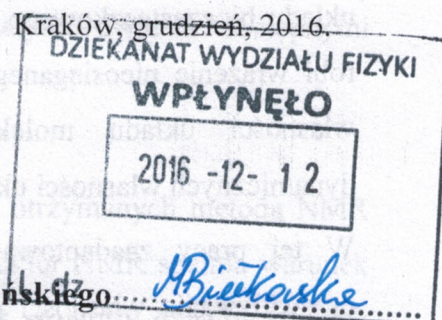




Recenzja

pracy doktorskiej mgra Macieja Dziubińskiego

pt.: „Development of selected mesoscopic physical models with the aid of machine learning methods and their applications in studies of molecular systems”

Praca doktorska mgr Macieja Dziubińskiego została wykonana pod kierunkiem prof. Bogdana Lesynga i dr Pawła Daniluka w Zakładzie Biofizyki na Wydziale Fizyki Uniwersytetu Warszawskiego. Jest to praca z pogranicza biofizyki teoretycznej i bioinformatyki i dotyczy takiego dostosowania metod uczenia maszynowego bez nadzoru by z ich pomocą skuteczniej prowadzić analizy strukturalne i dynamiczne białek. Rozprawa oparta jest na dwóch opublikowanych artykułach, w obu mgr Dziubiński jest pierwszym autorem, w drugiej jest też autorem korespondencyjnym. Napisana jest poprawnie w języku angielskim, zawiera streszczenie w języku angielskim, ale nie w polskim i składa się z trzech głównych części: rozdziału 1, stanowiącego wstęp, w którym doktorant omawia istotne elementy uczenia maszynowego oraz symulacji dynamiki molekularnej (MD), ale niestety niewystarczająco uwypukla biomolekularne aspekty zagadnień omawianych dalej; rozdziału 2, opartego na pierwszej publikacji i opisującego opracowaną przez autorów metodologię identyfikacji dynamicznych domen (sztywnych obszarów) białka, nazwaną *Residue Contacts analysis* (ResiCon); oraz rozdziału 3, opartego na drugiej publikacji i opisującego nowe podejście do analizy wewnętrznej mechaniki cząsteczki, tj. tendencji do zmian konformacyjnych w cząsteczce. Na końcu zamieszczone są dwa załączniki stanowiące *Supplementary Material* do artykułów z rozdziału 2 i 3 oraz spis literatury.

Rozprawa napisana jest starannie, ale w stylu typowym dla teoretyka, co nie ułatwia niespecjaliście czytania i szybkiego zrozumienia. Trudno byłoby ją wykorzystać jako materiał dydaktyczny dla studentów z niższych lat, czy materiał wprowadzający dla ewentualnego kontynuatora badań. Ilustracje w pracy stanowią znakomite dopełnienie tekstu i są bardzo starannie wykonane.

Cele pracy:

Autor podzielił cele pracy na (1) cel podstawowy (*primary aim*) i (2) cel ogólny (*overall aim*). Celem podstawowym autora było opracowanie metod umożliwiających wydobycie istotnych informacji z wielowymiarowych danych opisujących strukturalne i dynamiczne własności

układu biocząsteczkowego. Cel ogólny, opisany na stronie 14, nie jest do końca jasny, ale robi wrażenie nieosiąganego, gdyż tym celem jest uproszczenie złożonych dynamicznych własności układu molekularnego. Prawdopodobnie chodzi o opisanie złożonych dynamicznych własności układu w sposób uproszczony.

W tej pracy zaadaptowano i użyto metodę uczenia maszynowego bez nadzoru do przewidywania własności dynamicznych biocząsteczek i wskazania mechanizmów za nie odpowiedzialnych w oparciu o dane z pomiarów doświadczalnych lub też wygenerowane w symulacjach MD.

Metoda ResiCon

Celem badań opisanych w rozdziale drugim, opartym na publikacji: Dziubinski M., Daniluk P., Lesyng B. *ResiCon: a method for the identification of dynamic domains, hinges and interfacial regions in proteins*, Bioinformatics 32, 25–34, 2016, jest opracowanie metody ResiCon do identyfikacji dynamicznych domen w białku. Metoda ocenia, w oparciu o rozważania geometryczne, wielkość oddziaływań między resztami aminokwasowymi w białku dla zbioru wielu struktur tego samego białka, otrzymanych metodą MNR lub symulacji MD i tworzy macierz kontaktów. Kontakty między resztami i ich wielkość, zachowane we wszystkich strukturach ze zbioru struktur, wskazują, że dana grupa reszt aminokwasowych tworzy domenę dynamiczną. Podział białka na domeny dynamiczne przeprowadzany jest z pomocą algorytmu klastrującego wykorzystującego wartości własne macierzy kontaktów (klastrowanie spektralne).

