

Good Fences Make Distant Neighbors: Determining the Effect of Local Sequence Context on Pairwise Residue Interactions.



By: Connor Pitman¹, Matthew E. B. Hansen³, Grace Brannigan^{1,2}

¹Center for Computational and Integrative Biology, Rutgers—Camden, NJ, 08102, ²Department of Physics, ³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

Abstract

Residue-residue interactions extend beyond just oppositely charged residues; some hydrophobic residues can also undergo favorable interactions. One example of this kind of interaction occurs between two methionine residues. Recently, a study on the BDNF prodomain identified that the disease-causing effects of the Val66Met mutation stems from Met-Met interactions increasing the probability of contact between contiguously hydrophobic regions (termed “blobs”) not previously in contact [1]. However, there is a lack of knowledge about how widespread specialty hydrophobic interactions are in proteins. In this poster, we detect coevolving mutations with the goal of identifying intraprotein interactions in a large bacterial protein dataset of structured proteins. We find that coevolving residues are enriched in both residues found in h-blobs and between specialty hydrophobic residue pairs, and ultimately demonstrate that you and your neighbors can make (or break) whom you come into contact with.

Background and Approach

- A previous study done by our lab found that Methionine-Methionine (Met-Met) interactions in the mutant changed the frequency of contacts between methionine-containing blobs [1].
- Some hydrophobic residues can undergo specific interactions:

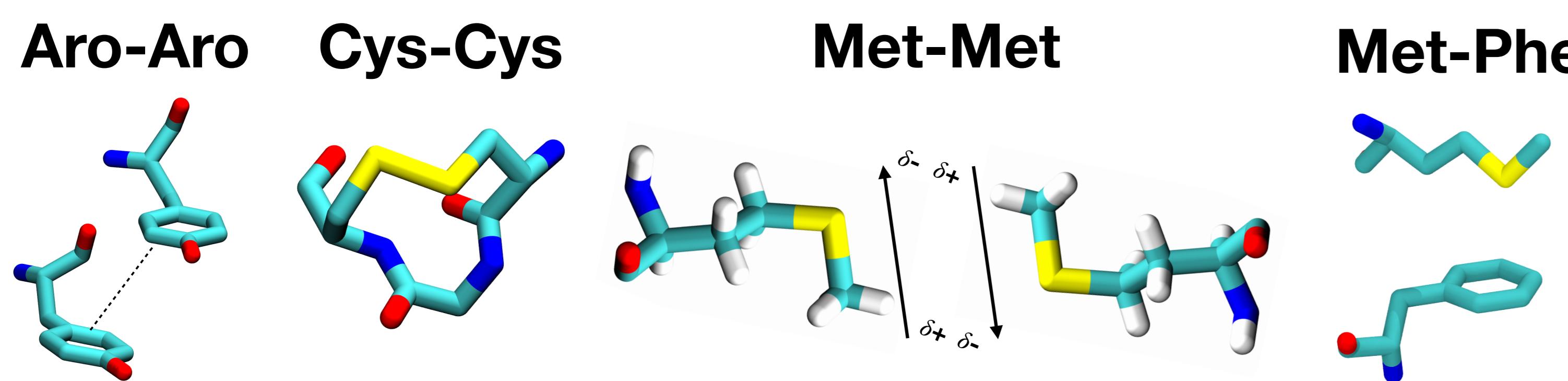


Figure 1: Pairwise specialty hydrophobic residue interactions.

- In a study on the structures of all proteins in the protein databank (PDB), hydrophobic residues with special properties were found in contact with other residues of the same amino acid type, shown in the table below (reproduced from [2]):

AA	Near Contacts
C	All AA
M	M, F
W	R, C, Q, G, Y, H, K, M, F, P, S, W
Y	P, F, M, K, L, I, E
F	Y, W, L, I

- Identifying contacts in proteins has been done using either solved protein structures or through simulations, but is there a more efficient way to detect contacts?
- Coevolving residues are often found in contact [3]
- However, coevolving doesn't necessarily mean spatially in contact

Research Questions:

- Do Residues in the same blob type tend to coevolve?
- Do specialty residue interaction pairs (Aro-Aro, Met-Met, Cys-Cys) tend to coevolve?

Approach:

Detecting Coevolving Pairs in a large Bacterial protein dataset (1630 protein families, with ~229 orthologs per family) of structured proteins using CoMap [4].

