

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

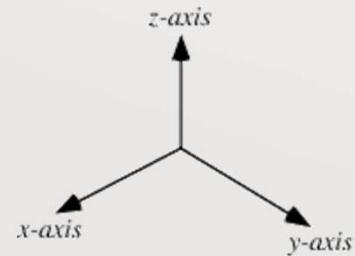
Why are some sites in proteins more variable than others?

Influenza virus hemagglutinin

```
... GTIKSWDESYIELKVEVP ...  
... GTIKSWDESYTELKVDVP ...  
... GTIKSWDENYTELKVDVP ...  
... GTIKSWDENYAELKVDVP ...  
... GTIKSWDESYTELKVEVP ...  
... GTIKSWDESYTELKVDVP ...  
... GTIKSWDESYTELKVDVP ...
```

What are the best structural predictors of protein's sequence evolution?

Influenza virus hemagglutinin



Space domain

... GTIKSWDES**Y**I**E**LKV**E**VP ...
... GTIKSWDES**Y**T**E**LKV**D**VP ...
... GTIKSWDE**N****Y**T**E**LKV**D**VP ...
... GTIKSWDE**N****Y**A**E**LKV**D**VP ...
... GTIKSWDES**Y**T**E**LKV**E**VP ...
... GTIKSWDES**Y**T**E**LKV**D**VP ...
... GTIKSWDES**Y**T**E**LKV**D**VP ...

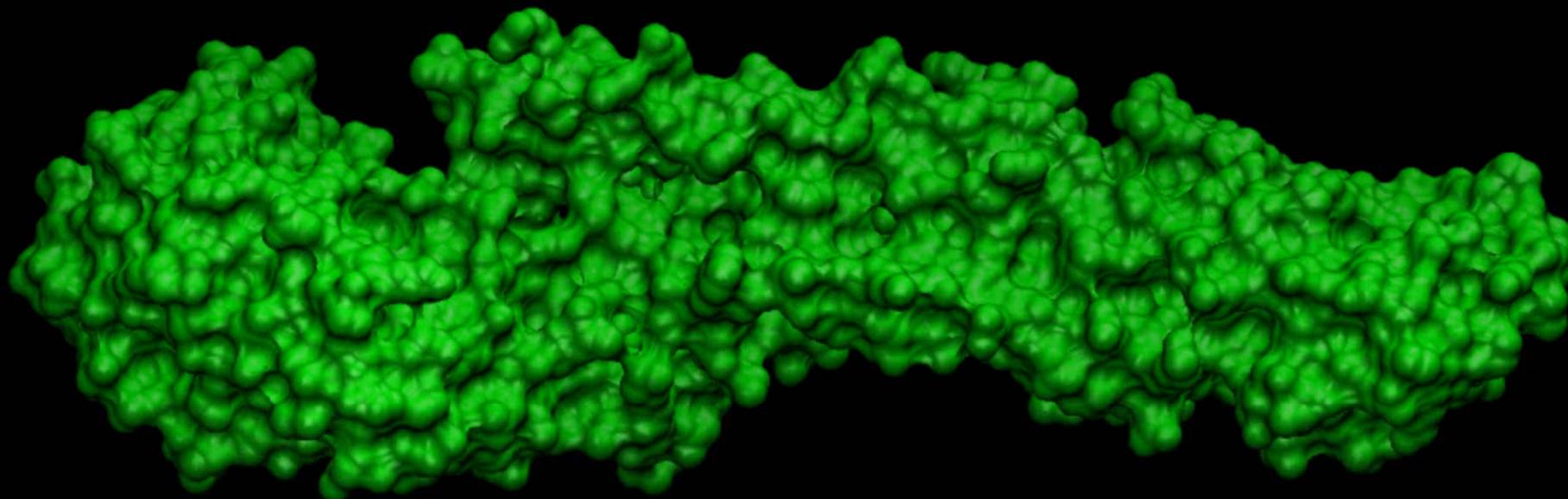
The Past The Present The Future



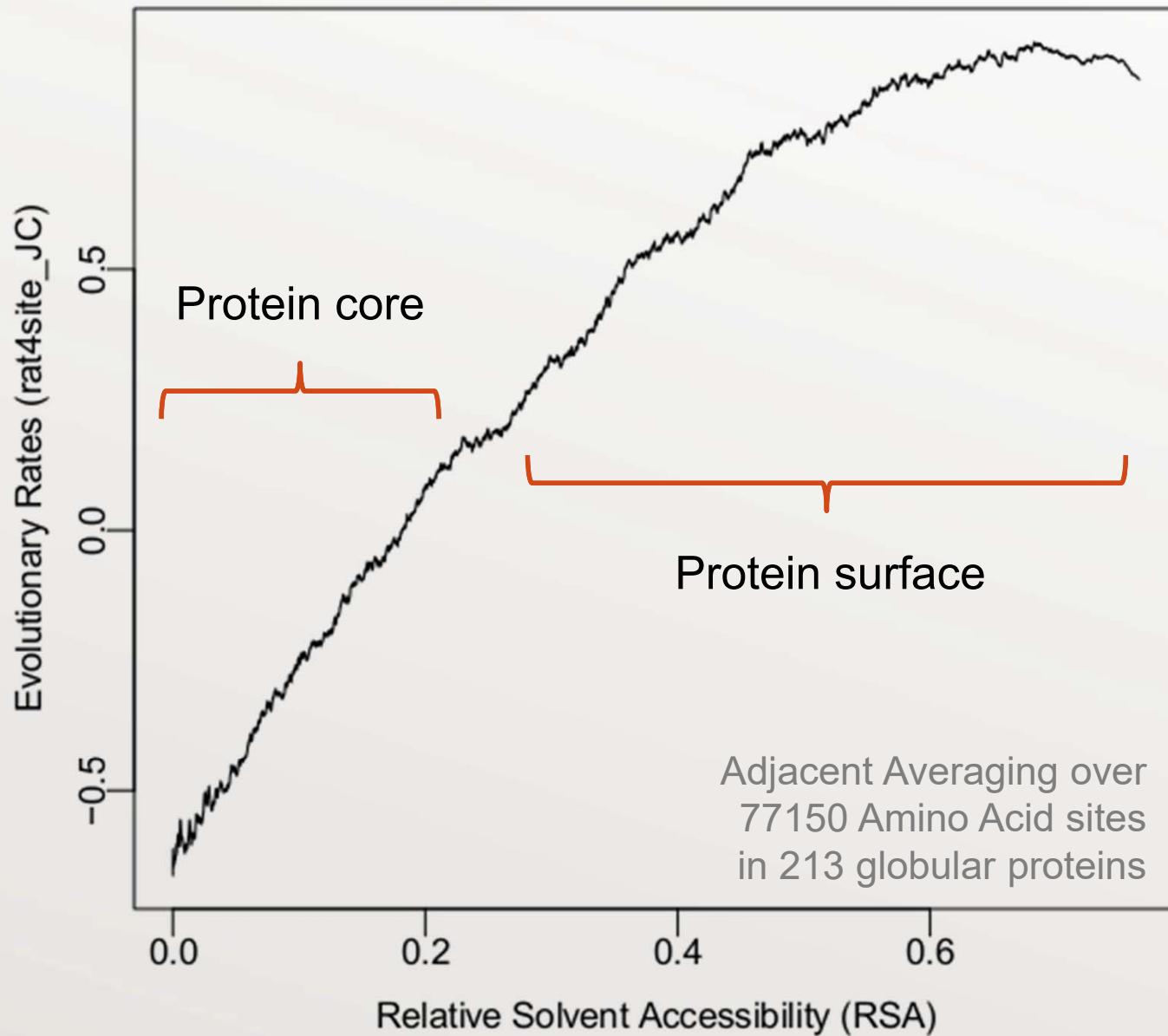
Time domain

Amino acid sites in the core of proteins have lower solvent accessible surface area

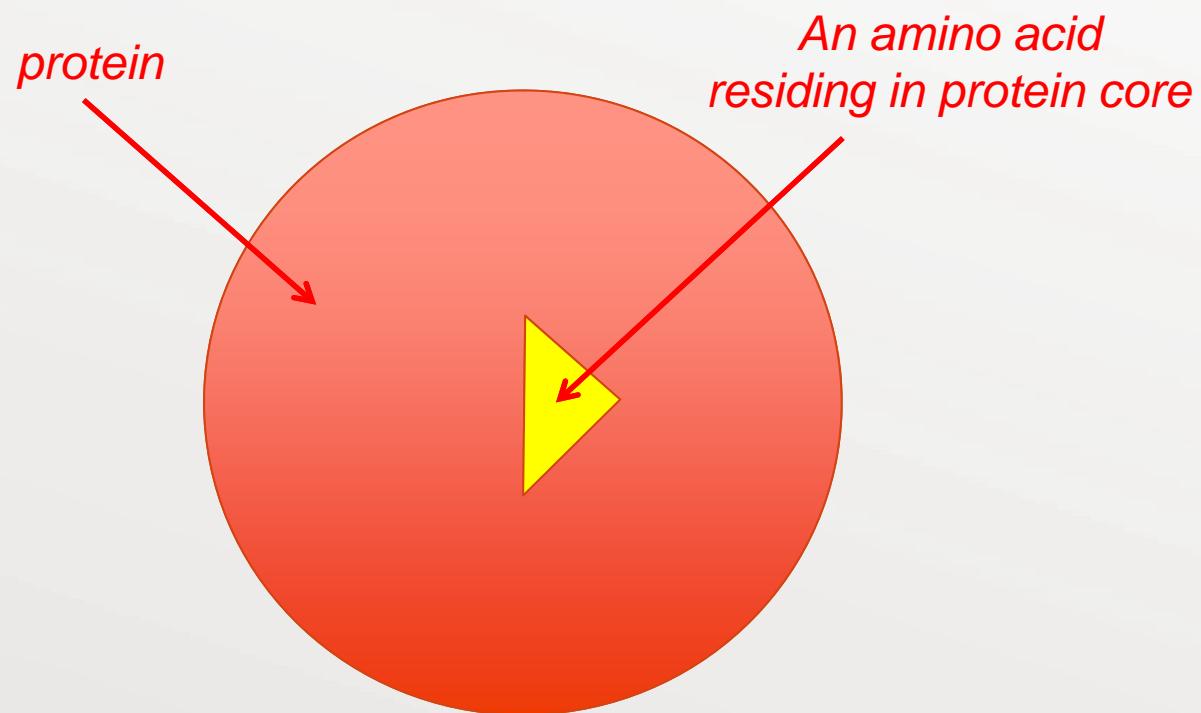
Influenza virus hemagglutinin



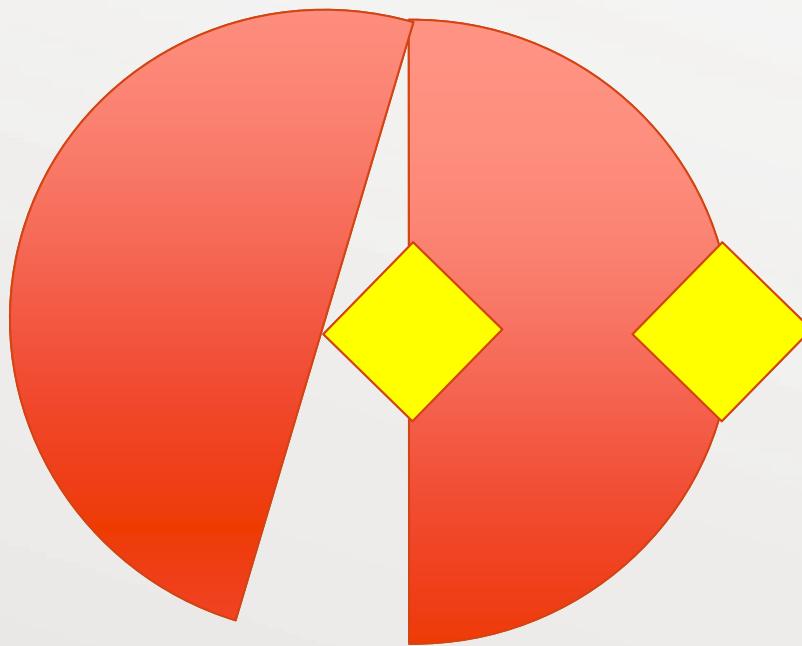
Buried residues (sites) evolve more slowly than exposed residues



