system for animal cells which do not have chloroplast was also developed in comparison to the system for plants. We collected proteins of plant genomes into four groups from Swiss-Prot 58.0. First, proteins targeted to ER were discriminated from proteins in other organelles. The average values of several physicochemical properties were calculated around maximum point of hydropathy index at N-terminus, by which a score for the prediction of ER was obtained. Second, we discriminated among chloroplast, mitochondria, and others based on the distribution of several physicochemical parameters around N-terminus. The prediction scores were obtained by canonical discriminant analysis. When tested in a five-fold cross-validation procedure, this system demonstrates 88.1% overall accuracy. Compared with TargetP and PredSL, the accuracies of this system were better than other systems. We will also report the predicted result of animal cells.

3P-294 触媒ドメインの連結が代謝系酵素へ与える機能的変化

Enzymes with multiple catalytic domains in metabolic pathways

Keiko Tan'ya (1), Motonori Ota (2), Ken Nishikawa (3). (1:Grad. Sch. Biosci&Biotech., Tokyo Inst. Tech.; 2:Grad. Sch. Info Sci., Nagoya univ.; 3:Dept. Bioinfom., Maebashi Inst.Tech.)

The mechanisms to produce proteins with novel or modified functions include sequence modification and domain recombination. Many previous studies focused on the role of the former, whereas the role of the latter is unclear. In this study, we have examined enzyme superfamilies in terms of domain combination, sequence diversity and EC number, and then investigated the correlation between the fusion of catalytic and periferal domains and the functional modification. We have found that the sequence modification at the catalytic domain is dominant to alter the protein function rather than the recombination of peripheral domains.

Trough these analyses we also found the fusion of catalytic domains is not rare. To investigate this event more estensively, we built an dataset of 38,310 enzyme sequences from 384 genomes by integrating three databases, GTOP[1], SCOPEC[2], and EzCatDB[3]. The data were grouped into one class if they share same catalytic domains at the superfamily level and EC number. Out of 343 classes, 61 classes (2,984sequences) are enzymes with more than one catalytic domain. Structural and functional features of fused catalytic domains were examined in terms of fold classes, EC number and the reaction in the metabolic pathways (KEGG).

[1]:a database of protein fold assignments to genome sequences (http://spock.genes.nig.ac.jp/~genome/gtop.html)

[2]:a database of catalytic domain of enzymes (*Bioinformatics*.(2004)4;20,i130-6.)

[3]:a database of enzyme catalytic mechanisms (http://mbs.cbrc.jp/EzCatDB/)

3P-295 蛋白質ー蛋白質間の相互作用残基ペアの配列からの予測

Prediction of interacting residue pairs in protein-protein complexes from single sequences and alignment profiles

Shandar Ahmad (1), Chioko Nagao (1), Kenji Mizuguchi (1). (1: National Institute of Biomedical Innovation, Osaka)

Prediction of protein-protein interface residues from sequence or structure of the unbound protein is a major challenge for scientists. Some machine learning methods have been recently developed, which predict interacting residues on single protein surfaces. On the other hand there are methods that try to predict pairs of proteins likely to interact with each other. Till date there is no method, which can use amino acid sequences of a pair of proteins and predict their likely interactions. We have developed a data set of protein-protein interacting sites mapped to protein sequence. We aim to analyze sequence-dependence of pairs of residues on interacting proteins and predict such interactions from sequence information. The greatest challenge in such a prediction method is a large amount of negative interaction pairs compared to a very small number of residue pairs, which interact. We are solving these problems by introducing data balancing techniques and Monte-Carlo like random selection. Preliminary results of predictions are promising and we hope that these methods will be successful in predicting interacting residue pairs from sequences at a high speed.

