# Výskyt sekundárních struktur v sekvenčním prostoru proteinů

Treťjačenko V1, Vymětal J2, Bednárová L2, Vondrášek J2 and Hlouchová K1,2

1 Univerzita Karlova v Praze, Přírodovědecká fakulta, Katedra biochemie, Hlavova 2030, 128 00 Praha 2

2Ústav organické chemie a biochemie , Flemingovo náměstí 2, 166 10, Praha 6

V proteinové databázi Uniprot je v současné době evidováno více než 50 milionů potvrzených a predikovaných sekvenci. Strukturní databáze PDB čítá nad 100 tisíc proteinových struktur reprezentujících přibližně 1300 známých proteinových uspořádání (z angl. folds). Tato čísla ovšem tvoří zanedbatelný zlomek vůči astronomickému počtu kombinaci představujících sekvence z 20 proteinogenních aminokyselin (pro 100 aminokyselinové řetězce je 10130 možných sekvenčních variant), které nedostaly šanci v biologickém kontextu (angl. „never born proteins“ - NBP). Dřívější hypotézy předpokládají, že třídimenzionální uspořádání proteinů je založeno na přírodní selekci vzácných sekvenci s danými vlastnostmi, zatímco sekvence nevystavené evolučnímu tlaku budou pravděpodobně neuspořádané.

Za účelem potvrzení nebo vyvrácení těchto spekulací jsme analyzovali knihovnu náhodných sekvenci. Naším cílem bylo vypracovat postup pro určení frekvence a podoby sekundárních a terciárních struktur. Prvním krokem byla bioinformatická analýza knihovny pomoci prediktoru sekundárních struktur, proteinové neuspořádanosti a ostatních fyzikálně-chemických vlastnosti (rozpustnost, hydrofobicita, agregační potenciál). S použitím těchto nástrojů jsme v knihovně náhodných sekvenci odhalili neočekávaně vysokou frekvenci výskytu sekundárních struktur.

Pro experimentální charakterizaci, s cílem ověření a následující analýzy našich výsledků, jsme připravili 45 proteinů z knihovny náhodných sekvencí. Sekvence byly vybírány tak, aby zapadaly do tří skupin, reprezentujících celé spektrum výskytu sekundárních struktur (vysoký, nízký a náhodný výběr z celé). Jednotlivé rozpustné proteiny byly exprimovány *in vitro*, purifikovány a charakterizovány biofyzikálními metodami (hmotnostní spektrometrie (MS), dynamicky rozptyl světla (DLS), cirkulární dichroismus (CD)). CD spektra potvrzují výskyt sekundárních struktur pro jednotlivé skupiny náhodných sekvencí a souhlasí se závěry bioinformatické analýzy o téměř totožně vysokém výskytu sekundárních struktur v náhodném sekvenčním prostoru v porovnání s jejích biologickými protějšky.

V budoucnu se zaměříme na analýzu různých aminokyselinových repertoáru podle stejného protokolu a na vyhodnocení vlivu aminokyselinového složení na tendenci proteinů pro tvorbu sekundárních struktur. Konkrétně je naším cílem pozorovat vliv evoluce genetického kódu (tedy postupného přidávání aminokyselin k dostupným kodonům) na potenciál proteinu pro tvorbu sekundárních struktur.