

Nástroj pro automatickou predikci bodových mutací z konsenzuální sekvence

Konsenzuální sekvence je překvapivě účinným nástrojem pro návrh termostabilizujících mutací [1]. Konsenzuální sekvenci můžeme získat například zhuštěním multičetného zarovnání blízké příbuzných proteinových sekvencí do jediné sekvence takovým způsobem, že na každou pozici umístíme buď aminokyselinu v tomto zarovnání převládající nebo symbol „-“ nemá-li pozice jasného vítěze. Ukazuje se, že záměna aminokyseliny daného proteinu za tu aminokyselinu, která se vyskytuje v konsenzuální sekvenci, bývá až v polovině případů termostabilizující [1, 2, 3].

Pro úspěšnost metody je klíčové mít k dispozici dostatečné množství (typicky více než 10) vysoce příbuzných sekvencí (sekvenční identita okolo 60 %). Jsou-li sekvence málo příbuzné, konsenzuální sekvence se odlišuje od jednotlivých proteinů v příliš mnoha pozicích a nedá se jednoznačně určit, které z nich vybrat. V takovém případě může proteinový inženýr vyloučit mutace, které podle jeho zkušeností mohou narušovat strukturu nebo funkci proteinu [3].

Cílem projektu je vyvinout a implementovat algoritmus, který na vstupu dostane protein (PDB soubor), vygeneruje konsenzuální sekvenci a s její pomocí seřadí rezidua podle perspektivnosti pro termostabilizující mutaci.

Algoritmus bude začleněn do nástroje [HotSpot Wizard](#), který prozatím slouží pouze k vytipování reziduí, jejichž mutace pravděpodobně vede k dosažení nových katalytických vlastností enzymu. Implementace bude využívat externí webovou službu pro vyhledání homologních sekvencí v databázi UniProt. Vícečetné zarovnání sekvencí může být prováděné buď lokálně nebo vzdáleně. Výstupem algoritmu bude seznam mutací seřazený podle funkce

$$\Delta \Delta G = -RT \ln \frac{f_{\text{consensus}}}{f_{\text{original}}} \quad [1],$$

ve kterém vystupují frekvence aminokyselin na dané pozici v zarovnání. Symbol $f_{\text{consensus}}$ označuje počet výskytů nejfrekvencovanější aminokyseliny v tomto zarovnání a f_{original} je frekvence aminokyseliny z našeho proteinu. Čím výše v seznamu je mutace uvedena, tím vyšší je pravděpodobnost, že je termostabilizující. Uživatel může experimentálně testovat přínos mutací shora a zastavit se podle klesající úspěšnosti nebo svých časových a finančních možností.

Další vylepšení (zřejmě již nad rámec projektu) mohou spočívat ve vážení sekvencí (blízce příbuzné sekvence se na konsenzu budou podílet menší vahou), využití fylogenetického stromu ke konstrukci konsenzuální sekvence a zpřesnění hodnotící funkce pomocí nástrojů pro výpočet energetického přínosu jednobodové mutace nebo pomocí strojového učení.

1. B. Steipe, Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes, *Methods Enzymol.* 388 (2004), 176–186.
2. M. Lehmann b, L. Pasamontes b, S.F. Lassen c, M. Wyss: The consensus concept for thermostability engineering of proteins, *Biochimica et Biophysica Acta* 1543 (2000) 408-415
3. K. M. Polizzi, J. F. Chaparro-Riggers, E. Vazquez-Figueroa and A. S. Bommarius: Structure-guided consensus approach to create a more thermostable penicillin G acylase, *Biotechnol. J.* 1 (2006), 531–536