148

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.

 Copyright © 1993 International Union of Crystallography

05-Molecular Modelling and Design for Proteins and Drugs

After scanning the sequence against the library of 3D profile fragments, each residue in the given sequence normally has numerous matches with residues from different profile fragments with different scores. The score indicates the preference of this residue adopting the conformation of the matched residue in the profile fragment. We can make a scatter plot of the main chain conformation (ϕ and ψ angles) of all the matched residues from the profile fragments in the library. The number of residues which fall into each of the three regions (α-helix, β-strand or coil) are calculated, weighted by the scores matched by that fragment. The distribution of the three states for all the structures in the profile fragment library are also calculated. The preference of each residue in one of the three conformation states is calculated using a statistical inference algorithm. The secondary structure state of a residue is assigned to be the one with the highest preference. Tests using 40 structurally non-homologous structures indicate that the method has an overall prediction accuracy of 64%.

MS-05.01.04 DESIGN STRATEGY FOR PROTEIN STABILITY: STABLE LOCAL CONFORMATION IN CONSISTENCY WITH THE GLOBAL CONFORMATION. By K. Ishikawa, S. Kimura, K. Morikawa, S. Kanaya and H. Nakamura, Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan.

Various mechanisms of the protein stability have so far been studied, and several strategies to enhance the stability are now proposed. We have recently made lots of mutant proteins of ribonuclease HI from *Escherichia coli* (*E.coli* RNase HI, 17.6kDa), and studied the conformational stability and their crystal structures. As a result, several of the mutant proteins obtained the remarkable thermal stability, due to only very local amino acid replacements; For example, Lys95→Gly is considered to stabilize the local left-handed α-helical conformation by a Gly residue (S. Kimura *et al.*, *J. Biol. Chem.*, 1992, **267**, 22014-22017). His62→Pro may stabilize the short turn structure (K. Ishikawa *et al., Protein Eng.*, 1993, **6**, 85-91). Val74→Leu, fills the cavity in the hydrophobic core (K. Ishikawa *et al., Biochemistry,* 1993, in press). All three mechanisms are localized, and the characteristic features of individual amino acids contribute the increase of the thermal stability. Analyses of the crystal structures of the wild-type and mutant proteins of *E.coli* RNase HI less than 1.8Å resolution reveal that the global conformations of all those mutant proteins deviate very little from that of the wild-type protein. It means that those local structural changes can be permitted and even suitable for the original global conformation. The additivity of the mutations was confirmed (S. Kimura *et al., J. Biol. Chem.*, 1992, **267**, 21535-21542), and the structural analysis of the associated protein from *Thermus thermophilus* show that these local mechanisms are used in the thermophilic protein (K. Ishikawa *et al.*, *J. Mol. Biol.*, 1993, **230**, 529-542).

MS-05.01.05 RULE-BASED APPROACHES TO COMPARATIVE MODELLING by Z.Y. Zhu , M. S.Johnson, H. Wako, R. Sowdhamini, N. Srinivasan, K. Guruprasad, Z.Sun, B. Reddy, S. Rufino, Y. Edwards, T. Blundell*. Imperial Cancer Research Fund Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

Comparative modelling can be envisaged as two steps. The first is to solve the inverse folding problem: to define all those sequences that can adopt a particular fold. Operationally this is more usefully posed as defining whether a new sequence belongs to any of the known folds.It involves projecting restraints from a three-dimensional structure on to a one dimensional sequence. For this step we have calculated amino acid substitution tables in terms of local structural environmental parameters, which can be used to generate sequence templates for secondary structures, structural motifs and tertiary folds. The second step is to use the sequence, together with the knowledge that the protein belongs to a family of known fold, to construct a model. This form of protein modelling and prediction involves placing constrains from a known fold on a related protein sequence. The two steps require similar knowledge of the structures of protein families, and this knowledge can be expressed as rules that relate both local and global three-dimensional structure to patterns in the sequence of amino acids in a polypeptide chain. The method is comparative but exploits a broader knowledge-base of non-homologous protein structures.

PS-05.01.06 COMPARATIVE STRUCTURAL AND STEREOELECTRONICAL STUDY OF PINACIDIL, DIAZOXIDE AND CROMAKALIM, POTASSIUM CHANNEL OPENERS BELONGING TO THREE DIFFERENT CHEMICAL CLASSES. By L. Dupont*[a], B. Pirotte, P. de Tullio, B. Masereel, M. Schynts and J. Delarge[b].
[a]Cristallographie, Institut de Physique, [b]Chimie Pharmaceutique, Institut de Pharmacie, University of Liège, B-4000 Liège, Belgium.

Over the past few years the number of chemical agents with K^+ channel opening properties has greatly expanded. They are separated into distinct chemical classes typically exemplified by cromakalim (a benzopyran), pinacidil (a pyridylalkyl-cyanoguanidine), diazoxide (a benzothiadiazine), nicorandil (a pyridinic nitro compound), minoxidil sulfate (a pyrimidine derivative) and RP 49356 (a pyridinic thioformamide). For the three best studied K^+ channel openers, the rank order of potency for vascular smooth muscle relaxation was found to be cromakalim > pinacidil > diazoxide whereas for their activity on insulin secreting cells, the order was diazoxide > pinacidil > cromakalim (Newgreen, D.T., Bray, K.M., McHarg, A.D., Weston, A.H., Duty, S., Brown, B.S., Kay, P.B., Edwards, G., Longmore, J. and Southerton, J.S., Br. J. Pharmacol., 1990, 100, 605-613; Lebrun, P., Antoine, M.H., Devreux, V., Hermann, M., Herschuely, A., J. Pharmacol. Exp. Theor., 1990, 255, 948-954). Of particular interest is the intermediate position of pinacidil between diazoxide and cromakalim. Moreover, pinacidil could be regarded as a fairly good structural analog of diazoxide.
The present work tries to evaluate the level of structural analogy between these three classes of compounds by using crystallographic and infographic data.
A systematic search was performed with SYBYL (Tripos Associates Inc., St Louis, Missouri, USA) starting from the X-ray conformation of pinacidil optimized by the Tripos force field maximin2 energy minimizer. The analysis of the search process exhibits four interesting low energy conformations. The four selected geometries have been optimized using the semiempirical method AM1 (MOPAC 5.0) and have been compared in terms of total energy calculation : the lowest energy conformation is actually the one found in crystal. A comparative study was undertaken on conformations and stereoelectronical properties of pinacidil, diazoxide and cromakalim to highlight the similarities which could be related to