

OPTIMA: A New Score Function For Distantly Related Protein Sequence Comparison

*Maricel Kann *, Bin Qian # and Richard Goldstein *#*

*Biophysics Research Division, University of Michigan.

Chemistry Department, University of Michigan.

mkann@umich.edu, bqian@umich.edu, richardg@umich.edu

The recent explosive expansion of the sequence databases has increased the need for ways to infer structural and functional information regarding newly sequenced proteins. One of the most effective ways to do this is to identify homologous proteins about which more is known.

Homolog identification methods rely on some score function to measure sequence similarity. While close evolutionary relationships can be confidently determined, the difficulty increases as the evolutionary relationship becomes more distant. The choice of score function is a critical aspect to improve the identification of distant relationships.

We describe a new method of determining the score function by optimizing the ability to discriminate between homologs and non-homologs. This new score function out-performs currently available score functions at identifying both distant and close homologies.