

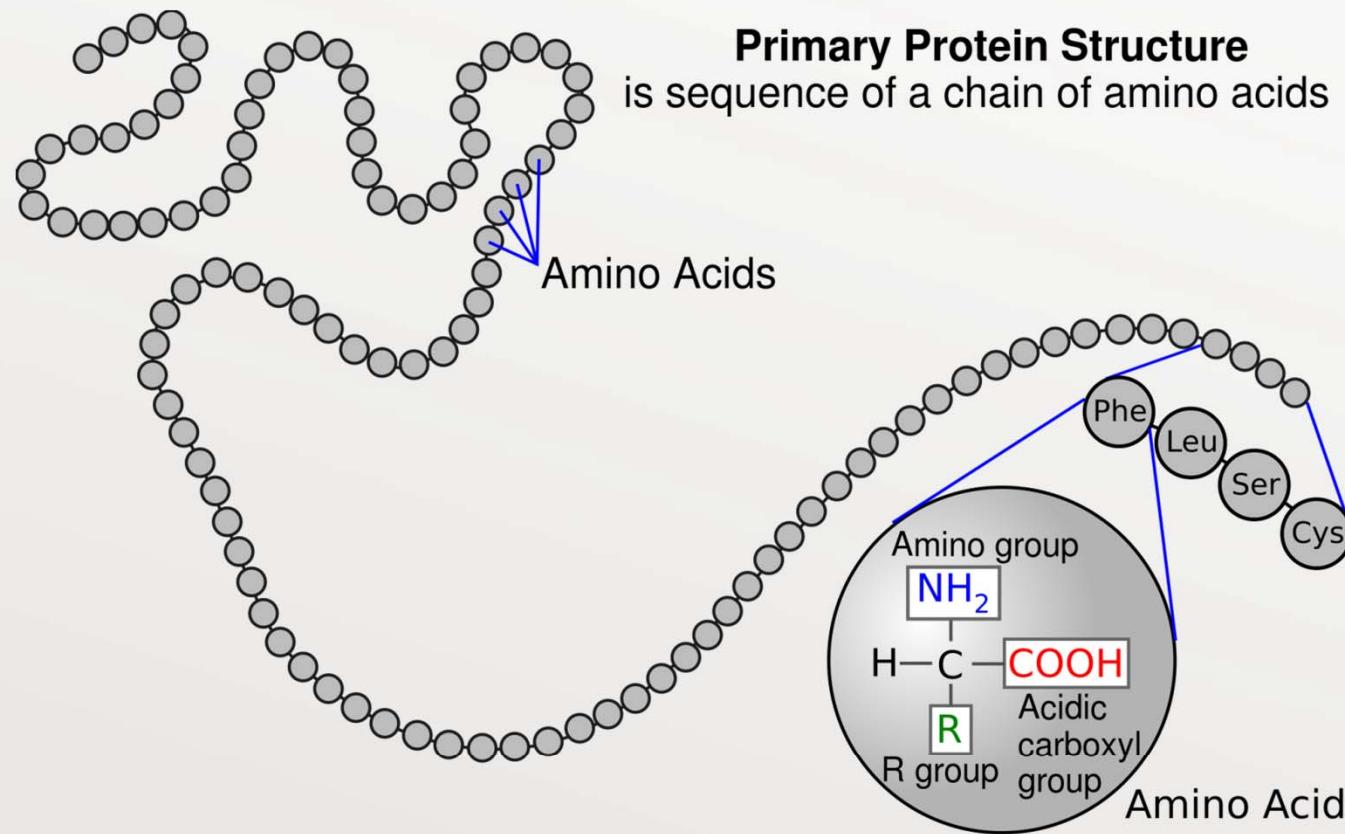
# Dissecting the relationship between protein structure and sequence evolution

Amir Shahmoradi

WilkeLab, Department of Integrative Biology  
University of Texas at Austin

BEACON Congress, August 2015

Proteins are large biomolecules consisting of one or more long chains of amino acid molecules woven together in an almost unique structure



Example protein sequence (primary structure)

MASQGTKRSYEQMETDGERQNATEIRASVGKMICGIGRFYIQMCTELKLSDYEGRLIQNSLTIERMVLSAFDERRNKYLEEHPSAGKDPKKTGGPIYKRV  
GKWMRELILYDKEEIRRIWRQANNGDDATAGLTHMMIWHSNLNQDATYQRTRALVRTGMDPRMCQLMQGSTLPRRSGAAGAAVKGIGTMVMELEMIRMIKR  
GINDRNFWRGENGKRKTRIAYERMCNILKGKFQTAACRQAMMDQVRESRNPNGNAEFEDLTFLARSALILRGSVAHKSCLPACVYGPVASGYDFEREGYSL  
VGIDPFKLLQNSQVYSLIRPNENPAHKSQLVMACHSAAFEDLRVSSFIRGTVIPRGKLSTRGVQIASNENMETMGSSTLELRSRYWAIRTRSGGNTNQ  
QRASAGQISIQPTFSVQRNLPDFRTTIMAFTGNTEGRTEIIRMMESARPEDVSFQGRGVFELSDEKATNPIVPSFDMSNEGSYFFGDNAEYDN

Proteins are large biomolecules consisting of one or more long chains of amino acid molecules woven together in an almost unique structure

Slightly different amino acid sequences



Slightly different protein structures



All perform **the same** biological function

Example protein sequence (primary structure)

MASQGTKRSYEQMETDGERQNATEIRASVGKMGIGRIFYIQMCTELKLSDYEGRLIQNSLTIERMVLSAFDERRNKYLEEHPGAGKDPKKTGGPIYKRVD  
GKWMRELILYDKEEIRRIWRQANNGDDATAGLTHMMIWHSNLNQDATYQRTRALVRTGMDPRMCQLMQGSTLPRRSGAAGAAVKGIGTMVMELEMIRMIKR  
GINDRNFWRGENGKRKTRIAYERMCNILKGKFQTAACRQAMMDQVRESRNPNGNAEFEDLTFLARSALILRGSVAHKSCLPACVYGPASGYDFEREGYSL  
VGIDPFKLLQNSQVYSLIRPNENPAHKSQVLWMACHSAAFEDLRVSSFIRGTVIPRGKLSTRGVQIASNENMETMGSSTLELRSRYWAIRTRSGGNTNQ  
QRASAGQISIQPTFSVQRNLPDFRTTIMAFTGNTEGRTEIIRMMESARPEDVSFQGRGVFELSDEKATNPIVPSFDMSNEGSYFFGDNAEYDN

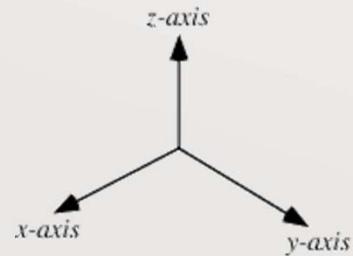
# Why are some sites in proteins more variable than others?

