

# Investigating the Link Between Intra-protein Interactions and Contiguous Hydrophobicity

By: Connor Pitman<sup>1</sup>, Anthony Geneva<sup>1,2</sup> Matthew E. B. Hansen<sup>4</sup>, Grace Brannigan<sup>1,3</sup>

<sup>1</sup>Center for Computational and Integrative Biology, Rutgers—Camden, NJ, 08102, <sup>2</sup>Department of Biology, <sup>3</sup>Department of Physics, <sup>4</sup>Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

## Abstract

Previously our lab has found that in proteins, clusters of hydrophobic residues ("hydrophobic blobs") are likely to form tertiary interactions even in the absence of structure and are more likely to contain disease-associated mutations across the human exome. However, whether certain individual residues and blobs are interacting with each other remains unclear. Identification of coevolving residues allows for an alternative approach to detecting residue-residue interactions using only protein sequence data. Here, we test whether the coevolution of certain residue pairs is context-dependent, and whether there is a difference between residues both found in the same blob types and those found in opposite blob types

## Research Questions

### Main Question:

Are evolutionary signatures for residue-level interactions correlated with hydrophobic clusters in bacterial orthologues?

### Sub-Questions:

1. Are hydrophobic blobs enriched for intra-protein coevolving sites?
2. Does the amino acid composition of coevolving sites differ between hydrophobic and non-hydrophobic blobs?

## Approach

1. Detected coevolving sites in a large Bacterial protein dataset (1630 protein families, with ~229 orthologs per family - previously used to investigate the role of structure in coevolution) using CoMap [1]
2. Blobulated all protein sequences (as in Figure 3)
3. Calculated enrichment of coevolving residues for all blob type pairings, and for all amino acid type pairings among varying blob types ("contexts"). Null expectation was generated using a permutation test. All detected coevolving sites were shuffled, with the amino acids found at each remaining unshuffled.

