



Silesian
University
of Technology

SILESIAN UNIVERSITY OF TECHNOLOGY
FACULTY OF AUTOMATIC CONTROL, ELECTRONICS AND
COMPUTER SCIENCE

Doctoral dissertation

Algorithms for the analysis of molecular protein structures and
drug-like ligands for modeling and simulation of residence time
drug-molecular target

Author: mgr inż. Magdalena Ługowska

Supervisor: prof. dr hab. inż. Marek Kimmel

Co-supervisor: dr inż. Marcin Pacholczyk

Gliwice 2023

Streszczenie rozprawy doktorskiej

*“Algorytmy analizy struktur białkowych i ligandów lekopodobnych
w celu modelowania i symulacji czasu przebywania leku w miejscu docelowym”*

mgr. inż. Magdalena Ługowska

1. Wstęp

Oprócz odpowiedniego stężenia leku w tkance efektora, warunkiem działania leku jest jego zdolność wiązania się z receptorem, czyli najczęściej odpowiednim białkiem regulującym wiązanie komórek lub wspomagającym transdukcję sygnału (czyli powinowactwo do receptora), a także zdolność aktywowania lub blokowania receptora. Powszechnie stosowanym kryterium wyboru w procesie projektowania leków jest równowagowe powinowactwo wiązania małych cząsteczek chemicznych (ligandów) do celu molekularnego (receptora). Powinowactwo to opisuje trwałość wiązania i siłę działania ligandu z receptorem. Ważnym czynnikiem działania liganda jest jego skuteczność. Parametr ten definiuje potencjał liganda do wydajnej aktywacji receptora w celu wytworzenia silnej odpowiedzi *in vivo*. Zatem powinowactwo jest miarą ilościowej oceny skuteczności leku i określenia korzyści wynikające z interakcji zachodzących między cząsteczką leku, a jego celem. Dlatego protokoły projektowania nowych leków opierają się głównie na cząsteczkach o wysokim powinowactwie wiązania. Jednak wiele leków posiada nierównowagową charakterystykę wiązania i to podejście nie zawsze odpowiada wyższej sile działania leku w warunkach *in vivo*. W ostatnich latach wykazano, że możliwe jest przewidywanie skuteczności nowego leku w warunkach *in vivo* poprzez pomiar kinetyki wiązania, a szybkość wiązania i dysocjacji leku z receptorem są głównymi cechami, obok właściwości farmakokinetycznych, które określają profil aktywności biologicznej leku (Copeland 2016, Vauquelin 2016). Wprowadzono również pojęcie czasu przebywania leku w celu molekularnym, które uwzględnia dynamikę konformacyjną cząsteczek docelowych wpływającą na wiązanie i dysocjację leku. Istotną obserwacją jest to, że czas ten, w niektórych przypadkach, lepiej koreluje ze skutecznością leku *in vivo* niż powinowactwo wiązania (Copeland et al. 2006). Dlatego czas przebywania liganda w miejscu docelowym (τ) stał się wiarygodnym wyznacznikiem skuteczności leku, który wraz z parametrami kinetycznymi reakcji zostaje brany pod uwagę w programach odkrywania leków. Warto jednak mieć na uwadze fakt, że czas rezydencji wykorzystywany jako jedyna miara skuteczności leku może dać ograniczony obraz kinetyki reakcji (Folmer 2018).

Serię występujących po sobie reakcji biochemicznych zachodzących w komórce rozpoczyna substrat - ligand (L), czyli mała cząsteczka chemiczna, na którą działa enzym - receptor (R). Ścieżki komórkowe mają swój początek poza komórką, gdy ligand (cząsteczka będąca pierwszym bodźcem) spotyka się ze specyficznym receptorem białkowym. Cząsteczki występują w ściśle dopasowanych parach, z receptorem rozpoznającym tylko jeden (lub kilka) specyficznych ligandów, i ligandem wiążącym się tylko z jednym (lub kilkoma) docelowymi receptorami. Wiązanie ligandu z receptorem zmienia jego kształt lub aktywność, umożliwiając mu przekazanie sygnału lub bezpośrednie wywołanie zmiany w komórce. Na przykład lek Ibuprofen, jeden z najczęściej stosowanych leków przeciwbólowych, przeciwgorączkowych i przeciwzapalnych, to nieselektywny inhibitor cyklooksygenazy (COX), enzymu odpowiedzialnego za przekształcanie kwasów tłuszczowych w prostaglandyny i należy do grupy leków NLPZ (niesteroidowe leki przeciwzapalne). Prostaglandyny to substancje biorące udział w procesie zapalnym. Zahamowanie działania COX prowadzi do zablokowania prostaglandyn. Oczekuje się, że dzięki temu ograniczeniu wytwarzania prostaglandyn lek Ibuprofen będzie zmniejszał gorączkę i ból towarzyszące stanowi zapalnemu. Należy jednak wziąć pod uwagę fakt, że zablokowanie enzymu COX ma skutki uboczne. Enzym ten pełni w organizmie bardzo istotne, korzystne funkcje takie jak na przykład ochrona błony śluzowej żołądka. Dlatego długotrwałe przyjmowanie leków blokujących COX będzie wiązało się ze szkodliwym efektem m.in. dla żołądka i jelit.

Zrozumienie oraz pełny opis kinetyki wiązania receptor-ligand (RL), a także zrozumienie molekularnych uwarunkowań tego dopasowania stanowi ważny element projektowania leków.

