

## 6. Graf vzdálenosti residuů od centra molekuly s vyznačenou sekundární strukturou proteinu

### Zadání

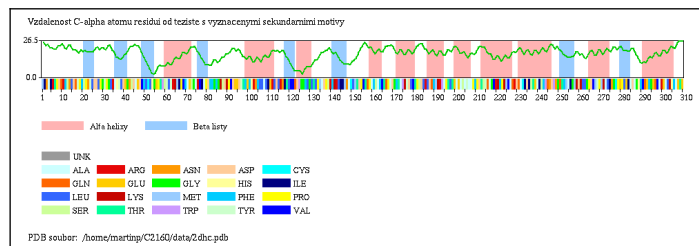
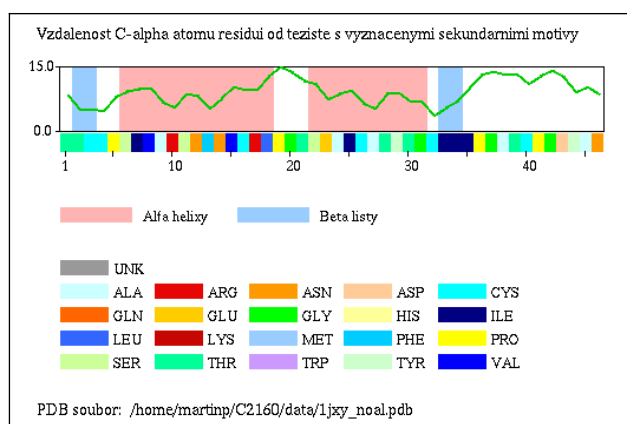
Vytvořte program, který spočítá vzdálenosti C-alfa uhlíkových atomů residuů od centra molekuly a navíc načte z PDB souboru informace o sekundární struktuře (alfa-šroubovicích a beta-listech). Potom zobrazí graf reprezentující vzdálenost C-alfa atomů od centra s vyznačením alfa-helixů a beta-listů. Program bude mít následující vlastnosti:

- Bude načítat PDB soubor s proteinem (pouze standardní residua z řádků ATOM)
- Z PDB souboru bude také načítat řádky HELIX a SHEET, na základě toho přiřadí jednotlivým residuům příslušnost k těmto sekundárním motivům
- 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 zobrazovat graf obsahující křivku charakterizující vzdálenost residuů od centra. Navíc budou obarvena některá políčka residuů, pro helixy světle červenou a pro beta-listy světle modrou barvou. Pod grafem bude barevně označen typ residua a uvedeno pořadí v sekvenci způsobem zobrazeným na obrázku
- 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 bodů křivky grafu.
8. Z PDB souboru z řádků HELIX a SHEET, které obsahují informace o sekundární struktuře proteinu ( $\alpha$ -helixech a  $\beta$ -listech).

Pro tyto účely definujeme strukturu SECONDARY\_MOTIF, která bude obsahovat 3 celočíselné proměnné: číslo prvního residua motivu (např. `first_residue`), číslo posledního residua motivu (např. `last_residue`) a číslo reprezentující typ motivu (např. `motif`, bude nabývat hodnoty 1 pro  $\alpha$ -helix a 2 pro  $\beta$ -list). Definujte pole těchto struktur (např. `motifs[]`) a globální proměnnou, která bude obsahovat aktuální počet sekundárních motivů (např. `motifs_count`).

Načítejte sekundární motivy ve funkci pro načítání PDB souboru. Kromě řádků začínajících ATOM (popř. HETATM) načítejte také řádky začínající HELIX nebo SHEET, z nich načítejte pouze čísla residuí, ostatní položky ignorujte. Číslo prvního residua se nachází na pozici 22-25 pro HELIX (číslováno od 1) a 23-26 pro SHEET, číslo posledního residua se nachází na 34-37 pro HELIX i SHEET (<https://www.wwpdb.org/documentation/file-format-content/format33/sect5.html>). Motivы ihned ukládejte do pole `motifs[]`.

Do struktury RESIDUES přidejte celočíselnou proměnnou (např. `motif`), která bude obsahovat typ motivu, jehož je residuum součástí. Nastavte hodnotu této proměnné. Nezapomeňte, že číslo residua v motivu není indexem v poli `residues[]`, ale číslo residua uvedené v PDB souboru (tj. hodnota `residues[].residue_number`). Hodnotu v proměnné `motif` použijte k obarvení příslušné oblasti v grafu.

## 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 (*1jxy\_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

Sekundární motivy ve struktuře crambinu (*1jxy\_noal.pdb*) jsou:

první  $\alpha$ -helix: SER6 – LEU18 (13 residuí)

druhý  $\alpha$ -helix: SER22 – GLY31 (10 residuí)

první  $\beta$ -list: THR2 – CYS3 (2 residua)

druhý  $\beta$ -list: ILE33 – ILE34 (2 residua)