

Tomasz Makarewicz

Promotor: dr hab. Rajmund Kaźmierkiewicz, prof. UG
Pracownia Symulacji Układów Biomolekularnych

Tworzenie i zastosowanie oprogramowania z graficznym interfejsem użytkownika do przeprowadzania, interpretacji i wizualizacji symulacji dynamiki molekularnej

Creation and application of software, with graphical user interface, that allows users to carry out, analyze and visualize molecular dynamics simulations.

Streszczenie

1. Streszczenie

Struktury biomolekuł ustalone przy pomocy krystalografii rentgenowskiej, magnetycznego rezonansu jądrowego lub kriomikroskopii elektronowej, są zapisywane w formie cyfrowej i przechowywane w archiwum np. w bazie struktur białek (ang. Protein Data Bank)[1]. Jednym ze standardowych formatów plików zawierających informację o strukturze przestrzennej molekuł jest PDB. Struktury zapisane w tym formacie mogą być bez problemu oglądane przy użyciu większości oprogramowania przeznaczonego do wizualizacji molekuł. Jednym z takich programów jest PyMOL[2]. Jest to popularna aplikacja napisana w języku programowania Python, która w prosty sposób umożliwia wizualizację i analizę struktury biomolekuł. Pomimo łatwej obsługi PyMOL daje badaczowi wiele zaawansowanych możliwości analizy badanych struktur. Ponadto udostępnia interfejs programistyczny, dzięki któremu można rozszerzać możliwości programu przy pomocy przygotowanych dla niego wtyczek. Jednakże sama analiza statycznej struktury zapisanej w formacie PDB jest niewystarczająca do uzyskania pełnej wiedzy o biomolekule. W układzie biologicznym cząsteczki te wykazują zachowanie dynamiczne, co ma kluczowe znaczenie dla ich funkcji biologicznej. Jednym ze sposobów uzyskania wiedzy o zmianach konformacji cząsteczki w czasie, jest poddanie statycznej struktury symulacji dynamiki molekularnej. Istnieje kilka zestawów oprogramowania umożliwiających przeprowadzenie takich obliczeń. Do najpopularniejszych można zaliczyć AMBER[3], CHARMM[4] i GROMACS[5]. Zwłaszcza ten ostatni może być uznany za dobry wybór dla naukowca, gdyż jego otwarty charakter zapewnia Nielimitowany dostęp do tego oprogramowania, a także swobodną jego modyfikację i jego używanie nie wiąże się z żadnymi kosztami. Jednakże, w przeciwieństwie do PyMOLa, oprogramowanie do przeprowadzania symulacji dynamiki molekularnej nie posiada, na ogół, graficznego interfejsu użytkownika, a złożoność charakteru obliczeń przekłada się na złożoność użytkowania aplikacji. Jest to uzasadnione głównie tym, iż długotrwałe obliczenia często wymuszają stosowanie wyspecjalizowanych komputerów, do których zdalny dostęp i tak nie umożliwia korzystania z programów z interfejsem graficznym. Niestety efektem ubocznym takiego podejścia jest utrudnione rozpoczęcie korzystania z oprogramowania do symulacji dynamiki molekularnej i wymaga ono przejścia przez naukowca wysokiego progu wiedzy i umiejętności. Ponadto dzisiejsze komputery osobiste mają już wystarczające możliwości obliczeniowe, żeby przeprowadzać prostszą symulację dynamiki molekularnej wprost na nich bez potrzeby korzystania ze specjalnych wyznaczonych do tego celu maszyn.

Celem niniejszej rozprawy jest zaproponowanie nowego programu działającego jako wtyczka do PyMOLa i umożliwiającego przeprowadzenie symulacji dynamiki molekularnej w prosty sposób za pośrednictwem interfejsu graficznego oraz wizualizację rezultatów w popularnym i łatwym w obsłudze PyMOLu. Napisany przeze mnie program nazwałem Dynamics PyMOL Plugin[6].