Calculating Enrichment of specialty interaction pairs in this dataset (Equation 1).

Blobulation

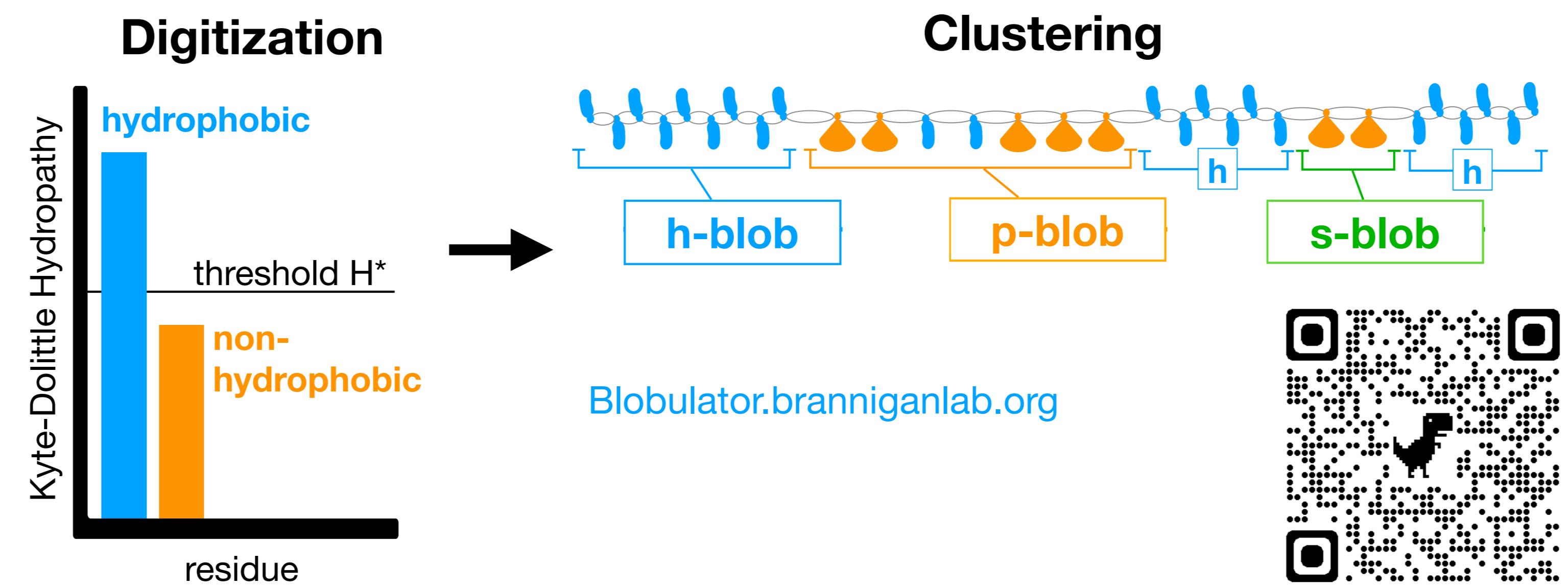


Figure 2: Blobulation, our sequence-based algorithm for generating topologies for IDPs and other proteins, involves two steps: digitization (left), and clustering (based on contiguous hydrophobicity). Figure adapted from [5].

