

Roman Bushuiev hledá nové léky v datech z přírody

5.1.2026 - Jakub Drahonský | Vysoká škola chemicko-technologická v Praze

Ing. Roman Bushuiev, doktorand VŠCHT Praha, získal v roce 2025 jako jediný z České republiky prestižní stipendium Google PhD Fellowship v kategorii Health research. Ve svém výzkumu propojuje strojové učení, chemii a biologii, aby pomocí AI odhaloval dosud neznámé biosyntetické dráhy rostlin. S bratrem Antonem vyvíjí modely, které mohou výrazně urychlit vývoj nových léků a proměnit způsob, jakým lidé objevují přírodní molekuly.

Co vás nejvíce ovlivnilo ve výběru bioinformatiky jako oboru, zejména když spolupracujete s bratrem Antonem na projektech jako DreaMS nebo MassSpecGym?

Ve výběru bioinformatiky mě nejvíce ovlivnil můj současný vedoucí Tomáš Pluskal, se kterým spolupracuji už šestým rokem. Předtím mě nejvíce bavila matematika, především abstraktní a do jisté míry až filozofické oblasti, jako je matematická logika nebo teorie kategorií. Při spolupráci s Tomášem jsem si postupně uvědomil, kolik zásadních a stále otevřených problémů v biologii lze řešit pomocí matematiky a informatiky – a že jejich řešení může mít reálný a často velmi praktický dopad. Postupně mě proto začaly více přitahovat aplikované než čistě teoretické směry. Dnes se věnuji strojovému učení, které přirozeně kombinuje matematiku a informatiku, a aplikuji ho na biologické problémy, zejména na objevování nových molekul z přírody. S mým bratrem Antonem v podstatě spolupracujeme na všech projektech a v mnoha ohledech se s ním velmi přirozeně doplňujeme.

Jaký byl přechod od studentského algoritmu pro biosyntetické reakce (oceněného cenou Via Chimica 2022) k publikacím v Nature Biotechnology?

Když jsem dokončil bakalářský projekt zaměřený na predikci biosyntetických reakcí, můj vedoucí Tomáš Pluskal mi doporučil začít pracovat na novém tématu a původní projekt předat nastupujícímu doktorandovi v naší skupině, Ramanovi Samusevichovi. Během bakaláře jsem pracoval hlavně s enzymy, zatímco následně jsem se začal věnovat datům z hmotnostní spektrometrie, kde je cílem anotace metabolitů. Tomuto tématu jsem se věnoval několik let a výsledkem byla mimo jiné publikace v Nature Biotechnology. Na začátku jsem, i kvůli absenci formálního biologického vzdělání, vůbec nevnímal propojení mezi těmito projekty. Až postupně mi došlo, že spolu velmi úzce souvisejí. Živé organismy lze chápat jako továrny produkující obrovské množství metabolitů a enzymy jako stroje, které v těchto továrnách krok za krokem přeměňují metabolity ze vstupních látek na finální produkty. Zpětně tak vnímám jako velkou výhodu, že mám zkušenosti s oběma těmito složkami – jak s enzymy, tak s metabolity – a dnes se je snažím propojit do jednotného přístupu k anotaci kompletních biosyntetických drah.

Jak byste porovnal první roky doktorátu na VŠCHT s předešlými zkušenostmi na UOCHB a ČVUT?

Upřímně řečeno jsem žádný zásadní rozdíl nepocítoval, kromě velké úlevy z toho, že už nemám tolik povinných předmětů. Už během magisterského studia mě práce na vlastních projektech pod vedením Tomáše Pluskala a Josefa Šivice natolik pohltila, že jsem se výuce věnoval spíše minimálně. První roky doktorátu pro mě znamenaly možnost se konečně soustředit téměř výhradně na výzkum. Bakalářské studium na ČVUT FIT ale považuji za velmi přínosné – program je na velmi vysoké úrovni, zejména v informatice a matematice, což jsou obory, které mě vždy bavily. Na ČVUT FIT jsem

se naučil programovat od úplných základů, což dnes využívám prakticky každý den.

Jak přesně váš model spojuje data o metabolitech, genech a enzymech, aby odhalil biosyntetické dráhy u rostlin - a jaký konkrétní léčivý metabolit jste už díky tomu identifikovali?

