112-T

Self-consistent Estimation of Inter-residue Protein Contact Energies Based on an Equilibrium Mixture Approximation of Residues S. Miyazawa[1], R. 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. Residue coordination numbers are optimized to obtain the best correlation between "input" and predicted values for the partition energies. The contact energies self-consistently estimated this way indicate that the partition 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. 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.
Reference: *Proteins*, 34:49-68 (1999)

113-S

FREE ENERGY TARGET FUNCTIONS FOR IMPROVED DISCRIMINATION OF PROTEIN DECOYS. Sheldon Dennis and Sandor Vajda. Boston University

Molecular mechanics energy functions are generally unable to distinguish between correct and misfolded protein conformations, and have been increasingly replaced by structure-based potentials. Here we discuss free energy potentials that combine a molecular mechanics energy function with empirical terms representing solvation and entropic effects. A key step in the calculation is the use of van der Waals normalization to reduce the sensitivity of the molecular mechanics energy to very small changes in the atomic coordinates [2]. The potential is applied to well-constructed protein decoys made available by Park and Levitt [3], and is shown to perform better than the purely empirical target functions applied to the same decoy set [3]. Successful discrimination generally requires both molecular mechanics and solvation/entropic terms. The molecular mechanics term becomes the dominant contribution to the free energy if the backbone is strained. The programs implementing the potential can be downloaded from the webpage http://bioinformatics.bu.edu/MERL/. **Acknowledgments:** This research has been supported by grants DBI-9630188 from the National Science Foundation and DE-F602-96ER62263 from the Department of Energy. References: 1. Vajda, S., Sippl, M., and Novotny, J. Empirical potentials and functions for protein folding and binding. Curr. Opinion Str. Biol. 1997, 7: 222 228. 2. Janardhan, A. and Vajda, S. Selecting near-native conformations in homology modeling: the role of molecular mechanics and solvation terms, Protein Science, 1998, 7:1772-1780. 3. Park, B. and Levitt, M. Energy functions that discriminate x-ray and near-native folds from well-constructed decoys. J. Mol. Biol. 1996. 258: 367-392.

114-M

Statistical analysis of amino acid pair correlation in transmembrane domains Alessandro Senes, Mark Gerstein and Donald M. Engelman. Dept. of Molecular Biophysics and Biochemistry, Yale University. New Haven. CT

We are interested in determining which patterns of amino acids occur with enhanced frequency in α-helical transmembrane domains relative to random expectation. To approach this question, we have analyzed the pairwise correlation of amino acids in putative transmembrane helices. Our analysis is based on the occurrence of each of the 400 possible pairs of amino acids, at separation registries ranging from 1 to 10 positions, using a homology-free database of transmembrane sequences annotated in the SwissProt protein database. The occurrences of each pair were compared with their expectancy calculated allowing all possible permutation of each single sequence. The most deviating pairs were ranked according to their distance from the expectancy mean expressed as a number of standard deviations. The number of significant cases exceeds the chance occurrence of deviating cases by several orders of magnitude. Within the most significant cases, pairs with spacing not exceeding one full turn of the helix are strongly prevalent. Among them, pairs with spacing 4 and 1 (which occur on the same face of an α-helix) are the most represented. Patterns can be recognized among groups of both overrepresented and underrepresented amino acids pairs. When residues of similar structure are substituted in the pairs, similar statistical biases are generally maintained. These patterns persist when the analysis is extended by the search for third and fourth residues that are associated with the pairs.

115-T

DOUBLET POSTIOTION SPECIFIC ANALYSIS OF ALPHA HELICES IN GLOBULAR PROTEINS. Z.Minuchehr, B.Goliaei. Institute of Biochemistry and Biophysics Tehran University, Tehran 13145-1384, Iran.

An analysis of the amino acid distributions at 10 positions N"-N', N'-Ncap, Ncap-N1, N1-N2, N2-N3, C3-C2, C2-C1, C1-Ccap, Ccap-C' and C'-C" in 286 different alpha helices revealed a unique characteristic in each position. The proteins were derived from the Brookhaven Protein Data Bank and different helices with more than 7 amino acids were tabulated in a file for further analysis. The expected frequency for all 400 doublets were calculated using IG suite package by creating 400 different keys and searching the whole Swiss Prot Data Bank for finding all possible matches. These data were used in order to calculate the Relative Preference Value for each doublet in all 10 positions. Our results showed that the doublets Asn-Pro and Asn-Asp were more frequent at the N'-Ncap positions and Phe, Asp, Ser and Pro were located more frequently with Ala at positions Ncap-N1 respectively. A striking result was that Gly was located at position C' with hydrophobic amino acids such as Val, Ile, Phe and Ala at position C" and His-Arg were mostly located at position C1-Ccap. We can therefore conclude that Asn is paired with acidic amino acids at N-Terminal and Gly with hydrophobic amino acids at C-Terminal.

116-S

Incorporating Hydrogen Bonds in Minimal Off-Lattice Models of α-Helices: Effect on Stability, Cooperativity and Kinetics D. K. Klimov, M. R. Betancourt, and D. Thirumalai. Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742

Using Langevin simulations we have examined the stability, cooperativity and kinetics of the helix-coil transition in minimal off-lattice model of 16-mer four turn α-helix, which explicitly includes hydrogen bonds (HBs). The characteristics of the native conformations of model helices coincide with those found in real proteins. The transition from coil to helix is quite broad and only weakly cooperative. However, the cooperativity measured by a dimensionless index Ω_c is enhanced when HBs are taken into account. The value of Ω_c for our model is consistent with that inferred from experiment for an alanine based helix. We show that the folding time τ_F ranges from 6 to 1000 ns within the temperature range $(0.7 - 1.9)T_F$, where T_F is the helix-coil transition temperature. These values are in excellent agreement with available experimental data. The temperature dependence of τ_F exhibits a nearly Arrhenius behavior. Thermally induced unfolding occurs on a time scale of $(40 - 170)ps$. We predict that τ_F sensitively depends on sequence composition. Our study shows that HBs not only affect the stability of α-helix but have a profound influence on the kinetics as well. The excellent agreement between simulation results and experiments implies that minimal off-lattice models can be successfully used to study various aspects of protein folding.

117-M

Root Mean Square Deviations of Computer Generated Structures Relative to Native Protein Structures Kaizhi Yue and Ken A. Dill. Department of Pharmaceutical Chemistry, Box 1204, University of California at San Francisco, California 94143

In a recent test of Geocore, an *ab initio* peptide folding algorithm, we studied 18 short molecules for which there are structures in the Protein DataBank (PDB) and found a high incidence of native-like structures within the best few hundred conformations generated by Geocore. Because of the short length of peptides (up to 31 monomers) and the use of disulphide bridge information in the search, we checked that the high incidence of native-like structures is not artifacts of the search method. The RMSD distributions of the generated conformations show that, despite the small radii of gyration, average RMSDs of small peptide conformations can be 7-8 Å for a chain of around 30 monomers. The data also show that, while disulphide bridges introduced strong constraints on peptide structures, the average RMSD of conformations generated with disulphide bond information is still around 1.6 Å from a native-like generated conformation. This implies that the Geocore energy function serves as a strong filter to reduce the population of good conformations by two or more orders of magnitude. Thus, native-like structures that are being found by Geocore are not due to some property of the constraints or the search, but are due to the energy function.