
Table of Contents

Introduction	1.1
1. Lectures	1.2
3.1 Introduction: Bioinformatics	1.2.1
3.2 Introduction: Structure	1.2.2
3.3 Alignments 1	1.2.3
3.4 Alignments 2	1.2.4
3.5 Comparative Modeling	1.2.5
3.6 Secondary Structure Prediction	1.2.6
3.7 Secondary Structure Prediction 2	1.2.7
3.8 Secondary Structure Prediction 3	1.2.8
3.9 Membrane Structure Prediction	1.2.9
3.10 TMSEG	1.2.10
3.11 Beta Membrane and Accessibility	1.2.11
2. Exercises	1.3
2.1 Introduction	1.3.1
2.2 Biological Background	1.3.2
2.3 Protein Structures	1.3.3
2.4 Alignments	1.3.4
2.5 Resources for Bioinformatics	1.3.5
2.6 Secondary Structure Prediction	1.3.6
2.7 Homology Modeling	1.3.7
2.8 Wrap Up	1.3.8
3. Exam Questions	1.4
3.1 Lecture Questions	1.4.1
3.2 Exercise Questions	1.4.2
3.3 Old Questions	1.4.3

Protein Prediction I

Course Summary, Summer Term 2017

tl;dr: This purpose of this document to collaboratively create a both concise and detailed course summary of the *Protein Prediction I* Lecture from 2017 Summer Term at TUM.

To learn as effective as possible, I would like to encourage everyone to engage in the discussion evolving around the content of this document. If you have questions or challenge what someone else wrote please do so in a **constructive way**. We are all new to the subject of Protein Prediction and mistakes happen. Let's learn from them together!

Official Lecture Resources

Lecture Homepage: <https://www.rostlab.org/teaching/ss17/pp1cs>

Lecture Wiki: https://i12r-studfilesrv.informatik.tu-muenchen.de/sose17/pp4cs1/index.php/Main_Page

Youtube Channel:

<https://www.youtube.com/channel/UCU6j8BG4RbEtTgylZJ6Vpow>

Getting Started

This document is set up a **Gitbook** and hosted on **Github**. When you read this, you were already granted access to the repository so the first step is done.

The easist way to start contributing is to download **Gitbook Editor** (available for Mac, Linux, Windows) from [here](#).

Before you add / change anything, please read through the Contribution Guide.

Contribution Guide

Tell others what you work on | Write meaningful commit messages | Push often | Use American English

Why is there a contribution guide? I think it is in everyone's best interest to keep this summary as easy to understand as possible for everyone. This guideline should help to maintain consistency across the entire document.

Each section may contain a short additional information on how to format things specific to that section. Please have a look there as well.

1. Adding new content

1.1 Adding minor updates

If you add minor updates, like the answer to a single question, you can do this on the `develop` branch directly. Make sure your commit has a meaningful message.

1.2 Adding major updates

If you add major updates, like several related changes (e.g. an entire lecture summary), go along as follows:

1. Add a new **issue** on Github, describing what you are working on
2. Create a `feature/<issue-name>` branch and add your changes
3. Open a pull-request to merge back into `develop` and add the other contributors as reviewers
4. Once the pull request is merged, delete your feature branch and close the issue by referencing the merge commit

Why so complicated? This way the issues reflect new changes and are transparent for all contributors.

2. Challenging existing content

If you find obvious mistakes (typos, clearly wrong statements) just change them directly.

If you are challenging statements, answers to questions etc. which might not be trivial to understand go along as follows:

1. Open a new **issue** on github.

2. Reference the the statement in question you consider to be wrong
3. Provide an explanation why you think it is wrong
4. Provide your correct solution.

3. Adding new contributors

The purpose of this document is to foster collaborative learning - hence to make this as inclusive as possible. This being said, too many collaborators would probably lead to chaos . If you know other students personally, you want to add to the project shoot me a message and we will figure it out.

1. Lectures

3.1 Introduction: Bioinformatics

02.05.2017 | [Slides](#) | [Lecture Recording](#)

3.2 Introduction: Structure

04.05.2017 | [Slides](#) | [Lecture Recording](#)

3.3 Alignments 1

11.05.2017 | [Slides](#) | [Lecture Recording](#)

3.4 Alignments 2

16.05.2017 | [Slides](#) | [Lecture Recording](#)

3.3 Comparative Modelling

18.05.2017 | [Slides](#) | [Lecture Recording](#)

3.6 Secondary Structure Prediction

?? .05.2017 | ??? | ???