## Summary

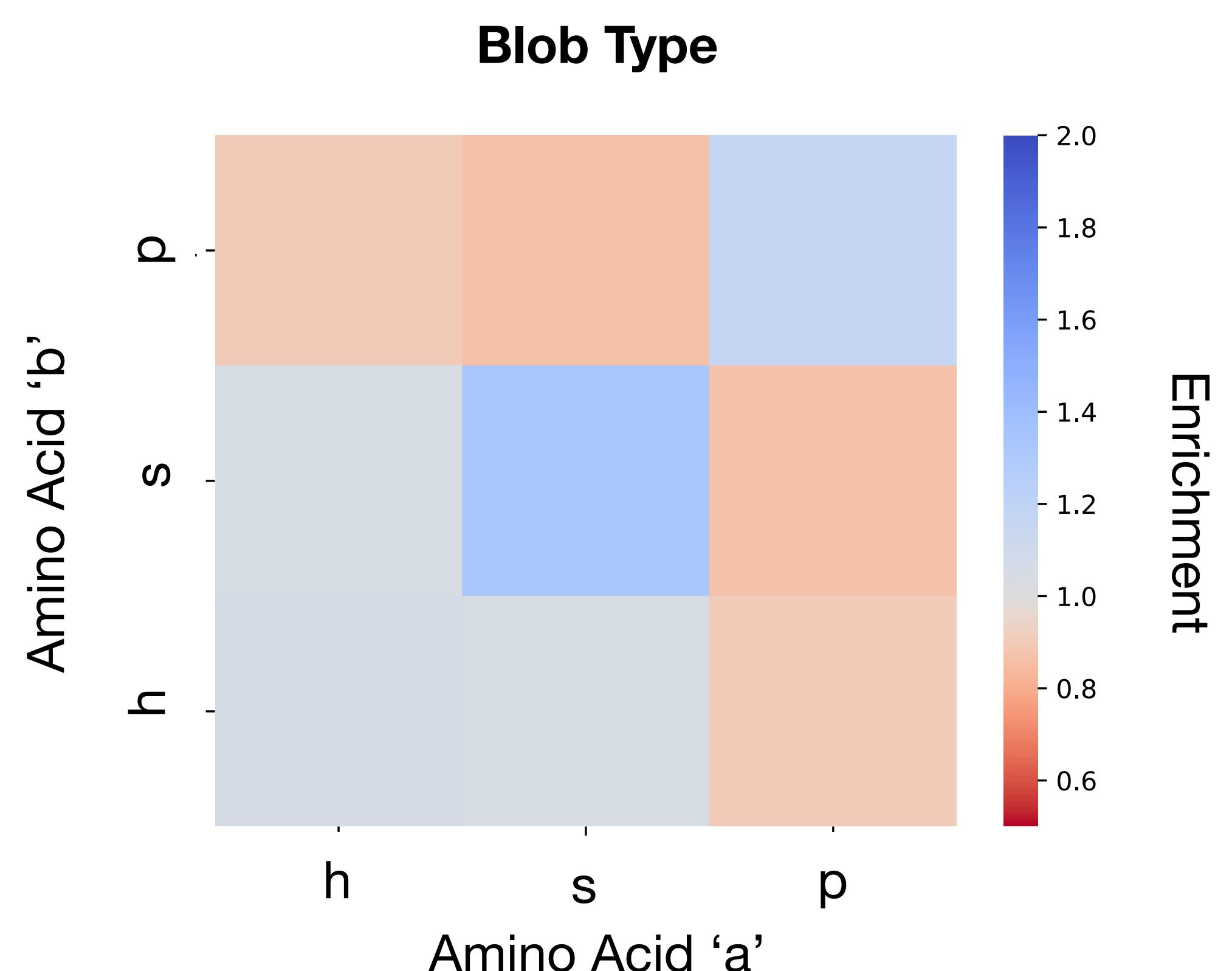
1. Pairs of similarly hydrophobic blobs (h-h, s-s, and p-p), as well as hydrophobic and short blobs (h-s), are enriched for coevolving residues
2. Pairs found in the same context (h-h and p-p) tend to be enriched for highly polar residues and some highly hydrophobic residues than those in opposite contexts (h-p), but there are notable exceptions (C-C, C-Y, and some aromatic pairs)
3. The amino acid composition among various blob pairs is distinct, though some residue pairings (ex. F-W, F-Y, and K-E) are enriched across all blob pairs

## Enrichment of Coevolving Pairs

$N_{ab}^{\text{obs}}$  = Number of detected coevolving pairs 'ab'  
 $N_{ab}^{\text{perm}}$  = Null frequency of pair 'ab', generated by random of sites (as in Approach)

