



Dr hab. Alina Plenis

KATEDRA I ZAKŁAD CHEMII ANALITYCZNEJ

Gdański Uniwersytet Medyczny

tel. (48) (58) 349 10 96 fax: (48) (58) 349 15 24

ul. Gen. J. Hallera 107, PL 80-416 Gdańsk

e-mail: aplenis@gumed.edu.pl

Gdańsk, dnia 21.07.2023 r.

Recenzja rozprawy doktorskiej mgr. Mariusza Zapadki pt. „Analiza i interpretacja wybranych deskryptorów molekularnych w modelowaniu zależności struktura-aktywność/właściwości fizykochemiczne”

będąca podstawą do ubiegania się o stopień doktora nauk medycznych i nauk o zdrowiu w dyscyplinie nauki farmaceutyczne pracownika Katedry i Zakładu Chemii Nieorganicznej i Analitycznej Wydziału Farmaceutycznego Collegium Medicum w Bydgoszczy, UMK w Toruniu

Promotor: dr hab. n. farm Bogumiła Kupcewicz, prof. UMK

W ostatnich latach odnotowywany jest systematyczny wzrost zainteresowania naukowców metodami chemoinformatycznymi, które dotyczą badań w zakresie projektowania leków o określonym kierunku działania terapeutycznego, przewidywania ich toksyczności, istotnych właściwości fizykochemicznych bądź farmakokinetycznych czy też prowadzenia wirtualnych badań przesiewowych ligandów *in silico*. Podstawą tych metod jest założenie, że właściwości biologiczne i fizykochemiczne substancji chemicznej wynikają z jej struktury molekularnej, a każdej zmianie w składzie chemicznym można przyporządkować odpowiadającą jej zmianę aktywności biologicznej/fizykochemicznej. Powiązanie wiedzy teoretycznej pozyskanej na podstawie analizy budowy cząsteczek opisanej za pomocą deskryptorów molekularnych z eksperymentalnie określonymi aktywnościami biologicznymi / właściwościami fizykochemicznymi związków chemicznych pozwala wskazywać wzajemne zależności, a to umożliwia predykcję nowych związków o pożądanym cechach biologicznych czy fizykochemicznych, bez konieczności prowadzenia długotrwałych oraz kosztownych badań eksperymentalnych. Przykładem mogą być ilościowe zależności struktura-aktywność (QSAR), które należą do modeli matematycznych stosowanych do przewidywania aktywności biologicznej projektowanych substancji chemicznych na podstawie budowy chemicznej podobnych związków, których aktywność została oceniona eksperymentalnie. Akronim QSPR jest używany wówczas, gdy modelowana jest jakakolwiek właściwość inna niż aktywność biologiczna, np. struktura – (chromatograficzna) retencja, której to podstawy teoretyczne opracował prof. Roman Kaliszan, wybitny naukowiec wywodzący się z Gdańskiego Uniwersytetu Medycznego. Pomimo rosnącego zainteresowania interpretacją modeli QSAR/QSPR, uzyskiwanie strukturalnie użytecznej wiedzy z modeli obliczeniowych jest jednak nadal dużym wyzwaniem. Dotychczas, w literaturze naukowej opisano ponad 5000 deskryptorów molekularnych, a mimo to znalezienie odpowiedniej kombinacji deskryptorów molekularnych w postaci ogólnej funkcji dla stworzenia modelu o przejrzystym, jasno wytłumaczalnym związku pomiędzy strukturą a aktywnością/właściwością jest

trudne, pomimo coraz większej dostępności danych doświadczalnych oraz rozwoju metod sztucznej inteligencji. Opracowany model QSAR/QSPR powinien mieć jasno zdefiniowany punkt końcowy, jednoznaczny algorytm, zdefiniowany obszar/dziedzinę zastosowania, właściwe miary dopasowania, odporności i predykcji, a także interpretację mechanistyczną, o ile jest to możliwe. W tych analizach, interpretację modeli QSAR/QSPR można oprzeć na dwóch podejściach: model→deskryptory→struktura oraz model→struktura, przy czym pierwsze podejście jest częściej stosowane.

W powyższy nurt badań wpisuje się przedłożona do oceny praca doktorska mgr. Mariusza Zapadki, której celem była analiza i interpretacja wybranych deskryptorów molekularnych klas RDF(R), GATSk(w), GETAWAY oraz 3D-MoRSE w modelowaniu zależności struktura-aktywność/właściwości fizykochemiczne kilku wytypowanych grup związków biologicznie aktywnych. Wybór deskryptorów należących do wyżej wymienionych klas wynikał z pogłębionej analizy danych literaturowych, które potwierdzały ich przydatność w modelowaniu różnorodnych punktów końcowych.