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 ...

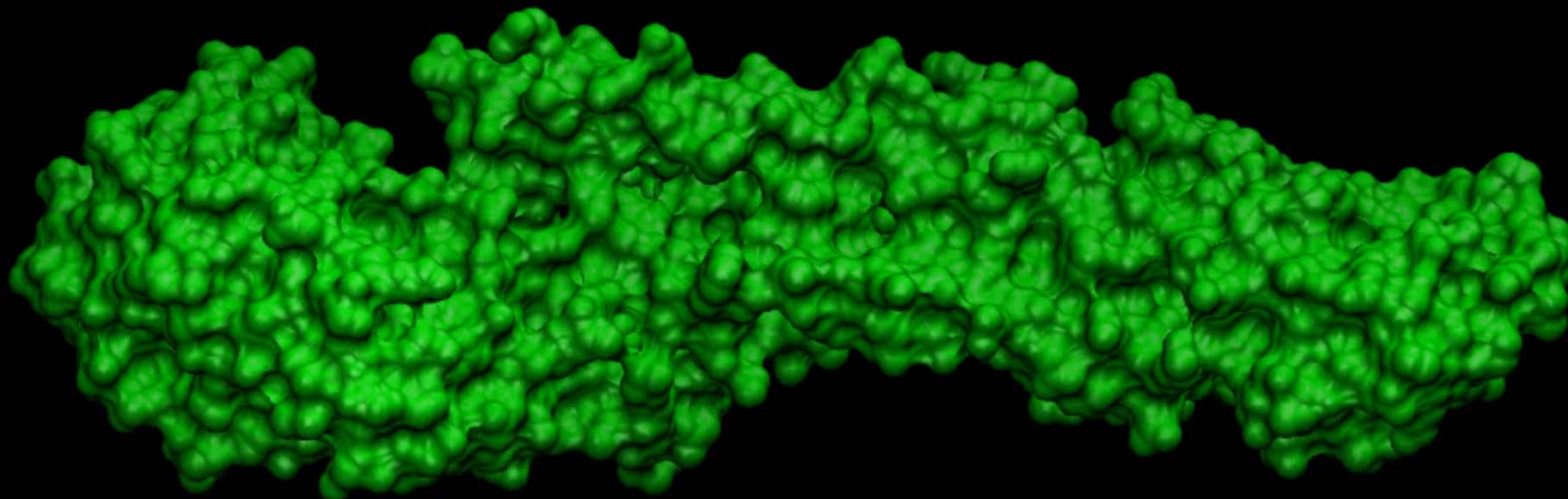
The Past      The Present      The Future



