



Warszawa, dn. 6.02.2018

Prof. dr hab. Sławomir Filipek,  
Wydział Chemii, Centrum Nauk Biologiczno-Chemicznych,  
Uniwersytet Warszawski,  
ul. Pasteura 1, 02-093 Warszawa  
Tel. 22-55-26405,  
E-mail: sfilipek@chem.uw.edu.pl

**Recenzja**  
**rozprawy doktorskiej mgr. Tomasza Makarewicza, pt.**

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

Rozprawa nie ma charakteru klasycznej rozprawy doktorskiej z podziałem na Wstęp, Metody, Wyniki i Dyskusję, a jest odzwierciedleniem otrzymanych i opublikowanych prac, dlatego dysertacja ta jest podzielona na 3 główne rozdziały: Streszczenie, pierwsza publikacja, i odpowiednio druga publikacja. Rozprawa jest napisana po polsku oraz po angielsku i liczy 35 stron razem z publikacjami i oświadczeniami. Obejmuje 10 odnośników literaturowych w części wstępnej (Streszczenie/Abstract), natomiast w obu publikacjach odpowiednio 35 i 50 odnośników. Prace Doktoranta zostały opublikowane w *Journal of Chemical Information and Modeling* w 2013 roku, oraz *Journal of Molecular Modeling* w 2016 roku. Obie prace są dwuautorskie, tylko Doktorant wraz ze swoim promotorem, co wskazuje na duży wkład Doktoranta w powstanie tych publikacji. Pierwsza z tych prac uzyskała już 17 cytowań, co pokazuje duży potencjał wytworzonego oprogramowania (w formie tzw. wtyczki do programu PyMOL) do badania dynamiki i stabilności białek. Niestety, żadne z tych cytowań (oprócz cytowania pochodzącego z drugiej publikacji Doktoranta) nie pochodzi z własnej pracowni. W takim razie sprawdza się powiedzenie „cudze chwalicie, swego nie znacie”. Publikacje, w których wykorzystano program Doktoranta nie są niestety przedstawione w wykazie publikacji, a tylko wymienione są



instytucje naukowe, które z niego korzystają. Publikacje na których jest oparty doktorat także nie są wymienione w Spisie publikacji Streszczenia. Jest to istotny mankament szczególnie, że wydruk pierwszej publikacji nie zawiera pełnego odnośnika a tylko litery XXX zamiast odpowiednich numerów.

W Streszczeniu Doktorant przedstawia najpierw programy używane do dynamiki molekularnej białek, bowiem napisany przez niego program dotyczy przeprowadzania symulacji dynamiki molekularnej oraz wizualizacji uzyskanych wyników w PyMOLu. Tymi wymienionymi programami są AMBER, CHARMM i GROMACS. Niestety, zabrakło w tym zestawieniu równie popularnego wśród naukowców programu NAMD, używanego także do symulacji dynamiki molekularnej białek. Program NAMD podobnie jak programy wymienione poprzednio nie posiada interfejsu graficznego ale istnieje program VMD, tych samych autorów co NAMD, do wizualizacji wyników uzyskanych za pomocą tego programu. Jak się wydaje, program NAMD jest znacznie bardziej popularny niż CHARMM. Poza tym, program NAMD, podobnie jak omawiany program GROMACS, jest także darmowy, czego nie można powiedzieć o dwu pozostałych wymienionych w pracy programach: AMBER i CHARMM. Szkoda także, że nie zostały wymienione inne programy do symulacji dynamiki molekularnej, np. coraz bardziej popularny program YASARA, który ma jednocześnie możliwość wizualizacji oraz przeprowadzania symulacji dynamiki molekularnej, i w którym można na bieżąco interweniować w dynamikę wymuszając ruch danej części cząsteczki co bywa przydatne w przypadku zmiany jej konformacji.

Takie zestawienie i omówienie programów zrobione na potrzeby doktoratu mogłoby być podstawą do napisania pracy przeglądowej na ten temat. Brakuje takich prac porównujących tego typu programy pod różnym kątem.