Amino acid substitutions in the core are disruptive

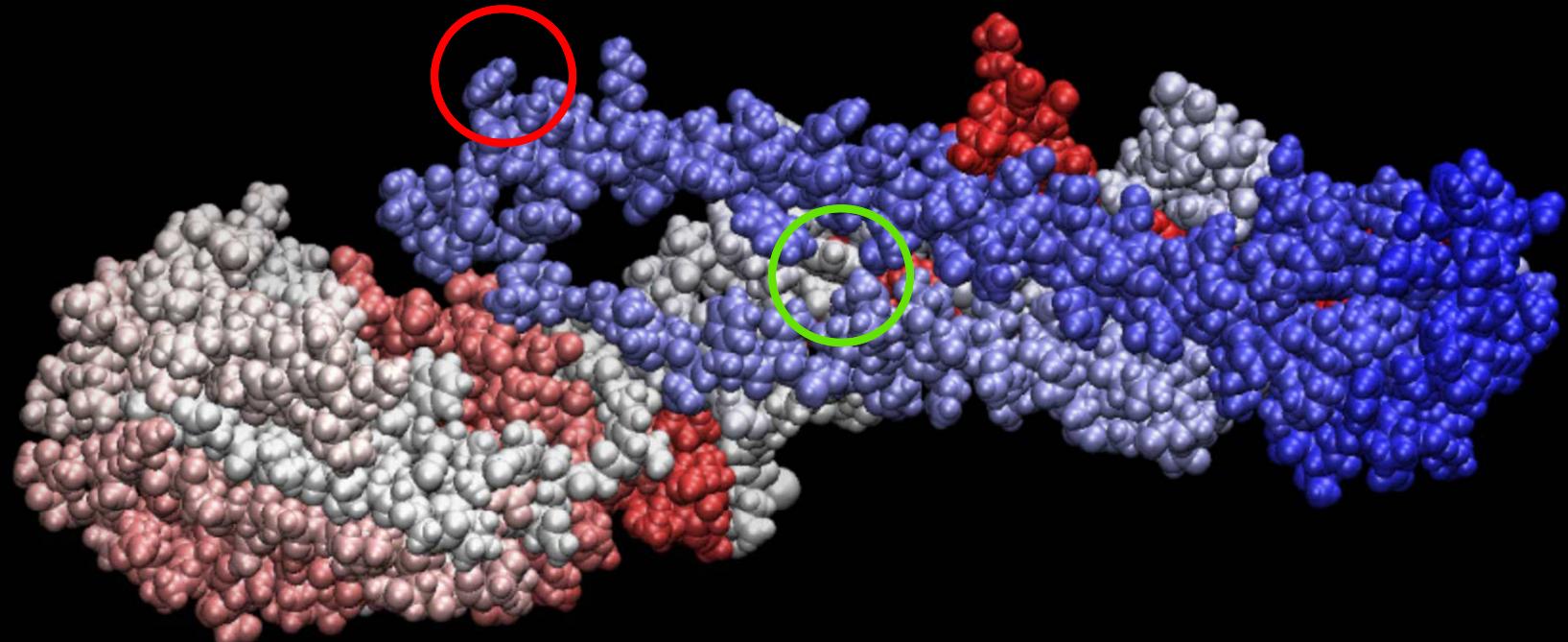


Amino acid substitutions in the core are disruptive

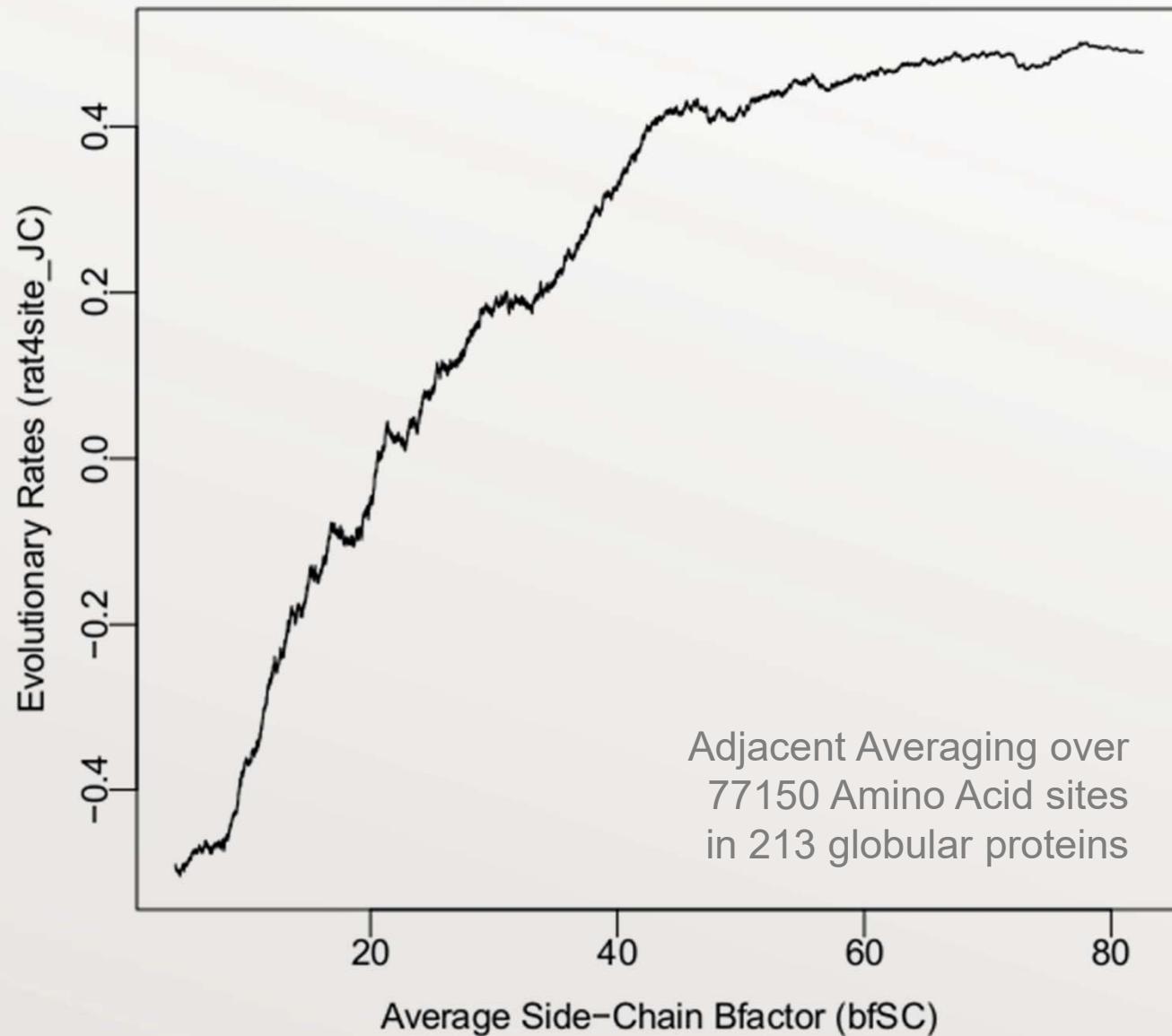


Proteins are dynamic 3D entities

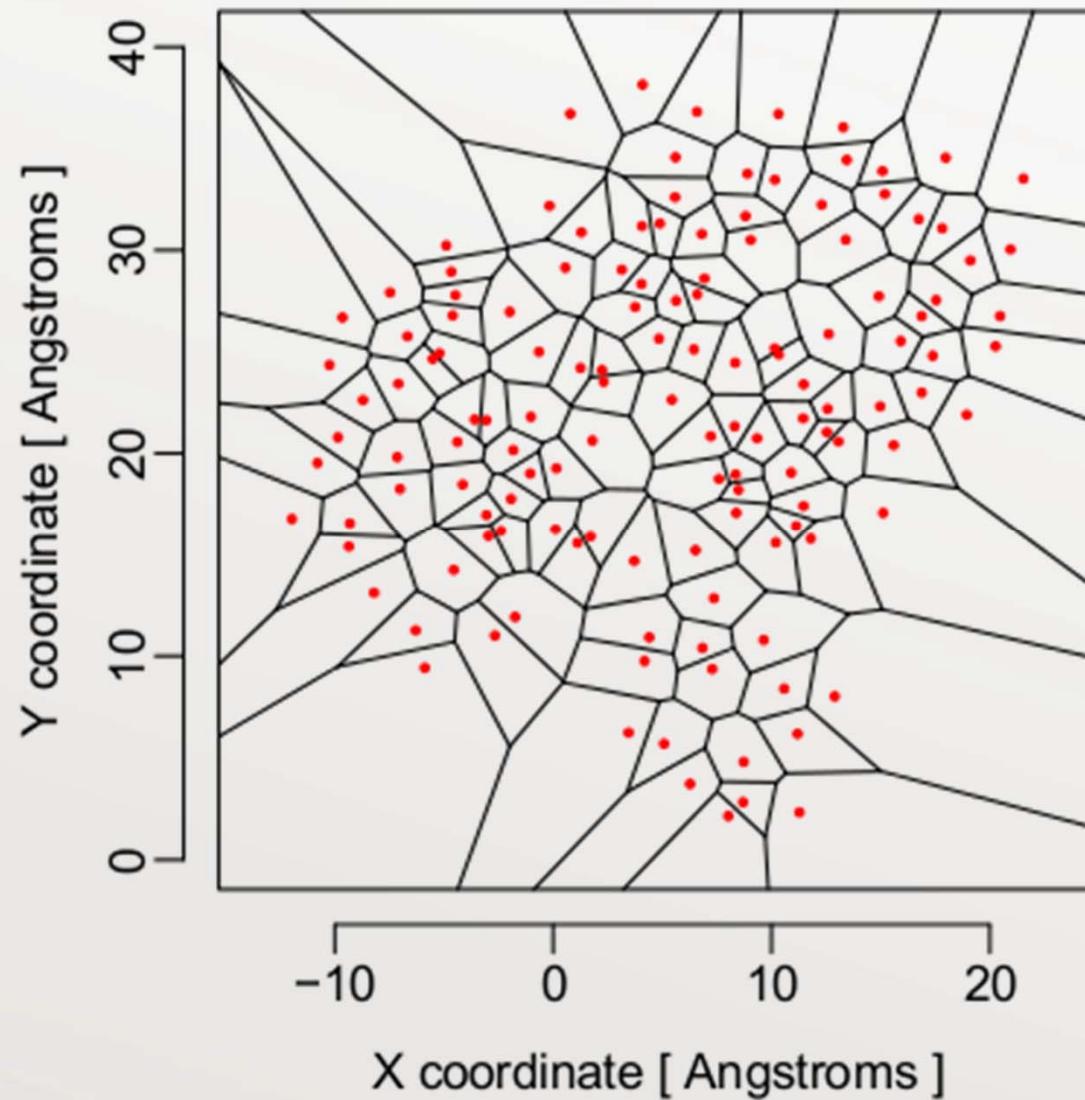
Influenza virus hemagglutinin



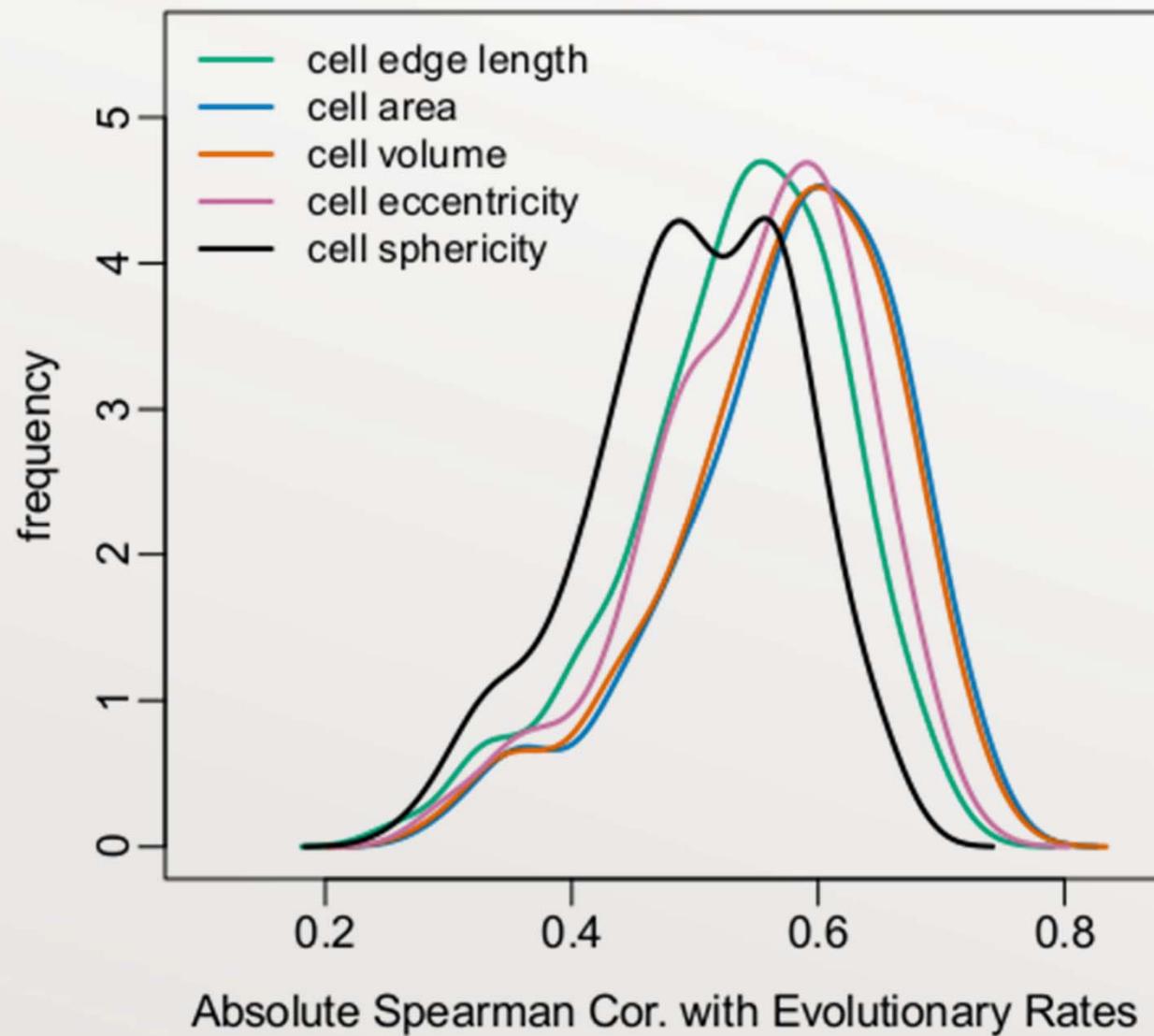
More flexible residues tend to evolve more rapidly



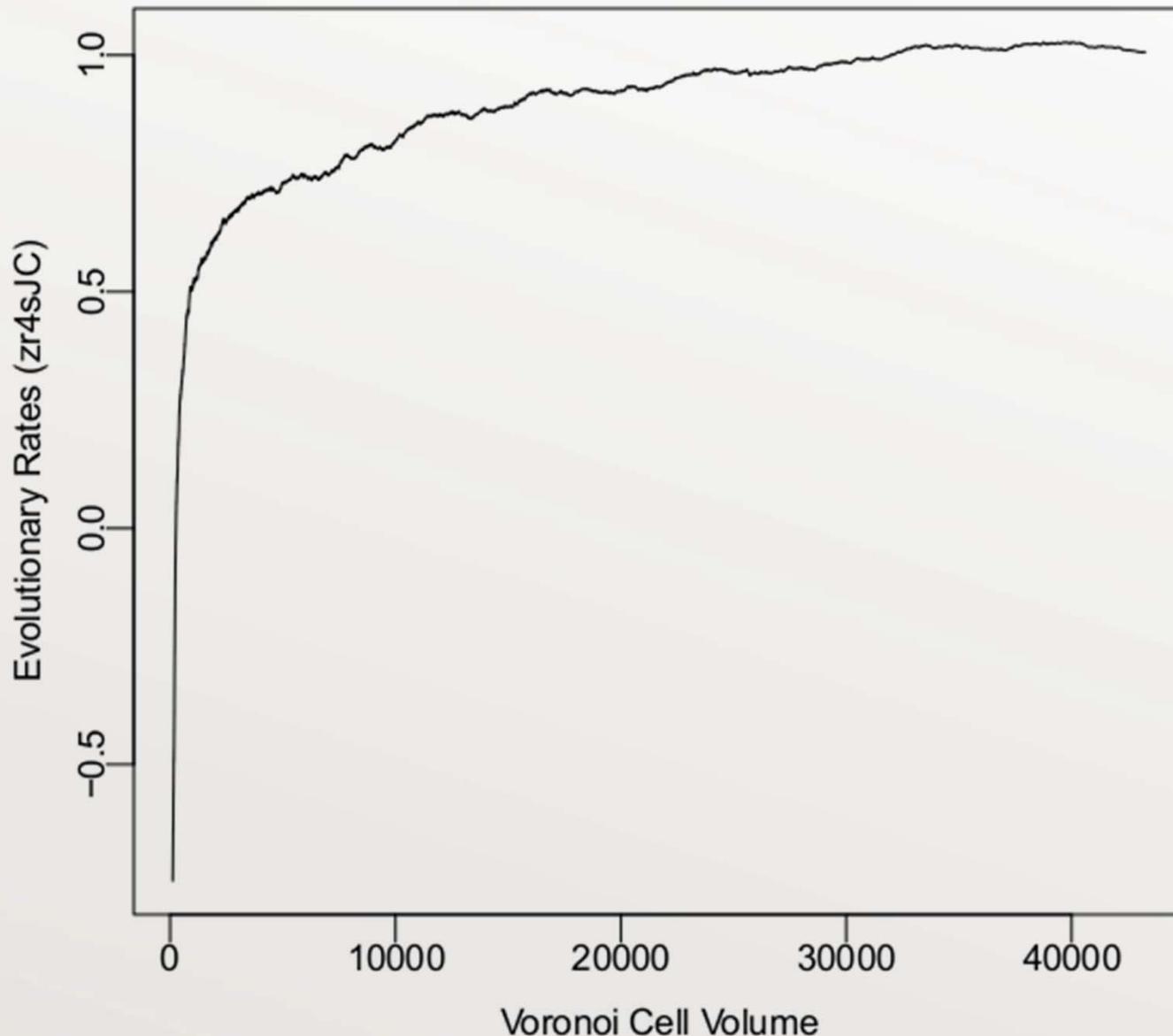
Voronoi tessellation as an unbiased parameter-free measure of site flexibility



