

Abstrakt

V posledních 20 letech se počet experimentálních proteinových struktur v databázi Protein Data Bank rychle zvyšuje, což motivuje vývoj nástrojů pro predikci protein-ligand vazebných míst. Strukturní predikce vazebných míst má mnoho důležitých aplikací; je klíčovým krokem v komplexním procesu návrhu léčiv, pomáhá objevovat vedlejší účinky léčiv, umožňuje chápat biologické funkce proteinů a je využívá se i v mnoha jiných oborech, jako je protein-ligand docking nebo molekulová dynamika. Pokud je autorce známo, dosud nebyla provedena studie, která by systematicky zkoumala obecné vlastnosti známých vazebných míst na velkých datasetech. Tato práce se zaměřuje na analýzu těchto vlastností, s využitím existujících experimentálních i predikovaných anotací primární a terciální struktury proteinu. Je zde představena metoda pro statistickou analýzu těchto anotací, která je založena na testování hypotéz a odhadu velikosti účinku. Metoda je implementována v jazyce Python a lze ji jednoduše rozšířit o nové anotace definované uživatelem. Použití je demonstrováno na 33 existujících anotacích a čtyřech různých datasetech. Praktická významnost výsledků je otestována s pomocí metody P2Rank. Výsledky i samotná metoda pro statistickou analýzu by mohly posléze přispět ke zlepšení úspěšnosti existujících nástrojů pro predikci vazebných míst.