Czas rezydencji jest miarą tego, ile czasu ligand spędza w miejscu wiążącym białka. Innymi słowy, to czas przebywania leku w danym miejscu docelowym. Lek jest aktywny farmakologicznie tak długo, jak długo pozostaje związany z receptorem. Zatem czas przebywania definiowany jest jako odwrotność stałej szybkości dysocjacji $\tau = 1/k_{off}$. Oznacza to, że stężenie ligandu nie wpływa na jego czas przebywania w celu, a leki o długim czasie przebywania mogą pozostać związane nawet, gdy ich stężenie spadnie poniżej wartości stałej równowagi dysocjacji K_d . Ma to istotne znaczenie szczególnie, gdy lek jest wydalany z organizmu, co powoduje jego zmienne stężenie in vivo. Model czasu rezydencji jest opisany przez dwustopniowy model kinetyki wiązania, który uwzględnia zmiany konformacyjne prowadzące do zwiększenia komplementarności cząsteczek. W tym modelu wolny lek napotyka swój cel w stanie konformacyjnym określonym przez kieszeń wiążącą lek, która jest nieoptymalna w stosunku do struktury cząsteczki leku. Pierwszy etap wiązania jest procesem asocjacji, w którym tworzony jest kompleks spotkań (RL*), określony przez stałą szybkości asocjacji (k_{on}), stałą szybkości dysocjacji (k_{off}) i stałą dysocjacji równowagi (K_d) (Gabdouline & Wade 1999, 2022). Po początkowym wiązaniu następuje kolejny etap, w którym system musi pokonać bariery energetyczne wytworzone na skutek zmian konformacyjnych receptora i ligandu tak, by możliwe było

utworzenie nowego stabilnego stanu (RL), w którym kieszeń wiążąca przyjmuje strukturę, która jest bardziej komplementarna do struktury cząsteczki leku.

2. Metody komputerowe do wyznaczania parametrów kinetycznych wiązania

Wraz z coraz większym zainteresowaniem czasem rezydencji i istotnością kinetyki wiązania leku w miejscu wiążącym celu, wzrasta znaczenie metod *in silico*. Szczególnie, gdy powszechnie stosowane metody eksperymentalne często są bardzo wymagające pod kątem zarówno czasowym jak i finansowym. Ponadto wykorzystanie metod komputerowych przewidujących czas przebywania jak i charakteryzujących kinetykę reakcji, może wspomóc medycynę personalizowaną. Symulacje dedykowane dla danego pacjenta mogą przyspieszyć decyzję lekarza o wyborze optymalnego z kilku potencjalnych farmaceutyków. Co więcej, takie obliczenia można przeprowadzać na związkach, które jeszcze nie zostały zsyntetyzowane, co znacznie wpływa na koszty i czas badań. Należy zauważyć, że opracowane metody *in silico* opierają się na danych doświadczalnych, co może potwierdzać ich wiarygodność.

Metody komputerowe pozwalające na oszacowanie czasu rezydencji oraz pozostałych parametrów kinetycznych można podzielić na dwie główne grupy. Pierwsza to zestaw metod dynamiki molekularnej o ulepszonym próbkowaniu. Druga to metody oparte na uczeniu maszynowym, często również wykorzystujące symulacje dynamiki molekularnej.

3. Motywacja i cel pracy

Charakterystyka i zrozumienie kinetyki reakcji dysocjacji ligandu z miejsca docelowego receptora jest ważnym elementem w procesie projektowania leków. Kluczowym narzędziem do opisanie ścieżki dysocjacji, przewidzenia parametrów kinetycznych, w tym czasu rezydencji, a także zdefiniowania determinantów strukturalnych są symulacje molekularne. Podejścia te często wykorzystują uproszczenia, takie jak zwiększone próbkowanie, co pozwala na zaobserwowanie wystąpienia rzadkich zdarzeń w czasie symulacji i zmniejszenie złożoności obliczeniowej. Zastosowanie uproszczeń powoduje, że symulacja nie jest tak szczegółowa, jak klasyczna dynamika molekularna i nie pozwala na wyjaśnienie, z wysoką dokładnością, zachowania układu zachodzącego w krótkiej skali czasowej, takich jak na przykład zmiany układu atomów w cząsteczce podczas etapu indukowanego dopasowania. Może być to traktowane jako pewne ograniczenie tych uproszczeń. Z drugiej strony stoi klasyczna dynamika molekularna, która pozwala na zrozumienie tych szybko zachodzących, ważnych zdarzeń, ale znów niemożliwe jest jej zastosowanie w dłuższych skalach czasowych, takich jak czas przebywania leku w celu, który może wynosić od kilku sekund do nawet godzin. Symulacje są też w większości przypadków informacją wejściową dla algorytmów opartych na

uczeniu maszynowym, jednak ich dokładność nie jest ustabilizowana. Motywacją przedstawionej pracy jest potrzeba opracowania i zastosowania efektywniejszych i dokładniejszych metod do analizy kinetyki wiązania receptor-ligand.

Celem pracy jest zastosowanie i zweryfikowane rozwiązań pozwalających na określenie czasu rezydencji leku w celu zbadania, czy można je stosować niezależnie od wielkości cząsteczek czy rodziny białek oraz analiza cech strukturalnych i oddziaływań w procesie wiązania w szczególności białka InhA i jego inhibitorów. Przedstawione badania mają na celu odpowiedź na następujące pytania:

- Czy istnieją i jakie są kluczowe interakcje receptor-ligand, które odróżniają ligandy o długim i krótkim czasie rezydencji?
- Czy procedura τ RAMD jest uniwersalna i może zostać zastosowana dla cząsteczek o różnej wielkości i różnym podobieństwie?
- Jaka jest korelacja między względnym czasem rezydencji, a zmierzonym eksperymentalnie?

Poza odpowiedziami na powyższe pytania, w trakcie pracy powstały dwa nowe narzędzia wspierające badania kinetyki wiązania liganda z receptorem:

- baza danych kinetycznych PDBrt dostępna publicznie pod adresem internetowym <https://pdbrt.polsl.pl/> oraz
- narzędzie do automatycznej analizy oddziaływań występujących między ligandem a receptorem w czasie symulacji.