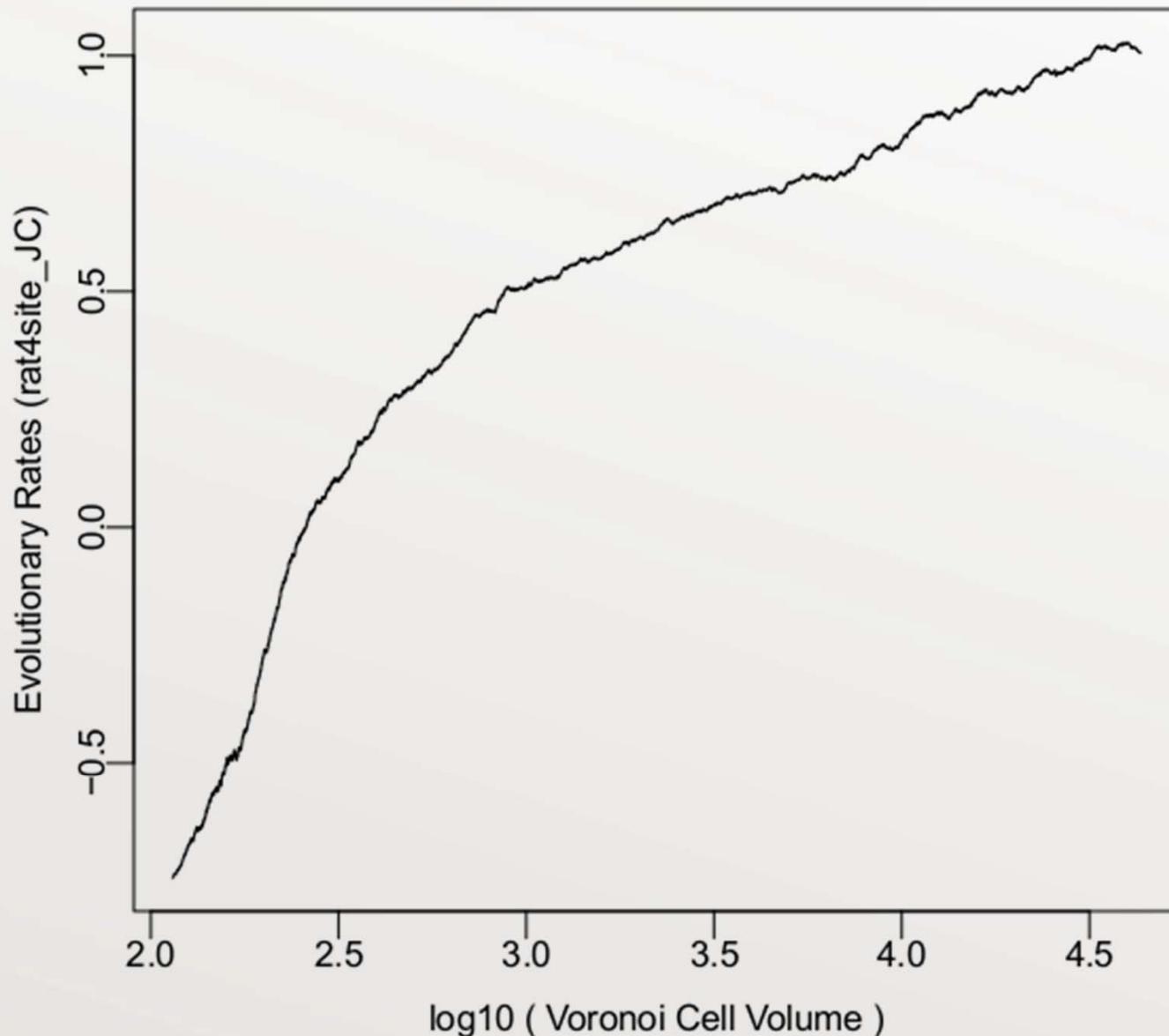
Voronoi cell volume & area as the best predictors of sequence evolution



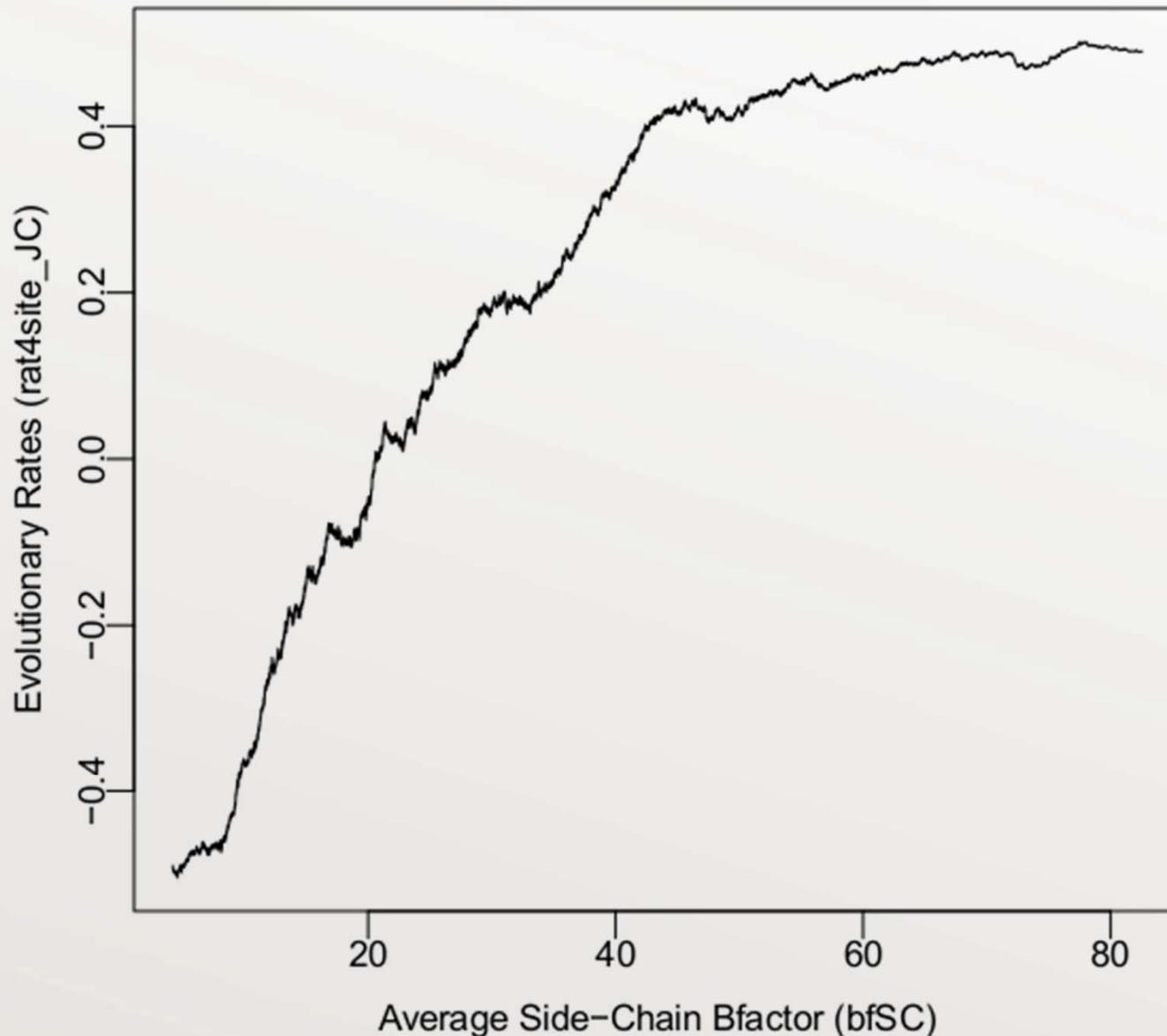
Site flexibility influences sequence evolution up to a certain threshold



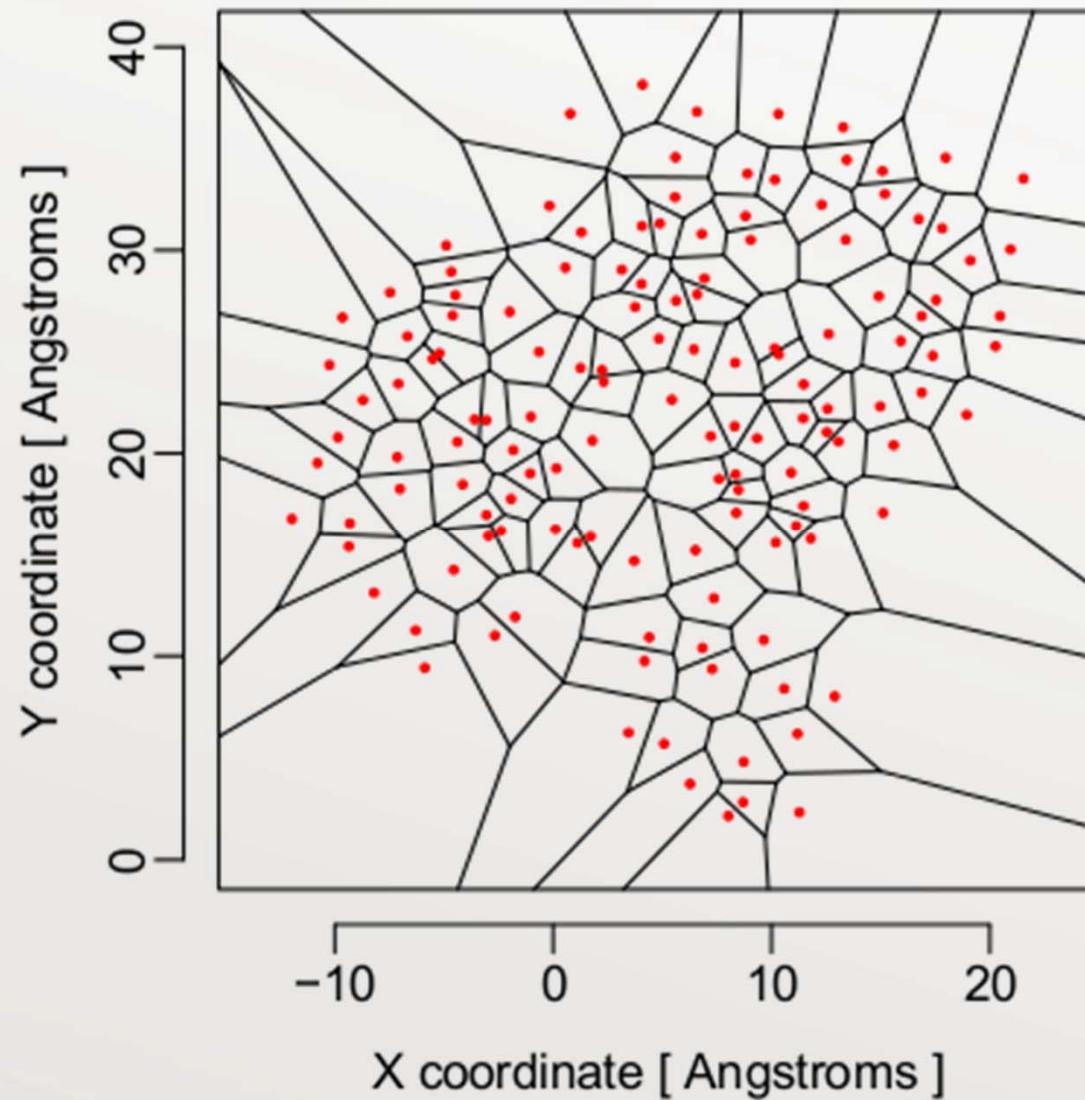
Site flexibility influences sequence evolution up to a certain threshold



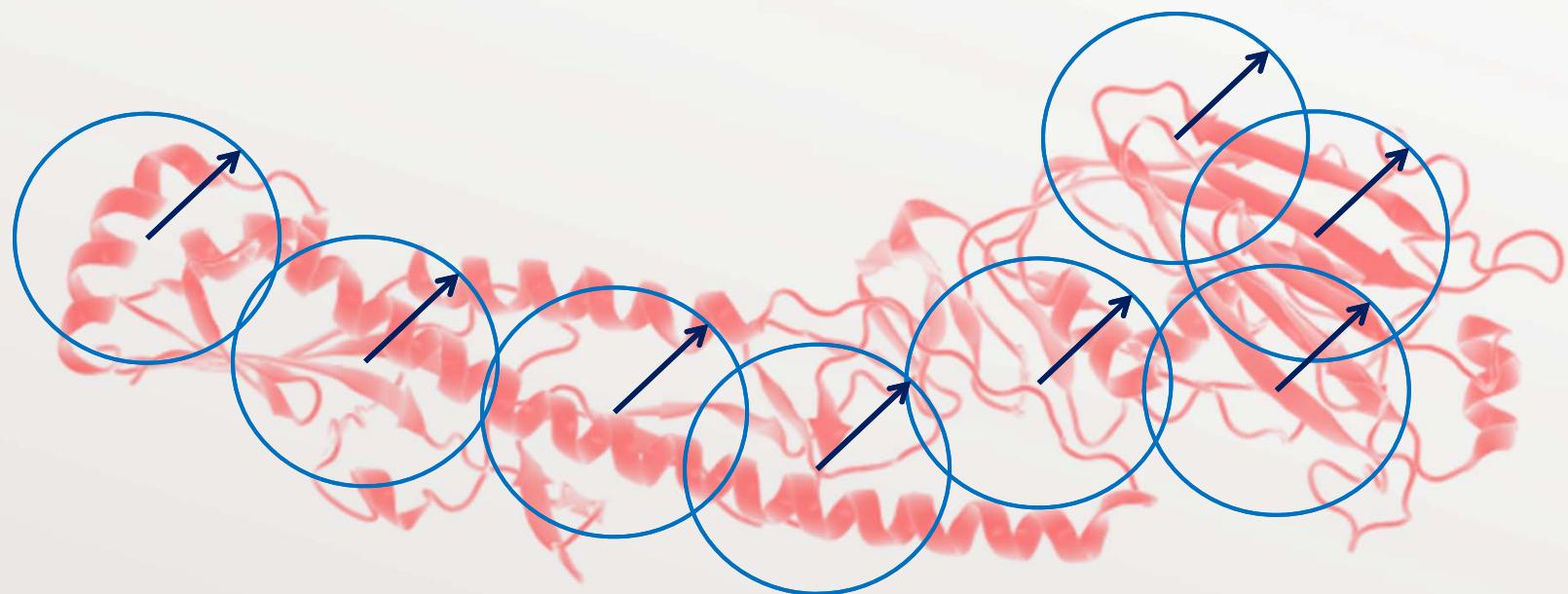
Site flexibility influences sequence evolution up to a certain threshold



