



UNIWERSYTET  
WARMIŃSKO-MAZURSKI W OLSZTYNIE

dr hab. Maciej Maciejczyk, prof. UWM

Katedra Fizyki i Biofizyki

Olsztyn, dnia 30 września 2020

**Recenzja rozprawy doktorskiej mgr. Marty Stępniewskiej-Dziubińskiej  
„Wykorzystanie sieci neuronowych do przewidywania aktywności związków  
niskocząsteczkowych w układach ligand-receptor” wykonanej w Zakładzie  
Bioinformatyki Instytutu Biochemii i Biofizyki PAN  
pod kierunkiem dr hab. Pawła Siedleckiego.**

Rozprawa doktorska składa się z czterech prac opublikowanych w czasopismach z tzw. listy filadelfijskiej wraz z oświadczeniami współautorów o ich udziale w przygotowaniu publikacji oraz autoreferatu zawierającego streszczenia (w języku polskim i angielskim) oraz wprowadzenie wraz z krótkim omówieniem wyników. Autoreferat przygotowano w języku angielskim.

Prace składające się na rozprawę doktorską opublikowane zostały w prestiżowych czasopismach o zasięgu międzynarodowym, których sumaryczny współczynnik oddziaływania (impact factor) wynosi 18,485, natomiast sumaryczna liczba punktów MNiSW wynosi 600 (2x200 i 2x100).

Na rozprawę doktorską składają się następujące publikacje:

- Stępniewska-Dziubińska, M.M.; Zielenkiewicz, P.; Siedlecki, P. *DeCAF—Discrimination, Comparison, Alignment Tool for 2D Pharmacophores*. *Molecules* 2017, **22**, 1128.
- Stępniewska-Dziubińska, M.M.; Zielenkiewicz, P.; Siedlecki, P., *Development and evaluation of a deep learning model for protein–ligand binding affinity prediction*, *Bioinformatics*, 2018, **34**, 3666.
- Wójcikowski, M., Kukiłka, M., Stępniewska-Dziubińska, M.M., Siedlecki, P., *Development of a protein–ligand extended connectivity (PLEC) fingerprint and its application for binding affinity predictions*, *Bioinformatics*, 2019, **35**, 1334.
- Stępniewska-Dziubińska, M.M.; Zielenkiewicz, P.; Siedlecki, P. *Improving detection of protein–ligand binding sites with 3D segmentation*. *Sci. Rep.*, 2020, **10**, 5035.



UNIWERSYTET  
WARMIŃSKO-MAZURSKI W OLSZTYNIE

dr hab. **Maciej Maciejczyk**, prof. UWM

Katedra Fizyki i Biofizyki

Pani Doktorantka jest pierwszym autorem w trzech publikacjach, w których brała udział w przygotowaniu koncepcji badań, implementacji modelu, przygotowaniu danych, przeprowadzeniu eksperymentów numerycznych, analizie wyników oraz przygotowaniu manuskryptu. Zestawienie wkładu Pani Doktorantki w powstanie tych publikacji z oświadczeniami współautorów, świadczy o tym, że wraz z Panem Promotorem mieli dominujący wkład w ich powstanie. W czwartej pracy (trzeciej w cyklu), w której Pani Doktorantka jest trzecim autorem, jej wkład polegał na przeprowadzeniu eksperymentów numerycznych, analizie wyników oraz przygotowaniu manuskryptu.

Pani Doktorantka w swojej rozprawie doktorskiej zajęła się tematem predykcji powinowactwa małych cząsteczek do białek. Problem ten jest niezwykle ważny, gdyż jego rozwiązanie pozwala na zastosowanie opracowanej metody we wstępnych fazach projektowania nowych leków. Metody komputerowego wspomagania projektowania leków można podzielić na dwie grupy – metody bazujące wyłącznie na właściwościach ligandów wiążących się do białek (ang. ligand-based design) oraz metody bazujące na strukturach białek do których mają wiązać się szukane ligandy (ang. structure-based design). W publikacjach składających się na rozprawę doktorską zastosowano obie metody. Warto podkreślić, że w trzech pracach cyklu do rozwiązania postawionego problemu zastosowano dynamicznie rozwijającą się w ostatnich latach metodę uczenia maszynowego opartego na sieciach neuronowych.

