MINING THE EVOLUTIONARY DYNAMICS OF PROTEIN LOOP STRUCTURE AND ITS ROLE IN BIOLOGICAL FUNCTIONS

PI:
Dr. Gustavo Caetano-Anollés
Professor of Bioinformatics (Crop Science / IGB)
University of Illinois at Urbana-Champaign

Presented By:
Fizza Mughal
Graduate Student (Informatics)
University of Illinois at Urbana-Champaign
Objectives

- Flexible-unstructured regions of proteins introduce fundamental heterogeneity for molecular function
- Exploring dynamics of loops to ascertain their role in protein function
- Identify protein motions exclusive to specific functions
- Examine biophysical properties (flexibility and fluctuations) in the light of evolution

Source: Kruse, E., et al. 2006. Genome Biology, 7(2), 206
Source: http://www3.mpibpc.mpg.de/groups/de_groot/compbio1/p5/index.html
Protein Structure

• Levels
  • Primary
  • Secondary
  • Tertiary
  • Quaternary

• Domains: folded stable units

• Structural Classification Of Proteins (SCOP)
  • Fold Families: recent common ancestry
  • Fold Super Families: distant common ancestor
  • Folds: common structural topology

Source: http://en.wikipedia.org/wiki/Protein_structure
Protein Molecular Function

- Gene Ontology (GO):
  - Cellular Component
    - intracellular or extracellular
  - Molecular function
    - Binding or catalysis
  - Biological Process
    - Operations critical to functioning of living units

Protein Evolution

- Assumption:
  - Most abundant = most ancient

- Phylogenomic reconstruction
  - Characters
  - Taxa

1. FF Assignment
2. Genomic Abundance calculation
3. Character states defined (N=most ancient; 0=most recent) and polarized
4. Tree construction using PAUP* (maximum parsimony)
5. Age (node distance, $nd$) calculated (0=most ancient; 1=most recent)
Approach

• Dataset
  • Aminoacyl-tRNA synthetase (aaRS) domain FFs
  • ArchDB loop classification
  • Annotation with nd values
  • 87 Classifications
    • Density Search (DS)
    • Lowest p-value
    • Loop length >2 AA
    • Sec struc length ≥ 8 AA
    • Overall length < ~40 AA

• MD Simulations (NPT)
  • NAMD 2.9
  • CHARMM36

Why Blue Waters?

• Computing capability

• Storage of temporary files

• Impact: International Collaboration

• Key Challenge: Output File Storage
The Journey So Far …

- 73 Simulations performed
- Associated molecular functions
- Example: b.40.4.4 (MyF domain)
- Global Parameters:
  - RMSD
  - Radius of gyration

<table>
<thead>
<tr>
<th>Classification</th>
<th>FF</th>
<th>Loop ID</th>
<th>Loop Length</th>
<th>GO Term</th>
<th>Molecular Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS.BN.3.13.1</td>
<td>b.40.4.4</td>
<td>1JMZ_A_182</td>
<td>3</td>
<td>GO:0020037</td>
<td>Heme binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GO:0046872</td>
<td>Metal ion binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GO:0009055</td>
<td>Electron carrier activity</td>
</tr>
<tr>
<td>DS.BN.4.2.13</td>
<td>b.40.4.4</td>
<td>1T77_A_2080</td>
<td>4</td>
<td>GO:0004930</td>
<td>G-coupled receptor protein activity</td>
</tr>
<tr>
<td>DS.BN.5.2.2</td>
<td>b.40.4.4</td>
<td>1FJR_A_36</td>
<td>5</td>
<td>GO:0008800</td>
<td>Beta-lactamase activity</td>
</tr>
<tr>
<td>DS.BN.6.69.1</td>
<td>b.40.4.4</td>
<td>4MLL_B_208</td>
<td>6</td>
<td>GO:0005524</td>
<td>ATP binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GO:0008658</td>
<td>Penicillin binding</td>
</tr>
</tbody>
</table>
Conformational diversity (RMSD) vs. evolutionary age (nd)
Radius of Gyration vs. age (nd)
• 1JMZ
Conclusion & Future Directions

- Identification of fundamental principles of molecular evolution is achieved by reconstructing past events
- Advances in synthetic biology and translational medicine
- Methods to predict future “evolutionary trajectories”
  - predict evolvability of viruses
  - treatment of viral diseases with interfering agents (Wilke, 2012 PLoS computational biology)
- Map motions specific to classification/function based on molecular dynamic simulations
- Energy analysis
- Expand the data set!
Acknowledgements

• NCSA Blue Waters
• Illinois Research Board Grant
• Dr. Frauke Gräter (Germany)
• Evolutionary Bioinformatics Lab Members
THANK YOU!

Questions/Comments/Suggestions