

## Posudek diplomové práce Kateřiny Břicháčkové „Využití anotací primární struktury pro strukturní predikci protein-ligand aktivních míst“

Oblastí zájmu posuzované diplomové práce je predikce protein-ligand vazebných míst. Konkrétně se práce zaměřuje na zkoumání obecných vlastností známých vazebných míst na velkých datasetech. V práci je představena obecná metoda, která je použita na testování 33 vybraných vlastností zahrnujících jak experimentálně naměřené, tak predikované vlastnosti. Pomocí statistických metod je pak zjišťováno, které z těchto vlastností jsou využitelné pro predikci vazebných míst. Konkrétně je pro každou vlastnost zkoumáno (pomocí testování hypotéz), zda se její hodnoty naměřené v aktivních místech a mimo ně statisticky významně liší. A dále zda je tento rozdíl hodnot dostatečný (pomocí měření velikosti účinku). Měření probíhalo na čtyřech různých datasetech. Praktický význam zjištěných výsledků byl následně ověřován využitím těchto zkoumaných vlastností v predikční metodě P2Rank.

Použité statistické metody jsou spíše základní. Sama práce poukazuje například na to, že použité statistické metody nejsou dost citlivé na to, aby dokázaly dopředu určit, který ze tří způsobů měření konzervace sekvence bude pro predikci ten nejlepší. Souhlasím však s prací, že metoda poskytuje dostatečné výsledky, aby určila, kterými vlastnostmi vazebných míst má smysl se zabývat a které jsou naopak pro predikci spíše nepodstatné.

Při posuzování praktického využití zkoumaných vlastností v metodě P2Rank bohužel nebylo dosaženo žádného přesvědčivého zpřesnění metody P2Rank (nad rámec již používaného rozšíření). Přidáním zkoumaných vlastností k základním vlastnostem již obsaženým v P2Rank došlo jen k malému zlepšení predikce (v jednom zkoumaném parametru dokonce došlo k mírnému zhoršení). Podstatného zlepšení bylo dosaženo jen přidáním vlastnosti měřící konzervaci sekvence, ale toto rozšíření již bylo s P2Rankem použito dříve. Některé prakticky prováděné experimenty mají dle mého názoru jisté nedostatky či omezení, pokud jde o jejich vypovídací hodnotu. Ovšem práce si jich je vědoma a explicitně je zmiňuje.

Při testu, kdy byly jednotlivě přidávány zkoumané vlastnosti k základním vlastnostem P2Ranku, vyšla jako užitečná přesvědčivě pouze míra konzervace sekvence. Práce správně konstatuje, že je to jistě důsledkem i toho, že v případě ostatních zkoumaných vlastností jsou tyto již ve stejné nebo podobné podobě zahrnuty mezi základními vlastnostmi užívanými P2Rankem. Nabízí se pak však jisté pochyby o účelnosti takového experimentu. Pokud je již velmi podobná vlastnost v P2Ranku zahrnuta, pak nám takový experiment nemůže mnoho říct o tom, jak moc je daná vlastnost pro predikci důležitá. A na druhou stranu zcela zřetelně takový experiment ani neodpovídá na otázku, zda je vhodné takovou vlastnost do P2Ranku přidat, neboť v praxi by asi mělo dojít místo případného přidání této vlastnosti spíše k nahrazení za onu velmi podobnou původní vlastnost.

Výsledky navíc ukazují, že přidáním menšího počtu vhodných zkoumaných vlastností dojde k většímu zlepšení predikce, než když jsou přidány všechny. A k úplně největšímu zlepšení predikce došlo v případě, že k základním vlastnostem používaným v P2Ranku byla přidána pouze vlastnost určující konzervaci sekvence. Nabízí se tedy otázka, zda byly predikce využívající větší počet (korelujících) vlastností vhodně trénovány. Práce si tuto otázku v závěru sice klade, ale dále ji nezkoumá. V této souvislosti by bylo jistě zajímavé seznámit se s tím, jak dobře predikce fungovaly na datasetu, na kterém byly trénovány, a zda se i zde projevil stejný pokles úspěšnosti predikce.

I přes tuto drobnou polemiku práce určitě splnila cíle a přinesla obecnou metodu, která pomůže snadno otestovat přínos nově navržených vlastností. Zároveň ukázala, že je možné nahradit jednu pomalou metodu počítající míru konzervace sekvence (která je aktuálně používána rozšířenou verzí P2Ranku) jinou, která je mnohem rychlejší a zároveň přináší stejnou nebo dokonce větší úspěšnost predikce. Což je jistě hezký praktický výsledek.

Celkově se mi práce líbila, hodnotím ji jako zdařilou a doporučuji k obhájení.

V Praze dne 14.2.2021

Jakub Galgonek