protocols such as mutagenesis. A Support Vector Machine has been applied to training models on a dataset of 81 RNA-binding proteins from the PDB (RNAset81). Models have been created using different combinations of features: position specific scoring matrix, residue interface propensity, predicted accessibility and hydrophobicity, over overlapping subsequences of various lengths. The method achieves a Matthews Correlation Coefficient (MCC) of 0.50 and an Area under a Receiver Operator Characteristic (AUC) of 0.86 in a 5-fold cross validation. The prediction models have been tested on a dataset of 42 RNA-binding proteins (RNAtestset42), which is non-homologous to RNAset82. The performance of the new method called PiRaNha outperforms all previous methods in the both the cross-validation and the test cases. PiRaNha is an accurate web-based prediction server for predicting specific RNA-binding residues from sequence. It enables users to upload sequences and to obtain information on potential RNA-binding residues as downloadable files. In addition, the server allows for the input of multiple sequences, thus predictions can be made for whole datasets.

2P-284 タンパク質の揺らぎを考慮したタンパク質間ドッキングシミュレーション

Protein-protein docking with protein flexibility from bioinformatic approach

Nobuyuki Uchikoga (1,2), Takatsugu Hirokawa (2). (1: JBIC, 2: CBRC, AIST)

Postdocking process is the bioinformatic subject of protein-protein docking analysis, which is combined with various methods for selecting near-native complex structures. We took protein flexibility into account for docking analysis, which is one of the essential event in process of protein-protein docking and can be divided into backbone and side-chains flexibilities. Backbone flexibility results in protein large conformational change, relating to shape complementarities, and side-chain is involved in local interactions. For this issue, bioinformatic approach is based on rigid-body algorithm and ensemble docking is applied. We, then, introduced novel technique of soft docking and comparison methods. Various complex structures of each pair of proteins were generated by ensemble docking for various structures in backbone level after trimmed side-chains. Side-chains were optimized after trimming process, referred to as soft docking, providing information of side-chain interaction. For comparing between decoys, instead of RMSD, we introduced interaction fingerprints (IFP), which is a bit sequence consisting of interacting side-chains. Because of independent of 3D structure data, IFP is convenient for comparing between molecules with large conformational change and for cluster analysis. We discuss about the results of cross docking analysis, comparison of native and non-native pairs of molecules of calmodulin as receptor, and CNG, CaMKK, and PMCA as ligands.

2P-285 ヒトタンパク質の構造ドメイン・天然変性領域への区分け

Molecular dichotomy of human proteins into structural domains and intrinsically disordered regions

Satoshi Fukuchi (1), Kazuo Hosoda (2) and Ken Nishikawa (2) (1: Center for Information Biology and DDBJ, Nat. Inst. Genetics; 2: Maebashi Inst. Tech.)

It has long been known that eukaryotic proteins have complex domain structures, and recently revealed that considerable number of eukaryotic proteins have the so-called intrinsically disordered (ID) regions, which occasionally extend more than a few hundred residues. Those proteins with long ID regions have been known to play crucial roles in the important cell activity such as signal transduction, transcriptional regulation etc. The structural domains homologous to those registered in PDB can be identified with highly sensitive homology searches such as PSI-BLAST and HMM, while ID regions can be predicted because of their biased amino acid composition in the sequence. We have constructed a pipeline to divide a protein molecule into structural domains and ID regions, which consists of the standard procedure of domain assignment, the ID-region prediction by the DISOPRED program, and a newly developed program (DICHOT) that discriminates domains from ID regions. The reliability was tested with human transcriptional factors, which are one of the typical proteins with long ID regions, showing a plausible result by referring some of the published domain-ID partitioning. All human proteins in the SwissProt database are loaded into the pipeline. The results indicate that 30% of total residues belong to ID regions, while 70% fall in structural domains including 20% of unknown structure. The dependence of ID-region fractions on the subcellular localization of proteins will also be presented

2P-286 βシグナルの特徴解析とそれを用いた新規ミトコンドリアβバレル型外膜タンパク質の探索

Searching for new mitochondrial β-barrel outer membrane proteins by refinement of the β-signal

Kenichiro Imai (1), M. Michel Gromiha (1) and Paul Horton (1). (1: Computational Biology Research Center, AIST)

