

Biological Data

Prof. Dr. Silvio Tosatto, Dr. Damiano Piovesan

Group project, ver. 1.2

1. Purpose

The analysis of a protein family is both the most creative part and closest to reality among the practical exercises. For this reason, we deliberately do not provide a standard *"recipe"* for the project, as each case can be different. In all cases, follow the course program and websites seen during the practicals to analyze the protein. This is however only the starting point and you will be evaluated in particular with regard to critical thinking and adapting the analysis to the peculiarities of the case. The result of the analysis should be a brief report describing and linking the various analyses made, combining critical information and providing a coherent picture of the protein family. In particular, those interested should consider to use additional online tools not described in class suitable to characterize the protein family. Any additional information will be rewarded a bonus in the evaluation.

2. How to Proceed

For the analysis, it is advisable to proceed along the lines of what was done in the practicals, passing from simple to sophisticated information. In general, analysis of sequence, phylogeny, structure, function and interactions are recommended. Depending on the specific case, they should to be integrated with specific methods. To find the web sites you can use the links from the practicals page or just search with Google. Data should be used to build the final report to be accompanied by appropriate documentation (see next point).

3. Documentation

The supporting documentation of the report consists of source code, data and outputs methods used to analyze the protein family. Screenshots and other figures should be imported into Word, grouped after the main text and treated as figures for the report. Every figure must contain a caption with a serial number and a concise sentence saying what can be seen (e.g. *"Figure 1. The UniProt database entry for protein XX"*). The serial number is to be quoted in the main report text to support the assertions made, e.g. *"GO function is XX (Figure 2)"*. **Only include figures to support the main considerations.**

Source code, data and outputs should be indicated as Supplementary Information, numbered progressively with a concise sentence saying what is included (e.g. *"Supplementary Information 1 – Python source code"*). Please include a section at the end of the report listing all supplementary information items with their brief description (concise sentence). The files themselves should also be named accordingly and collected in a single ZIP archive.

4. Submission

Reports and supporting files are to be submitted via e-mail to biocomp@bio.unipd.it. Please include your names in the submitted document. **The report submission deadline is ten days before the exam date.** The report must be in Word format (DOC, DOCX) or PDF. Please do not submit other formats such as OpenOffice, etc.

The report, should be written with a font size of at least 12 points, margins of at least 2 cm on each side and line spacing 1.5. The report should be **at least two (2) and not more than five (5) pages of text**, excluding figures and supporting documentation. (**NB:** Reports that do not adhere to these criteria may be subject to penalties)