W części teoretycznej dysertacji Doktorant przedstawił ogólne pojęcia dotyczące deskryptorów molekularnych i dualizmu QSAR-u, a także regulacje prawne i zalecenia OECD w odniesieniu do QSAR oraz scharakteryzował metody interpretacji modeli QSAR.

Następnie, przedstawił wspomniany powyżej nadrzędny cel pracy badawczej. Jego realizację podzielił na pięć projektów (celów szczegółowych):

- Budowa interpretowalnego modelu zależności struktura-aktywność przeciwgrzybicza przeciwko *Candida albicans* (ATCC 2091) dla wybranej grupy pochodnych tiazolu o strukturach chemicznych oraz aktywnościach biologicznych zaczerpniętych z danych literaturowych (Model 1),
- Ocena przydatności deskryptora molekularnego HATS5m jako zmiennej zależnej w modelu zależności struktura-aktywność antytrypanosomalna grupy pochodnych 4-tiazolidynonu (132 związki), których struktury oraz aktywność biologiczna także została zaczerpnięta z literatury,
- Ocena lipofilowości flawonoidów opisaną wartością log kw na podstawie modelowania ilościowych zależności struktura-retencja (QSRR) w oparciu o zastosowanie metody wielokrotnej regresji liniowej (Model 2 i 3),
- Opracowanie podejścia substrukturalnego do obliczeń deskryptorów molekularnych,
- Opracowanie metod wizualizacji deskryptorów molekularnych w celu usprawnienia procesu interpretacji modeli.

W części *Materiały i metody* Doktorant opisał badane związki chemiczne, a także oprogramowania, którymi posługiwał się w trakcie realizacji pracy doktorskiej (Tabela 8), z uwzględnieniem oprogramowania InterDescPy, które posłużyło do interpretacji deskryptorów klasy RDF(R), GETAWAY, 3D-MoRSE zastosowanych do oceny aktywności biologicznej i właściwości fizykochemicznych, a powstało we współpracy z mgr. Przemysławem Dekowskim, pracownikiem Działu Nowych Technologii firmy Softmaks.pl.

Kolejne rozdziały opisują metodologię obliczeń i interpretacji deskryptorów w Modelach 1-3, przy czym każda procedura została szczegółowo opisana i zilustrowana graficznie w postaci schematu/ów. To bardzo dobrze opracowana część pracy doktorskiej, która została dodatkowo rozszerzona o Załączniki 1-5, wnoszące dodatkowe informacje o przebiegu prowadzonych eksperymentów badawczych.

Następnie, Doktorant przedstawił wyniki badań prowadzonych zgodnie z procedurami opisanymi w Rozdziale *Metody i materiały*, zgodnie z kolejnością wyznaczonych celów szczegółowych.

Praca ta zawiera także streszczenie w języku polskim i angielskim, piśmiennictwo oraz spis rycin i tabel, a także życiorys zawodowy Doktoranta.

