

Katowice, 16 listopada 2020r.

Prof. dr hab. Adam Smoliński  
Główny Instytut Górnictwa  
Plac Gwarków 1, 40-166 Katowice  
[asmolinski@gig.eu](mailto:asmolinski@gig.eu)

**Recenzja rozprawy doktorskiej**  
**mgr inż. Roksany Duszkiewicz**  
**pt. „Ocena liganda jako potencjalnego leku na**  
**próbie wybranych baz wielkich danych”**

Podstawą opracowania niniejszej recenzji było pismo Dyrektora Instytutu Chemii Wydziału Nauk Ścisłych i Technicznych z dnia 9 października 2020 roku. Promotorem rozprawy jest prof. dr hab. inż. Jarosław Polański.

**Tematyka rozprawy doktorskiej i trafność jej wyboru**

Praca poświęcona jest zagadnieniom związanym z szeroko rozumianym, złożonym procesem projektowania leków. Opracowanie nowego leku jest procesem nie tylko bardzo czasochłonnym, ale przede wszystkim kosztownym. Dlatego też kluczową rolę w tym procesie odgrywa możliwość wykorzystania w procesie projektowania leków narzędzi badawczych w postaci oprogramowania bioinformatycznego. Oczywiście

obecnie dostępne metody nie zastąpią w pełni kosztownych badań laboratoryjnych i nie pozwolą na jednoznaczne i pewne przewidzenie struktury nowego, bezpiecznego i skutecznego leku. Metody te umożliwią jednak znaczne obniżenie kosztów oraz skrócenie czasu trwania projektowania nowego związku, z czym wiąże się szybsza identyfikacja potencjalnych kandydatów na leki, które trafiają do badań przedklinicznych. Oprogramowanie bioinformatyczne do modelowania molekularnego pozwala obecnie na efektywne przeszukiwanie baz danych cząsteczek, analizę i porównanie konfiguracji przestrzennej badanych molekuł, identyfikację konformacji aktywnej, a także na analizę oddziaływań ligand-receptor i przewidywanie mechanizmów działania leków. Dostępne dane zebrane w formie dużych baz danych w połączeniu z zaawansowanym oprogramowaniem bioinformatycznym pozwalają obecnie na projektowanie nowych związków o określonych właściwościach, oraz na optymalizację badanej molekuly pod kątem oddziaływania z receptorem.

Badania prowadzone w ramach pracy doktorskiej obejmowały ocenę liganda jako potencjalnego leku. Doktorantka skupiła się na analizie estymatora wydajności liganda (LE) jako jednego z typów reprezentacji właściwości substancji stosowanych w projektowaniu leków. W ramach prowadzonych prac badawczych Doktorantka analizowała zestawy danych z dużych baz danych molekularnych ChEMBL i PubChem.

Uważam, że zaproponowany temat i zakres pracy wpisuje się bez wątplenia w obszar ważnych i aktualnych problemów dotyczących projektowania leków. Podjęcie tematu należy uznać za w pełni uzasadnione.

## **2. Ogólna charakterystyka rozprawy**

Recenzowana praca doktorska liczy 143 strony i zawiera 39 Rycin, 5 Tabel i 106 pozycji bibliograficznych, wśród których dominują artykuły naukowe w języku angielskim z ostatnich kilku lat. Zasadniczą część rozprawy doktorskiej zawarto na 102 stronach. Ponadto do rozprawy dołączono cztery załączniki (przedstawione na str.103-143),

tj. życiorys Doktorantki, wykaz publikacji, oświadczenia współautorów publikacji dotyczące ich wkładu w powstanie artykułów oraz kopie publikacji, których tematyka jest ściśle powiązana z pracą doktorską. Prace te opublikowane zostały w prestiżowych czasopismach naukowych, takich jak: Chemometrics and Intelligent Laboratory Systems, Scientific Reports, Letters in Drug Design Discovery, Current Protein and Peptide Science i Acta Poloniae Pharmaceutica.