Na str. 2 Streszczenia napisano „PyMOL ... aplikacja napisana w języku programowania Python”. Nie jest to poprawne bowiem PyMOL jest napisany w językach C, C++ i Python. Głównym językiem jest jednak Python, stąd nazwa tego programu PyMOL. Na str. 3: „program napisałem w języku programowania Python 2 i dzięki temu jest on całkowicie kompatybilny z bieżącą wersją PyMOLa”. Mam dwie uwagi do tego zdania: (1) obecnie jest dostępna wersja Python 3 wobec tego czy wciąż jest zachowana kompatybilność? (2) nie podano nr wersji PyMOLa nazywanej w tym zdaniu *bieżąca*. Jest to istotne, bo nowa wersja PyMOLa może już nie być kompatybilna. Napisany przez Doktoranta program jest umieszczony na stronie internetowej GitHub, gdzie znajduje się także szczegółowy opis



programu i jego wszystkich opcji. Program nosi nazwę DYNAMICS PyMOL Plugin i jest rozwijany na zasadach wolnego oprogramowania - jest udostępniony na licencji GPL3 (*general public license*). Strona GitHub, grupująca 20 milionów osób zajmujących się rozwojem oprogramowania, stanowi wygodną i potężną platformę do wspólnego rozwoju oprogramowania, głównie darmowego typu OpenSource. Z kolei PyMOL jest to program używany na całym świecie, którego rozwój poprzez różne wtyczki, m.in. z wykorzystaniem GitHub, jest regularnie raportowany w czasopiśmie naukowych.

Pierwsza publikacja Doktoranta omawia przepływ danych pomiędzy PyMOLEM a GROMACSEM z wykorzystaniem napisanego przez Doktoranta programu oraz opisuje poszczególne opcje wykorzystywane w tym programie/wtyczce. Ta wtyczka pozwala znacznie zwiększyć możliwości PyMOLA poprzez wykonanie symulacji dynamiki molekularnej, co pozwala wyjść poza statyczny obraz białka, oraz na wizualizację uzyskanej trajektorii. Wtyczka używa języka Python oraz języka Tk dla grafiki. Wtyczka ta pozwala w łatwy sposób zamienić plik PDB (*Protein Data Bank*) na wymagany przez GROMACSA. Potem inne skrypty z GROMACSA tworzą pudło periodyczne i wypełniają je wodą. Można użyć pudła periodycznego o różnym kształcie, nie tylko prostopadłościennego, ale np. ośmiościanu czy dwunastościanu, oraz kilku modeli wody: TIP3P, TIP4P, SPC i innych. Można nakładać więzy na badany układ oraz wykonywać minimalizację energii układu jako wstępu do symulacji dynamiki molekularnej aby obniżyć energię układu wynikającą ze zbyt bliskich kontaktów między atomami. Po wystartowaniu symulacji można śledzić jej postęp oraz wykonywać restarty aby przedłużyć symulację.

Druga publikacja Doktoranta przedstawia istotne rozszerzenie wtyczki na symulacje w ciągłym modelu rozpuszczalnika (typu uogólniony Born GB) zamiast w formie cząsteczkowej, co pozwala na szybsze choć mniej precyzyjne obliczenia symulacyjne. Doktorant pisze, że przy ciągłym modelu rozpuszczalnika pudło periodyczne nie jest konieczne. Są jednak takie układy, np. kilka lub kilkanaście oddziałujących ze sobą peptydów, dla których pudło periodyczne zapewni pozostawanie w jednym miejscu i nie rozbieganie się do nieskończoności. Drugim istotnym rozszerzeniem jest wykorzystanie innego programu typu Open Source ProDy (od słów *Protein Dynamics*) do wizualizacji ruchów całego białka. Program ProDy wykorzystuje metody typu NMA (*Normal Mode Analysis*) i PCA (*Principal Component Analysis*). Aby najszybciej zobrazować jakie są możliwe ruchy białka wykorzystano metodę ANM (*Anisotropic Network Model*) z tego



programu, która bazuje na gruboziarnistym przybliżeniu struktury białka, gdzie każdy z aminokwasów jest reprezentowany za pomocą jednej kulki. Pozwala to w bardzo szybki sposób policzyć i zobrazować najwolniejsze, a więc angażujące całą cząsteczkę ruchy białka. Dodatkowo, wtyczka ta pozwala na wizualizację mapy kontaktów oraz macierzy cross-korelacji (czyli korelacji wzajemnych ruchów aminokwasów). Drgania atomów można także zamienić na wektory w statycznym obrazie białka, pokazujące kierunek największego ruchu danego atomu. Szkoda, że Doktorat nie porównał swojego rozszerzenia wtyczki z już istniejącą wtyczką PyANM, która także wykonuje analizę drgań normalnych białka i przedstawia wyniki w formie filmiku lub wektorów ruchu.