3P-296 DNA一蛋白質間の塩基特異的相互作用と相互作用様式の予測

Prediction of base-specific contacts and contact-types in DNA-binding proteins

Munazah Andrabi (1, 2), Kenji Mizuguchi (1, 2), Akinori Sarai (3) and Shandar Ahmad (1,2) (1: National Institute of Biomedical Innovation, Osaka; 2: Graduate School of Frontier Bioscience, Osaka University; 3: Kyushu Institute of Technology, Fukuoka)

There are a number of methods to predict DNA-binding residues from amino acid sequence or their evolutionary information. They use machine-learning approaches such as neural networks or SVM. However none of the existing methods predicts the target nucleic acids to which a given residue environment will interact. Thus they do not take advantage of the detailed contact information available in complex structures. We have created data sets of proteins with residue-wise information of their binding to each of the four bases, their nature of contacts and calculated the propensity scores for each combination. We also developed neural network models for each one of them. Results of prediction show subtle differences between the propensity scores and base specificity of binding sites. We have examined backbone-backbone, side-chain-side-chain and side-chain-backbone contacts between protein and DNA and analysed their predictability from single sequence as well as their evolutionary profile. We find that the best predicted contacts between protein and DNA are between side chain of amino acid and backbone of DNA. Contacts with DNA side chain are fewer and more difficult to predict.

3P-297 タンパク質機能予測法 FCANAL のさまざまな酵素への適用

Extention of FCANAL, structure-based protein function prediction method, to various types of enzymes.

Yuichi Watanabe(1), Tadashi Ando(2), Ichiro Yamato(2) and Satoru Miyazaki(3).(1:Dept BiolSci and Tech,Tokyo Univ of Sciences;2:Dept of Pharma Sci and Tech,Tokyo Univ of Sci)

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. Now, FCANAL can predict 72 types of enzyme functions. However, during extending this method to various types of enzyme functions, it was difficult to predict some groups of functions with our previous method. When we analyzed their structures, these groups were classified into a few sub-groups based on structural differences. Based on this view, score matrices for respective sub-groups were generated and applied for the enzymes, resulting in increase of prediction power. We are interested in the evolutionary relationships of these sub-groups. With this improvement, FCANAL is now useful to detect functional sites in protein structures deposited in PDB with high accuracy. In this respect, FCANAL would help identifying the functions of newly determined structures.

3P-298 On-web GP II：マイクロバイオーム解析と普遍的病原微生物診断法のためのウェブデータベース自動化

Revision of web database for microbiome analysis and universal pathogen diagnosis: On-web GP II

Takayuki Tsuruoka, Ahmed Shamim, Miho Suzuki and Koichi Nishigaki. (Graduate School of Science and Engineering, Saitama Univ.)

Genome profiling (GP) is a random PCR-based methodology used for identification of species by genotype. This requires computer-based processing and database managing also. Due to the advantageous properties in species identification, GP is highly useful in analyzing microbiomes (or metagenomes) and applying to universal pathogen diagnosis. A diagnostic system aimed to detect TB, AIDS, and the other pathogenic agents at a single access. It has already been applied successfully to identifying and classifying insects (more than 50 species) and the other various organisms (viruses to mammals). However, to perform microbiome analysis and/or universal pathogen diagnosis, the current web system (*On-web GP*) needs to be sophisticated to be able to process data without human interruption. In this context, we developed a program that makes clients needless to process data in a man-machine mode and contributes to raise the experimental accuracy. Through these improvements, we could construct a comfortable and usable web database (*On-web GP II*) for a public use.

3P-299 ゲノム距離測定になぜ緩和型相同検索法'Oligostickiness'が有効か？

Why is the relaxed measure of sequence similarity, oligostickiness, more effective in obtaining the genome distance?

Shamim Ahmed, Takayuki Tsuruoka, Sachika Tsuji, Miho Suzuki, Naoto Nemoto, Manish Biyani and Koichi Nishigaki. (Graduate School of Science and Engineering, Saitama Univ.)

Genome distance, d, can be defined from the genome similarity, s, between two organisms ($d = 1\text{-}s$). However, the most straightforward approach which is based on sequence alignment and Hamming distance measurement is effective only in the DNA of a gene size (kb or so) but cannot be applied to giga-base genome sequences due to the computing explosion. So, any elaboration is required.

