

Recenzja

pracy doktorskiej magistra Tomasza Wireckiego

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

Problem zwijania liniowego łańcucha peptydowego do natywnej struktury białka jest jednym z głównych, lecz nadal nierozwiązanych problemów biologii strukturalnej. Gdyby proces zwijania polipeptydu do struktury natywnej zachodził na drodze losowego przeszukiwania przestrzeni konformacyjnej, to jego czas byłby niemożliwie długi. Ponieważ zwijanie następuje w skali czasowej rzędu sekund lub krótszej, w procesie tym musi mieć miejsce uporządkowana hierarchia zdarzeń, tj. preferowana ścieżka zwijania. Jednakże, nadal brakuje ogólnej informacji jaka jest sekwencja i waga zdarzeń na tej ścieżce.

Znajomość przestrzennej struktury białka jest niezwykle przydatna w projektowaniu i prowadzeniu badań *in-vitro* oraz *in-vivo* na poziomie molekularnym i interpretacji ich wyników i jest niezbędna przy racjonalnym projektowaniu leków. Metody eksperymentalne nie są w stanie rozwiązać struktur przestrzennych wszystkich białek, których sekwencje aminokwasowe są znane – tylko dla mniej niż 0.2% znanych sekwencji znane są struktury. Dlatego rośnie zapotrzebowanie na przewidywanie struktur białek metodami komputerowymi. Przewidywanie struktur wymaga jednak zarówno wielkiej mocy jak i zaawansowanych metod obliczeniowych.

Przedstawiona mi do recenzji praca doktorska poświęcona jest komputerowym metodom przewidywania struktur przestrzennych białek. Mgr Tomasz Wirecki zaimplementował w pakiecie UNRES nowy algorytm i przeprowadził szereg testów i analiz oceniających jego skuteczność. Ten algorytm, opracowany wysiłkiem kilku grup badawczych, łączy dwie zaawansowane metody obliczeniowe w celu polepszania efektywności i poprawności przewidywania struktur przestrzennych peptydów i białek wychodząc z praw podstawowych. Oprócz oceny nowego algorytmu, doktorant przeprowadził porównanie starej i nowej parametryzacji pola siłowego UNRES. Doktorant określił również stałe czasowe procesu zwijania badanych białek. W tym celu zaadaptował program *g_kinetics* z pakietu GROMACS. Na koniec, sprawdził skuteczność pakietu UNRES w eksperymencie CASP10, gdzie przewidywał struktury kilku białek, dla których znane

były jedynie sekwencje aminokwasowe, a także wykorzystał go do oszacowania aktywności przeciwdrobnoustrojowej wybranych bakteriocyn.

Ponieważ pole UNRES jest grubo-ziarnistym polem siłowym, końcowym etapem każdej symulacji prowadzonej pakietem UNRES była konwersja modelu grubo-ziarnistego cząsteczek do modelu pełno-atomowego.

Przed rozpoczęciem prac nad oceną nowego algorytmu zwijania łańcucha peptydowego, doktorant przeprowadził porównawcze symulacje dynamiki prostych układów zbudowanych z atomów oddziałujących jedynie poprzez potencjały Lennarda-Jonesa. Do porównania użył metody klasycznej symulacji dynamiki molekularnej (MD) z dwoma sposobami termostatowania układu symulacyjnego: Berendsena (BER) i Nosé-Hoovera (NH), oraz trzech hybrydowych metod Monte Carlo (MC): Hybrid Monte Carlo (HMC), Shadow Hybrid Monte Carlo (SHMC) oraz Separable Shadow Hybrid Monte Carlo (S2HMC). Na podstawie Appendix'u można wywnioskować, że programy do symulacji metodami HMC i S2HMC napisał doktorant. Dla każdej z metod sprawdził jej ergodyczność, a dla metod hybrydowych sprawdził zależność wskaźnika akceptacji Metropolis'a od liczby kroków czasowych symulacji oraz od liczby L. W zasadzie wszystkie algorytmy działały poprawnie, jednak wśród hybrydowych metod MC najlepsza była metoda S2HMC, a wśród metod MD – metoda NH.

Niestety, zjawisko skokowego porządkowania się układu atomów symulowanego praktycznie każdą z metod nie zostało w pracy wyjaśnione, czy choćby skomentowane i jest dla mnie niezrozumiałe. W znanych mi układach, takie zachowanie jednoznacznie wskazywałoby na niepoprawną parametryzację atomów.