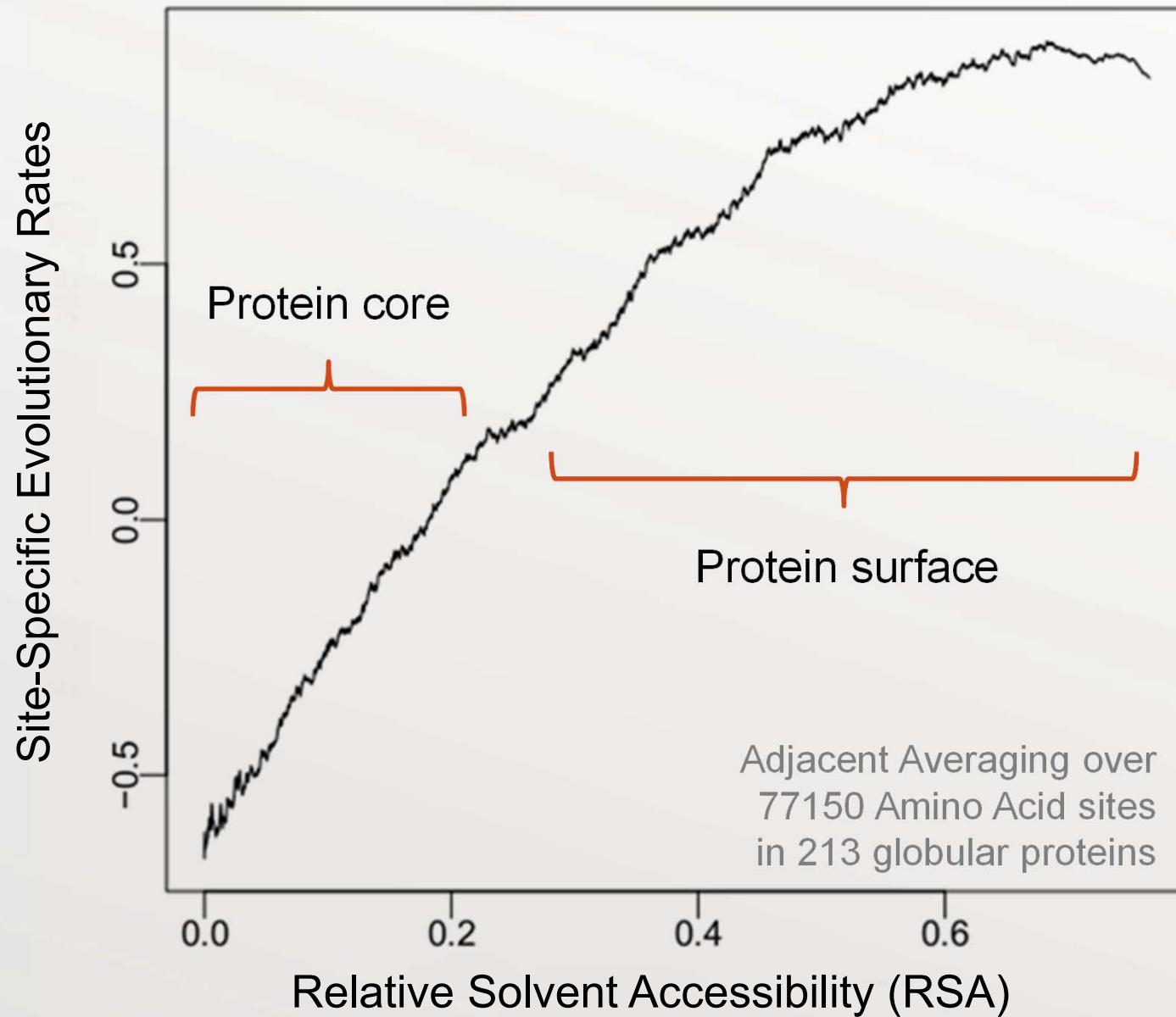
Time domain

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

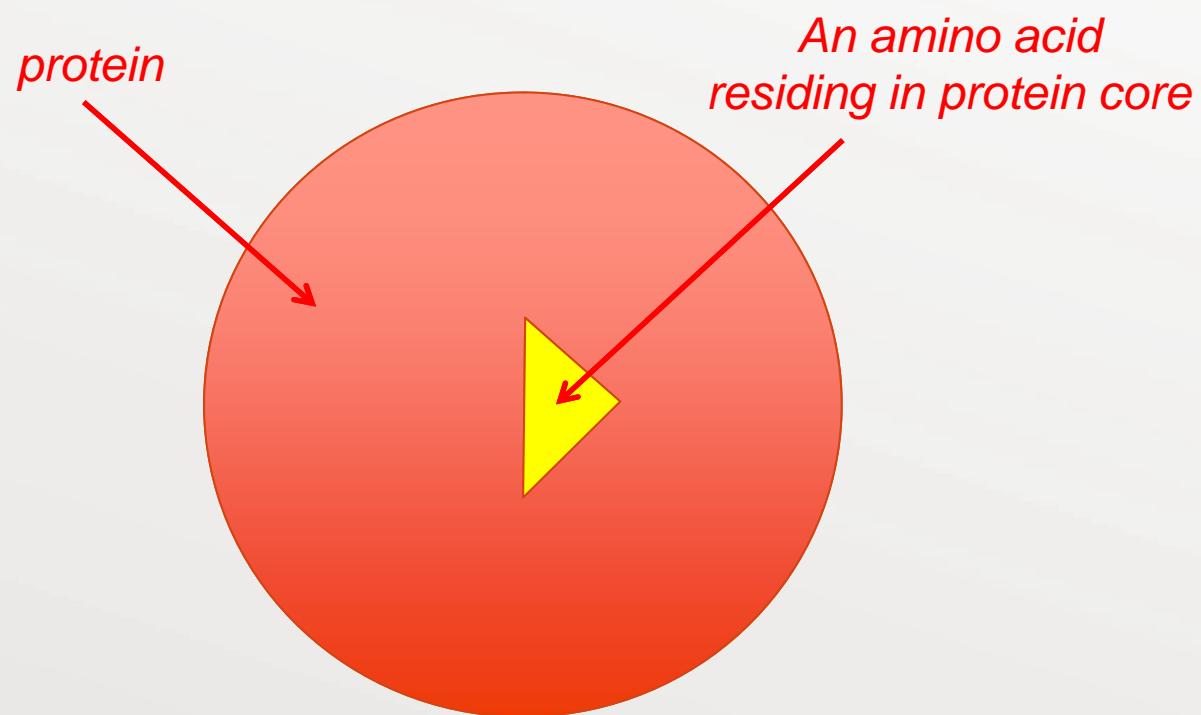
Influenza virus hemagglutinin



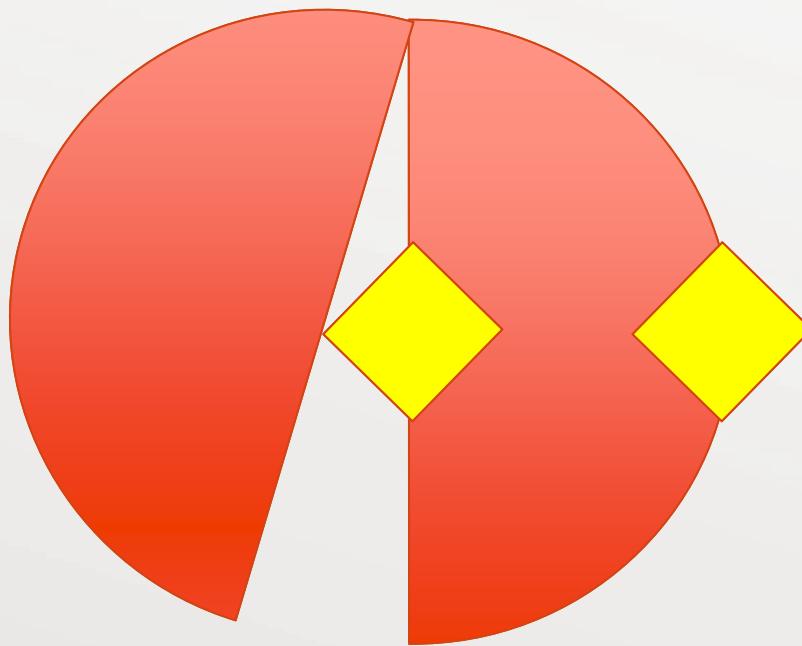
Buried residues (sites) evolve more slowly than exposed residues



