# KP201

# Development of application tools for protein-ligand interaction analyses in Protein Data Bank

Mihoko SAITO, Tsuyoshi SHIRAI

*Department of Bioscience, Nagahama Institute of Bioscience and Technology,*
1266 Tamura, Nagahama, Shiga 526-0829 JAPAN
*m_saito@nagahama-i-bio.ac.jp*

## 1. Introduction

The Protein Data Bank (PDB) is the representative database of biomolecular structures, and it contains the richest information of molecular interactions in atomic detail. The number of current entries in PDB has become 52,535, and they contain a total of 8,499 different types of non-protein molecule (ligand) as of Aug, 2008.

Various methods have been proposed to take advantage of the accumulated structural information in protein science or medicinal chemistry; however, several problems still remain. For example, the variety of molecules in PDB is highly biased reflecting the experimental feasibility, and the data of structure and intermolecular interaction are not classified systematically in the database.

We have developed a full-automatic classification system of Protein Data Bank, and a series of tools for analyzing protein-ligand interactions. This system (Structure -Interaction Relational Database; SIRD) processed entire PDB into the 2ry database, in which the protein structures were divided and classified into structural units. The ligand structures were also compared by using a fast graph-match algorithm, and a total of 124 types of atom-groups, which were frequently found in many different types of ligands were defined as fundamental fragments. These fragments were used as a unit for analysing the inter-molecular interaction in this system. The interaction data were extracted as fragment-protein complexes, and the redundancy in data were removed by referring to the structural classification in the processed 2ry database.

For a comprehensive presentation of the information in this 2ry database, two application tools have been developed The first one (SIRD-Motif) extracts structural motifs of interacting proteins, and the second one (SIRD-Isoster) collects bioisoster information from the database. Bioisoster is the equivalent atom groups of ligand that can occupy a same interaction site on protein. These tools were tested on the current PDB, and several characteristic features of protein-ligand interactions were analysed.

## 2. Method

1) The tool for interaction motifs (SIRD-Motif) superposed the fragment-protein complex structures in data set using the fragments as template. To evaluate the patterns of interaction, protein atoms were classified into 15 different classes according to their bond patterns and elements. The spatial distribution of protein atoms around the fragment (within 12.0 Å) was obtained, and the frequencies of atoms were rendered into Z scores. The high-score regions (top 5% positions) were

defined as ‘motif’. In this study, a total of 41 fundamental fragments, which were found in more than 10 different protein complexes, and had a little internal flexibility (single bond) were used.

2) The tool for bioisosteric information (SIRD-Isoster) sought for the complex structures, which were non-redundant (sequence identity >95%) and contained different types of ligands in the PDB. The structures of proteins and ligands were superposed, and different types of fragments that occupied an equivalent site of proteins were identified, then the bioisoteric information: fragment type, ligand type, and interacting amino acid, were recorded. In this study, seven types of fragment that consisted nucleotide ligands were used considering their biological importance and abundance of complex structures in the PDB.

## 3. Results and Discussion

1) Interaction motif

We tried to define the fragment-protein interaction motifs for 41 types of fundamental fragments, based on the empirical rules of spatial distribution of protein atoms around fragments. As a result, heterocyclic fragments like nucleotide bases were found to prefer distinct interaction motifs to each other. Also sugar fragments or aromatic fragments showed the motifs depending on their structural features.

We compared these interaction motifs with those derived from CSD, which were mainly composed of the crystal structures of small molecules. These comparisons demonstrated that the modes of protein-small molecule interactions were different from those between small-molecules in terms of stacking of aromatic groups or preference in hydrogen bonds patterns.

2) Bioisoster

The bioisosters for nucleotide fragments were explored in PDB with the developed tool (SIRD-Isoster), Nucleotide ligands were divided into seven types of fragment: five for bases (adenine, guanine, cytosine, thymine and uracil), one for ribose, and one for mono-phosphate. The data sets contained 633 protein complexes for adenine, 130 for guanine, 42 for cytosine, 39 for thymine, 96 for uracil, 446 for ribose, and 296 for phosphate fragments.

As a result, bioisoster-partners were found in 12% (adenine) ~ 38% (cytosine) of the complexes in the data set. Only one or two bioisoster fragments were detected for a binding site in most cases. A total of 182 bioisoster fragments from 77 types of adenine binding sites, 60 from 34 guanine sites, 30 from 16 cytosine sites, 38 from 22 thymine sites, 46 from 20 uracil sites, 205 from 136 ribose sites, and 84 from 66 phosphate sites were detected in the analyses.

The bioisosteric information was annotated by the interaction patterns. The detected interaction patterns showed a tendency that the isosters used similar key-interactions to each other.

**References**

1. Saito M, Go M, Shirai T. An empirical approach for detecting nucleotide-binding sites on proteins. *Protein Eng Des Sel.* **2006** Feb;19(2):67-75.