

Poster I-20

Effects of amino acid composition, finite size of proteins and sparse statistics on distance-dependent statistical pair potentials



Authors:

Dmitry Rykunov (*Albert Einstein College of Medicine*)

Andras Fiser (*Albert Einstein College of Medicine*)

Short Abstract: Statistical distance dependent pair potentials are frequently used in a variety of folding, threading and modeling studies of proteins. We explored the possible origin and extent of false positive signals in these types of potentials and a new definition for reference state, free of the observed systematic errors was suggested.

Long Abstract:

Statistical distance dependent pair potentials are frequently used in a variety of folding, threading and modeling studies of proteins. The applicability of these types of potentials is tightly connected to the reliability of statistical observations. We explored the possible origin and extent of false positive signals in statistical potentials by analyzing their distance dependence in a variety of randomized protein decoys. While on average potentials derived from such decoys are expected to equal zero at any distance, we demonstrate that systematic and significant distortions exist.

Our simulations on a variety of native and randomized protein models demonstrated the following: (i) Atom-based potentials contain significant distortions due to correlation among distances measured from atoms of one residue to atoms composing another residue. The false positive signal that originates from these artifactual correlations is dominating the statistical potentials. (ii) The small volume accessible for interacting pairs at both small and large pair separations (the latter due to the finite size of a protein) is responsible for errors in potentials at these distances. Routinely applied correction scheme for sparse statistics (Sippl 1990, JMB) cannot fully compensate for these errors. (iii) The diversity in the size of proteins used to derive potentials and the size-dependent change of average amino acid composition result in a systematic increase of potential values with increasing pair separation distance. (iv) Only at most ~80% of residue-based potential values are statistically significant at any distance range. (v) Use of shuffled decoys as reference state for deriving potentials allows room to significantly improve the performance of the potentials.