Vytěžování znalostí ze souborů chemických dat na základě interpretace QSAR modelů

Pavlo Polishchuk

Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University and University Hospital in Olomouc, Hněvotínská 1333/5, 779 00 Olomouc, Czech Republic

Nedávno vyvinutý univerzální přístup pro interpretaci QSAR modelů může být využit k extrakci vztahů mezi strukturou a účinkem zachycených jakýmkoliv modelem bez ohledu na to, jaké učící metody a deskriptory byly využity k jeho konstrukci. Díky tomu je tento přístup velice perspektivní ve vytěžování znalostí z dostupných souborů chemických dat. Oproti metodám, které extrahují informace pouze na základě chemické struktury a příslušných vlastností (často se vyskytující a emergentní vzorce, korespondující páry molekul, atd.), tvorba QSAR modelů určitou míru jistoty, že daný vztah mezi strukturou a aktivitou byl opravdu pozorován a zachycen modelem. Vyvinul jsem open-source software pro strukturní a fyzikálně-chemickou interpretaci (SPCI, structural and physico-chemical interpretation) QSAR modelů, který jsem následně úspěšně aplikoval na několik datových souborů zahrnujících různé biologické a fyzikálně chemické parametry jako jsou prostupnost přes hematoencefalickou bariéru, toxicita, mutagenicita, rozpustnost a další ADMET parametry, inhibiční aktivita vůči fibrinogenovému receptoru, atd. Nástroj prezentuje vztah mezi strukturou a účinkem způsobem srozumitelným z chemického hlediska a demonstruje výhody vyvinutého přístupu oproti metodě korespondujících párů zejména pro relativně malé soubory dat. Program má grafické uživatelské rozhraní a může být používán výzkumníky s pouze základní znalostí QSAR modelování, protože všechny kroky tvorby modelu a jeho validace jsou automatizovány. Interpretace modelů může být také prováděna automaticky pomocí vyvinutých schémat případně pomocí výpočtu příspěvků vybraných strukturních motivů k sledované aktivitě. Vizualizace výsledků je možná jak přímo v programu, tak pomocí externích nástrojů. Nástroj je k dispozici na <http://qsar4u.com/pages/sirms_qsar.php>.

[1] Polishchuk, P.G.; Kuz'min, V.E.; Artemenko, A.G.; Muratov, E.N. Mol. Inf. 32 (2013) 843-853.