

Jak szybko i łatwo stworzyć nowe leki? Penicylina, jeden z najpowszechniej stosowanych antybiotyków XX wieku, została odkryta przypadkiem, kiedy próbki Flaminga, przez nieuwagę, zostały zakażone przez pleśń. Jednak w świecie nowych technologii, sukces jest zazwyczaj planowany, a nie pozostawiany szczęściu. Przez wiele lat leki powstawały w oparciu o medycynę ludową, związki syntetyczne, naturalne ligandy (związki aktywne względem konkretnego białka), a tak właśnie wyszukiwano nowe zastosowania dla istniejących już leków. Testowano dziesiątki tysięcy związków i cząsteczek, co związane było nie tylko z ogromnymi kosztami, ale również z bardzo długim czasem koniecznym na wprowadzenie skutecznego leku na rynek. Wyjątkowo pomocna w tym procesie okazała się technologia komputerowa: komputerowo wspomagane projektowanie leków (computer aided drug design, CADD). Pozwoliła ona nie tylko na eliminację związków o bardzo małej szansie działania w danym schorzeniu, ale również na odrzucenie tych substancji, które mogłyby być ligandami dla niepożądanych białek.

Aby dany związek mógł stać się lekiem, powinien oddziaływać z konkretnym białkiem, stymulując lub hamując jego działanie. Określenie prawdopodobieństwa takiego oddziaływania wspomagane jest przez metody tzw. wirtualnego skriningu. Opierają się one na obliczaniu i analizie właściwości związku chemicznego i dzielą się na dwie grupy: metody oparte na strukturze i metody oparte na ligandzie. Metody oparte na strukturze wykorzystują informacje o budowie związku chemicznego oraz docelowego białka i mają na celu obliczanie energii ich kompleksów w różnych warunkach. Im niższa energia, tym wyżej jest stabilność kompleksu, co może świadczyć o aktywności związku chemicznego względem docelowego białka. Najbardziej popularne metody oparte na strukturze to: dokowanie ligandów oraz dynamika molekularna. Drugie podejście, oparte na ligandach, wykorzystuje informacje jedynie o strukturze związku chemicznego oraz jego właściwościach. Pozwala to na określenie prawdopodobieństwa jego aktywności na podstawie podobieństwa do już istniejących, opisanych leków o znanym działaniu. Z uwagi na fakt, iż metody oparte na ligandach wykorzystują znacznie mniejsze ilości informacji, koszt obliczeniowy potrzebny do ich użycia jest znacznie niższy niż w przypadku metod opartych na strukturze. Przykładami takich metod opartych na ligandzie są fingerprinty oraz farmakofory.

Fingerprinty, czyli „odciski palców”, to metody opisywania obiektów seriami wartości liczbowych, co pozwala na ich analizę, klastrowanie (grupowanie) i klasyfikację poprzez różne algorytmy. Są one powszechnie używane w życiu codziennym, np. przy rozpoznawaniu słów, ochronie baz danych etc. W przypadku fingerprintów chemicznych, taki „odcisk palca” związku chemicznego może reprezentować fragmenty jego struktury lub własności fizykochemiczne.

Jednym z głównych typów fingerprintów są fingerprinty podstrukturalne, które opisują występowanie określonych grup funkcyjnych (podstruktur) w konkretnym związku chemicznym. Jest to prosta, powszechnie używana metoda przedstawiania struktury związku, jednak posiada ona pewne wady. W fingerprintach podstrukturalnych nie jest zawarta informacja na temat kolejności ułożenia podstruktur wewnątrz cząsteczki, przez co dwa lub więcej różnych związków chemicznych może być opisanych tym samym ciągiem znaków.

W Projekcie „Fingerprinty Podstrukturalnych i Maszyny Ekstremalnej Entropii: Nowa Metoda Reprezentacji i Analizy Związków Chemicznych” zaproponowano rozwiązanie powyższego problemu. Dodanie do fingerprintów podstrukturalnych dodatkowych informacji na temat poszczególnych podstruktur powinno pozwolić na dokładniejszy opis cząsteczek i wzrost skuteczności ich analizy. Badania wstępne wykazały, że już bardzo prosta implementacja tej metodologii skutkuje wyraźnym wzrostem trafności klasyfikacji związków chemicznych pod względem ich aktywności. Dodatkowo Projekt zakłada wprowadzenie bardziej wyszukanych metod analizy fingerprintów, takich jak grafowe funkcje J-dra i maszyny ekstremalnej entropii, co również powinno poprawić skuteczność procesu.

Projekt składa się z kilku etapów, których celem jest uzyskanie kompletnej metodologii tworzenia i analizy fingerprintów podstrukturalnych (FPP). Będzie ona dostępna do pobrania w formie darmowej aplikacji pozwalającej na użycie FPP przez pozostałych badaczy.

Pierwszy etap będzie skupiony na stworzeniu narzędzia do konstrukcji FPP. Kolejne kroki zakładają eksplorację różnych podejść do tworzenia zestawów związków, parametrów generacji FPP oraz różnych metod ich analizy. W Projekcie założone jest wykorzystanie maszyn ekstremalnej entropii, nowatorskiego podejścia do klasyfikacji danych poprzez metody uczenia maszynowego. W celu oceny skuteczności metodologii FPP w odróżnianiu związków aktywnych od nieaktywnych, będzie ona porównywana z obecnie dostępnymi fingerprintami podstrukturalnymi oraz metodami uczenia maszynowego, użycie tego wielu dostępnych baz związków chemicznych. Wszystkie to procedury przeprowadzone zostaną dla licznych białek, w tym dla receptorów sprzężonych z białkiem G oraz kinaz białkowych.

Powyższy projekt stworzy nowe metody reprezentacji związków chemicznych wraz z zestawem algorytmów do jej analizy, które zostaną przetestowane pod względem wydajności w porównaniu do obecnie dostępnych metod. Przyczyni się to do znacznego rozwoju wirtualnego skriningu opartego na ligandach, co usprawni proces wyszukiwania nowych leków.