Failures of Inverse Folding and Threading with Gapped Alignment

Gordon M. Crippen
gcrippen@umich.edu

College of Pharmacy, University of Michigan
Ann Arbor, MI 48109, U.S.A.

Abstract

In order to calculate the tertiary structure of a protein from its amino acid sequence, the thermodynamic approach requires a potential function of sequence and conformation that has its global minimum at the native conformation for many different proteins. Here we study the behavior of such functions for the simplest model system that still has some of the features of the protein folding problem, namely two-dimensional square lattice chain configurations involving two residue types. First we show that even the given contact potential, which by definition is used to identify the folding sequences and their unique native conformations, is unable to always correctly select which sequences will fold to a given structure. Second, we demonstrate that the given contact potential is not always able to favor the native alignment of a native sequence on its own native conformation over other gapped alignments of different folding sequences onto that same conformation. Because of these shortcomings, even in this simple model system where all conformations and all native sequences are known and determined directly by the given potential, we must reexamine our expectations for empirical potentials used for inverse folding and gapped alignment on more realistic representations of proteins.