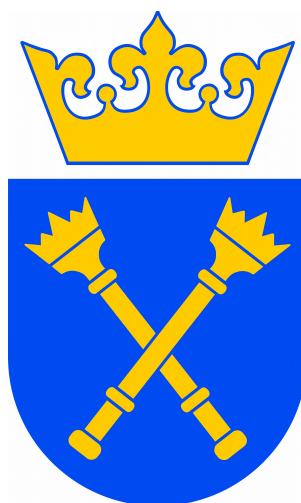


**CONFORMATIONAL AND METHODOLOGICAL ASPECTS OF  
SELECTED GPCRS' MOLECULAR DOCKING STUDIES FOR  
IMPROVEMENT OF COMPUTER-AIDED DRUG DESIGN**

**STRESZCZENIE ROZPRAWY DOKTORSKIEJ**

**Katarzyna Rzęsikowska**



**Rozprawa doktorska wykonana na  
Wydziale Chemii Uniwersytetu Jagiellońskiego  
pod kierunkiem prof. dr hab. Krzysztofa Lewńskiego**

**Kraków, 2019**

Receptory sprzężone z białkiem G (z ang. G-protein coupled receptors; GPCR) należą do jednej z największych i najbardziej zróżnicowanych rodzin białek transbłonowych. Są one odpowiedzialne za regulację wielu procesów fizjologicznych w organizmie ludzkim, np. uczenie się, regulacji ciśnienia krwi czy kontrolę funkcjonowania zmysłów, takich jak: wzrok, zapach, czy smak. Z tego powodu receptory GPCR są głównym białkowym celem wykorzystywanym w projektowaniu leków. W latach 2012-2017, 69 nowych leków działających na różne receptory GPCR zostało zatwierdzonych przed Agencją Żywności i Leków (z ang. Food and Drug Administration; FDA). Szacuje się, że około 34% leków dostępnych na rynku działają na 108 różnych receptorów GPCR. Obecnie, coraz więcej tego typu związków znajduje się obecnie w fazie badań klinicznych. Pomimo wielkiego zainteresowania receptorami z tej rodziny, leki dostępne na rynku działają jedynie na ok. 13% wszystkich znanych GPCR. Taki wynik stwarza duże możliwości w dziedzinie projektowania leków działających na cele molekularne tego typu. Nadal istnieje jednak wiele ograniczeń, takich jak niewystarczająca ilość informacji strukturalnych lub brak/niedobór danych dotyczących funkcji tych receptorów. Dodatkowo, zagadnienie modelowania konformacji ligandów w kieszeni aktywnej receptora również nie jest proste.

Obecnie metody komputerowo-wspomagane projektowania leków (z ang. computer-aided drug design; CADD) stały się nieodzownym elementem procesu projektowania nowych leków. Wykorzystywane są one, między innymi do modelowania nieokreślonych eksperymentalnie struktur receptorów, przewidywania powinowactwa potencjalnych ligandów do wybranego receptora, czy określenia sposobu oddziaływania liganda z białkiem. CADD pozwala na przyspieszenie procesu projektowania leków, jednocześnie wpływając na zmniejszenie jego kosztów produkcji. Jednakże, obecnie dostępne narzędzia do projektowania leków w dalszym ciągu nie są wystarczająco dokładne. Z tego powodu, poszukiwanie słabych punktów w metodologii i próba znalezienia rozwiązań prowadzących do ulepszenia stosowanych metod jest celem przedłożonej rozprawy. Podczas realizowanych badań weryfikowano różne aspekty doświadczeń opartych o procedury dokowania molekularnego pod kątem poprawy działania narzędzi stosowanych w CADD. Skupiono się na: (1) doborze matrycy i sposobie budowania modeli przy użyciu modelowania homologicznego, (2) rozkładzie potencjału elektrostatycznego i analizie oddziaływań międzycząsteczkowych w kompleksach białko-ligand, (3) wpływie wybranych parametrów dokowania molekularnego i sposobu generowania początkowej

konformacji ligandów na otrzymane wyniki predykcji oraz (4) sposobie definiowania funkcji oceny i jego wpływie na wyniki otrzymane w procesie dokowania.

Początkowe etapy przeprowadzono dla dwóch białek GPCR, należących do klasy A: 5-HT<sub>1A</sub> i 5-HT<sub>7</sub>. Badania rozpoczęto od skonstruowania modeli tych receptorów. Jakość otrzymanych modeli była sprawdzana przy użyciu takich parametrów jak: aLogAUC, EF1% oraz analizując wykresy Ramachandrana i sprawdzając ilość aktywnych ligandów oddziałujących z resztą Asp3.32. Wspomniane oddziaływanie jest znanym i charakterystycznym wiązaniem jonowym obserwowanym dla większości kompleksów receptorów serotoniny z aktywnymi ligandami. Zastosowanie zróżnicowanej metodologii i matryc pozwoliło na wygenerowanie modeli obydwu receptorów, które w procesie dokowania pozwalają na rozróżnienie ligandów aktywnych od nieaktywnych. Zarówno stopień podobieństwa pomiędzy sekwencjami matrycy i modelowanego receptora, jak i sposób budowy modeli nie wykazały jednoznacznego wpływu na jakość otrzymanych modeli. Zaskakującym jest, że w niektórych przypadkach modele budowane w oparciu o matrycę o niższym stopniu podobieństwa sekwencji uzyskiwały lepsze wyniki w porównaniu z tymi, otrzymanymi w oparciu o matrycę wykazujące bardzo wysoki stopień podobieństwa sekwencji.

