

prof. dr hab. inż. JAN MAZERSKI  
WYDZIAŁ CHEMICZNY  
POLITECHNIKA GDAŃSKA  
ul. Narutowicza 11/12  
80-233 Gdańsk  
☎: (58) 347 11 16 347 18 06  
Fax: (58) 347 11 44  
e-mail: jan.mazerski@pg.gda.pl.

---

Gdańsk dn. 31.12.2015r

Recenzja rozprawy doktorskiej

mgr TOMASZA WIRECKIEGO

pt.

*„Molecular Dynamics and Monte Carlo methods and their extensions for  
coarse-grained simulations of peptide folding”*

promotor pracy: **dr hab. Cezary Czaplewski**

promotor pomocniczy: **dr Artur Gieldoń**

Jednym z istotnych zagadnień współczesnej biochemii strukturalnej i biotechnologii jest zdobycie umiejętności przewidywania trzeciorzędowej struktury polipeptydów i białek o zadanej sekwencji aminokwasów. Projekt Ludzkiego Genomu nie tylko pozwolił na poznanie sekwencji aminokwasowych zakodowanych w nim białek, ale również przyczynił się do opracowania i komercjalizacji prostych i względnie tanich metod ustalania genomu prostych organizmów, takich jak np. bakterie. Obecnie liczba białek o znanych sekwencjach wzrasta w tempie prawie wykładniczym. Jednakże liczba białek dla których określono doświadczalnie strukturę 3D (trzeciorzędową) ciągle jest bardzo mała i wzrasta w zasadzie jedynie liniowo. W tej sytuacji istnieje pilna potrzeba opracowania metody lub metod obliczeniowych pozwalających na szybkie i w miarę wiarygodne przewidzenie struktury 3D co najmniej polipeptydów i krótkich białek.

Recenzowana rozprawa mgr Tomasza Wireckiego wykonana została w Katedrze Chemii Teoretycznej Wydziału Chemii Uniwersytetu Gdańskiego w zespole naukowym kierowanym przez dr hab. Cezarego Czaplewskiego. Zespół ten zajmuje się m.in. badaniami nad symulacją struktury biopolimerów z wykorzystaniem technik gruboziarnistych. Zespół ma na tym polu szereg istotnych osiągnięć na poziomie międzynarodowym. Świadczy o tym m.in. zaproszenie zespołu do udziału w

międzynarodowym projekcie badawczym CASP, którego celem jest opracowanie satysfakcjonującej metody przewidywania struktury trójwymiarowej białek jedynie na podstawie ich sekwencji aminokwasowej. Rozprawa doktorska mgr Wireckiego stanowi próbę oceny przydatności na tym polu jednego z nowych algorytmów symulacji, MREHMC, i stanowi naturalną kontynuację badań prowadzonych dotychczas w zespole.

Przedstawiona do recenzji rozprawa doktorska napisana została w całości w języku angielskim, nawet bez streszczenia w języku polskim. Składa się ona z 5 rozdziałów merytorycznych i 4 części pomocniczych. Rozdziały merytoryczne zatytułowane są: *Theory and methods*, *Aim of the studies*, *Results*, *Conclusions* i *Literature*. Na część pomocniczą składają się wykazy: stosowanych skrótów, uzyskanych grantów i komputerów wykorzystywanych do symulacji, publikacji w których Doktorant jest współautorem oraz materiały uzupełniające.

Rozprawa pozbawiona jest jakiegokolwiek wstępu czy wprowadzenia, co istotnie utrudnia Czytelnikowi zorientowanie się w bardzo obszernej części teoretycznej. W pewnym stopniu funkcję tę pełni podrozdział *2.1 Protein folding mechanisms*, nie jest jednak w stanie w pełni zastąpić poprawnego wprowadzenia w tematykę rozprawy.

Rozdział *Theory and methods* przedstawiony został na 60 stronach i podzielony jest na 6 głównych podrozdziałów: *Protein folding mechanisms*, *Conformational space search*, *Protein folding simulations*, *UNRES force field*, *Bacteriocins* i *Genom comparison*. W 4 pierwszych podrozdziałach przedstawione zostały w zwartej formie podstawowe problemy związane z symulacją fałdowania się białek, w tym algorytmy przeszukiwania przestrzeni konformacyjnej oraz gruboziarniste pole siłowe UNRES. Omawiany materiał opiera się na 241 pozycjach literaturowych.

