

Paper Reading

The paper is somewhat hard to read, so you may want to read some other sources such as http://monod.uwaterloo.ca/~cs882/Old_courses/W03/Lecture3.pdf first.

Note that the original paper formulated the problem with arbitrary gap costs, compared with now commonly presented linear gap costs problem (which can be solved more efficiently).

A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins

by Saul B. Needleman and Christian Wunsch

First, the paper describes the matching algorithm. Note that this is a dynamic programming algorithm.

1. What is the maximum match?
2. What representation is used to store the matches?
3. What do the entries of the array represent? How is this related to the pathways through the array?
4. The simple method assigns a value of one to each match. This can be generalized to increase its sophistication. How?
5. How do we find the maximum-match pathway?
6. Reduce this to a longest path problem in a graph. What does the graph look like?

The algorithm can be used in experiments to determine if two proteins are related. In the experiment, there is always a chance that the observation is due to random events rather than a true relationship between the proteins.

7. What is the procedure used to find out whether the match is significant or may have occurred by chance?
8. What are the two possibilities of chance events that are proposed?
9. What is the null hypothesis?

Experiments are carried out to demonstrate that the proposed method gives reasonable answers.

10. What are the biological factors taken into account in constructing the model?
11. Two experiments are done: one with a pair of proteins expected to show homology and one with a pair of proteins that is not expected to show homology. Why isn't one experiment enough?
12. Read and interpret Table 1 on the results of the β -hemoglobin-myoglobin experiments.
13. Read and interpret Table 2 on the results of Ribonuclease-lysozyme experiments.