Zaproponowany program napisałem w języku programowania Python 2 i dzięki temu jest on całkowicie kompatybilny z bieżącą wersją PyMOLa. Tak samo, jak PyMOL korzysta z biblioteki Tk do wyświetlania grafiki, dzięki czemu nie wymusza żadnych dodatkowych zależności i już sama poprawna instalacja PyMOLa umożliwia instalację i skorzystanie z napisanego przeze mnie programu. Podstawowa zasada działania programu Dynamics PyMOL Plugin polega na tym, że pozwala on wybrać biomolekułę załadowaną w PyMOLu, następnie, w zaprojektowanym przeze mnie oknie graficznym, ustawić dla niej

opcje do przeprowadzenia symulacji dynamiki molekularnej, a następnie uruchomić tę procedurę i wyświetlić rezultaty z powrotem w PyMOLu. Do przeprowadzenia obliczeń symulacji dynamiki molekularnej moja aplikacja wykorzystuje program GROMACS. W związku z tym także GROMACS musi być zainstalowany w systemie, co nie powinno być wyzwaniem, gdyż jest on dostępny standardowo w repozytoriach wielu popularnych dystrybucji Linuksa. Napisany przeze mnie program Dynamics PyMOL Plugin jest kompatybilny z GROMACSem w wersjach 4.x, 5.x oraz 2016.x. Sposób działania wtyczki zobrazowałem na poniższym schemacie 1.

Zgodnie ze schematem 1 przepływ komunikacji z wykorzystaniem Dynamics PyMOL Plugin wygląda w sposób następujący:

1. Po uruchomieniu PyMOLa i wczytaniu pliku zawierającego strukturę badanej molekuly należy uruchomić wtyczkę z paska menu.
2. W oknie, które się pojawi (ilustracja 1), należy ustawić wszelkie potrzebne opcje potrzebne do przeprowadzenia symulacji dynamiki molekularnej. A następnie kliknąć przycisk OK.
3. Wtyczka skorzysta z narzędzi GROMACSa `pdb2gmx` lub `x2top`, żeby przekształcić załadowany w PyMOLu plik PDB do formatu trajektorii i topologii, który może zostać wykorzystany do dalszych obliczeń.
4. Następnie `editconf` i `gmx solvate` zostaną wykorzystane, żeby umieścić modelowaną cząsteczkę w rozpuszczalniku, a `genion` pozwoli na dodanie określonych jonów do środowiska symulacji.
5. Narzędzie `genrestr` doda ograniczenia w swobodzie ruchu elementów badanej molekuly wybranych przez badacza.
6. Faktyczne obliczenia dynamiki molekularnej zostaną przeprowadzone przez narzędzia `grompp` i `mdrun`.
7. Finalny wynik jest przekształcany z powrotem do formatu PDB przez `trjconv` i wyświetlony w PyMOLu.
8. Jeżeli program ProDy jest zainstalowany, to umożliwi on dalsze pogłębienie analizy wyników.

Pomimo tego, że program jest łatwy w obsłudze dzięki graficznemu interfejsowi użytkownika i domyślnej konfiguracji, która na ogół nie wymaga wielu modyfikacji w celu przeprowadzenia symulacji dynamiki molekularnej, to jednak oferuje on duże możliwości zmiany domyślnych ustawień GROMACSa. Dzięki temu staje się możliwe przeprowadzenie obliczeń w bardzo specyficznych warunkach ustawionych przez naukowca. Już z głównego okna wtyczki, można wybrać najważniejsze opcje – którą biomolekulę należy poddać obliczeniom, jaką jej część oraz przy użyciu którego pola siłowego. Wszelkie dostępne możliwości są bezpośrednio szczytywane z programów PyMOL oraz GROMACS, a interfejs użytkownika automatycznie dostosowuje się do tego, co owe aplikacje w danej wersji umożliwiają. Ponadto wtyczka oferuje szerokie możliwości dostosowania modelu rozpuszczalnika dla danej symulacji. Wybór zaczyna się od ustalenia, czy należy zastosować ciągły model rozpuszczalnika, czy raczej w formie cząsteczkowej. Ten pierwszy jest mniej precyzyjny, ale dużo oszczędniejszy obliczeniowo i łatwiejszy do prawidłowego skonfigurowania (np. zawsze pozwala uniknąć warunków brzegowych). Obecnie GROMACS pozwala na zastosowanie metod Still[7], HCT (Hawkins-Cramer-Truhlar)[8] oraz OBC (Onufriev-Bashford-Case)[9]. Wszystkie one są zaimplementowane we wtyczce. W przypadku modelu rozpuszczalnika w formie cząsteczkowej można określić model cząsteczki wody. Zakres dostępnych do wyboru modeli rozpuszczalnika zależy od wyboru empirycznego pola sił, a wtyczka automatycznie przedstawia dostępne modele cząsteczek wody dla zaznaczonego pola siłowego, np. dla pola AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003) są to: TIP3P, TIP4P, TIP4P-Ew, TIP5P, SPC i SPC/E, a dla pola GROMOS96 54a7 force field (Eur. Biophys. J. (2011), 40,, 843-856, DOI: 10.1007/s00249-011-0700-9) są to: SPC i SPC/E. Przykładowy sposób wyboru modelu wody przedstawia ilustracja 2.

