

MASTERS FINAL ORAL EXAMINATION

**Tuesday, November 15th
9:00a.m. @ 223 Atanasoff**

Ce Zhang

Major Professor: Guang Song

Protein Wild-Type and Mutant Ensemble Database

Protein structures have been determined and deposited into Protein Data Bank at an increasing rate. In this work, we organize all the protein structures in the PDB and form a mutant structure database. The database groups the wild type and mutant structures of the same proteins together. A direct benefit of the database is thus the easy accessibility of the structure ensembles of these proteins. Such ensembles are known to be highly useful for representing the native states of proteins and for understanding their function. For each protein, mutants are sorted by the number of mutations and the location(s) of the mutations.

What distinguishes our work from other mutation database is that it is structure-based and includes all the existing structure of the PDB. Synchronization with the PDB will be maintained. As an application, we carry out an experimental structures-based statistical analysis of the effect of mutations. A key question we address in this work is: is it valid to use mutant structure (or variants from different species) to represent a native state sample of the wt? Our results indicate that mutations in general cause an undetectable structure changes to the protein but there are exceptions. This justifies the uses mutation structures as a valid samples of the conformation space of the wild-type for most proteins and most mutations.

IOWA STATE UNIVERSITY
Department of Computer Science