S I - P 16

ESTIMATION OF THE AVERAGE ENERGY INCREMENT BY AN AMINO ACID EXCHANGE IN PROTEINS AND ITS USE TO EVALUATE A HOMOLOGY SCORE MATRIX

S. MIYAZAWA,, R. L. JERNIGAN*, NATIONAL INSTITUTE OF GENETICS, MISHIMA, SHIZUOKA 411, JAPAN; *NATIONAL CANCER INSTITUTE, NATIONAL INSTITUTES OF HEALTH, BETHESDA, MD 20892, USA.

The energy change caused by an amino acid exchange in protein structures has been estimated in average, and used to evaluate a score matirix[1], which is used to measure the similarity between protein sequences, In a statistical sense, each type of amino acid residue is found at the particular location in the three dimensional structure of proteins; non-polar residues are more often found in non-polar environment of protein core and polar residues on protein surface. Residues surrounding an amino acid in protein structures are specific to the type of the amino acid. We consider a typical or average protein which satisfies such statistical features observed in a large set of protein structures, and the average energy increment caused by an amino acid exchange in such an average protein. Miyazawa and Jernigan[2] estimated the effective interresidue contact energy of each type of amino acid pair for proteins in solution from 18192 residue-residue contacts observed in 42 globular proteins, and also compiled the types of residues and their average numbers in contact with each type of amino acid. In the present report, the average energy increment by an amino acid exchange in the average protein is evaluated for each of 20 by 20 types of amino acid residue pairs by using their[2] contact energies and statistical data of surrounding residues. The average energy increment by an amino acid exchange would be a good measure of the structural instability caused by an amino acid exchange, and therefore we assume that the probability of an amino acid exchange to be tolerable in the evolutional process is proportional to the Boltzmann factor of the average energy increment by the amino acid exchange. The substitution process of amino acids consists of two steps, mutation and selection at the DNA level and selection at the protein level. The process of codon substitution is assumed to be in equilibrium and for simplicity its rate at the DNA level is assumed to be proportional to the equilibrium frequency of codon. Then a transition matrix of codon substitution for a long time interval which corresponds to 250 PAM is generated from that for a short time interval, and the log of each element of the 250 PAM matrix divided by the amino acid frequency is calculated as a score matrix in the same way as the reference 1. The substitution matrix which is the transition matrix multiplied by the amino acid frequency is similar to the mutation probability matrix compiled by Dayhoff et al.; the correlation coefficient is about 0.73. The correlation coefficient between the present score matrix and Dayhoff's score matrix (MDM_{78})[1] calculated from the 250 PAM mutation probability matrix is about 0.55 for all off-diagonal but 0.82 when infrequent amino acids of met, trp, cys and tyr are excluded. This score matrix has demonstrated the same degree of detection power of sequence homology as the Dayhoff's score matrix[1] These results indicate that the average energy increments by amino acid exchange estimated here reflect the structural instability caused by the amino acid exchange.

References; (1) R. M. Schawartz and M. O. Dayhoff, Atlas of Protein Sequence and Structure, Vol. 5, Supplement 3 (1978) 353-358. (2) Sanzo Miyazawa and Robert L. Jernigan, Macromolecules, 18 (1985) 534-552.