P15/16-221

Probabilistic Alignment for Protein Sequences and Structures

○Ryotaro Koike[1,2], Kengo Kinoshita[2], Akinori Kidera[2] ([1]Institute for Protein Research, Osaka University, [2]Graduate School of Integrated Science, Yokohama City University, Japan)

A number of equally optimal alignments inherently exist in the sequence and the structure comparisons between proteins. To represent those suboptimal alignments systematically, we developed a method to generate probabilistic alignments for sequences and structures, in which the correspondence between a pair of residues is evaluated in a probabilistic manner. The method uses the periodic boundary condition to avoid the entropy artifact favoring full-length matches. In the structure comparison, the environmental effects are incorporated by the mean-field approximation. We applied the method to the comparisons of two pairs of proteins with internal symmetry; the first is proteins of TIM-barrel fold and the second is those of β-trefoil fold. It was shown that the sequence and structure alignments are consistent with each other and that the alignments with the highest probability represent circular permutation.

P15/16-222

Long- and short-range interactions in native protein structures are consistent/minimally-frustrated in sequence space

○ Sanzo Miyazawa[1], Robert L. Jernigan[2] ([1]Faculty of Technology, Gunma University, [2]L. H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, USA, Japan)

We show that long- and short-range interactions in almost all protein native structures are actually consistent with each other for coarse-grained energy scales; specifically the long-range inter-residue contact energies and the short-range secondary structure energies based on peptide dihedral angles, which are potentials of mean force evaluated from residue distributions observed in protein structures. This consistency is observed at equilibrium in sequence space rather than in conformational space. Statistical ensembles of sequences are generated by exchanging residues for each of 797 protein structures. It is shown that adding the other category of interaction to either the short- or long-range interactions decreases the means and variances of those energies for essentially all protein structures, indicating that both interactions consistently work by more-or-less restricting sequence spaces available to one of the interactions. Evidence is provided that protein native sequences can be regarded approximately as samples from the statistical ensembles of sequences with these energy scales, and that all proteins have the same effective conformational temperature. Designing protein structures and sequences to be consistent and minimally-frustrated among the various interactions is a most effective way to increase protein stability and foldability. (*Proteins* 50:35-43, 2003)

Poster Sessions

P15/16-223

Prediction of Protein-Protein Interaction Sites using Residue Interface Propensity

○Naoshi Fukuhara, Takeshi Kawabata, Nobuhiro Go (Graduate School of Information Science, Nara Institute of Science and Technology, Japan)

Protein-protein interaction plays a key role in biological processes, and it is important to understand its mechanism. In this study, we tried to predict protein-protein interaction sites from the unbound structures without assuming interacting partner proteins. It is a realistic problem because deducing interacting partner protein is sometimes very difficult.From the statistics of the representative 3D structural data of complexes, the residue interface propensities were estimated. We found that hydrophobic or aromatic residues tend to have high propensity values. Next, we constructed a simple prediction algorithm by using this propensity. For each residue on the protein surface, an averaged propensity score is calculated using the structurally neighboring residues with ASA weighting. The residue is predicted as the interface when this score is over the threshold value.To evaluate the accuracy of our predictions, we prepared 41 hetero-complexes whose bound and unbound structure had already been solved. The interface sites were predicted for the unbound structure, and their prediction accuracies were evaluated using the corresponding bound structure. This test revealed that the correlation coefficient of observed and predicted sites was 0.197. From now on, we are going to improve prediction method by using other parameter except residue interface propensity.

P15/16-224

The novel algorithm for finding pockets on protein surface using small and large probe spheres

○Takeshi Kawabata[1], Nobuhiro Go[1,2] ([1]Grad. School of Information Science, Nara Institute of Science and Technology, [2]CCSE, Japan Atomic Energy Research Institute, Japan)

It has been important to predict putative binding sites from protein 3D structure data, with progress of the structural genomics project. The simplest way for the purpose is just to find "pocket" (concave, cleft)-shape regions on the protein surface by a pure geometric way. It is based on the well-known fact that small ligands like to bind "pocket" region, because of getting large binding energy and excluding waters which disturb catalytic activities. Many algorithms for finding pocket region have been already proposed, but each method has its own weakness in its definition and algorithm. We proposed the new definition for pockets, which is an improvement of previous definition of the PASS algorithm (Brady and Stouten, 2000). In our algorithm, small and large probes are placed on protein VdW surface, and pockets is defined as the small probes which does not overlap with any large probes. Probe spheres should be tangent to three protein atoms. Pocketness for each atom or residue is defined the volume of the remaining small probe spheres. Our definition needs only two arbitrary parameters : small and large radius for the probes. We tested several protein structures, and observed a correlation with our pocketness and actual binding sites. We believe that this algorithm provides good candidates of binding sites for successive docking methods with physicochemical score function.