



ÚOCHB AV
ČR

Ústav organické chemie a biochemie
Akademie věd České republiky, v. v. i.



TISKOVÁ ZPRÁVA

Vědci z ÚOCHB AV ČR a CIIRC ČVUT objevují neznámé molekuly s pomocí AI

27. 5. 2025

Letošní nositel Ceny Neuron pro mladé nadějně vědce Dr. Tomáš Pluskal z Ústavu organické chemie a biochemie AV ČR spolu se svým studentem Romanem Bushuievem a kolegy z Českého institutu informatiky, robotiky a kybernetiky ČVUT (CIIRC ČVUT), Dr. Josefem Šivicem a Antonem Bushuievem, vyvinuli model strojového učení nazvaný DreaMS, který výrazně urychluje analýzu dosud neznámých molekul. Studii publikoval vlivný vědecký časopis *Nature Biotechnology*.

Příroda je plná chemických látek, které zatím čekají na své objevení. Předpokládá se, že neznáme naprostou většinu přírodních molekul. Jejich popis může přitom otevřít cestu k novým lékům, šetrnějším pesticidům, hlubšímu porozumění biologických procesů nebo k pokročilejšímu výzkumu života ve vesmíru.

Každá látka má unikátní vzor, podobný lidskému otisku prstu, kterému se říká hmotnostní spektrum a který dokáže zachytit metoda zvaná hmotnostní spektrometrie. Ačkoliv lze touto cestou získat obrovské množství dat, mimořádně obtížné je jim porozumět a odhalit přesné molekulární struktury. Získané datové sady se často jeví jen jako rozsáhlé tabulky čísel bez zjevného významu.

K rozluštění tajemství neznámých molekul použil tým z ÚOCHB AV ČR a CIIRC ČVUT umělou inteligenci. Podobně jako se velké jazykové modely typu ChatGPT učí porozumět jazyku, i když dopředu neznají význam slov, pokouší se model DreaMS porozumět hmotnostním spektrům, aniž by znal jejich chemické struktury. „*ChatGPT dokáže z velkého množství slov odvodit jejich význam a souvislosti mezi nimi a neuronová síť DreaMS pomocí samořízeného strojového učení zase rozpoznává, jaké molekulární struktury se za spektry skrývají. Využívá přitom údaje z milionů spekter,*“ vysvětluje Josef Šivic.

“Model DreaMS trénoval na desítkách milionů spekter z různých organismů a prostředí – z rostlin, mikrobů, potravin, tkání i půdních vzorků. Díky tomu dokáže nacházet skryté podobnosti mezi spektry, mezi nimiž na první pohled žádná souvislost není,” popisuje Tomáš Pluskal. Výsledkem je propojená síť usnadňující orientaci v obrovském množství chemických dat. Tuto síť, kterou je možné si představit jako internet hmotnostních spekter, nazvali vědci DreaMS Atlas. Každé spektrum je jako webová stránka spojená s ostatními. Na tomto „internetu spekter“ lze vyhledávat, sledovat objevené souvislosti a klást si nové otázky. Například, co mají společného pesticidy, potraviny a lidská kůže? DreaMS mezi nimi totiž odhalil nečekané chemické podobnosti a navrhl hypotézu, že určité pesticidy by mohly souviset s autoimunitním onemocněním, jako je lupénka.

Kromě propojování spekter z různých studií lze DreaMS využít i k dalším praktickým úlohám. Například k odhadu, kolik má molekula určitých fragmentů nebo jestli obsahuje konkrétní chemické prvky. „Překvapilo nás zejména, že se model naučil detekovat fluor,“ říká Roman Bushuiev: „Fluor se vyskytuje přibližně ve třetině všech léčiv a agrochemikálií, z hmotnostního spektra jsme ho dřív ale prakticky nedokázali odhalit. DreaMS, předtrénovaný na milionech spekter, jsme doladili na několika tisících příkladech molekul obsahujících fluor a najednou to fungovalo.“

Vědci teď pracují na dalším kroku. Učí model předpovídat celé molekulární struktury. Pokud se jim to podaří, zásadně to ovlivní náš pohled na chemickou rozmanitost, ať už na planetě Zemi, nebo dokonce ve vesmíru.

Původní článek: R. Bushuiev, A. Bushuiev, R. Samusevich, C. Brungs, J. Sivic and T. Pluskal, Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS Nature Biotechnology (2025)

<https://doi.org/10.1038/s41587-025-02663-3>

Ústav organické chemie a biochemie AV ČR / ÚOCHB (www.uochb.cz) je přední mezinárodně uznávaná vědecká instituce, jejímž hlavním posláním je základní výzkum v oblasti chemické biologie a medicínské chemie, organické a materiálové chemie, chemie přírodních látek, biochemie a molekulární biologie, fyzikální chemie, teoretické chemie a analytické chemie. Nedílnou součástí poslání ÚOCHB je přenos výsledků základního výzkumu do praxe. Důraz na mezioborové zaměření výzkumu ústí do řady aplikací v medicíně, farmacii a dalších odvětvích.

--- KONEC TISKOVÉ ZPRÁVY ---

KONTAKT PRO NOVINÁŘE:

Veronika Sedláčková (ÚOCHB – Komunikace): veronika.sedlackova@uochb.cas.cz

mob: +420 602 160 135