Robust Classification for Microarray Gene Expression Data Analysis

Conference Paper · January 2013

CITATIONS
0

READS
25

6 authors, including:

Md. Matiur Rahaman
Zhejiang University
21 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Asif Ahsan
Zhejiang University
7 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Md Mamun Monir
Zhejiang University
9 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Md Alamin
Zhejiang University
21 PUBLICATIONS **7** CITATIONS
SEE PROFILE

SCP-18

Biophysical analysis of protein solubility and aggregation kinetics using short peptide tags

Monsur A. Khan[1*], Islam M. Monirul[2], and Kuroda Yutaka[1]

Low protein solubility or protein aggregation is a concern in several areas of biotechnology, and protein solubility becoming the focus of biophysical studies. Here we report the effect of 10 representative amino acids on the solubility and aggregation kinetics of proteins. The effect on protein solubility was determined by measuring the solubility of a simplified bovine pancreatic trypsin inhibitor (BPTI) variant, to which a short artificial tags containing the amino acid of interest was added at its C-terminus. As anticipated, positively charged residues significantly increased the solubility of the model protein, at both pH 4.7 and 7.7, whereas very hydrophobic poly-Ile markedly reduced the solubility of BPTI. Poly-Asp and poly-Glu barely affected BPTI solubility at pH 4.7, but induced an eight to ten-fold increase at pH 7.7, attributable to the ionization of their side chains. Regarding protein aggregation kinetics, we observed that the simplified BPTI precipitated promptly when the initial concentration exceeded some critical concentration. This critical concentration, as well as the rate of protein aggregation, was dependent on the type of amino acid composing the tags. These observations clearly demonstrated that the amino acid type, the initial protein concentration and the equilibrium time need to be considered when defining protein solubility. To the best of our knowledge, this is the first attempt to systematically assess the contribution of individual amino acids to protein solubility or to aggregation kinetics and propensities. These findings could eventually lead to the calculation of a polypeptide's relative solubility from its amino acid sequence.

Keywords: Biophysics, Protein solubility and Protein aggregation kinetics.

Correspondence to: YK
E-mail: ykuroda@cc.tuat.ac.jp

Authors' Affiliations: [1]Department of Biotechnology and Life Sciences, Graduate School of Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan; and [2]Department of Biochemistry and Molecular Biology, University of Chittagong, Chittagong-4331, Bangladesh.

SCP-19

Robust classification for microarray gene expression data analysis

Md. Matiur Rahaman[1*], Md. Asif Ahsan[1], Md. Mamun Monir[1], Md. Alamin[1], Md. Abdul Mazid[2] and Md. Nurul Haque Mollah[1]

A key challenge in recent time is the classification of microarray gene expression data into their expression categories or gene groups. Recently tumor/cancer classification based on gene expression data or gene classification based on gene expression on normal cell and cancer cell playing the significant role in cancer research. This is done by first learning how to classify, based on a training gene expression dataset containing labeled, and then predicting the label of a new gene or patient. There is several learning algorithm use for microarray gene expression data classification, as for example Bayes classifier, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Boosting etc. But there is no classifier that is superior over the other. Again, the gene expression dataset is very different from any one of the dataset. It is very high dimensional, contains thousands to tens of thousands of genes and publicly available data size is very small. Also, this data contains various labels of outliers and lead to unreliable and low accuracy results as well as the high dimensionality problem. Most of the current learning algorithms are not robust enough to handle these types of data properly. Therefore, we robustify the most popular Bayes classifier using the minimum β-divergence estimator for proper classification of microarray gene expression data. The performance of the proposed method is investigated using both the simulated and real microarray gene expression datasets. It is observed that the proposed method shows the better performance than traditional Bayes classifier as well as other existing classifier in presence of outliers. Otherwise, it produces almost same results.

Keywords: Microarray Gene expression, Classification, β –divergence and Robustness.

Correspondence to: MMR
E-mail: matiur.stat@gmail.com

Authors' Affiliations: [1]Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh; and [2]Institute of Biological Sciences, University of Rajshahi, Rajshahi-6205, Bangladesh.