Rozprawę doktorską rozpoczyna krótki opis celu pracy oraz wstęp, w którym mgr inż. Duszkiewicz w sposób syntetyczny przedstawia zakres pracy i zasadność podjęcia tematu badawczego. Jak słusznie zostało zauważone projektowanie leku wymaga ustalenia dwóch grup zmiennych. Pierwszą grupę stanowią deskryptory molekularne związane z cząsteczkową reprezentacją związku chemicznego. Druga grupa zmiennych związana jest z biologiczną aktywnością leku (właściwościami chemicznymi substancji). Deskryptory można obliczyć w operacjach *in silico*, co nie jest operacją kosztowną w odróżnieniu od działań mających na celu określenie właściwości chemicznych substancji, wymagających wielu pomiarów. Dlatego też wyniki takich badań gromadzone są w dużych zestawach danych i analiza tych danych jest pierwszym etapem na drodze poszukiwania związków chemicznych o właściwościach leczniczych. Doktorantka w swojej pracy analizuje duże zestawy danych zgromadzonych w bazach ChEMBL i PubChem oraz dąży do wyjaśnienia kontrowersji związanych ze wskaźnikiem efektywności – wydajności liganda, jako jednego z typów reprezentacji molekularnej, a także jego przydatności w projektowaniu leków.

We wprowadzeniu do Rozdziału I, poświęconego podstawom teoretycznym opartym na analizie dostępnej literatury przedmiotu badań Doktorantka stwierdza, że w przypadku ligandów egzo- jak i endogennych siła działania jest podstawową cechą determinującą interakcje z receptorem. Jednocześnie stwierdza, że „precyzyjne znaczenia pojęć związanych z oddziaływaniem ligandów wciąż są dalekie od precyzyjnego zrozumienia”. Wymienione przez Doktorantkę pojęcia takie jak powinowactwo, czyli siła działania liganda z receptorem przy danym stężeniu leku, siła działania leku, czyli ilość leku niezbędnego do wywołania efektu o określonej wartości czy skuteczność jako miara maksymalnej odpowiedzi farmakologicznej leku,

gdy zachodzi interakcja z receptorem są jednak zdefiniowane w literaturze. Dlatego wątpliwości budzi, co Doktorantka rozumie pod pojęciem „precyzyjnego zrozumienia tychże pojęć”?

W podrozdziale I.2 Doktorantka omawia pojęcia deskryptora i właściwości, których poprawne zdefiniowanie jest kluczowe w procesie projektowania molekularnego. Zwraca słusznie uwagę na fakt, że właściwości odnoszą się do substancji natomiast deskryptory obliczane są na podstawie reprezentacji molekularnej. Powoduje to szereg problemów i nieścisłości widocznych choćby przy porównaniu nomenklatury IUPAC i nomenklatury stosowanej w chemoinformatyce. Jednocześnie mgr inż. Duszkiewicz wskazuje,

że dostępne są liczne prace, w których projektowanie leków oparte jest na spektrum właściwości. Są tu wspomniane prace nad zrobotyzowanymi metodami pomiaru siły działania lub nowymi metodami skutecznego i niedrogo pomiaru profili analizowanych lipidów z udziałem kilku tysięcy lipidów. Zabrakło jednak komentarza Doktorantki ma temat tych metod.

W podrozdziale I.3 Doktorantka omawia „lekoopodobności” jako główne cechy, które powinny posiadać związki będące kandydatami na przyszłe leki, podczas gdy w podrozdziale I.4 przedstawia stan wiedzy na temat aktywności biologicznej (termodynamika wiązania ligand-receptor, zależności charakteryzujące aktywności biologiczne, rodzaje i sposób pomiaru oraz znaczenie aktywności biologicznych) w projektowaniu leku. Omawiając szczegółowo termodynamikę wiązań ligand - receptor Doktorantka stwierdza m.in. słusznie, że odwrotnością stałej równowagi procesu asocjacji,  $K_r$  jest stała dysocjacji  $K_d$  jednak równanie jest zapisane błędnie. Zgodnie z opisem w pracy powinno być  $K_d=1/K_r$  a nie  $K_r=1/K_d$ . Opisując w podrozdziale I.4.2 zależności parametrów  $K_d$  oraz  $K_i$  Doktorantka wskazuje na różnice w definicji stałej hamowania  $K_i$  i stałej dysocjacji  $K_d$ . Obie wielkości służą do opisu powinowactwa wiązania cząsteczki lub makrocząsteczki do enzymu lub receptora. Zabrakło w tekście rozprawy jasnego wyjaśnienia różnicy pomiędzy tymi wielkościami. Doktorantka stwierdza jedynie, że stała dysocjacji jest pojęciem bardziej ogólnym natomiast stała hamowania „reprezentuje również stałą dysocjacji”. Ponadto każda zmienna powinna