# Amino acid substitutions in the core are disruptive

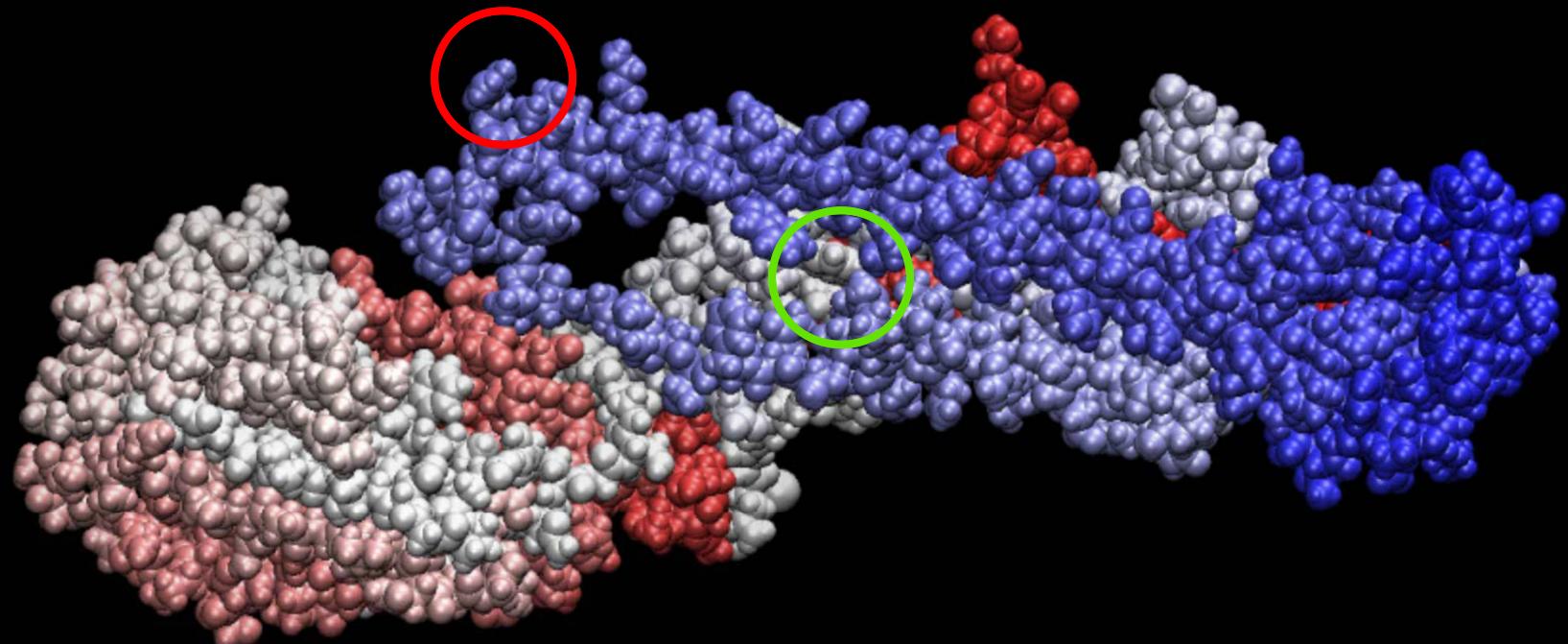


Amino acid substitutions in the core are disruptive

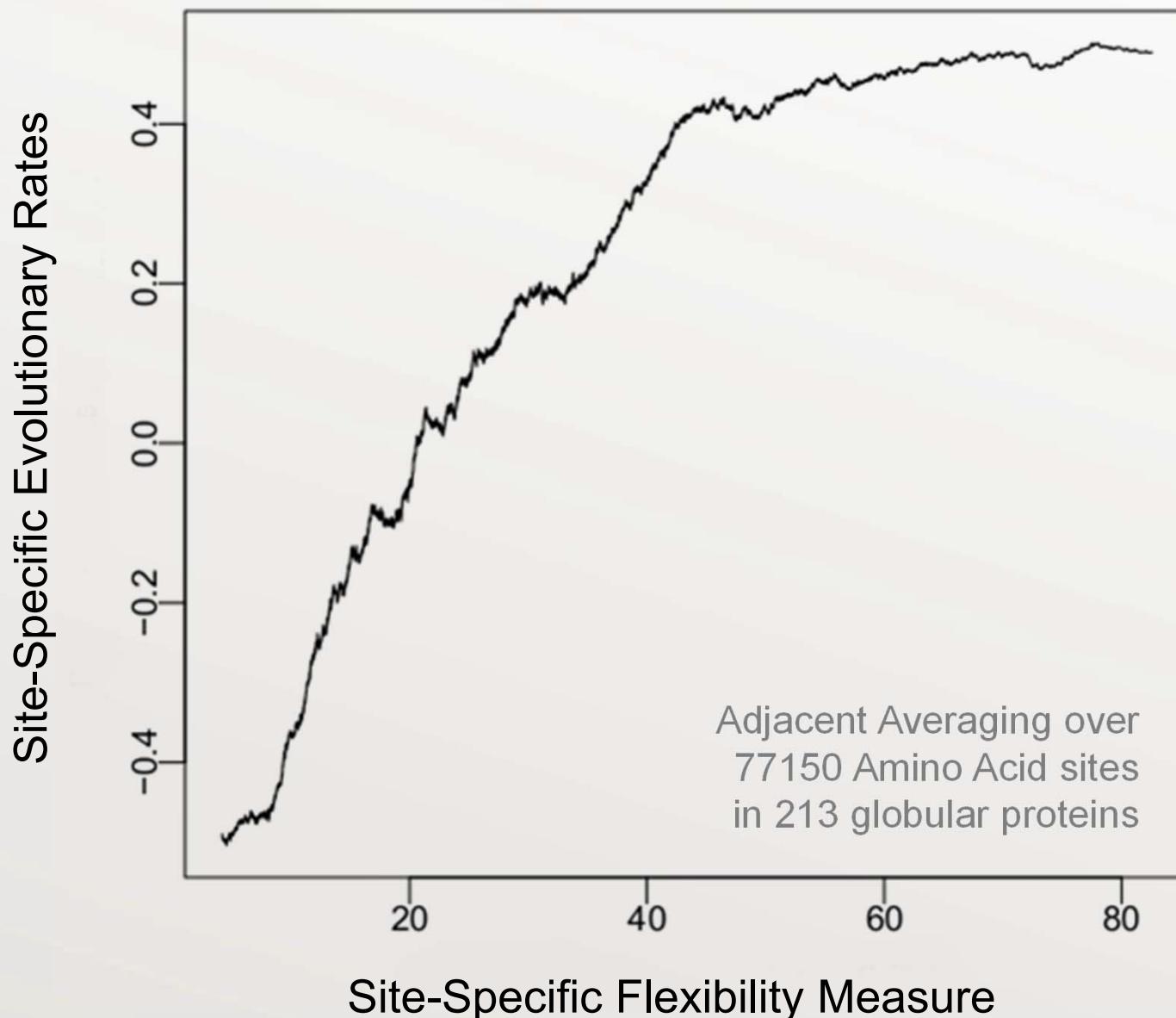


# Proteins are dynamic 3D entities

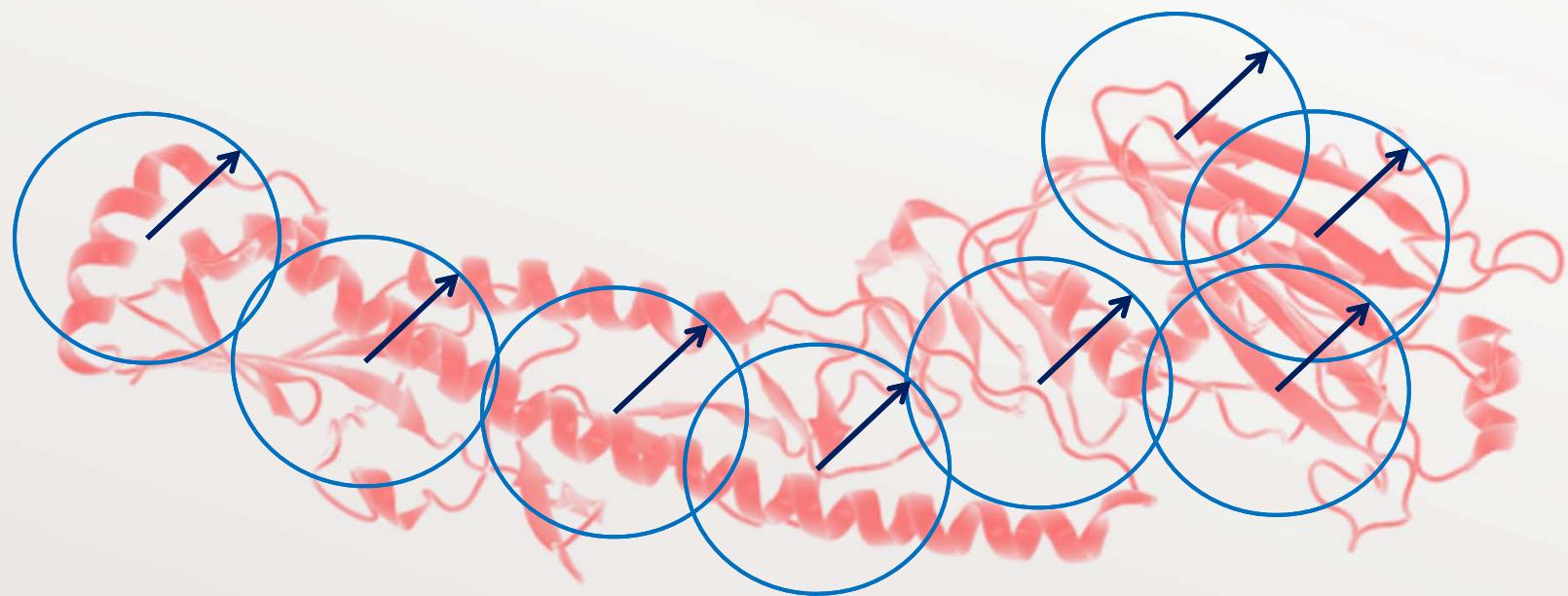
