1P254 Protein Data Bank に蓄積されたリガンドー蛋白質相互作用情報解析ツールの開発
○齊藤 美保子¹、宇都 昌志²、細見 祐輔¹、白井 剛¹
¹長浜バイオ大・生命情報、²三重大院・医科学専攻
Development of analyzing tools for protein-ligand interaction in Protein Data Bank
Mihoko Saito(1), Masashi Uto(2), Yusuke Hosomi(1), Tsuyoshi Shirai(1) (1:
Department of Bioscience, Nagahama Institute of bioscience and technology; 2:
Graduate School of medicine, Mie university)

The Protein Data Bank (PDB) is the richest source of information of molecular interactions in atomic detail. We have developed a computer system that can automatically generate a binding-motif database out of the whole data set of PDB. The developed system consists of three parts. The first is the system to compare the structures of non-protein molecules (ligands) found in the PDB by using a fast graph-match algorithm. By using this system, 7,500 ligands currently held in the PDB were classified into groups based on their forms. The second system superposes all the protein structures in the PDB that bind a certain group of ligand, and extracts the empirical rules of spatial distribution of protein atoms around each group of ligand. The empirical rules are then interpreted into a set of amino acid residues that represents the typical binding-motif for the ligand group. The last system compares the binding-motifs and ligand structures to detect possible analogous ligands for the specified binding-motifs/sites. This system can be used for a systematic analysis and prediction of protein-ligand interaction. We applied this system for the analysis of the interactions between proteins and five types of nucleotide. This system was able to define the binding-motifs, which were consistent with the previously proposed motifs and also contain several additional features, for these ligand. Also this system was used to find "bio-isosteres" that might replace a certain chemical group in the ligands with a comparable fitness to the binding-motifs.

1P256 機械学習による DNA 結合サイトの予測：方法の改良とベンチマーク
○Andrabi Munazah^{1,2}, Ahmad Shandar¹、水口 賢司^{1,2}、皿井 明倫³
¹National Institute of Biomedical Innovation, 7-6-8, Saito Asagi, Ibaraki-shi, Osaka 567-0085, Japan、²Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871 Japan、³Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka-shi, Fukuoka, 820-8502, Japan
Prediction of DNA-binding sites from sequence and structure using machine learning
Munazah Andrabi(1,2), Shandar Ahmad(1), Kenji Mizuguchi(1,2) and Akinori Sarai(3): National Institute of Biomedical Innovation, 7-6-8, Saito Asagi, Ibaraki-shi, Osaka 567-0085, Japan; 2: Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871 Japan; 3: Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka-shi, Fukuoka, 820-8502, Japan

Application of machine learning methods for predicting DNA-binding residues from their amino acid sequence and evolutionary information were first made by Ahmad et al. (2004) and Ahmad and Sarai (2005) resulting in sequence based prediction server DBS-PRED and the PSSM based prediction DBS-PSSM. Since then there have been a number of papers reporting similar approaches with improved performance. Most of these methods have shown improvements due to larger number of trainable parameters, more permissive definition of binding sites and inclusion of additional experimental and predicted descriptors. In some cases the claimed performance improvement is misleading as the data sets are redundant or unavailable and performance evaluation is biased-leading to poor validation. We have compiled blind non-redundant data sets of DNA-binding proteins from PDB by selecting proteins whose coordinates have been submitted after the publication of these servers and are evaluating the prediction performance of these web servers and methods. In addition, we report the improvements in our own methods, leading to better accuracy. Whereas our first models used linear neural networks with only two neighbor information, to justify the number of parameters with a relatively smaller data set available at that time, we have now expanded our data sets and obtained performance better than other servers. We also search for a suitable definition of binding site and see the impact of this definition on prediction performance.

1P255 タンパク質リガンド結合部位の構造比較方法の開発とタンパク質ー薬剤相互作用予測への応用
○葉袋 良一^{1,2}、松尾 洋^{1,2}、廣田 洋^{1,2}
¹横浜市立大、²理研 GSC
A New method for comparing the structures of protein ligand-binding sites and its application for predicting protein-drug interactions
Ryoichi Minai(1,2), Yo Matsuo (1,2) and Hiroshi Hirota (1,2). (1: Supramolecular Biology, International Graduate School of Arts and Sciences, Yokohama City University; 2: RIKEN Genomic Sciences Center)

Recent studies have been revealing many examples of unexpected off-target interactions of existing drugs. Some of these interactions may be cooperative with producing the desirable therapeutic effects, while others may lead to adverse side effects. We developed a new method for the structural alignment of atoms in the solvent-accessible surface of proteins using feature vectors, which reflect the local geometrical and physicochemical arrangements of atoms, and we used it to compare 48,347 potential ligand-binding regions of proteins from a non-redundant protein structure subset (nrPDB, provided by NCBI). We found 2,415 known ligand-binding region pairs whose structures were similar despite having different global folds. Of these, we found 119 region pairs that had similar ligands with similar binding modes. These proteins are good examples of convergent evolution. We also found some regions that were similar to known drug-binding sites, suggesting possible off-target sites for drug interactions (e.g., lovastatin, flurbiprofen, metyrapone, and salicylic acid). In these cases, the binding simulation for 63 drugs using AutoDock obtained 207 proteins in total with pKd >3.0 and r.m.s.d. ≤4.0Å. Additionally, we will also discuss the web database on the results of all-against-all calculation.

1P257 DNA 結合蛋白質における保存されたアミノ酸のネットワーク：ホットスポットの類似と予測への応用
○Ahmad Shandar¹, Keskin Ozlem¹, Nussinov Ruth³、皿井 明倫²
1aD10 ¹National Institute of Biomedical Innovation, Osaka, Japan、²Kyushu Institute of Technology, Fukuoka, Japan、³Koc University, Turkey; National Cancer Institute, Fredrick, USA
Clusters of conserved residue in DNA-binding proteins: Analogy with hot spots and implications to prediction
Shandar Ahmad (1), Ozlem Keskin (2), Ruth Nussinov (3), Akinori Sarai (4). (1. National Institute of Biomedical Innovation, Osaka; 2. Kyushu Institute of Technology, Iizuka, Fukuoka; 3. Koc University, Turkey; 4. NCI-Fredrick, USA

Structurally conserved, tightly packed residues forming geometric clusters have been widely studied - both experimentally and theoretically - and shown to be the key residues of protein structure, function and stability. They have been termed as hot spots due to their crucial role in the overall protein sequence and structure. There is enough experimental evidence of the occurrence of hot spots in proteins in general and protein-protein interfaces in particular. We have examined structural aspects of potential hot-spot residues in three-dimensional structures of DNA-binding proteins. We explore the packing density, conservation scores, solvent accessibility and other aspects of these residues and investigate their pattern of clustering, propensity to bind DNA and their role in predictability of DNA-binding sites on proteins. We observe that most clusters (networks) of conserved residues fall in the DNA-binding regions. All DNA binding proteins have at least one hot-spot cluster in contact with DNA. Idea of hot-spot residues with suitable solvent exposed cutoff can be used to identify some residues (24% sensitivity in reference to all binding residues) with high precision (89% specificity). The balance between specificity and sensitivity can be adjusted by changing the cutoff parameters. We will discuss possible functional roles of the clusters of conserved residues in DNA-binding proteins.