

Comparative Genomics by Computational Exploration: a Probabilistic Approach

(Steinar Thorvaldsen)

For a better understanding of environmental adaptations of proteins, orthologous sequences from different habitats may be explored and compared. We present statistical methods and trend-tests that are useful when the protein sequences in alignments can be divided into two or more groups based on known phenotypic traits such as preference of temperature, radiation, pH, salt concentration or pressure. The methods have been successfully applied in research on extremophilic organisms, and have special relevance for membrane proteins, since it is very difficult to obtain structural models of this important group. The approach is also suitable for high throughput sequence analysis.

The computational methods are implemented in *DeltaProt*, a software toolbox that may be useful in importing, analysing and visualizing data from multiple alignments of proteins. This is done by applying comparative statistical methods to comparisons of extremophile proteins versus orthologous genes from organisms of normal (e.g. mesophilic) habitats. Equipped with the algorithms, we are capable of visualizing and comparing various datasets of protein sequences. The framework of the algorithms also enables easy incorporation of extra information on structure, if such data is available.