Fingerprint Sąsiadujących Podstruktur jako Nowa Metoda Reprezentacji Związków Chemicznych oraz Jego Wykorzystanie w Poszukiwaniu Związków Selektywnych dla Wybranych Receptorów Serotoninowych.

Krzysztof Rataj, Instytut Farmakologii PAN, ul. Smętna 12 31-343 Kraków

Chemoinformatyka jest domeną naukową łączącą zagadnienia chemii, fizyki, matematyki i informatyki. Wykorzystuje ona stale rosnącą moc obliczeniową nowoczesnych komputerów aby poznawać i symulować własności związków chemicznych, co pozwala na późniejsze zastosowanie uzyskanych informacji np. do projektowania materiałów o określonych właściwościach, czy też aktywnych biologicznie związków. W poszukiwaniach ligandów aktywnych wobec określonego białka cheminformatyka wykorzystuje różne algorytmy z zakresu przeszukiwania danych, pól siłowych, fizyki kwantowej, czy klasyfikacji. Wiele podejść z dziedziny komputerowo wspomaganego projektowania leków (Computer Aided Drug Design, CADD) opiera się w znacznej mierze na poprawnej reprezentacji związków chemicznych, tzn. takiej, która pozwalałaby na łatwy zapis struktury i własności molekuł w postaci interpretowalnej przez algorytmy komputerowe. Fingerprinty molekularne to zestaw metod mających na celu zapisanie związku w postaci ciągu bitów, w którym każdy bit reprezentowałby jakąś cechę strukturalną lub fizykochemiczną cząsteczki. Taki zapis pozwala z kolei na wykrycie ukrytych powiązań i zależności pomiędzy tymi właściwościami w związkach o niskim podobieństwie strukturalnym i np. klasyfikację tych związków na aktywne i nieaktywne względem wybranego celu biologicznego z użyciem algorytmów uczenia maszynowego (Machine Learning, ML). Obecnie najbardziej wydajnym podejściem są fingerprinty haszowane, gdzie pojedyncze bity nie reprezentują wyszczególnionej właściwości lecz raczej zestaw skorelowanych cech strukturalnych lub fizykochemicznych. Taka reprezentacja praktycznie uniemożliwia fizyczna

poprawną interpretację danych, gdyż nie można określić jaka cecha lub cechy odpowiadała za wartość wybranego bitu. Drugim podejściem są fingerprinty oparte na kluczach, w których każdy pojedynczy bit odpowiada dokładnie jednej, uprzednio zdefiniowanej podstrukturze cząsteczki. Te fingerprinty można znacznie łatwiej interpretować, jednakże osiągają gorsze wyniki w testach klasyfikacyjnych.

W celu ulepszenia obecnie dostępnych fingerprintów opartych na kluczach zaproponowana została nowa reprezentacja: Fingerprinty Sąsiadujących Podstruktur (Neighboring Substructure Fingerprint, NSFP). W tym podejściu poszczególne bity nie reprezentują pojedynczych podstruktur lub abstrakcyjnych grup, lecz pary sąsiadujących ze sobą podstruktur. Taka reprezentacja jest znacznie bardziej spójna i pozwala na lepsze odwzorowywanie związków chemicznych. Testy klasyfikacji aktywności związków ukazały, że nowa reprezentacja osiąga lepsze wyniki od wszystkich innych fingerprintów podstrukturalnych oraz niektórych haszowanych.

Druga część pracy dotyczy wykorzystania metodologii NSFP w celu wyszukania nowych, selektywnych ligandów dla receptorów serotoninowych 5-HT_{1B} i 5-HT_{2B}. W ramach badań użyto również szeregu innych metod CADD, w tym tych opartych na strukturze (wykorzystujących dodatkowo strukturę trójwymiarową badanego białka), takich jak dokowanie molekularne czy fingerprinty interakcji. W końcowej fazie badań, po wielopoziomowej kaskadzie przesiewowej wybrano 28 związków, wśród których kilka wykazało wysoką aktywność i selektywność względem 5-HT_{2B}.