

Posudek oponenta bakalářské práce

Jméno a příjmení uchazeče/ky : Evžen Wybitul

Název práce: Differential discovery of protein features using tandem mass spectrometry

A. Bodové hodnocení jednotlivých aspektů práce (označte právě jednu z možností)

1. Rozsah BP a její členění	
✓	A - přiměřené, odpovídají charakteru BP a významu jednotlivých částí
	B - nevyrovnané, členění není logické n. rozsah jednotlivých částí nekoresponduje s jejich významem
	C - uspokojivé, rozsah některých částí nedostačuje
	N - nedostatečné

2. Odborná správnost	
✓	A - výborná, bez závažnějších připomínek
	B - velmi dobrá, s ojedinělými drobnými závadami (nejasnost výkladu, chyby ve vzorcích nebo chemických názvech, nedokonalý popis metod nebo výsledků)
	C - uspokojivá, s čtenějšími drobnými závadami
	N - nevyhovující, s hrubými chybami

3. Uvedení použitých literárních a j. zdrojů	
✓	A - bez připomínek, všechny převzaté údaje s citací zdroje, celkový počet citací odpovídá charakteru práce
	B - uspokojivé, s občasnými neobratnostmi zejm. v umístění odkazů, nebo s celkově nižším počtem citací
	C - s vážnějšími závadami, např. převažují "nestandardní" odkazy na učebnice, přednášky, webové stránky, nebo se ojediněle vyskytuje opominutí odkazu na zdroj převzatých dat
	N - nevyhovující, velmi málo citací, ev. rysy plagiátu (časté opomíjení odkazu na zdroj převzatých dat, popř. opsání velkých částí textu)

4. Jazyk práce	
	A - výborný, práce je napsána čtivě a srozumitelně, bez závažnějších gramatických n. pravopisných chyb
✓	B - velmi dobrý, ojedinělé stylistické neobratnosti, gramatické n. pravopisné chyby
	C - uspokojivý, čtenější slohové neobratnosti, gramatické n. pravopisné chyby, ojediněle se vyskytují obtížně srozumitelné n. nejednoznačné formulace
	N - nevyhovující, s četnými hrubými chybami

5. Formální a grafická úroveň práce	
✓	A - výborná, bez překlepů a chyb ve formátování
	B - velmi dobrá, ojedinělé chyby formátu citací, překlepy, chybějící zkratky apod.
	C - uspokojivá, s ojedinělými většími (např. vynechání stránky) nebo čtenějšími drobnými chybami
	N - nevyhovující, s četnými hrubými chybami

Případný slovní komentář k bodům 1. až 5.:

Práce je prakticky a implementačně zaměřená a adresuje poměrně složitý problém identifikace disulfidických můstků v proteinech pomocí hmotnostní spektrometrie. Navzdory praktickému zaměření práce jsou aplikované algoritmy dobře teoreticky popsány a zdůvodněny (např. pomocí teorie grafů). Samotný přístup k řešení problému pomocí generování databáze všech možných fragmentů a jejich porovnání s experimentálními hmotnostními spektry odpovídá obecným trendům v proteomice, není tedy ani zastaralý ani výjimečně inovativní. Za velkou výhodu představeného softwaru Dibby považuji možnost využití dat jak ze standardních proteinových extraktů, tak z redukovaných proteinů bez disulfidických můstků („RAT“ přístup). Po formální a grafické stránce je práce na velmi dobré úrovni, s ojedinělými gramatickými či faktickými chybami. Úvodní kapitola o hmotnostní spektrometrii by mohla být trochu stručnější (např. úvod do různých hardwarových architektur hmotnostních spektrometrů a ionizace je v kontextu práce celkem zbytečný) a lépe organizovaná (autor často užívá pojmy, které jsou představeny až později v práci).

Obsahuje-li práce i vlastní výsledky uchazeče/ky (nejsou povinnou součástí práce), pak prosíme o Vaše stanovisko k následujícímu:

Jsou řádně stanoveny a vysvětleny cíle experiment? **Ano.**

Je množství experimentů adekvátní k cílům? **Ano.**

Je dokumentace výsledků dostačující? **Částečně (dokumentace analýz je dobrá, dokumentace softwaru nikoli).**

Jsou výsledky diskutovány a zasazeny do kontextu existující literatury? **Ano.**

Případný další slovní komentář k výsledkům autora:

Cíle vyvinutého softwaru Dibby jsou dobře popsány a výsledky jsou důkladně zhodnoceny. Slabší stránkou práce je dokumentace samotného softwaru, která spočívá pouze v jednom README souboru na GitHubu, který popisuje, jak se program spouští. Chybí informace o vyžadovaných knihovnách či o přesném formátu vstupních a výstupních dat. Zdrojové kódy samotného programu neobsahují žádné komentáře, vyznat se v nich tedy může do budoucna jen autor sám.

B. Obhajoba

Dotazy k obhajobě (povinná část posudku)

1. Vzhledem k praktickému zaměření práce, nebylo by vhodnější software Dibby implementovat jako modul do nějakého široce používaného toolboxu pro proteomiku jako např. OpenMS? Práce tím mohla získat výrazně větší publikum a impakt.
2. Vektory váhových koeficientů pro prekurzor wP a fragmenty wF jsou v práci natvrdo zakódovány jako $wP = (32,4,4,4)$ a $wF = (16, 4, 4)$. Racionální zdůvodnění těchto hodnot chybí, což považuji za slabinu. Šlo by tyto hodnoty optimalizovat například učením se z existujících dat? Kolik takových dat je k dispozici a jak byste to implementoval?
3. Při testu na datech dobře charakterizovaných proteinů program odhalil pouze část známých disulfidických můstků (3 ze 4 u LYS a 1 ze 3 pro LIP). Ve kterém kroku celého algoritmu myslíte, že je slabé místo, a máte nějakou představu, jak ho dále vylepšit?

Stanovisko k opravě chyb v práci:

opravný lístek/oprava v textu **JE** / **NENÍ** (zakroužkujte) podmínkou přijetí práce

C. Celkový návrh

Práci **doporučuji** k přijetí k dalšímu řízení: **ANO** / **NE**

Navrhovaná celková klasifikace: **1**

Datum vypracování posudku: 10.9.2021

Jméno a příjmení, podpis oponenta (SIS): Tomáš Pluskal