Aby realizować pierwszy cel badawczy Doktorant wykorzystał dostępne dane literaturowe opisujące struktury chemiczne oraz aktywność przeciwgrzybiczą przeciw *Candida albicans* (ATCC 2091), wyrażoną jako logarytm minimalnego stężenia hamującego (pMIC) pochodnych tiazolu. Związki te podzielił na różne klasy aktywności, zgodnie z zakresami wartości pMIC, przyjmując pMIC jako zmienną zależną w Modelu 1. Procedurę tę podzielił na kilka etapów, które dotyczyły: modelowania molekularnego – optymalizacji geometrii badanych związków (1); obliczania i wstępnej obróbki macierzy zmiennych niezależnych: deskryptorów molekularnych (2); przygotowania macierzy zmiennej zależnej, pMIC (3); opracowania i walidacji modelu QSAR (4); interpretacji deskryptorów molekularnych (5). Wynikiem tych prac było zbudowanie przez Doktoranta modelu QSAR, opartego na algorytmie genetycznym oraz wielokrotnej regresji liniowej (ang. *genetic algorithm - multiple linear regression*, GA-MLR), którego parametry walidacyjne przedstawił w Rozdziale 3.3 i omówił w Rozdziale 4.1. W powyższym modelu, zmiennymi niezależnymi było pięć deskryptorów molekularnych (RDF100e, I_{TH} , $R4+(m)$, RDF120s i GATS8e), gdzie zmienną zależną (modelowaną) był wspomniany powyżej pMIC (Równanie 20). Wszystkie współczynniki regresji w Modelu 1 były ujemne skorelowane z pMIC, wobec czego dodatnie wartości standaryzowanych deskryptorów molekularnych powodowały obniżenie wartości pMIC. Powyższy model pozwalał wyjaśnić zmienność pMIC na poziomie 85,14%, a uzyskane wyniki oceniające jakość tego Modelu 1 dowiodły, że stanowił kompromis pomiędzy akceptowalnym dopasowaniem modelu do danych eksperymentalnych i dobrymi własnościami predykcyjnymi. Do mechanistycznej interpretacji deskryptorów molekularnych w Modelu 1 (I_{TH} , $R4+(m)$, RDF100e, RDF120s i GATS8e) Doktorant użył programu InterDescPy, a uzyskane informacje opisał w Rozdziale 4.1.1. Cztery deskryptory tj. RDF100e, RDF120s, I_{TH} , $R4+(m)$ kodowały strukturę trójwymiarową cząsteczek, podczas gdy GATS8e charakteryzował dwuwymiarową przestrzeń pochodnych tiazoli. Ponadto, trzy deskryptory (RDF100e, RDF120s, GATS8e) wykorzystywały strukturę elektronową atomów do opisu ich właściwości fizykochemicznych. Dodatkowo, RDF100e oraz GATS8e uwzględniały tendencję do przyciągania elektronów przez poszczególne atomy związku chemicznego (elektroujemność), a RDF120s uwzględniał także topologię cząsteczki. Dla odmiany, deskryptor I_{TH} nie opisywał właściwości fizykochemicznych atomów w cząsteczce, podczas gdy wyłącznie w przypadku $R4+(m)$ stosowana była masa atomowa. W powyższych obliczeniach, struktury pochodnych tiazolu zostały podzielone na cztery części, tj. wspólny rdzeń (F1, (Φ)) oraz trzy fragmenty oznaczone jako F2, F3 oraz F4, wyróżnione kolorystycznie.

Interpretację RDF100e Doktorant przeprowadził z wykorzystaniem trzech podejść: analizy rozmieszczenia i istotności par atomowych (1), metody substrukturalnej (2) oraz analizy powierzchni molekularnej zmapowanej wartościami kumulatywnych wkładów poszczególnych atomów (Ω) (3). W pierwszym podejściu interpretacyjnym analizie podlegały wkłady poszczególnych par atomowych (γ) w wartość deskryptora, natomiast w drugim sposobie sumy wkładów par atomowych należących do wyodrębnionych fragmentów strukturalnych, czyli wkłady wewnątrz-fragmentowe (ϵ). Ponadto, Doktorant wykonał analizę wkładów między-fragmentowych (δ), a także określał wkład i -tego atomu (Ω_i), będącego sumą wkładów par atomowych (γ), których elementem jest i -ty atom. Szczegółowe wyjaśnienie przyjętych podejść interpretacyjnych przyjętych przez Doktoranta zostało przedstawione w Rozdziałach 3.3.1 i 3.5, a opis uzyskanych wyników w Rozdziale 4.1.2. Metodologia obliczania i

interpretacji $R4+(m)$ została przedstawiona na Schemacie 3 (Rozdział 3.3.2) i polegała na identyfikacji kluczowych par atomowych o największym wkładzie w wartość deskryptora $R4+(m)$, a także podejściu „analiza rozmieszczenia i istotności par atomowych”. Wyniki tych badań Doktorant opisał w Rozdziale 4.1.4. Analiza pozostałych dwóch zmiennych niezależnych (I_{TH} , $R4+(m)$, GATS8e) polegała na identyfikacji kluczowych motywów strukturalnych na podstawie oceny wpływu zmienności parametrów wejściowych na wartość deskryptora. W części eksperymentalnej Doktorant przedstawił metodologię ich obliczeń oraz interpretacji, a uzyskane wyniki szczegółowo opisał w Rozdziałach 4.1.3 i 4.1.5.

W Rozdziale 5 Doktorant przedstawił na Ryc. 52 podsumowanie zależności pomiędzy strukturą a aktywnością przeciwgrzybiczą przeciw *Candida albicans* (ATCC 2091) pochodnych tiazolu na podstawie interpretacji modelu QSAR (Model 1), a zestawienie wniosków z interpretacji modelu QSAR zgodnie z paradygmatem model→deskryptory→struktura opisał w Tabeli 24. Jak wcześniej wspomniano, wszystkie deskryptory molekularne Modelu 1 były ujemnie skorelowane z pMIC.

