

Programování v jazyce C pro chemiky (C2160) – závěrečné cvičení

1. Kruhové schéma vzdáleností residuů od centra molekuly

Zadání

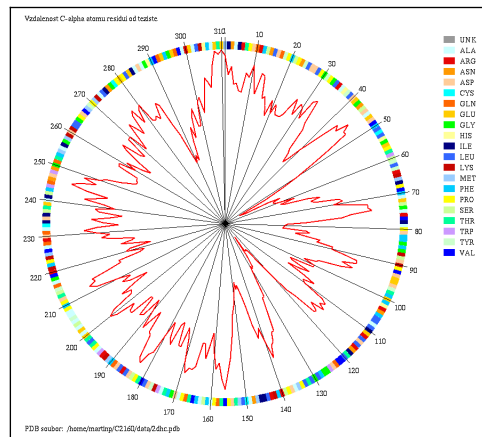
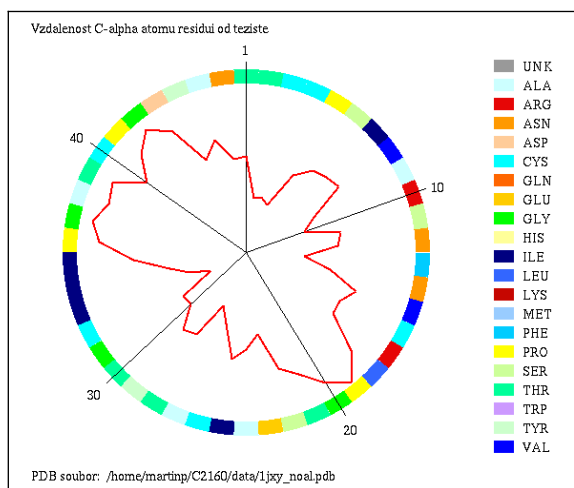
Vytvořte program, který spočítá střed struktury proteinu, vzdálenosti C-alfa uhlíkových atomů residuů od středu a vykreslí graf s barevným značením residuů a křivkou reprezentující vzdálenost C-alfa atomů od centra. Program bude mít následující vlastnosti:

- Bude načítat PDB soubor s proteinem (pouze standardní residua z řádků ATOM)
- Střed struktury bude spočítán jako těžiště struktury, kde však všechny atomy budou považovány za stejně hmotné
- Budou určeny C-alfa atomy aminokyselin a spočítána jejich vzdálenost od centra
- Bude vykresleno kruhové schéma, kde na okraji budou residua vyznačena barevně a také budou vyznačena čísla residuů (tj. pořadí v sekvenci jak je uvedeno v PDB souboru) – viz. obrázek níže
- Uvnitř grafu bude křivka reprezentující vzdálenost residuů od centra molekuly, střed kruhu bude odpovídat pozici v centru molekuly, okraj kruhového schématu bude odpovídat pozici nejvzdálenějšího residua
- Graf bude obsahovat barevnou legendu, nadpis a jméno PDB souboru
- Jméno vstupního PDB souboru bude specifikováno jako parametr na příkazovém řádku
- Program bude uživatele informovat o chybě při otevření souboru, načítání konfiguračního souboru, překročení maximální přípustné velikosti polí a pod.
- Zdrojový kód programu bude opatřen komentáři

Nepovinné rozšíření (+5 bodů):

- Program bude načítat konfigurační soubor, ve kterém bude specifikováno jméno vstupního PDB souboru na řádku ve formátu "INPUT_FILE = jmeno_pdb_souboru", dále bude v konfiguračním souboru na samostatném řádku specifikována velikost okna ve formátu "WINDOW_SIZE = sirka, vyska"
- Název konfiguračního souboru bude předán programu jako parametr na příkazovém řádku

Program otestujte se strukturou crambinu (*1jxy_noal.pdb*) a enzymu haloalkan dehalogenáza (*2dhc.pdb*), které najdete mezi studijními materiály v IS MU ve složce „data“.