Głównym celem doktoranta była ocena skuteczności i wydajności nowo zaimplementowanej w pakiecie UNRES metody hybrydowego Monte Carlo ze zwielokrotnioną wymianą replik (*Multiplexed Replica Exchange Hybrid Monte Carlo*, MREHMC) w przewidywaniu struktur przestrzennych białek. W tej części pracy symulacje i analizy prowadził na 15 białkach o różnych liczbach reszt aminokwasowych i znanych strukturach natywnych. Wzajemne podobieństwo sekwencji wybranych białek było niskie. Jako metody porównawczej użył symulacji dynamiki molekularnej ze zwielokrotnioną wymianą replik (*Multiplexed Replica Exchange Molecular Dynamics*, MREMD), która już wcześniej była częścią pakietu UNRES. W testowych obliczeniach użył obu parametryzacji pola siłowego UNRES, E0LL2Y i GAB, jednak wyniki otrzymane dla GAB nie były zadowalające. Dla wszystkich 15 białek, symulacje rozpoczynał od struktury natywnej (z bazy PDB) oraz rozwiniętej. Ponieważ symulacje startujące z rozwiniętej struktury dawały lepsze wyniki (str. 82), w dalszych symulacjach użył tej struktury jako startowej oraz

parametryzacji EOLL2Y. Przeprowadził też szereg innych testów, niektóre takie jak w pierwszej części rozprawy, by ustalić optymalne parametry symulacji.

Ocenę wyników symulacji przeprowadził wyznaczając dla każdego białka krzywe pojemności cieplnej oraz pierwiastka ze średniego kwadratu odległości (RMSD) względem struktury natywnej w funkcji temperatury i czasu symulacji. Otrzymane w symulacjach struktury zostały podzielone na klastry wg kryterium temperatury; dla każdej struktury została obliczona wartość RMSD, a dla każdego klastra, średnia wartość RMSD; wyznaczono też *TM-score*. Zależnie od kryterium, albo metoda MREMD albo MREHMC dawała lepsze wyniki.

Liczbowe oceny oraz graficzne przedstawienie niektórych przewidzianych struktur nie wskazują na ich wysokie podobieństwo do struktur natywnych (pdb). Doktorant nie podał kryterium „akceptacji” przewidzianej struktury dla RMSD. Ogólnie przyjmuje się, że dwie struktury są podobne, jeśli RMSD jest nie większe niż błąd metody, którą rozwiązano strukturę. Dla *TM-score*, doktorant przyjął wartości nie mniejsze niż 0.4. Przyjmując jednak $\text{RMSD} \leq 3.5 \text{ \AA}$ jako zakres akceptowalnych wartości, z grubsza, 6 na 15 struktur białek zostało akceptowalnie przewidzianych metodą MREMD oraz 5 na 15 metodą MREHMC. Taka liczba przewidzianych struktur zapewne odpowiada statystyce otrzymanej z eksperymentów CASP i jest dużym sukcesem. Biorąc pod uwagę kryterium *TM-score*, tylko jedna struktura spełnia je, ale jego wysoka wartość nie odpowiada małej wartości RMSD. Na wybranych przykładach, doktorant stara się przekonać czytelnika o wyższości metody MREHMC nad metodą MREMD, ale nie do końca skutecznie.

Opisując analizy wyników obliczeń prowadzonych w tej części pracy, doktorant nie wyjaśnił, dlaczego wyznacza krzywe pojemności cieplnej białka i o czym te krzywe mówią. Czy jedynym powodem ich wyznaczania jest określenie temperatury „klastrowania”? Skąd wzięło się takie kryterium? W przypadku metody MREHMC, dla różnych fragmentów trajektorii, maksima krzywych pojemności cieplnej są dla różnych temperatur, a niektóre krzywe mają dwa równoważne maksima – którą temperaturę należy wtedy wybierać?

Są też dwa różne uzasadnienia wyboru struktury rozciągniętej białka jako startowej; na str. 82 „bo dawała lepsze wyniki”, na str. 87 „...struktura natywna dawała lepsze wyniki ... ale celem symulacji jest otrzymanie natywnej struktury na podstawie znajomości tylko sekwencji” (luźne tłumaczenie recenzenta).

