Bioinformatics Center
- Biological Information Networks -

http://www.bic.kyoto-u.ac.jp/takutsu/index.html

Prof
AKUTSU, Tatsuya
(D Eng)

Assist Prof
UEDA, Nobuhisa
(D Eng)

Assist Prof
HAYASHIDA, Morihiro
(D Inf)

Guest Res Assoc
NACHER, Jose C.
(Ph D)

PD
MATSUDA, Setsuro
(D Eng)

PD
OCHIAI, Tomoshiro
(D Sc)

Students

FUKAGAWA, Daiji (D3)
K. C., Dukka Bahadur (D3)
SAIGO, Hiroto (D3)
TAMADA, Yoshinori (D3)
MEIRELES, Lidio (M2)
BROWN, John (M1)
OOTAKA, Ryu (M1)
SAKAI, Daisuke (M1)
TAKEUCHI, Shigeki (M1)
POOLSAP, Unyanee (RS)

Visitors

Prof ZHANG, Xiang-Sun	Chinese Academy of Sciences, 16 May 2005
Dr VERT, Jean-Philippe	Ecole des Mines de Paris, France, 11 - 21 October 2005
Dr YAMANISHI, Yoshihiro	Ecole des Mines de Paris, France, 31 October - 2 November 2005
Mr MAHÉ, Pierre	Ecole des Mines de Paris, France, 5 - 17 December 2005

Scope of Research

Due to rapid progress of the genome projects, whole genome sequences of organisms ranging from bacteria to human have become available. In order to understand the meaning behind the genetic code, we have been developing algorithms and software tools for analyzing biological data based on advanced information technologies such as theory of algorithms, artificial intelligence, and machine learning. We are recently studying the following topics: systems biology, scale-free networks, protein structure prediction, inference of biological networks, chemo-informatics, discrete and stochastic methods for bioinformatics.

Research Activities (Year 2005)

Presentations

Clique Based Algorithms for Protein Threading with Profiles and Constraints, K.C. D, Tomita E, Suzuki J, Horimoto K, Akutsu T, The 3rd Asia-Pacific Bioinformatics Conference, 18 January.

On Construction and Transformation of Scale-free Networks, Akutsu T, The 50th NIBB Conference on Structure and Dynamics of Complex Biological Networks, 8 February.

Inferring a Graph from Path Frequency, Akutsu T, Fukagawa D, The 16th Annual Symposium on Combinatorial Pattern Matching, 22 June.

On Transformation and Construction of Scale-free Networks, Akutsu T, International Workshop on Complex Networks, 24 June.

Kernel-based Approaches to Classification and Design of Protein Sequences and Chemical Compounds, Akutsu T, 2005 International Joint Conference of InCoB, AASBi, and KSBi, 22 September.

On Structures of Metabolic Networks and Protein-Domain Networks, Akutsu T, Mathematical Analysis of Complex Phenomena in Life Sciences, 26 October.

Grants

Akutsu T, Miyano S, Maruyama O, Ueda N, Algorithms for Extracting Common Patterns from Structured Biological Data, Grant-in-Aid for Scientific Research (B), 1 April 2004 - 31 March 2008.

Akutsu T, Mathematical Analysis of Structure and Dynamics of Biological Information Networks, Grant-in-Aid

A Novel Representation of Protein Sequences for Prediction of Subcellular Location Using Support Vector Machines

As the number of complete genomes increases, accurate methods to automatically predict subcellular locations of proteins are increasingly helpful to annotate their biological functions. In order to improve predictive accuracy of the many prediction methods developed to date, we have proposed a novel representation of protein sequences. This representation involves local compositions of amino acids and twin amino acids, and local frequencies of distance between successive (basic, hydrophobic, and other) amino acids. For calculating the local features, each sequence is split into the N-terminal, middle, and C-terminal parts. The N-terminal part is further divided into four regions to consider ambiguity in the length and position of signal sequences. We tested this representation with support vector machines on two data sets extracted from the SWISS-PROT database. Despite simplicity of the representation, overall accuracies of our method were more than 87% and 91% for eukaryotic and prokaryotic proteins, respectively. These are almost the highest accuracy among the methods using sequence information alone. Based on the above methodology, we have also developed a web server, SLP-Local (http://mint.kuicr.kyoto-u.ac.jp/~smatsuda/slplocal.html).

