

BAYESIAN UNSUPERVISED STRUCTURAL CLASSIFICATION OF PROTEINS

P. Marttinen¹, J. Corander¹

¹*University of Helsinki, Helsinki, Finland*

Email: *pekka.marttinen@helsinki.fi*

We introduce a novel Bayesian model-based method for simultaneously classifying proteins into evolutionary groups and identifying conserved areas in the amino acid sequences within the groups. Biologically relevant information of the characteristics of conserved areas is used in the formulation of the model to significantly increase the power to detect different groups. Our model recognizes the areas in the sequences that are relevant for classification purposes and regards other areas of the sequence as noise. We have implemented the method using a fast stochastic optimization algorithm that yields a classification associated with the maximum estimated posterior probability. The algorithm has been shown to have high specificity and sensitivity in simulated and real data analyses.