Exhaustive structural comparison of interaction interfaces of proteins

To understand the function of a protein, it is necessary to understand the interactions between the protein and other molecules. To understand the interactions, in turn, it is useful to compare the structures of interaction sites and to identify the structural patterns or motifs. It has been assumed for a long time that proteins with similar sequences share similar structures and functions. However, a wide variety of protein functions for a particular protein family forces us to recognize the importance of the diversity within the universality of protein families. Protein structure comparison is known to be a computationally hard problem so that most studies use a set of “representative” structures based on sequence similarities as well as some coarse-grained representations of structures. In order to examine the diversity of protein structures and interaction patterns, we needed to develop an extremely efficient and detailed method for structure comparison. The GIRAF method we have developed is the first to accomplish the truly exhaustive all-against-all comparison of all the interaction site structures at atomic resolution in the Protein Data Bank. Since the publication of the BIOPHYSICS paper in 2007, the method has been improved significantly and applied to exhaustive comparative studies of interaction interfaces of small ligands, proteins and nucleic acids, which have lead to the notion of the composite motif to annotate protein functions in terms of the differential combination of multiple structural motifs.