

## C2110 Operační systém UNIX a základy programování

### Data k úkolům jsou v adresáři:

/home/kulhanek/Documents/C2110/Tests/Final/Data/Scripts/WeekB

### UPOZORNĚNÍ:

Jednu zvolenou úlohu řešte v adresáři C2110-kolokvium, který bude mít plně zakázaný přístup pro skupinu a pro ostatní uživatele!

### Zadání A

V adresáři „PDB“ se nacházejí soubory ve formátu PDB obsahující strukturní informace biomolekul určených různými experimentálními metodami. Proveďte jejich analýzu jejichž výsledkem bude soubor „methods.txt“ obsahující na jednotlivých řádcích PDB ID struktury (jméno souboru PDB bez přípony) a název experimentální metody, která byla použita k určení její struktury. Experimentální metoda je uvedena v souboru vedle klíčového slova EXPDTA (**10 bodů**). Dále ze získaného souboru methods.txt vytvořte souhrnnou statistiku obsahující experimentální metodu a počet struktur, pro jejichž určení byla metoda použita. Dále celkový počet struktur obsažených v adresáři. Vytvořený přehled vypište na standardní výstup (**15 bodů**). Jméno adresáře obsahující struktury zadá uživatel interaktivně po spuštění skriptu. Skript ověří, zdali adresář existuje. Pokud ne, tak vypíše chybové hlášení a ukončí se (**5 bodů**).

Ukázka výstupního souboru s počty molekul získanými jednotlivými metodami:

SOLUTION_NMR	8
ELECTRON_MICROSCOPY	2
X-RAY_DIFFRACTION	10
TOTAL	20

### Zadání B

Pomocí kvantově-chemického výpočtu metodou HF/6-31G(d) byla provedena optimalizace geometrie molekuly. Výsledkem optimalizace je nalezení geometrického uspořádání atomů, které má nejnižší energii. Úkolem je ze zaznamenaného průběhu optimalizace vyextrahovat pro každou geometrii její energii a ověřit, že poslední geometrie má skutečně nejnižší energii. Energie pro každý optimalizační krok je uvedena za klíčovým výrazem: „SCF Done:“. Jednotkou energie je Hartree (**15 bodů**). Průběh energie zobrazte ve formě grafu, který uložíte do souboru s názvem „optimization.png“. Řádně popište osy grafu (**10 bodů**). Název analyzovaného souboru zadá uživatel jako první argument skriptu. Skript ověří, že zadaný soubor skutečně existuje (**5 bodů**).

Testovací soubor s názvem „aceticacid.log“ je dostupný v adresáři „QM“.

### Zadání C

Výsledkem literární rešerše je soubor, který obsahuje informace důležitosti aminokyselin pro funkci membránového receptoru. Soubor ve formátu tsv obsahuje sloupce oddělené tabulátorem. Popis sloupců je uveden na prvním řádku. Typ aminokyseliny je zapsán ve sloupci „residue“ pomocí jednopísmenného kódu. Proveďte statistickou analýzu důležitostí jednotlivých aminokyselin, pro které jsou dostupné anotace. Výsledkem bude soubor aminoacids.txt, který bude v prvním sloupci obsahovat jednopísmenný kód aminokyseliny a ve druhém počet záznamu pro tuto aminokyselinu (**15 bodů**). Na standardní výstup dále vypište aminokyselinu s největším a nejmenším zastoupením (**10 bodů**). Jméno vstupního souboru k analýze uživatel zadá jako první argument skriptu. Skript ověří, zdali soubor existuje. Pokud ne, tak vypíše chybové hlášení a ukončí se (**5 bodů**).

Testovací soubor **2020.tsv** je dostupný v adresáři Annotations.