4. Trajektorie dysocjacji liganda z miejsca wiążącego receptora metodą τ Random Acceleration Molecular Dynamics

τ RAMD to procedura rozszerzonego próbkowania symulacji dynamiki molekularnej. Została opracowana w celu obliczenia względnego czasu przebywania związków o charakterze farmakologicznym w ich celach molekularnych, a także zbadania drogi dysocjacji ligandów z miejsc wiążących receptorów (Kokh et al. 2018). Symulacje RAMD przeprowadzane są dla układów receptor-ligand zanurzonych w rozpuszczalniku, podczas których do środka masy liganda przykładana jest niewielka losowo zorientowana siła, która przyspiesza jego wyjście z miejsca aktywnego receptora. Po określonym kroku czasowym sprawdzany zostaje ruch liganda. Jeśli zmiana położenia była mniejsza niż zdefiniowana odległość progowa, następuje losowa zmiana kierunku siły. Symulacja kończy się, gdy ligand opuści miejsce wiążące receptora. Ten warunek jest zdefiniowany poprzez określenie w pliku konfiguracyjnym odległości liganda od miejsca wiążącego, która odpowiada jego uwolnieniu. Czas symulacji zależy od czasu przebywania liganda w celu. Ligandy o dłuższym czasie przebywania, dłużej opuszczają cel (czas symulacji jest dłuższy) lub wymagają

zastosowania większej siły, by opuścić cel w z góry określonym czasie symulacji. Co więcej, metoda ta nie wymaga wcześniejszej znajomości ścieżki dysocjacji ani też ekstensywnej parametryzacji. Jedynym parametrem, który należy uważnie zdefiniować jest wielkość siły, która nie powinna zaburzać obliczonych względnych czasów rezydencji. Innymi słowy należy zwrócić szczególną uwagę, by zadana siła nie była za duża, co powoduje gwałtowne wyszarpięcie ligandu z miejsca wiążącego, a tym samym niezależnie od rzeczywistej szybkości dysocjacji, estymowane wartości względnych czasów przebywania będą dla każdego ligandu przybliżone. Opisane wyżej cechy charakteryzujące metodę powodują, że τ RAMD jest nie tylko wydajnym, ale też relatywnie niewymagającym narzędziem do estymowania względnych czasów rezydencji leków w celach molekularnych.

W niniejszej pracy została wykonana analiza opublikowanych wyników przez Kokh et al. rozszerzona o cząsteczkę geldanamycin w kompleksie z HSP90. Następnie protokół powtórzono dla 11 ligandów receptora InhA i po 1 ligandzie receptorów ENR, EGFR i HIV-1. Celem analizy było sprawdzenie uniwersalności i powtarzalności metody τ RAMD.

Do przeprowadzenia symulacji τ RAMD jako struktury wyjściowe wykorzystano struktury krystalograficzne kompleksów HSP90-inhibitor, HSP90-geldanamycin, InhA-inhibitor, ENR-triclosan and EGFR-lapatinib w stanie związanym. Przygotowanie modeli molekularnych do symulacji obejmowało: protonację układów wejściowych z wykorzystaniem narzędzia PyMOL (Schrodinger & DeLano 2020), parametryzację przy użyciu pakietu AmberTools (Case et al. 2022), w tym przypisanie ładunków cząstkowe poszczególnym atomom liganda oraz zmiana typów atomów na rozpoznawalne przez pakiet Amber14, obliczenie częściowych ładunków atomowych ligandów metodą AM1-BCC (Jakalian et al. 2000, 2002), neutralizację ładunków (dodanie jonów Na⁺ lub Cl⁻) oraz zanurzenie układu w rozpuszczalniku przy użyciu wybranego pola siłowego i utworzenie plików z topologią układu. Następnie wykorzystano narzędzie pmemd do uruchomienia obliczeń minimalizacji energii, ogrzewania i ekwilibracji układów (Maier et al. 2015). Tak wygenerowane współrzędne atomów układów zostały wykorzystane jako dane wejściowe do symulacji dynamiki molekularnej przeprowadzonej za pomocą oprogramowania NAMD (Phillips et al. 2020).. Symulacje o długości 2 ns uwzględniały dynamikę Langevina dla stałej temperatury (300K) i ciśnienia (1 atm). Następnie uzyskane współrzędne atomów i ich prędkości były wykorzystane jako dane wejściowe do symulacji RAMD. Koniec symulacji, a tym samym uwolnienie liganda obserwowano, gdy odległość pomiędzy środkiem masy liganda i receptora przekroczyła 40 Å. W przypadku, gdy nie zaobserwowano uwolnienia liganda w czasie 2 ns, symulacja była zatrzymywana. Współrzędne trajektorii zapisywano co 100 fs. Dla każdego układu powtórzono 5-krotnie etapy symulacji dynamiki molekularnej z użyciem oprogramowania NAMD, które były traktowane jako pliki startowe dla τ RAMD. Z każdego pliku startowego został wygenerowany zestaw 50 trajektorii dysocjacji, co daje w sumie 250 symulacji dysocjacji dla każdego układu. Zgodnie z opublikowaną procedurą, czas rezydencji został

zdefiniowany jako czas symulacji wymagany do dysocjacji liganda w przynajmniej 50% trajektorii. Dla każdego symulowanego powtórzenia zastosowano procedurę bootstrap do obliczenia przejściowego czasu przebywania jako średniej z rozkładu (t_r). Następnie właściwy względny czas rezydencji został oszacowany jako średnia ze wszystkich symulowanych powtórzeń dla danego układu. Dla każdego układu powtórzono 5-krotnie etapy symulacji dynamiki molekularnej z użyciem oprogramowania NAMD, które były traktowane jako pliki startowe dla τ RAMD. Z każdego pliku startowego został wygenerowany zestaw 10 trajektorii dysocjacji, co daje w sumie 50 symulacji dysocjacji dla każdego układu. Zgodnie z opublikowaną procedurą, czas rezydencji został zdefiniowany jako czas symulacji wymagany do dysocjacji liganda w przynajmniej 50% trajektorii. Dla każdego symulowanego powtórzenia zastosowano procedurę bootstrap do obliczenia przejściowego czasu przebywania jako średniej z rozkładu (t_r). Następnie właściwy względny czas rezydencji został oszacowany jako średnia ze wszystkich symulowanych powtórzeń dla danego układu.

