Dissecting the relationship between protein structure and sequence variation

Amir Shahmoradi

Department of Physics
Institute for Cellular and Molecular Biology
University of Texas at Austin

APS March Meeting 2015
Why are some sites in proteins more variable than others?

Influenza virus hemagglutinin

```plaintext
... GTIKSWDESYIELKVEVP ...
... GTIKSWDESYTELKVDVP ...
... GTIKSWDENYTELKVDEVP ...
... GTIKSWDENYAEELKVDEVP ...
... GTIKSWDENSYTELKVDEVP ...
... GTIKSWDENSYTELKVDEVP ...
```
What are the best structural predictors of protein's sequence evolution?

Influenza virus hemagglutinin

... GTIKSWDESYTELKVDVP ...
... GTIKSWDESYTELKVDVP ...
... GTIKSWDESYTELKVDVP ...
... GTIKSWDESYTELKVDVP ...
... GTIKSWDESYTELKVDVP ...
... GTIKSWDESYTELKVDVP ...

Space domain

Time domain

The Past  The Present  The Future
Amino acid sites in the core of proteins have lower solvent accessible surface area

Influenza virus hemagglutinin
Buried residues (sites) evolve more slowly than exposed residues

Adjacent Averaging over 77150 Amino Acid sites in 213 globular proteins
Amino acid substitutions in the core are disruptive
Amino acid substitutions in the core are disruptive
Proteins are dynamic 3D entities

Influenza virus hemagglutinin
More flexible residues tend to evolve more rapidly

Adjacent Averaging over 77150 Amino Acid sites in 213 globular proteins
Voronoi tessellation as an unbiased parameter-free measure of site flexibility
Voronoi cell volume & area as the best predictors of sequence evolution
Site flexibility influences sequence evolution up to a certain threshold
Site flexibility influences sequence evolution up to a certain threshold
Site flexibility influences sequence evolution up to a certain threshold.
Voronoi tessellation as a parameter-free measure of Local Packing Density
The traditional definitions of local packing density involve adjustable parameters.
There is degeneracy in the definition of the Local Packing Density.
What set of atomic coordinates best represent individual amino acids in proteins?
Side Chain center-of-mass coordinates are the best representation of Protein 3-dimensional structure.

Moving away from side chain coordinates.
Side Chain center-of-mass coordinates are the best representation of Protein 3-dimensional structure.
Average side chain B factor is best representative of residue local flexibility in protein.
Side Chain center-of-mass coordinates are the best representation of Protein 3-dimensional structure for LPD calculation.
Other cell characteristics appear to have minor independent contributions.
Sequence divergence & H-bond homogeneity as the main determinants of sequence-structure correlation

<table>
<thead>
<tr>
<th></th>
<th>r.r4s.wcn</th>
<th>sd.se</th>
<th>sd.r4s</th>
<th>sd.hbe</th>
<th>mn.helix</th>
<th>mn.betas</th>
</tr>
</thead>
<tbody>
<tr>
<td>r.r4s</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd.se</td>
<td>0.62</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd.r4s</td>
<td>-0.25</td>
<td>-0.02</td>
<td>-0.1</td>
<td>-0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd.hbe</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.75</td>
<td></td>
<td>-0.84</td>
<td></td>
</tr>
<tr>
<td>mn.helix</td>
<td>-0.06</td>
<td>-0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mn.betas</td>
<td>-0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sequence divergence appears to be the primary determinant of evolution-structure correlations.
Summary

- Voronoi tessellation of protein structure can provide an unbiased measure of site-specific flexibility and a parameter-free measure local packing density.

- Do not use CA atomic coordinates.
  There is a better choice: the center-of-mass of the side chains.

- Two primary factors affecting sequence-structure correlation strengths:
  - Sequence divergence
  - Homogeneity of hydrogen-bond strengths in the entire protein structure

- https://github.com/shahmoradi/cordiv
- amir@physics.utexas.edu
Proteins exhibit a diverse range of sequence-structure correlations.
There is also degeneracy in the definition of the structural properties; consider Contact Number as an example.
Sequence divergence appears to be the primary determinant of evolution-structure correlations.
There is also degeneracy in the definition of the structural properties consider Contact Number as an example.
There is also degeneracy in the definition of the structural properties. Consider Contact Number as an example.
There is also degeneracy in the definition of the structural properties. Consider Contact Number as an example.
There is also degeneracy in the definition of the structural properties. Consider Contact Number as an example.
There is also degeneracy in the definition of the structural properties. Consider Contact Number as an example.
There is also degeneracy in the definition of the structural properties. Consider Contact Number as an example.
Sequence entropy appears to be determined more by local interactions, compared to B factors.
References


@article{shahmoradi2014predicting,
    title={Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design},
    author={Shahmoradi, Amir and Sydykova, Dariya K and Spielman, Stephanie J and Jackson, Eleisha L and Dawson, Eric T and Meyer, Austin G and Wilke, Claus O},
    journal={Journal of molecular evolution},
    volume={79},
    number={3-4},
    pages={130--142},
    year={2014},
    publisher={Springer US}
}
@article{shahmoradi2014predicting,
  title={Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design},
  author={Shahmoradi, Amir and Sydykova, Dariya K and Spielman, Stephanie J and Jackson, Eleisha L and Dawson, Eric T and Meyer, Austin G and Wilke, Claus O},
  journal={Journal of molecular evolution},
  volume={79},
  number={3-4},
  pages={130--142},
  year={2014},
  publisher={Springer US}
}
@article{jackson2016intermediate,
    title={Intermediate divergence levels maximize the strength of structure--
    sequence correlations in enzymes and viral proteins},
    author={Jackson, Eleisha L and Shahmoradi, Amir and Spielman, Stephanie J
    and Jack, Benjamin R and Wilke, Claus O},
    journal={Protein Science},
    volume={25},
    number={7},
    pages={1341--1353},
    year={2016}
}
@article{shahmoradi2016dissecting,
  title={Dissecting the roles of local packing density and longer-range effects in protein sequence evolution},
  author={Shahmoradi, Amir and Wilke, Claus O},
  journal={Proteins: Structure, Function, and Bioinformatics},
  volume={84},
  number={6},
  pages={841--854},
  year={2016}
}