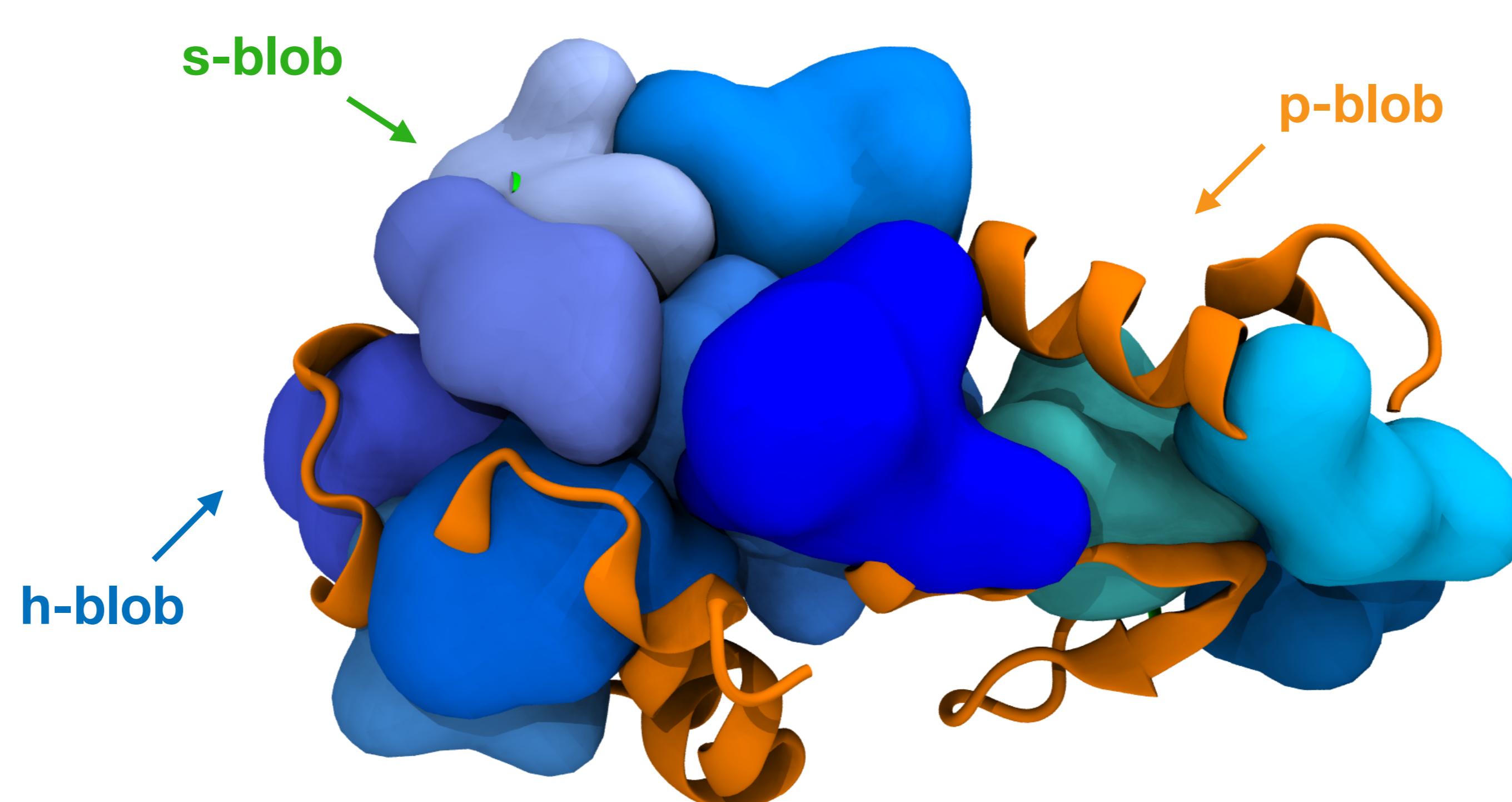


Figure 3: Molecular image of Lysozyme colored by blob type (Uniprot: P00720, PDB:2LZM). Lysozyme was blobulated using the settings $H^* = 0.4$ and a minimum length threshold of 4. H-blobs are blue, p-blobs are orange, and s-blobs are green.

Enrichment

$$N_{ab} = \sum_{J \in \text{Ortholog}} \sum_{K \in \text{Family}} n_{JK}[a, b]$$

$$f_{ab} = \frac{N_{ab}}{\sum_{ab} N_{ab}} \quad f_{ab}^* \text{ is the same, but restricted to significant coevolving pairs}$$

$$\text{Enrichment} = \frac{f_{ab}^*}{f_{ab}} \quad \text{Equation 1: Formula used to calculate the enrichment of coevolving pairs}$$

References

- Ruchi Lohia, Reza Salari, and Grace Brannigan. Sequence specificity despite intrinsic disorder: How a disease-associated val/met polymorphism rearranges tertiary interactions in a long disordered protein. *PLOS Computational Biology*. 15(10):e1007390, oct 2019.
- Faure G, Bonnot A, de Brevern AG. Protein contacts, inter-residue interactions and side-chain modelling. *Biochimie*. 2008; 90(4):626–639. <https://doi.org/10.1016/j.biochi.2007.11.007> PMID: 18086572
- Shilpi Chaurasia and Julien Y. Dutheil. The structural determinants of intra-protein compensatory substitutions. *Molecular Biology and Evolution*, March 2022.
- Duthell, J., Galtier, N. Detecting groups of coevolving positions in a molecule: a clustering approach. *BMC Evol Biol*. 7, 242 (2007). <https://doi.org/10.1186/1471-2148-7-242>

Identifying Coevolving Residue Pairs

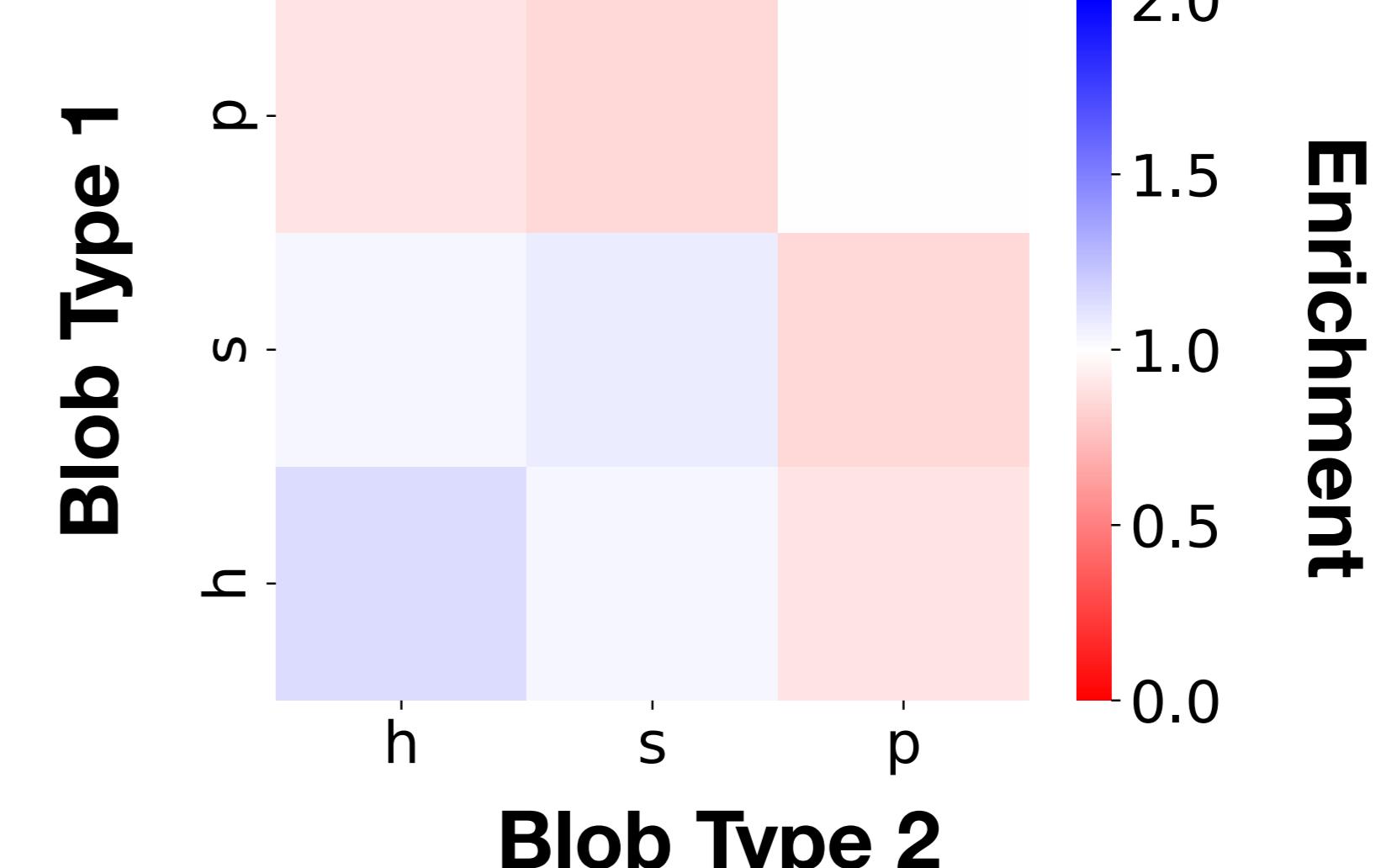


Figure 4: Enrichment of coevolving residues by blob type. Enrichment was calculated as in Equation 1, with each residue being classified by the blob that contains it.

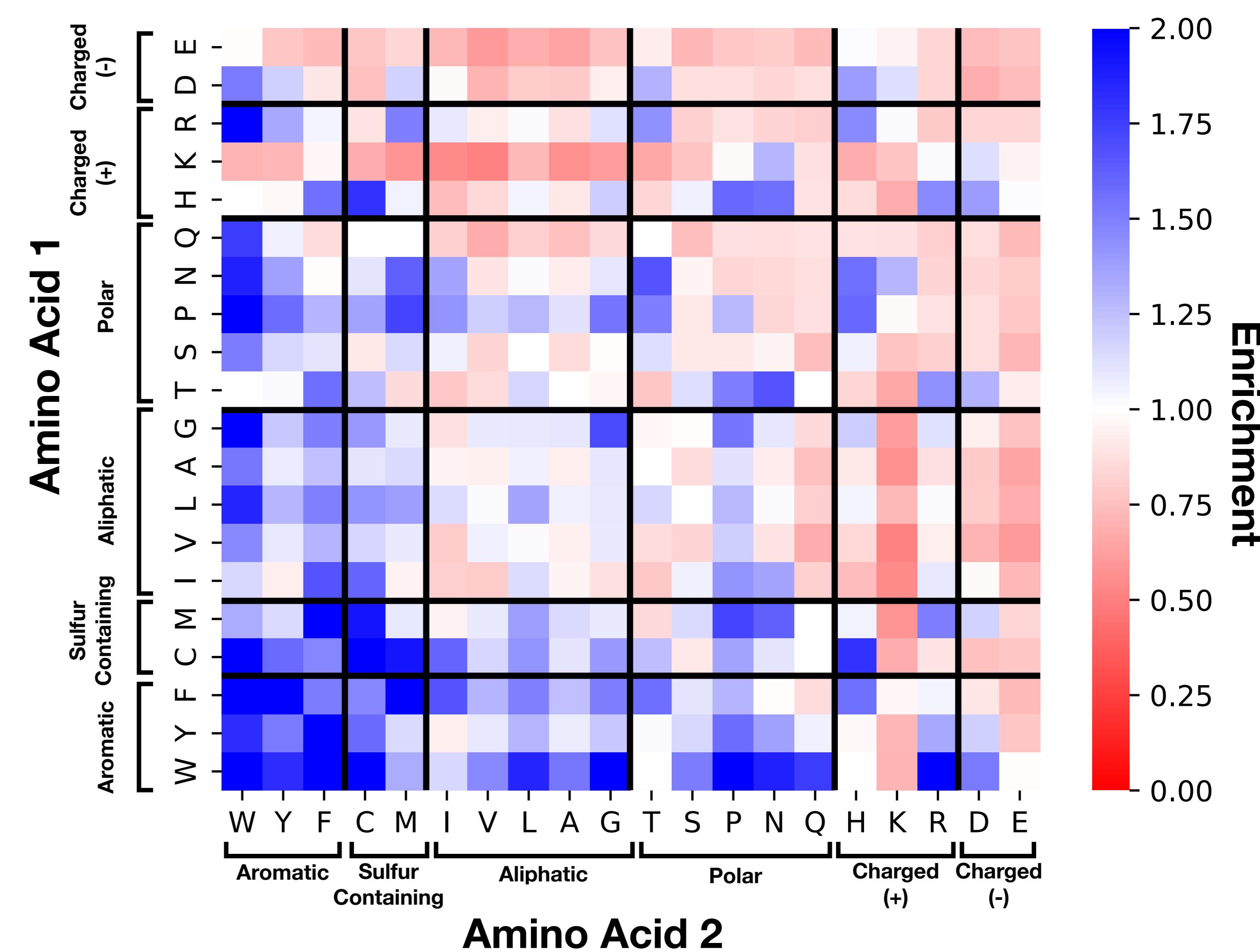


Figure 5: Enrichment of coevolving amino acid residue pairs. Residues are labeled by their biochemical properties. Enrichment was calculated as in Equation 1.

Summary

- Coevolving residues are enriched in specialty residue interaction pairs, though not very strongly enriched for Met-Met pairs
- Coevolving residues are enriched in h-h blob pairs and slightly enriched in h-s blob pairs

Acknowledgements

- Rutgers Office of Advanced Research Computing (OARC)
- Busch Biomedical Foundation
- NRT, NSF DGE 2152059