論文内容の要旨

論文題目

A knowledge-based method for predicting ligand binding sites and conformations using three-dimensional structure of protein

（既知タンパク質-リガンド複合体立体構造情報を利用した結合部位および結合様式予測手法の開発）

氏名　笠原　浩太

In recent years, structural databases of proteins have been rapidly growing, due to structural genomics projects. Revealing the nature of ligand recognition by proteins, based on the databases, is one of the main goals of structural biology. Toward this end, statistical, data-mining studies of the databases have revealed several structural patterns in protein-ligand interactions. There are two major methods to detect the patterns from the databases. The first treats the patterns in elemental units of interactions, such as inter-atomic and functional groups, that commonly appear in a variety of complexes. The second focuses on the structural patterns in larger units consisting of some residues, called “binding motifs”, that are conserved among particular types of proteins. Although information about these patterns has been adopted to predict ligand recognition, the prediction methods are based on only one of the two types of patterns. This is because the methods focusing on the patterns in larger units may overlook those defined by elemental units, and *vice versa*. In order to consider both patterns, the size of the unit of interactions has to be defined suitably.

This thesis is composed of two studies, considering both types of patterns. First, I defined molecular fragments as the units of interactions, and developed a new prediction method for binding sites and ligand conformations, based on information about the interactions between the fragments. Second, I performed data-mining analyses on the spatial distributions of the fragment-fragment interactions based on a Gaussian mixture model, to clarify the patterns that contributed to the predictions.

Evaluations of the prediction method revealed that it could accurately predict the binding sites and conformations of chemically diverse ligands. I found that predictions for nucleotide binding were primarily supported by information about interactions in the binding motifs. On the other hand, other diverse ligands were predicted mainly from the information about interactions in the elemental units observed among a wide range of complexes. In addition, as a result of the data-mining analyses, various patterns of interactions were defined as Gaussian distributions. Basically, the patterns observed only among particular types of proteins and/or ligands had a small variance in their Gaussian distribution. In total, the majority of the ligands were recognized by the patterns commonly observed among various complexes.

In conclusion, I successfully developed a prediction method based on the patterns of interactions. In addition, the data-mining study based on the Gaussian mixture model clarified the two patterns: that is, those commonly observed among various complexes, and those appearing only in particular types of proteins and/or ligands with strictly conserved structures. Both types of interaction patterns of interactions play crucial roles in the ligand recognition by proteins.