Zastosowanie dla inhibitorów białka HSP90

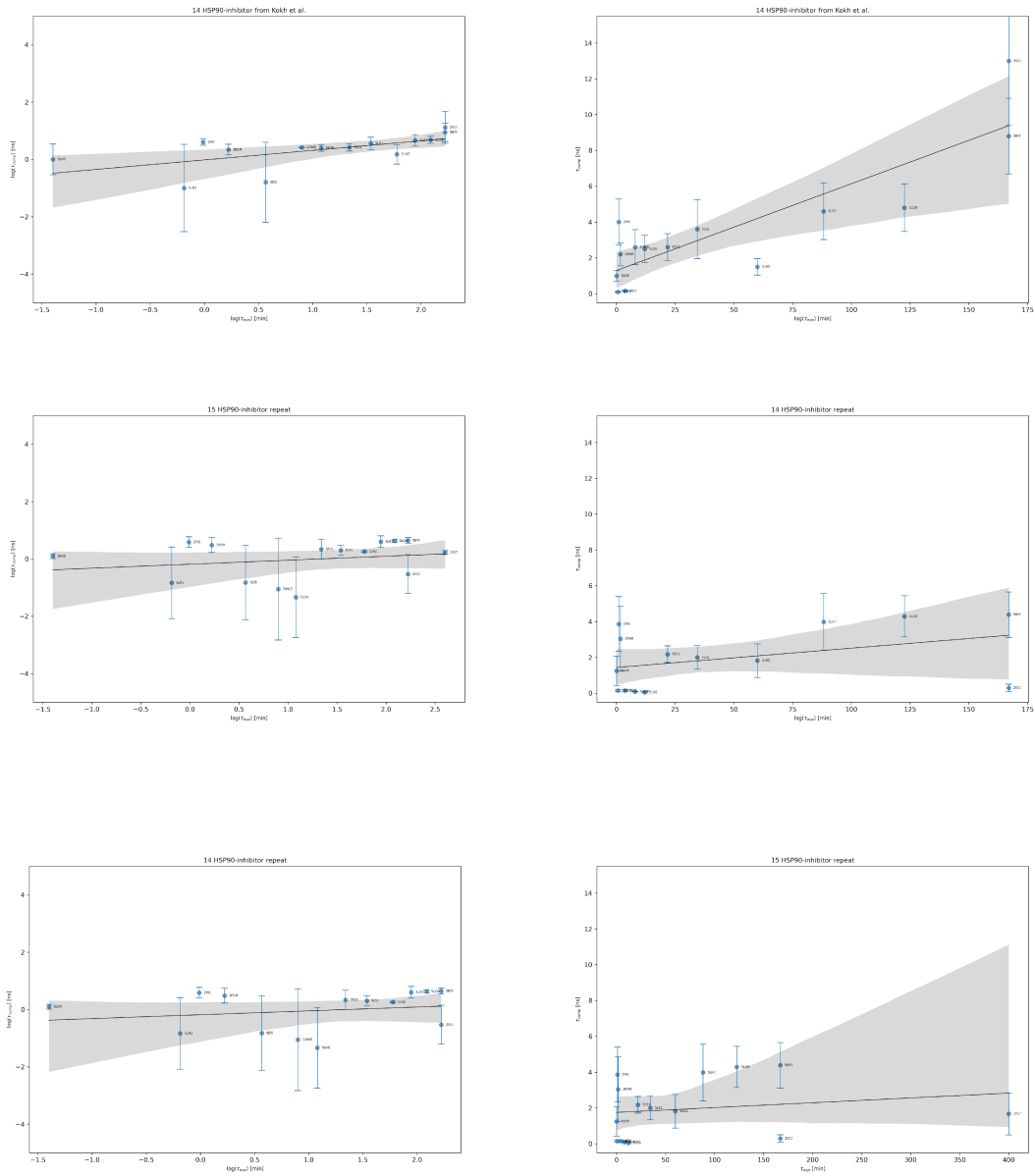
Opublikowane zastosowanie τ -RAMD dla zbioru 70 ligandów HSP90 o różnym składzie chemicznym pokazało, że dla 55 z nich istnieje silna korelacja pomiędzy średnią długością trajektorii dysocjacji (τ_{comp}) a zmierzonym eksperymentalnie czasem rezydencji. Co więcej zaobserwowano, że istnieje zależność pomiędzy τ_{comp} i eksperymentalnie wyznaczonym czasem przebywania w kongenerowych, czyli o podobnej strukturze, zestawach ligandów. Stwierdzono, że τ -RAMD jest wydajną metodą obliczeniową o szerokim zastosowaniu do poprawy czasu przebywania leku w celu.

Aby zweryfikować powtarzalność wyników, a tym samym sprawdzić wiarygodność metody, analizę z wykorzystaniem protokołu τ -RAMD przeprowadzono dla 15 inhibitorów HSP90, z których 14 znajduje się w zbiorze opublikowanych danych, a 1 nie został poddany analizie przez autorów metody. Wszystkie badane związki zostały skryształizowane w kompleksie inhibitor-HSP90, a modele molekularne są publicznie dostępne. Oznacza to, że można użyć do badań dokładnie tej samej początkowej struktury wyjściowej.