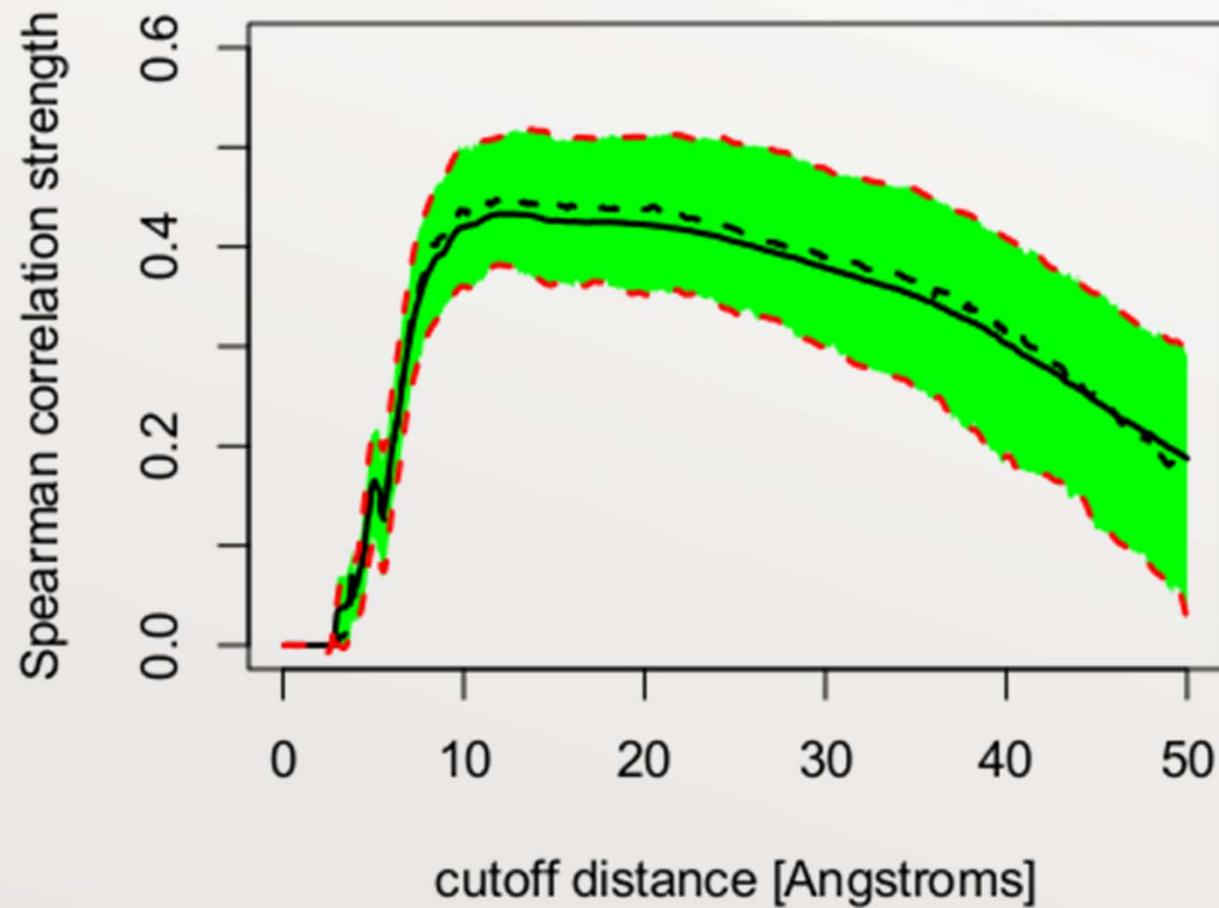
Voronoi tessellation as a parameter-free measure of Local Packing Density



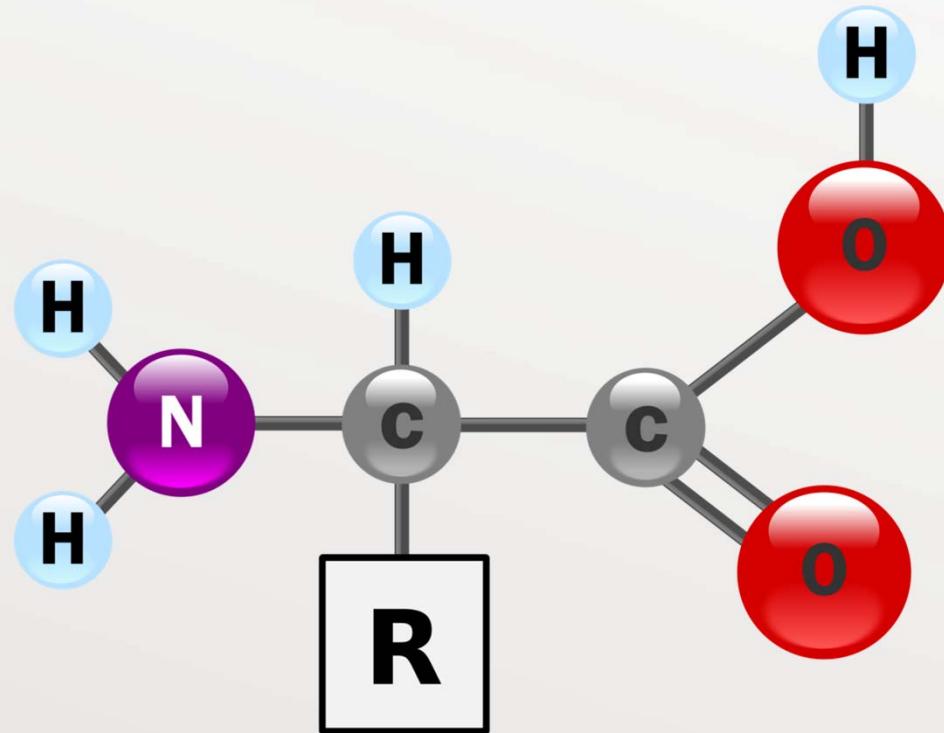
The traditional definitions of local packing density involve adjustable parameters



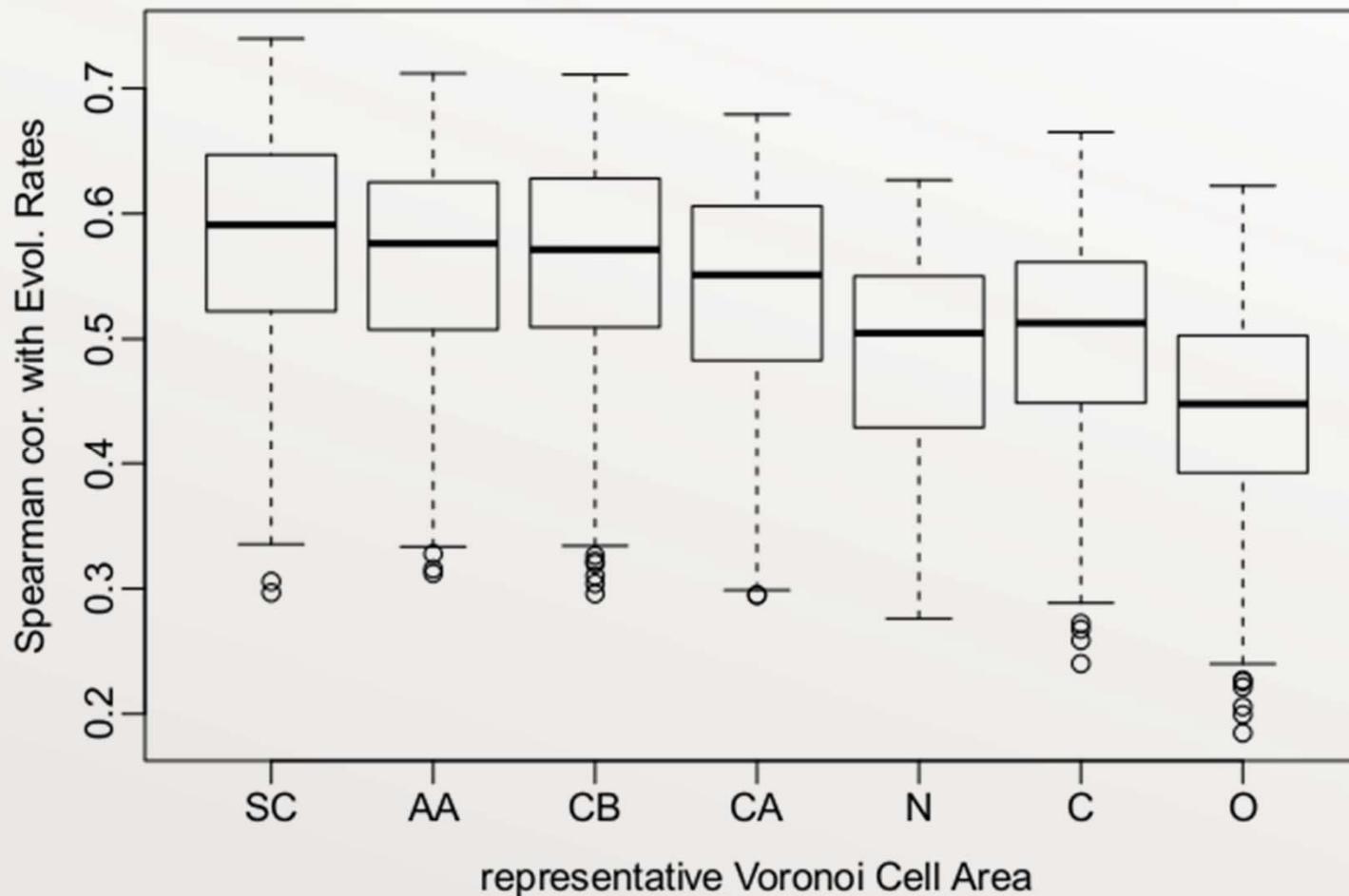
There is degeneracy in the definition of the Local Packing Density
(Contact Number)



What set of atomic coordinates best represent individual amino acids in proteins?

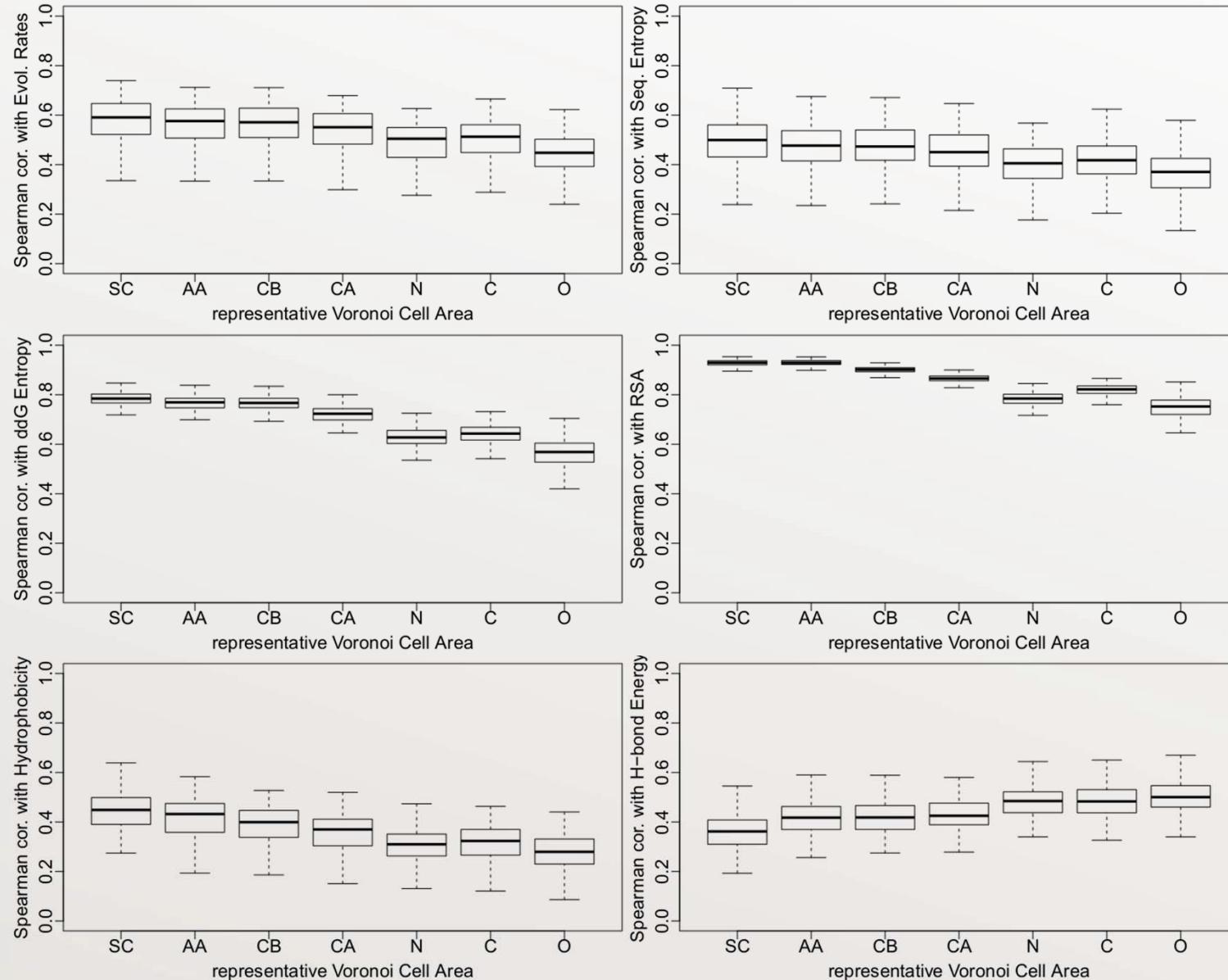


Side Chain center-of-mass coordinates are the best representation of Protein 3-dimensional structure

