

Softwarewerkzeuge der Bioinformatik

Project Report

The reports presents the results and thus constitutes the basis of the assessment. It should be written in a way that the approach, results and discussion is easy to understand without the reader having to consult additional sources. All reports have to adhere to the following guidelines.

1 General Remarks

Language: The project report can be written in **German** or **English**.

Text: The report should be written in full sentences with proper paragraphs and follow the spelling and grammar rules of the respective language.

Format: There are **no** specific rules regarding font sizes, line spacing, etc. as long as it is readable. However, make sure that the report is clear and well-structured.

Division of Tasks In a group you can of course assign the tasks to different group members. However, coordinate with each other, so that you use the same approach for similar tasks and make sure that the final report is consistently formatted. That also means that if a certain information is relevant for several tasks, that this information should **not** be procured and presented in three different ways, and certainly not with three different results.

Length: The report should contain the required information, no more, no less. The length of the reports can differ between projects. There are no bonus points for long reports, and unnecessarily long reports mean more work for everyone involved. It is therefore better to be concise and precise. It is possible to fully answer each task in 0.25 to 2 pages of pure text (excluding tables and figures).

Figures and Tables: Figures and results should be included in the text and not just linked. Figures and tables must be of a sufficient quality and be readable. All used figures and tables must be numbered, have a caption, and be referenced and sufficiently explained in the text.

Keep in mind that while tables and figures are important to give the reader a good overview of the results, they are **not** a replacement for a textual summary of the results, or a comparison or discussion.

Screenshots: Screenshots should be treated like figures and you should only present the relevant parts. That means, you should **not** include a screenshot of an entire website. Before you include a screenshot, ask yourself, if it is really needed. A report quickly becomes confusing and hard to read when several screenshots are included for a single task. For example, instead of including a screenshot of the PRINTS or UniProt website, summarise the relevant information in a table or in 1-2 sentences.

Software Tools and Databases: If possible and not otherwise stated, you should use the software tools and databases that were introduced in the tutorial, so have a look at the relevant tutorials.

Plagiarism: Solutions and code should be written by you in your own words. Parts that were copied from other groups, the internet or submissions of previous years will be marked with 0 points.

Questions: You can ask questions during the tutorial or contact the responsible tutor via **email**. The email address is given in each project.

Submission: The reports have to be submitted as a single **PDF** file via email to the specified tutor **before** the deadline given in the project. This should happen as an actual attachment, not as a link. Each report should contain the name of the project, the group number, as well as the names, matriculation numbers and email addresses of the group members.

2 Structure

Document and justify your approach in your approach and specify which software tools (including chosen parameters), data bases (including accession numbers) or other sources were used. Additionally present and discuss your results as specified in each task.

For each task the following structure should be followed:

2.1 Introduction

Describe the goal of the task in 1-2 sentences. For example:

- In this task, we examine...
- The goal of this task was...

2.2 Approach

Describe your approach thoroughly. State which software tools, databases and websites you have used, and with which settings and parameters, so that your approach is reproducible. Additionally, you should always mention the accession numbers of proteins, motifs, structures etc. For example:

- To solve this task, we applied tool X to protein (YYYYYY) with default settings.
- As a threshold we chose an *E-value* of 0.0001.

2.3 Presentation of the Results

Results have to be presented in a **well-structured** and **comprehensible** manner. All results and claims have to be **reproducible** with the stated approach and must be summarized in figures and tables when appropriate. For example, the referenced sequence positions have to be clearly stated and marked in figures (e.g. motifs in sequence alignments). Additionally, key information must be addressed in the text, and it has to be clear **what the text is referencing**. For example, it should not be left as an exercise for the reader to count the transmembrane helices in a figure. If necessary, a sensible selection should be made and described (e.g. select and discuss 3–5 proteins of 200).

2.4 Discussion and Comparisons

Some tasks ask for a discussion or a comparison of the results. This should only reference results that were presented in the report. Results that were not presented cannot be discussed or used in comparisons.

3 Example

3.1 Task

Compare the PRINTS- **and** PROSITE-motifs of your protein.

3.2 Proposed Solution

In this task we examined the motifs of the given protein (P123456). Subsequently, we compared the identified motifs. We used PRINTS and PROSITE to identify the motifs by following the corresponding links on the UniProtKB site of our protein.

The search in PRINTS and PROSITE returned several motifs for the given protein. PRINTS contained information about two motifs with the accession XXXX, whereas PROSITE yielded only one motif with the accession YYYYY. Table 1 presents the identified PRINTS- and PROSITE-motifs.

	PRINTS	Pos.	PROSITE	Pos.
Accession	XXXX		YYYY	
Motif 1	GGLIGSIYGV	6-15	L-x(2)-S-I-x-[FYGWIV]-{K}	8-15
Motif 2	VTFCIYPLE	21-29		

Table 1: Identified PRINTS- and PROSITE-motifs for our protein (P123456) with their positions in the protein sequence.

PRINTS-motif 1 consists of ten amino acids, whereas the PROSITE-motif is two amino acids shorter in comparison. Additionally, PRINTS-motif 2 is nine amino acids long.

Comparing PRINTS-motif 1 with the PROSITE-motif reveals a match between positions 3-10 of PRINTS-motif 1 (GG**LIGSIYGV**) and the entire PROSITE-motif (**L-x(2)-S-I-x-[FYGWIV]-{K}**). PRINTS-motif 2 on the other hand differs from the PROSITE-motif. This is also reflected in their positions in the protein sequence (Table 1), which we highlighted in the multiple sequence alignment from Task 1 (Figure 1).

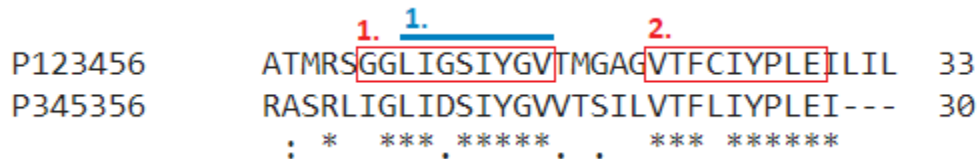


Figure 1: Multiple sequence alignment of our protein (P123456). Marked are complete (asterisk), highly (colon) and slightly (fullstop) conserved areas. The two PRINTS motifs of our protein are marked in red and the PROSITE motif in blue. [In the actual project report you should of course use more sequences for a multiple sequence alignment.]