Otrzymane wyniki zostały zobrazowane na wykresie punktowym w skali logarytmicznej i zwykłej (Rys. 1) z wykonanym dopasowaniem liniowym. Czarna linia pokazuje regresję liniową wszystkich punktów poza szarym obszarem, który wskazuje region w obrębie resztkowego odchylenia standardowego dopasowania liniowego obliczonego przy progu ufności 0.95. Słupki błędów pokazują odchylenia standardowe obliczonych czasów rezydencji.

Korelacja danych została określona za pomocą współczynnika Pearsona. Dla opublikowanych wyników (Rys. 1 a,b) wynosi ona $R^2=0.86$ co wskazuje na bardzo silną zależność między danymi obliczonymi a doświadczalnymi. Dla powtórzonych symulacji dokładnie tych samych układów (Rys. 1 c,d), które zostały opublikowane, współczynnik korelacji przyjął wartość $R^2=0.4$, co sugeruje

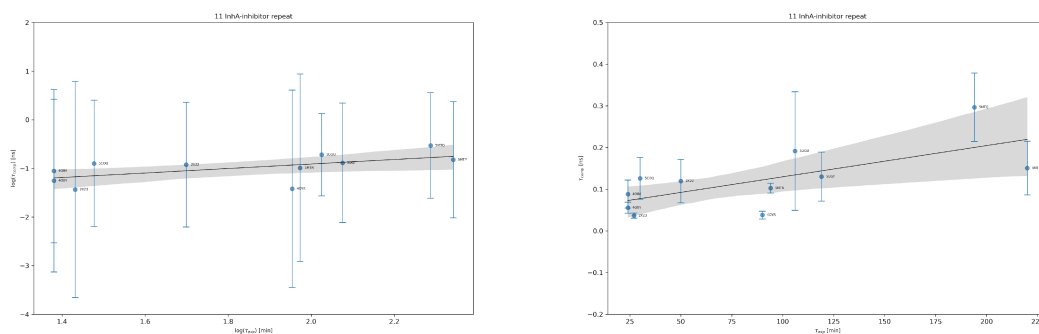
umiarkowanie pozytywną zależność. Natomiast dane z dodatkowym kompleksem (Rys. 1 e,f) można określić jako słabo skorelowane, ponieważ współczynnik korelacji wynosi $R^2=0.2$.



Rysunek 1. Wykres korelacji τ_{comp} z τ_{exp} w skali logarytmicznej (po lewej) i porządkowej (po prawej) dla (a, b) 14 inhibitorów HSP90 i opublikowanych wyników (c, d) 14 inhibitorów HSP90 i wyników powtórzenia symulacji (e, f) 15 inhibitorów HSP90.

Zastosowanie dla inhibitorów białka InhA

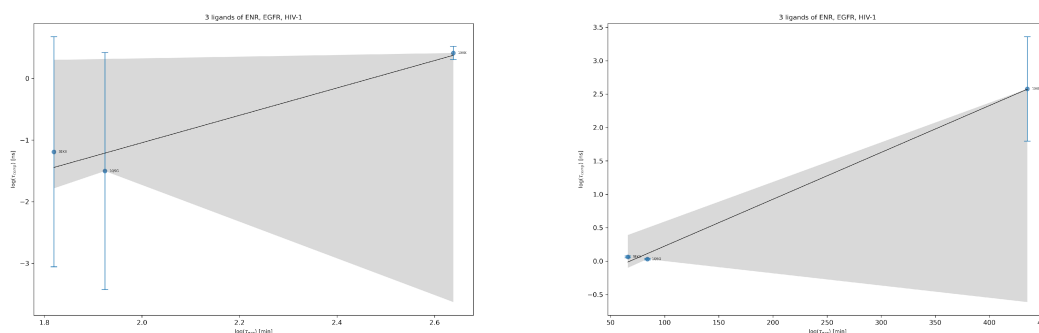
Analiza τ RAMD została przeprowadzona na zbiorze 10 ligandów białka InhA. Korelacja danych została określona za pomocą współczynnika Pearsona, którego wartość wynosi $R^2=0.68$ co wskazuje na silną zależność między danymi obliczonymi a doświadczalnymi.



Rysunek 2. Wykres korelacji τ_{comp} z τ_{exp} w skali logarytmicznej (po lewej) i porządkowej (po prawej) dla 10 inhibitorów InhA.

Zastosowanie dla ligandów białek ENR, EGFR i HIV-1

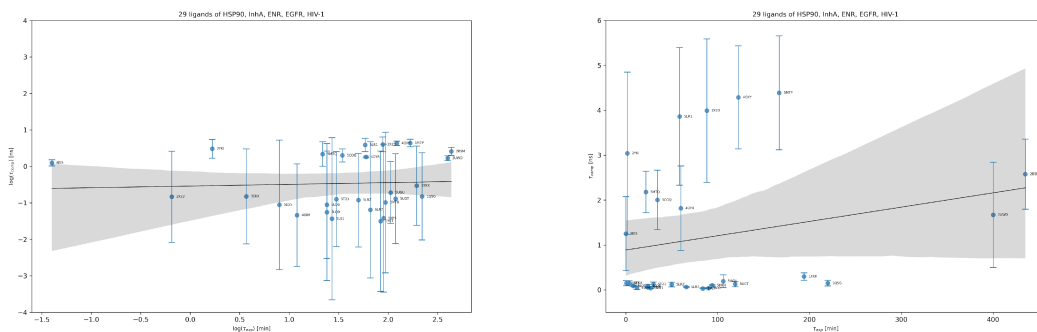
Korelacja danych została określona za pomocą współczynnika Pearsona, którego wartość wynosi $R^2=0.99$ co wskazuje na bardzo silną zależność między danymi obliczonymi a doświadczalnymi.



