RDITT-apl. 4.07.2023

UNIWERSYTET WARSZAWSKI | CeNT CENTRUM NOWYCH TECHNOLOGII

15 czerwca, 2023

Prof. dr hab. Joanna Trylska
e-mail: joanna@cent.uw.edu.pl
telefon (22) 55 43 600
http://bionano.cent.uw.edu.pl

Prof. dr hab. inż. Andrzej Polański
Przewodniczący Rady Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Ul. Akademicka 16
44-100 Gliwice

Recenzja rozprawy doktorskiej mgr inż. Magdaleny Ługowskiej

Rozprawa doktorska mgr inż. Magdaleny Ługowskiej o tytule w języku polskim *„Algorytmy analizy struktur białkowych i ligandów lekopodobnych w celu modelowania i symulacji czasu przebywania leku w miejscu docelowym"* (ang. *Algorithms for the analysis of molecular protein structures and drug-like ligands for modeling and simulation of residence time drug-molecular target)* została wykonana na Wydziale Automatyki, Elektroniki i Informatyki Politechniki Śląskiej. Promotorami tej rozprawy są prof. dr hab. inż. Marek Kimmel oraz dr inż. Marcin Pacholczyk (jako promotor pomocniczy).

Dla zachodzenia wielu procesów w komórkach konieczne jest, aby cząsteczki biorące w nich udział stworzyły odpowiedni kompleks bimolekularny. Często tworzone są kompleksy mniejszych ligandów jako substratów reakcji, z większymi białkami (enzymami), czy kwasami nukleinowymi. Oddziaływanie ligand-receptor rozpoczyna się zwykle od asocjacji komponentów w zatłoczonym środowisku komórki, gdzie ostatecznie dochodzi do stworzenia charakterystycznego i często unikalnego kompleksu. Utworzenie specyficznego kompleksu ligand-receptor można zablokować poprzez zaprojektowanie cząsteczki (inhibitora), która oddziałuje z większym powinowactwem z receptorem niż naturalny ligand. Na takiej podstawie działa wiele leków celujących w białka. Jednak nie tylko powinowactwo, czyli siła oddziaływania inhibitor-receptor ma znaczenie, ale także czas przebywania inhibitora w miejscu wiążącym receptora. Temat badawczy poruszony w rozprawie dotyczył właśnie obliczania czasów przebywania (rezydencji) różnych ligandów w ich miejscach wiążących w receptorach.

Zwykle oddziaływania ligand-receptor określa się poprzez stałą równowagową dysocjacji (K_d) lub asocjacji (K_a). Stała K_d jest natomiast ilorazem stałej szybkości dysocjacji (k_{off}) i asocjacji (k_{on}) komponentów kompleksu. Dla dokładniejszego opisu oddziaływania ligand-receptor używa się też czasu rezydencji liganda w miejscu wiążącym. Co ważne, zauważono, że czas rezydencji może

korelować ze skutecznością leku *in vivo*. Ten właśnie istotny parametr autorka badała dla serii białek i ich ligandów używając metod dynamiki molekularnej oraz jej analizy. Tematyka poruszona w rozprawie mgr Magdaleny Ługowskiej jest więc istotna z punktu widzenia metod obliczeniowych dotyczących projektowania leków oraz porównywania ich wyników z badaniami doświadczalnymi.

Przedstawiona mi do recenzji rozprawa doktorska mgr Magdaleny Ługowskiej jest oparta o wyniki badań opublikowanych w trzech artykułach, ale zawiera też wyniki jeszcze nieopublikowane, zwłaszcza te dotyczące reduktazy InhA. W dwóch artykułach pani Ługowska jest pierwszym z dwóch autorów (drugim autorem jest promotor pomocniczy), co zwyczajowo oznacza, że jej wkład w powstanie tych artykułów był dominujący. Jeden z artykułów to praca przeglądowa w *Curr. Prot. Mol. Biol.*, która podczas składania przez kandydatkę rozprawy była w recenzji. Drugi artykuł został już opublikowany w *F1000Research*. Są to czasopisma o raczej niskim współczynniku oddziaływania jak na tę dyscyplinę. Trzeci artykuł, w którym kandydatka jest szóstym z siedmiu autorów został opublikowany w *Bioinformatics*, wysoko notowanym czasopiśmie, i dotyczy programu AQUA-DUCT, który ułatwia wyszukiwanie kanałów w białkach. Dorobek publikacyjny kandydatki nie jest więc imponujący, jednak zaznaczam, że przedmiotem mojej oceny jest przedstawiona mi rozprawa doktorska, a nie zbiór artykułów. Wyniki rozprawy pani Ługowska prezentowała na sześciu konferencjach, co jest dobrym osiągnięciem.

Rozprawa została napisana w języku angielskim, ale zawiera także dwa streszczenia, jedno i kilkustronicowe, w języku polskim. Rozprawa wraz z bibliografią liczy około 100 stron i jest napisana poprawnie językowo, jeśli chodzi o wersję anglojęzyczną. Podzielona jest na sześć rozdziałów, po których umieszczono wykaz odnośników oraz skróty. Rozprawa ma układ standardowy. Po stronie tytułowej następują kolejno spis treści, wykaz rysunków oraz tabel, lista publikacji autorki rozprawy i konferencji oraz streszczenie. Następnie autorka przedstawia główne rozdziały.

W pierwszym rozdziale zatytułowanym Wstęp (*Introduction*) autorka wprowadza czytelnika w tematykę kinetyki oddziaływania bimolekularnego. Podkreśla, że czas rezydencji leku w miejscu wiążącym receptora jest ważną miarą, która często koreluje z aktywnością leku *in vivo*. Następnie opisuje w skrócie podstawy symulacji klasycznej dynamiki molekularnej, uwzględniając metody wzmocnionego próbkowania, które zostały wykorzystane do obliczeń czasów rezydencji. Opis ten jest raczej praktyczno-techniczny a nie teoretyczny, brakuje choćby równania ruchu czy równania opisującego tzw. pole siłowe. Na pochwalenie zasługuje Tabela 1.1, w której autorka przedstawia kompleksy molekularne, dla których wyznaczono metodami obliczeniowymi czasy rezydencji ligandów i porównuje je z danymi doświadczalnymi (wraz z odnośnikami). Dziwi mnie jednak fakt, że autorka nie cytuje oryginalnej pracy z grupy R. Wade, gdzie wprowadzono technikę RAMD (*J Mol Biol.* Elsevier; 2000; 303: 797–811).

W podrozdziale 1.4 autorka rozprawy przedstawia jej cele, które są jasno określone. Z powodu potrzeby przeanalizowania i opracowania metod służących do obliczania czasów rezydencji ligandów w miejscach wiążących receptorów, autorka postawiła sobie następujące cele:
- sprawdzenie wiarygodności i uniwersalności metody τRAMD dla wcześniej testowanych i nowych kompleksów ligand-receptor,
- sprawdzenie jak wyznaczone relatywne czasy rezydencji korelują z czasami wyznaczonymi doświadczalnie,

- wyznaczenie najistotniejszych oddziaływań ligand-receptor, które przyczyniają się do długości przebywania cząsteczki w miejscu wiążącym receptora.

Do realizacji celów postawiono jeszcze dwa cele szczegółowe:
- stworzenie ogólnodostępnej bazy danych kinetycznych dla białek, dla których wyznaczono doświadczalnie czasy rezydencji i dla których struktury kompleksów zostały wyznaczone metodami doświadczalnymi i znajdują się w bazie danych Protein Data Bank,
- stworzenie narzędzia do automatycznej analizy oddziaływań ligand-receptor z symulacji dynamiki molekularnej.
Cele pracy zostały przez panią Ługowską w całości zrealizowane.

Drugi rozdział to Metody (*Methods*). W tym rozdziale kandydatka opisuje w skrócie użyte metody badawcze oraz oprogramowanie. Z metod najważniejszą jest τRAMD, która została rozwinięta w grupie prof. R. Wade w Heidelbergu. Pani Ługowska przedstawia schemat obliczeniowy stosowany w tej metodzie. Opisuje też oddziaływania, które brała pod uwagę jako najistotniej przyczyniające się do określonych czasów rezydencji. Szkoda, że nie podano jakie kryteria zostały zastosowane w przypadku wiązań wodorowych (np. odległość donor-akceptor, kąt) czy jak obliczano *stacking*. Czasy rezydencji zależą od bliskozasięgowych oddziaływań, takich jak *stacking* czy wiązania wodorowe, więc ich ocena może zależeć od przyjętych kryteriów geometrycznych czy energetycznych.

W rozdziale 3 przedstawione są już wyniki uzyskane przez autorkę rozprawy. W pierwszej części tego rozdziału pani Ługowska opisuje stworzoną przez siebie bazę PDBrt zawierającą struktury kompleksów z *Protein Data Bank* (PDB), dla których zostały wyznaczone dane kinetyczne, typu czasy rezydencji, wraz z możliwością oglądania tych struktur oraz odnośnikiem do publikacji. Baza zawiera około 60 kompleksów, gdyż tylko tyle odpowiednich rekordów autorka znalazła w PDB. W rozdziale 3 autorka wymienia też białka i ich ligandy, dla których przeprowadzała symulacje dynamiki molekularnej typu τRAMD. Pani Ługowska badała kompleksy 10 inhibitorów z reduktazą InhA, biorącą udział w szlaku syntezy kwasów tłuszczowych. Badała też 15 ligandów oddziałujących z białkiem szoku cieplnego (*heat-shock protein*) Hsp90 oraz kilka innych kompleksów m. in., inhibitora proteazy wirusa HIV-1 oraz receptora czynnika wzrostu naskórka (*epidermal growth factor receptor*) EGFR. Wszystkie białka mają znaczenie medyczne, gdyż są istotne z punktu widzenia projektowania leków, zwłaszcza przeciwnowotworowych.

W rozdziale 4 autorka najpierw sprawdza podobieństwo badanych ligandów używając współczynnika Tanimoto jako miary tego podobieństwa. Wyniki podsumowuje w tabeli, na histogramach i tzw. *heat maps* i stwierdza, że zestaw ligandów jest różnorodny. Następnie w rozdziale 4.2 przechodzi do analizy największego zestawu ligandów wiążących się z Hsp90. Porównuje otrzymane przez siebie czasy rezydencji z tymi uzyskanymi w roku 2018 przez Darię Kokh i współpracowników oraz ich względne wartości z danymi doświadczalnymi. W pracy referencyjnej D. Kokh et al. przeanalizowano 70 ligandów tego białka, a autorka pracuje na mniejszym zestawie - 15 ligandów. Ciekawa jestem jakie były kryteria doboru tych ligandów, dlaczego wybrano akurat te?

Nie jest dla mnie jasne, czy symulacje autorki były powtórzone dokładnie takim samym protokołem jak w pracy D. Kokh et al., do której autorka się porównuje. Przykładowo, czy było stosowane takie samo pole siłowe, czy ten sam program do symulacji τRAMD?

Niektóre otrzymane przez autorkę czasy rezydencji odbiegają od tych wcześniejszych, a także od doświadczalnych. Jako możliwe uzasadnienie autorka podaje zbyt małą liczbę symulacji τRAMD. Czy nie mogła ich zwiększyć chociaż dla tych kontrowersyjnych przypadków? Jednak ostatecznie autorka uznaje, że odtwarza poprzednie wyniki z wystarczającą dokładnością, aby móc zastosować metodę τRAMD do nietestowanych jeszcze przypadków ligand-receptor. Następnie pani Ługowska przedstawia wyniki dla kolejnego enzymu - InhA. Czasy rezydencji uzyskane przez autorkę korelują z danymi doświadczalnymi na poziomie R^2=0,68 i taki wynik autorka uznaje za zadowalający.

Każdy zestaw otrzymanych danych jest porównywany z zestawem doświadczalnym i podlega dokładnej analizie statystycznej. Mimo, iż zestawy danych zawierają maksymalnie 15 liczb, należy tę dbałość o analizę statystyczną wyników docenić. Choć taka analiza tylko dla 3 punktów przedstawiona w podrozdziale 4.4 nie jest wiarygodna.

W sumie symulacje τRAMD zostały przeprowadzone dla zestawu 28 kompleksów ligand-białko. Kandydatka stwierdza, że dla podobnych ligandów metoda ta się sprawdza w obliczaniu względnych czasów rezydencji, ale dla zestawu ligandów różnorodnych metoda okazała się mało przydatna.

W rozdziale 5 autorka rozprawy przedstawia właściwości ligandów białka InhA, które wpływają na ich czasy rezydencji w kompleksach, czyli analizuje mechanizmy dysocjacji uzyskane w trajektoriach τRAMD. Wskazanie tych kluczowych oddziaływań może być przydatne w projektowaniu leków. Autorka identyfikuje kontakty, oblicza ich częstość występowania i ostatecznie wybiera 24 istotne cechy, które dalej analizuje, między innymi wykorzystując analizę PCA (*principal component analysis*). Wyniki są przedstawione w formie tabel i wykresów, oddziaływania są pogrupowane i sklasyfikowane. Wymienione są czynniki, które są odpowiedzialne za zróżnicowanie czasów rezydencji tych ligandów. Jednak szkoda, że kontakty i aminokwasy, które autorka wymienia jako najistotniejsze (np. na stronie 89) nie zostały przedstawione w formie trójwymiarowej bazując na strukturach kompleksów. Ułatwiłoby to zrozumienie oddziaływań prowadzących do utworzenia i utrzymania kompleksu, które odpowiadają też za dysocjację ligandów (czyli czasy rezydencji).

Rozdział 6 jest zatytułowany Dyskusja i Wnioski i jest dosyć krótki. Wynika to z tego, że wcześniej na końcu każdego rozdziału autorka umieściła krótki podrozdział zatytułowany Podsumowanie i Konkluzja.

Poniżej wymieniam moje uwagi do rozprawy. Ciekawa jestem zdania autorki dotyczącego kilku poniższych zagadnień.