Dodržujte následující pravidla

- Dbejte na správné odsazování textu
- Pro reálné proměnné používejte typ `double`, ne `float`
- Při každém použití operátoru dělení si ujasněte zdali dochází k celočíselnému nebo reálnému dělení a jaký typ dělení požadujete
- Proměnné vždy inicializujte vhodnou hodnotou
- Při použití funkcí pro práci s řetězci a při práci s poli dbejte na to, aby nedošlo k překročení velikosti pole
- Dobře zvažte, které proměnné budou lokální a které globální
- Názvy globálních proměnných volte tak, aby z nich byl jasný význam proměnné, volte raději delší názvy
- Názvy funkcí volte tak, aby z nich bylo jasné, jakou činnost funkce vykonává
- Pro překlad programů používejte nástroj `make` (tj. vytvořte si příslušný `Makefile`)
- Z programu odstraňte veškerý kód, který není nutný pro splnění zadání (např. pozůstatky z minulých cvičení, zakomentované části kódu). Ponechat můžete funkci pro zápis PDB
- Program nesmí při překladu vypisovat žádné varovné hlášky (při použití parametrů `-Wall -pedantic`)
- Na začátek programu umístěte stručný komentář obsahující jméno autora, rok vytvoření, popis funkce programu, parametry příkazového řádku, popř. formát konfiguračního souboru popisující činnost programu
- Všechny funkce a proměnné opatřete komentářem

Nápověda

1. Upravte funkci pro načítání PDB souboru tak, že bude načítat pouze řádky ATOM a nikoliv HETATM.
2. Střed struktury proteinu spočítejte tak, že sečtete souřadnice všech atomů (zvláště pro x, y, a z) a vydělíte je počtem atomů.
3. Ve struktuře proteinu vyhledejte pro každé residuum atom C-alfa, tj. atom se jménem " CA " (vč. mezery na začátku a na konci).
4. Pro pohodlnější práci s C-alfa atomy, přidejte do struktury RESIDUE proměnnou (např. `atom_c_alpha`) která bude obsahovat index atomu C-alfa v poli atomů (tj. pořadí v poli atomů). Hodnotu proměnné nastavte pro každé residuum ve funkci pro vyhledávání residuí nebo v samostatné funkci.
5. Pro pohodlnější práci se souřadnicemi C-alfa atomů přidejte do struktury RESIDUE proměnnou která bude obsahovat souřadnice atomu C-alfa.
6. Vytvořte funkci (např. `get_points_distance()`), které předáte souřadnice dvou bodů a ona vrátí vzdálenost mezi těmito body. Tuto funkci využijete při výpočtu vzdálenosti mezi centrem a C-alfa atomy.
7. Na začátku kreslící funkce spočítejte maximální hodnotu vzdálenosti C-alfa atomu od centra. Tuto hodnotu využijete pro výpočet pozice bodů křivky uvnitř grafu (spočítaná maximální hodnota bude odpovídat okraji grafu).
8. Kreslete kruh s barevnými kódy residuí podobně jako v úloze 1. ze cvičení 11. Je vhodné začít v horní části kruhu (tj. úhel 90° resp. $\pi/2$) a pokračovat po směru hodinových ručiček.

Testovací data

Souřadnice centra pro strukturu crambinu (*1jxy_noal.pdb*):
x = 9.338197, y = 9.687547, z = 6.947231 Å

Vzdálenosti C-alfa atomů od středu pro prvních 10 residuí pro strukturu crambinu (*crambin_noal.pdb*):

THR1:	8.590359
THR2:	4.900823
CYS3:	5.161508
CYS4:	4.661459
PRO5:	8.042055
SER6:	9.413193
ILE7:	9.862124
VAL8:	10.162746
ALA9:	6.802787
ARG10:	5.521469