Strona 102, zdanie dotyczące białka 1CLB (76 reszt aminokwasowych): „...MREHMC is able to locate the global minimum of the conformational space and provide cluster with native structure” budzi wątpliwości, bo skąd wiadomo, że globalne minimum zostało osiągnięte? Wprawdzie *TM-score* ma wartość 0.41, ale RMSD wynosi 4.22. Dla tego RMSD energia układu ma wartość niską,

ale dla RMSD ~12 ma podobną lub jeszcze niższą; z kolei dla stanu natywnego białka energia może być jeszcze mniejsza – czy doktorant porównał energię przewidzianą i natywnej struktury?

W następnej części pracy doktorant przeprowadził analogiczne symulacje dla mniejszego zbioru białek, ale z użyciem znacznie większych mocy obliczeniowych, co umożliwiło zwiększenie liczby replik i temperatur oraz wydłużenie czasu symulacji. W symulacjach użył dwóch parametryzacji pola UNRES – poprzedniej E0LL2Y oraz nowo opracowanej MAXLIK. Dla obu metod symulacyjnych i starego pola siłowego, otrzymane struktury były bliższe strukturom natywnym niż w obliczeniach opisanych w poprzedniej części pracy. Jak poprzednio, doktorant stara się przekonać czytelnika o wyższości metody MREHMC nad metodą MREMD.

Obie metody dały jeszcze lepsze wyniki, gdy użyto nowego pola siłowego MAXLIK; w tym jednak przypadku, metoda MREMD okazała się generalnie lepsza niż MREHMC.

Doktorant wyznaczył stałe czasowe dynamicznego procesu zwijania białek dla tego samego zbioru białek, co wyżej, z symulacji prowadzonych trzema metodami: MD, MREMD i MREHMC, z użyciem dwóch parametryzacji: E0LL2Y oraz MAXLIK. Parametry symulacji były podobne do tych powyżej. Do przeprowadzenia analiz czasowych wykorzystał program g_kinetics z pakietu GROMACS, użycie którego wymagało odpowiedniego skonwertowania zbiorów wynikowych wygenerowanych przez programy z pakietu UNRES. Dla każdej z metod symulacyjnych i każdego pola siłowego wyznaczył stałą czasową zwijania i rozwijania łańcucha peptydowego do lub ze struktury zwiniętej. Kryterium stanu zwiniętego i rozwiniętego była wartość RMSD – progowa wartość RMSD została wyznaczona z histogramu wartości RMSD względem struktury natywnej wzdłuż trajektorii każdego białka.

Wyznaczone stałe czasowe zwijania i rozwijania łańcucha peptydowego są względnie podobne dla metod MD i MREMD i obu pól siłowych i są istotnie krótsze od stałych wyznaczonych w symulacjach MREHMC. Ponadto, dla symulacji MREHMC z polem siłowym MAXLIK wyznaczone stałe są dłuższe niż z polem E0LL2Y. Większość stałych, przynajmniej dla symulacji MD i MREMD, jest rzędu ps, tj. są bardzo krótkie. Czy to znaczy, że w czasie symulacji białko zwija się i rozwija wielokrotnie? Profile czasowe na Rys. 49 raczej nie wskazują na czasy zwijania w skali ps – jak należy je rozumieć? Poza tym, wyniki otrzymane z analiz stałych kinetycznych nie wydają się być w zgodzie z wynikami przewidywania struktur przestrzennych białek opisanymi w paragrafie 4.4.2, gdzie różnica wyników otrzymanych w symulacjach MREMD i MREHMC nie była znaczna. Podsumowując tę część pracy należy stwierdzić, że wyniki dotyczące stałych kinetycznych nie są wystarczająco przedyskutowane.

Na koniec części pracy oceniającej nowe elementy pakietu UNRES, doktorant opisał udział w eksperymencie CASP10 i odniesione tam sukcesy. Nie wspomniał jednak, która parametryzacja pola siłowego UNRES była użyta.

Ostatnia część pracy doktorskiej mgr Tomasza Wireckiego dotyczy przewidywania struktur przestrzennych trzech bakteriocyn. Celem tych badań było znalezienie związku między brakiem lub posiadaniem odporności bakterii na własne bakteriocyny a różnicami w ich strukturach oraz strukturach ich białek odpornościowych. Niestety, przewidziane struktury nie różniły się na tyle, by na tej podstawie wskazać różnice w ich działaniu. Dalsze poszukiwania związku prowadzone były na poziomie analizy genomowej i nie miały bezpośredniego powiązania z główną tematyką pracy doktorskiej.