For this purpose we have introduced a physicochemical property, *oligostickiness* (a measure of affinity of an oligonucleotide to a particular DNA), and then we have constructed how to measure the genome distance experimentally (Genome profiling (GP)) or computationally (*Oligostickiness* (OS) analysis), and have demonstrated their validities (**GP**: 'General method to identify species' *Gene*, 2002; 'Universal classification of speceis' *Int. J. Plant* Genomics 2006; 'Mutagen analysis system of a high sensitivity J. Biochem. 2007; **OS**: '*Oligostickiness* analysis of genomes' *Bioinformatics*, 2002; ' Chromosome analysis' *J. Comp. Chem. J.* 2004 , and others).

In this paper, such effectiveness is theoretically interpreted. In brief, if we consider that the genome sequences are incessantly changing due to the environmental fluctuation, under such circumstances "relaxed" sequence analysis is more robust and more informative than "strict" one caused by the difference in the degree of degeneracy of state.

3P-300 オリゴスティッキネス：宿主と寄生種の類縁関係発見の強力なツール

A Relaxed Measure of Sequence Similarity, Oligostickiness: Its Usage

Shamim Ahmed, Ayumu Saito, Takayuki Tsuruoka, Miho Suzuki, Naoto Nemoto and Koichi Nishigaki. (Graduate School of Science and Engineering, Saitama University, saitama 338-8570, Japan)

Oligostickiness (σ) is defined to be the binding affinity of an oligonucleotide (probe sequence) to a genomic DNA, conforming to the constraint that the deltaG of binding should be lower than a threshold value. This parameter is known to be useful for revealing veiled sequence characteristics (chromosome texture [Ayumu Saito and koichi Nishigaki 2004, JCCJ] and others). In this paper, we further

derived a parameter SOSS (Set of oligostickiness similarity score) from a set of oligostickiness obtained by 12 different probes and used it to calculate the similarity of genomes (n = 33) together with the G + C content and the codon usage analyses termed COUSS (Codon usage similarity score) to evaluate the similarity of two genomes. We found (i) The host-parasite relationship could be clearly revealed by SOSS values providing distinctively higher values than that of non-related bacteria and phages. (ii) Lysogenic phages showed higher SOSS to their host than nonlysogenic phages. (iii) Lysogenic phages tend to have a similar G + C content to that of their host; however, this property is not diagnostic in discriminating or characterizing the phages. (iv) Lysogenic phages tend to score higher COUSS than nonlysogenic phages to their host with a few exceptions. These results indicate that oligostickiness analysis is most significant in revealing host-parasite relationship. Therefore, oligostickiness analysis can be used to discover unknown host-parasite relationships among a diversity of wild viruses, especially offered by the metagenome analysis.

3P-301 蛋白質進化のコドン置換モデル

A Codon-based Model for Evolution of Protein-coding DNA Sequences

Sanzo Miyazawa. (Graduate School of Engineering, Gunma University)

Reliability of phylogenetic analyses strongly depends on models assumed for the substitution processes of nucleotides and amino acids. Here we propose a new codon-based model in which codon replacements with single and multiple base changes occur in the same order of time, and a new evaluation of the selective restraints on amino acid replacements is used. The codon substitution process is modeled as the time-reversible Markov process and the codon mutation rates are assumed to be proportional to the equilibrium codon frequencies. All codon replacements are assumed to be lethal or neutral, and the mean fraction of neutral replacements is approximated by the exponent of a linear combination of two terms, mean increment of contact energies between an amino acid and surrounding amino acids due to an amino acid exchange in a protein structure and volume difference between an amino acid and its replacement. This codon-based model is examined by maximizing its likelihoods to the amino acid substitution matrices, JTT tallied by Jones et al., mtREV estimated for mitochondria- and cpREV for chloroplast-encoded proteins. The maximum likelihoods indicate that the present evaluation of the selective restraints better fit the JTT than both Grantham'sand Miyata's distances. The Akaike information criterion scores indicate the significance of multiple base changes. The estimated values of parameters confirm transition/transversion rate bias and codon usage bias, and reproduce characteristic features in the temporal changes of the log-odd matrix.

