Genetic Analysis of the Metazoan and Fungal Heat Shock Protein Subfamily

Benjamin Cooper$^{1,2}$
Mentor: Dr. Hugh B. Nicholas Jr.$^3$

$^1$Bioengineering & Bioinformatics Summer Institute, Center for Computational Biology & Bioinformatics, University of Pittsburgh, Pittsburgh, PA 15261
$^2$The Pennsylvania State University, State College, PA 16802
$^3$Biomedical Initiative, Pittsburgh Supercomputing Center, Pittsburgh, PA 15213

Overview

- Sequence Retrieval
  - Choosing a Protein
  - iProclass
- Aligning Sequences
  - T-Coffee Multiple Sequence Alignment
  - MEME Conservation Motifs
  - GeneDoc
- Group Determination
  - TreeView
  - GEnt
  - Sequence Space
- Structural Analysis
  - PDB
  - ClustalW
  - RasMol
Choosing a Protein


- Heat Shock Proteins change expression levels over 100 fold during gastrulation

Gastrulation
Heat Shock Protein

iProclass
Aligning the Sequences

- T-Coffee Algorithm
- Yields highly accurate results
- Solves large problems but sacrifices time
  - T-Coffee: 36 hours
  - ClustalW: 6 minutes

MEME

- MEME uses Log Odds scoring using an amino acid similarity matrix
- Created 20 Motifs
Group Determination

- TreeView
- Sequence Space
- GEnt

TreeView
Sequence Space

- Uses a component of the PHYLIP software suite to display phylogenetic data based on the sequence alignment.
- Focuses on:
  - Substitutions
  - Deletions
  - Insertions
GEnt

- Runs on the bioinformatics cluster of the PSC
- Analyzes columns of the GeneDoc Alignment that are unbroken
- Displays them with the first letter corresponding to the letter that is found in the group and the second letter is the amino acid that is found in the overall Subfamily
- Shows the most distinguishing amino acids of a group

Gent First Run
Gent Final Groups

GEnt

GEnt Results
B3 Group
Structural Analysis

- No proteins in the two Kingdoms in the PDB
- Selected a PDB file from *Methanococcus Jannaschii* an Archaebacteria
- Imported into my sequence alignment and aligned it with the protein which had the closest amount of homology with it
- ClustalW has capability of a fast accurate alignment of two amino acid sequences

Visualization

- GeneDoc has a built-in RasMol script writer
- First made an image displaying the conservation throughout the Subfamily
Conservation

MEME Patterns
Important Residues (GEnt B3 group)

Combining the Sequences
The Future

- The findings provide data to be used in Molecular Modeling Simulations and potential wet-lab applications
- Do some 3-D analysis on the combined sequence
- Potentially combine the entire family into one alignment file

Acknowledgements

- Thanks to:
  - BBSI
  - NSF/NIH
  - Dr. Nicholas
  - Everyone in Room 220
  - Rajan Munshi