**PC-6**

## An Energy Potential and Alignment Method for Identifying Protein Sequence – Structure Compatibilities

Sanzo Miyazawa
miyazawa@smlab.sci.gunma-u.ac.jp
Faculty of Technology, Gunma University, Japan

We develop a method for sequence – structure alignments and examine how effectively simple potential functions previously developed can identify compatibilities between sequences and structures of proteins for database searches. The stabilities of structures are assumed here as a primary requirement for compatibilities between sequences and structures. The stabilities of conformations depend on not only their conformational energies but the whole ensemble of conformations. The amino acid composition dependencies of the latter are taken into account. The potential function consists of pairwise contact energies, repulsive packing potentials of residues for overly dense arrangement and short-range potentials for secondary structures, all of which were estimated from statistical preferences observed in known protein structures (Proteins, 34:49-68, 1999). In the preceding paper (Proteins, 36:357-369, 1999), it was shown that this simple potential function can distinguish native structures from alternate folds and also recognize native sequences from non-native sequences by threading sequences into other structures in all possible ways without gaps. Here, it is more thoroughly examined by allowing deletions and insertions in sequence - structure alignments (Protein Eng. 13:459-475, 2000).

Pairwise contact interactions in a sequence-structure alignment are evaluated in a mean field approximation on the basis of probabilities of site pairs to be aligned. To obtain the self-consistent values of alignment probabilities of site pairs, an iterative method is employed. Gap penalties are assumed to be proportional to the number of contacts at each residue position, and as a result gaps will be more frequently placed on protein surfaces than in cores. In addition to minimum energy alignments, we use probability alignments (Protein Eng. 8:999-1009, 1995) that are made by successively aligning site pairs in order by pairwise alignment probabilities and provide information of how reliable each aligned site pair is.

Results show that the present energy function and alignment method can detect well both folds compatible with a given sequence and, inversely, sequences compatible with a given fold, and yield mostly similar alignments for these two types of sequence and structure pairs. Probability alignments consisting of most reliable site pairs only can yield extremely small root mean square deviations, and including less reliable pairs increases the deviations. Also it is observed that secondary structure potentials are usefully complementary to yield improved alignments with this method. Remarkably, by this method some individual sequence-structure pairs are detected having only 5-20 % sequence identity.