3.7 Secondary Structure Prediction 2

01.06.2017 | [Slides](#) | [Lecture Recording](#)

3.8 Secondary Structure Prediction 3

08.06.2017 | [Slides](#) | [Lecture Recording](#)

3.9 Membrane Structure Prediction 2

13.06.2017 | [Slides](#) | [Lecture Recording](#)

3.10 TMSEG

20.06.2017 | [Slides](#) | [Lecture Recording](#)

3.11 Beta Membrane and Accessibility

22.06.2017 | [Slides](#) | Lecture Recording

2. Exercises

2.1 Introduction

11.05.2017 | [Slides](#) | [Wiki](#)

2.2 Biological Background

11.05.2017 | [Slides](#) | [Wiki](#)

2.3 Protein Structure

18.05.2017 | [Slides](#) | [Wiki](#)

2.4 Alignments

01.06.2017 | [Slides](#) | [Wiki](#)

2.5 Resources for Bioinformatics

08.06.2017 | [Slides](#) | [Wiki](#)

2.6 Secondary Structure Prediction

22.06.2017 | [Slides](#) | [Wiki](#)

2.7 Homology Modeling

29.05.2017 | [Slides](#) | [Wiki](#)

2.8 Wrap Up

06.07.2017 | Slides | Wiki

3. Exam Questions

This section contains possible exam questions compiled from different sources.

Contribution Guide: Exam Questions

Since all questions here are answered by students, there might be some mistakes in them. Hence a few more words on how to best handle this section.

1. Adding a new question

Just add the question in the respective file. Optimally, you can already provide an answer.

2. Answering a question

To clearly distinguish questions from answers, please put answers in **blockquotes** right under the respective question.

Example:

- How can 1D secondary structure information be used to derive a 3D model?
- It is not possible to derive a 3D model from 1D information. (Trick Question)

3. Updating an answer

If you think an answer does not properly answer a question (e.g. it is wrong or the answer is not sufficient), mark the answer and open a new **issue** on Github to discuss the question and share your improved answer.

(Use the 🚩 emoji to mark the possibly wrong answer inline)

Example:

- How can 1D secondary structure information be used to derive a 3D model?

- It is not possible to derive a 3D model from 1D information. (Trick Question)

3.1 Lecture Questions

This section contains possible exam questions asked Professor Rost in the lectures he dedicated to answering student questions. They are **highly relevant**, because he will sample exam questions from this pool.

3.2 Exercise Questions

This section contains possible exam questions asked and answered as part of the exercises.

3.3 Old Questions

This section contains possible exam questions sourced from students of previous Protein Prediction I lectures.

3.3.1 Exam Structure: 2016ST

We were able to obtain last year's exam structure. Let's try to answer all of the concrete questions :-)

Part 1 is mandatory, for the rest choose 3 out of 4.

1. Multiple Choice (5 questions, 10 points)

- Secondary Structure
- RMSD - Protein Similarity
- Hydrogen Bonds (α -helix, β -sheets, long/short bonds)
- 100% sequence identity => same structure? (PIDE)
- Can modern prediction methods correctly predict structure in the midnight zone?
- About "Cryo-Microscope"
- About "X-Rays"

2. Sequence Alignment (10 points)

- Explain each of the following alignment techniques and provide one method for each
 - Sequence - Sequence
 - Sequence - Profile
 - Profile - Profile
- General scoring BLOSUM62 matrix vs. PSSM
- Why is the sequence information valuable?
- How BLAST speed up pairwise alignments?
- Global vs Local alignment

3. Sequence Structure (10 points)

- What data is needed to predict the structure with ML?
- Which tools and db you will use?
- How to prepare data for ML
- Which 2-3 features will help to predict?
- Would you apply method to all protein (query)?
- Which measure would you use to evaluate your method?

4. Protein Structure (10 points)

- Why it is important to know 3D structure?
- Why is it so hard to compare 3D structure?
- Most successful ML algorithm for predicting structure, steps
- Method for experimental structure determination. Short explanation. How many structures are experimentally known?

5. Machine Learning (10 points)

- General definition of Machine Learning
- Cross validation
- What is 'feature'?
- ETP explain, example
- Name and describe one ML method
- Name and describe "sequence" in context of PP
- Discuss how to predict Protein Structure from amino-acid sequence using ML
- Q2: which is better, how to prove your's is better, which value you will publish?
(What is Q2)