V současnosti se umělá inteligence na metabolity, geny a enzymy většinou aplikuje odděleně. Neznám žádný obecný výpočetní přístup, který by tyto modality systematicky propojoval, a i když se o to některé metody snaží, obvykle jsou navrženy pro velmi konkrétní molekulu nebo jednotlivou biosyntetickou dráhu. Vývoj univerzální metody je naším dlouhodobým cílem, mimo jiné i v rámci Google PhD Fellowship. Společně s bratrem současně vyvíjíme dva nové modely: DreaMS-Mol pro predikci struktur metabolitů a EnzymeseeK pro predikci enzymatických aktivit. Už od začátku je navrhujeme tak, aby mohly být vzájemně propojené do jednoho systému, který dokáže anotovat celé biosyntetické dráhy na základě kombinace dat o proteinech, například z RNA-seq experimentů, a dat o metabolitech, například z hmotnostní spektrometrie LC-MS/MS.

Pokud jde o konkrétní léčivé metabolity, je důležité říct, že jejich potvrzení je běh na dlouhou trať – typicky se mluví o zhruba deseti letech, než se molekula skutečně ověří jako léčivo. V této fázi náš výzkum spíše automaticky identifikuje velmi slibné kandidáty. Zajímavým příkladem je spolupráce s doktorandem z naší skupiny Andrejem Tekelem, který se zaměřuje na rostlinné látky obsahující atomy fluoru. Přibližně třetina současných léčiv obsahuje fluor, který je ale do molekul obvykle přidáván chemicky, zatímco biologických mechanismů je známo jen minimum. Andrejovi se podařilo identifikovat desítky nových takových molekul pocházejících z rostlin, mimo jiné i s pomocí našeho modelu DreaMS (zatímco dříve bylo známo jen několik relativně jednoduchých molekul). Jde o velmi zajímavý biologický objev s výrazným aplikačním potenciálem, například ve farmaceutickém průmyslu.

V čem spočívá největší riziko, že AI zkreslí predikce biosyntetických drah kvůli šumu v datech z hmotnostní spektrometrie, a jak ho řešíte v MassSpecGym benchmarku?

Je pravda, že jedním z hlavních problémů je, že data z hmotnostní spektrometrie často obsahují velké množství šumu. To může vést k tomu, že modely takzvaně halucinují a produkují biologicky nesmyslné predikce. V projektech MassSpecGym a DreaMS tento problém řešíme pomocí sady algoritmů, které se zaměřují na detekci a odstraňování šumu. Tyto postupy aplikujeme jak na trénovací a evaluační data, tak i na data od spolupracovníků, kteří modely používají v praxi. Díky tomu je výsledné nasazení modelů výrazně spolehlivější.

Co vás jako první přesvědčilo, že AI dokáže nahradit pokus-omyl v objevování přírodních molekul?

Myslím si, že umělá inteligence dokáže nahradit jakýkoli proces založený na pokusu a omylu, pokud existuje dostatek dat, která tyto pokusy reprezentují. V oblasti objevování přírodních molekul tuto roli plní databáze hmotnostních spekter, které se v posledních letech rozšiřují exponenciálním tempem. Naše laboratoř navíc vyvíjí vlastní experimentální metody, které umožňují efektivní generování nových dat a tento růst dále urychlují.

Jak by váš přístup mohl zkrátit vývoj léků na mrtvici nebo Parkinsonovu chorobu, a co byste potřeboval od farmaceutických firem pro reálné nasazení?

Vývoj nových léků obvykle začíná hledáním jedné správné molekuly mezi miliony až miliardami možností, což je extrémně náročný proces. Přírodní látky mají v tomto ohledu velkou výhodu, protože jejich struktury byly evolučně optimalizovány pro interakci s biologickými cíli, proto z nich vychází

značná část dnes používaných léčiv. Naše metody umožňují tyto molekuly z přírody objevovat automaticky, a tím výrazně urychlit úplně první fáze vývoje léčiv. Zároveň je objevování nových molekul důležité nejen pro vývoj nových léků, ale i pro diagnostiku onemocnění, například při hledání nových biomarkerů. U konkrétních nemocí, jako je mrtvice nebo Parkinsonova choroba, je klíčová úzká spolupráce s kliniky a farmaceutickými firmami, které se těmto onemocněním věnují desítky let. Jsem si jistý, že naše AI modely mohou být pro jejich výzkum užitečným nástrojem.