W jednostronicowym rozdziale *Aim of the studies* mgr Wirecki przedstawia główny cel swojej pracy doktorskiej oraz zadania badawcze, które zrealizował w jej ramach. Celem tym było porównanie nowego algorytmu przeszukiwania przestrzeni konformacyjnej z wykorzystaniem hybrydowej metody Monte Carlo (MREHMC) z jej poprzednią wersją opartą na klasycznej dynamice molekularnej (MREMD). Przedstawione w tym rozdziale zadania badawcze wykraczają jednak wyraźnie poza tak sformułowany cel rozprawy doktorskiej. W tym miejscu Recenzentowi nasuwa się wątpliwość czy przedmiotem oceny jest rozprawa doktorska czy sprawozdanie z działalności badawczej Doktoranta. Do problemu tego powrócę jeszcze w dalszej części recenzji.

W kolejnym rozdziale zatytułowanym *Results* mgr Wirecki omówił na 67 stronach wyniki badań w których uczestniczył jako doktorant. Rozdział ten składa się z 7 podrozdziałów o tytułach: *Initial tests on Lennard-Jones fluid*, *Proteins used for simulations*, *Multiplexed Replica Exchange Hybrid Monte Carlo*, *MREHMC – large-scale simulations*, *Retrieving the kinetic information from MREMD and MREHMC simulations*, *CASP experiment results* i *Bacteriocins*. Cztery pierwsze

podrozdziały potraktować można jako kolejne etapy realizacji głównego celu pracy, czyli porównania klasycznego algorytmu MREMD z jego nową modyfikacją MREHMC. Kolejny podrozdział przedstawia ocenę możliwości uzyskania informacji kinetycznych z trajektorii MREMD/MREHMC przy zastosowaniu pola siłowego UNRES. Dwa ostatnie podrozdziały przedstawiają przykłady potwierdzające zakres zastosowania algorytmu MREHMC. Zwłaszcza materiał zawarty w ostatnim z nich budzi moje wątpliwości. Jak wynika z tytułu rozprawy oraz z Rozdziału 3 jej celem jest ocena przydatności i możliwości zastosowania określonego algorytmu przeszukiwania przestrzeni konformacyjnej polipeptydów i białek. W tej sytuacji przedstawianie w rozprawie wyników określania genomu pewnych szczepów bakterii oraz wyników porównywania tych genomów z danymi literaturowymi jest może ciekawe z punktu widzenia Doktoranta lecz nie wnosi niczego do realizacji celu rozprawy. Tym bardziej, że wyniki symulacji struktury białek sugerują, niezależnie od zastosowanego algorytmu, brak istotnych różnic w strukturze bakteriocyn i ich białek ochronnych pochodzących z różnych szczepów bakterii. Uważam za bardzo dobry, uświęcony tradycją akademicką zwyczaj przedstawiania w rozprawie doktorskiej tylko tych wyników, które związane są z realizacją celu rozprawy. Rozumiem, że wynikający z tego zwyczaju brak możliwości zaprezentowania części uzyskanych wyników może to być czasami trudny dla Doktoranta, ale wyniki te można przecież opublikować w literaturze fachowej.

Jak przypuszczam, że względu na bardzo obszerne omówienie uzyskanych wyników mgr Wirecki zdecydował się na umieszczenie w pracy 2-stronicowej części o nazwie *Conclusions*. Uważam to za bardzo dobry pomysł. Dzięki temu mógł przedstawić w zwartej formie podstawowe wnioski wpływające z jego badań.

Rozprawę kończy wykaz cytowanej literatury obejmujący łącznie 285 pozycji. Są to w większości prace najnowsze, w znacznej części opublikowane po roku 2000.

W ocenianej rozprawie odczuwam wyraźny brak dyskusji uzyskanych wyników, w tym również porównania jakości uzyskanych przewidywań struktury testowanych białek z przewidywaniami innych autorów. Czyżby żadne z wybranych białek nie było dotychczas przedmiotem symulacji?

Przystępując do oceny merytorycznej rozprawy mgr Tomasza Wireckiego należy podkreślić, że nie jest to praca doświadczalna, lecz „czystej wody” praca obliczeniowa. Jej celem jest porównanie nowej wersji algorytmu symulacji z jego wcześniejszą, niejako klasyczną wersją. Na uwagę zasługuje bardzo staranne zaplanowanie hierarchicznego toku oceny jakości symulacji: poprawność działania algorytmu i jakość uzyskanych przy jego pomocy wyników sprawdzana jest kolejno na:

- układzie modelowym,
- układach rzeczywistych o znanym poprawnym rozwiązaniu (dane doświadczalne) i wzrastającym stopniu komplikacji

- układach rzeczywistych o nieznanym rozwiązaniu (projekt CASP i bakteriocyny)

Taki tok postępowania zgodny jest z zasadami dobrej praktyki obliczeniowej.

