

Untangling the role of contiguous hydrophobicity and specialty hydrophobic interactions in residue coevolution



By: Connor Pitman¹, Anthony Geneva^{1,2} Matthew E. B. Hansen⁴, Grace Brannigan^{1,3}
¹Center for Computational and Integrative Biology, Rutgers—Camden, NJ, 08102, ²Department of Biology, ³Department of Physics, ⁴Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA



Abstract

Co-variation of sites across a phylogeny is a powerful means to detect coevolution between positions in a protein sequence. Coevolving pairs have been integrated into many sequence-based studies and methods to identify pairwise interactions that are critical for protein structure and function. Previous research on the amino acids that coevolve has focused on pairs of residues that undergo mechanically straightforward pairwise interactions, as in electrostatic interactions or disulfide bonds. Less attention has been paid to other kinds of pairwise interactions, particularly between non-aliphatic hydrophobic residues. Additionally, identifying the “local sequence context” surrounding coevolving pairs has been intractable using conventional definitions, as they tend to be found outside of secondary structure elements. Here, we detect coevolving residues across a bacterial dataset composed of ~1600 protein families, and test whether pairs of non-aliphatic amino acids are as likely to be coevolving as oppositely charged pairs or disulfide-bonded pairs. We use blobulation to define the “local hydrophobic context” surrounding coevolving residues, segmenting protein sequences into hydrophobic and polar regions (“blobs”). We have previously found that hydrophobic blobs are enriched disease-causing mutations, despite weak correlation between h-blobs and secondary structure. We find that coevolving pairs undergo a variety of previously uninvestigated pairwise residue interactions, and that they tend to be found in pairs of blobs of the same type – either both in hydrophobic blobs, or both in polar blobs. Additionally, we find that hydrophobic blobs contain many types of coevolving pairs, including polar and oppositely charged residues; while coevolving pairs in polar blobs tend to be limited to oppositely charged and a few other polar residue pairs. Ultimately we observe that coevolving amino acid pairs are associated with certain blob types, indicating the need to account for the hydrophobic context in protein evolutionary studies.

Coevolving Positions

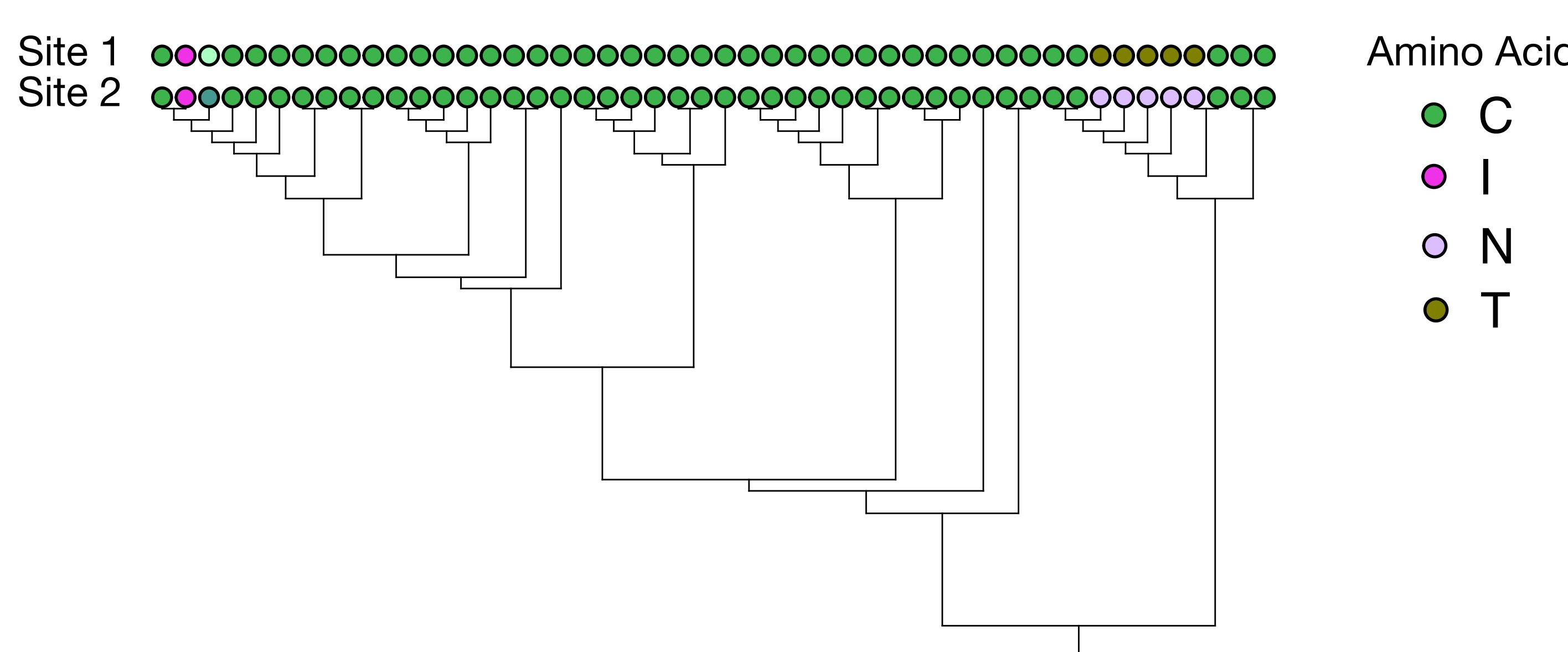


Figure 1: Coevolving positions. Pairs of positions that co-vary across evolutionary time are considered coevolving. The amino acid identities of two coevolving sites across an example set of species are shown.

