

Warszawa, 28/8/2016

Dr hab. Dariusz Plewczyński, prof. UW  
Laboratorium Genomiki Funkcjonalnej i Strukturalnej  
Centrum Nowych Technologii  
Uniwersytet Warszawski  
Ul. Banacha 2c, 02-097 Warszawa, Polska

#### RECENZJA

rozprawy doktorskiej  
mgr Sabiny Podlewskiej  
zatytułowanej

*„Development of Machine Learning-based Tools  
for Computer-Aided Drug Design”*

wykonanej w Zakładzie Chemii Leków  
w Instytucie Farmakologii Polskiej Akademii Nauk w Krakowie  
pod kierunkiem promotora prof. dr hab. Andrzeja J. Bojarskiego  
oraz promotora pomocniczego dr Rafała Kurczaba

Rozprawa doktorska mgr Sabiny Podlewskiej została przygotowana pod kierunkiem prof. dr hab. Andrzeja J. Bojarskiego który wskazał doktorantce interesującą tematykę badawczą i wspierał realizację jej prac, oraz promotora pomocniczego dr Rafała Kurczaba, który miał swój istotny wkład w przygotowanie rozprawy doktorskiej, jak również współpracował z doktorantką. Tematyka pracy skupia się na rozwoju metod komputerowego wspomaganie procesu projektowania leków, poczynając od wstępnej identyfikacji związków aktywnych, przez ich optymalizację oraz badanie aspektów strukturalnych ich aktywności przy użyciu metod uczenia maszynowego.

Nie jest moim celem przekonywanie Szacownej Rady Wydziału, jak kluczowym dla współczesnych nauk farmaceutycznych jest wykorzystanie zaawansowanych metod obliczeniowych, szczególnie przy próbie opisu oddziaływań między ligandami i białkami. Jednak chciałbym zwrócić uwagę na szczególny przypadek, w którym metody te są niezbędne do osiągnięcia postawionego sobie celu. Mianowicie w kontekście dużych zbiorów danych, stanowiących punkt wyjścia do wirtualnego przesiewania nie jest możliwe zastosowanie powolnych metod fizyko-chemicznych z uwagi na skalę problemu. Dlatego też opracowano szereg metod informatycznych, które umożliwiają sprawdzanie dużych bibliotek związków małowymiarowych z uwagi na ich powinowactwo do wybranych białkowych celów terapeutycznych. Chciałbym też podkreślić odwagę doktorantki w nietrywialnym połączeniu metod bazujących na strukturze i własnościach liganda, oraz metod wykorzystujących strukturę trójwymiarową celu terapeutycznego (białka). W ten sposób stworzona została metodologia oceny wyników dokowania przy użyciu algorytmów uczących się. Moim zdaniem jest to udana implementacja rozwiązania obliczeniowego przy użyciu różnorodnych typów reprezentacji biomolekuł, jak również różnych metod statystycznego uczenia się. Praca doktoranta jest wartościowym krokiem w kierunku zrozumienia efektów interakcji między białkami i ligandami.

Wybór celu i obiektu prowadzonych badań jest w pełni uzasadniony z uwagi na fakt, że strukturalne aspekty oddziaływania ligand-białko stanowią fundament zrozumienia aktualnie używanych i wykorzystywanych leków małowymiarowych. Przy odpowiedniej selekcji cech ligandów dla ustalonych celów białkowych uzyskany za pomocą leku efekt terapeutyczny może być szansą na przeżycie dla osoby chorej.

Najogólniej charakteryzując dysertację, trzeba zauważyć jej szeroki zakres tematyczny, co prowadzić musiało do zwięzłości wypowiedzi w części wstępnej (która jednak i tak liczy aż 84 strony), co jednak znajduje rozwinięcie w siedmiu publikacjach doktorantki składających się na całość rozprawy. Rozprawa przygotowana jest w języku angielskim, została podzielona w sposób klasyczny na wprowadzenie (7 str.), opis celu pracy (4 strony), część literaturową (4 str.), opis publikacji składających się na rozprawę doktorską (40 str.), podsumowanie wraz z najważniejszymi osiągnięciami (4 str.), opis dalszych planów badawczych (4 str.), spis literatury obejmujący 97 pozycji, oraz część główną złożoną z 7 publikacji ze znaczącym wkładem współ-autorskim doktorantki. Całość poprzedzona jest streszczeniem w języku polskim i spisem stosowanych skrótów, a zakończona listą literatury, oraz załącznikami będącymi publikacjami ze współautorstwem doktorantki (od strony 85).