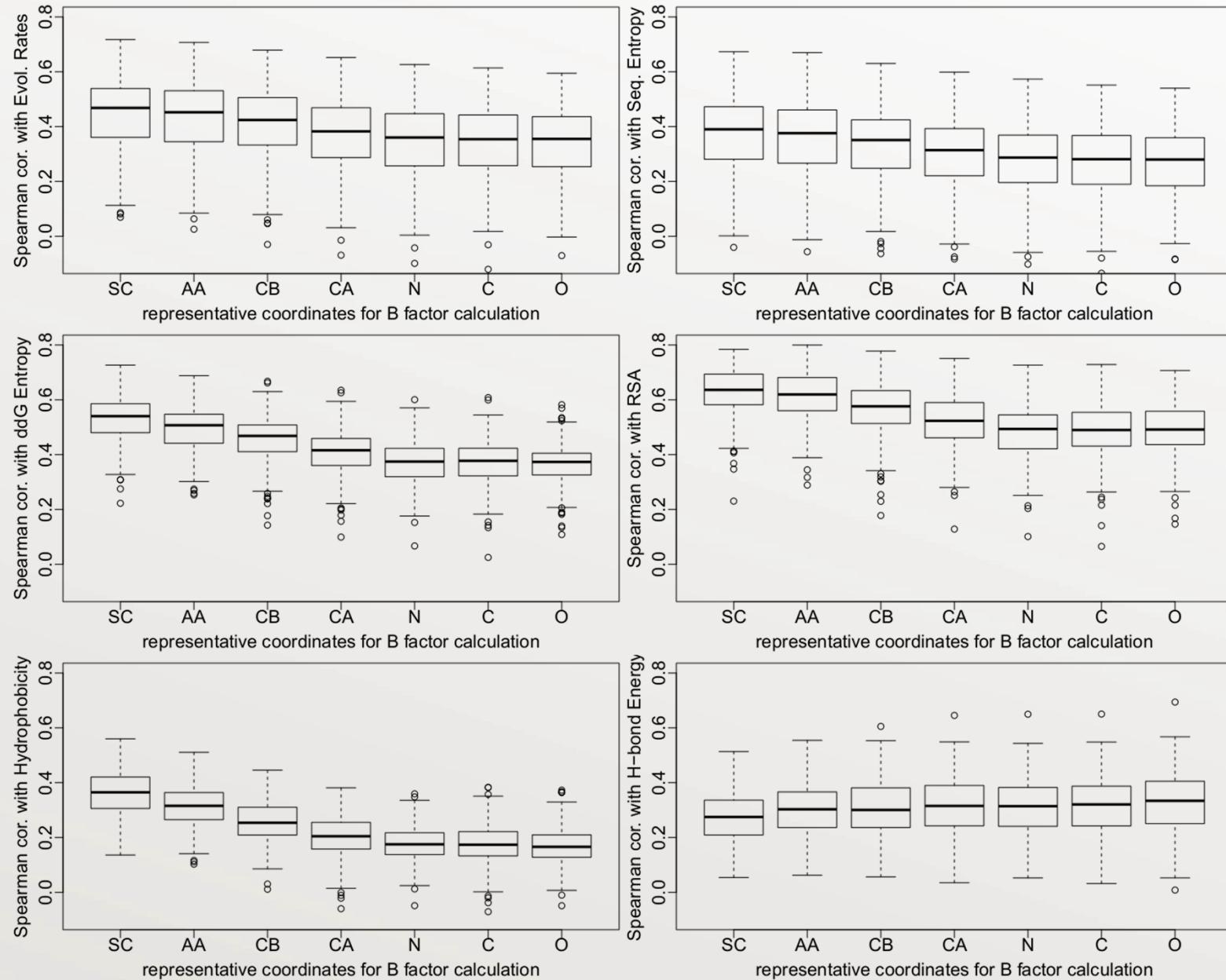


Moving away from side chain coordinates

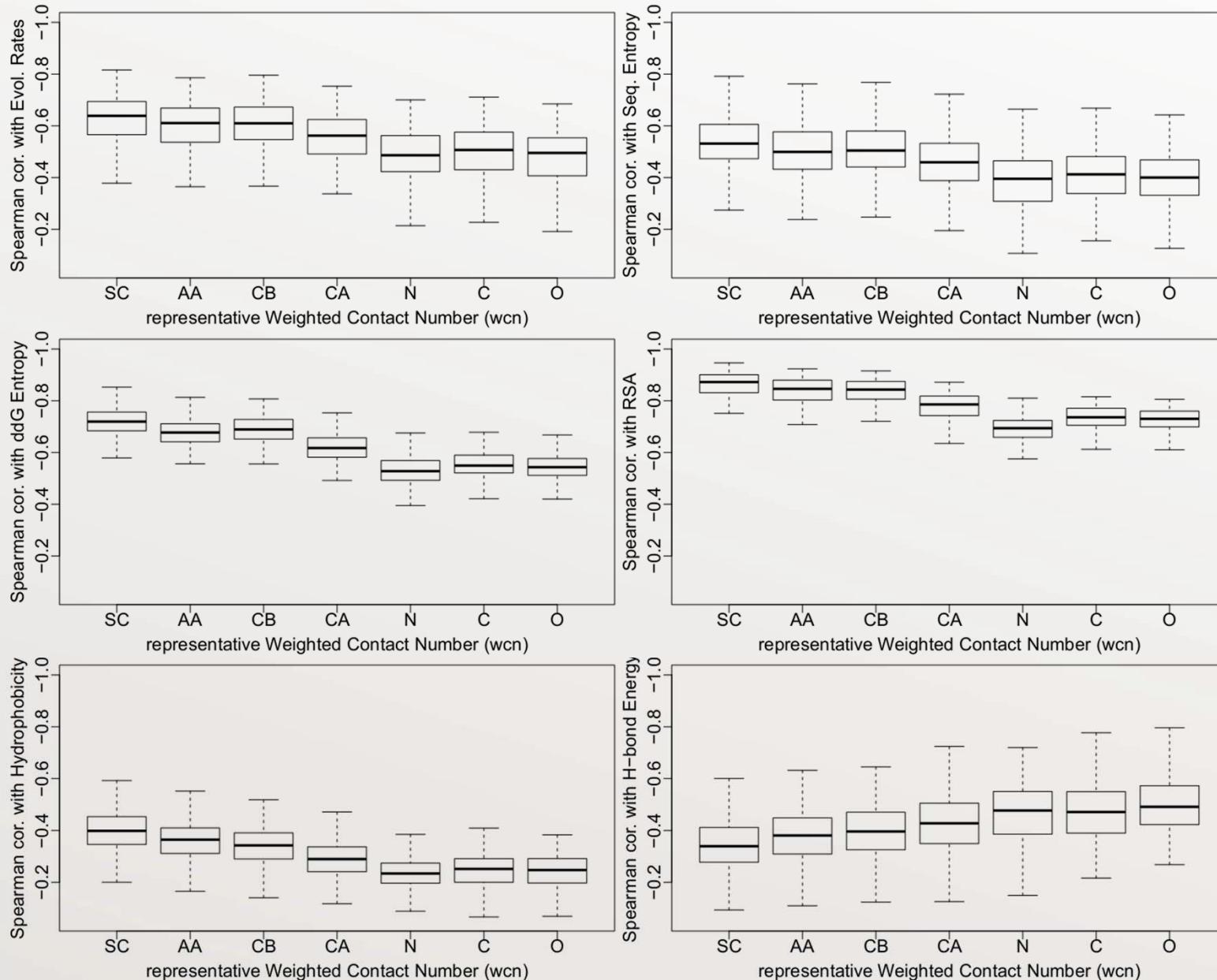
Side Chain center-of-mass coordinates are the best representation of Protein 3-dimensional structure



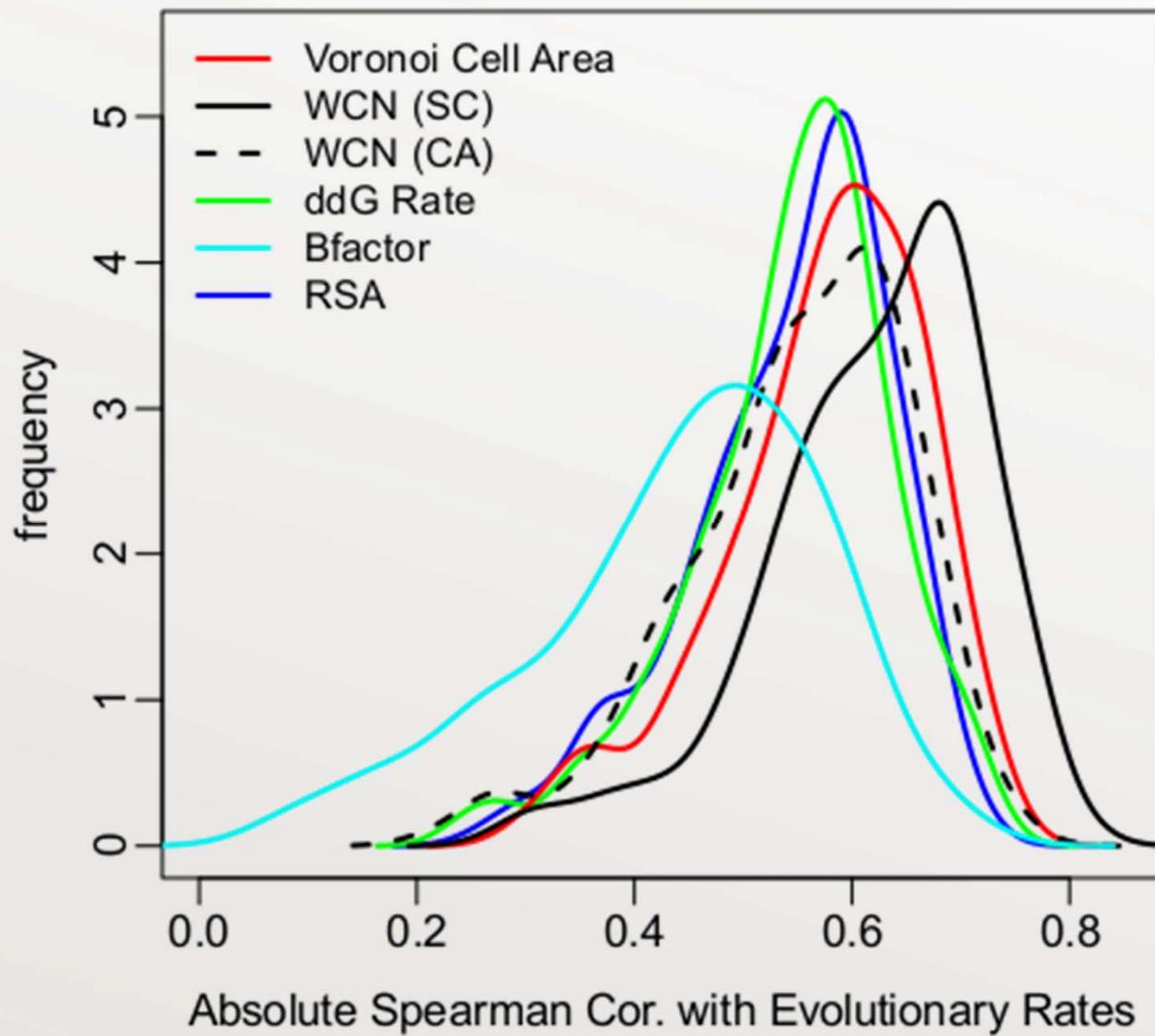
Average side chain B factor is best representative of residue local flexibility in protein



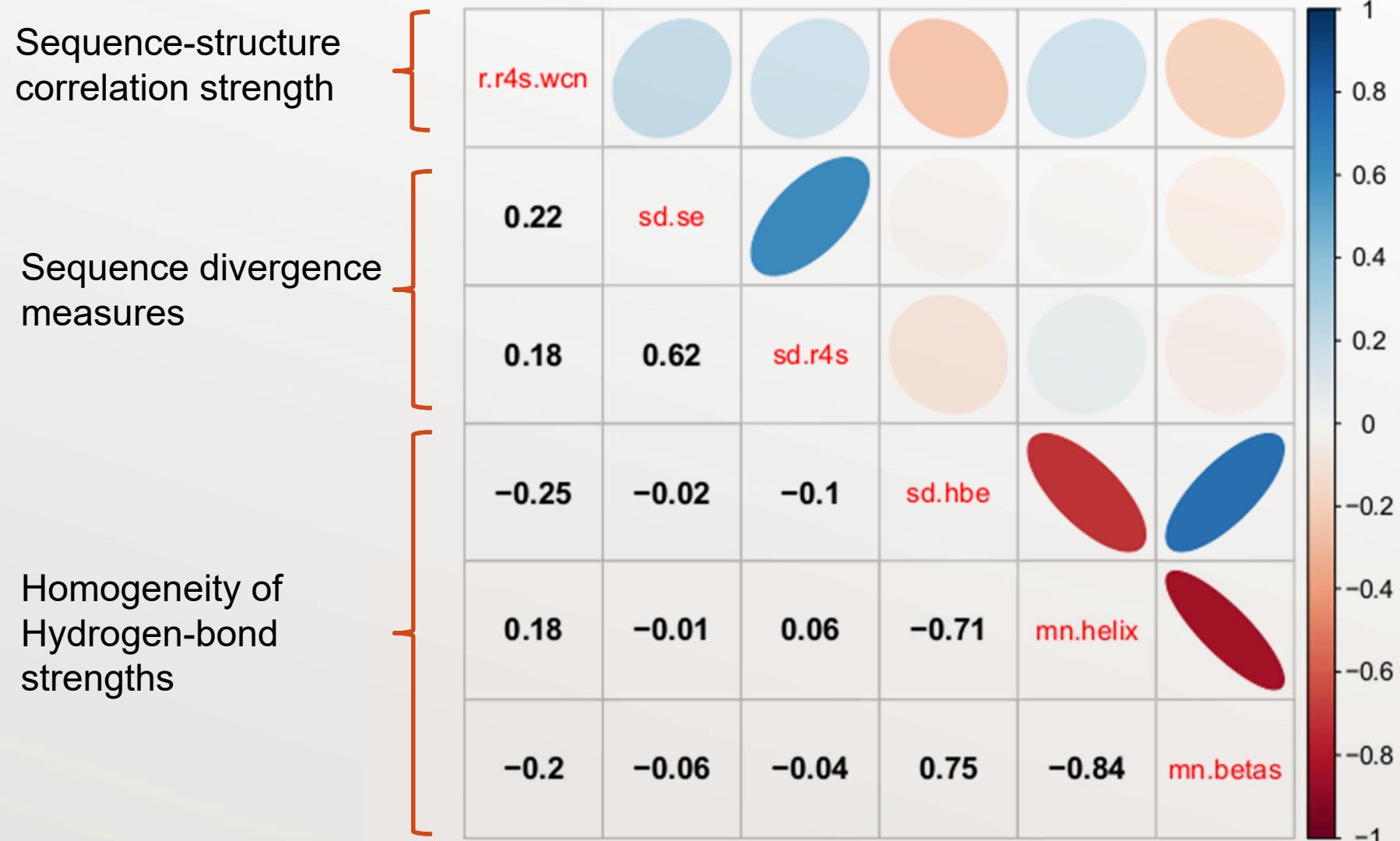
Side Chain center-of-mass coordinates are the best representation of Protein 3-dimensional structure for LPD calculation



