

Bioinformatics Center -Biological Information Network-



Prof
AKUTSU, Tatsuya
(D Eng)



Assoc Prof
SUGISAKI, Hiroyuki
(D Sc)

Students

DUKKA BAHADUR, K. C. (M1)

FUKAGAWA, Daiji (M1)

SHIBATA, Yuzo (M1)

SAIGO, Hiroto (M1)

YAMAMURA, Masaki (M1)

Scope of Research

Due to rapid progress of the genome projects, whole genome sequences of many organisms and a draft of human genome sequence have been already determined. But, the determination of the whole genome sequence does not mean the end of analysis of genetic code. In order to understand the meaning behind the genetic code, we have been developing algorithms for analyzing proteomics data and genomics data. Recently, we focus on the following topics: protein structure prediction, classification of protein structures, motif extraction, inference of metabolic pathways and genetic networks, and analysis of two-dimensional electrophoresis gel images. We also conduct experimental studies on 4-Hydroxy-4-methyl-2-oxoglutarate (HMG) aldolase.

Research Activities (Year 2001)

Presentations

Selecting informative genes for cancer classification using gene expression data, Akutsu T, Miyano S, 2001 IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, 5 June.

A local search algorithm for local multiple alignment: special case analysis and application to cancer classification, Akutsu T, Int. Conf. Parallel and Distributed Processing Tech. and Appl., 26 June.

A Gibbs sampling algorithm for numerical sequences: detection of subtle motifs from protein sequence and structures, Asian Workshop on Protein Informatics, Akutsu T, Horimoto K (Saga Med.), 14 December.

Local multiple alignment of numerical sequences: detection of subtle motifs from protein sequences and structures, Akutsu T, Horimoto K (Saga Med.), Int. Conf. Genome Informatics, 18 December.

Grants

Akutsu T, Miyano S, Algorithms for finding common patterns in bioinformatics, Grant-in-Aid for Scientific Research (C) (2), 1 April 2001 - 31 March 2005.

Akutsu T, Genome Informatics Science (a member of the project), Grant-in-Aid for Scientific Research Priority Areas (C) , 1 April 2000 - 31 March 2005.

Topics

Local multiple alignment of numerical sequences: detection of subtle motifs from protein sequences and structures

Motif extraction is one of the well studied problems in Bioinformatics. We developed a new method to find motifs from multiple protein sequences and multiple protein structures (see Fig.1). The method consists of two parts: quantification and local multiple alignment. In the former part, protein sequences and protein structures are transformed into sequences of real numbers and real vectors respectively. In the latter part, fixed length regions having similar shapes are located. A variant of the Gibbs sampling algorithm [1], which can be applied to sequences of real numbers/vectors, is newly developed for finding common regions.

We also study a related problem: given positive and negative examples (sequences), find a PSSM (Position Specific Score Matrix)[1] which correctly discriminates between positive and negative examples (see Fig. 2). We proved some theoretical results on the computational complexity of this problem.

1. R. Durbin et al., Biological Sequence Analysis. Probabilistic Models of Protein Structures and Nucleic Acids, Cambridge University Press (1998).

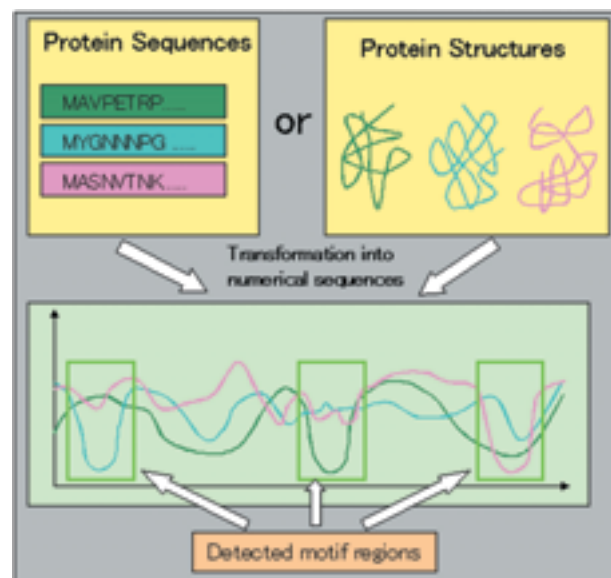


Figure 1. Motif detection using local multiple alignment of numerical sequences.

Cloning, sequencing, and expression of the gene encoding 4-hydroxy-4-methyl-2-oxoglutarate aldolase from *Pseudomonas ochraceae* NGJ1

4-Hydroxy-4-methyl-2-oxoglutarate (HMG) aldolase is involved in the α -keto acid pathway for degradation of meta-fission products of protocatechuate in bacteria. A DNA fragment that carried the proA gene encoding HMG aldolase was cloned from the chromosomal DNA of *Pseudomonas ochraceae*, and the coding region was assigned to the nucleotide sequence based on the N-terminal amino acid sequence of the purified from the organism. The proT and proH genes encoding putative transporter and 4-oxalomesaconate hydratase, respectively, were upstream, and the 3'truncated proL gene encoding 2-pyrone-4, 6-dicarboxylate lactonase was downstream from the proA gene in the same orientation on the DNA fragment. These enzymes are also members of the enzymes responsible for the α -keto acid pathway for protocatechuate degradation.

1. K. Maruyama et al., Cloning, sequencing, and expression of the gene encoding 4-hydroxy-4-methyl-2-oxoglutarate aldolase from *Pseudomonas ochraceae* NGJ1, *Bioscience, Biotechnology, and Biochemistry*, **65**, 2701-2709 (2001).

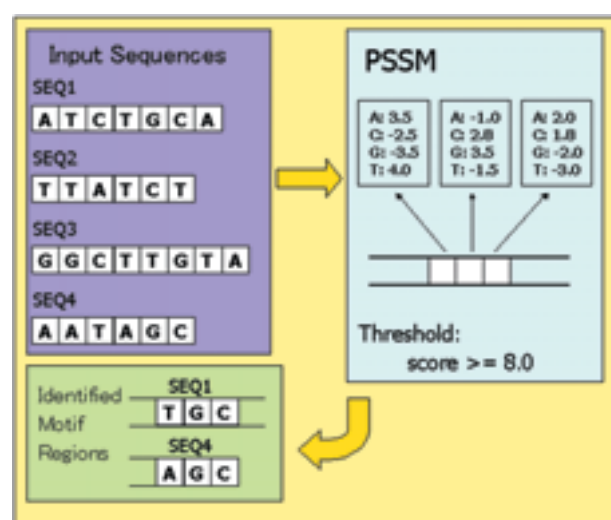


Figure 2. Motif detection using a position specific score matrix.