

„Algorytmy dynamiki molekularnej i Monte Carlo oraz ich rozszerzenia w gruboziarnistych symulacjach zwijania polipeptydów.”

Streszczenie.

Pierwsza część mojego projektu obejmowała przetestowanie algorytmu Hybrid Monte Carlo (oraz jego dwóch rozszerzeń: Shadow Hybrid Monte Carlo i Separable Shadow Hybrid Monte Carlo) na prostym modelu płynu Lennarda-Jonesa oraz porównanie ich z dynamiką molekularną przy użyciu dwóch różnych algorytmów termostatujących: termostat Berendsena oraz termostat Nose-Hoovera. Przeprowadziłem analizę histogramów temperaturowych dla wszystkich metod, testy ergodyczności w celu zbadani wydajności algorytmów oraz testy akceptowalności w zależności od wartości kroku czasowego MD jak i ilości kroków dynamiki pomiędzy kolejnymi testami dla metod hybrydowych. Wyniki pokazały iż metody HMC oraz S2HMC są obiecujące w kontekście dalszych badań.

Główną częścią mojego projektu była implementacja algorytmu HMC do pola siłowego UNRES, oraz połączenie go z już istniejącym w programie algorytmem Multiplexed Replica Exchange Molecular Dynamics. Nowozaimplementowany algorytm nazwałem Multiplexed Replica Exchange Hybrid Monte Carlo. Nowa metoda została kompleksowo przetestowana i porównana z 'klasycznym' algorytmem MREMD. Klaster obliczeniowy pozwala na przeprowadzanie symulacji około 20 replik na symulację w taki sposób aby obliczenia zostały wykonane w rozsądnym czasie. Obydwie metody zostały przetestowane na zestawie 15 małych białek (12 - 97 reszt aminokwasowych) o różnej sekwencji i strukturze trzeciorzędowej. Algorytm MREHMC był testowany w wielu różnych warunkach w celu dobrania optymalnych wartości specyficznych dla HMC parametrów. Przeprowadziłem testy akceptowalności testu Metropolisa wprowadzonego przez hybrydowy algorytm, zdefiniowałem najlepszą strukturę startową dla symulacji (rozwinęta lub natywne), oszacowałem optymalną wartość parametru L oraz długości symulacji jaka jest potrzebna do jej uzbieźnienia oraz porównałem wydajność nowej metody z algorytmem MREMD. Przeprowadziłem szczegółową analizę i porównanie obydwu metod co zawarłem w swojej pracy doktorskiej. Podsumowując wydajność metody MREHMC była lepsza od starszego algorytmu w odniesieniu do wartości C^α RMSD w stosunku do struktury natywnej dla najlepszych klastrów z symulacji. Dodatkowo udowodniłem wyższość mojej

metody w przypadku symulacji białka 1CLB dla którego MREMD nie było w stanie zlokalizować minimum globalnego w przestrzeni konformacyjnej, natomiast metoda hybrydowa była zdolna je wyszukać.

Symulacje przeprowadzone na superkomputerze TRYTON znajdującego się w Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej w Gdańsku pozwoliły na przeprowadzenie symulacji około 200 replik na białko na symulację w rozsądnym czasie (24 - 48 godzin). Dzięki takiemu podejściu możliwe było ocenienie różnic pomiędzy algorytmami MREMD i MREHMC w przypadku dużo większych i dużo bardziej kosztownych obliczeniowo symulacji. Nie tylko ilość replik (oraz temperatur) została zwiększona ale także całkowity czas symulacji został podwojony. Kolejnym powodem dla tych dodatkowych testów było przetestowanie nowej wersji parametryzacji pola siłowego UNRES - parametryzacji 'MAXLIK' - oraz porównanie jej wydajności względem starej wersji parametryzacji ('1L2Y'). Symulacje na nowym superkomputerze zostały przeprowadzone na zestawie 10 małych białek (12 - 95 reszt aminokwasowych) różniących się sekwencją i strukturą. Ostatecznie dwie metody symulacji (MREMD i MREHMC) w dwóch wersjach parametryzacji pola siłowego zostały porównane (w efekcie dało to cztery różne typy symulacji). Wydajność nowej wersji parametryzacji okazała się być dużo lepsza niż w przypadku starej wersji, natomiast różnice pomiędzy MREMD i MREHMC zostały zmniejszone (z tego względu iż przy tak dużych symulacjach metoda MREMD była w stanie bardziej efektywnie niż dotychczas przeszukiwać przestrzeń konformacyjną białek).

Drugą częścią mojego projektu było znalezienie sposobu na odzyskiwanie informacji kinetycznych z symulacji MREMD/MREHMC w polu siłowym UNRES. W tym celu wykorzystałem oryginalny algorytm `g_kinetics` z pakietu Gromacs. Pliki wyjściowe z symulacji w polu UNRES musiały być odpowiednio zmodyfikowane na potrzeby programu `g_kinetics`. Dla testów wydajności algorytmu zostały wykorzystane obie wersje parametryzacji pola siłowego, oraz trzy typy symulacji: MD, MREMD i MREHMC. Wydajność trzech metod została oceniona pod kątem szybkości zwijania białka. Metoda `g_kinetics` pozwoliła na uzyskanie informacji dla których została zaprojektowana co pozwoliło mi na wykonanie porównania metod na podstawie czasu zwijania białek. Otrzymane wyniki pozwoliły również na zbadanie zależności czasu zwijania białka od jego wielkości. Dodatkowo wyniki pozwoliły na porównanie czasu zwijania białka w symulacjach w polu siłowym UNRES z wynikami eksperymentalnymi i oszacowanie różnicy

pomiędzy czasem zwijania białka w symulacji oraz prawdziwym czasem zwijania białka.

Dodatkowo zastosowałem symulacje zwijania białek w celu uzyskania struktur kilku bakteriocyn. W tej części projektu próbowałem odpowiedzieć na pytanie związane z pewnym szczepem bakterii, który był wrażliwy na produkowaną przez niego bakteriocynę. Symulacje zostały przeprowadzone w celu uzyskania struktur dla białek, które nie posiadają jeszcze struktur natywnych wyznaczonych eksperymentalnie i wyszczególnienie różnic w ich strukturach. Oprócz tego wykonałem analizę genomów blisko spokrewnionych szczepów w poszukiwaniu różnic w ich kodzie genetycznym związanych z brakiem oporności w badanym szczepie.