How Function Shapes Dynamics in Evolution of Proteins

Gustavo Caetano-Anolles (PI)\textsuperscript{1}, \textbf{Fizza Mughal}\textsuperscript{1*} and Frauke Gräter\textsuperscript{2}

\textsuperscript{1}Evolutionary Bioinformatics Laboratory, Department of Crop Sciences, University of Illinois at Urbana-Champaign, 1101 W Peabody Drive, Urbana IL 61801. Phone: 217-333-8172. Email: gca@illinois.edu

\textsuperscript{2}Molecular Biomechanics, Heidelberg Institute for Theoretical Studies (HITS) gGmbH, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany. Phone: +49-6221-533267. Email: frauke.graeter@h-its.org
Relationship between protein dynamics and function

• Dynamics possess information that cannot be deciphered just by inspecting the structure.
• Proteins are found to be in various conformations as a dynamic ensemble (Orozco et. al, 2011).
• Evolution seems to constrain dynamics and flexibility (Micheletti, 2013), possibly to preserve function
  • Similar fluctuation in dynamics of Rossmann-like fold proteins (Keskin et al., 2000; Pang et al., 2005).
  • Common dynamics shared among members of the same protein family and superfamilies in comparison with unrelated/distant proteins (Maguid et.al, 2006; Maguid et.al 2008).
  • Common dynamics properties can be detected in proteins with different structure but common function (Ramanathan & Agarwal, 2011).

Tracing the evolutionary origins of protein function

Project activities using Blue Waters

• 203 Molecular Dynamics simulations
  • Aminoacyl tRNA synthetases
    • 87 loops – 10 ns each
  • Metaconsensus Enzymes
    • 116 loops – 50-70 ns each
  • Single domain proteins

Dynasome

- Construction of a dynamics space based on MD data
- Functionally similar proteins tend to cluster together.
- Dynamics space is continuous.
- Combining structure and dynamics information for prediction of function yielded better results than using either of the two alone.

Variables in the modified Dynasome

- nd values
- Mean RMSD
- Mean radius of gyration
- Eigenvalues of top 5 principal components
- Diameter
- Average path length
- Clustering coefficient
- Maximum modularity score
- Power law fit – alpha values
- Power law fit – KS statistic
- Bartels test statistic
- Mean betweenness values for N and C termini secondary structures as well as loop region
- Mean closeness values for N and C termini secondary structures as well as loop region
- Mean degree centrality values for N and C termini secondary structures as well as loop region
## GO Level 1 Term Description

<table>
<thead>
<tr>
<th>Description</th>
<th>Number of loops</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA binding TF activity</td>
<td>3</td>
</tr>
<tr>
<td>Binding</td>
<td>135</td>
</tr>
<tr>
<td>Catalytic activity</td>
<td>76</td>
</tr>
<tr>
<td>Electron carrier activity</td>
<td>1</td>
</tr>
<tr>
<td>Molecular function regulator</td>
<td>2</td>
</tr>
<tr>
<td>Signal transducer activity</td>
<td>3</td>
</tr>
</tbody>
</table>

### Variables in the modified Dynasome

![Residue Cross Correlation](image)

Visualizing the design space of dynamics community networks


Adapted from: https://en.wikipedia.org/wiki/Zooko's_triangle
Qualitative Space of Dynamics Community Networks
Future directions

• Clustering of all simulations based on the dynasome variables
• Devising methods for classifying community structure patterns
• Construct a “structure-evolution” space to complement the dynamics space
• Sampling more protein loops representative of the GO level 1 classes not covered in our current experiments.
• Longer timescale simulations are pertinent in capturing “slower” dynamics.
Acknowledgments

• NCSA Blue Waters
• Dr. Frauke Gräter (Germany)
• Evolutionary Bioinformatics Lab Members