wymagany aparat matematyczny oraz stosowne algorytmy obliczeniowe, które mogą być zastosowane do opisu bardziej skomplikowanych układów.

Uzupełnieniem rozdziału czwartego jest dodatek D, w którym Kandydat szczegółowo omawia rozszerzenie analizy Grangera, w szczególności algorytmu autoregresji na zmienne nieskalarno (wektorowe i kwaterniony).

Mam następujące uwagi oraz pytania do rozdziału czwartego. Kandydat pisze, że do rozwiązywania równań ruchu zastosował metodę Runge-Kutty drugiego rzędu. Dlaczego nie zastosował któregoś z algorytmów typu Verleta, które są symplektyczne, jeżeli działają tylko siły potencjalne? Wprawdzie Kandydat włączył siły niekonserwatywne ale z własnych (przykrych) doświadczeń wiem, że niezadbanie o zachowanie energii w symulacjach mikrokanonicznych skutkuje znaczącymi często błędami wartości średnich w symulacjach dynamiki Langevina. Co gorsza, badacz często sobie z tych błędów nie zdaje w ogóle sprawy, ponieważ termostat Langevina działa tak znakomicie, że rozkład temperatury jest prawie perfekcyjnie zgodny z uogólnionym na daną liczbę stopni swobody rozkładem Maxwella-Boltzmann. W nawiązaniu do tej uwagi, czy Kandydat wyznaczył rozkład temperatury i porównał go z uogólnionym rozkładem Maxwella-Boltzmann? W jaki sposób były traktowane w symulacjach siły lepkości i stochastyczne? Czy dodawano je wprost do prawej strony równań ruchu czy stosowano pre-całkowanie jak w algorytmie Ciccottiego (np. *Mol. Phys.*, **2003**, 101, 1302; *Philos. Trans. R. Soc. London, Ser. A*, **2004**, 362, 1583)?

W rozdziale 5 Kandydat podaje perspektywy dalszego rozwoju opracowanych w ramach rozprawy zastosowań analizy Grangera do analizy symulacji biomolekularnych na uwzględnienie kanałów mieszanych i uczenie maszynowe. Natomiast oprócz rozwoju metodologii wydaje się, że należałoby wspomnieć o możliwości przyszłych zastosowań opracowanej przez Kandydata

analizy trajektorii do bardziej skomplikowanych białek, gdzie tworzenie struktury drugo- i trzeciorzędowej nie współbrzmia ze sobą. W pracy grupy Shawa, z której Kandydat korzystał znajdują się np. dane dla 20-resztowej klatki tryptofanowej, która zawiera dwa elementy spakowane ze sobą: helisę α i tzw. helisę poliprolinową. Dla tego białka również zaobserwowano wielokrotne zwijanie i rozwijanie. Nieco większe białka są również obecne w tym zestawie symulacji. Do symulacji zwijania białek 100- i więcej resztowych konieczne są metody gruboziarniste i to raczej nie te oparte na potencjałach typu Gō.

Z obowiązku recenzenta muszę wymienić nieliczne nieścisłości i błędy językowe oraz edytorskie, które zauważyłem czytając rozprawę; podaję je w kolejności:

str. 50 wiersz 3 pod równaniem 3.27: ?? zamiast odnośnika do dodatku; błąd $\text{\TeX}$ ologii.

str. 65, drugi akapit (mniej więcej w połowie strony): Tak zdefiniowany potencjał średniej siły jest bezwymiarowym potencjałem średniej siły, o czym należało napisać.

str. 78, linie 3 i 4 od dołu: „wzdłuż gradientu energii swobodnej” – chyba „wzdłuż gradientu potencjału średniej siły”, ponieważ energia swobodna jest funkcją zmiennych stanu, do których nie należą współrzędne.

str. 87, ostatnie zdanie sekcji 1: „we wnętrzu wnęki kulistej” (to sformułowanie powtarza się jeszcze w paru miejscach) – brzmi niezręcznie; lepiej by było „we wnęce kulistej” albo „w obrębie wnęki kulistej” jeżeli koniecznie trzeba zaznaczyć, że obiekt nie może się poza nią znaleźć.

Rozdział 4.3, str. 109: zamiast „zrezygnowano z jednostek fizycznych” należałoby napisać „przyjęto jednostkowe ładunki i stałą dielektryczną próżni równą 1”.

