TEMPLATE-FREE PROTEIN STRUCTURE PREDICTION BY COARSE-GRAINED HIERARCHICAL FOLDING

By

Patrick Buck

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the degree of

DOCTOR OF PHILOSOPHY

Major Subject: Biology

The original of the complete thesis is on file
In the Rensselaer Polytechnic Institute Library

Examining Committee:

Dr. Chris Bystroff, Thesis Adviser

Dr. Angel Garcia, Member

Dr. George Makhatadze, Member

Dr. Mohammed Zaki, Member

Rensselaer Polytechnic Institute
Troy, New York
July, 2010
(For Graduation August 2010)

ABSTRACT

Predicting a protein structure from its amino acid sequence is a long-standing unsolved problem in computational biology. While significant advancements in structure prediction have been made, substantial slowing of progress in template-free structure prediction calls for new ideas and approaches. Based on knowledge that short sequence segments can initiate folding through structural preferences independent of their three-dimensional context in proteins, our efforts have focused on folding proteins from local to global, so called hierarchical folding.

The first and last studies report on the construction of a differentiable coarse-grained knowledge-based force field in which the energy terms are conditional on local sequence patterns. Carbon-alpha force field (CALF) builds sequence specific statistical potentials based on database frequencies for alpha-carbon virtual bond opening and dihedral angles, pair-wise contacts and hydrogen bond donor-acceptor pairs. Constant temperature simulations were performed to fold 27 peptides selected as putative folding initiation sites, each 12 residues in length, representing several different local structure motifs. To assess the adequacy of the energy function on non-local interactions, 11 full length native structures were relaxed using Brownian dynamics. A simple energy potential that folds proteins locally and stabilizes proteins globally may enable realistic modeling of folding pathways.

The second study addresses the need to reduce sampling time necessary to fold full length sequences. Local native structure is constrained using rigid body dynamics at all possible sequence windows for Chymotrypsin Inhibitor-2 and the SH3 domain and monitored for changes in the folding pathway and rate. Based on the results of previous studies, we attempt to show that locally constraining proteins can promote structural polarization which may reduce sampling time necessary to reach the folded state. Accurate local structure prediction, based on the results of folding putative initiation sites, makes this approach attractive as a way to encode a hierarchical structure prediction program.