być w całym tekście oznaczana takim samym symbolem. W tytule podrozdziału stała hamowania i stała dysocjacji oznaczane są jako  $K_I$  i  $K_D$  natomiast w tekście jako  $K_i$  i  $K_d$ . W podrozdziale I.5 Doktorantka skupiła się na przeglądzie stanu wiedzy na temat zastosowania typu LE w projektowaniu leku. Wielkość cząsteczek ma istotny wpływ na efektywność opracowania nowych leków. Rozmiar liganda odgrywa istotną rolę w dopasowaniu i wiązaniu ligand-cel i właśnie ocena tej roli jest ważnym narzędziem w poszukiwaniu bardziej skutecznych sposobów projektowania leków.

Przegląd stanu wiedzy kończą dwa podrozdziały I.6 i I.7, w których Doktorantka przedstawia duże bazy danych zawierających informacje na temat substancji chemicznych, ich aktywności biologicznych, właściwości wiązania, funkcjonalności oraz właściwości ADMET dużej liczby bioaktywnych związków lekopodobnych z baz danych ChEMBL i PubChem oraz przedstawia matematyczne aspekty związane z projektowaniem leków.

Zabrakło krótkiego, syntetycznego podsumowania części poświęconej przeglądowi stanu wiedzy, uzasadniającego cel prowadzenia podjętych badań i stanowiącego nawiązanie do kolejnej części rozprawy, przedstawiającej część badawczą. Nie do końca właściwe moim zdaniem jest umieszczenie opisu metodyki badań na końcu rozprawy, po podsumowaniu. Rozdział zatytułowany „Część eksperymentalna” powinien stanowić pierwszy rozdział części „Badania własne”.

Przedmiotem badań własnych Doktorantki były statystyki molekularne modelowane w dwóch bazach danych ChEMBL i PubChem, które są obecnie największymi dostępnymi zbiorami informacji o wartościach aktywności biologicznych związków chemicznych. Dodatkowo w ramach pracy przeanalizowano dane dla wybranych serii leków zebrane w Binding Database oraz w bazie Psychoactive Drug Screening Program. W podrozdziale II.2 Doktorantka wyjaśnia, że w najszerszym znaczeniu statystycznym wydajność liganda (LE) jest transformacją  $pAC_{50}$  i HAC. Podane w pracy równanie na str. 47 jest niekompletne. Powinno być  $LE = 1,37 * (pAC_{50}/HAC)$ . Wzór ten poprawnie jest zapisany w publikacji [D1] (patrz str. 110 pracy doktorskiej). Należy również nadmienić, że w pracy doktorskiej zabrakło numeracji równań.



W dyskusji wyników swoich badań mgr inż. Duszkiewicz stawia pytanie o optymalną siłę działania leku, które pozostaje bez odpowiedzi. Doktorantka pisze, że w praktyce dążenie wyłącznie do maksymalizacji wartości siły oddziaływania nie zawsze jest podejściem właściwym. W rzeczywistości o potencjalne kandydatów na leki decyduje szereg tzw. właściwości „lekoopodobnych” np. ADMET i właściwości lipofilowe, które prognozowane są *in silico* i rzadko mierzone. Wielkości ADMET czy logP są deskryptorami molekularnymi wyliczanymi na podstawie struktur odpowiednich cząsteczek w celu umożliwienia symulacji prognozowanych wartości właściwości nieznanymi projektowanych struktur. Podczas publicznej obrony proszę jednak o krótki komentarz na temat optymalnej siły działania leku.

