

A DATA-DRIVEN, SYSTEMATIC SEARCH ALGORITHM FOR STRUCTURE DETERMINATION OF DENATURED OR DISORDERED PROTEINS

Lincong Wang
Dartmouth Computer Science Department
Hanover, NH 03755, USA

Bruce Randall Donald
Dartmouth Computer Science Department, Dartmouth Chemistry Department
Dartmouth Department of Biological Sciences
Hanover, NH 03755, USA

Traditional algorithms for the structure determination of native proteins by solution nuclear magnetic resonance (NMR) spectroscopy require a large number of experimental restraints. These algorithms formulate the structure determination problem as the computation of a structure or a set of similar structures that best fit the restraints. However, for both laboratory-denatured and natively-disordered proteins, the number of restraints measured by the current NMR techniques is well below that required by traditional algorithms. Furthermore, there presumably exists a heterogeneous set of structures in either the denatured or disordered state. We present a data driven algorithm capable of computing a set of structures (ensemble) directly from sparse experimental restraints. For both denatured and disordered proteins, we formulate the structure determination problem as the computation of an ensemble of structures from the restraints. In this formulation, each experimental restraint is a distribution. Compared with previous algorithms, our algorithm can extract more structural information from the experimental data. In our algorithm, all the backbone conformations consistent with the data are computed by solving a series of low-degree monomials (yielding exact solutions in closed form) and systematic search with pruning. The algorithm has been successfully applied to determine the structural ensembles of two denatured proteins, acyl-coenzyme A binding protein (ACBP) and eglin C, using real experimental NMR data.