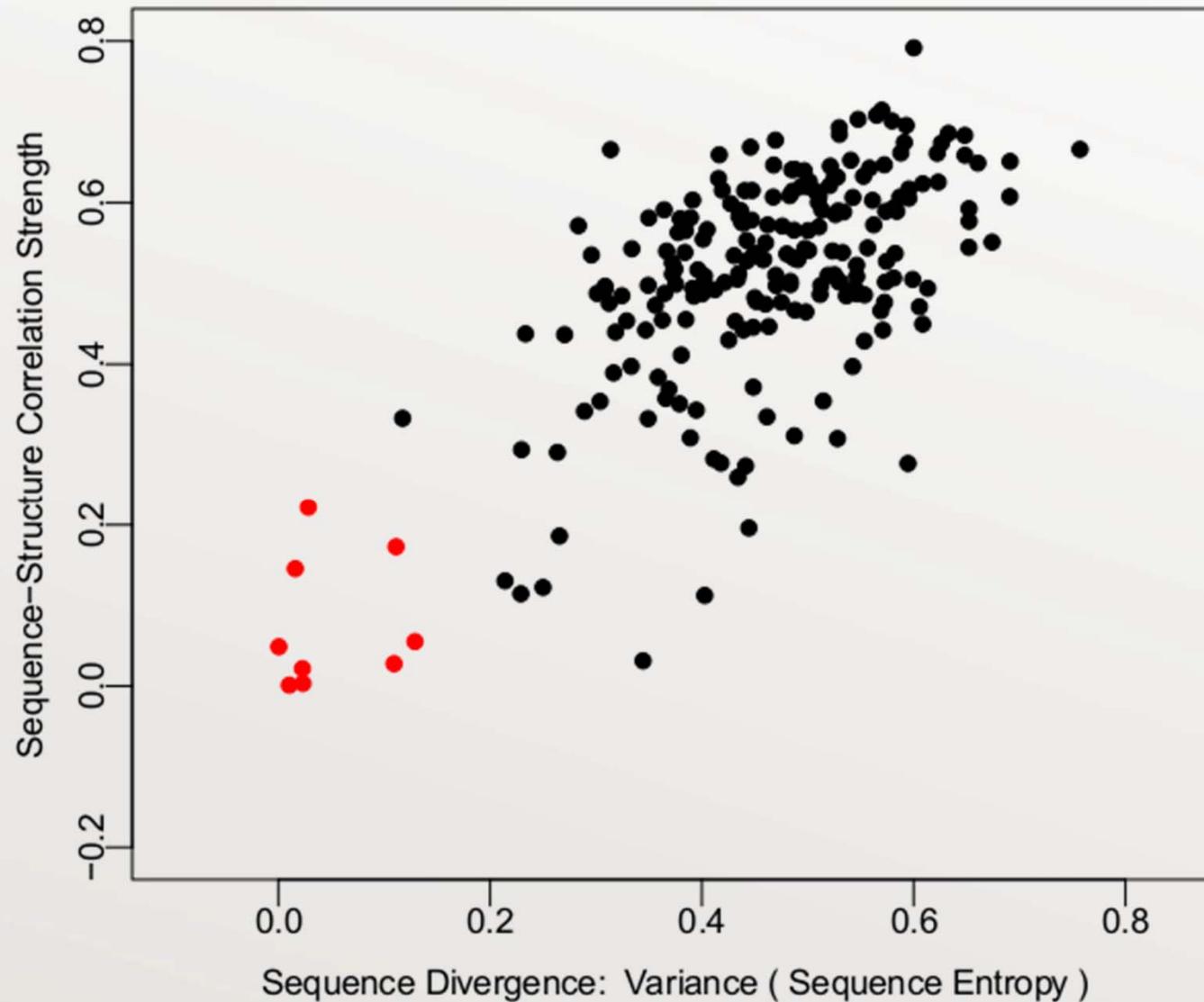
Other cell characteristics appear to have minor independent contributions



Sequence divergence & H-bond homogeneity as the main determinants of sequence-structure correlation



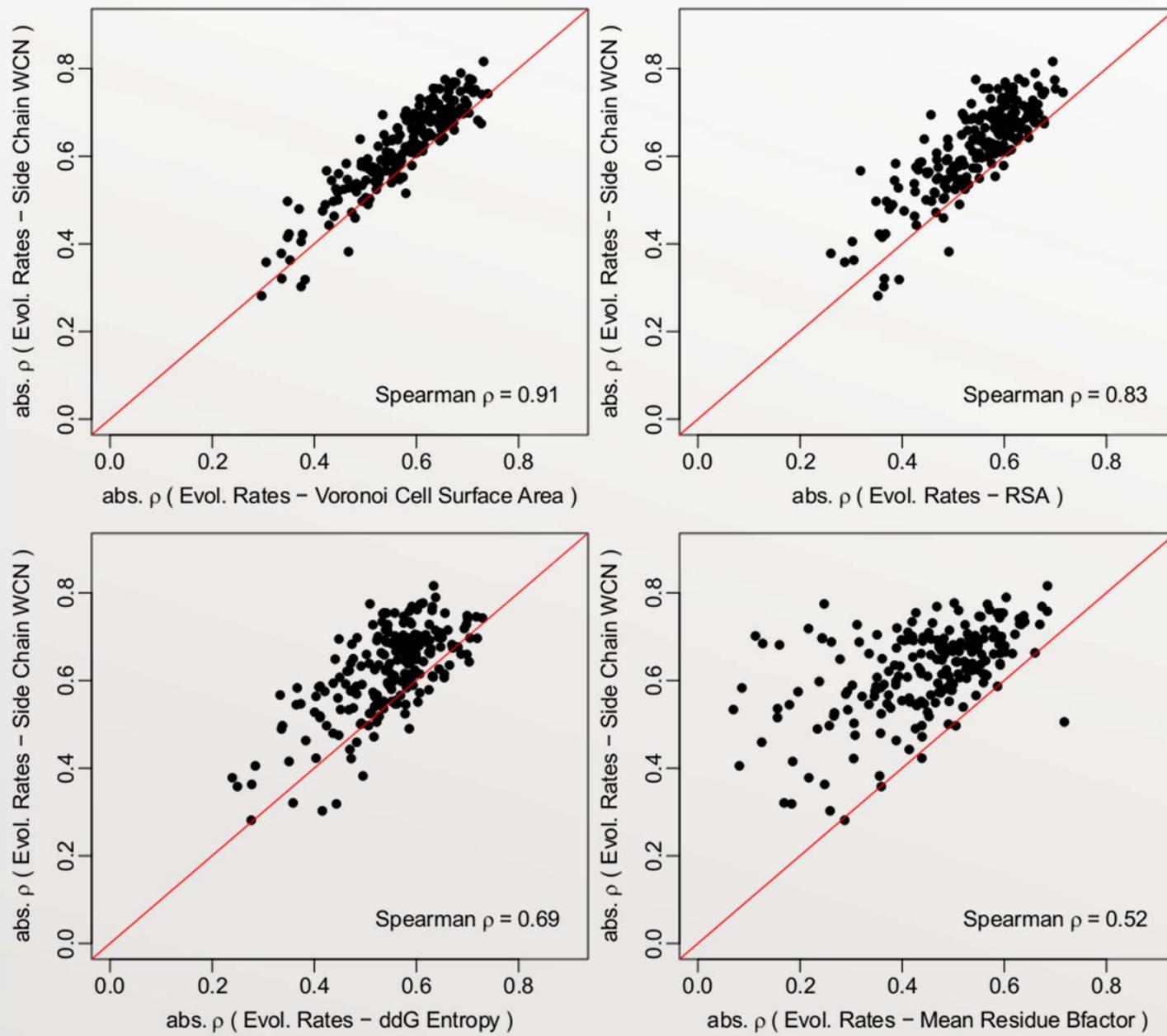
Sequence divergence appears to be the primary determinant of evolution-structure correlations



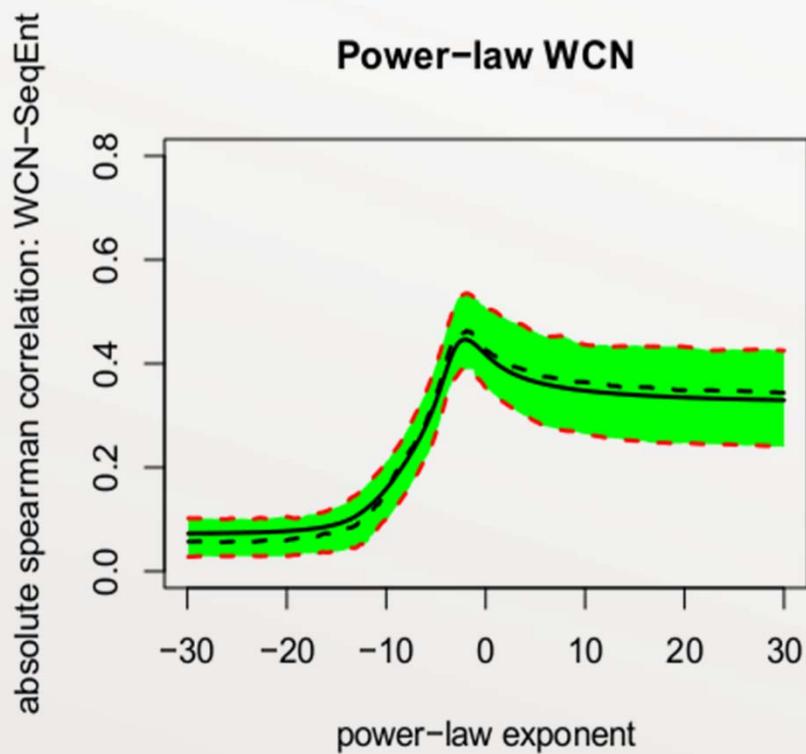
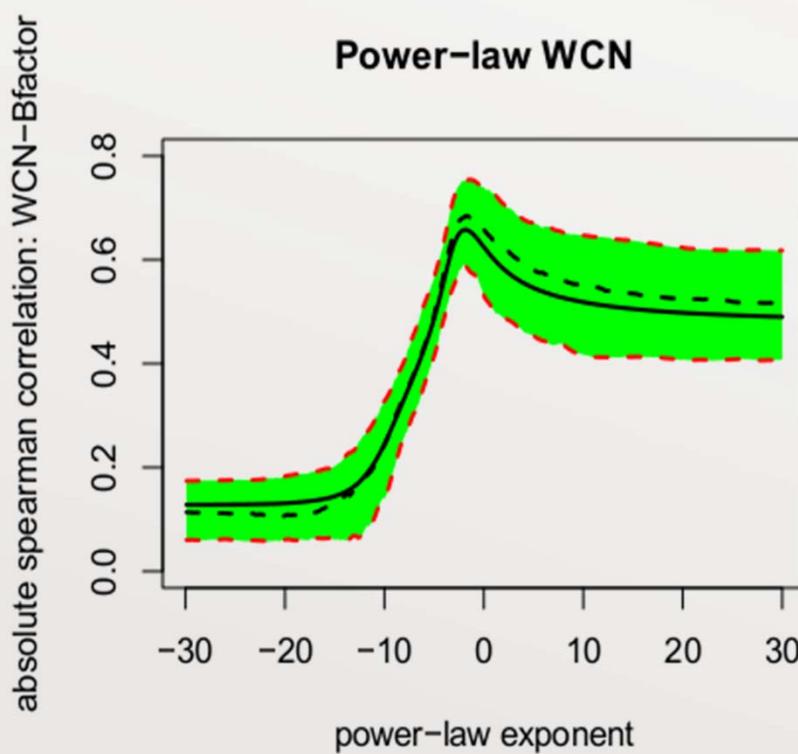
Summary

- 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

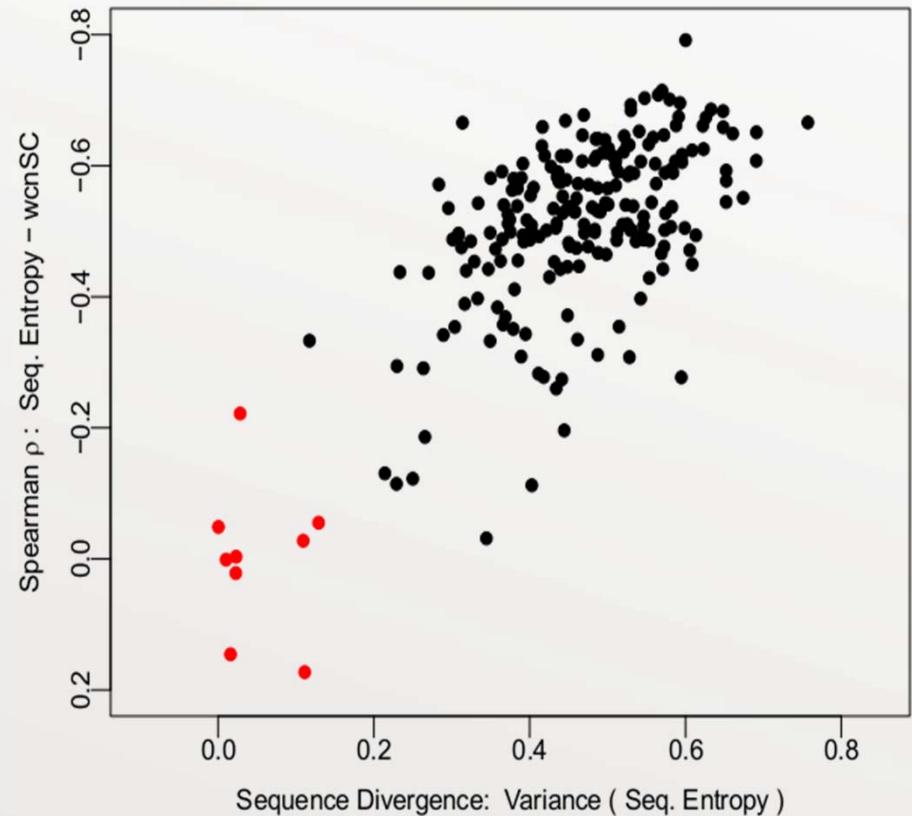
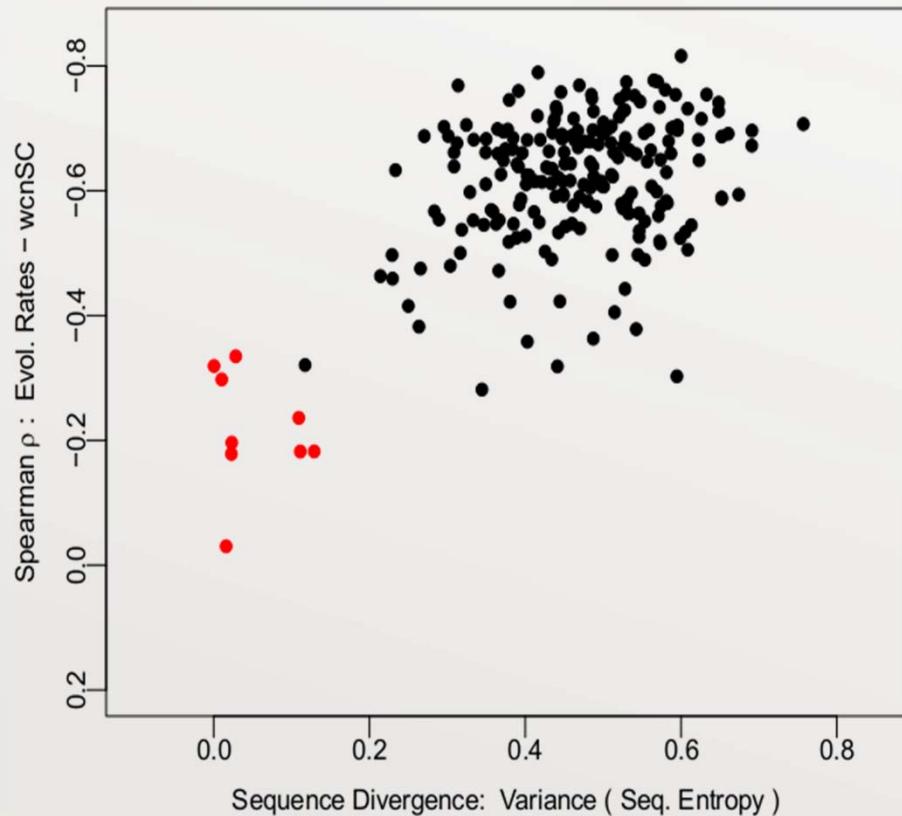
Proteins exhibit a diverse range of sequence-structure correlations