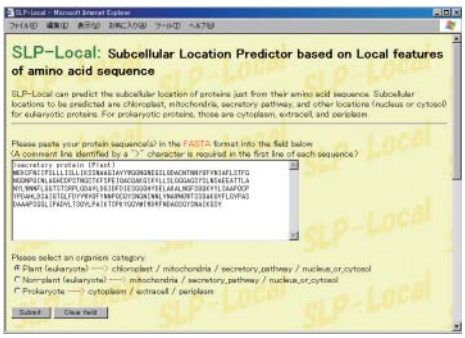

Figure 1. A snapshot of the top page of SLP-Local.

Matsuda, S., Vert, J.-P., Saigo, H., Ueda, N., Toh, H., and Akutsu, T. (2005) *Protein Sci.*, **14**, 2804-2813.

Protein Domain Networks: Scale-Free Mixing of Positive and Negative Exponents

Proteins are essential molecules and responsible for most cellular processes. A protein region with well-defined structural and functional properties is called *protein domain*.

Here, we present a theoretical model for studying the protein domain networks, where one node of the network corresponds to one protein and two proteins are connected if they contain the same domain. The resulting distribution of nodes with a given degree, k, shows not only a power-law with negative exponent $\gamma = -1$, but it resembles the superposition of two power-law functions, one with a negative exponent and another with a positive exponent $\beta = 1$. We call this distribution pattern "*scale-free mixing*". To explain the emergence of this superposition of power-laws, we propose a basic model with two main components: (1) mutation and (2) duplication of domains. Precisely, duplication gives rise to complete subgraphs (i.e., cliques) on the network, thus for several values of k a large number of nodes with degree k is produced, which explains the positive power-law branch of the degree distribution.

The results of our model were compared with protein domain networks of six organisms generated with data from the Uniprot Knowledgebase-Swissprot database for protein sequences and using InterPro, Pfam and Smart for domain databases. Our results indicate that the signal of this positive power-law branch of the measured distribution is observed in experimental data and it is conserved among organisms from *E. coli* to *H. sapiens*.

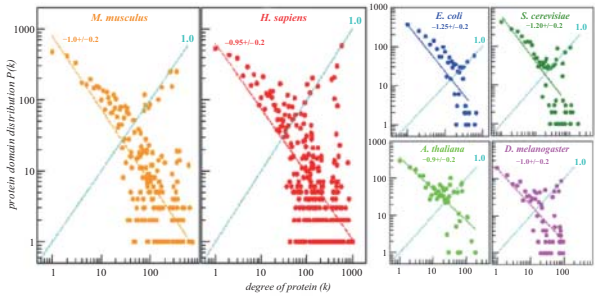

Figure 2. The degree distribution $P(k)$ of the protein domain network in *M. musculus, H. sapiens, E. coli, S. cerevisiae, A. thaliana and D. melanogaster* organisms exhibits a *scale-free mixing* pattern.

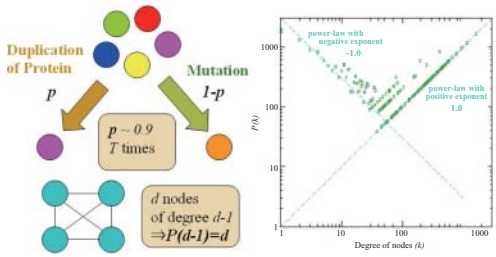

Figure 3. Scheme of the process that generates a *scale-free mixing* distribution, and the results of our proposed model.

Nacher J. C., Hayashida M., Akutsu, T., Protein domain networks: scale-free mixing of positive and negative exponents, *Physica A*, in press.

for Priority Area Research, 1 April 2005 - 31 March 2010.

Ueda N, Statistical Language Models that Generate a Pair of Sequences for Sequence Analysis, Grant-in-Aid for Encouragement of Young Scientists, 1 April 2003 - 31 March 2006.