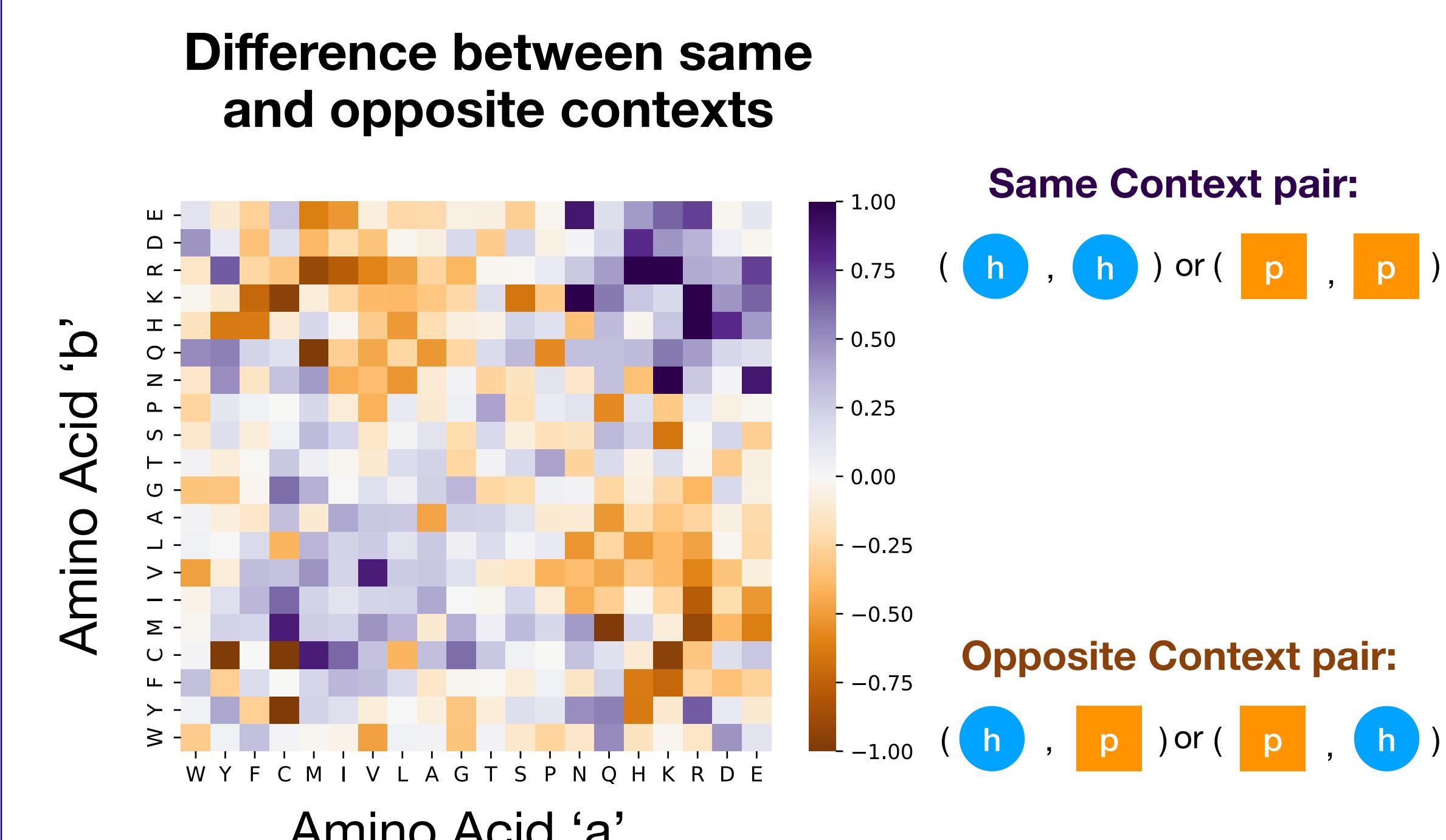
$$\text{Enrichment} = \frac{N_{ab}^{\text{obs}}}{N_{ab}^{\text{perm}}}$$