Napisany przeze mnie program umożliwia następnie zmodyfikowanie parametrów pudełka z wodą, w którym odbędzie się symulacja. W tym miejscu można ustawić jego kształt, rozmiar oraz gęstość rozpuszczalnika. Dodatkowo ustalić można, czy w cząsteczkach wody będzie „zwykły” wodór, czy może deuter lub jeszcze cięższy izotop tego pierwiastka. W przypadku cząsteczkowego modelu rozpuszczalnika ustalić też można, czy ładunek elektryczny układu ma być zneutralizowany poprzez dodanie do rozpuszczalnika jonów oraz jakie ma być ich stężenie. W celu dalszego, bardziej szczegółowego, ustawienia parametrów symulacji dynamiki molekularnej GROMACS korzysta z tekstowych plików konfiguracyjnych w formacie .mdp. Napisana przeze mnie wtyczka umożliwia graficzną edycję parametrów w domyślnie nazwanych plikach dla: minimalizacji energii (em.mdp), symulacji dynamiki molekularnej samego rozpuszczalnika (pr.mdp) oraz właściwej symulacji dynamiki molekularnej (md.mdp). Możliwości edycji parametrów w pliku md.mdp przedstawiam na ilustracji 3.

Jeżeli naukowiec przygotował te pliki wcześniej i chce je zastosować bezpośrednio (bez ich edycji), to wystarczy, że je skopiuje do katalogu z projektem, a zostaną one zastosowane zamiast domyślnych plików wtyczki. Napisany przeze mnie program umożliwia także ograniczenie ruchu fragmentów biomolekuły, co jest przydatne w momencie, gdy naukowca interesuje dynamika jedynie jej fragmentu, np. centrum aktywnego. Szczególnie przydatną tutaj możliwością jest zaznaczenie w PyMOLu fragmentu cząsteczki, którego ruch ma być ograniczony, a następnie wtyczka umożliwia poinformowanie GROMACSa, żeby dokonał ograniczenia ruchu dla właśnie zaznaczonych atomów. Moim zdaniem warto też wspomnieć o kolejnej dużej funkcji programu. Jeżeli biblioteka ProDy jest zainstalowana w systemie, to jest możliwe jej wykorzystanie w celu przekształcenia animacji dynamiki molekularnej na statyczną migawkę cząsteczki z zaznaczonymi wektorami przesunięcia. Przewaga takiej metody wizualizacji jest taka, że daną symulację dynamiki molekularnej można wydrukować jako pojedynczą ilustrację i zastosować np. w publikacjach naukowych. Warto tu dodać, że Dynamics PyMOL Plugin jest obecnie jedynym sposobem na wizualizację rezultatów obliczeń biblioteki ProDy w PyMOLu. Przykładowy wynik zaprezentowałem na ilustracji 4.

Ponadto biblioteka ProDy umożliwia wizualizowanie zarówno za pomocą wykresu, jak i bezpośrednio w oknie PyMOLa tak zwanych map kontaktów i korelacji krzyżowych. Przykładową mapę korelacji krzyżowych przedstawia wykres 1.

Symulacja dynamiki molekularnej nie musi być przeprowadzana w całości na jednym komputerze. Napisany przeze mnie program Dynamics PyMOL Plugin umożliwia wygodne zapisywanie i wczytywanie rezultatów pracy, a następnie dalsze kontynuowanie pracy już w innym środowisku. Ponadto, gdyby symulacja się nie powiodła, to wtyczka daje prosty dostęp do przeglądania logów GROMACSa, co umożliwia wykrycie problemu, korekcję ustawień symulacji, a następnie ponowne jej przeprowadzenie, bądź wznowienie od przerwanej chwili.

