Based on a manually-curated database of NESs built in 2003, an analysis of the amino acid composition of NES regions revealed statistical over-representation of glutamate, aspartate, and serine besides the consensus hydrophobic residues. To find new insights on this signal, we are now analyzing differences between the NES regions and the hydrophobic core regions. Several features potentially related to this issue were found and then used to make a classifier to discriminate between these two regions sharing similarity sequence characteristics but different functions.

- 2P-242 立体構造を用いたタンパク質機能予測法 FCANAL のさまざまなタンパク質への応用
FCANAL, structure-based protein function prediction method, applied to various types of proteins.

Yuuichi Watanabe (1), Kousuke Kaido (1), Takashi Ando (1), Ichiro Yamato (1), Satoru Miyazaki (2) ((1) Department of Biological Science and Technology, Tokyo University of Science; (2) Department of Pharmaceutical Science and Technology, Tokyo University of Science)

Structural genomics projects are beginning to produce protein structures with unknown functions, thereby requiring high-throughput methods to predict functions of these proteins. Although sequence-based function prediction methods have been used extensively, structure-based prediction is believed to provide higher specificity and sensitivity because functions are closely related to the three-dimensional structures of functional sites. We have developed FCANAL, a method to predict functions using score matrices obtained from the distances between C α atoms and frequencies of amino acid appearance around functional sites of proteins with known functions (A. Suzuki, et al., (2005) Chem-Bio Info. J, Vol. 5, pp39-55). Now, FCANAL can predict protein functions of many proteins, binding proteins of low molecular weight substrates, sugar binding proteins, DNA binding proteins, enzymes, etc. We further extended our FCANAL to predict functional sites for protein-protein interaction. Using only one key residue at the interacting site to make score matrix was not enough to give accurate prediction. Thus we used a few key residues and the prediction accuracy was improved. In case of enzyme function prediction, we sometimes faced difficulty, since we found different structures even if the enzyme function was the same. We analyzed their structures, and we found that they were classified into a few sub-groups. Based on this finding, score matrices for respective sub-groups were generated and applied, resulting in increase of prediction accuracy. With these improvements, FCANAL is now useful to detect functional sites in various protein structures deposited in PDB with high accuracy. In this respect, FCANAL would help identifying the functions of newly determined structures.

- 2P-243 カーネル密度推定を用いたベイズ分類器による配列上の蛋白質間相互作用部位予測
The prediction of protein-protein interaction sites in sequences using a naïve bayes classifier with a kernel density estimation

Yoichi Murakami (1), Kenji Mizuguchi (1) ((1) National Institute of Biomedical Innovation)

Identification of protein-protein interaction sites is not only critical for understanding protein functions, but also helpful in developing new drugs. In spite of much effort in structural determination of many proteins, the number of three-dimensional protein structures is still much smaller than those of protein sequences (e.g. by May 5, 2009, there are 466,739 identified protein sequence in UniprotKB/Swissprot and only 57,706 known protein structures in PDB). Thus, it is now significantly important to identify protein-protein interaction sites from sequence information only. A Naïve Bayes Classifier (NBC) with a Kernel Density Estimation (KDE) has been applied to the prediction of protein-protein interaction sites in sequences. Prediction models have been created using local sequence information; position specific scoring matrix (PSSM) and predicted accessibility (pASA), over overlapping subsequences of various lengths (or sliding windows), and NBC is trained to distinguish between interaction residues and non-interacting residues. Previously published methods use other machine learning techniques such as Support Vector Machine or Neural Network, but these machining learning techniques are known to be sensitive to class skew in the training dataset and tend to become over-fitted to the imbalanced dataset. However, NBC is insensitive to such dataset. In addition, NBC is known to be simple, fast and easy to implement, and performs well. In this research, we present a more reliable and accurate method to predict protein-protein interaction sites in sequences using PSSM and pASA based on NBC with KDE.

- 2P-244 ヒトゲノムからの電荷周期性を持つ DNA 結合タンパク質の解析
Analysis of DNA binding proteins with charge periodicity from human genome

Runcong Ke (1), Noriyuki Sakiyama (1), Shigeki Mitaku (1) ((1) Dept. of Appl. Phys., Grad Sch. of Eng., Nagoya Univ.)

Recently we found that there are many proteins with significantly charge periodicity of 28 residues in amino acid sequences from human genome. It was

also found that this type of proteins increased in vertebrate genomes and approximately 3% proteins from human genome belonged to this type. Furthermore, the intracellular localization of many of those proteins was nucleus. However, the details of the functions of those proteins have not been studied.