Realizacja drugiego celu badawczego obejmowała wyznaczenie relacji między strukturą a aktywnością antytrypanosomalną pochodnych 4-tiazolidynonu (132 struktury) na podstawie analizy deskryptora HATS5m, należącego do klasy H-GETAWAY. HATS5m określa podobieństwo geometrii, topologii i masy atomowej struktur molekularnych (analogicznie do $R4+(m)$, I_{TH}) i koduje strukturę molekularną w trojaki sposób: fizykochemiczny, topologiczny (2D) i geometryczny (3D), uwzględniając rzeczywistą geometrię układu za pomocą macierzy wycelowanych współrzędnych kartezyjskich M_w . Metodologię obliczania i interpretacji HATS5m przedstawił na Schemacie 6 i 7 (Rozdział 3.3.5). W oparciu o podejście substrukturalne i analizę istotności i rozmieszczenia par atomowych ocenił, w jaki sposób schemat ważenia atomów, funkcjonal Delta-Diraca, rozmiar oraz kształt cząsteczek wpływają na wartość HATS5m, a uzyskane wyniki omówił w Rozdziale 4.2. Interpretację deskryptora przeprowadził w oparciu o dwa podejścia: analizę wkładów par atomowych w obrębie całej cząsteczki oraz z podziałem na fragmenty (substruktury). W powyższych analizach, pochodne 4-tiazolidynonu podzielił na 5 grup w zależności od rodzaju największego wspólnego motywu strukturalnego (Φ). Następnie każda cząsteczka została podzielona na pięć fragmentów (substruktur): wspólny motyw Φ oraz podstawniki R1, R2, R3 i R4, a poszczególne fragmenty molekularne wyróżnił innym kolorem. Wpływ sposobu wyboru schematu ważenia na wartość deskryptora określił na podstawie analizy wpływu masy atomowej na przykładzie wybranych czterech związków Les1, Les2, Les3 i Les4, których cząsteczki zostały pocięte w taki sposób, aby uzyskać porównywalne wielkością substruktury. Interpretacji deskryptorów typu HATS_m dokonał na podstawie identyfikacji par atomowych o największym wkładzie (γ) w wartość deskryptora oraz analizie porównawczej wkładów wewnątrz- (ϵ) i między-fragmentowych (δ). W kolejnych etapach badań sprawdził także wpływ rodzaju schematu ważenia tj. objętości van der Waalsa (v), polaryzowalności (p), masy atomowej (m), elektrojemności (e), potencjału jonizacji (i) na wartość HATS5(w). Analizę tę przeprowadził na przykładzie substruktury 4 (Załącznik 1). Ta sama substruktura (4) została także wykorzystana do oceny wpływu funkcjonu Delta Diraca na wartość HATS_k(m), podczas gdy związek Les3 posłużył do oceny wpływu wielkości cząsteczki na wartość HATS5 m . Ponadto, na przykładzie serii dwudziestu konformerów związku Les5 (Les-3834) określił przydatność HATS5 m do oceny podobieństwa/różnorodności molekularnej pochodnych 4-tiazolidynonu. Dodatkowo, zestaw 132 źródłowych pochodnych 4-tiazolidynonu Doktorant poddał modyfikacji w taki sposób, aby nowe cząsteczki zawierały wspólny motyw strukturalny i tylko jeden podstawnik. Takie podejście pozwoliło scharakteryzować wpływ podstawników (R1, R2 i R4)

na wartość HATS5m, uwzględniając zarówno wkłady wewnątrz- jak i między fragmentowe. To pozwoliło uszeregować podstawniki pod względem ich wpływu na wartość HATS5m.

Wynikiem tych badań było wytypowanie dla grupy pochodnych 4-tiazolidynonu elementów strukturalnych odpowiedzialnych za wzrost/spadek wartości deskryptora HATS5m, przy czym Doktorant wykazał, że zarówno niskie jak i wysokie wartości HATS5m były związane ze wzrostem aktywności antytrypanosomalnej badanych pochodnych 4-tiazolidynonu. Jednakże, interpretacja HATS5m nie była łatwa, gdyż narzucała konieczność jednoczesnego uwzględnienia trzech składowych: schematu ważenia, wskaźników wpływu oraz funkcjonału Delta-Diraca.