Influenza virus hemagglutinin



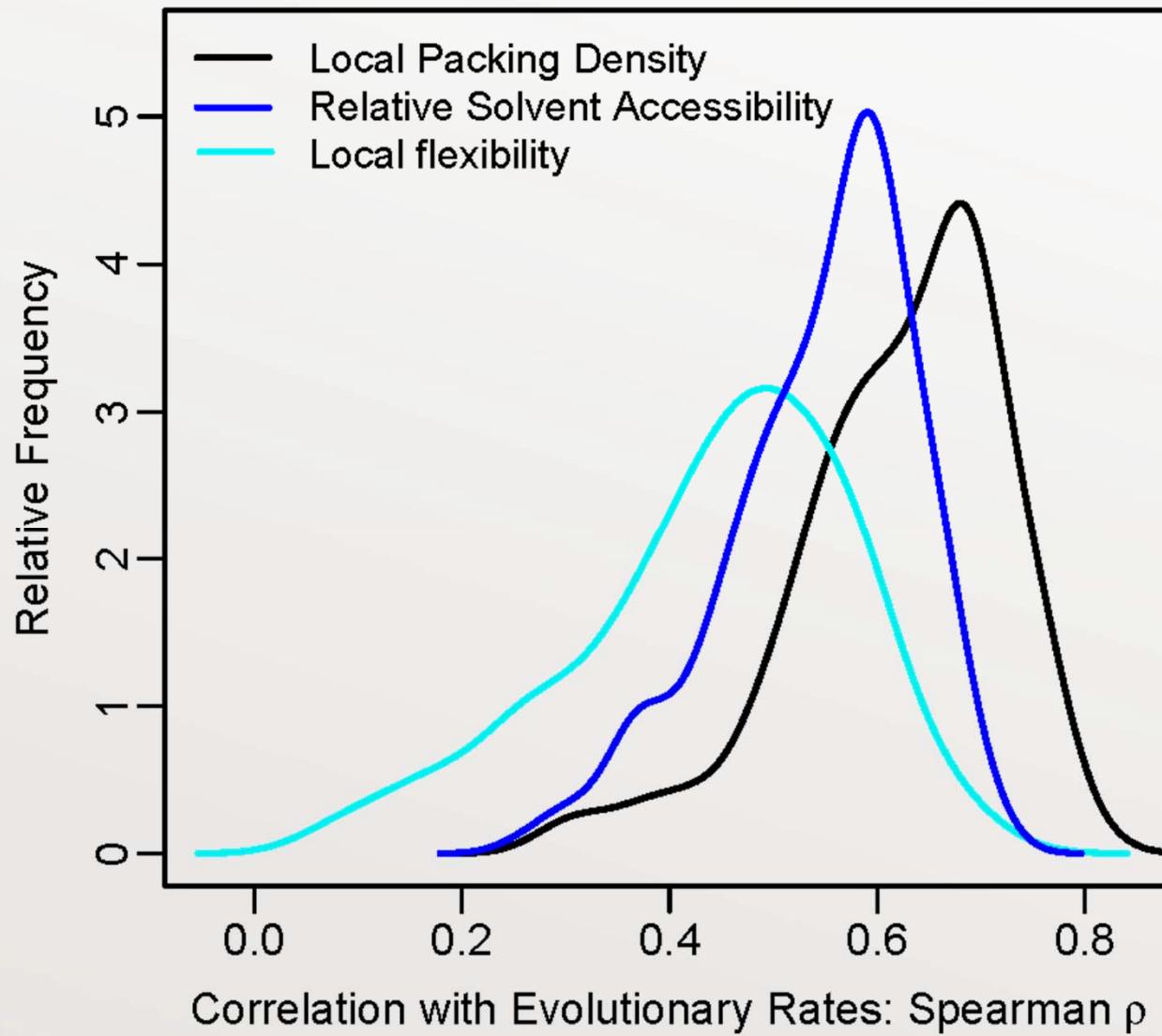
More flexible residues tend to evolve more rapidly



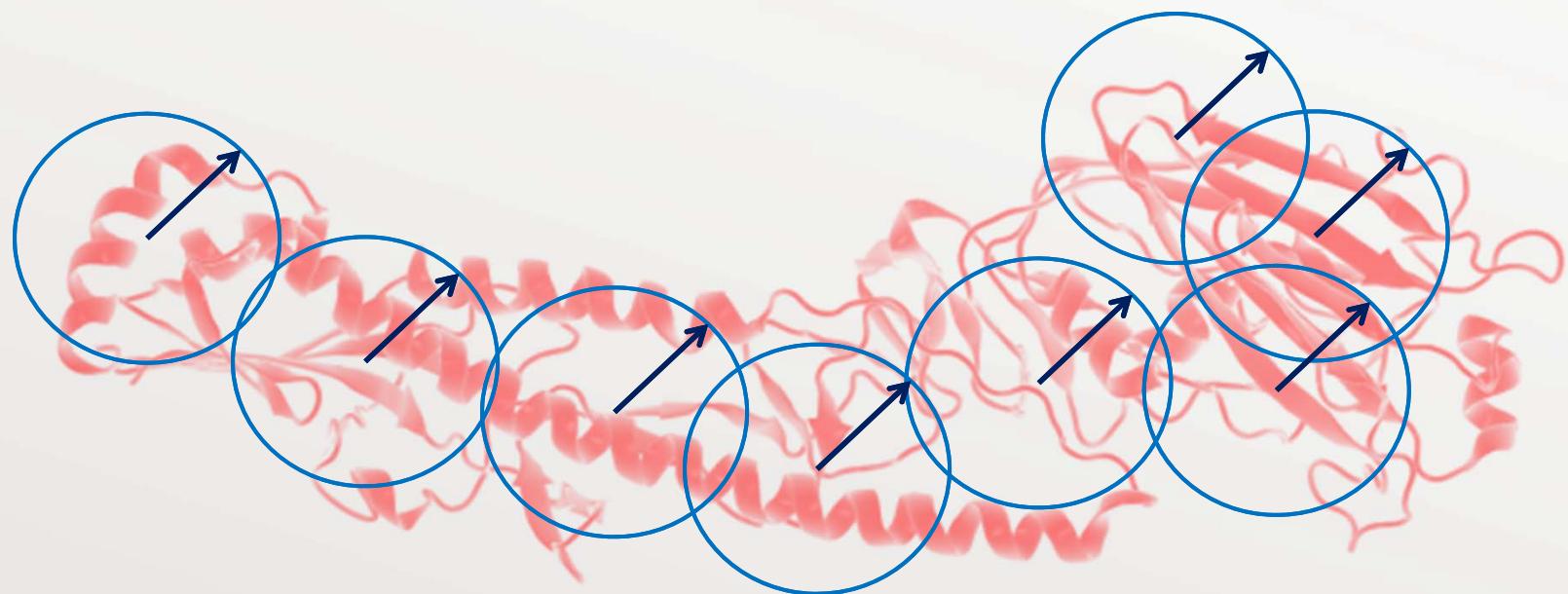
Denser regions in proteins also tend to evolve less frequently



