N3a

Self-consistent Estimation of Inter-residue Protein Contact Energies Based on an Equilibrium Mixture Approximation of Residues

Sanzo Miyazawa[1], Robert L. Jernigan[2]
[1]Gunma Univ., Fac. Tech., [2]NIH, NCI, LECB

Pairwise contact energies for 20 types of residues are estimated self-consistently from the actual observed frequencies of contacts with regression coefficients that are obtained by comparing "input" and predicted values with the Bethe approximation for the equilibrium mixtures of residues interacting. This is premised on the fact that correlations between the "input" and the predicted values are sufficiently high although the regression coefficients themselves can depend to some extent on protein structures as well as interaction strengths.

Residues in each native protein structure are shuffled to generate the equilibrium mixture of residues. The relative hydrophobic energies $\Delta e_{ir}(\equiv e_{ir} - e_{rr})$ and the intrinsic pairwise energies $\delta e_{ij}(\equiv e_{ij}+e_{rr}-e_{ir}-e_{rj})$ are predicted from the accumulated numbers of contacts over proteins with the Bethe approximation; e_{rr} is a collapse energy and cannot be evaluated for this system. Residue coordination numbers are optimized to obtain the best correlation between "input" and predicted values for the relative hydrophobic energies. Regression coefficients between the "input" and predicted values for both Δe_{ir} and δe_{ij} are calculated and used to obtain the better estimates of relative contact energies $\Delta e_{ij}(\equiv e_{ij} - e_{rr})$.

The contact energies self-consistently estimated this way indicate that the relative hydrophobic energies predicted with the Bethe approximation should be reduced by a factor of about 0.3 and the intrinsic pairwise energies by a factor of about 0.6. This equilibrium mixture approximation of residues for proteins is supported at least to the extent that the observed distribution of contacts can be approximated with a small relative error of only about 0.08 as an equilibrium mixture of residues, if many proteins were employed to collect more than 20,000 contacts. Including repulsive packing interactions and secondary structure interactions further reduces the relative errors. These new contact energies are demonstrated by threading to have improved their ability to discriminate native structures from other non-native folds. (This work will be published in Proteins (1998).)

---

[1] Kiryu, Gunma 376-8515, Tel: 0277-30-1940, Fax: 0277-40-1029,
e-mail: miyazawa@smlab.sci.gunma-u.ac.jp

[2] NIH, Room B-116, Bldg. 12B, MSC 5677, Bethesda, MD 20892-5677, U.S.A.,
Tel: 001-1-301-496-4781, Fax: 001-1-301-402-4724, e-mail: jernigan@structure.nci.nih.gov