Na koniec, trudno się zgodzić z konkluzjami doktoranta. Ogłosił sukces zaimplementowanego przez siebie algorytmu MREHMC (*Multiplexed Replica Exchange Hybrid Monte Carlo*) dla gorszych symulacji i gorszego pola siłowego, a krótko skwitował zdecydowanie większe, ale podobne obu metod symulacyjnych (MREMD i MREHMC) sukcesy, w znacznie dłuższych i pełniejszych symulacji oraz lepszego pola siłowego. Zrozumiałe jest zadowolenie z własnej pracy i sukcesów, ale w nauce obowiązuje obiektywizm; nadmierna reklama i akcentowanie tylko pewnych wyników przy braku podkreślenia innych, które są wielkim sukcesem, uważam za graniczące z wszechobecną dziś propagandą.

Oceniając całość rozprawy doktorskiej, stwierdzam, że mgr Tomasz Wirecki wykonał gigantyczną pracę przeprowadzając wielką liczbę symulacji różnymi metodami, dla różnych parametryzacji pola siłowego UNRES i dla różnych układów cząsteczkowych. Jego opanowanie warsztatu, zrozumienie fizycznych podstaw metod oraz ich aspektów informatycznych jest godne podkreślenia. Zwraca również uwagę samodzielność doktoranta w doborze narzędzi, analizie wyników i ich prezentacji.

Zastrzeżenia budzi natomiast niedoskonałość tekstu napisanego w nie-ojczystym języku angielskim, w którym jest sporo błędów (mylenie liczby pojedynczej i mnogiej; niepoprawne używanie rodzajników [ich brak, albo niepotrzebne użycie]; błędy gramatyczne [np. „much more slower”] itp.). Pracy brakuje wstępu, który zazwyczaj wprowadza w tematykę i prezentuje aktualny stan wiedzy dotyczącej omawianych w pracy zagadnień. Ponadto, ułatwiłby on czytelnikom i recenzentom „wejście” w tematykę.

Poważnym błędem formalnym pracy jest brak streszczeń w języku polskim i angielskim, wymaganych *Ustawą o stopniach naukowych i tytule naukowym* (Art. 13, pkt. 6).

W pracy jest trochę drobnych błędów, m. in. w używanych skrótach; opis rysunku 50 jest niedokładny; w tekście jest odniesienie do rysunku 12 zamiast 11; na końcu tytułu nie powinno być kropki.

Poza tym, na str. 118 wprowadzone są stałe czasowe τ_f i τ_u , które, jak mówi pierwsze zdanie pod Rys. 48, są obliczone dokładnie tak, jak opisano w sekcji 2.4.4.1. Niestety, w tej sekcji parametry takie nie występują. Czy na str. 121, „s” w „log(s)”, oznacza sekundę?

Odczuwam też brak odniesienia się w dyskusji do roli wody w przewidywaniu struktur białek – na ile pole siłowe i oprogramowanie UNRES jest w stanie poprawnie przewidzieć strukturę przestrzenną białka traktując wodę w sposób niejawną? Taka dyskusja pozwoliłaby „osobie z zewnątrz” lepiej ocenić sukcesy osiągnięte przez doktoranta.

Zgodnie z *Ustawą o stopniach naukowych i tytule naukowym*, rozprawa doktorska powinna stanowić oryginalne rozwiązanie problemu naukowego oraz wykazywać ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej. Przedstawiona mi do oceny praca doktorska mgr Tomasza Wireckiego bezwzględnie wskazuje na samodzielne i umiejętne rozwiązanie postawionego przed nim problemu naukowego oraz jego ogólną i szczegółową wiedzę teoretyczną w dziedzinie szeroko rozumianej biologii strukturalnej, a także umiejętności w zakresie zastosowań zaawansowanych metod chemii obliczeniowej do rozwiązywania bardzo trudnych problemów strukturalnych i dynamicznych biocząsteczek. Wnoszę więc do Rady Wydziału Chemii Uniwersytetu Gdańskiego o dopuszczenie mgra Tomasza Wireckiego do dalszych etapów przewodu doktorskiego.

Masekiewicz Aetula