Rysunek 3. Wykres korelacji τ_{comp} z τ_{exp} w skali logarytmicznej (po lewej) i porządkowej (po prawej) dla 3 ligandów białek ENR, EGFR i HIV-1

Zastosowanie dla wszystkich badanych systemów

Aby sprawdzić uniwersalność metody, sprawdzono korelację τ_{comp} względem eksperymentalnie wyznaczonego czasu dla wszystkich badanych układów. Współczynnika Pearsona przyjął wartość 0.23 ($R^2=0.23$), co uznaje się za nieznaczną korelację. Można zatem stwierdzić, że dokładność metody dla związków o większym zróżnicowaniu strukturalnym jest mniejsza.



Rysunek 4. Wykres korelacji obliczonego czasu rezydencji τ_{comp} (ns) z wyznaczonym doświadczalnie τ_{exp} (min) dla odpowiednio wszystkich badanych układów.

5. Właściwości ligandu, które wpływają na czas przebywania

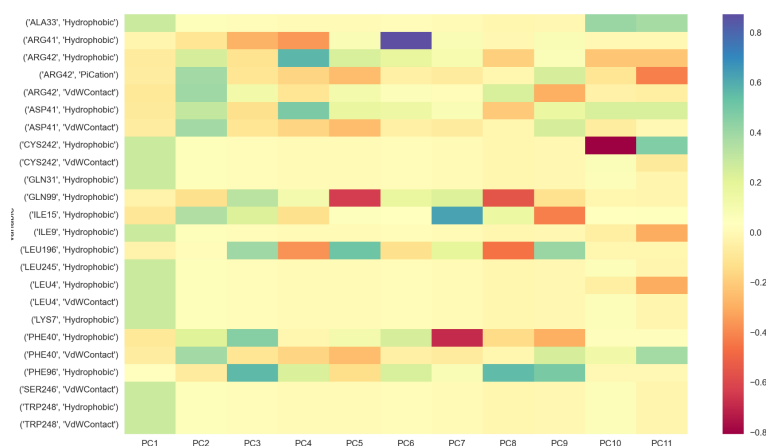
Ważnym czynnikiem pozwalającym na zrozumienie kinetyki wiązania ligandów z receptorami jest analiza oddziaływań między nimi, jak również poznanie ich cech strukturalnych. Zaproponowane zostało nowe podejście mające na celu identyfikację kluczowych interakcji występujących w trajektoriach dysocjacji τ RAMD, które wpływają na czas przebywania leku w jego celu molekularnym dla badanych inhibitorów enzymu InhA.

Interakcje występujące pomiędzy receptorem a ligandem zostały wyodrębnione z trajektorii dysocjacji τ RAMD w następujący sposób: (i) z każdej klatki (punktu czasowego) trajektorii uzyskano pozycję środka masy liganda i współrzędne atomów tworzących cały układ i zapisano je do osobnych plików *.pdb* za pomocą przygotowanego skryptu tcl dla narzędzia VMD uruchomionego skryptem pythona; (ii) otrzymane współrzędne położenia atomów w przestrzeni posłużyły jako dane wejściowe do zidentyfikowania oddziaływań między ligandem a receptorem z wykorzystaniem bibliotek RDKit i ProLif z Pythona (kategorie oddziaływań: hydrophobic, π -stacking, π -cation and cation- π , anionic and cationic, and H-bond donor and acceptor); (iii) każde oddziaływanie zostało oznaczone jako „1”, gdy zaobserwowano wystąpienie danego oddziaływania lub „0”, czyli jego brak; (iv) powtarzające się oddziaływania zostały zsumowane w celu określenia częstości występowania danego kontaktu oraz uśrednione względem zarówno pojedynczej trajektorii dysocjacji jak i wszystkich powtórzeń (trajektorii) dla danego układu (co pozwoliło też na wyeliminowanie zależności czasowej danych); (v) na podstawie częstości występowania oddziaływań zdefiniowano próg, który pozwalał na oddzielenie stanu związanego od stanów przejściowych i całkowitego uwolnienia – do dalszej analizy usunięte zostały konformacje układu, w których zostało wykryte oddziaływanie występujące w przynajmniej 20% pojedynczej trajektorii dysocjacji.

Standaryzacja zbiorów danych jest ważnym krokiem w analizie, ponieważ usuwa ona tendencyjność z oryginalnych zmiennych. Zmienne standaryzowane mają podobną wariancję.

Analiza tabeli oddziaływań badanych kompleksów z wykorzystaniem PCA pozwoliła na wyodrębnienie głównych czynników dla wszystkich badanych układów inhibitorów InhA. Rysunek 5 pokazuje, jak zmienne główne korelują ze składową główną.

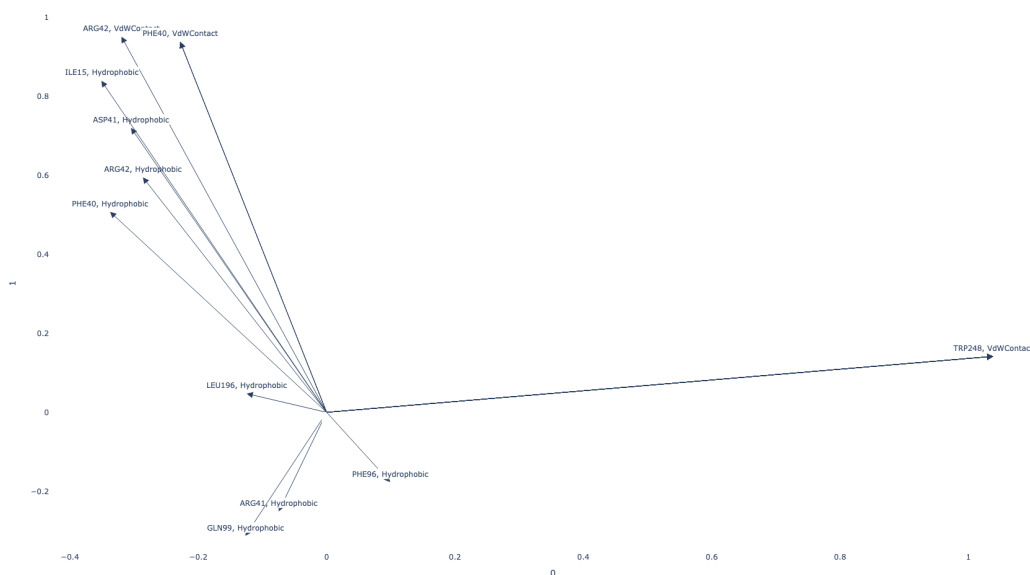
Pierwsze dwa czynniki modelują największą wariację w danych, najlepiej opisując jego strukturę. Dla badanych układów czynniki te opisują jedynie około 50% całkowitej wariacji danych. Mimo to analiza PCA dostarcza cennych informacji.



