

Abstrakt

Ve farmaceutickém výzkumu a vývoji hrají enzymy důležitou roli při syntéze léků a molekul souvisejících s léky. Pro vyšší účinnost a zvýšenou produkci je důležité optimalizovat výtěžek těchto enzymů, což je úkol často řešený proteinovým inženýrstvím a designem. Tento proces optimalizace enzymu se však může stát zdlouhavým s obrovskými možnostmi mutací pro každý jednotlivý protein.

Ke zlepšení procesu optimalizace enzymů lze použít sekvenční homologii a metody strojového učení. Ty výrazně snižují manuální úsilí při přepracování proteinů a mohou pomoci při hledání nejvhodnějšího enzymu pro daný úkol, čímž se zvyšuje efektivita celého procesu vývoje léčiv. S ohledem na tento cíl tato práce shrnuje výběr existujících metod a jejich možné aplikace při optimalizaci enzymů. Testování dvou prediktivních modelů s různou složitostí na 4 souborech dat ve snaze optimalizovat absorpci, enantiosektivitu, nebo teplotní stabilitu zjistilo pouze mírnou korelaci mezi skutečnými cílovými hodnotami a jejich předpokládanými hodnotami: průměr Pearsonova R 0,20775 a 0,5188. Porovnání pravděpodobnostních vzorů vložení proteinových sekvencí vedlo ke korelačnímu skóre 0,815 se substituční maticí BLOSUM62, což potvrzuje intuici jazykového modelu o přirozené frekvenci různých typů mutací.

Zatímco výsledky prediktivních modelů nejsou slibné pro výrazné zvýšení efektivity pomocí výpočetních přístupů, prezentované modely ponechávají prostor pro zlepšení a naznačují možné zvýšení prediktivního výkonu doladěním reprezentace sekvencí pro danou úlohu.

Klíčová slova: optimalizace enzymů, řízená evoluce, redesign proteinů, farmacie, strojové učení, sekvenční homologie