Exercise 'Protein Prediction I'
Summer Term 2011
Sheet 3

General information

- Our course homepage, containing lecture slides and exercise sheets: http://www.rostlab.org/cms/teaching/teaching-overview/pp0/
- Time and place: Thursday, 11:30 – 12:15, room MI 01.08.021 (Computer room)
  - Groups of 2-3 people
  - Grading („Schein“):
    - 40% exercise (theory/programming), 60% final exam (lecture/exercise content)
      - minimum of exercise points required for exam attendance
- Contact: hampt@rostlab.org, schaefer@rostlab.org, vicedo@rostlab.org
- Send an email (one per group) to the three of us including the paths only (no files as attachments!) to your results (answers, program code, figures,...) until May 26, 10:00 am. Scripts should be executable for us so that we can reproduce your results. Everything has to be readable by us, so please check the permissions of your directories/files.

Exercise 1: Pfam general Information (10 Points):

Referring to the paper


and the pfam website:

http://pfam.sanger.ac.uk/
please answer the following questions:

a) what are Hidden Markov Model profiles and protein domain families? (2)

b) describe shortly: what is Pfam? how does it work? which sources does it use? which different methods? what are its advantages or inconveniences? (4)

c) how does Pfam define domains? (2)

c) which scores are used to evaluate the alignments in Pfam? Explain why they are using them (2)

Exercise 2: Analyse profiles and families predictions with Pfam (20 Points):

Let's look at what PFAM has to say about the proteins O59398 and P5CS_HUMAN:

a) How many trusted profiles match? What is the name of this profile and its accession number? What is the E-Value of the match? (5)

b) How many PFAM-A families match? What is the E-Value of the match? What is the most common domain architecture found amongst multi-domain proteins sharing that particular domain? (5)

c) For the protein O59398, how many members does the family have? How many sequences were used to build the profile to describe this family? What is the function of the members of this protein family? What does this protein have in common with the protein P5CS_HUMAN? (5)

Can you find the summary page for the metalloenzyme Family? Are there any GO terms available? How many sequences and structures are present? View a representative three-dimensional structure and rotate it to observe different portions of the molecule. Explain what you see (you can add a picture) (5)