Environment for Probabilistic Amino Acid Analysis

Shtephan, Michael
Borovitinov, Ilya

Software environment for probabilistic comparison of protein sequences has been developed. It contains algorithms and tools for finding matrix of alignment probability for aminoacid pairs as well as performing probabilities-based sequence alignment. Among other features, it allows to calculate distance between sequences and to perform search in homologous sequence database. This environment may be used in various knowledge domains including biochemistry, evolutionary biology and pharmacology (homology tree restoration, evolution analysis, protein structure discovery and functionality). Use of probabilistic method provides higher precision, especially for distant homologues. There are some projects that offer tools for protein comparison, but they don't use probabilistic models or fail to implement them properly. We haven't found any direct analogs of our project. Protein comparison model is based on interpreting sequence alignment presented in graph form and analysing it with probabilistic methods in order to find aminoacid pairs kinship probability matrix. Further processing of these results allows to obtain in-depth data (distance, most probable alignment etc). These data, in turn, can be used to perform database search. In order to create user friendly and fast Web service, Java language environment (in particular, Google Web Toolkit and HTML 5 technologies) was used.

Awards Won: