Structural Variation and Site Variability in Proteins

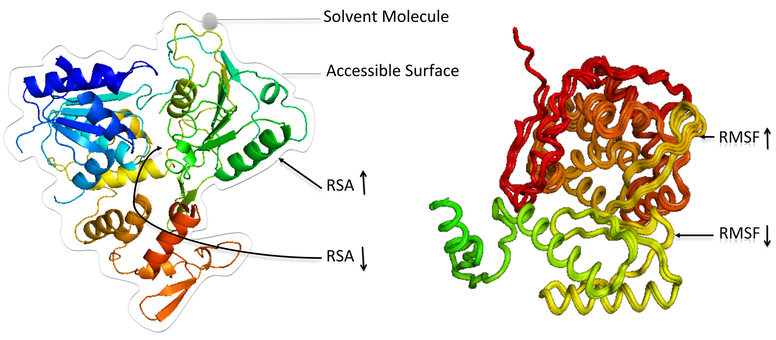
Dariya K. Sydykova, Amir Shahmoradi, Eric T. Dawson, Stephanie J. Spielman, Claus O. Wilke.
Institute for Cellular and Molecular Biology, The University of Texas at Austin.

Introduction

What does protein structure tell us about site variability? We seek to answer this question by analyzing the relationship between the variability at individual sites in alignments of viral sequences to properties of those sites in the three-dimensional structures of the corresponding proteins.

Structural Variation

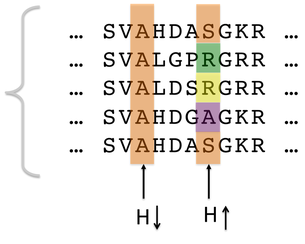
We considered two structural properties of proteins; relative solvent accessibility (RSA) and root mean square fluctuations (RMSF). RSA measures the surface area of a residue that comes in contact with solvent. Generally, residues that are more buried will have lower RSA, while the ones closer to the surface will have higher RSA. RMSF measures the structural variation at a site of superposed proteins. Areas with more structural differences will have higher RMSF, while the ones with fewer differences will have smaller RMSF.



RSA **RMSF**

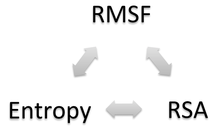
Site Variability

We capture site variability by measuring sequence entropy (H). Certain sites in protein alignments are more conserved than others. Those sites with less variation have lower entropy than the more variable ones.



Hypotheses

We know from prior work that RSA has a positive correlation with evolutionary rate, therefore we predict that RSA will have a positive correlation with sequence entropy. Similarly, RSA should correlate with RMSF. Because residues vary more on the surface, we also predict that RMSF will have a positive correlation with sequence entropy.



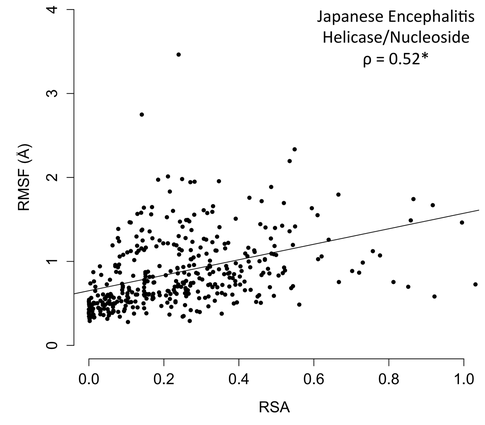
PDB Structure Availability

After searching the Protein Data Bank for homologous viral proteins, we found very few distinct structures.

Protein	PDB Structure	Total PDB Structures	Unique Sequences	Unique Sequences with Pairwise Difference ≥10%	≥5%	≥2%
HIV – 1 Reverse Transcriptase	2HMI_B	373	31	2	2	3
West Nile Protease	2FP7_B	68	9	2	6	6
Hepatitis C Protease	4AEX_A	410	20	4	5	10

RSA increases with RMSF

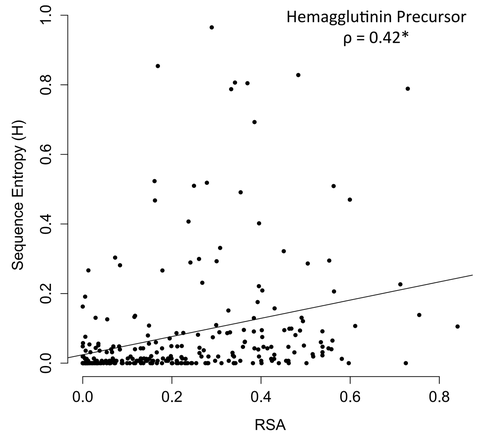
We observe that RMSF is positively correlated with RSA.



Protein	Spearman Correlation
Japanese Encephalitis Helicase/Nucleoside	0.52*
Hemagglutinin Precursor	0.38*
West Nile Protease	0.43*
Hepatitis C Protease	0.38*

Sequence Entropy Increases with RSA

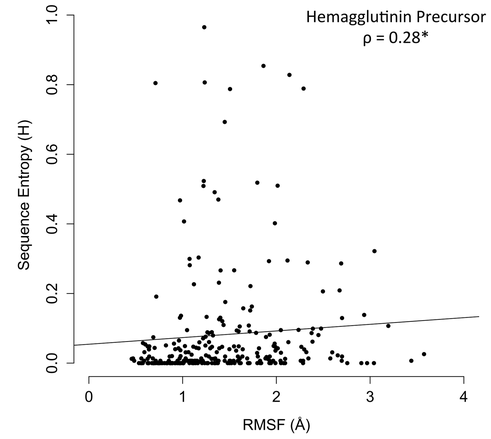
We observe that sequence entropy is positively correlated with RSA.



Protein	Spearman Correlation
Japanese Encephalitis Helicase/Nucleoside	0.26*
Hemagglutinin Precursor	0.42*
West Nile Protease	0.18
Hepatitis C Protease	0.26*

Sequence Entropy Increases with RMSF

We observe that sequence entropy is positively correlated with RMSF.



Protein	Spearman Correlation
Japanese Encephalitis Helicase/Nucleoside	0.16*
Hemagglutinin Precursor	0.28*
West Nile Protease	0.07
Hepatitis C Protease	0.23*