Když porovnáváte spolupráci s Tomášem Pluskalem (chemie) a Josefem Šivicem (AI/robotika), jaké nečekané lekce z jedné oblasti jste aplikovali do druhé?

Možnost spolupracovat současně s Tomášem Pluskalem a Josefem Šivicem je skutečně unikátní. Překvapivě pro mě nejdůležitější lekcí bylo vnímat chemii/biologii a AI jako jeden společný obor, například jako "výpočetní biologii". To je důležité například proto, že pro biologa nebo chemika je v praxi často jedno, zda používá sofistikovaný AI model nebo jednodušší algoritmus – zásadní je, aby byl nástroj přesný, spolehlivý a snadno použitelný. Když člověk přichází z výpočetního prostředí, jako já, a většinu času vyvíjí nové AI metody, je překvapivě snadné uváznout v technických detailech a zapomenout na to, co biologové nebo chemici skutečně potřebují k řešení praktických problémů. Toto uvědomění pro mě bylo klíčové a dnes na něj při vývoji nových metod kladu velký důraz.

Jste jediným oceněným Google PhD Fellowship z ČR. Co byste poradil doktorandům a doktorandkám na VŠCHT, aby zvýšili své šance na úspěch?

Z mé zkušenosti je zásadní pracovat s lidmi, se kterými se dlouhodobě výborně spolupracuje, a udržet si optimistický mindset. Všechny projekty dělám společně s bratrem Antonem a mám pocit, že když pracujeme spolu, naše produktivita neroste lineárně, ale spíše superlineárně – není dvakrát, ale třeba pětikrát či šestkrát vyšší. Za velké štěstí považuji také to, že jsem od začátku mohl pracovat s Tomášem Pluskalem a Josefem Šivicem, od kterých jsem se naučil obrovské množství věcí a v podstatě drtivou většinu toho, co vědecky umím.

Druhá věc je přístup k doktorátu samotnému. Často slyším, že je doktorandské studium extrémně stresující, že ho lidé „přežívají“ nebo že je to „peklo“. Já ho naopak vnímám jako velký luxus – dává člověku velkou svobodu v tom, čemu se chce věnovat, jaké otázky chce zkoumat a jakým směrem svůj výzkum rozvíjet. Pozitivní a optimistický přístup považuji za klíčový, aby člověk nezůstal uvězněn v „problémech“, které jsou ve skutečnosti jen zatím nevyřešenými výzkumnými otázkami.

Proč jste si vybrali Google PhD Fellowship v Health research před jinými kategoriemi?

Objevování biosyntetických drah má podle mě největší aplikační potenciál právě ve farmaceutickém výzkumu, takže volba kategorie Health research byla přirozená.

Jak vám stipendium změnilo priority ve výzkumu oproti standardnímu doktorandovi na VŠCHT?

Google DeepMind jsem vždy vnímal jako špičkové místo pro výzkum v oblasti AI. Projekt AlphaFold z DeepMind, který nedávno získal Nobelovu cenu, je pro mě velkou inspirací i pro můj současný projekt DreaMS-Mol: AlphaFold řeší predikci struktur velkých biomolekul, jako jsou proteiny, a my vyvíjíme metodu pro predikci struktur malých molekul. Mým cílem je proto do budoucna s Google DeepMind spolupracovat. Zároveň stipendium poskytuje výrazný finanční bonus, který nemění jen výzkumné priority, ale do určité míry i život obecně. Dává větší svobodu a prostor soustředit se na dlouhodobé a ambiciózní cíle.

Jaké máte další plány do budoucna?

Určitě chci pokračovat v biologickém strojovém učení. Zatím ještě nevím, v jaké formě – zda se s bratrem vydáme cestou vlastního startupu, zůstaneme v akademickém prostředí, nebo budeme pracovat v průmyslu, například v týmu typu DeepMind. V každém případě chci v této oblasti zůstat. Mám pocit, že jsme teprve na úplném začátku kariéry a že potenciál tohoto směru je obrovský. Věřím, že během příštích deseti let dokážeme udělat v biologii, a hlavně v objevování molekul, skutečně výrazný dopad.

<https://www.vscht.cz/popularizace/rozhovory/roman-bushuiev>