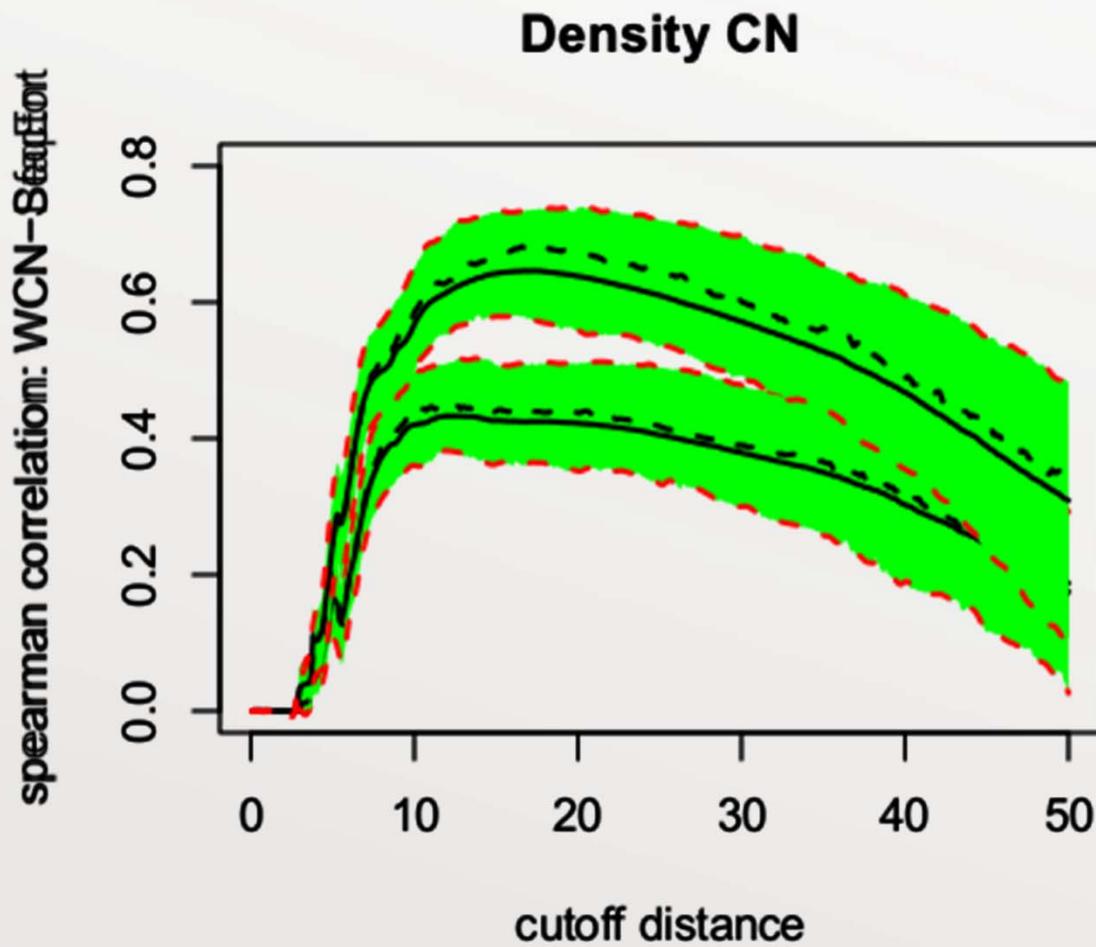
There is also degeneracy in the definition of the structural properties
consider Contact Number as an example



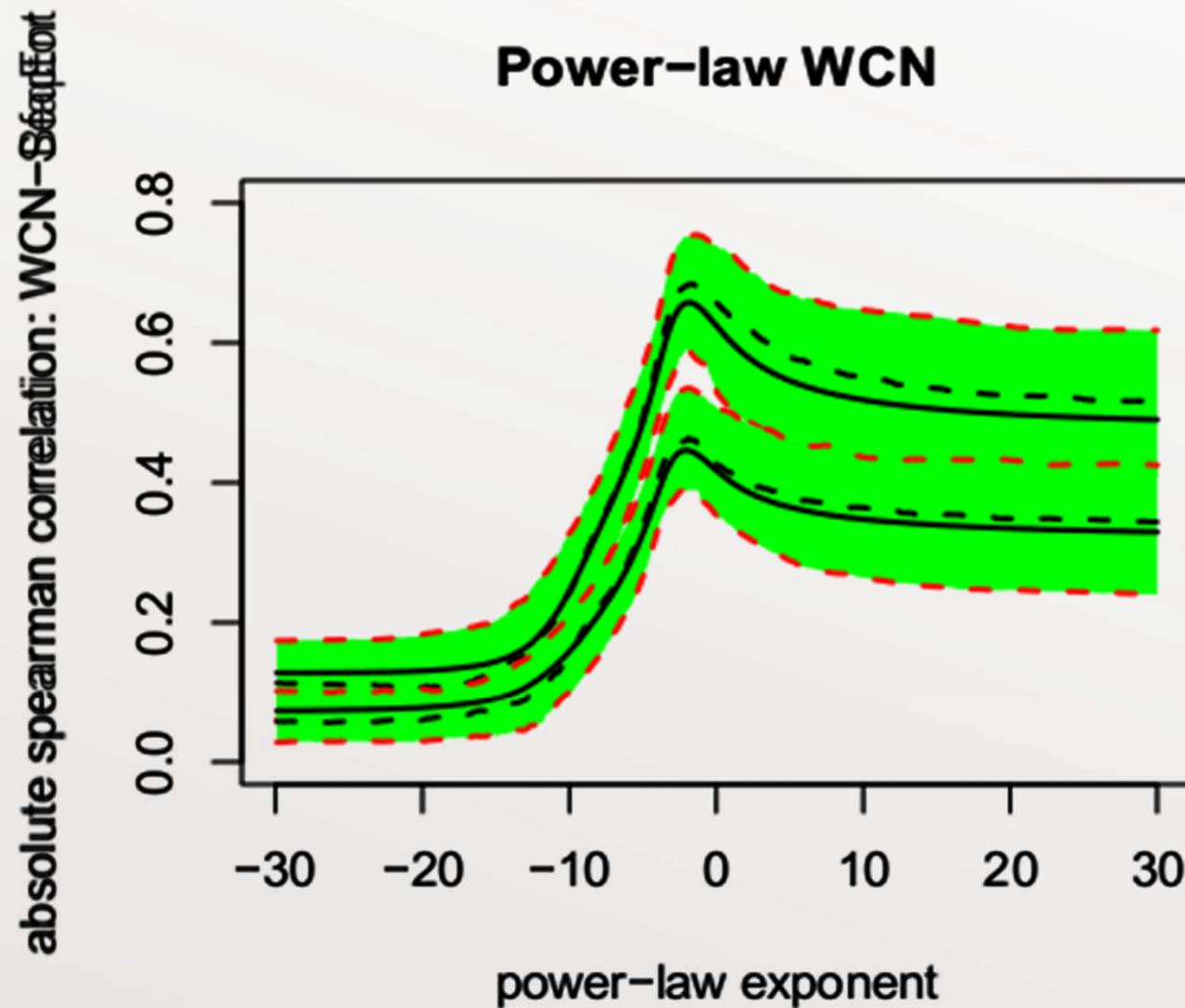
Sequence divergence appears to be the primary determinant of evolution-structure correlations



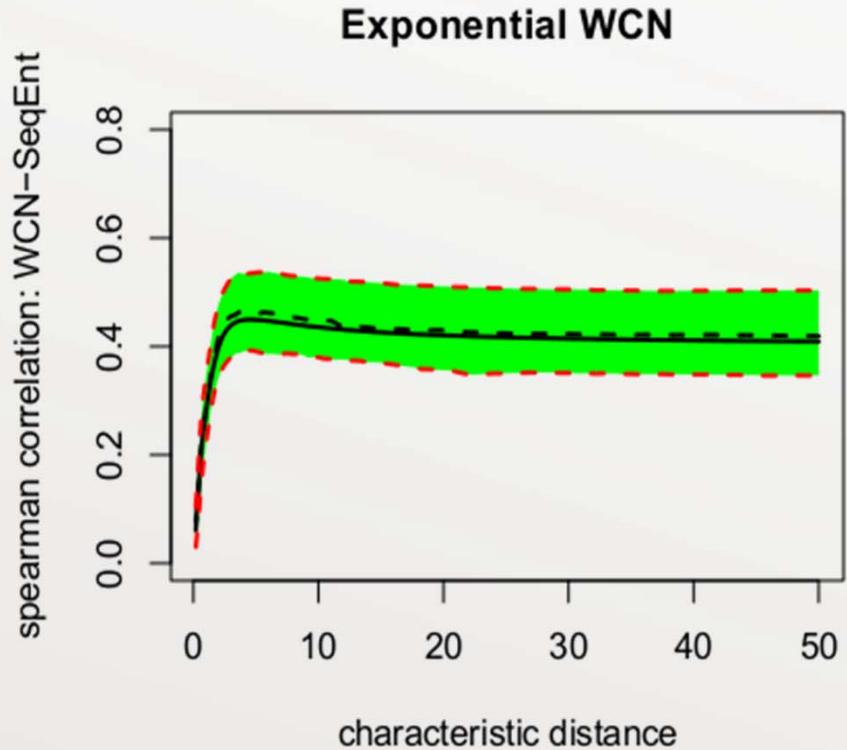
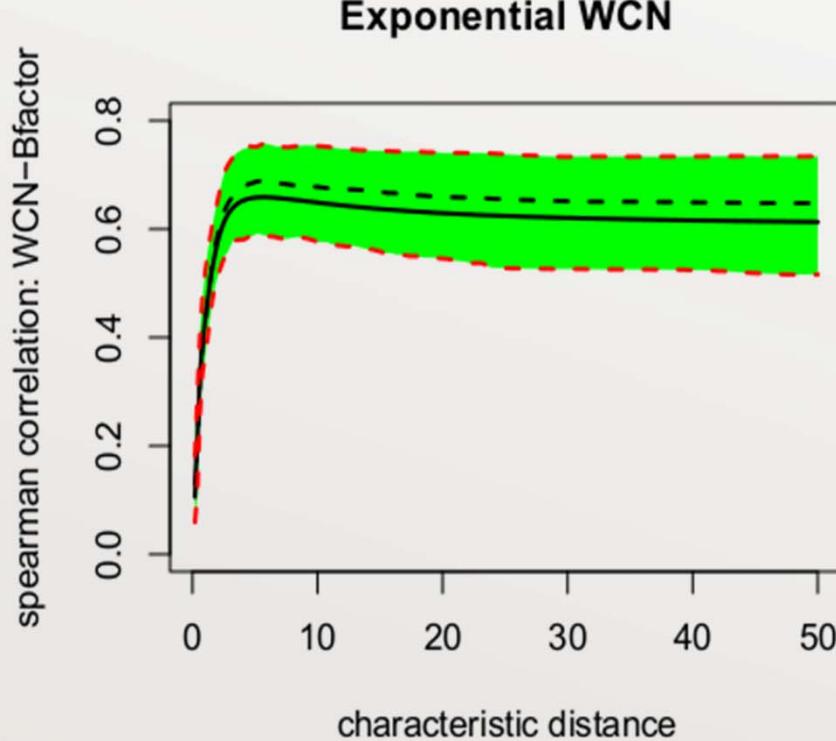
There is also degeneracy in the definition of the structural properties
consider Contact Number as an example



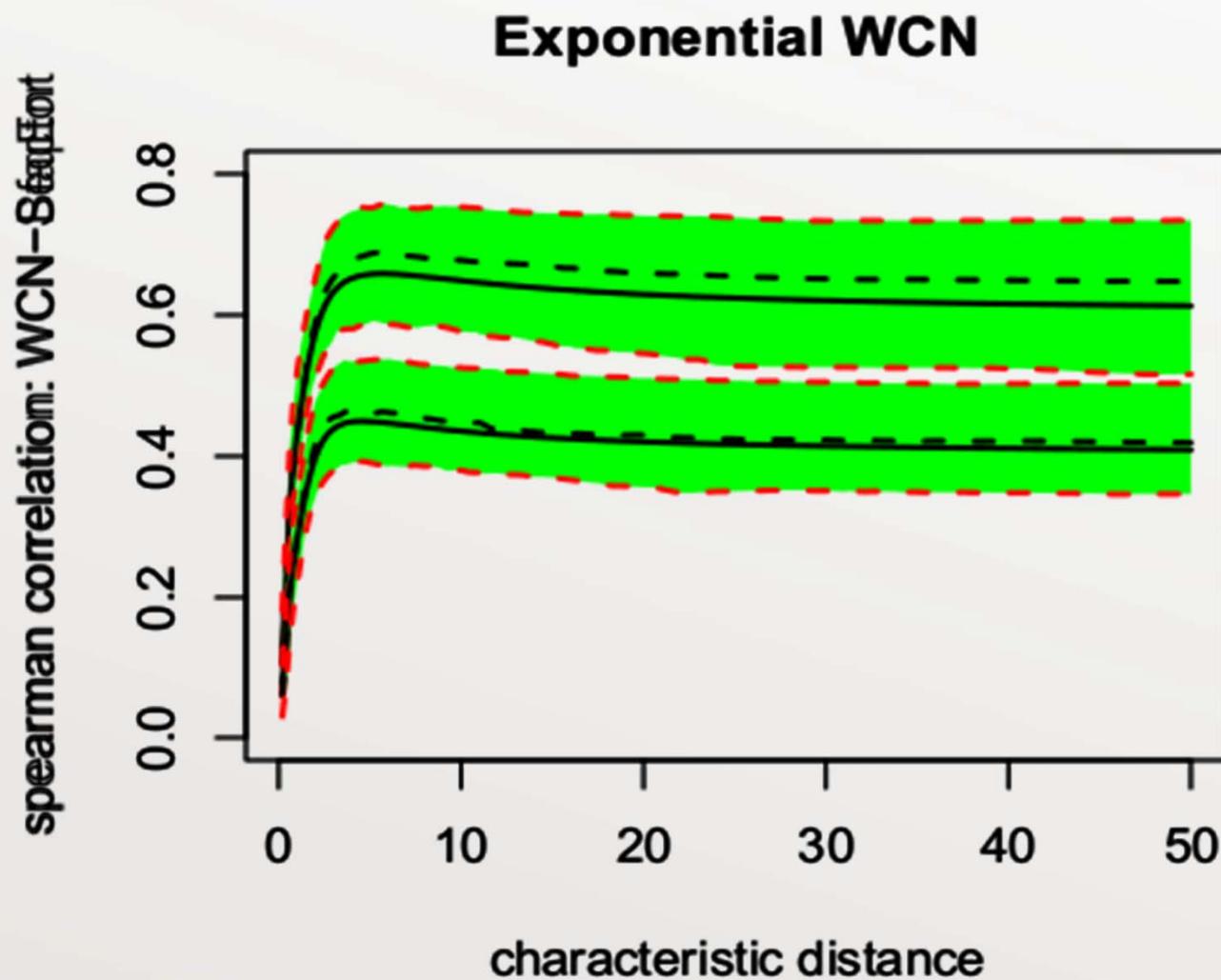
There is also degeneracy in the definition of the structural properties
consider Contact Number as an example



