Comparison and Analysis of Heat Shock Proteins in Organisms of the Kingdom Viridiplantae

Emily Germain$^{1,2}$
Mentor – Dr. Hugh Nicholas$^3$

1 Bioengineering & Bioinformatics Summer Institute, Department of Computational Biology, University of Pittsburgh, Pittsburgh, PA 15261
2 Departments of Biomedical Engineering and Biology, Rensselaer Polytechnic Institute, Troy, NY 12180
3 Biomedical Initiative, Pittsburgh Supercomputing Center, Pittsburgh, PA 15213

Objectives

- Identify characteristics and motifs of protein family
- Determine residues essential to structure-function relationship
- Organize proteins into subfamilies
- Locate residues unique to particular subfamilies
- Make predictions regarding the protein’s evolution
Heat Shock Proteins

- Present in all living cells in cytoplasm and nuclei
- Transcriptionally upregulated when cell is stressed
  - Extremes of temperature
  - Toxins
  - Oxygen or nutrient deprivation
- Chaperone refolding of denatured proteins
- Transport other proteins within the cell
- Possible role in the immune response

Model Heat Shock Protein

*Methanococcus jannaschii*
Model Heat Shock Protein

Multiple Sequence Alignment

- 190 Viridiplantae HSP sequences extracted from iProClass sequence database
  - Remove fragments, 167 sequences remain
- Align sequences with T-Coffee
  - Perform global multiple alignment for all sequences
- Run MEME to locate motifs
  - 20 highly conserved patterns identified
- View T-Coffee and MEME results together and refine alignment by hand
  - Remove sequences not displaying multiple motifs, 161 sequences in final alignment
Residues Highly Conserved Over Family

MEME Patterns in Group HSP 17
**PHYLIP Bootstrap and Sequence Space Analyses**

- Input refined MSA into algorithms
- SeqSpace calculates clusters, defines similarity vectors from origin
- PHYLIP iterations created 1000 trees, compiled to create consensus tree
- Combined output of PHYLIP and SeqSpace used to define five subfamilies

**Phylogenetic Tree**
Sequence Space Output

Group Entropy

- Used PSC’s GEnt program
- Calculates the group entropy distance for each defined subfamily
- Gives a best fit match for sequences still ungrouped
- Residues with higher scores are unique to a particular subfamily and essential to its specific function

Group Entropy Distance = $S\left((p_i \cdot q) \times \log_2(p_i/q_i)\right)$

- $p_i$ – foreground residue frequency
- $q_i$ – background residue frequency
Group Entropy

High group entropy indicates conserved amino acid unique to subfamily.
Conclusions

- Evolutionary relationships suggest that different variations resulted from gene duplication.
- HSPs are more closely related to others in species similar to the one in which they are found, rather than to others of comparable molecular weights in more distantly related species.
- HSPs are highly conserved over the whole family, very specific residue alterations give particular subfamilies their individual properties.
- The data collected in this study can be further analyzed by comparing the highly conserved residues found in each group. This can be matched up with data regarding the specific functions of each heat shock protein to generate hypothesis regarding how these specific residues contribute to functional specificity and biochemical properties.
Resources


Acknowledgements

- Dr. Hugh Nicholas Jr.
- Pittsburgh Supercomputing Center
- Rajan Munshi
- National Institutes of Health
- National Science Foundation
- Everyone in BBSI