

„V přírodě jsou miliony proteinů, které ještě neznáme,“

říká bioinformatička Monika Rosinská

Jen málokdo se dokáže pohybovat na pomezí dvou tak rozdílných a zároveň náročných světů, jako jsou informatika a živé vědy. Vyžaduje to totiž analytické myšlení, hluboké znalosti obou oborů a především schopnost nacházet nové souvislosti. Vystudovaná back-end developerka Monika Rosinská se před dvěma lety pustila do studia životního prostředí a spojení těchto dvou oborů si nemůže vynachválit. Zabývá se vývojem nástrojů pro analýzu a návrh proteinů, letos se objevila mezi oceněnými v rámci Ph.D. Talent Jihomoravského kraje, programu JCMM, který podporuje mladé vědecké talenty.



V rámci své doktorské práce pracujete na nástroji analyzující proteiny. Můžete popsat, v čem konkrétně váš výzkum spočívá?

Zabývám se enzymy, což jsou druhy proteinů. Jsou to molekuly, které se vyskytují všude v přírodě i v lidském těle. A provádí opravdu hodně práce. Spousta lidí ví, že proteiny jsou důležité pro tvorbu svalů, nezbytné jsou však i pro spoustu jiných procesů, třeba trávení. Na činnosti enzymů je založena například také spousta pracích prostředků. Většina enzymů, které se v přírodě běžně vyskytují, jsou stvořeny na přirozené podmínky. Pokud je však chceme využít v produktech, různých chemických procesech a podobně, musíme je upravit, aby fungovaly tak, jak chceme. Laboratorně je tedy měníme, mutujeme. V nedávné době se také ověřilo, že v přírodě se nachází spousta proteinů a enzymů, které ještě neznáme a které jsou přizpůsobeny náročnějším podmínkám. Může jít o proteiny z organismů, které žijí v nějakých vřídlech a podobně. Existují databáze těchto proteinů, které se sekvenovaly z organismů, ale nikdo je nikdy neanalyzoval, a tak nevíme, jaké mají vlastnosti a v jakém prostředí fungují. A nástroj EnzymeMiner, který zpracovávám v rámci svého doktorského studia, tyto databáze

prohledává a snaží se najít ty nejvhodnější. Zaměřuje se na konkrétní části proteinů, které jsou pro zadavatele důležité. Následně nabídne takové, které mohou vykazovat požadovanou aktivitu, a vypíše i další anotace, třeba teplota nebo pH, ve kterém fungují.

Kolik takovýchto ještě neprobádaných proteinů může existovat? Stovky, tisíce?

Bavíme se dokonce o milionech! Momentálně vznikají metagenomické databáze, kdy výzkumník vezme vzorek z půdy nebo jiného prostředí a začne sekvenovat molekuly, které tam najde. Občas to jsou celé proteiny, občas je to DNA, někdy zase jen nějaký fragment. A právě v těchto databázích se mohou ukrývat ty skryté diamanty, které mohou mít přesně takové účinky, které potřebujeme.

Pokud už databáze najde odpovídající proteiny, jak dlouho trvá experimentální ověření, jestli skutečně fungují tak, jak chceme?

Minimálně několik měsíců. Je potřeba vytvořit gen pro daný protein, gen vložit do buněk, dostat ten protein z buněk. A až pak začít s ověřováním, což je samo o sobě dlouhodobá záležitost. V průběhu toho také můžeme přijít na to, že ten protein v našich podmínkách nejde vyprodukovat, nebo že nejde vyprodukovat v organismu, který používáme.

V jakých organismech proteiny produkuje?

Nejčastěji se používají *E.coli* buňky, případně kvasinky. Protein se dá samozřejmě koupit i laboratorně vytvořený, ale to je podstatně nákladná záležitost, takže je to otázka rozpočtu, který na to vyhradíme. Ověřování proteinu je ale hodně individuální, někdy může trvat i roky.

Jakým způsobem v práci využíváte umělou inteligenci a strojové učení?

Verze EnzymeMineru, kterou jsem pomáhala vytvořit, využívá na anotaci proteinů hodně nástrojů strojového učení. Vezme sekvenci proteinu a na jejím základě se snaží predikovat jeho vlastnosti, například v jaké teplotě bude protein fungovat. Tyhle věci je poměrně těžké dělat statisticky. U některých je to naprosto nemožné, takže strojové učení může poskytnout užitečný pohled. Člověk musí mít na paměti, že to není stoprocentní, ale může to filtrovat ty zajímavé případy vhodné k dalšímu testování.