## Are hydrophobic blobs enriched for intra-protein coevolving sites?



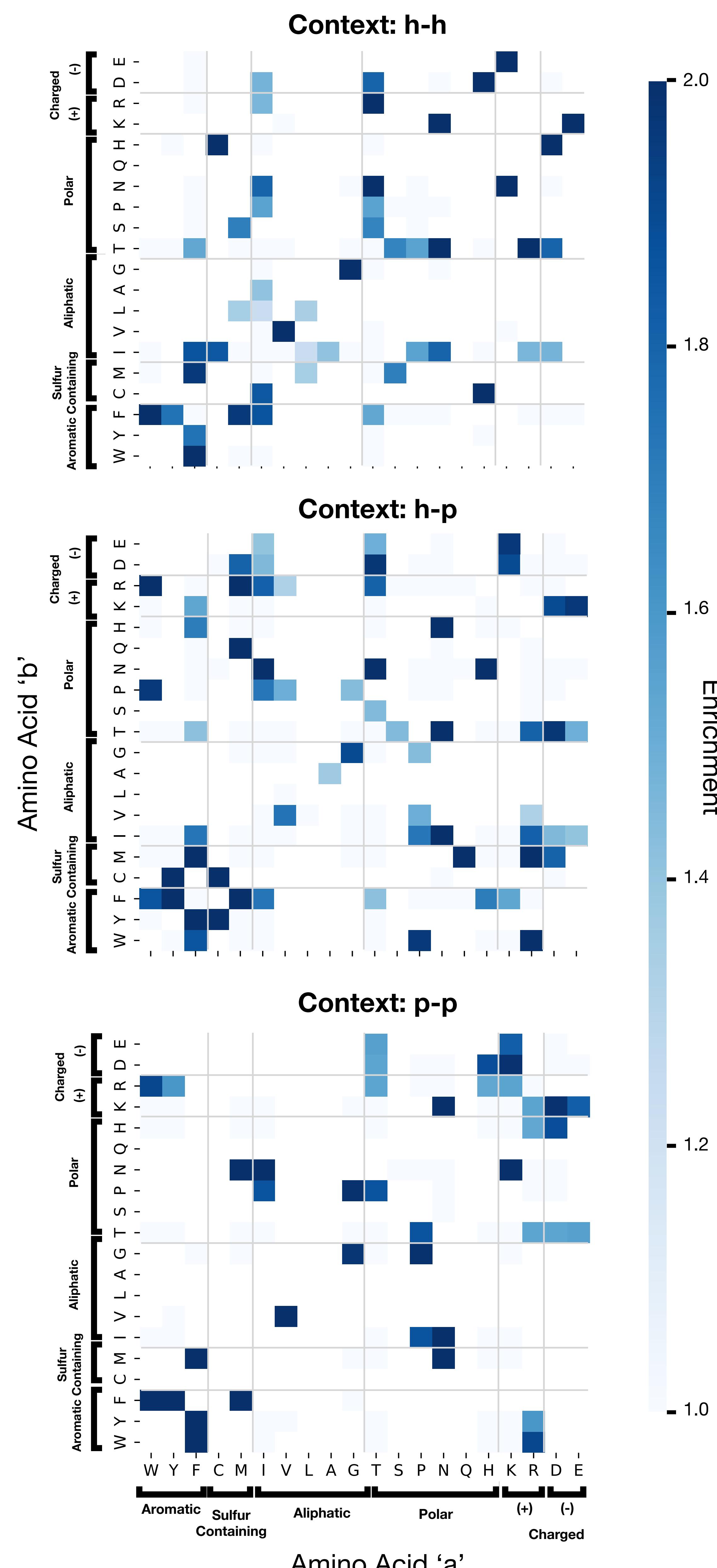
**Figure 1 :** Enrichment of each blob-type pairing for coevolving residues. Blobulation was done using a hydrophobicity threshold of 0.4 and a minimum length requirement of 4.

