Technische Universität München
Lehrstuhl XII: Bioinformatik

Prof. Burkhard Rost
Christian Schäfer
Tobias Hamp

Exercises to the Lecture ``Protein Prediction´´
Summer Term 2010

Sheet VI

General information

- This sheet is due on **Friday June 25, 9 am** along with the programming part from sheet V. Please send an email (one per group) to `hampt@rostlab.org` **and** `schaefer@rostlab.org` including the paths to your results (answers, scripts, datasets). Scripts should be executable for us so that we can reproduce your results!
- Grading: There will be either 60 or 120 reachable points in the exam (depending on the yet to be determined amount of questions). In order to apply the weighting of 40% for the exercises, all of your exercise points will be devided by either 5.75 or 2.875 and rounded up to the next best half point. This number will then be added to the points achieved in the exam, so that there will be exactly 100 or 200 overall points which determine the final grade.

Exercise 14: Threading (20 points)

Referring to the lecture and paper found at `/mnt/home/rost/schaefer/exercise6/threading.pdf`:

a) What is meant by the *inverse folding* problem? Explain the *tertiary template* approach in this context. What are the drawbacks of that method?
b) Explain in your own words the advantages of the method proposed in the paper over the *tertiary template* method.
c) Give a short but precise wrap-up of how the 3D compatibility search works. Specifically address the issues of 3D structure profiles generation, environment classes and 3D-1D scores. What is the question that the method tries to answer?
d) Why is the graph in figure 4 triangular and not rectangular?
e) Discuss figure 6. How was the graph obtained? What does the Z-score express and why doesn't sperm whale myoglobin show up as the highest ranking sequence although its structure served in the generation of the structure profile?
f) Finally, try to relate the method described here to *homology modeling* with respect to the hssp curve.

Exercise 15: Contacts, Neural Networks and Sec. Struct. Prediction (8 points)

Referring to the slides and videos of lecture 12:

a) Define: 1D, 2D and 3D structure prediction
b) What is meant by local and long range contacts? Give an example for both types in the context of secondary structure.
c) Given a contact map and a distance cutoff, how can you distinguish between local and long-range contacts?
d) Come up with an own definition of a contact which tries to improve the definition of 8A from the lecture.
e) Outline the two major approaches to contact prediction
f) Considering the Neural Network of slide 33 in lecture 12:
   Given input points (1,0) and (1,1) with class values 0.9 and 0.0, respectively:
   What values would J11 and J12 adopt after successfully training the Neural Network? (Assume J11 and J12 are discrete variables)
   What would be the prediction for input points (0,1) and (0.0)?
g) Discuss the following statement: "Strands can not be predicted with an accuracy significantly higher than random."