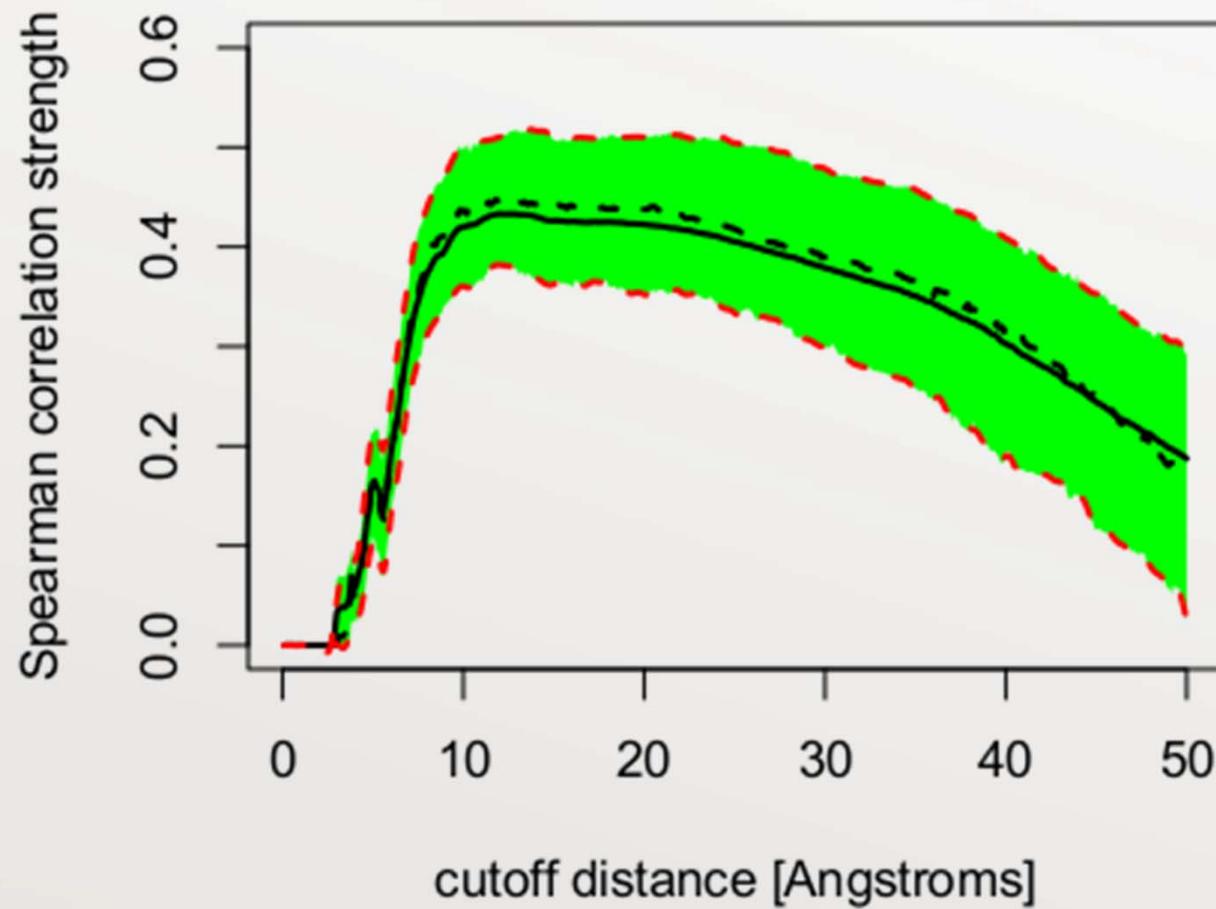
Local Packing Density outperforms other structural characteristics in predicting sequence evolution



