

C2184 Úvod do programování v Pythonu

Domácí úkol 3

Stáhněte si ze studijních materiálů `Data/pdb_report_100.py` a nakonec programu napište program, který

A) vypíše seznam PDB ID, které splňují sadu podmínek (**vyberte jednu, 4 body**):

- i) exp. metoda (`Exp. Method`) zjištění struktury je Rentgenová difrakce s rozlišením (`Resolution`) menším a rovným 2.5
- ii) exp. metoda (`Exp. Method`) zjištění struktury je NMR s počtem residuí (`Residue Count`) menším a rovným 150
- iii) sekvence proteinu (`Sequence`) obsahuje tyto po sobě jdoucí aminokyseliny:
`serin-methionin-leucin`

CHALLENGE:¹

- iv) sekvence proteinu (`Sequence`) obsahuje tento vzor: `RE*I` (návod: `cyklus+find`)

B) vypočítá a vypíše následující informace (**vyberte jednu, 4 body**):

- i) průměrný rok vydání (`Rel. Date`)
- ii) průměrné rozlišení (`Resolution`), s prázdnými hodnotami nepočítejte a vypište také počet prázdných hodnot
- iii) průměr a medián počtu residuí (`Residue Count`, návod: pro medián můžete použít funkci `sort`)
- iv) počet proteinů (`Macromol. type`), které mají spojitost s apoptózou (ve sloupečku `Classification` mají klíčové slovo `apoptosis`²)

Poznámka: Úkol do odevzdáárny nahrajte jako jeden soubor `homework03.py`, vaše jméno a UČO se přidá automaticky.

¹kdo ze seminární skupiny první odevzdá správné řešení vyhraje specialní cenu

²na velikosti písmen nezáleží