How many amino-acid sequences can fold to a given protein structure?

ArticleGrants	:	
ArticleContext	:	130591122

Rachel Brem

Abstract

Information about the set of amino-acid sequences that will fold to a given target protein structure can be predicted without assuming complete knowledge of the possible conformational space they can occupy.

Significance and context

Given a three-dimensional protein structure, the 'protein design problem' asks which amino-acid sequences will fold to it spontaneously. Zou and Saven have devised a theory for predicting the average amino-acid composition of the group of sequences that are guaranteed to fold to a given target structure. This theory is the first to make such predictions without assuming knowledge of all unfolded or partly folded conformations of every possible amino-acid sequence. This is an important development, as it is not possible to compute such an ensemble of conformations for a real protein.

Key results

Zou and Saven first assume that they can describe protein conformation in terms of two types of energies: first, the 'propensity' of an amino-acid type to be in a structural environment, for example helix, sheet, or coil; second, the pairwise interaction free energy, E, between two amino acids that are close in space. Next they define a quantity, δ, which for a given amino-acid sequence and a target structure is the difference between the E of the sequence when it folds into the target conformation (E_1, say) and the average E of the sequence in many other compact, but unfolded, conformations (E_2, say). The authors hypothesize that if δ for a given sequence and target is large and negative, the sequence is likely to fold stably into the target conformation. To compute E_2 over the ensemble of unfolded compact conformations, they use the following assumptions. For the propensity energy, the authors compute the average environment of a sequence position i in all unfolded conformations. Then for a given amino-acid type j placed in position i, they score j's propensity for this average environment. The propensity term in E_2 for a sequence (of fixed js) is then the sum of all such scores over all i. Next the authors compute the average pairwise interaction energy between an amino acid j in sequence position i and all other nearby residues, taken over all unfolded conformations and all sequences. The interaction energy

term in E_2 for a sequence (of fixed j values) is then the sum of all such scores over all i. Zou and Saven use these assumptions, in a new statistical mechanical formalism, to find information about the set of sequences that will fold to a given target with a given value of δ. In particular, their method finds the probability that, for the set of folding sequences, each position i will contain amino acid j. Zou and Saven test their theory on a single target structure in a three-dimensional lattice model of proteins, in which residues can only occupy certain positions in space and sequences are composed of only two amino-acid types which interact on a close-range basis. The authors calculate average sequence ensembles and related quantities for the target structure by exhaustive enumeration of all sequences and conformations, and compare this exact result with those from their analytic method. The results are almost identical.

Conclusions

Zou and Saven conclude that it is possible, in principle, to identify sets of amino-acid sequences that will fold to a target structure, given only a set of unfolded conformations. The method does not identify a specific folding sequence, just sets of probabilities.

Reporter's comments

Of the three hypotheses on which the theory rests, two are not tested. First, Zou and Saven assume that their designed lattice-model sequences will fold to the target conformation, but they never check to make sure. It will be important to do this check by exhaustively enumerating all conformations of the designed sequences, and seeing whether the target conformation is their lowest-energy state. The second untested hypothesis is that an ensemble of compact unfolded conformations is a good substitute for the entire set of unfolded states, which contains non-compact conformations too. Also, during their lattice test, Zou and Saven used the set of 'all' compact unfolded conformations. In real proteins one could never generate such a complete set. The authors will need to check the robustness of their method with respect to incomplete ensembles. Nevertheless, as the first theory to make predictions without knowing all possible conformations, this provides an exciting starting point for further study.

Table of links

JournalofMolecularBiology

References

1. Zou J, Saven JG: Statistical theory of combinatorial libraries of folding proteins: energetic discrimination of a target structure. J Mol Biol. 2000, 296: 281-294. 0022-2836