

# COMP 364: Mid-term Practice

*Submit all solutions in the form of python script files that one can run. Zip your solutions up into a compressed file called `<your name>_practice.zip` and email it to `jerome.waldispuhl@mcgill.ca` (indicate “COMP364 practice” in the subject).*

## 1 Python (3 points)

Write a script that takes a series of integers on the input, store them into a list and print the sum of these numbers.

## 2 Find the bug (2 points)

Download the python file at : <http://www.cs.mcgill.ca/~jeromew/comp364/data/R1Q2.txt> (This is a python file. Once downloaded, you can change locally the extension of this file to `.py`). Run it. Identify the bug in the program file, indicate and explain it with a comment in the same file and fix it.

## 3 Sequence analysis (5 points)

Download the FASTA file at : <http://www.cs.mcgill.ca/~jeromew/comp364/data/R1Q3.fasta>.

- Parse this file using Biopython and print the number of sequences found inside this file. (1 point)
- Scan all sequences in this file and create a dictionary named `ntcount` in which you will store the number of occurrences of each nucleotide in the data set. Print the content of this dictionary and the type of the most common and uncommon nucleotide. (2 points)
- Print a bar chart named `R1Q2_barchart.pdf` showing the distribution of the nucleotides (each bar must be labelled with the corresponding nucleotide). (2 points)

## 4 Image manipulation (5 pts)

Download the PNG file at : <http://www.cs.mcgill.ca/~jeromew/comp364/data/R1Q4.png>.

- Import the module `matplotlib.image` and use the function `imread()` to read the image and create a RGB array storing the pixel values. What is the size (in pixels) of this image? (2 points)

- Display this image using `imshow()` and save it in a file named `R1Q4_Q2.png` (1 point)
- Duplicate the image matrix but set to 0 the blue component of the RGB pixels. Save the results in a file named `R1Q4_Q3.png` (1 point).
- Duplicate the image matrix and color in red all pixels above the value (0.8,0.8,0.8). Save the results in a file named `R1Q4_Q4.png` (1 point).

## 5 Analyzing protein structures (5 pts)

Download the PDB file at : <http://www.cs.mcgill.ca/~jeromew/comp364/data/R1Q5.pdb>.

- Parse the structure using the PDBparser implemented in Biopython. What is the length of this protein? Print this number (1 point).
- Calculate the distance between the  $C_\alpha$ 's of all residues separated by 4 positions in the chain (i.e. residue contact  $(i, i + 4)$ ). Print all contacts below 8 Å (2 points)
- In a file named `R1Q5_dist.pdf`, Print these distances using the function `plot` of `matplotlib`. The x-axis will represent the index of the first residue index and the y-axis will plot the distance between  $i$  and  $i + 4$ . (2 points)