- Coevolving positions (Figure 1) are residue pairs that evolve in tandem, often indicating structurally and functionally important residue contacts.
- Previous work has largely focused on oppositely charged pairs and disulfide bonds
- It is unknown whether non-aliphatic hydrophobic residues play a role in protein coevolution.
- The role of the local sequence context has been intractable – coevolving pairs tend to be found outside of secondary structure elements [1].

Favorable pairwise amino acid interactions

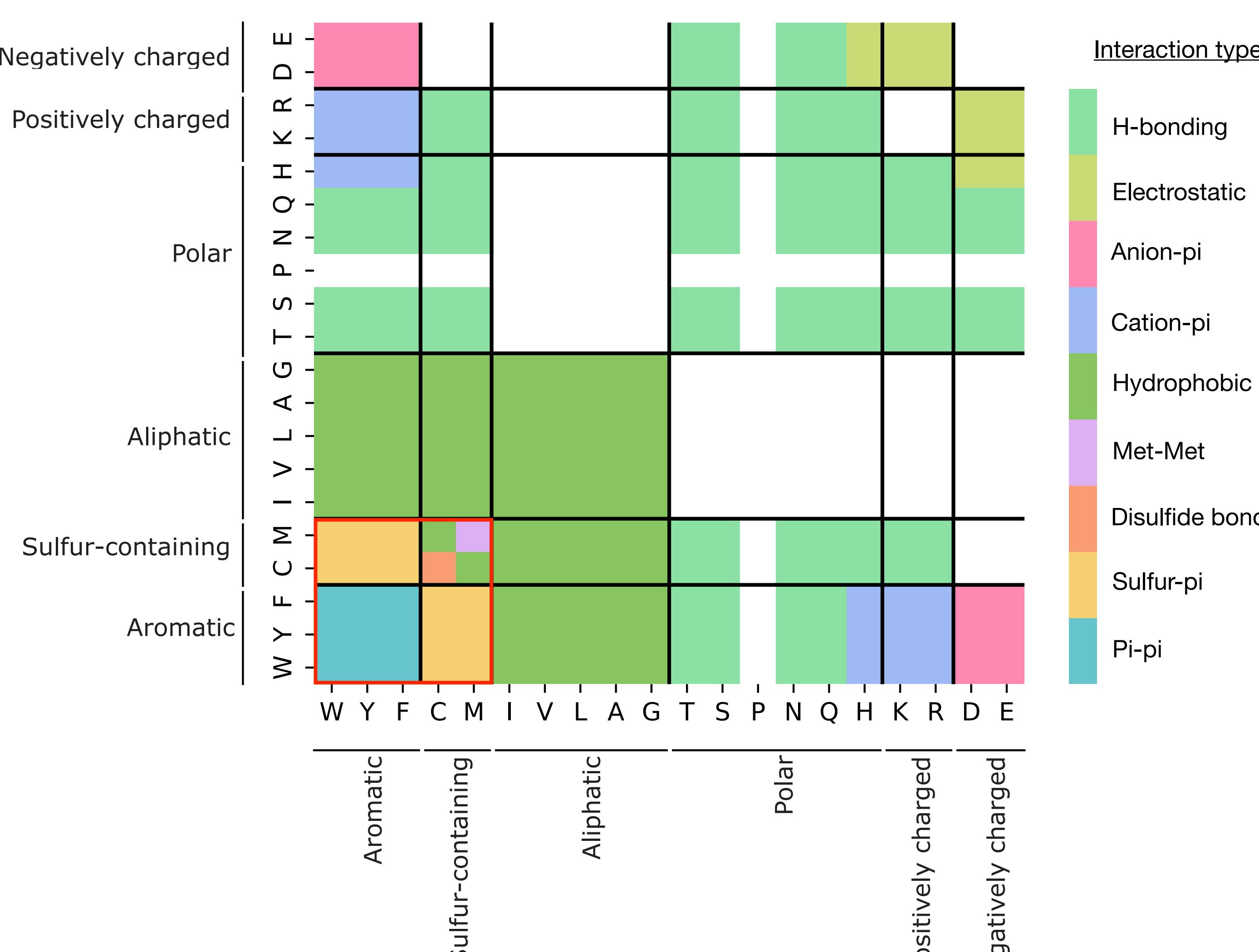


Figure 2: Schematic showing favorable interactions that occur between amino acid pairs. Interactions between non-aliphatic hydrophobic residues are indicated by a red box.

