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, Japan, [2]L. H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, USA)

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 we mean 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 native 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 native structures with the Metropolis method. 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 native structures, indicating that both interactions consistently work by more-or-less restricting sequence spaces available to one of the interactions. In addition to this consistency, independence of these interaction classes is also indicated by the fact that there are almost no correlations between them when equilibrated using both interactions and significant but small, positive correlations at equilibrium using only 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)