W Streszczeniu opisującym drugą publikację, na str. 6 jest napisane: „ciągły model rozpuszczalnika ... zawsze pozwala uniknąć warunków brzegowych” – jest to mało precyzyjne stwierdzenie. Zapewne chodziło o brak potrzeby stosowania pudła periodycznego. Także na str. 6 jest mowa o „pudełku z wodą, w którym odbędzie się symulacja” – jest to wyrażenie żargonowe. Powinno być: pudła periodycznego, w którym odbędzie się symulacja w roztworze wodnym. Na str. 7: „daną symulację dynamiki molekularnej można wydrukować jako pojedynczą ilustrację” – zapewne chodziło o drgania normalne cząsteczki i pojedynczy mod tych drgań jak na to wskazuje rys. 4. Z kolei na wykresie 1 przedstawiono tzw. mapę korelacji krzyżowych ale nie wyjaśniono (ani w tekście pracy ani pod wykresem) na czym ta mapa polega ani do czego te korelacje się odnoszą. Doktorant planuje także wykonanie innych rozszerzeń swojej wtyczki na m.in. symulacje gruboziarniste w polu siłowym MARTINI (także w programie GROMACS), oraz dokowanie molekularne – zapewne chodzi o ligandy małowcząsteczkowe, chociaż tutaj już kilka wtyczek jest zrobionych np. AUTODOCK/VINA. Na str. 9 napisano „zaimplementowanie systemu ciągłej integracji” – natomiast nie wyjaśniono na czym ten system miałby polegać.

W przedstawionym Streszczeniu brakuje także opisu, nawet skromnego, np. w formie tabelki, innych wtyczek, które zostały skonstruowane aby także rozszerzyć możliwości PyMOLa, a jest ich ok. 40. Dotyczą różnych kategorii: do wizualizacji i analizy białek, do modelowania układów ligand-białko, do przeszukiwania baz danych ligandów (*virtual screening*), oraz do dynamiki molekularnej. Jest nawet baza danych na wzór Wikipedii, PyMOLWiki gdzie jest przedstawiony program PyMOL oraz wtyczki rozszerzające jego możliwości. Niestety, wtyczka Doktoranta DYNAMICS nie jest tam opisana (może warto to zrobić?) oraz zabrakło przedstawienia linku do tej bazy (<https://pymolwiki.org>) w tej



dysertacji. Z mojej strony, ponieważ zajmujemy się białkami błonowymi, brakuje możliwości konstruowania dwuwarstwy lipidowej i umieszczenia tam białka. Tego na razie nie robi żadna z istniejących wtyczek do PyMOLa, chociaż odpowiednie skrypty w programie GROMACS istnieją.

Wtyczka DYNAMICS tak bardzo się rozrasta, że można by rozważyć napisanie osobnego programu, jak np. zrobiono w przypadku programu LiGRO (napisany także w Pythonie, opublikowany w 2017 roku), do parametryzacji małych cząsteczek i symulacji dynamiki molekularnej układów białko-ligand w programie AMBER. Program LiGRO używa okna z PyMOLa do wyświetlania wyników. Dobrze było rozważyć zalety i wady takiego oddzielnego programu.

Biorąc pod uwagę dwie prace opublikowane przez Doktoranta oraz szerokie wykorzystanie napisanej przez niego wtyczki do PyMOLa przez innych naukowców, pomimo braku własnych prac wykorzystujących napisany program do badań naukowych, Doktorant wykazał dużą znajomość zagadnień naukowych w których się porusza, aby stworzyć program, który w sposób jasny i przejrzysty umożliwia łatwe i bezpieczne (tj. bez błędów związanych z ręczną modyfikacją skryptów) wykonywanie skomplikowanych symulacji dynamiki molekularnej białek. Wyżej wymienione w recenzji niedociągnięcia nie przesłaniają znaczącej wartości napisanego programu oraz bardzo dobrego przygotowania programistycznego Doktoranta. Przedstawiona mi do oceny rozprawa doktorska spełnia zatem warunki określone w art. 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 z późn. zm.), dlatego przedkładam wniosek do Rady Naukowej Międzyuczelnianego Wydziału Biotechnologii Uniwersytetu Gdańskiego i Gdańskiego Uniwersytetu Medycznego o dopuszczenie mgr. Tomasza Makarewicza do dalszych etapów przewodu doktorskiego.

Sławomir Filipek