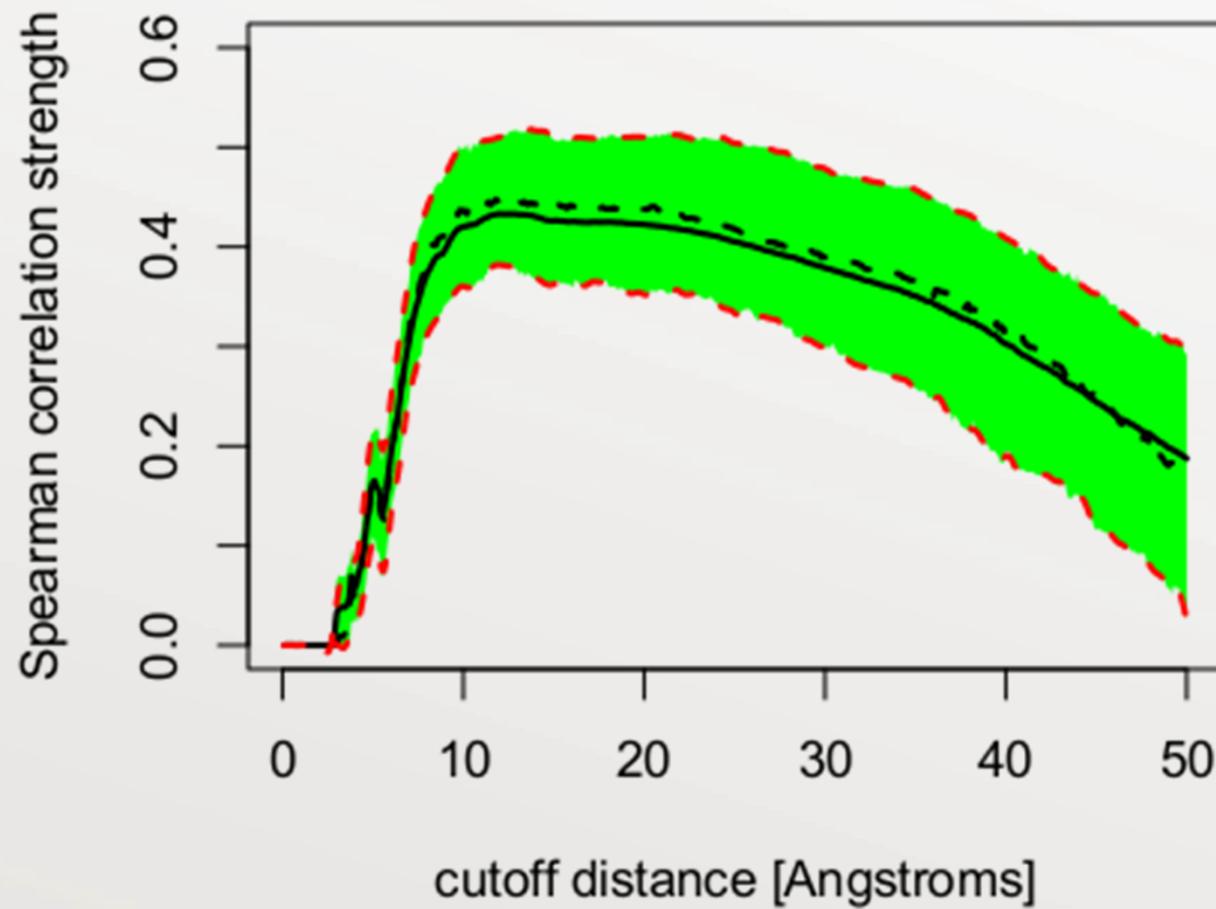
The traditional definitions of local packing density involve adjustable parameters: The radius of neighborhood



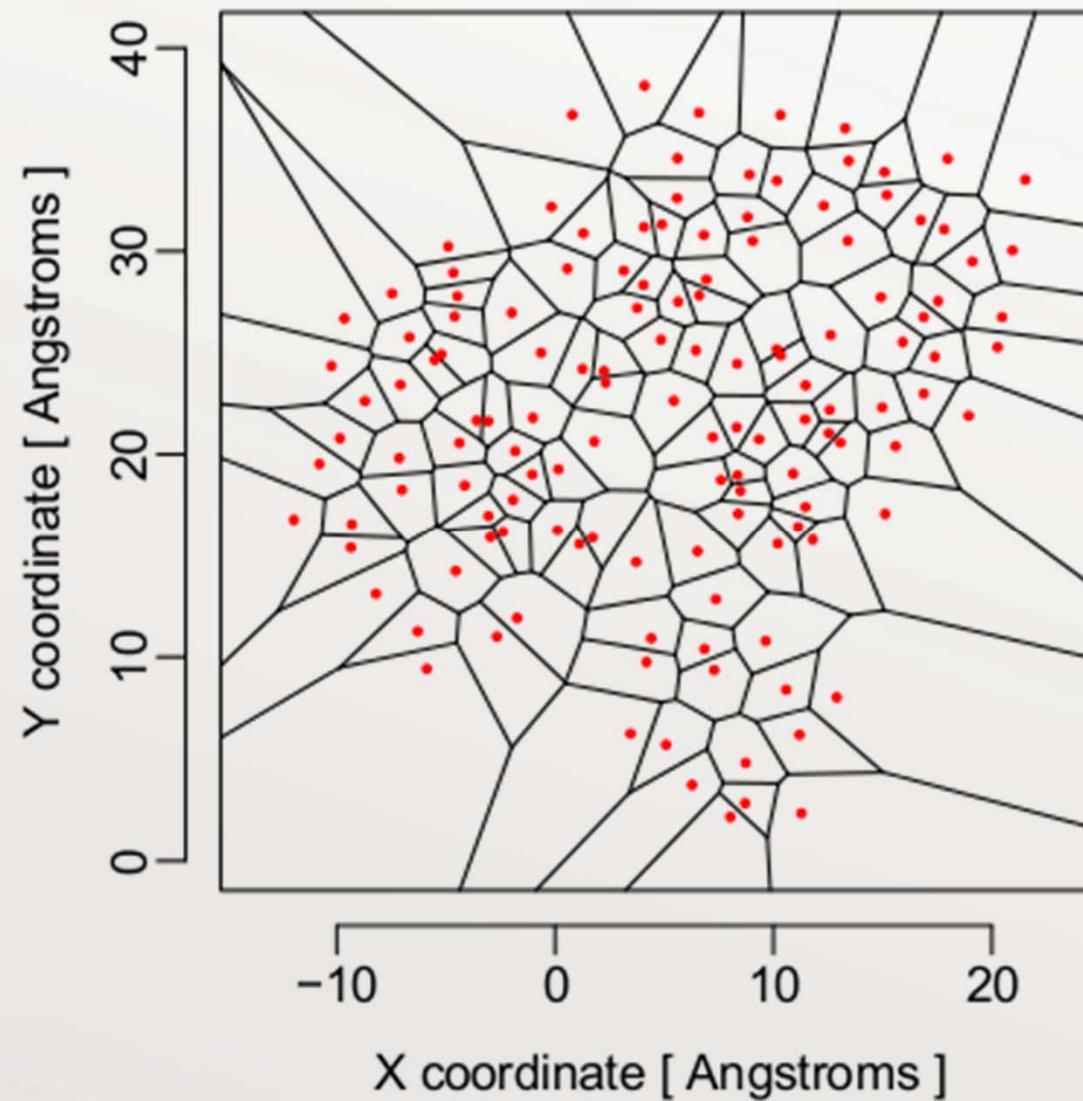
A neighborhood radius of 12-14 Angstroms results in the best predictions of sequence evolution