We extracted proteins with charge periodicity of 28 residues (PCP28) from human genome, leading to the total number of 972. Among those, 561 proteins are zinc-finger proteins and 411 proteins are non-zinc-finger proteins according to the genome annotations. The three-dimensional structures of some zinc-finger proteins have been reported, showing the charge periodicity of around 28 residues, corresponding to the zinc finger motif. All human PCP28 were analyzed by the prediction system SOSUI and SOSUIpcp28 (not open to internet) to classify into the three categories of PCP28: nuclear, cytoplasmic and membrane/secretion PCP28. The ratios of nuclear, cytoplasmic and membrane/secretion PCP28 were 75%, 14% and 11%, respectively. Cytoplasmic PCP28 included myosin, tubulin and cytosolic proteins. There were 85 unknown PCP28 which are predicted to localize in nucleus but non-zinc-finger type, inferring that many of those proteins were DNA binding proteins.

- 2P-245 発生過程の細胞上皮系にかかる力の推定
Estimating forces in the growing epithelial cells

Shuji Ishihara (1), Kaoru Sugimura (3), Atsushi Miyawaki (3) ((1) Graduate School of Arts and Sciences, Univ. of Tokyo; (2) PRESTO, JST; (3) RIKEN BSI)

In the course of development, tissues undergo deformation, which is driven by the mechanical forces regulated by the activity of constituent cells. However, there is no proper experimental method to directly measure the forces in the cell population inside the animal body. Here we propose a method to estimate the forces from observed cell shapes.

Consider that the geometry of epithelial cell population is represented by an assemblage of 2D polygons. Forces in the cell population are described by the isotropic pressures of respective cells and tensions applied in the adherent surface between cells. By considering the geometrical forms of the cells, we obtained balance equations of the afore-mentioned forces. We can estimate the unknown values of forces by solving the balance equations, however, the number of conditions is less than that of unknowns. So, we applied an inverse problem technique that is a generic method to deal with such situations. We are now trying to evaluate the validity of the estimated force in the Drosophila epithelial tissue. The procedure would be applicable to analyze various feature of growing tissue.

- 2P-246 二温度系におけるロバストネスの進化の統計力学的研究
A statistical-mechanical study of evolution of robustness in noisy environment

Ayaka Sakata (1), Koji Hukushima (1), Kunihiko Kaneko (2) ((1) Graduate School of Arts and Sciences, University of Tokyo; (2) Graduate School of Arts and Sciences, University of Tokyo, Complex Systems Biology Project, ERATO, JST)

In biological systems, expression dynamics that can provide fitted phenotype patterns with respect to a specific function have evolved through mutations. This has been observed in the evolution of proteins for realizing folding dynamics through which a target structure is shaped. We study this evolutionary process by introducing a statistical-mechanical model of interacting spins, where a configuration of spins and their interactions $\mathbf{J}$ represent a phenotype and genotype, respectively. The phenotype dynamics are given by a stochastic process with temperature T_S under a Hamiltonian with $\mathbf{J}$. The evolution of $\mathbf{J}$ is also stochastic with temperature T_I and follows mutations introduced into $\mathbf{J}$ and selection based on a fitness defined for a configuration of a given set of target spins. Below a certain temperature T_S^{c2} , the interactions $\mathbf{J}$ that achieve the target pattern evolve, whereas another phase transition is observed at $T_S^{c1} < T_S^{c2}$. At low temperatures $T_S < T_S^{c1}$, the Hamiltonian exhibits a spin-glass like phase, where the dynamics toward the target pattern require long time steps, and the fitness often decreases drastically as a result of a single mutation to $\mathbf{J}$. In the intermediate-temperature region, the dynamics to shape the target pattern proceed rapidly and are robust to mutations of $\mathbf{J}$. The interactions in this region have no frustration around the target pattern and results in funnel-type dynamics. (We call this region "local Mattis states".) These results imply that the funnel-type dynamics is a consequence of evolution subjected to thermal noise beyond a certain level; this also leads to mutational robustness of the fitness. We will explain the mechanism of the appearance of local Mattis states with the statistical-mechanical method.

- 2P-247 細胞性粘菌の細胞運動を支配する位相方程式と最小記述
The phase equation and minimal description for cell locomotion of dictyostelium

Miki Yamamoto (1), Yusuke Maeda (2), Hirokazu Tanimoto (1), Masaki Sano (1) ((1) Department of Physics, the University of Tokyo; (2) Rockefeller University)

We concern an amoeba locomotion that is a characteristic motion of eukaryotic cells. In this presentation, we do not mention a lot about realistic mechanisms to enable cells to locomote, but mention about the constraints that cells should obey. Cell is an extended object that occupies a finite region of circumstances, so that the motions of cells must obey to some laws that satisfy the symmetries that the