

Introduction to proteomics and the statistical aspects of protein research

Kanti V. Mardia*

University of Leeds, Leeds, UK K.V.Mardia@leeds.ac.uk

Michael G. Schimek*

Medical University of Graz, Graz, Austria michael.schimek@medunigraz.at

It would not be a great exaggeration to say that whereas the last century belonged to the physical sciences then this century must belong to the life sciences! This century will see the continuation of biomedical science, building on advances from the second half of the 20th century, which allowed us to characterize DNA and protein structure (represented by the coordinates of the atoms). Already, there are large and rapidly expanding databases of DNA and protein sequences, and related databases of the structure and function of biomolecules; such as proteins.

Protein bioinformatics is a rapidly developing field. Proteins are the workhorses of all living systems, and are a byproduct of DNA, but where DNA is like a recording system, protein is a playback system. A protein is a complex object in the sense that it takes a three-dimensional shape from a one-dimensional sequence of amino acids. One of the central problems of current life sciences is how this folding takes place. Although, there is now considerable statistical literature on gene expression (microarray) data analysis, protein bioinformatics has so far had little exposure to statistical thinking. This session brings together current developments and new challenges in this fast growing domain.

Key Words: Bayesian methods, computational statistics, folding, protein bioinformatics, structure prediction.