05-Molecular Modelling and Design for Proteins and Drugs

05.01 – Protein Structure Prediction and Design

MS–05.01.01
SIMULATING ANTIGEN-ANTIBODY RECOGNITION
By J. Janin, J. Cherfils & S. Duquerroy, Laboratoire de Biologie Structurale, CNRS and Université Paris-Sud, 91405-Orsay, France

Antibody-lysozyme are reconstituted by rigid-body docking the antigen onto the combining site of the antibody. Simulated annealing using a crude energy function where the attractive component is proportional to the interface area, yields clusters of orientations in which steric fit between the two protein components is achieved over a large contact surface. In nearly all cases, the native complex is among the ones selected, and often near the top of the list. However, after submitting artificial complexes created by the docking procedure to conformational energy refinement with full atomic detail, we find many that form large interfaces with correct packing and electrostatic interactions. These criteria are therefore necessary, but not sufficient, in defining specific recognition.

MS–05.01.02 THE CORRELATION BETWEEN STRUCTURE FLEXIBILITY OF PEPTIDE FRAGMENTS IN PROTEINS AND THEIR SEQUENCE DISTRIBUTION IN DATABASE. Luhua Lai*, Xiaopeng Gong, Xiaojie Xu, Department of Chemistry, Peking University, Beijing 100871, CHINA

It was reported that some peptide fragments adopt a common conformation in different proteins, while others have different conformations in different proteins. Criteria for distinguishing conformational constant peptide fragments from conformational variable ones would enhance the development of protein structure prediction, since conformational constant peptide fragment could be used in constructing unknown protein structures. The structure flexibility of peptide fragments was correlated to the properties of their sequences by studying their sequence distribution in protein sequence database.

Normalized probabilities of tripeptides and tetrapeptides in protein sequence data base were calculated. In order to compare sequence probabilities with structure variances, the Protein Data Bank was searched for those tetrapeptides which occur more than three times. These tetrapeptides were classified according to the structure similarity of a same sequence in different proteins into: (1)conformational constant–showing common structure in different proteins; (2)conformational variable–adopting different structures each time it occurs in protein; (3)between (1) and (2). 21 out of 32 conformational constant tetrapeptides have sequence proabilities above 1.2, while 15 out of 28 conformational variable tetrapeptides have sequence probilities below 1.2. This shows some correlation between sequence probilities and structural variance. Hydrophobicity and other parameters were also included to test their improvement on the correlation.

MS–05.01.03 PROTEIN SECONDARY STRUCTURE PREDICATION USING THE 3D PROFILE METHOD.
By Kam Y. J. Zhang* and David Eisenberg, Molecular Biology Institute & Department of Chemistry and Biochemistry, University of California, Los Angeles, CA90024-1570, USA.

A new method for the prediction of protein secondary structures has been developed. The method uses the 3D profile (Bowie, Lüthy & Eisenberg, *Science*, 1991, 253, 164-170) to represent the three dimensional (3D) structure by the environments of each residue. Environmental properties include the buried area, fraction polarity and secondary structure. We have modified the 3D profile method to use a Fourier series to represent the residue preferences as a function of these properties of the environment. The advantage is that the environment is treated as a continuous function rather than as 18 discrete states.

We calculated 3D profiles for a representative set of 40 well refined non-homologous 3D structures selected from Protein Data Bank (Bernstein, *et al*, *J. Mol. Biol.*, 1977, 112, 535-542). These 3D profiles are separated into fragments in order to create a large ensemble of structures. A profile fragment library is created by dividing these 3D profiles into 10 residue long fragments with an overlap of 8 residues. A total of 4638 profile fragments were created from these 40 3D profiles.

A sequence from one these 40 structures is each in turn scanned against the library of 3D profile fragments for matches using dynamic programming algorithm, which enables insertion and deletions (Needleman & Wunsch, *J. Mol. Biol.*, 1970, 48, 443-453; Smith & Waterman, *J. Mol. Biol.*, 1981, 147, 195-197). Those profile fragments which correspond to the structure of the sequence for which we want to predict its secondary structure is deleted from the profile fragment library. Since all the structures used in our work are non-homologous, no structurally homologous structures were used in the secondary structure prediction. The match of a fragment(of known structure) indicates the corresponding region in the sequence may adopt the conformation of the profiled fragment. The score of the match indicates the likelihood that the region of sequence adopts the conformation of the profile fragment.