## Does the amino acid composition of coevolving sites differ by blob type?

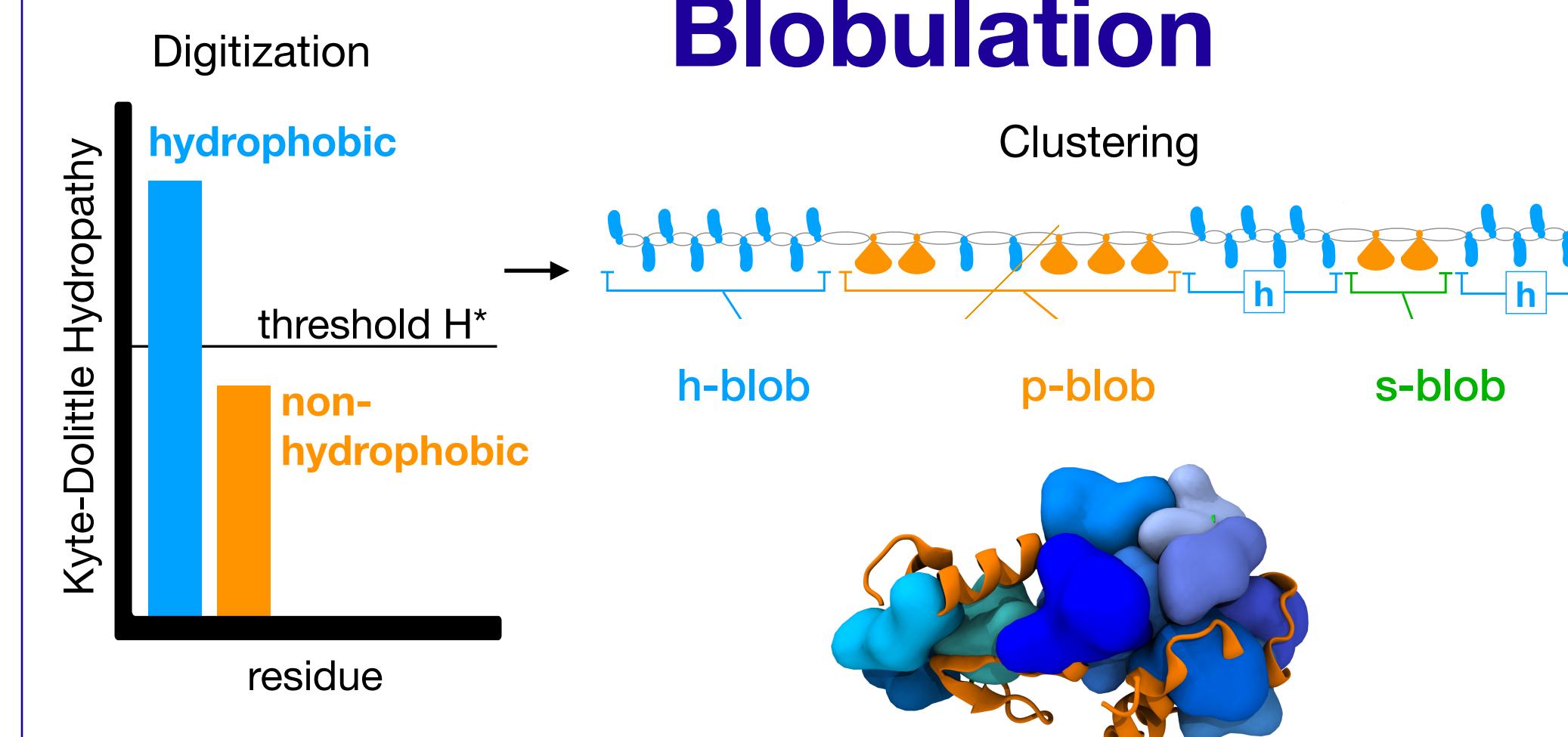


**Figure 2 :** Difference in Enrichment between residues found in the same (h-h and p-p), and opposite (h-p and p-h) contexts. Blobulation was done using the same settings as Figure 1.