W pierwszej pracy znajdującej się w Rozdziale 4 przedstawiono narzędzie DeCAF, służące do tworzenia dwuwymiarowych farmakoforów dla badanego zbioru molekuł. Celem autorów było stworzenie reprezentacji molekuły oddającej jej właściwości fizyko-chemiczne oraz względne relacje grup funkcyjnych, bez konieczności kosztownej generacji wielu konformerów. W tym celu zastosowano nowatorskie podejście polegające na przetransformowaniu molekuły do postaci nieukierunkowanego grafu, w którym węzłami są atomy posiadające określone właściwości farmakoforowe, a krawędzie oddają odległości pomiędzy nimi. Następnie opracowano specjalny algorytm pozwalający na dopasowanie modeli molekuł poprzez znalezienie ich największego wspólnego podgrafu. Skuteczność opracowanej metody zbadano przeprowadzając dwa testy – klasyczny test wyszukiwania

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Katedra Fizyki i Biofizyki

ul. Michała Oczapowskiego 4, 10-718 Olsztyn

tel. (89) 523 32 34    [maciej.maciejczyk@uwm.edu.pl](mailto:maciej.maciejczyk@uwm.edu.pl)

fax (89) 523 38 61    [www.uwm.edu.pl/wnz/node/508](http://www.uwm.edu.pl/wnz/node/508)





UNIWERSYTET  
WARMIŃSKO-MAZURSKI W OLSZTYNIE

dr hab. **Maciej Maciejczyk**, prof. UWM

Katedra Fizyki i Biofizyki

molekuł o podobnej aktywności oraz bardziej wyrafinowany test oparty na metodzie SEA (ang. Similarity Ensemble Approach). Metodę DeCAF porównano wyników uzyskanych za pomocą 14-tu innych tzw. odcisków palców molekuly (ang. fingerprint). Autorzy pokazali, że wyniki testów są podobne lub lepsze od pozostałych testowanych metod. Szczególnie interesujące i wartościowe jest porównanie metody DeCAF do metody USRCAT, wymagającej kosztownej generacji wielu konformerów. Również w tym porównaniu DeCAF wypada podobnie lub lepiej, przy czym, ze względu na znacznie mniejszy koszt obliczeniowy metoda może być zastosowana do większych zbiorów danych niż metoda USRCAT. Metoda zaimplementowana została w języku Python i jest publicznie dostępna.

W drugiej pracy znajdującej się w Rozdziale 5 przedstawiono model o nazwie Pafnucy, służący do przewidywania powinowactwa układów ligand-receptor. Model należy do grupy modeli bazujących na znanych strukturach kompleksów białko-ligand oraz znajomości ich względnego powinowactwa. Do rozwiązania postawionego problemu autorzy zastosowali dynamicznie rozwijającą się w ostatnich latach metodę uczenia maszynowego, opartą na głębokiej konwolucyjnej sieci neuronowej. Dane wejściowe dla sieci neuronowej stanowiły struktury kompleksów białko-ligand przetransformowane do postaci cztero-wymiarowych tensorów zawierających współrzędne punktów siatki oraz specjalnie wybrane własności fizyko-chemiczne najbliższego atomu. Autorzy pokazali, że testy przeprowadzone za pomocą opracowanego modelu dają lepsze średnie wyniki w porównaniu z pięcioma innymi uznanymi metodami m.in. X-Score i ChemScore. Kod źródłowy metody jest publicznie dostępny i wygląda na to, że metoda spotkała się ze znaczącym uznaniem środowiska naukowego, gdyż praca w ciągu zaledwie dwóch lat od ukazania się, cytowana była aż 32 razy (źródło Web of Science). Nie bardzo rozumiem dlaczego na stronie 16 rozprawy pojawia się stwierdzenie, że manuskrypt opisujący model Pafnucy cytowany był ponad 50 razy. Prosiłbym Panią Doktorantkę o wyjaśnienie tej rozbieżności.