Napisany przeze mnie program - Dynamics PyMOL Plugin - jest rozwijany na zasadach wolnego oprogramowania i udostępniłem go na licencji GPL-3. Projekt korzysta z platformy GitHub i znajduje się pod adresem: <https://github.com/tomaszmaekarewicz/Dynamics>. Taki model tworzenia oprogramowania powoduje, że tak długo, jak są osoby nim zainteresowane, jest on dalej rozwijany, a korzystać może z niego każdy. Ponadto należy dodać, że zarówno PyMOL, jak i GROMACS także stanowią otwarte oprogramowanie, więc naukowiec dysponuje dostępem do pełnego kodu źródłowego wszystkich narzędzi. W przyszłości planuję dalej zajmować się projektem i zaimplementować między innymi: ograniczenia ruchu oparte o dystanse między atomami (obecnie zaimplementowane są oparte o pozycję atomów), potencjał Martini[10], symulacje dokowania molekularnego, działanie wtyczki na zasadzie klient → serwer (w takim przypadku ustawienia użytkownika są wysyłane na serwer, który dokonuje obliczeń, a następnie odsyła użytkownikowi rezultaty). Jednak tak naprawdę projekt będzie się rozwijał, tak jak użytkownicy sobie tego życzą. Projekt przyjmuje zgłoszenia błędów, życzeń oraz nawet kodu w sformalizowanym systemie śledzenia problemów. I tak np. dzięki donacji kodu z dnia 11.04.2017 od dr Ajit Bikram Datta z Department of Biochemistry, Bose Institute, Indie udało się zaimplementować obsługę

modułu GROMACSa genion oraz opcji do ustawiania masy wodoru, a dzięki donacji kodu z dnia 31.05.2017 od Manish Sud z projektu MayaChemTools udało się zaimplementować obsługę systemu Microsoft Windows. W związku z tym planuję skupić się głównie na zarządzaniu projektem, rozwiązywaniem problemów użytkowników oraz kontaktem z programistami, którzy pragną zaimplementować kolejne funkcje. Szczególnie istotnym kamieniem milowym na tym polu będzie zaimplementowanie systemu ciągłej integracji.

W ciągu ostatnich sześciu lat mój program Dynamics PyMOL Plugin wyrósł z „niczego”, do pokaźnego projektu używanego na całym świecie, którego rozwój raportowany jest w czasopiśmie naukowych. Do znanych użytkowników należy zaliczyć osoby z takich instytucji jak: Bowdoin College, KU Leuven, Harvard Medical School, NCBI. Projekt otrzymał bezpośrednie kontrybucje kodu od członków takich instytucji jak: Bose Institute, MayaChemTools. Obecnie projekt ma już prawie trzy tysiące linii kodu, które zostały do niego włączone przez ponad dwieście czterdzieści zaakceptowane zmiany. Oprogramowanie obecnie wspiera systemy Linux, Windows oraz macOS i wiele różnych wersji GROMACSa. Wtyczka dostępna jest w głównych repozytoriach dystrybucji Linuksa Gentoo oraz repozytoriach ppa dla Ubuntu.

Uważam, że Dynamics PyMOL Plugin dobrze sprawdza się jako program, który obniża dla naukowca próg wejścia w techniki symulacji dynamiki molekularnej przez prosty w obsłudze interfejs graficzny, wstępnie skonfigurowane środowisko obliczeniowe oraz prosty dostęp do logów z działania GROMACSa. Jednocześnie program daje bardzo szerokie możliwości konfigurowania warunków symulacji dynamiki molekularnej i może być stosowany nawet przez naukowców biegłych w obsłudze takiego oprogramowania z linii poleceń i oczekujących dostępu do wielu opcji. Sprawdza się tu doskonale do przeprowadzenia symulacji niedużych układów biomolekuł, co nie wymaga olbrzymich możliwości obliczeniowych komputera oraz do wstępnego sprawdzenia, czy proponowane warunki obliczeń są prawidłowe, zanim zastosuje się je na właściwym serwerze obliczeniowym. Liczę na to, że projekt Dynamics PyMOL Plugin będzie nadal rozwijany, jego użyteczność będzie rosła, a w konsekwencji adaptacja wśród naukowców i programistów. Mam nadzieję, że program skutecznie ułatwi naukowcom rozpoczęcie pracy w dziedzinie symulacji dynamiki molekularnej, a dzięki temu przyczyni się do wzrostu zainteresowania tą dziedziną nauki.

Efektom mojej pracy, poza samym programem, są dwie publikacje naukowe.