Rysunek 5. Korelacja cech z głównymi składowymi.

Projekcja wag na przestrzeń zdefiniowaną przez dwie pierwsze składowe główne po klasteryzacji danych metodą k-means przedstawiono na Rysunku 6 w celu zbadania, które czynniki są odpowiedzialne za zróżnicowanie próbek. Każda zmienna (interakcja) jest reprezentowana w postaci wektora, którego kierunek i długość określają, jak bardzo dana zmienna wpływa na każdą składową główną. Można więc stwierdzić, że największy wkład w tworzenie pierwszego czynnika (wartości współczynników są najwyższe) jest oddziaływanie van der Waalsa z aminokwasem Trp248. Kontakty hydrofobowe z Leu196, Arg41, Gln99 i Phe96 również przyczyniają się do powstania pierwszego czynnika głównego. Trzy ostatnie wraz z hydrofobowymi oddziaływaniami z Phe40, Arg42, Asp41, Ile15 oraz oddziaływaniami van der Waalsa z Arg42 i Phe40 tworzą drugi główny składnik. Z projekcji wagowych można wywnioskować, że oddziaływania hydrofobowe z Gln99 i Arg41 są silnie skorelowane. Interakcje hydrofobowe z aminokwasami Phe40, Arg42, Asp41 i Ile15 oraz oddziaływania van der Waalsa z Arg42 i Phe40 są również dodatnio skorelowane ze sobą. Ten zestaw interakcji jest ujemnie skorelowany z hydrofobowymi oddziaływaniami z Gln99, Arg41 i Phe96. Oddziaływanie hydrofobowe pomiędzy ligandem a Leu196 jest ujemnie skorelowane z oddziaływaniem van der Waalsa z aminokwasem Trp248, które nie wykazuje korelacji z innymi

oddziaływaniami. Oddziaływanie z Trp248 obserwowano tylko w przypadku liganda w kompleksie 5MTQ (czas przebywania 119 min), a z Leu196 z ligandem w kompleksie 5UGU (194 min).



Rysunek 6. Projekcja wag na przestrzeń dwóch pierwszych składowych głównych.

6. Dyskusja i wnioski

Celem pracy było zastosowanie i przetestowanie rozwiązań przewidywania czasu przebywania leków w celu określenia, czy mogą one być stosowane niezależnie od wielkości cząsteczek i rodziny białek, a także identyfikacja właściwości molekularnych cząsteczek zaangażowanych w proces dysocjacji inhibitorów białka InhA. Najważniejsze pytania na które starano się odpowiedzieć w prezentowanych badaniach to: Jakie są główne oddziaływania receptor-ligand, które wyróżniają ligandy o długim i krótkim czasie przebywania? Czy metoda τ RAMD jest uniwersalna i możliwa do zastosowania dla cząsteczek o różnej wielkości i podobieństwie strukturalnym? Jaka jest korelacja pomiędzy względnym czasem przebywania a pomiarami eksperymentalnymi?

Ponieważ nie było dostępnej bazy danych kinetyki wiązania ligandów, dane zostały zebrane z literatury i opublikowane w internetowej bazie danych (<https://pdbrt.polsl.pl>). Baza danych PDBrt zawiera 59 wpisów ligandów dla 7 różnych rodzin białek. Baza będzie stale poszerzana o nowe dostępne dane.

Metoda τ RAMD została zastosowana dla 5 różnych systemów białkowych do oszacowania względnego czasu rezydencji. Aby sprawdzić powtarzalność metody, wykonano analizę na zestawie 15 inhibitorów białka HSP90, z których 14 zostało opublikowanych. Natomiast w celu sprawdzenia uniwersalności, przeprowadzono badania dla 11 inhibitorów białka InhA, a także 3 ligandów białek ENR, EGFR HIV-1. Cały zestaw składał się zatem z 29 ligandów o zróżnicowanej strukturze.

Przeprowadzona analiza pozwoliła na stwierdzenie, że τ RAMD jest powtarzalna, jednak należy wykonać więcej zestawów symulacji w niektórych przypadkach, aby zmniejszyć rozproszenie danych wokół średniej. Zaobserwowano, że wyniki τ RAMD wykazują dobrą lub silną korelację z doświadczalnie wyznaczonym czasem rezydencji dla związków o podobnej strukturze tj. różniących się niewielką modyfikacją np. przesunięciem grupy funkcyjnej. Natomiast dla zróżnicowanego strukturalnie zestawu ligandów metoda okazała się mieć niską skuteczność, co sugeruje ograniczone zastosowanie τ RAMD w procesie odkrywania leków.

Aby określić właściwości molekularne białka InhA i jego inhibitorów podczas procesu dysocjacji charakterystycznego dla danego czasu przebywania, zastosowano podejście, które uwzględnia stany przejściowe układu oraz oddziaływania zachodzące między cząsteczką białka a cząsteczką liganda. Podejście to nie wymaga wcześniejszej wiedzy na temat mechanizmu wiązania i opiera się na analizie PCA i klasteryzacji k-średnich.