Opis metody dynamiki molekularnej zawiera protokół symulacji, ale brakuje informacji o teorii. Przydałaby się jakaś wzmianka o równaniu ruchu, trajektorii czy polu siłowym, skoro jest to główna technika używana przez autorkę.

Jakość wykresów pozostawia wiele do życzenia, gdyż są nieczytelne. Rozumiem, że chodziło o przedstawienie ogólnego trendu, ale chociaż podpisy na osiach powinny być czytelne.

Czasem sformułowania nie są jasne lub błędne. Np. str. 18 *Cellular pathway originates outside the cell with the ligand (a molecule that is the initial stimulus) approaching a specific protein.* Proces asocjacji ligandu odbywa się raczej wewnątrz komórki.

Str. 18 *Binding of the ligand to the receptor changes receptor's shape or activity, allowing it to transmit the signal or directly induce a change in the cell.* Jest to zbyt ogólnikowe zdanie, w którym nie rozumiem jak wiązanie liganda ma zmieniać aktywność receptora.

Str. 39 pierwszy paragraf, jest powtórzone to samo zdanie zaczynające się od „The only parameter..."

Str. 40 Dlaczego stosowano starszą wersję Ambera, czyli Amber14. Czy wynikało to z tego, że taką stosowano w pract Kokh et al. 2018, dla celów porównania?

Str. 41 Wskazany krok czasowy to 2 fs. Nie widzę jednak informacji o tym, aby stosowany był algorytm Shake czy Rattle dla wiązań z atomami wodoru. Ten krok czasowy wydaje się więc zbyt długi, jeśli nie zastosowano odpowiednich rozwiązań.

Str. 52, rys. 3.1 nie są zdefiniowane skróty M3R oraz DPP4.

Str. 68 i 68, rysunek 4.4 nie ma rysunku 4.4a, 4.4b, 4.4c i 4.4d, do których autorka odwołuje się na stronie 68

Str. 73 Tabela 4.3 Liczby w tabeli powinny mieć kropki (np. 0.09) a nie przecinki (0,09), gdyż rozprawa jest napisana w języku angielskim a nie polskim.

Autorka wymienia i w skrócie opisuje obliczeniowe metody wzmocnionego próbkowania służące do wyznaczania stałych szybkości asocjacji czy dysocjacji, ale nie wspomina o dynamice Brownowskiej. Czy ta technika nie nadaje się do badania czasów rezydencji?

Brakuje mi chociażby krótkiego opisu metod doświadczalnych stosowanych do wyznaczania czasów rezydencji. Skoro autorka sprawdza korelacje wyników symulacji z wynikami doświadczeń, przydałoby się znać także ograniczenia metod doświadczalnych i chociażby wspomnieć jak te czasy rezydencji zostały zmierzone. Częścią rozprawy było też stworzenie bazy danych czasów rezydencji uzyskanych doświadczalnie, więc informacje o technikach doświadczalnych wyznaczania τ czy k_{on} wzbogaciłyby tę rozprawę.

Pomimo powyższych uwag uważam, że badania mgr inż. Magdaleny Ługowskiej dotyczą ważnego obszaru dotyczącego oddziaływania ligand-receptor, który jest często pomijany przy projektowaniu leków. To jak długo dany ligand przebywa w miejscu wiązania może korelować z jego skutecznością blokowania tego miejsca. Metoda symulacji została dobrze dobrana do realizacji celu pracy. Metoda ta pozwala na wyznaczanie względnych czasów rezydencji dla różnych ligandów. Użytecznym i praktycznym efektem badań kandydatki jest baza danych zawierająca doświadczalne czasy rezydencji oraz inne wartości kinetyczne wyznaczone doświadczalnie. Baza danych zmierzonych czasów rezydencji ligandów wraz z podstawowymi wartościami i odnośnikami, którą stworzyła autorka

zawiera czasy rezydencji dla 60 kompleksów. Baza ta będzie użyteczna dla innych badaczy i mam nadzieję, że uzupełniana o nowe dane doświadczalne, jeśli się pojawią.

Podsumowując stwierdzam, że rozprawa doktorska mgr Magdaleny Ługowskiej stanowi oryginalne rozwiązanie problemu naukowego, potwierdza jej wiedzę teoretyczną w dziedzinie nauk inżynieryjno-technicznych, dyscyplinie informatyka techniczna i telekomunikacja oraz umiejętność samodzielnego prowadzenia pracy naukowej. Przedstawiona mi do oceny rozprawa doktorska spełnia warunki ustawy z dnia 20 lipca 2018 roku Prawo o szkolnictwie wyższym i nauce, w szczególności art. 187 tej Ustawy. W związku z tym wnoszę o dopuszczenie mgr Magdaleny Ługowskiej do dalszych etapów przewodu doktorskiego.