Recenzowana rozprawa zawiera, zdaniem recenzenta, dwa podstawowe elementy nowości naukowej o zdecydowanie różnym charakterze:

i) *element ogólny* - wykazanie, że algorytm MREHMC w połączeniu z polem siłowym UNRES pozwala w przypadku polipeptydów a nawet małych białek na efektywne przeszukiwanie przestrzeni konformacyjnej

ii) *element porównawczy* – algorytm MREHMC w porównaniu z algorytmem MREMD jest w większości przypadków jedynie trochę efektywniejszy. Różnice te zanikają przy przedłużeniu czasu symulacji oraz w przypadku zastosowania ulepszonej parametryzacji MAXLIK.

Przy lekturze rozprawy nasunęło mi się szereg wątpliwości i uwag o charakterze dyskusyjnym. Najważniejsze z nich przedstawiam poniżej mając nadzieję, że Doktorant ustosunkuje się do nich podczas publicznej obrony.

1. Na str. 82 znajduje się stwierdzenie „*Finally the extended state starting point was chosen for further tests since it gave better results.*” Nie jest dla mnie jasne dlaczego rozpoczynanie symulacji od struktury natywnej daje gorsze wyniki niż start od nierealistycznej struktury „rozciągniętej”. Ponadto stwierdzenie to stoi w sprzeczności z wynikami opisanymi na str. 87.
2. Porównanie obydwu algorytmów przedstawione w p. 4.3.5 zyskałoby znacznie na wiarygodności, gdyby zostało wsparte nawet najprostszymi testami statystycznymi dla par wiązanych. Wizualna analiza rysunków 28 ÷ 31 nie jest wcale taka jednoznaczna jak sugeruje to Doktorant. Potwierdza to dodatkowo Tab.4. Analogiczna uwaga dotyczy danych z Tab.5 (p.4.4.1) i Tab.6 (p. 4.4.2).
3. Nie jest dla mnie jasne dlaczego do symulacji białek chroniących przed bakteriocynami wybrano algorytm MREMD a nie podobno lepszy algorytm MREHMC (p.4.7.1.2, str. 127). Mam nadzieję, że Doktorant wyjaśni tą kwestię podczas publicznej obrony.
4. Z rozprawy nie wynika (albo nie udało mi się znaleźć tej informacji) czy podczas symulacji białek z bazy PDB (p. 4.3 i 4.4) również stosowano więzy struktury II-rzędowej (*secondary restrains*). Jeżeli nie, to jakie mogą być zdaniem Doktoranta fizykochemiczne lub strukturalne przyczyny konieczności ich zastosowania w przypadku bakteriocyn i białek chroniących przed nimi.

Do obowiązków recenzenta należy też ocena strony językowej i edytorskiej rozprawy. Recenzent nie czuje się kompetentny do oceny poprawności stylistycznej i gramatycznej angielskiego tekstu rozprawy. Użyte słownictwo fachowe nie budzi wątpliwości. Zamieszczony w pracy materiał ilustracyjny jest bogaty i poprawny pod względem technicznym. Znacznie gorzej wygląda strona

edytorska tekstu i opisy rysunków. Można tu znaleźć bardzo liczne przykłady błędów i omyłek. Poniżej podam tylko dwa reprezentatywne przykłady:

- oś pionowa na rysunkach 12, 13, 15, 16, 27 zatytułowana jest „*Probability of Acceptance [%]*” podczas gdy wartości liczbowe zawarte są w przedziale od 0,0 do 1,0.
- w tytułach tabel 5 i 6 występuje zwrot „... *of all 15 proteins ...*” podczas gdy obliczenia przeprowadzono jedynie dla 10 polipeptydów. Również w tabelach do których odnoszą się te tytuły znajdują się dane tylko 10 polipeptydów.

Opisy rysunków są często na tyle niejednoznaczne, że nawet korzystanie z tekstu rozprawy nie rozwiewa wszystkich wątpliwości, np. co oznacza zielona linia na rys. 49?

Recenzent odnosi wrażenie, że tekst rozprawy przygotowywany był w pośpiechu i bez dostatecznej staranności.

Podsumowując moje uwagi stwierdzam, że recenzowana rozprawa doktorska spełnia zwyczajowe i ustawowe wymogi stawiane pracom doktorskim i wnoszę do Rady Wydziału Chemii Uniwersytetu Gdańskiego o dopuszczenie **mgr Tomasza Wireckiego** do dalszych etapów przewodu doktorskiego.

prof. dr hab. inż. Jan Mazerski