# Počítačová predikce syntetické dostupnosti

*Milan Voršilák 1, Daniel Svozil 1,2*

*1 CZ-OPENSCREEN & Laboratoř informatiky a chemie, Fakulta chemické technologie, VŠCHT Praha, Technická 5, 166 28 Praha 6*

*2 CZ-OPENSCREEN, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4*

Počet generovaných struktur při počítačovém *de novo* návrhu běžně dosahuje desítek či stovek tisíc, a proto je potřeba pro další zpracování odfiltrovat takové, které nemají požadované vlastnosti. Mezi tyto vlastnosti patří např. biologická aktivita, kdy v první fázi pouze rozhodujeme o tom, zda je či není daná látka aktivní, toxicita či syntetická dostupnost. Pro takové filtrování se běžně používají klasifikační metody, pro jejichž naučení je třeba mít k dispozici trénovací množinu obsahující příklady z obou dvou tříd. Takovou množinu je relativně snadné zkonstruovat v případě biologické aktivity či toxicity, kteréžto vlastnosti je možné experimentálně určit. Syntetická dostupnost je ovšem vlastnost, kterou nelze experimentálně charakterizovat a pro konstrukci klasifikátoru tak nejsou k dispozici příklady struktur obtížně syntetizovatelných. Proto jsme navrhli metodu Nonpher pro návrh synteticky nedostupných sloučenin založenou na algoritmu Molpher. Za použití Molpheru byly synteticky dostupné struktury transformovány pomocí jednoduchých strukturních změn (např. přidej/změň/uber atom/vazbu) na sloučeniny mnohem komplexnější a tudíž obtížně syntetizovatelné. Nonpher metoda byla porovnána s dalšími možnými přístupy pro návrh obtížně syntetizovatelných látek. Pro porovnání byla použita klasifikace pomocí *random forest* klasifikátoru. Kvalita jeho predikce byla vyhodnocena za použití externí testovací množiny sestávající z publikovaných látek, obtížnost jejichž syntézy byla konsensuálně určena několika medicinálními či organickými chemiky. Námi navržený postup Nonpher dosáhl ze všech testovaných postupů nejvyšší přesnosti (85%) a může proto významným způsobem přispět k dalšímu rozvoji počítačové predikce syntetické dostupnosti.