

**Posudek doktorské práce Miloše Musila s názvem
“Automatizovaný návrh stabilních proteinů”
Vypracoval: Ing. Matej Lexa, PhD**

Posuzovaná práce Miloše Musila patří do oblasti bioinformatiky proteinů, jedná se o kombinaci aplikované informatiky s doménovými znalostmi struktury proteinů a evoluce proteinů, respektive proteinových sekvencí. Formálně práce představuje 123-stranový text (83-stran bez příloh) obsahující jako přílohy klíčové publikace, na kterých se uchazeč podílel významnou mírou. Text je psán v anglickém jazyce na výborné úrovni, s výjimkou opakovaných prohrěšků nadměrného a tím i nevhodného užívání určitých členů. Tyto se vyskytují hlavně v anotaci, ale několik případů jsem zaznamenal i v textu práce. Jako přílohy obsahuje dizertační práce 4 publikace a CD se soubory dalších publikací s hodnocením některých z nich.

Po obsahové stránce svědčí úvod textu o vynikající orientaci uchazeče v problematice stability proteinů, práce se sekvencemi proteinů, existenci nástrojů pro práci s tímto typem dat a taky výhodami a nevýhodami jednotlivých přístupů. Většina problematiky je prezentovaná srozumitelně a systematicky do míry, že tuto část textu by šlo lehce využít jako učební materiál pro výuku principů konstrukce fylogenetických stromů či faktorů ovlivňujících stabilitu proteinů. Text obsahuje jasně definované cíle dizertace a následně pak stručný popis jejich dosažení, který pomáhá spojit tuto část s přílohami, popisujícími několik nástrojů, které student vytvořil nebo se podílel na jejich tvorbě. Podíl je dán procentuálně a krátkými zmínkami o způsobu participace. Deklarované cíle jsou (kromě orientace v problematice stability proteinů) sběr výchozích dat, generování stabilnějších variantů pro libovolný protein a jako mezikrok této činnosti práce s rodinami příbuzných sekvencí za účelem nalezení jejich nejpravděpodobnější výchozí (“ancestrální”) podobu. Z textu je zřejmé, že hlavním autorským dílem i objektem bádání je prohlížečová verze softwaru FireProt a hodnocení jeho úspěšnosti v generování stabilních variantů vstupních proteinů.

Z hlediska proteinového inženýrství je zajímavé, že ve FireProt se povedlo vytvořit výpočetní postup využívající několik vzájemně ortogonálních přístupů k hodnocení mutací – strukturní, evoluční a statistický a za pomoci asi 16 existujících nástrojů tyto integrovaně vyhodnotit k předpovědi stability mutovaného proteinu. Tato část práce je ale zřejmě dlem jiného autorského kolektivu. V prezentované verzi by mohl být problémem malý vzorek proteinů, na kterých byl nástroj validován. Úzké spektrum validačních experimentů mírně snižuje kvalitu a hodnotu jinak nesporně zajímavé a úspěšné práce. Na druhé straně je součástí práce i integrovaná sbírka dat o stabilitě proteinů, která samotná je příspěvkem k vyšší kvalitě práce s daty podobného typu. Je taky pravda, že nástroj má dle slov autorů sloužit k prioritizaci mutací a jako takový může být přínosem i když schopnost posuzovat mutace a jejich kombinace není 100% spolehlivá. Kvalitu dílčích výstupů kromě jiného dokládá i jejich publikace v renomovaných vědeckých časopisech. V případě klíčových prací k software FireProt se jedná o Nucleic Acids Research a Briefings in Bioinformatics, oba s IF přes 10.

K některým částem práce a tvrzením v textu mám drobné připomínky, některé k případné diskusi:

1) Ke schémě a použitému postupu implementovanému ve FireProt:

- “Correlation v. conservation” Může být někdy problém rozeznat funkčně důležité korelace od korelací z evolučních důvodů (pokud například nastal v evoluci na některých pozicích tzv. “bottleneck”)?

- Jelikož obě větve výpočtů využívají FoldX a Rossettu, v čem je principiální rozdíl zaručující, že předpovědi ve dvou větvích nebudou stejné?

První část schémy je velice podobná HotspotWizardu. Nevzniká tím nějaký problém, kdy data ve FPdb jsou anotována HotspotWizardem a podobný postup je zároveň používán k analýze těchto dat?

2) V práci se u motivace několikrát opakuje, že proteiny v buňce jsou vystaveny mírným podmínkám oproti podmínkám například industriálním. Proteiny jsou ale molekuly, které mohou existovat ve velkém rozsahu teplot, tlaků, koncentrace solí a jiných podmínek prostředí. Objektivnější shodnocení situace by podle mě bylo konstatování, že proteiny, které se v rámci evoluce vyvinuly pro funkce v relativně dobře definovaném prostředí buňky, nemají složení vhodné pro jiné situace, bez odkazování se na “mírné” či “drsné” podmínky.

3) V práci je zmínka o velkém množství sekvencí proteinů (oproti strukturám), které jsou k dispozici díky novým metodám sekvenování DNA. Určitým problémem při charakterizaci mutací by mohla být existence pseudogenů či jiných sekvencí, které lze sekvenovat, považovat za kódující, ale ve skutečnosti nemusí plnit plnohodnotné biologické funkce. Jaký vliv by měli takové sekvence na analýzu stability, zejména s využitím mnohočetného zarovnání?

Vzhledem k uvedeným skutečnostem považuji práci za **aktuální**, vypracovanou v některých aspektech na **dobré a v jiných na vysoké úrovni a demonstrující schopnost uchazeče samostatně řešit vědecké problémy. Doporučuji proto uznat práci jako dizertační a udělit na její základě titul PhD.**

Další otázky k obhajobě:

1. V práci uvádíte, že většina programů neřeší změnu stability proteinů s více mutacemi. V tabulce 5.3 a 5.3 nicméně uvádíte software jako MAESTRO, ELASPIC, Prethermut, FRESCO a PROSS. Některé existují již mnoho let. Mají tyto přístupy nějaké slabiny, které bylo potřeba řešit novým postupem a které z nich se Vám povedlo vyřešit?

2. Jak je Váš software využíván vědeckou komunitou a jakou zpětnou vazbu jste od ní dostal?

V Brně 9.8.2021

podpis