

Od przeszło 70 lat deskryptory molekularne są wykorzystywane do transformacji cech molekularnych związków chemicznych na wartości liczbowe. Jednym z narzędzi stosowanych do ustalenie ilościowej relacji pomiędzy strukturą a aktywnością biologiczną/właściwościami fizykochemicznymi (QSAR/QSPR) związków chemicznych jest model predykcyjny. Zgodnie z zaleceniami OECD, prawidłowo opracowane i zweryfikowane modele predykcyjne należy interpretować o ile jest to możliwe, tzn. wyjaśniać związek przyczynowo-skutkowy między zastosowanymi w modelu deskryptorami molekularnymi a modelowanym punktem końcowym. Wraz z rozwojem metod sztucznej inteligencji dokonał się postęp w chemometrii i chemoinformatyce, który doprowadził do nowych strategii znajdowania matematycznych zależności w analizie QSAR/QSPR. Natomiast pogłębiona interpretacja deskryptorów pozwala wnioskować, w jaki sposób cząsteczka wywołuje określony efekt w układzie biologicznym.

Celem badań w niniejszej pracy była analiza i interpretacja wybranych deskryptorów molekularnych klas RDF(R), GATSk(w), GETAWAY oraz 3D-MoRSE wykorzystanych w modelowaniu zależności struktura-aktywność/właściwości fizykochemiczne trzech grup związków biologicznie aktywnych: pochodnych tiazolu, pochodnych 4-tiazolidynonu oraz flawonoidów. Badane deskryptory molekularne są powszechnie stosowane w modelach QSAR/QSPR. W badaniach podjęto udaną próbę opracowania metodologii obliczeń deskryptorów molekularnych połączonej z dekompozycją ich wartości na wkłady par atomowych. Ponadto celem pracy było wprowadzenie podejścia substrukturalnego, opartego o addytywny charakter deskryptorów jako narzędzia ułatwiającego odpowiedź na pytanie jakie fragmenty strukturalne badanych związków w największym stopniu wpływają na wartości deskryptorów. Kolejnym celem pracy było zaproponowanie metod wizualizacji deskryptorów molekularnych w celu usprawnienia procesu interpretacji modeli QSAR.

Do realizacji założonych celów przeprowadzono optymalizację geometrii cząsteczek z wyznaczeniem macierzy deskryptorów molekularnych i analizę QSAR/QSRR z zastosowaniem wielokrotnej regresji liniowej z algorytmem genetycznym lub z selekcją zmiennych wprzód. Skonstruowano model QSAR pozwalający prognozować aktywność przeciwgrzybiczą przeciw *Candida albicans* (ATCC 2091) na podstawie struktury pochodnych tiazolu oraz model QSRR do predykcji lipofilowości flawonoidów. Z kolei dla pochodnych 4-tiazolidynonu zinterpretowano wybrany deskryptor klasy GETAWAY, który w literaturze był wykorzystywany jako zmienna zależna w modelu zależności struktura-aktywność

antytrypanosomalna. Modele QSAR i QSRR zostały zweryfikowane na podstawie szeregu metryk walidacyjnych. Interpretacja deskryptorów klas RDF(R), GATSk(w), GETAWAY oraz 3D-MoRSE została oparta o analizę (1) rozmieszczenia par atomowych i ich wkładu w wartość deskryptora (2) fragmentów cząsteczek (substruktur) oraz (3) powierzchni molekularnych mapowanych wartościami kumulatywnych wkładów atomów. Interpretacja deskryptorów molekularnych została przeprowadzona przy użyciu autorskiego narzędzia InterDescPy i wzbogacona o szereg elementów graficznych umożliwiającą wizualizację wyników oraz wspomagających proces wnioskowania.

Wyniki uzyskane z interpretacji wybranych deskryptorów wskazują, że (1) aktywność przeciwgrzybicza pochodnych tiazolu silnie zależy od wielkości cząsteczki, elektroujemności podstawników pierścienia fenyłowego oraz jego wzajemnych relacji przestrzennych w stosunku do reszty cząsteczki oraz (2) lipofilowość flawonoidów zależy od rozmieszczenia grup hydroksylowych, obecności reszty cukrowej, typu wiązania glikozydowego i nasycenia pierścienia γ -pironu. Z kolei na podstawie analizy deskryptora HATS5m pochodnych 4-tiazolidynonu wykazano jaki wpływ na mechanistyczną interpretację deskryptora ma schemat ważenia, wartość funkcjonatu Delta-Diraca oraz wskaźniki wpływu.

Słowa kluczowe: QSAR, deskryptory molekularne, RDF(R), GETAWAY, interpretacja

*Mariusz
Zapadka*