**Project in Molecular Life Science (KB8024 / KB8025)**

*Prof. Arne Elofsson*

**Project plan**

*By Ramakant Mishra*

|  |  |
| --- | --- |
| Introduction | Protein Secondary structure prediction has come a long way since the first method proposed by Chothla and Levitt in the year 1976. The prediction accuracy has improved from less than 60% to more than 85% over these years. This has been made possible recently due to implementation of machine learning algorithms such as SVM and Deep Learning methods in developing state of art predictors. |
| Goal | The goal of this project is to develop one such functional predictor using SVM and other suitable algorithms to predict secondary structure of a protein, and in the process, learn about latest developments on application of machine leaning to the field of bioinformatics. |
| Main methods to be used | SVM, Random Forest, Neural Network (optional) |
| Workplan | The following workplan will be used:   1. Extracting the feature from the dataset 2. Creating cross-validated data 3. Training with different algorithms starting with SVM 4. Experimenting with different window sizes 5. Adding evolutionary information 6. Optimising the parameters 7. Comparing performances of different models 8. Applying on test data 9. Comparing results with published data 10. Review the Predictor 11. Making a presentation 12. Writing a report |