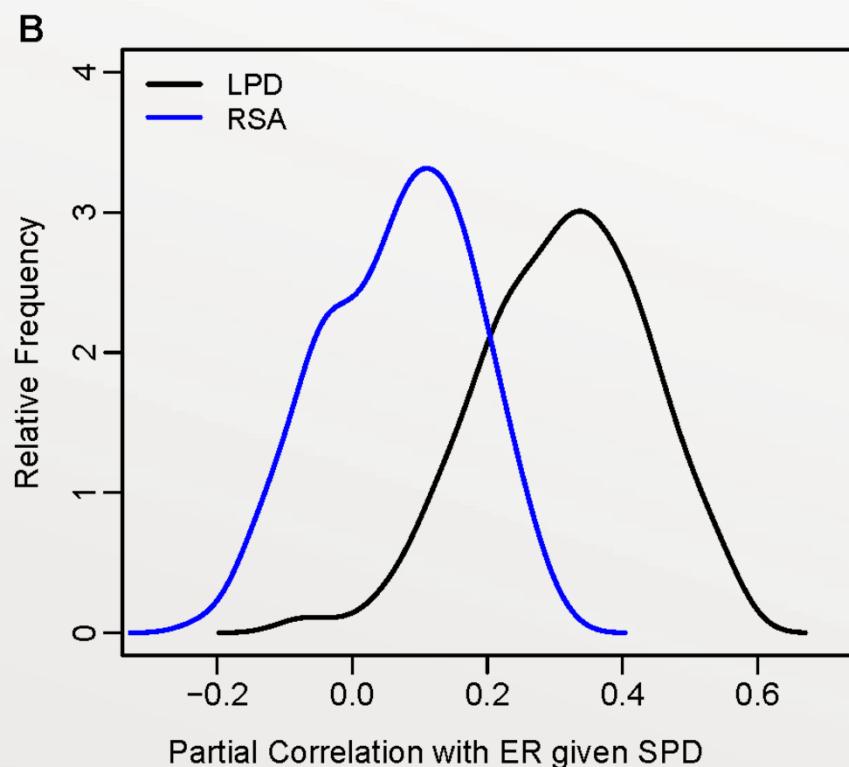
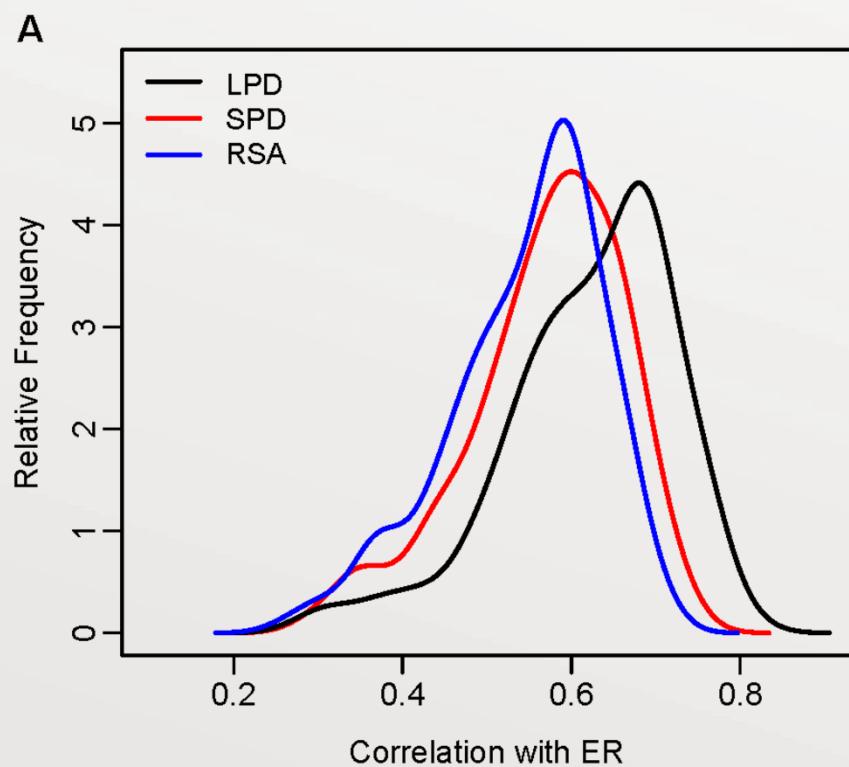
Non-local effects are important in sequence evolution



Voronoi tessellation provides a parameter-free measure of local packing density



Packing density, once corrected for non-local effects, has almost the same predictive power about sequence evolution as solvent accessibility (RSA)



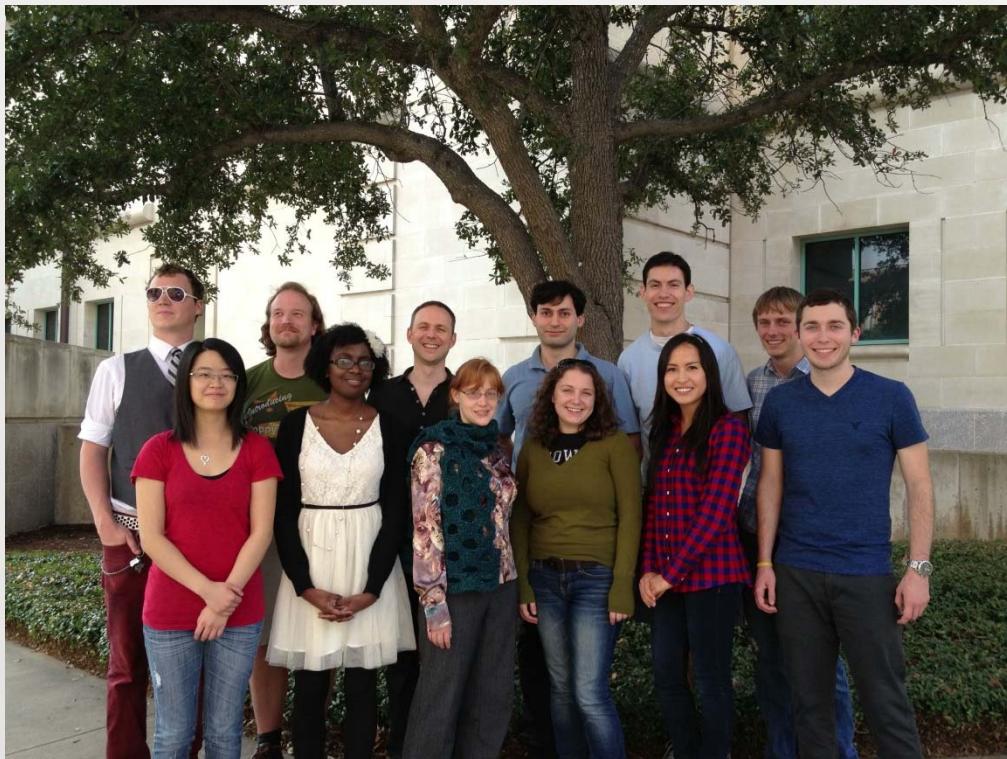
# Summary

- Packing density (excluding non-local effects) can on average explain 34% of sequence variability
- Non-local effects can explain on average 10% of sequence variability, independently of local packing density
- Solvent accessibility and local packing density (excluding non-local effects) represent the same local properties of sites in proteins
- Data available at: <https://github.com/shahmoradi/cordiv>
- Shahmoradi, Wilke, “Dissecting the roles of local packing density and longer-range effects in protein sequence evolution”  
<http://biorxiv.org/content/early/2015/07/30/023499>

## Acknowledgments

---

Claus Wilke



The WilkeLab members

# 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}  
}
```