

Neural Net Representations of Empirical Protein Potentials

Tal Grossman
Theoretical Division and CNLS
MS B213
Los Alamos National Laboratory
Los Alamos, NM 87544, USA
email: tal@t13.lanl.gov

Robert Farber
Theoretical Division
Los Alamos National Laboratory
and The Santa Fe Institute
1399 Hyde Park Rd.
Santa Fe, NM 87501, USA

Alan Lapedes
Theoretical Division
Los Alamos National Laboratory
and The Santa Fe Institute
1399 Hyde Park Rd.
Santa Fe, NM 87501, USA

Abstract

Recently, there has been considerable interest in deriving and applying knowledge-based, empirical potential functions for proteins. These empirical potentials have been derived from the statistics of interacting, spatially neighboring residues, as may be obtained from databases of known protein crystal structures.

In this paper we employ neural networks to redefine empirical potential functions from the point of view of discrimination functions. This approach generalizes previous work, in which simple frequency counting statistics are used on a database of known protein structures. This generalization allows us to avoid restriction to strictly pairwise interactions. Instead of frequency counting to fix adjustable parameters, one now optimizes an objective function involving a neural network parameterized probability distribution.

We show how our method reduces to previous work in special situations, but also allows extensions to include orders of interaction beyond pairwise interaction. Given the close packing of proteins, steric interactions etc., the inclusion of higher order interactions is critical for developing an accurate potential. A key feature in the approach we advocate is the development of a representation to describe the spatial location of interacting residues that exist in a sphere of small fixed radius around each residue. This is a "shape representation" problem that has a natural solution for the interaction neighborhoods of protein residues. We demonstrate in a series of numerical experiments that the neural network approach improves discrimination over that obtained by previous methodologies limited to pair-wise interactions.

1 Introduction

Recently there has been considerable interest in knowledge-based, empirical potential functions for proteins. The idea of using a database of known protein structure/sequence pairs to derive an empirical potential describing the interaction between residues has a relatively long history. Some early work considered the frequency with which pairs of amino acids appeared within a certain "contact" distance of each other (Miyazawa & Jernigan 1985), and used a quasi-chemical approximation to relate this frequency to an approximate free-energy of interaction of a "gas" of residue