Zidentyfikowane grupy tworzą odciski palców interakcji charakterystyczne dla ligandów o określonym czasie przebywania, a także grupy wyróżniające ligandy na podstawie długości czasu przebywania w celu molekularnym.

Dodatkowa analiza deskryptorów molekularnych, które pozwalają na głębszy wgląd we właściwości molekularne układu białko-ligand byłyby zapewne interesującym rozszerzeniem metody. Jest to obszar dalszego rozwoju metody.

Analiza częstości występowania danej interakcji w kompleksach pozwoliła na zidentyfikowanie kluczowych aminokwasów, które mogą mieć znaczący wpływ na zróżnicowanie czasu przebywania w kompleksach:

- oddziaływania hydrofobowe z aminokwasem Leu196, a także oddziaływania van der Waalsa z Phe40, Asp41, Arg42 oraz π -kation z Arg42 sprzyjają dłuższemu czasowi przebywania, gdyż zostały one zidentyfikowane tylko w kompleksach z ligandami o najdłuższych czasach przebywania w badanym zbiorze danych (106, 194 i 220 min),
- dla związków o stosunkowo krótkich czasach przebywania (30 i 50 min), oddziaływanie hydrofobowe oddziaływanie pomiędzy ligandem a aminokwasem Gln99 jest charakterystyczne.

Analiza głównych składowych (PCA) pozwoliła zidentyfikować czynniki odpowiedzialne za zróżnicowanie czasów przebywania ligandów. Czynniki te są następujące:

- oddziaływania van der Waalsa z aminokwasem Trp248,
- oddziaływania hydrofobowe z aminokwasami Gln99 i Arg41,
- oddziaływania hydrofobowe z aminokwasami Phe40, Arg42, Asp41, Ile15 oraz -oddziaływania van der Waalsa z Arg42 i Phe40.

Dla związków o podobnej strukturze, badania wykazały, że względny czas przebywania koreluje dobrze z czasem wyznaczonym doświadczalnie. Zaproponowany algorytm może być wykorzystany do identyfikacji kluczowych cech molekularnych dla szybkości dysocjacji ligandów z miejsca wiązania dla związków o podobnej strukturze.

7. Bibliografia

Case, D., Aktulga, H., Belfon, K., Ben-Shalom, I., Berryman, J., Brozell, S., Cerutti, D., III, T. C., Cisneros, G., Cruzeiro, V., Darden, T., Duke, R., Giambasu, G., Gilson, M., Gohlke, H., Goetz, A., Harris, R., Izadi, S., Izmailov, S., Kasavajhala, K., Kaymak, M., King, E., Kovalenko, A., Kurtzman, T., Lee, T., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Machado, M., Man, V., Manathunga, M., Merz, K., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., O'Hearn, K., Onufriev, A., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Roe, D., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shajan, A., Shen, J., Simmerling, C., Skrynnikov, N., Smith, J., Swails, J., Walker, R., Wang, J., Wang, J., Wei, H., Wolf, R., Wu, X., Xiong, Y., Xue, Y., York, D., Zhao, S., & Kollman, P. (2022), 'Amber 22'. URL: <https://ambermd.org/AmberTools.php>

Copeland, R. A. (2016), 'The drug-target residence time model: A 10-year retrospective', *Nature Reviews Drug Discovery* 15, 87–95. URL: <http://dx.doi.org/10.1038/nrd.2015.18>

Copeland, R. A., Pompliano, D. L. & Meek, T. D. (2006), 'Drug-target residence time and its implications for lead optimization', *Nature Reviews Drug Discovery* 5, 730–739

Dahl, G. & Akerud, T. (2013), 'Pharmacokinetics and the drug-target residence time concept'.

Dowling, M. R. & Charlton, S. J. (2006), 'Quantifying the association and dissociation rates of unlabelled antagonists at the muscarinic m₃ receptor', *British Journal of Pharmacology* 148, 927–937

Folmer, R. H. (2018), 'Drug target residence time: a misleading concept', *Drug Discovery Today* 23, 12–16. URL: <https://doi.org/10.1016/j.drudis.2017.07.016>

Gabdoulline, R. R. & Wade, R. C. (1999), 'On the protein-protein diffusional encounter complex'.

Gabdoulline, R. R. & Wade, R. C. (2022), 'Biomolecular diffusional association gabdoulline and wade 205'

Jakalian, A., Bush, B. L., Jack, D. B. & Bayly, C. I. (2000), 'Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method', *Journal of Computational Chemistry* 21, 132–146.

Jakalian, A., Jack, D. B. & Bayly, C. I. (2002), 'Fast, efficient generation of high-quality atomic charges. am1-bcc model: II. parameterization and validation', *Journal of Computational Chemistry* 23, 1623–1641

Kokh, D. B., Amaral, M., Bomke, J., Gradler, U., Musil, D., Buchstaller, H. P., Dreyer, M. K., Frech, M., Lowinski, M., Vallee, F., Bianciotto, M., Rak, A. & Wade, R. C. (2018), 'Estimation of drug-target residence times by τ -random acceleration molecular dynamics simulations', *Journal of Chemical Theory and Computation* 14, 3859–3869.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. (2015), 'ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb', *Journal of Chemical Theory and Computation* 11(8), 3696–3713. PMID: 26574453. URL: <https://doi.org/10.1021/acs.jctc.5b0025>

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. & Schulten, K. (2020), 'Scalable molecular dynamics with namd'. URL: www.ks.uiuc.edu

Schrodinger, L. & DeLano, W. (2020), 'Pymol'. URL: <http://www.pymol.org/pymol>

Swinney, D. C. (2004), 'Biochemical mechanisms of drug action: what does it take for success?'. URL: www.nature.com/reviews/drugdisc

Vauquelin, G. (2016), 'Effects of target binding kinetics on in vivo drug efficacy: koff, kon and rebinding', *British Journal of Pharmacology* pp. 2319–2334