W trzeciej pracy znajdującej się w Rozdziale 6 opracowano oryginalny odcisk palca (ang. fingerprint) odzwierciedlający siatkę oddziaływań pomiędzy białkiem i ligandem. Odcisk o nazwie PLEC został następnie użyty do konstrukcji kilku modeli opartych na uczeniu maszynowym, które następnie zastosowano do szacowania powinowactwa ligandów do receptorów. Ponownie testy przeprowadzone na bazie struktur PDBbind oraz na zbiorze



UNIWERSYTET  
WARMIŃSKO-MAZURSKI W OLSZTYNIE

dr hab. Maciej Maciejczyk, prof. UWM

Katedra Fizyki i Biofizyki

struktur Astex Diverse Set wykazały wyższość opracowanej metody nad innymi dostępnymi metodami takimi jak X-Score czy RF-Score. Ponadto autorzy pokazali, że najprostszy model liniowy daje wyniki bardzo zbliżone do bardziej wyrafinowanych metod uczenia maszynowego, co wskazuje na dobrze przemyślaną i skuteczną konstrukcję reprezentacji (fingerprint) siatki oddziaływań białko-ligand. Również w tym wypadku kod źródłowy metody jest publicznie dostępny.

W czwartej pracy znajdującej się w Rozdziale 7 autorzy zajęli się problemem wyszukiwania miejsc wiążących na powierzchni białek, który w sposób oczywisty mocno związany jest z problemem predykcji powinowactwa białko-ligand. W modelu Kalasanty zastosowano metodę głębokiego uczenia opartej na modelu U-Net, stworzonemu do segmentacji obrazów dwuwymiarowych, odpowiednio zaadoptowanemu do siatek trójwymiarowych, odzwierciedlających strukturę i właściwości układów białkowych. Model trenowano oraz sprawdzono na bazie struktur sc-PDB, zawierającej struktury białek ze znanymi miejscami wiążącymi. Wyniki następnie porównano do DeepSite – innej wiodącej metody wykorzystującej uczenie maszynowe. Autorzy wykazali bardzo znaczącą przewagę opracowanego przez nich modelu nad modelem DeepSite, zarówno jeśli chodzi o precyzję lokalizacji centrum kieszeni wiążącej, jak również o oddanie kształtu kieszeni wiążącej. Kod źródłowy metody jest publicznie dostępny.

Mam do Pani Doktorantki dwa pytania związane z opracowywaniem modelu Pafnucy. Na stronie 3668 publikacji z Rozdziału 5 autorzy piszą „All complexes in this study were protonated and charged using UCSF Chimera with Amber ff14SB for standard residues and AM1-BCC for non-standard residues and ligands”. Dalej autorzy piszą, że procedura ta została dobrana w ten sposób, aby móc porównać model Pafnucy z innymi wiodącymi modelami służącymi do predykcji powinowactwa białko-ligand. W cytowanym zdaniu nie jest *explicite* napisane w jaki sposób określone były stany protonacyjne reszt miareczkowalnych. Zakładam, że przyjęto najbardziej prawdopodobne stany protonacyjne (modelowe) dla eksponowanych reszt modelowych w neutralnym pH. W wielu układach białko-ligand stan protonacyjny grup miareczkowalnych może mieć znaczący wpływ na ułożenie liganda w kieszeni wiążącej oraz oczywiście na stałą asocjacji. Czy Pani Doktorantka rozważała możliwość zastosowania metod numerycznego miareczkowania do





UNIWERSYTET  
WARMIŃSKO-MAZURSKI W OLSZTYNIE

dr hab. Maciej Maciejczyk, prof. UWM

Katedra Fizyki i Biofizyki

przygotowania kompleksów stanowiących bazę danych dla modelu Pafnucy? Czy w ocenie Pani Doktorantki wprowadzenie tej modyfikacji może w sposób istotny poprawić otrzymane wyniki?