Co Doktorantka rozumie pod sformułowaniem „dostroić” wartości  $a$  i  $b$  w funkcji SCORE w celu przyjęcia strategii preferowanej w konkretnym projekcie rozwoju leku. W jaki sposób dobiera się te wartości  $a$  i  $b$  (patrz str.65).

Ciekawym etapem pracy są analizy 2 milionów cen związków chemicznych dostępnych w katalogu Abamachem. Zaskakującym jest na tym etapie stwierdzenie Doktorantki, że „analiza danych ekonomicznych jest zdecydowanie bardziej złożona niż danych chemicznych”. Na Rycinie 29 Doktorantka przedstawia binowane ceny 2 milionów związków chemicznych dostępne w bazie Abamachem, na podstawie których stwierdza, że „efekt hiperboliczny obserwowany jest dla „ceny w skali wagowej” przy niskich masach cząsteczkowych. Co jest przyczyną nieliniowego zniekształcenia trendu przy niskich masach cząsteczkowych?

W Rozdziale 7 Doktorantka analizuje zależności wydajność liganda od temperatury. Na str. 79 Doktorantka stwierdza, że temperatura jest związana ze średnią energią cząsteczek w układzie. Oznacza to, że temperatura jest związana ze średnią energią pojedynczej cząsteczki, co z kolei oznacza, że temperatura jest związana ze skalą pojedynczej cząsteczki. Co Doktorantka rozumie pod stwierdzeniem: „Jednak temperatura jest tylko skalą kodującą temperaturę wrzenia, złożoną właściwość cieczy, która jest skorelowana z ciepłem parowania lub energią potrzebną do przekształcenia substancji z cieczy w fazę pary.”?

Czytając rozprawę doktorską zauważono pewne niedociągnięcia o charakterze redakcyjnym. Wybrane z nich przedstawiono poniżej:

- ✓ W podpisie Ryciny 2 fragment „Często nie rozróżnia tych różnych typów aktywności biologicznej ligandów, zmodyfikowane wg [1,2]” powinien znaleźć się w tekście, a nie w podpisie ryciny;
- ✓ Na str. 20 jest „kryteria lekopodobieństwa”, powinno być „kryteria lekopodobieństwa”;
- ✓ W Tabeli 5 przedstawiającej liczebność dużych zbiorów danych dotyczących aktywności biologicznej liczebność danych określono na podstawie „\* Suthan 2018”. W tabeli nie zaznaczono „\*\*” natomiast w legendzie Tabeli widnieje „\*\*” który oznacza „na podstawie [D3]”.
- ✓ Na str. 53 jest „ Podobne wykresy jak w dyskutowane powyżej można....” Powinno być „Podobne wykresy jak dyskutowane powyżej można...”.
- ✓ Na rycinach 22 i 23 na osi OY jest pAC50/LE/pPLE, powinno być pAC<sub>50</sub>/LE/pPLE
- ✓ Na str. 80 jest „Formalnie BP nie jest reprezentacją ani molową, ani wagą; jednak; BP zależy of MW. . Kiedy cząsteczki...” powinno być „Formalnie BP nie jest reprezentacją ani molową ani wagą, jednak BP zależy of MW. Kiedy cząsteczki...”
- ✓ Ryciny 34 i 38 praktycznie przedstawiają ten sam interfejs graficzny oprogramowania Instant JChem.

### **Wniosek końcowy**

Stwierdzam, że przedłożona mi do recenzji rozprawa doktorska mgr inż. Roksany Duszkiewicz spełnia wymagania Ustawy z dnia 14 marca 2003r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65. Poz. 595 z późniejszymi zmianami). Zwracam się do Rady Instytutu Chemii z wnioskiem o dopuszczenie mgr inż. Roksany Duszkiewicz do dalszych etapów przewodu doktorskiego.