## Does the amino acid composition of coevolving sites differ by blob type?



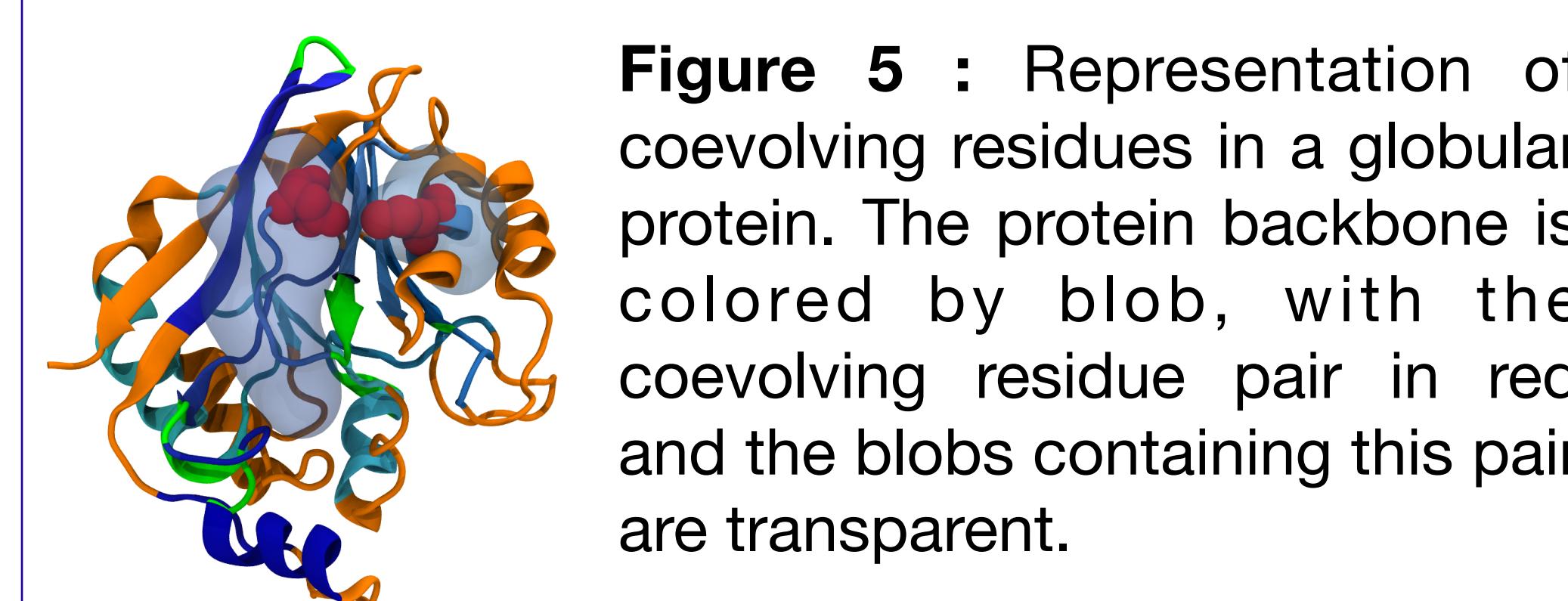
**Figure 3 :** Enrichment of coevolving amino acid residue pairs as in 1. Residue pairs are further broken down by the blob that contains each. Blobulation was done using the same settings as Figure 1. Grey lines delineate between aromatic, sulfur containing, aliphatic, polar, and charged residues.



**Figure 4:** Blobulation, our algorithm for detecting intrinsic modularity in protein sequences based on hydrophobicity. The algorithm involves two steps: digitization using hydrophobicity threshold  $H^*$  (left), and clustering (middle). Figure adapted from [2]. Example representation made in VMD of Lysozyme blobs (right, Uniprot: P00720, PDB:2LZM).

## Detecting interactions from sequence

- Pairs of residues found at coevolving sites (two positions in orthologous proteins consistently co-occurring across evolutionary history) are often found in contact [1]
- Using coevolution allows us to infer contacts using only protein sequences
- Additionally, the properties such as hydrophobicity and the charge class of residue groups provide information about a protein's ensemble [3, 4]
- Highly charged regions either attract or repel each other, while neutral regions tend to be globular [3]



**Figure 5 :** Representation of coevolving residues in a globular protein. The protein backbone is colored by blob, with the coevolving residue pair in red and the blobs containing this pair are transparent.

## Acknowledgements

- Rutgers Office of Advanced Research Computing (OARC)
- NRT, NSF DGE 2152059
- NIH 1R35GM134957



Blobulatorbranniganlab.org  
Blobulate your own protein here!

## References

- [1] S. Chaurasia and J. Dutheil. Molecular Biology and Evolution, 2022.
- [2] R. Lohia, M. Hansen, G. Brannigan. PNAS, 2022.
- [3] R. Das and R. Pappu. PNAS, 2013.
- [4] V. Uversky, J. Gillespie, A. Fink. Proteins, 2000.