

5. Graf prostorové vzdálenosti aminokyselin v proteinu

Zadání

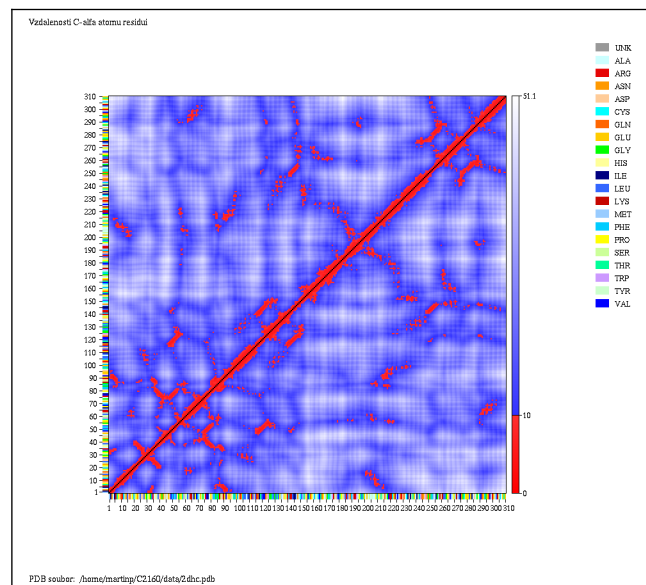
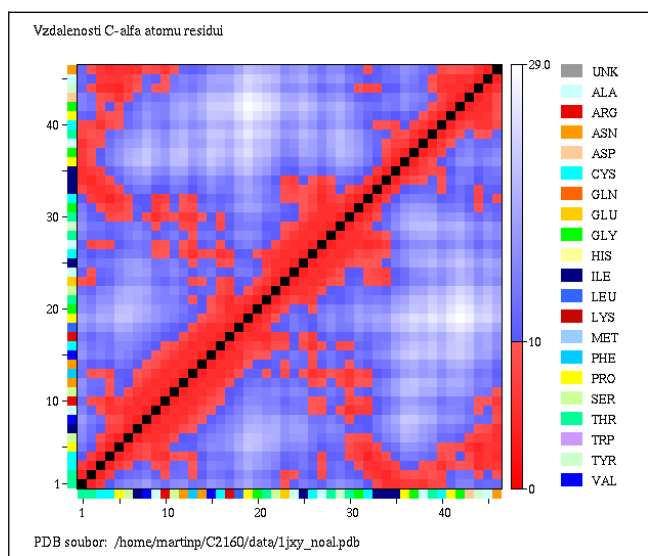
Vytvořte program, který zobrazí XY graf prostorových vzdáleností aminokyselin s barevným označením blízkých aminokyselin. Program bude mít následující vlastnosti:

- Bude načítat PDB soubor s proteinem (pouze standardní residua z řádků ATOM)
- Budou určeny C-alfa atomy aminokyselin
- Vzdálenost residuů bude aproximována jako vzdálenost jejich C-alfa atomů
- Bude zobrazovat graf kde na ose x a y budou residua ve stejném pořadí jako v sekvenci, uvnitř grafu budou jednotlivé body označeny různými odstíny modré, pro nejvzdálenější residua bude barva bílá pro nejbližší modrá, navíc atomy, jejichž vzdálenost bude menší než 10 Å, budou označeny odstínem červené
- Osy x a y v grafu budou obsahovat barevný kód residua a čísla pořadí v sekvenci způsobem uvedeným na obrázku
- Graf bude obsahovat barevnou legendu pro aminokyseliny a pro barevné odstíny použité v grafu, dále 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. Ve struktuře proteinu vyhledejte pro každé residuum atom C-alfa, tj. atom se jménem " CA " (vč. mezery na začátku a na konci).
3. 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.
4. Pro pohodlnější práci se souřadnicemi C-alfa atomů přidejte do struktury RESIDUE proměnnou která bude obsahovat souřadnice atomu C-alfa.
5. 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 C-alfa atomy residuí.
6. Na začátku kreslící funkce spočítejte maximální hodnotu vzdálenosti C-alfa atomů. Tuto hodnotu využijete pro výpočet odstínu barev.
7. Vykreslování grafu může být pomalé, zejména u větších struktur. Knihovna g2 totiž po zavolání funkce pro kreslení čeká, až dojde k vykreslení na obrazovku, což proces kreslení zpomaluje. Pro urychlení zavoláme ihned po otevření okna funkci `g2_set_auto_flush()`, kde první parametr specifikuje číslo zařízení (tj. číslo vracené funkcí pro otevření okna) a jako druhý parametr použijeme 0. Po dokončení kreslících operací potom zavoláme `g2_flush()` (přijímá číslo zařízení jako jediný parametr). Více na:
http://ncbr.chemi.muni.cz/~martinp/g2/group_device.html
http://ncbr.chemi.muni.cz/~martinp/g2/group_control.html

Testovací data

Vzdálenosti C-alfa atomů vybraných dvojic residuí pro strukturu crambinu (*1jxy_noal.pdb*):

| | | | | |
|------|---|------|---|-----------|
| THR1 | - | THR2 | : | 3.757827 |
| THR1 | - | CYS3 | : | 6.668096 |
| THR1 | - | CYS4 | : | 10.153580 |
| THR1 | - | PR05 | : | 13.523983 |
| THR1 | - | SER6 | : | 15.584773 |
| THR2 | - | CYS3 | : | 3.823774 |
| THR2 | - | CYS4 | : | 6.711954 |
| THR2 | - | PR05 | : | 10.285453 |
| THR2 | - | SER6 | : | 12.237011 |
| CYS3 | - | CYS4 | : | 3.769718 |
| CYS3 | - | PR05 | : | 6.888749 |
| CYS3 | - | SER6 | : | 9.133217 |
| CYS4 | - | PR05 | : | 3.811543 |
| CYS4 | - | SER6 | : | 5.683572 |
| PR05 | - | SER6 | : | 3.796576 |