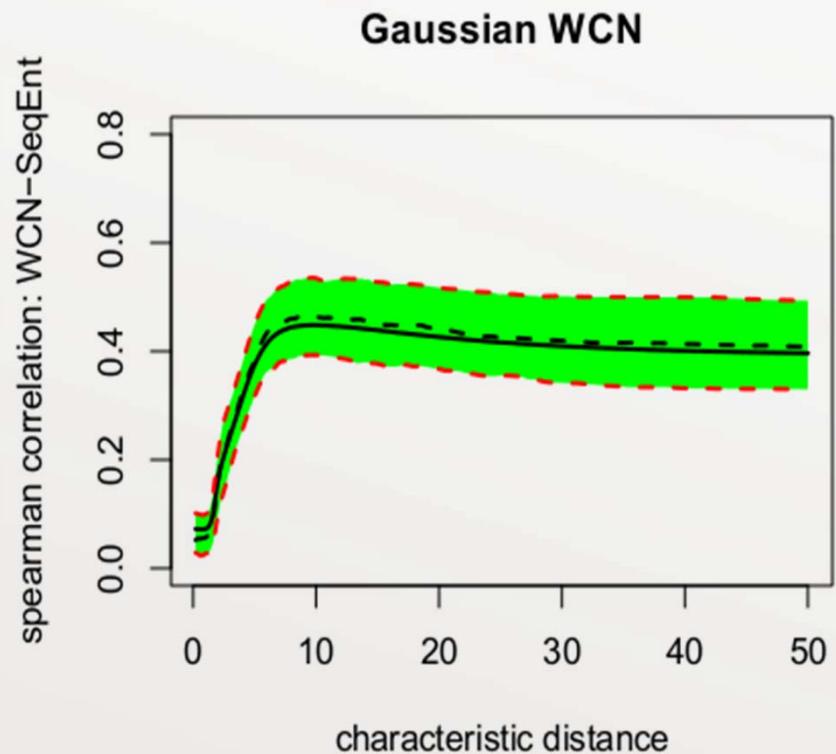
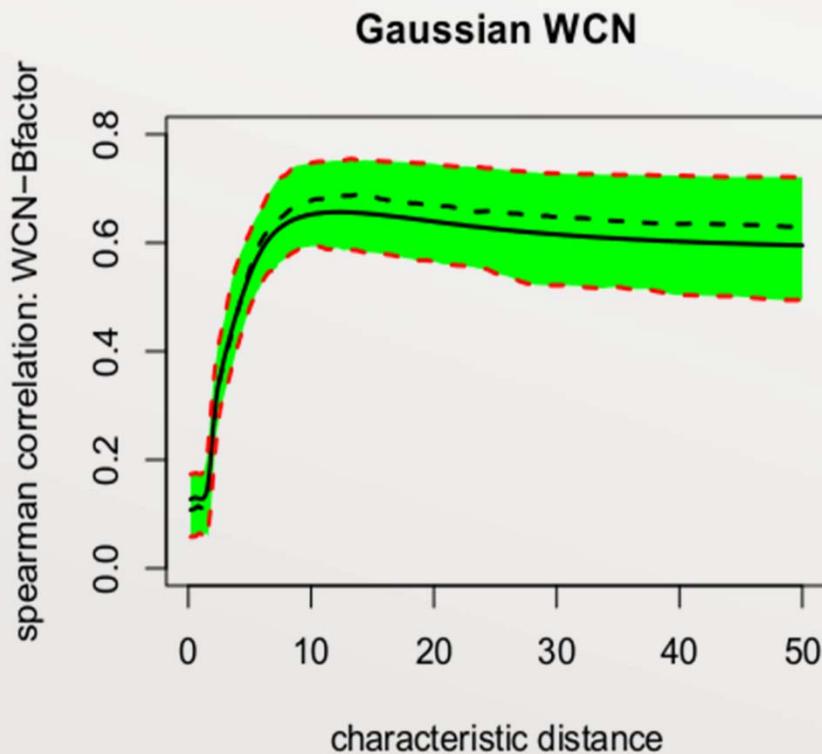
There is also degeneracy in the definition of the structural properties
consider Contact Number as an example



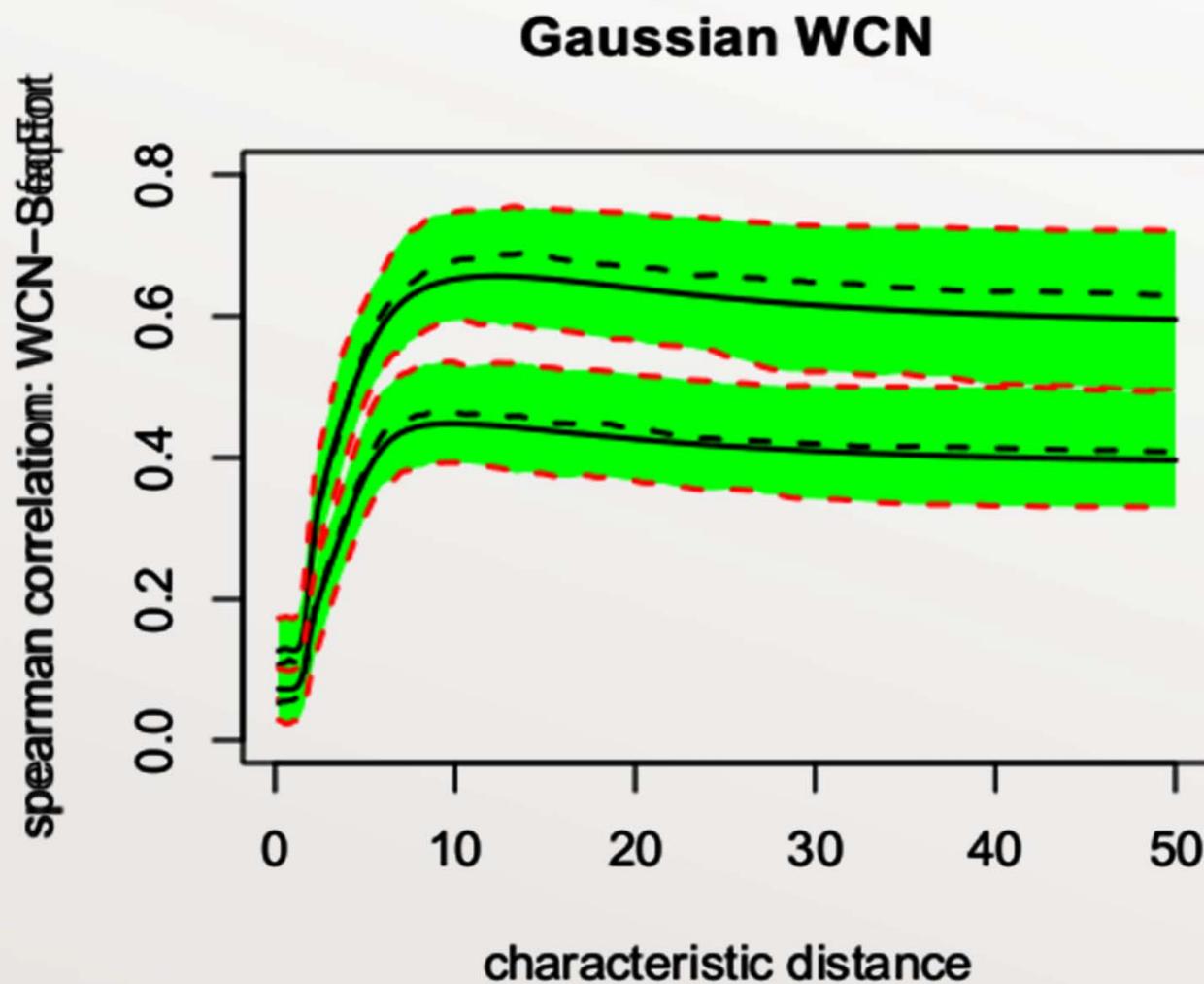
There is also degeneracy in the definition of the structural properties
consider Contact Number as an example



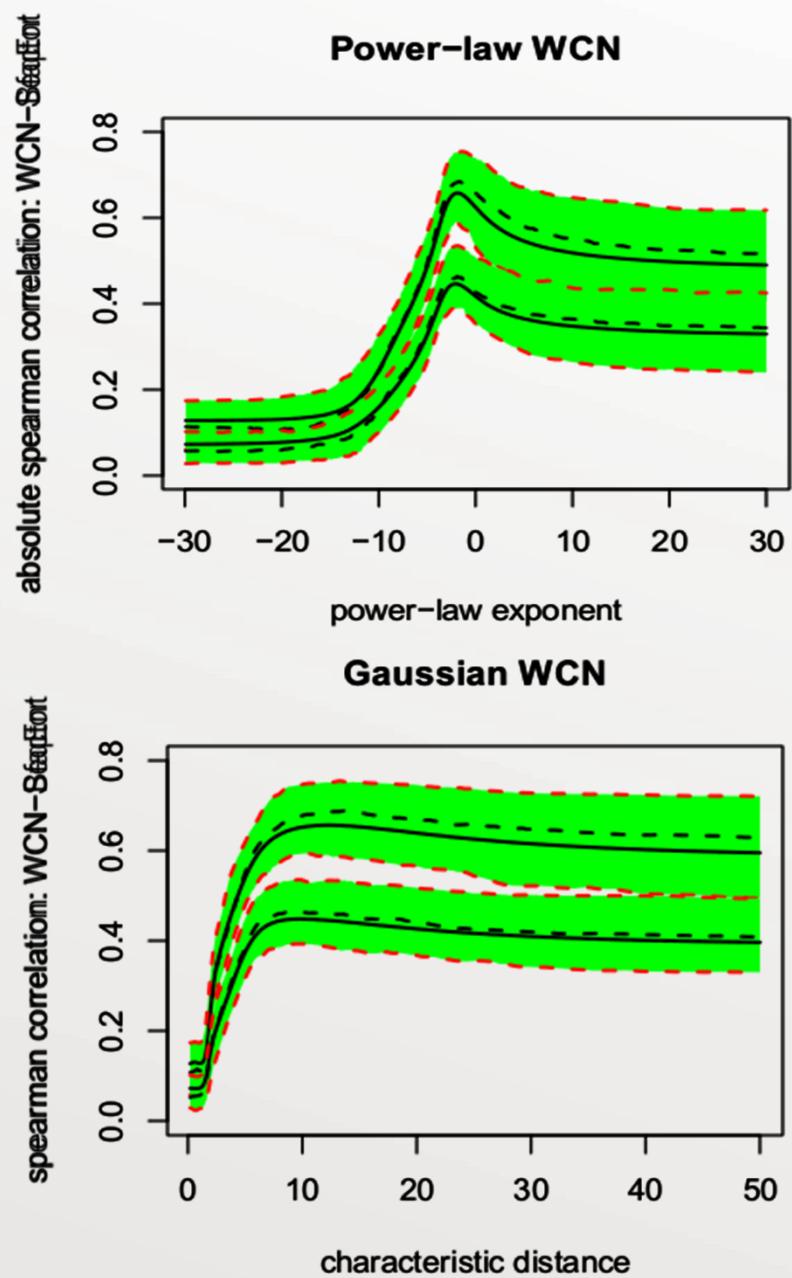
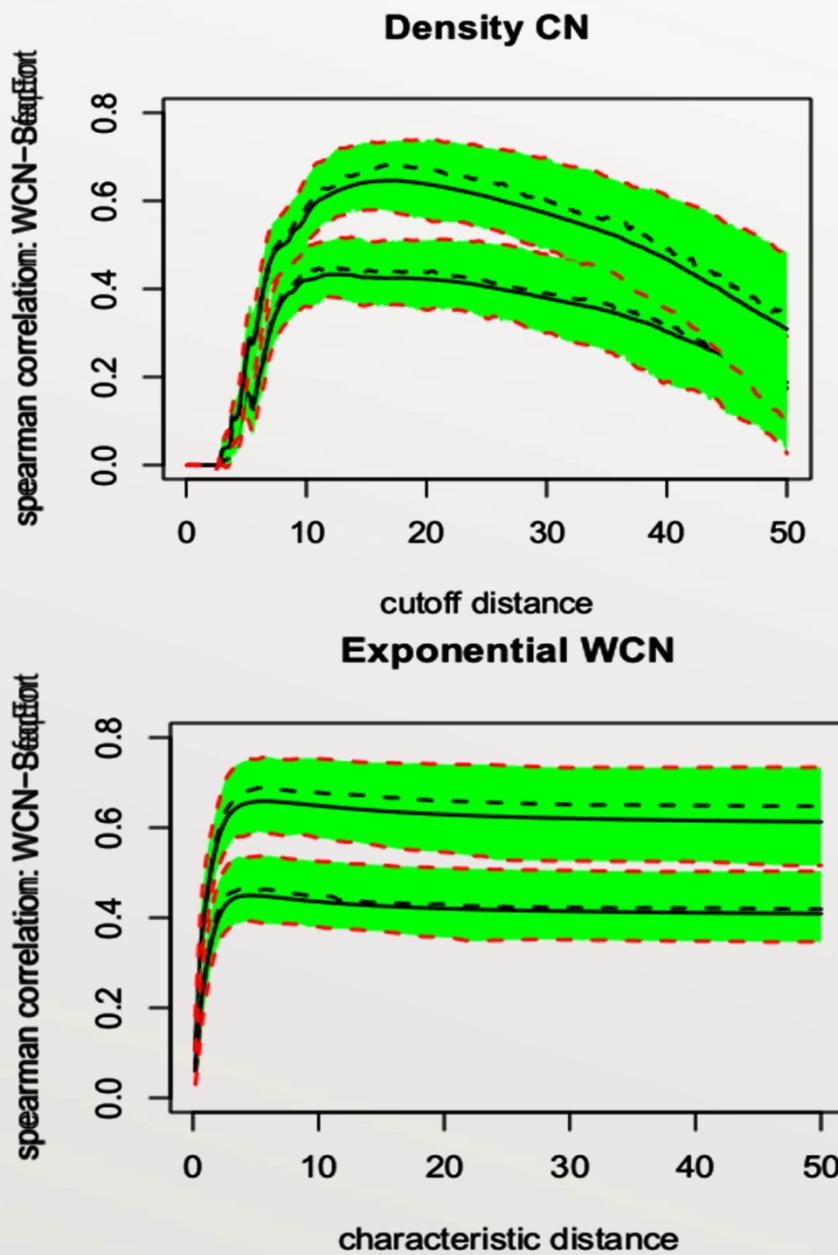
There is also degeneracy in the definition of the structural properties
consider Contact Number as an example



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

1. Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design, 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 of Molecular Evolution (JME)* 79, 3-4, 130–142, 2014.
2. Intermediate divergence levels maximize the strength of structure--sequence correlations in enzymes and viral proteins, Jackson, Eleisha L and Shahmoradi, Amir and Spielman, Stephanie J and Jack, Benjamin R and Wilke, Claus O, *Protein Science*, 25, 7, 1341-1353, 2016.
3. Dissecting the roles of local packing density and longer-range effects in protein sequence evolution, Shahmoradi, Amir and Wilke, Claus O, *Proteins: Structure, Function, and Bioinformatics*, 84, 6, 841-854, 2016.

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}  
}
```

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}  
}
```

References

```
@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}  
}
```

References

```
@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}  
}
```