

Cpts 350

Hongqi Guo

011552159

08/31/2020

A: The ultimate function of the algorithm is to improve productivity, so the corresponding algorithm can be developed to perform database filtering. It can sift through a database of molecules, drastically reducing the number of candidate's researchers will test, while minimizing the need to test hundreds or thousands of compounds. The goal is to reduce the workload and increase efficiency.

B: I chose the sequence of amino acids in the polypeptide chain of the protein as the input structure. Because amino acids are linked by peptide bonds in the order of the genetic code, then they form polypeptide chains. We only need to input the corresponding amino acid sequence to know the corresponding protein.

C: 1. By comparing the amino acid sequence of proteins.

The advantage of this algorithm is to ensure the accuracy of the data, because the amino acid sequence of the protein is stable and cannot be changed. The disadvantage of this algorithm is that it is inefficient because the amino acid sequence of the protein is long. The basic principle of this algorithm is to mark each amino acid with a number, and then

draw a string and then compare the two strings. We can set a threshold of similarity, and when the data exceeds that threshold, we can tell that these two proteins are similar.

2. Compare pictures. Advantage is simple and fast, only need to compare the similarity of the picture. The disadvantage is that the difference rate can be very large. This algorithm USES image contrast and sets a threshold for image similarity to determine whether two proteins are similar or not.