Aby zrealizować ambitnie sformułowany cel rozprawy, doktorantka wykorzystwała odpowiednio dobrane metody badawcze modelowania wspomaganego komputerowo, oraz wirtualnego przesiewania. W toku prowadzonych badań używała metod uczenia maszynowego. Większość obliczeń została wykonana z zastosowaniem kilku podstawowych algorytmów uczących się (np. metoda wektorów wspierających), nie jest jednak jasne dlaczego został dokonany ten, a nie inny wybór i jak na wyniki wpływa decyzja dotycząca się użytej metody statystycznej. Pewien niedosyt pozostawia brak wystarczającej dyskusji w tym względzie.

Tym niemniej cele rozprawy udało się zrealizować. Należy w pełni zgodzić się z doktorantką, iż z sukcesem: opracowano zestaw narzędzi do przewidywania aktywności biologicznej biomolekuł opartych o metody uczenia maszynowego. Ustalono nowatorską metodologię wyboru przykładów negatywnych (tj. związków nieaktywnych) co stanowi jeden z kluczowych problemów chemoinformatyki współczesnej. Dokonano ciekawej analizy baz danych: ZINC, MDDR oraz DUD. Opracowano nową metodę optymalizacji parametrów metody maszyny wektorów wspierających.

Jako bardzo ważny rezultat przeprowadzonych badań należy uznać połączenie informacji strukturalnej o ligandzie z informacją o celu białkowym, oraz odpowiedź na pytanie o niespójność danych doświadczalnych, co doktorantka zaadresowała opracowując protokół uwzględniający niepewność wyników biologicznych, wprowadzając swego rodzaju algorytm fuzji danych z różnych źródeł. Algorytm ten przetestowano na zbiorze 25 przykładowych celów terapeutycznych.

Aż prosi się by doktorantka zaproponowała eksperymenty syntetyczne pozwalające potwierdzić bądź zaprzeczyć tej tezie. Niewątpliwie taką rolę częściowo pełni analiza ligandów receptora serotoninowego 5-HT<sub>6</sub>, mimo wszystko sądzę, że taka dyskusja byłaby wskazana w trakcie publicznej obrony. Jako elementy do omówienia należy postawić pytania dlaczego nie przeprowadzono badań doświadczalnych? O ile można zgadywać, że doktorantka koncentrowała się na zagadnieniach obliczeniowych to jednak badania choćby na wybranych celach terapeutycznych znacząco podniosłyby wartość naukową pracy. Czy nie warto byłoby również zbadać bardziej szczegółowo charakter fizyczny i naturę biochemiczną tworzących się między ligandem i białkiem wiązań?

Rozprawa przedstawiona przez doktorantkę, pomimo drobnych (i niewartych wymienienia) usterek językowych, napisana jest interesująco i świadczy o bardzo dobrym zrozumieniu stawianych zadań badawczych. Doktorantka pokazała, że dobrze rozumie i umiejętnie używa różne metody badawcze, odważnie formułuje hipotezy naukowe i je rygorystycznie weryfikuje tak więc ogólną wiedzę teoretyczną doktorantki należy ocenić bardzo wysoko. Nie sposób nie zauważyć, że całość wyników badań doktorantki było poddane szczegółowej ocenie przez recenzentów wybranych przez edytorów czasopism naukowych. W renomowanych czasopismach, są to zwykle bardzo rygorystyczne oceny.

### Ocena końcowa

Podsumowując moją recenzję rozprawy doktorskiej pani magister Sabiny Podlewskiej stwierdzam, że zaprezentowaną rozprawę oceniam bardzo wysoko. Biorąc pod uwagę wysoki poziom naukowy rozprawy doktorskiej, udane połączenie różnorodnych metod komputerowych, a także walory aplikacyjne oceniam rozprawę doktorską mgr Sabiny Podlewskiej jako ważny wkład w rozwój do naszej wiedzy o chemii leków. Uważam, że rozprawa ta spełnia zwyczajowe i ustawowe wymogi, stawiane rozprawom doktorskim, stanowi oryginalne rozwiązanie problemu naukowego, demonstruje ogólną wiedzę teoretyczną kandydatki oraz pokazuje umiejętność samodzielnego prowadzenia pracy naukowej. Zgłaszam Radzie Wydziału Chemii Uniwersytetu Jagiellońskiego w Krakowie wniosek o dopuszczenie pani magister Sabiny Podlewskiej do dalszych etapów przewodu doktorskiego. Dodatkowo – biorąc pod uwagę bardzo wysoki poziom naukowy rozprawy – wnoszę o jej wyróżnienie.



Dr hab. Dariusz Plewczynski, prof. UW