Wśród biochemików, biologów strukturalnych, farmakologów, bioinformatyków i biofizyków obliczeniowych jest ciągle zapotrzebowanie na nowe sposoby analizy struktury i dynamiki wewnętrznej makrocząsteczek, a przede wszystkim białek. Każde nowe opracowanie otwiera nowe możliwości poznawania ich własności. Opracowana w ramach tej części doktoratu metoda dedykowana jest rygorystycznej identyfikacji domen dynamicznych białka. Przez dynamiczną domenę rozumie się grupę reszt aminokwasowych, które podlegają uzgodnionym ruchom, w czasie których odległości między tymi resztami w zasadzie nie ulegają zmianom. Ruchy każdej z domen są słabo sprzężone z ruchami pozostałych. Ruchy domen w białku są często związane z jego funkcjami takimi jak kataliza, transport, sygnalizacja, odpowiedź immunologiczna i inne. Zidentyfikowanie dynamicznych domen oraz wskazanie miejsc zawiasowych i „powierzchni styku” umożliwia przewidywanie funkcji białka i wskazanie podstawowych mechanizmów za nie odpowiedzialnych. Metoda ResiCon niewątpliwie wzbudzi zainteresowanie i będzie wykorzystana innych badaczy, gdyż, jak pokazują testy

przeprowadzone przez doktoranta, w porównaniu z pozostałymi dwiema dostępnymi metodami jest skuteczniejsza.

Uwagi do rozdziału 2:

Metoda ResiCon została zastosowana m. in. do zbioru struktur otrzymanych metodą NMR (struktury NMR) i symulacji MD (struktury MD). Każda ze struktur NMR spełnia warunek więzów wyznaczonych z wyników pomiarów (nie są to jednak struktury dynamiczne, jak częściowo sugeruje doktorant). Nasuwa się więc pytanie:

czy macierzy kontaktów nie można zbudować na podstawie zbioru więzów? lub przynajmniej wykorzystać więzy w celach szacunkowych?

Symulacje MD generują wiele zbiorów konfiguracyjnych i umożliwiają prześledzenie zmian czasowych w strukturze cząsteczki. W opisanych analizach z trajektorii wybrano 200 reprezentatywnych struktur symulowanego białka. W przypadku struktur MD proteazy HIV-1, która jest stosunkowo niewielkim białkiem, do stworzenia macierzy kontaktów można by wykorzystać wszystkie wygenerowane struktury lub przynajmniej wyniki dla nich otrzymane mogłyby posłużyć do oceny wartości otrzymanych dla wybranego podzbioru struktur reprezentatywnych – czy w ocenie doktoranta wykorzystanie całej roboczej trajektorii miałoby wpływ na poprawę otrzymanych wyników? W programach symulacyjnych oblicza się oddziałujące pary i co ileś kroków czasowych tworzy „listę par”; te obliczenia można by praktycznie bez straty czasu wykorzystać do tworzenia macierzy kontaktów – czy tak?

Brakuje mi też ogólniejszej dyskusji nad związkiem między zidentyfikowanymi domenami dynamicznymi w badanych białkach a funkcjami biologicznymi tych białek.

Tryby molekularne

Rozdział trzeci bazuje na publikacji: Dziubinski M., Lesyng B. *Towards the identification of molecular cogs*, J. Comput. Chem. 37, 848–860, 2016. Celem tej części pracy jest identyfikacja w cząsteczce takich grup atomów, które są odpowiedzialne za przeprowadzenie w niej zmian konformacyjnych; te grupy nazwane zostały „trybami molekularnymi”. W celu identyfikacji trybów molekularnych opracowana została metoda analizująca kooperatywne oddziaływania mogące indukować przejścia konformacyjne. Przejścia zachodzą wzdłuż współrzędnej reakcji i odpowiada im zmiana energii swobodnej układu. Ponieważ w pracy proponuje się zupełnie nowe podejście, a w zasadzie opracowanie unikalnej analizy, testowanym układem jest mała cząsteczka, dla której człon entropowy energii swobodnej można zaniedbać. Tak więc problem ogranicza się do członu energetycznego energii swobodnej. Upraszczając zagadnienie, dla każdej pary atomów człon energetyczny jest wyrażony poprzez oddziaływania dwuciałowe między atomami; z niego wylicza się wagę

krawędzi w nieskierowanym grafie, którego wierzchołkami są te atomy; wagi dla poszczególnych krawędzi są miarą kooperatywności atomów. Grupy kooperujących atomów identyfikowano genetycznym algorytmem klastrującym.

Opracowaną metodologię doktorant sprawdził – z pozytywnym wynikiem – na 11-atomowej cząsteczce oraz na cząsteczkach pochodnych. Przeprowadzona analiza pokazała udział poszczególnych oddziaływań w przejściach konformacyjnych i wskazała, że największy przyczynek pochodzi od oddziaływań niewiążących.