The β-signal (kutik et al, 2008), which is the first specific sorting signal for mitochondrial β-barrel outer membrane proteins (MBOMP), was discovered by mutational analysis of some MBOMPs. The β-signal, consisting of the motif: P₀xGxxH₂xH₂ (P₀ = polar, G = Gly, H₂ = large hydrophobic residue), located near the C-terminal last β-strand. We attempted to investigate the refinement of the β-signal through statistical analysis of known five MBOMPs (Sam50, Porin, Tom40, Mdm10 and Mmm2) and their homologs. Multiple sequence alignment revealed the motif was conserved between known MBOMPs and their homologs and the pattern was refined to P₀xGH₂xH₂xH₂, which has a periodic of hydrophobic residues, which may reflect the repeat structure of β-strand. More

interestingly, the motif of Mmm2 is far from the C-terminal unlike the other four MBOMPs, and secondary structure and disorder prediction suggested Mmm2 has a β-rich region in N-terminal and the position of motif corresponds to the most C-terminal β-strand while the C-terminal region following the motif is mostly unstructured. The new β-signal provides the opportunity to search for new MBOMPs, we are trying to identify promising MBOM candidate from Eukaryotic proteins of Uniprot by using analysis of evolutionary conservation of the new β-signal, Gene Ontology and secondary structure prediction, and we have found at least one promising candidate at the present. If this candidate is identified as a MBOMP experimentally, it will be the sixth MBOMP ever discovered.

2P-287 Motion tree 法による蛋白質構造変化の階層的記述

Hierarchical description of protein conformational change with motion tree

Ryotaro Koike (1,2), Motonori Ota (1) and Akinori Kidera (3,4). (1: Graduate School of Information Science, Nagoya Univ; 2: BIRD, JST; 3: International Graduate School of Arts and Sciences, Yokohama City Univ; 4: Research Program for Computational Science, RIKEN)

Proteins are flexible molecules, and their structures often change upon external stimuli such as ligand binding. Since these protein motions are relevant to the protein functions, the way to describe the motions is the central issue to grasp the mechanism of the protein functions. We developed a procedure to illustrate the protein motions in a hierarchical and multi-resolution manner, where all the motions of loops, secondary structure elements and domains in a protein are displayed in a tree. We call this 'motion tree'. By applying our procedure to proteins, we obtained a new picture of protein motions that cannot be detected by the conventional descriptors based on the identification of rigid domains. The tree description of protein motions allows a straightforward extension to the comparison of protein motions using the algorithm for comparing trees. We performed an extensive comparison of protein motions to present the entire view of protein conformational changes.

2P-288 蛋白質相互作用サイトデータベースの構築

Development of database for protein-protein interaction sites

Takashi Ishida (1), Miho Higurashi (2), Kengo Kinoshita (1). (1: Institute of Medical Science, Univ of Tokyo, 2: Mitsui Knowledge Industry Co.)

Most of proteins show their functions by interacting other proteins. These days, large-scale experiments, such as yeast two-hybrid, provide large amount of information about which proteins interact with a target protein. However, they cannot provide the information about actual interaction sites in proteins. Such information about protein-protein interaction sites in atomic resolution is important for better understanding of protein interactions. Recently, we have a large amount of structural data in PDB thanks to the progress of several structural genomics projects. Thus, we have developed the database of protein-protein interaction sites based on the structure of protein complexes in PDB. Usual practice to identify the interaction site has been done using a representative complex in PDB. However, for the proteins with multiple partners, non-interaction sites identified by using a single complex structure is not enough, because some part of the non-binding sites may be involved in the interaction sites with another proteins. Therefore, the real interaction sites should be obtained by using all of the binding states in PDB. For the purpose, we have identified the binding sites by searching the same proteins with different binding sites in PDB at first, and then mapping the binding sites onto the query proteins. As a result, the database contains so much information for each protein that we have also developed a web-interface of this database with 3D graphics for more intuitive understanding of protein-protein interactions.

2P-289 タンパク質の金属結合部位に関する電子波動関数の解析

Analysis of electronic wave function in the metal-binding sites of proteins

Naoaki Hariu, Kentaro Fukushima and Minoru Sakurai. (Center for Biological Resources and Informatics, Tokyo Institute of Technology)

Structural genomics has revealed a huge number of protein structures, and currently in silico methods are required to determine their functions from their three dimensional structures. Usually, metals are important components of enzymes. However metals are often detached from their binding sites during protein crystallization. In such cases, the missing metals must be restored in the resultant X-ray structures. It is therefore desirable to develop in silico methods for predicting the location of metal-binding site. Here, to develop such a method, in the framework of quantum mechanics, we performed several benchmark calculations for representative Zn-binding enzymes. At first for the structures obtained from molecular dynamics simulation, we performed linear-scaling semi-empirical MO calculation (MOZYME) with a solvent model known as COSMO. For the canonical MOs obtained, we investigated the relationship between their locations and the metal-binding sites determined experimentally. Then, if the electron density of a given MO has a particularly large value (more than 0.2) on a special amino acid, we assumed that this MO is localized on that residue. As a result, it was found that such localized orbitals tend to be located near the HOMO/LUMO band gap in energy and to be located near the metal-binding sites in structure. This finding is great informative to predict protein functional sites.