Rys. 4.3: Należałoby napisać, że kolumna/wiersz odpowiada indeksowi ładunku a nie ładunkowi. Sformułowanie „ładunek” kojarzy się z wielkością ładunku i początkowo dziwiłem się, dlaczego w panelu w lewym dolnym rogu jest zawsze „1”.

str. 117, 3 linia od końca drugiego akapitu: „...ta druga właściwość jest wygodne...” — błąd gramatyczny.

str. 117-118: „Zamiast relacje pomiędzy kontaktów...” – błąd gramatyczny.

str. 127, początek drugiego akapitu: „Miarą transfer entropii...” – błąd gramatyczny.

Rozprawę doktorską Pana mgra Marcina Sobieraja oceniam bardzo wysoko. Cel badawczy został w pełni zrealizowany. Wymienione uwagi krytyczne, z których przynajmniej część ma charakter dyskusyjny, nie umniejszają w znaczącym stopniu jej wartości. Kandydat wykonał ogromną pracę, wymagającą bardzo dużej wiedzy w zakresie matematyki, w tym matematyki numerycznej, fizyki, chemii oraz metod symulacyjnych. Wykazał biegłość w operowaniu zaawansowanymi metodami matematycznymi i konstrukcji oryginalnych algorytmów analizy danych. Cała przedstawiona w pracy analiza oraz interpretacja wyników są profesjonalne. Rozprawa zawiera bardzo duży ładunek nowości naukowej, w szczególności pod względem zastosowania analizy Grangera do analizy trajektorii symulacji białek. Opracowana metodologia może stać się standardowym narzędziem analizy wyników symulacji biomolekularnych. Rozprawa spełnia z dużą nawiązką wymagania stawiane rozprawom doktorskim przez Ustawę z

dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (tekst jednolity: DzU z 2022 r. poz. 574 z późn. zm.), jak również zwyczajowe standardy stawiane rozprawom doktorskim w dziedzinie nauk ścisłych i przyrodniczych. Dlatego z pełnym przekonaniem wnoszę do Wysokiej Rady Dyscypliny Nauki Fizyczne Uniwersytetu Warszawskiego dopuszczenie Pana mgra Marcina Sobieraja do dalszych etapów przewodu doktorskiego. Jednocześnie, z uwagi na bardzo dużą wartość naukową, w szczególności duży ładunek nowości naukowej rozprawy, wnoszę o jej wyróżnienie. Uzasadnienie wyróżnienia załączam.



prof. dr hab. Józef Adam Liwo



Uniwersytet
Gdański



Związek
Uczelni
Fahrenheita

Wydział Chemii
ul. Wita Stwosza 63
80-308 Gdańsk

Gdańsk, dnia 28.08.2023

**Uzasadnienie wyróżnienia rozprawy doktorskiej
Pana mgra Marcina Sobieraja
pt. „Nowe metody analizy mechanizmów funkcjonowania układów
biomolekularnych w oparciu o symulacje dynamiki molekularnej”**

Rozprawa doktorska pana mgra Marcina Sobieraja dotyczy opracowania zaawansowanych metod analizy symulowanych trajektorii zwijania oraz innych rozległych zmian konformacyjnych białek i innych makromolekuł biologicznych. Z uwagi na to, że istotne informacje o układzie, które można uzyskać z symulacji giną w szumach, jest to „gorące” zagadnienie i mimo ponad 30 lat intensywnych badań (począwszy od pierwszych zastosowań metody czynników głównych) nie opracowano jeszcze zadowalającego algorytmu. Kandydat stworzył oryginalną implementację algorytmu Grangera, który umożliwia uwzględnienie wzajemnego wpływu poszczególnych zdarzeń tak właśnie jak to się dzieje w symulacjach dynamiki. Z powodzeniem zastosował opracowaną metodologię do analizy trajektorii zwijania/rozwijania minibiałka o strukturze spinki β (chignoliny). Ponadto rozszerzył analizę na zmienne wektorowe, stosując zaawansowany matematycznie formalizm kwaternionów i przeprowadził pierwszą próbę jej zastosowania do prostego układu ładunków dodatnich we wnęce kulistej. Opracował też obiektywną metodę, opartą na teorii informacji, określania odległości obciążenia dla kontaktów międzyresztowych w białkach. Opracowana metodologia może stać się standardem analizy trajektorii dynamiki molekularnej. Ze spisu literatury wynika, że część materiału rozprawy została opublikowana w renomowanych czasopismach międzynarodowych. W mojej opinii rozprawa p. mgra Marcina Sobieraja w pełni zasługuje na wyróżnienie.

prof. dr hab. Józef Adam Liwo