Podsumowanie zależności pomiędzy strukturą a wartością deskryptora HATS5m pochodnych 4-tiazolidynonu o aktywności antytrypanosomalnej Doktorant zestawiał w Rozdziale 5 na Ryc. 53, a zestawienie wniosków z interpretacji deskryptora HATS5m opisał w Tabeli 25.*

Trzeci z postawionych celów badawczych dotyczył oceny lipofilowości flawonoidów opisanej wartością $\log k_w$ na podstawie modelowania ilościowych zależności struktura-retencja (QSRR) określonych metodą wielokrotnej regresji liniowej (Model 2 i 3). Realizacją tego zadania wymagała wyznaczenia logarytmu dziesiętnego współczynnika retencji, ekstrapolowanego do 100% wodnej fazy ruchomej ($\log k_w$) dla grupy badanych flawonoidów. Parametr ten określono na podstawie pomiarów chromatograficznych techniką wysokosprawnej chromatografii cieczowej w odwróconym układzie faz (RP-HPLC). Pomiarzy przeprowadzono z wykorzystaniem dwóch kolumn, Synergi Polar-RP oraz Synergi Fusion-RP, różniących się powierzchnią fazy stacjonarnej a tym samym charakterem oddziaływań międzycząsteczkowych w układzie chromatograficznym. W oparciu o szereg parametrów walidacyjnych Doktorant wybrał dwa najwyżej ocenione modele QSRR. Dla kolumny POLAR-RP-80A był to model z trzema deskryptorami molekularnymi: HBD, Mor13v oraz MlogP2 (Model 2, Równanie 21). W przypadku kolumny FUSION-RP-80A, najlepszym okazał się model oparty o cztery deskryptory molekularne (Mor13u, Mor27v, ISH oraz RDF130p) (Model 3, Równanie 22) (Rozdział 4.3).

Zarówno dla Modeli 2 jak i 3, metodologia opracowania modelu QSPR obejmowała cztery etapy: specyfikację modelu (1); estymację parametrów (2); weryfikację modelu (3); walidację modelu (4). Decyzja o wyborze ostatecznej postaci modelu QSRR została podjęta przez Doktoranta na podstawie obliczonych miar dopasowania (1); walidacji wewnętrznej (2), walidacji zewnętrznej (3) oraz weryfikacji dziedziny modelu (4) (Rozdział 3.3.6). W Tabeli 12 Doktorant zestawiał wyznaczone parametry walidacyjne opracowanego modelu QSRR z zastosowaniem wielokrotnej regresji liniowej z selekcją zmiennych wprzód, a uzyskane wyniki omówił w Rozdziale 4.3. Na podstawie współczynnika determinacji modelu dla kolumny POLAR-RP-80A określił, iż zmienność $\log k_w$ była wyjaśniona na poziomie 90,2 %, natomiast dla kolumny FUSION-RP-80A opracowany model wyjaśniał 88,4 % zmienności $\log k_w$. W przypadku kolumny Fusion-RP-80A, wprowadzenie do modelu zmiennych niezależnych, które nie były ze sobą skorelowane, zapewniało poprawne oszacowanie wariancji i błędów standardowych estymatorów. Natomiast dla kolumny POLAR-RP-80A została utworzona nowa zmienna niezależna będącą ilorazem dwóch deskryptorów (HBD, Mor13v) (Równanie 21). Ten zabieg pozwolił uniknąć problemu współliniowości zmiennych niezależnych przy jednoczesnym zachowaniu informacji chemicznej. Ponadto, Doktorant wskazał na dobrą jakość predykcji Modelu 2 (POLAR-RP-80A) oraz umiarkowaną Modelu 3 (FUSION-RP-80A). Badania dowiodły także, że struktura molekularna niektórych związków ma istotny wpływ na parametry modelu dla kolumny POLAR, a względny błąd predykcji modelu dla tej kolumny jest znacznie większy w porównaniu do FUSION-RP-80A.

Nasuwa się zatem pytanie, czy utworzenie w Modelu 1 zmiennej niezależnej będącą ilorzem dwóch deskryptorów (HBD, Mor13v) oraz jednoczesne oparcie tego modelu na Mor13v jako kolejnej zmiennej niezależnej było w pełni uzasadnione i nie wpływało negatywnie (przeszacowanie wpływu jednego z deskryptorów) na jakość tego modelu?