Podjęcie badań nad trybami molekularnymi (stosując skrót używany przez doktoranta) uważam za niezwykle uzasadnione. Spontanicznie zachodzące zmiany konformacyjne w cząsteczce nierozzerwalnie związane są z bardzo elementarnymi ruchami termicznymi atomów tworzących cząsteczkę. Jednakże przejście między tymi elementarnymi ruchami a fluktuacjami strukturalnymi jest niejasne i trudne do formalnego opisania. Uważam, że doktorant podjął bardzo trudne wyzwanie, wymagające wiedzy z różnych dziedzin nauki oraz sprawności w posługiwaniu się nią a także narzędziami używanymi w tych dziedzinach. Realizując je, wykazał się nie tylko wiedzą, zrozumieniem i sprawnością, ale też samodzielnością. Jego praca dała obiecujące początki, i sędzę, że będzie kontynuowana i rozszerzona na większe układy wymagające bardziej złożonego opisu. Jednak nawet dla niewielkiej cząsteczki, aparat zaangażowany w opis spontanicznych zmian strukturalnych w cząsteczce jest trudny zarówno koncepcyjnie jak i algorytmicznie. W swojej pracy doktorskiej mgr Dziubiński skupił się przede wszystkim na stronie teoretycznej i algorytmicznej badanych zagadnień, co naturalnie wynika z problematyki badawczej grupy prof. Bogdana Lesynga, promotora rozprawy, oraz z wiedzy i zainteresowań doktoranta. Mnie bardziej interesuje, jaką nową/dodatkową wiedzę o biocząsteczkach możemy zdobyć stosując metody komputerowe, więc chciałam podsumować istotność metodyki opracowanej przez doktoranta w badaniach biocząsteczek.

A więc uważam, że wyniki osiągnięte metodą ResiCon są istotne dla bionauk, gdyż informacje o dynamice wewnątrzcząsteczkowej, takie jak zidentyfikowane domeny dynamiczne, odgrywają zasadniczą rolę w przewidywaniu funkcji białka, o ile jej nie znamy, czy w zrozumieniu sposobu jego funkcjonowania, jeśli znamy jego funkcje.

Analiza i opis procesu „globalnych” przejść konformacyjnych w cząsteczce, przedstawione w rozdziale 3, satysfakcjonują naszą naturalną ciekawość poznawczą i umożliwiają nowe spojrzenie na to bardzo powszechne zjawisko, jednakże nie jestem w stanie stwierdzić, czy mają bezpośrednie przełożenie praktyczne.

Ustawa o stopniach naukowych i tytule naukowym wymaga, aby rozprawa doktorska stanowiła oryginalne rozwiązanie problemu naukowego, wykazywała ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej oraz umiejętność samodzielnego prowadzenia pracy naukowej przez doktoranta. Mgr Maciej Dziubiński wykazał się wymaganymi umiejętnościami i zdolnościami w stopniu ponadstandardowym. Wnoszę więc do Rady Wydziału Fizyki Uniwersytetu Warszawskiego o dopuszczenie go do dalszych etapów przewodu doktorskiego.

M. Wasecki

Praca doktorska mgr Macieja Dziubińskiego została wykonana pod kierunkiem prof. Bogdana Lesyng i dr Pawła Janickiego w Zakładzie Biofizyki na Wydziale Fizyki Uniwersytetu Warszawskiego. Jest to praca z pogromienia biofizyki teoretycznej, bioinformatyki i dotyczy takiego zastosowania metod uczenia maszynowego bez nadzoru by z ich pomocą skutecznie przewidywać zmiany strukturalne i dynamiczne białek. Rozprawa została opublikowana w dwóch opublikowanych artykułach, w obu mgr Dziubiński jest pierwszym autorem, w drugiej jest też autorem korespondencyjnym. Napisana jest poprawnie w języku angielskim, zawiera streszczenie w języku angielskim, ale nie w polskim i składa się z trzech głównych części: rozdziału 1, zawierającego wstęp, w którym doktorant przedstawia istotne elementy uczenia maszynowego oraz symulacji dynamiki molekularnej (MD), ale nie jest to wystarczające wprowadzenie do tematu; rozdziału 2, opartego na pierwszej publikacji i opisującego opracowaną przez autora metodologię identyfikacji dynamicznych domien (sztywnych obszarów) białka, nazwaną *Residue Contact analysis* (ResCot); oraz rozdziału 3, opartego na drugiej publikacji i opisującego nowe podejście do analizy wewnętrznej mechaniki cząsteczki, tj. tendencji do zmian konformacyjnych w cząsteczce. Na koniec zamieszczone są dwa załączniki stanowiące *Supplementary material* do artykułów z rozdziału 3 i 4 oraz spis literatury.

Rozprawa napisana jest starannie, ale w stylu typowym dla teoretyka, co nie ułatwia specjalnie czytania i szybkiego zrozumienia. Też nie udało się wykorzystać jako materiału dydaktyczny dla studentów z dziedziny fizyki, czy chemii, czy biologii, czy biofizyki, co jest istotnym mankamentem tej pracy. Ilustracje w pracy stanowią ważną część tekstu i są bardzo starannie wykonane.

Cele pracy:

Autor rozdzielił cele pracy na (1) cel podstawowy (*primary aim*) i (2) cel ogólny (*overall aim*). Celem podstawowym autora było zaimplementowanie metody umożliwiającej wydobycie istotnych informacji z wielowymiarowych danych dotyczących strukturalnych i dynamicznych właściwości