3P-302 酵母のタンパク質間相互作用ネットワークにおける進化的拘束としてのモジュラー性

Modularity as an evolutionary constraint on the yeast protein interaction networks.

Masataka Kikuchi(1), SoichiOgishima(1), Hiroshi Tanaka(2).(1: Dept. of Bioinformatics, Medical Research Institute , Tokyo Medical and Dental University; 2: Dept. of Bioinformatics, School of Biomedical Science, Tokyo Medical and Dental University;)*

Recently, comprehensive data of protein interactions have been accumulating based on large-scale hybridization methods, and makes it possible to understand the structure and the design of cellular networks. Protein interaction networks (PINs) are known to show remarkable global structures with scale-free, and hub proteins are reported to show significantly low evolutionary rates [Fraser (2002) Science]. PINs are also known to show “small-world” properties, in other words “modularity”, thus we hypothesized that it should be an evolutionary constraint on PINs. To prove our hypothesis, we examined the Spearman correlation coefficient between clustering coefficients and evolutionary rates (dN/dS) of the corresponding genes in the yeast PINs [Guldener (2005) NAR], and showed significantly negative correlation between them in the yeast (Spearman corr. coef.=-0.113, P=1.53×10-7). We also examined the difference in evolutionary rates between party hubs (which interact with most of their partners simultaneously) and date ones (which bind their different partners at different times or locations) [Han (2004) Nature].We found that party hubs exhibit significantly lower evolutionary rate than date ones (Welch t test, P=6.245×10-6).These results suggest that “modularity” might be an evolutionary constraint in the evolution of yeast PINs.

3P-303 ユーグレナ植物のコハク酸脱水素酵素は新規の超分子呼吸鎖複合体である

Succinate dehydrogenase in the Euglenozoa is a novel supramolecular respiratory complex

Tatsushi Mogi (1) and Kiyoshi Kita (1)(1: Dept Biomedical Chemistry, Graduate School of Medicine, Univ of Tokyo)

Mitochondrial succinate dehydrogenase (Complex II) is a four-subunit membrane-bound enzyme (~130 kDa). Flavoprotein SDH1 oxidizes succinate and transfers electron to iron-sulfur subunit SDH2, consisting of the 2Fe-2S plant ferredoxin (Ip_N) and the 4Fe-4S plus 3Fe-4S bacterial ferredoxin (Ip_C) domain. At a boundary between SDH2 and membrane anchor subunit SDH3 and SDH4, the reduction of ubiquinone takes place. Recently we purified Complex II from the protozoan parasite *Trypanosoma cruzi*, the causative agent of Chagas disease and found that the Trypanosomatidae Complex II was a 550-kDa dimer and composed of six each of hydrophilic and hydrophobic subunits (1). Notably SDH2 is heterodimeric; $SDH2_N$ and $SDH2_C$ contain the Ip_N and Ip_C domain, respectively. Here we examined properties and the structure of Complex II in *Euglena gracilis* and *Astasia longa*, the evolutionary cousins in the Euglenozoa. We found that the enzymes were resistant to quinone-related inhibitors, due to divergent SDH3 and SDH4 subunits. The current database suggests that the Euglenida Complex II is an eight-subunit enzyme lacking soluble subunit SDH5~SDH7. Sequence analysis suggests that Euglenozoa-specific subunits have evolved by duplication and degeneration of canonical subunits. If such degenerated duplicons retain the protein network-forming potential, they would be organized as a novel supercomplex. This could be one of mechanisms for overweighting the mitochondrial respiratory complexes.1) J. Morales, T. Mogi, K. Kita (2008) *Biochim. Biophys. Acta* 1777, S94.