Jak wspomniano powyżej, Model 2 opracowany dla kolumny POLAR-RP-80A był oparty na trzech deskryptorach tj. HBD, Mor13v oraz MlogP2. Deskryptory HBD oraz MlogP2 należą do kategorii właściwości molekularnych. HBD jest miarą zdolności cząsteczki do tworzenia wiązań wodorowych i wyraża liczbę możliwych donorów wiązań wodorowych. MlogP2 jest podniesionym do kwadratu obliczeniowym współczynnikiem podziału $\log P$, którego wartość charakteryzuje oddziaływania hydrofobowe pomiędzy flawonoidami a powierzchnią fazy stacjonarnej (im wyższa wartość tym oddziaływania są silniejsze). Sposób ich wyznaczenia Doktorant opisał w Rozdziale 3.3.7. Deskryptor Mor13v opisywał rozmieszczenie atomów w trójwymiarowym układzie współrzędnych kartezjańskich, a jego wartość była wynikiem sumowania wkładów wszystkich par atomowych cząsteczki (γ). W oparciu o metodykę obliczeń deskryptorów klasy 3D-MoRSE przedstawioną w Rozdziale 3.3.9 Doktorant zdekomponował wartości Mor13v analizowanych struktur na wkłady pary atomowych ($\gamma + i$ i $\gamma -$) oraz wkłady wewnątrz- (ϵ) i między-fragmentowe (δ). Następnie, identyfikował zestawy par atomowych o najmniejszych i największych wartościach wkładów (γ) w wartość deskryptora. Wyznaczona dodatnia wartość współczynnika regresji Mor13v wskazała, że wraz ze wzrostem wartości deskryptora rośnie wartość $\log kw$.

W przypadku kolumny FUSION-RP-80A, model QSRR zawierał cztery deskryptory molekularne (RDF130p, Mor13u, Mor27v oraz I_{SH}), przy czym cztery z nich: Mor13u, I_{SH}, Mor27m oraz RDF130p kodowały geometrię przestrzenną flawonoidów. RDF130p opisywał trójwymiarową strukturę flawonoidów na podstawie zbioru atomów oddalonych od siebie na odległość wynoszącą w przybliżeniu 13 Å, a do opisu fizykochemicznego flawonoidów wykorzystywał polaryzowalność, przez co odzwierciedlał wpływ rodzaju reszty cukrowej i wiązania glikozydowego na wartość $\log kw$ flawonoidów. W pracy doktorskiej interpretacja RDF130p została oparta na przykładzie analizy roifoliny i izoroifoliny. Deskryptor Mor13u nie uwzględniał właściwości fizykochemicznych atomów, ale pomimo różnic w zastosowanych wagach atomów interpretacja Mor13u prowadziła do analogicznych wniosków jak Mor13v. Kolejnym deskryptorem molekularnym klasy 3D-MoRSE był Mor27m, którego interpretacji Doktorant dokonał także na przykładzie roifoliny i izoroifoliny. Na podstawie wyników tych analiz przedstawił, w jaki sposób Mor27m koduje grupy hydroksylowe, reszty cukrowe, przestrzenne ułożenie aglikonu i reszty cukrowej oraz wzajemne relacje przestrzenne pomiędzy podjednostkami reszty cukrowej. Deskryptor I_{SH} należy do deskryptorów geometrycznych GETAWAY i koduje informację o symetrii struktur molekularnych, ale jego wartość nie jest determinowana wielkością struktury molekularnej. Dla wyznaczonego Modelu 3, wartości trzech deskryptorów tj. I_{SH}, Mor27m oraz RDF130p były dodatnio skorelowane z wartością $\log kw$, podczas gdy Mor13u był skorelowany ujemnie. W pracy doktorskiej została szczegółowo opisana metodologia obliczeń wszystkich deskryptorów ujętych w Modelu 3, przy czym metodologia obliczeń deskryptorów klasy 3D-MoRSE na przykładzie Mor13v mgr Mariusz Zapadka opisał także na Schematach 8, 9 i 10 (Rozdział 3.3.9). Omówienie wyników dotyczących interpretacji wszystkich wyżej wymienionych deskryptorów przedstawił w Rozdziałach 4.3.1-4.3.4.

Reasumując, Doktorant wykazał, iż analiza QSRR wraz z interpretacją deskryptorów molekularnych pozwala na wskazanie elementów w strukturze flawonoidów odpowiedzialnych za ich lipofilowość opisaną wartością $\log kw$. Sumaryczne zestawienie zależności pomiędzy strukturą a $\log kw$ flawonoidów na podstawie interpretacji modeli QSAR (Model 2 i 3 (nie 1 i 2)) zostało

przedstawione na Ryc. 54, zestawienie wniosków z interpretacji deskryptorów Modelu 2 i 3 (nie 1 i 2) zawiera Tabela 26 (Rozdział 5).

