

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

Projektowanie leków jest złożonym procesem, który wciąż obciążony jest dużym ryzykiem i niepewnością. W jego wczesnym etapie skupiano się na równowagowym powinowactwie wiązania leku z określonym celem, które opisuje stała dysocjacji równowagowej (K_d). Jednak istnieje wiele leków o nierównowagowych właściwościach wiążących. Dlatego coraz bardziej kładzie się nacisk na optymalizację pozostałych parametrów kinetycznych: stałej szybkości dysocjacji (k_{off}) i stałej szybkości asocjacji (k_{on}), aby zwiększyć dokładność pomiarów skuteczności leków *in vivo*. W tym celu opracowano koncepcję czasu przebywania leku w celu (τ), która uwzględnienia ciągłą eliminację leku, brak warunków równowagi i dynamikę konformacyjną cząsteczek docelowych. Dlatego czas przebywania okazał się lepszym surogatem skuteczności życiowej niż powinowactwo wiązania w stanie równowagi i jest uznawany za kluczowy parametr w procesie projektowania leków. Należy jednak pamiętać, że czas przebywania jako pojedyncza miara skuteczności leku może dać ograniczony obraz kinetyki wiązania i powinowactwa. Czasochłonność, pracochołonność i często wysoki koszt metod eksperymentalnych służących do wyznaczania parametrów kinetyki wiązania wraz z szybkim rozwojem technologicznych wpływają na coraz większą potrzebę opracowania wysokoprzepustowych metod *in silico*, których przeznaczeniem będzie zarówno przewidzenie parametrów jak i określenie kluczowych determinantów kinetyki wiązania lek-receptor.

Niniejsza praca opisuje podstawowe założenia oraz zastosowania dostępnych metod obliczeniowych, które pozwalają na zrozumienie i analizę kinetyki (nie)wiązania ligandów, a także wyznaczenie czasu przebywania leku w celu. Są to metody dynamiki molekularnej (MD), uczenia maszynowego (ML) i ich połączenie oraz wykorzystanie modeli stanów Markova. Ponieważ dostępne bioinformatyczne bazy danych nie przechowują kompleksowych informacji na temat danych kinetycznych kompleksów o znanej strukturze krystalograficznej, dane te zostały zebrane z opublikowanych źródeł i przekształcone w publicznie dostępną internetową bazę danych, nazwaną PDBrt. Badania wykonano dla inhibitorów InhA, the enoyl acyl carrier protein reductase z organizmu *Mycobacterium tuberculosis* - jednego z kluczowych enzymów uczestniczących w szlaku biosyntezy kwasów tłuszczowych typu II u *M. tuberculosis*. Do sprawdzenia powtarzalności τ RAMD wykorzystano również inhibitory białek szoku cieplnego HSP90 i ligandy białek ENR, EGFR i HIV-1.

Zastosowanym podejściem MD jest analiza τ RAMD w celu sprawdzenia, czy metoda jest powtarzalna i możliwa do zastosowania w obliczaniu różnych względnych czasów przebywania związków podobnych do leków. Zaobserwowano, że τ RAMD zapewnia najdokładniejsze przewidywanie czasów przebywania, gdy jest stosowane do serii podobnych związków. Uzyskane wyniki pozwoliły na stwierdzenie, że τ RAMD jest powtarzalna - uzyskane wyniki są podobne do tych, które zostały opublikowane. Co więcej, badania wykazały, że metoda τ RAMD nie pozwoliła na oszacowanie względnych czasów przebywania, które dobrze korelują z wartościami doświadczalnymi dla związków zróżnicowanych strukturalnie. Można zatem stwierdzić, że zastosowanie metody jest ograniczone i nie może zostać ona wykorzystana dla dowolnego zestawu związków.

Kolejnie zaproponowany został algorytm wykorzystujący metody uczenia maszynowego do określenia cech molekularnych mających wpływ na kinetykę wiązania białko-ligand dla zestawu podobnych związków. Dane do trenowania modelu zostały uzyskane z symulacji dynamiki molekularnej τ RAMD. Badania potwierdziły, że τ RAMD dostarcza informacji o charakterystyce ścieżki dysocjacji, ponieważ z otrzymanych trajektorii dysocjacji możliwe jest zidentyfikowanie występujących oddziaływań oraz zmian konformacyjnych układu w kolejnych punktach czasowych. Te informacje zostały wykorzystane do dalszej analizy prowadzącej do zdefiniowania kluczowych własności molekularnych dla serii inhibitorów enzymu InhA. Zaproponowany algorytm pozwolił na uzyskanie informacji o kontaktach białko-ligand specyficznych dla ich czasów przebywania.