Dala jste se do studia oboru Životní prostředí a zdraví, kde zkoumáte chemické procesy v přírodě a to, jak mohou jednotlivé chemikálie ovlivnit člověka i celý ekosystém. Proč jste si po studiu informatiky vybrala právě tento obor?

Mým podoborem je Matematická biologie, bioinformatika a modelování, což se týká i právě analýzy enzymů, proteinů a DNA v organismech. Byla to pro mě možnost, jak zůstat u

informatiky, kterou jsem studovala, a zároveň proniknout více do světa bioinformatiky. Obor mi navíc umožňuje studovat i pracovat zároveň. Pokud bych pracovala třeba v nějakém IT korporátu, skloubit to s doktorátem není jednoduché... Informatika mě baví, zároveň bych ráda pomohla přírodě i lidskému zdraví.

Je to poměrně unikátní kombinace, lidí s tímto záběrem není mnoho... A řekla bych, že i jako žena jste mezi informatiky musela být v menšině!

To jsem byla. (smích) Poslední roky už počet žen v tomto oboru stoupá. Když jsem ale nastoupila, byly jsme jedna studentka na padesát studentů.

Jak se vám v takovémto kolektivu studovalo?

Mně tam bylo moc fajn. Mezi kluky jsem vyrůstala, rozumím si s nimi. Občas bylo trochu těžké si vybojovat místo v týmových projektech, spousta spolužáků si rádo dělalo věci sami a po svém. Že se chci podílet na důležitých částech práce, jsem si musela prosadit. Často měli představu, že já napíšu dokumentaci a udělám testování a do té vyložené algoritmické části se pouštěli hlavně oni.

Co vás na tom baví nejvíc?

Já hrozně ráda řeším problémy. Baví mě, když na řešení člověk přichází postupně, v oboru informatiky si člověk může vyzkoušet spoustu cest. Pořád je co objevovat.

Je to ohromně rychlý a proměnlivý obor, dá se vůbec zvládat to tempo a se všemi novinkami držet krok?

Za začátku to bylo hodně náročné, přece jen jsem se k té bioinformatice dostala až na magisterském studiu. Zvládnout celou tu teorii, celý ten obsáhlý základ není jednoduché. Zjistila jsem, že člověk se opravdu musí soustředit jen na velmi úzkou a specifickou část tohoto oboru a s tou pak držet krok a ve volném čase si rozšiřovat obzory a dostávat se do širšího kontextu oboru. Pro mě zpočátku bylo opravdu nemožné obsáhnout nějaký širší kontext už té samotné bioinformatiky zaměřené na proteiny.

Říkáte, že je ještě hodně proteinů, respektive jejich funkcí a možností, zůstává stále neobjevená. O jak velké části spektra mluvíme?

Obecný názor je, že většinu reakcí už objevenou máme. Nemyslím si, že bychom objevili vyloženě novou reakci, která je naprosto odlišná od ostatních. Ale asi mohou existovat enzymy, které mají odlišný mechanismus od těch, které momentálně známe.

A pokud už známe funkci proteinu, můžeme na základě toho odvodit veškeré biologické procesy, které se v něm odehrávají?

Ne. My vidíme, jak reaguje, že konkrétní protein provádí to, co chceme, ale výzkum samotného mechanismu je velmi náročný. Sledovat protein „v akci“ je velmi těžké. Používají se simulace, je tam kvantová mechanika a je to opravdu složitý proces. Na jednom proteinu pracuje několik lidí několik let, pokud se to vůbec podaří. Je to práce s dost nejistým výsledkem.

Kam si myslíte, že se tohle téma, ať už konkrétně vaše práce nebo obor jako takový, posune za 5, za 10 let?

Tohle je opravdu těžká otázka. Nevíme ani, kde bude za rok. Určitě by bylo úžasné, pokud bychom vytvořili nástroj strojového učení, který by dokázal pomoci buď s designem proteinu, nebo právě s jeho anotací, tedy že bychom na základě odebrané sekvence zvládli určit, jakou funkci protein má. Může se to stát hned, nebo se to nemusí podařit ani za deset let... Těžko říct.

Když zrovna nepracujete, čím si nejraději vyčistíte hlavu?

Sportuji, jezdím na kole nebo cvičím. Teď bych si zase ráda našla čas na to jezdit zase na koních. Taký čtu knížky, od detektivek, přes romány až po science fiction.

martina.kopecka@fnusa.cz