Celem badań Doktoranta było także opracowanie metody opartej na podziale cząsteczek analizowanych związków na nienakładające się fragmenty strukturalne (metoda substrukturalna) i wykorzystanie tego podejścia do interpretacji deskryptorów. Metoda substrukturalna pozwalała Doktorantowi ocenić wkład poszczególnych fragmentów cząsteczki w wartość deskryptora oraz znaczenie ich wzajemnego rozmieszczenia w przestrzeni. Tym samym, możliwe było uzyskanie odpowiedzi na pytanie, jakie fragmenty strukturalne badanych związków w największym stopniu wpływają na wartości deskryptorów.

Kolejnym kierunkiem badań było opracowania metodologii obliczeń deskryptorów molekularnych połączonej z dekompozycją ich wartości na wkłady par atomowych. Takie podejście umożliwiało interpretację deskryptorów molekularnych na różnych poziomach szczegółowości. Najbardziej ogólną metodą interpretacyjną było przedstawienie wartości deskryptora jako wkłady wewnątrz- (ϵ) i między-fragmentowe (δ), podczas gdy najbardziej szczegółową była analiza macierzy wkładów par atomowych (γ_i).

Wyżej opisane podejścia Doktorant zastosował do interpretacji deskryptorów klas RDF(R), GETAWAY oraz 3D-MoRSE. Metodę obliczeń deskryptorów klasy RDF(R) na przykładzie RDF100e (Model 1) opartą na analizie rozmieszczenia i istotności par atomowych Doktorant przedstawił na Schemacie 1, podczas gdy Schemat 2 ilustruje metodę interpretacji deskryptorów klasy RDF(R) z wykorzystaniem rozmieszczenia i istotności par atomowych oraz metody substrukturalnej (Rozdział 3.3.1). Metodę obliczeń HATS5m w przejrzysty sposób opisał na Schemacie 6; sposób jego interpretacji z wykorzystaniem rozmieszczenia i istotności par atomowych oraz metody substrukturalnej przedstawił na Schemacie 7 (Rozdział 3.3.1) i omówił w Rozdziale 4.2. W przypadku deskryptorów bazujących na dyfrakcji elektronowej (3D-MoRSE), ich metodę obliczania Doktorant przedstawił w Rozdziale 3.3.9. Na przykładzie Mor13v wyjaśnił sposób ich wyznaczania (Schemat 8) oraz interpretacji z użyciem analizy rozmieszczenia i istotności par atomowych (Schemat 9) oraz podejścia substrukturalnego (Schemat 10).

Dodatkowo, mgr Mariusz Zapadka zaproponował zastosowanie metod interpretacji deskryptorów za pomocą powierzchni molekularnych mapowanych wartościami kumulatywnych wkładów atomów (Ω), które łączyło w sobie informacje z macierzy wkładów poszczególnych par atomowych z podejściem substrukturalnym do budowy cząsteczki. Metodyka analizy powierzchni molekularnych zmapowanych wartościami kumulatywnych wkładów atomowych została opisana w Rozdziale 3.5. Wynikiem jej zastosowania było przedstawienie struktury analizowanego związku w postaci powierzchni molekularnej, którą mapuje się wartościami wkładów kumulatywnych zgodnie z przyjętą skalą barw. To pozwalało intuicyjnie określić wpływ poszczególnych atomów na wartość deskryptora, a przez to dosyć dokładnie wskazać fragmenty struktury związku chemicznego o małym, średnim i dużym wkładzie w wartość danego deskryptora. Wyżej opisane podejście także zastosował do interpretacji wybranych deskryptorów klas RDF(R), GETAWAY oraz 3D-MoRSE.

Wyniki analiz z użyciem wszystkich wyżej wymienionych metod wizualizacji oraz uzyskane informacje z interpretacji tych wybranych deskryptorów Doktorant opisał w Rozdziałach: 4.1.2 (RDF100e, RDF120s), 4.2.1-4.2.5 (HATS5m), 4.3.1 (Mor13v) oraz 4.3.3 (RDF130p). Wyniki tych badań dowiodły, że zaprezentowane w rozprawie doktorskiej sposoby interpretacji deskryptorów molekularnych mogą być przydatne do przeprowadzenia walidacji modeli QSAR oraz lepszego

zrozumienia złożonej natury procesów biologicznych czy fizykochemicznych. Jednakże nasuwają się następujące pytania:

- 1) Które z proponowanych rozwiązań Doktorant uważa za najbardziej optymalne do wizualizacji i interpretacji deskryptorów molekularnych?
- 2) Czy uzyskane wyniki zależnie od użytej metody znacząco się różniły?
- 3) Jakie były wady i zalety poszczególnych podejść metodycznych?

