Homework Assignment #2
Due: 10/11/2004 in class

1. (20 points) This task will consist of gathering the information from Protein Data Bank (PDB) and Swiss-Prot. Our protein of interest is Human Calcineurin (PDB code 1aui). Please answer the following questions:
   a. (3 points) What is the experimental method by which the structure of Human Calcineurin is determined?
   b. (2 points) Who are the authors of the paper in which the structure of 1aui was published?
   c. (2 points) When was the protein deposited in PDB?
   d. (3 points) How many chains (subunits) does the protein have, and how many residues are in each chain?
   e. (3 points) What are the (x, y, z) coordinates of the alpha carbon of the alanine #20 in the A subunit?
   
   Now, find the DBREF field and find a corresponding Swiss-Prot entry for the A subunit.
   f. (3 points) What is the Swiss-Prot ID of the A subunit of 1aui?
   g. (4 points) By looking at the Swiss-Prot keywords, what could you say about protein’s functional properties? How informative do you think keywords are (provide evidence for your statement)?

2. (5 points) Textbook, question 2.1.

3. (8 points) Textbook, question 2.2.

4. (8 points) Textbook, question 2.4. Show all your work.

5. (10 points) Given two nucleotide sequences \( s_1 = \text{AATCATT} \) and \( s_2 = \text{CAGTAG} \), a scoring matrix \( S \) and gap penalty of \(-3\), find the highest scoring subsequence alignment using Smith-Waterman algorithm. Show all your work (complete table with arrows and numbers). Scoring matrix adds 5 for a match and subtracts 4 for a mismatch (see lecture notes).

6. (10 points) Find an optimal global alignment for the sequences and scoring system from question 5. Show all your work.

7. (6 points) Describe in a few sentences how BLOSUM62 matrix was derived.

8. (6 points) Explain “functional constraints” in gene evolution? Do all genes evolve at the same rate?

9. (6 points) What is the difference between mutations and substitutions? What is the difference between synonymous and non-synonymous substitutions? Finally, what are alleles?
10. (6 points) Textbook, question 3.4. Why is the number of true substitutions different from the number of observed substitutions between two species since they last shared the common ancestor?

11. (4 points) What is saturation mutagenesis?

12. (7 points) What are the main conclusions you can draw from Figure 3.1, page 60. Explain in a few sentences.

13. (4 points) What is fixation?

Extra credit:

14. (15 points) For the following multiple sequence alignment:

<table>
<thead>
<tr>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>A</td>
</tr>
</tbody>
</table>

calculate the sequence profile (PSSM). Assume that the scoring matrix is +5/−4 for match/mismatch. Show how you calculated at least one full row of the profile.

Why are profiles generally better as a scoring system than general scoring matrices such as those from PAM or BLOSUM series?