Drugie pytanie również związane jest z cytowanym wcześniej zdaniem. Atomom w kompleksach przypisano ładunki cząstkowe z pola siłowego Amber ff14SB (białko) oraz AM1-BCC (niestandardowe reszty aminokwasowe oraz ligandy). Jak wiadomo istnieje wiele pól siłowych, a ładunki cząstkowe można obliczać za pomocą bardziej wyrafinowanych metod chemii kwantowej. Twórcy programu Amber sugerują stosowanie stosunkowo prostej metody HF/6-31G\* do parametryzacji niestandardowych reszt aminokwasowych oraz ligandów. Jak duży wpływ na wyniki ma dobór ładunków cząstkowych w modelu Pafnucy? Czy ich zmiana może doprowadzić do istotnej poprawy wyników?

Trzecie pytanie dotyczy metod przedstawionych w pracach 1-3. Jak wiadomo wiele ligandów aktywnych biologicznie może znajdować się w równowadze tautomerycznej, a tautomery mniejszościowe takich związków mogą mieć znaczenie przy szacowaniu stałej asocjacji. (patrz np. Park, M.; Gao, C.; Stern, H.A. *Estimating binding affinities by docking/scoring methods using variable protonation states*, Proteins Struct. Funct. & Bioinfo. (2011), 79, 304). Czy zaprezentowane modele w jakikolwiek sposób uwzględniają problem równowagi tautomerycznej? Jeśli nie, to czy w ocenie Pani Doktorantki dałoby się w jakiś sposób uwzględnić ten efekt w proponowanych modelach i czy mogłoby to w sposób istotny wpłynąć na poprawę uzyskanych wyników?

Oceniając pozostały dorobek Pani Doktorantki należy podkreślić, że poza pracami z rozprawy doktorskiej Pani Doktorantka jest współautorem w czterech publikacjach z listy JCR oraz jednego streszczenia pokonferencyjnego. Wszystkie prace, w których współautorem jest Pani Doktorantka cytowane były 74 razy (71 bez autocytowań), a jej index H wynosi 4. Uważam, że są to bardzo dobre wyniki jak na ten etap kariery naukowej Pani Doktorantki.

Podsumowując, uważam, że cel rozprawy został przez Panią Doktorantkę zrealizowany, a opracowane metody z pewnością będą stosowane przez naukowców zajmujących się szeroko pojętym modelowaniem molekularnym, a zwłaszcza projektowaniem nowych małowcząsteczkowych związków biologicznie aktywnych. Pokazano wyższość opracowanych metod nad innymi uznanymi metodami, co niewątpliwie stanowi



UNIwersytet  
WARMIŃSKO-MAZURSKI W OLSZTYNIE

dr hab. **Maciej Maciejczyk**, prof. UWM

Katedra Fizyki i Biofizyki

znaczący sukces naukowy zarówno Pani Doktorantki, jak również Pana Promotora. Rozprawa doktorska Pani mgr. Marty Stępniewskiej-Dziubińskiej zawiera wiele nowatorskich rozwiązań i pomysłów oraz potwierdza jej umiejętność prowadzenia badań naukowych.

W mojej ocenie praca spełnia warunki stawiane rozprawom doktorskim określone w Ustawie o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (art. 13 ustawy z dnia 14 marca 2003 nr 65 poz. 595 z późniejszymi zmianami). W związku z tym zwracam się do Rady Naukowej Instytutu Biochemii i Biofizyki PAN o dopuszczenie mgr. Marty Stępniewskiej-Dziubińskiej do dalszych etapów przewodu doktorskiego. Ze względu na znaczący wkład prac składających się na przedstawioną rozprawę doktorską w rozwój bioinformatycznych metod projektowania nowych leków wnoszę również o wyróżnienie przedstawionej rozprawy doktorskiej.

Z wyrazami szacunku

UNIwersytet WARMIŃSKO-MAZURSKI W OLSZTYNIE  
Katedra Fizyki i Biofizyki

ul. Michała Oczapowskiego 4, 10-718 Olsztyn  
tel. (89) 523 32 34    maciej.maciejczyk@uwm.edu.pl  
fax (89) 523 38 61    www.uwm.edu.pl/wnz/node/508