3P-304 様々な粒径・元素のナノ粒子プローブセットで標識された標的DNAの無増幅定量検出法の開発

Use of a nano-particle probe set for quantitative detection of target DNA without amplification: identification through size and element

Hyonchol Kim (1), Hiroyuki Takei (1) and Kenji Yasuda (1,2). (1: Kanagawa Academy of Science and Technology; 2: Inst Biomaterial Bioengineering, Tokyo Medical and Dental Univ)

Monitoring functions and variations of expression profiles in a cellular system is a key to understanding the mechanisms of living organisms, especially to identifying those phenotypical characteristics. The DNA microarray technology has been widely used for such a purpose; however, this method is associated with some technical problems. One of them is lack of the quantification capability due to nonlinear sample amplification. The obtained result is an averaged value of many cells and monitoring of the expression profile in a single cell is in general difficult using this method. These problems are in principle caused by the underlying optical detection methodology based on fluorescent dyes; therefore a new detection system should be devised to measure the expression profile in single cell level. We tried to carry out quantitative and highly sensitive detection of expressed biomolecules using nano-particle probes as labels instead of fluorescent dyes. Nano-particles were prepared different metals in various sizes and probe DNAs were immobilized on the particles. Target DNAs were labeled with the nano-particle probes and both kind and amount of the labeled particles were quantitatively measured using a field emission scanning electron microscope (FE-SEM). Sensitivity of our method was tested using custom ordered DNA microarrays and targets in the femto molar concentration range could be detected without any amplification. These results indicated the advantage of the use of nano-particles as labels.

3P-305 極低温電子顕微鏡による高分解能像撮影条件の探索

Which temperature is optimal for high-resolution Cryo-EM ?

Masamichi Ashihara (1), Takayuki Kato (1), Hideyuki Matsunami (2), Tomoko Miyata (1) and Keiichi Namba (1). (1: Graduate School of Frontier Biosciences, Osaka University; 2: Okinawa Institute of Science and Technology)

The principal resolution limitation in electron cryomicroscopy of frozen-hydrated biological material is low contrast and low signal-to-noise-ratio (SNR) that results from the radiation damage to biological specimen by the electron beam. Increase in the electron dose limit that specimens can tolerate would be the simplest way to improve the SNR. It has been demonstrated that cooling the specimen to 4 K with liquid helium reduces the damage and allows statistically better images to be recorded. In fact an atomic model of the bacterial flagellar filament has been built from images recorded at 4 K. However, 3-D reconstructions with resolutions of 5 Å or better from images recorded at ~ 100 K have recently been reported. We therefore decided to reexamine how much the liquid-helium cooling is advantageous for recording high-resolution data to higher temperatures and determine optimal electron dose at each resolution. We collected electron diffraction data from 2-D crystals of bacteriorhodopsin (bR) at 4 K, 50 K and 80 K, and quantitatively evaluated the radiation damage at each specimen temperature by following the decay in diffraction intensities with dose accumulation. We found that the radiation damage was smallest at 50 K. The same evaluation with thin 3-D crystals of a flagellin fragment is under way to confirm the result. We also compared 3-D image reconstructions of the Hepatitis B virus core from images collected at different temperatures and accumulated dose to see the temperature effect on image quality.

3P-306 原子間力顕微鏡を用いた抗体-抗原相互作用に対する温度の影響の計測

Measurement of influence of temperature on antibody-antigen interaction by AFM

Jun'ichi Wakayama (1), Shigeru Sugiyama (1). (1: Nano-Biotechnology Laboratory, National Food Research Institute)

Molecular recognition such as interaction between receptor and ligand, as well as antigen and antibody, can be characterized by an energy landscape consisting of potentially energetic barriers along a reaction coordinate. Dynamic force spectroscopy (DFS), which consists of measurements of the rupture forces between a molecule of interest and a complementary molecule at different loading rate by use of atomic force microscopy (AFM), is useful tool to investigate the energy landscape of the interaction at the single molecule level. However, in the measurements of rupture force between antibody and antigen, several studies showed that the forces are often obscured by nonspecific interaction. To study exactly the energy landscape of antibody-antigen interaction by means of DFS, one should reduce the nonspecific interaction. Recently, we demonstrated that the use of an experimental solution containing with both a detergent and a