Otrzymane modele zostały użyte do zaprojektowania protokołu pozwalającego na znalezienie selektywnych ligandów receptora 5-HT<sub>7</sub>. Badania skupiły się na analizie rozkładu potencjału elektrostatycznego oraz oddziaływań niekowalencyjnych występujących pomiędzy ligandem i białkiem. Skuteczność zaproponowanego algorytmu została sprawdzona w eksperymentach klasyfikacyjnych, przeprowadzonych dla 3 zbiorów związków, spośród których dwa (1A-active i 7-active) były zaprojektowane w taki sposób, aby odwzorować warunki wirtualnego przesiewania bibliotek związków (z ang. virtual screening; VS) – zbiory zawierały małą liczbę ligandów aktywnych i bardzo duży zbiór związków nieaktywnych. Zastosowanie protokołu zwiększyło ponad pięciokrotnie prawdopodobieństwo wyselekcjonowania związku aktywnego. Co więcej, prawie wszystkie etapy algorytmu nie wymagają użycia bardziej wyspecjalizowanych zasobów obliczeniowych. W związku z tym, zaproponowane podejście może być zastosowane do przesiewania bibliotek związków w celu poszukiwania selektywnych ligandów receptora 5-HT<sub>7</sub>.

W następnym kroku skupiono się na umiejętności prawidłowego przewidywania konformacji liganda w kieszeni wiążącej receptora (z ang. sampling power). Badania przeprowadzono na

trzech receptorach GPCR ( $A_{2A}$ ,  $\beta_1$  and  $\beta_2$ ), dla których w bazie PDB można znaleźć przynajmniej 10 struktur krystalicznych z różnymi ligandami. Celem tej części prac badawczych było wyznaczeniu optymalnych warunków dokowania, skupiając się na następujących parametrach: ilość cykli algorytmu genetycznego, rodzaj funkcji oceniającej, aktywacja lub brak opcji generowania różnorodnych rozwiązań oraz poziom wydajności przeszukiwania. Wykazano, że czynnikiem mającym największy wpływ na otrzymane pozy był rodzaj zastosowanej funkcji oceny.

Zoptymalizowane w poprzednim etapie warunki dokowania wykorzystano do zbadania wpływu sposobu generowania początkowej konformacji ligandów na uzyskane wyniki dokowania. Użyto konformacji referencyjnych, obserwowanych w wybranych strukturach krystalicznych kompleksów ligand-białko oraz geometrii wygenerowanych z formatu SMILES korzystając z dostępnych serwerów i programów (CACTUS, OpenBabel i OpenEye Omega) a także uzyskanych przy pomocy obliczeń kwantowo-mechanicznych. Zaobserwowano pozytywny wpływ na otrzymane wyniki dokowania, gdy początkowe geometrie ligandów były uzyskane za pomocą oprogramowania OpenEye Omega. Co więcej, dalsza optymalizacja geometrii otrzymanych przy użyciu tego programu z zastosowaniem metod kwantowo-mechanicznych nie doprowadziła do uzyskania lepszych wyników. W związku z tym, wydaje się, że nie ma potrzeby stosowania bardziej czasochłonnych metod obliczeniowych do generowania startowej konformacji badanych związków aby otrzymać wiarygodne wyniki dokowania.

Wyniki otrzymane w trakcie optymalizacji warunków dokowania sugerują znaczący wpływ rodzaju funkcji oceniającej (SF) na otrzymane pozy ligandów. W związku z tym, w ostatniej części tej pracy skupiono się na sposobie definiowania SF. Sprawdzano, czy zdefiniowanie SF dla konkretnego białka, klasy czy rodziny, pozwoli na uzyskanie lepszych wyników dokowania. W tym celu badano jak dana funkcja oceny (domyślna i zaprojektowana) radzi sobie z klasyfikacją ligandów pod kątem ich powinowactwa do badanego receptora. Zaobserwowano, że użycie SF z wygenerowanymi dla poszczególnych białek czynnikami wagowymi prowadzi do otrzymanie wyników bardziej zgodnych z danymi eksperymentalnymi.

Wyniki uzyskane w ramach przedstawionej rozprawy doktorskiej pozwoliły na głębsze zrozumienie niedoskonałości stosowanych powszechnie metodologii i zaproponowanie

rozwiązań mających na celu ulepszenie procesu poszukiwania nowych, biologicznie aktywnych związków przy użyciu technik dokowania molekularnego.