

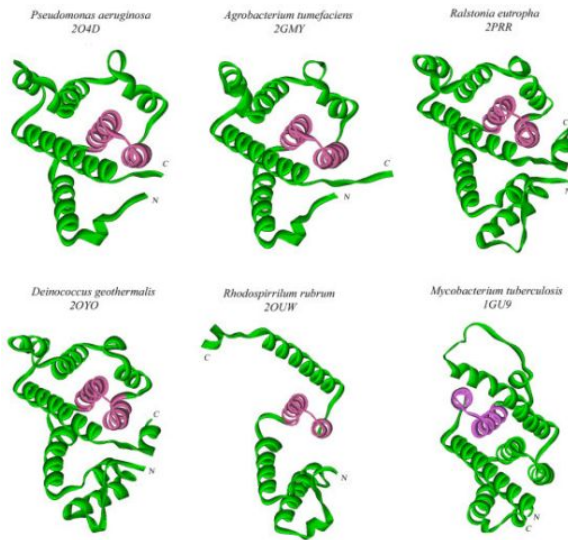
PFAMexplorer

A visual tool for the exploring correlation between
protein families and protein function

By Daniel Meyer and Matthew Carrington-Fair

Problem and Motivation

- Large databases consisting of information about proteins
- PFAM database creates “families” of proteins according to evolutionarily conserved regions, which often correspond to a structural motif or “domain”
- Hypothesis: some of these families are highly correlated to a specific function
- Motivation: if we find any of these linked to a function associated with a particular disease, we could more quickly find proteins involved in disease pathways that were previously unknown to be related to that disease



Dataset and Visualization

- Dataset is currently on a mysql server in Halligan, set up with a data model built in Python using sqlalchemy
- Goal is to build a visualization that clearly shows the relationships between all proteins of a family and their functions
- Improve upon existing visualization

