COMP364: PDB & Biopython

Jérôme Waldispühl, McGill University

Working with structure objects

- Choose a model (E.g.: `first_model=structure[0]`).
- Choose a chain (E.g.: `chain_A=model["A"]`).
- Choose a residue (E.g.: `res10=chain[10]`).
- Choose a atom (E.g.: `atom=res10[”CA"]`).
- Retrieve Atom attributes:

```
a.get_name()         # atom name (spaces stripped, e.g. "CA")
a.get_id()           # id (equals atom name)
a.get_coord()        # atomic coordinates
a.get_bfactor()      # B factor
a.get_occupancy()    # occupancy
a.get_altloc()       # alternative location specifie
a.get_sigatm()       # std. dev. of atomic parameters
a.get_siguij()       # std. dev. of anisotropic B factor
a.get_anisou()       # anisotropic B factor
a.get_fullname()     # atom name (with spaces, e.g. ".CA.”)
```

Example

```
from Bio.PDB.PDBParser import PDBParser

parser=PDBParser()

# parse PDB file and store it in structure object
structure=parser.get_structure("test", "1fat.pdb")

# print the coordinate of CA atoms with B factor > 50
for model in structure.get_list():
  for chain in model.get_list():
    for residue in chain.get_list():
      if residue.has_id("CA"):
        ca=residue["CA"]
        if ca.get_bfactor()>50.0:
          print ca.get_coord()
```

Problem 1

Go to the PDB (http://www.rcsb.org) and download the structural data for the myoglobin.

Using the code of the previous example:

- parse the file and create a structure object storing the data,
- Print the list of all residues (index and amino acid type) in the structure and display the 3D coordinates of their C_α (N.B.: You can retrieve the index of a residue with the following command: `hetflag, resseq, icode=residue.get_id()`).
- Make an bar chart where the x-axis represent the residue index and the y-axis plot the B-value.
- Modify this bar chart such that the bar corresponding to residues in alpha helix are red and the others green.

Problem 2

Continue with the previous structure data. We are now looking at spatial contacts.

- Compute the average distance between all pairs of residues in your protein.
- Calculate the list of residue pairs that are distant by at most 8 Å (You will calculate the distance between the C_α)