Cennym byłoby wskazanie konkretnych przykładów przy udzielaniu odpowiedzi na te pytania.

Jednakże należy zaznaczyć, że część wynikowa pracy doktorskiej mgr. Mariusza Zapadki jest bardzo dobrze udokumentowana, a jej cennym uzupełnieniem jest szereg tabel, rycin, wykresów czy filmów, które są zamieszczone w części dysertacji, jak i dołączone w formie Załączników 1-5. To dowodzi, że wnioski płynące z tych badań mają solidną podstawę eksperymentalną.

Podsumowując, na podstawie przeprowadzonych badań Doktorant sformułował wnioski, które są w pełni poparte danymi eksperymentalnymi, a założone cele badawcze zostały w pełni zrealizowane. Wprawdzie Autorowi nie udało się uniknąć niektórych błędów edytorskich w trakcie edycji dysertacji, ale te drobne potknięcia nie umniejszają jej wartości naukowej. Na wyróżnienie zasługuje bardzo przejrzysty sposób prezentacji danych, wzbogacony o szereg elementów graficznych umożliwiających wizualizację wyników oraz wspomagających proces wnioskowania. Część wyników uzyskanych w trakcie realizacji niniejszej pracy została już opublikowana w dwóch czasopismach z listy JCR (*International Journal of Molecular Sciences* oraz *Journal of Pharmaceutical and Biomedical Analysis*) o łącznym współczynniku oddziaływania IF wynoszącym 9,417. Doktorant jest pierwszym autorem tych prac, a oświadczenia współautorów tych publikacji wskazują, że udział Doktoranta w ich powstaniu był znaczący. Wymiernym efektem realizacji tego projektu było także opracowanie we współpracy z mgr Przemysławem Dekowskim, oprogramowania InterDescPy o nowatorskich rozwiązaniach, które może stanowić cenne narzędzi do zastosowań chemioinformatycznych w zakresie nauk farmaceutycznych.

Doktorant nabył duże doświadczenie w prowadzeniu badań naukowych jako pracownika Katedry i Zakładu Chemii Nieorganicznej i Analitycznej Wydziału Farmaceutycznego CM w Bydgoszczy, UMK w Toruniu, które świadczą o wysokich kompetencjach, umiejętnościach i wiedzy mgr. Mariusza Zapadki w projektowaniu i realizacji badań naukowych. Jest autorem dziewięciu publikacji naukowych o zasięgu krajowym i zagranicznym o sumarycznym współczynniku oddziaływania IF 21,315 (sumaryczna wartość punktów MNiE 685). Brał udział w licznych krajowych sympozjach i konferencjach, a ponadto jest laureatem kilku stypendiów i nagród za działalność naukową. Mgr Mariusz Zapadka jest także zaangażowany w działalność dydaktyczną i popularyzującą naukę.

Podsumowując ocenę rozprawy doktorskiej należy podkreślić wysoki poziom naukowy badań prowadzonych przez Doktoranta oraz ich wymierny aspekt praktyczny, a dobór narzędzi i zaplanowany tok prowadzonych badań świadczą o profesjonalizmie Doktoranta.

W mojej ocenie uzyskane wyniki badań dowodzą, iż rozprawa doktorska mgr. Mariusza Zapadki spełnia wymogi stawiane rozprawom doktorskim przez obowiązujące przepisy i tym samym wnoszę o jej dopuszczenie do dalszych etapów postępowania doktorskiego celem uzyskania stopnia doktora nauk medycznych i nauk o zdrowiu w dyscyplinie nauki farmaceutyczne.

Z uwagi na wysoki poziom prowadzonych badań i umiejętny sposób prezentacji uzyskanych wyników, a także ich znaczący aspekt praktyczny, wnioskuję o wyróżnienie przedstawionej do oceny

rozprawy doktorskiej mgr. Mariusza Zapadki. Część wyników badań zawartych w dysertacji została już opublikowana w renomowanych czasopismach z listy JCR, co dodatkowo potwierdza dojrzałość naukową mgr. Mariusza Zapadki oraz wysoką jakość zrealizowanych przez Niego badań.

Gdańsk, 21.07.2023 r.

Katedra i Zakład Chemii Analitycznej
dr hab. Alina Plenis

Kierownik