Blobulation

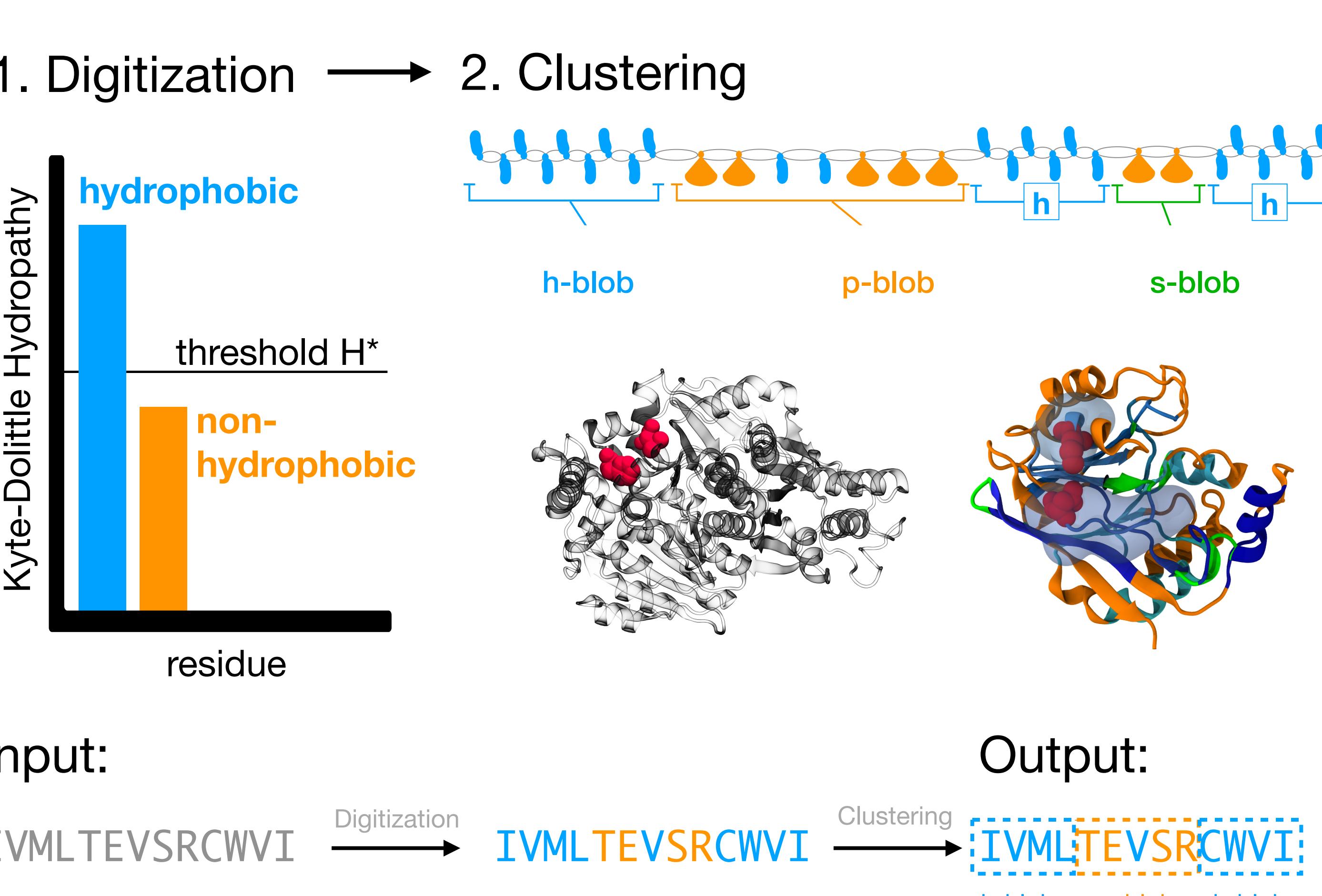


Figure 3: Blobulation, our algorithm for detecting intrinsic modularity in protein sequences based on hydrophobicity. The algorithm involves two steps: digitization using hydrophobicity threshold H^* (left), and clustering (middle). Figure adapted from [2]. Example of coevolving residues (red) in an unblobulated protein (left) and a blobulated protein (right). Input and output for example sequence (bottom).

Research Questions

1. Do non-aliphatic hydrophobic pairs tend to coevolve?
2. What is the local sequence context around coevolving pairs?
3. What is the effect of the local sequence context (blob) on coevolving pairs?

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 [3]
2. Blobulated all protein sequences (as in Figure 1)
3. Calculated enrichment and χ (enrichment of a blob type for a given pair)

Calculating Enrichment

$$N_{ab}^{obs} = \text{Number of detected coevolving pairs 'ab'}$$

$$N_{ab}^{perm} = \text{Null frequency of pair 'ab' generated by permutation of sites}$$

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

Enrichment captures how much more likely than random chance an amino acid pair is to coevolve

Do non-aliphatic hydrophobic pairs tend to coevolve?

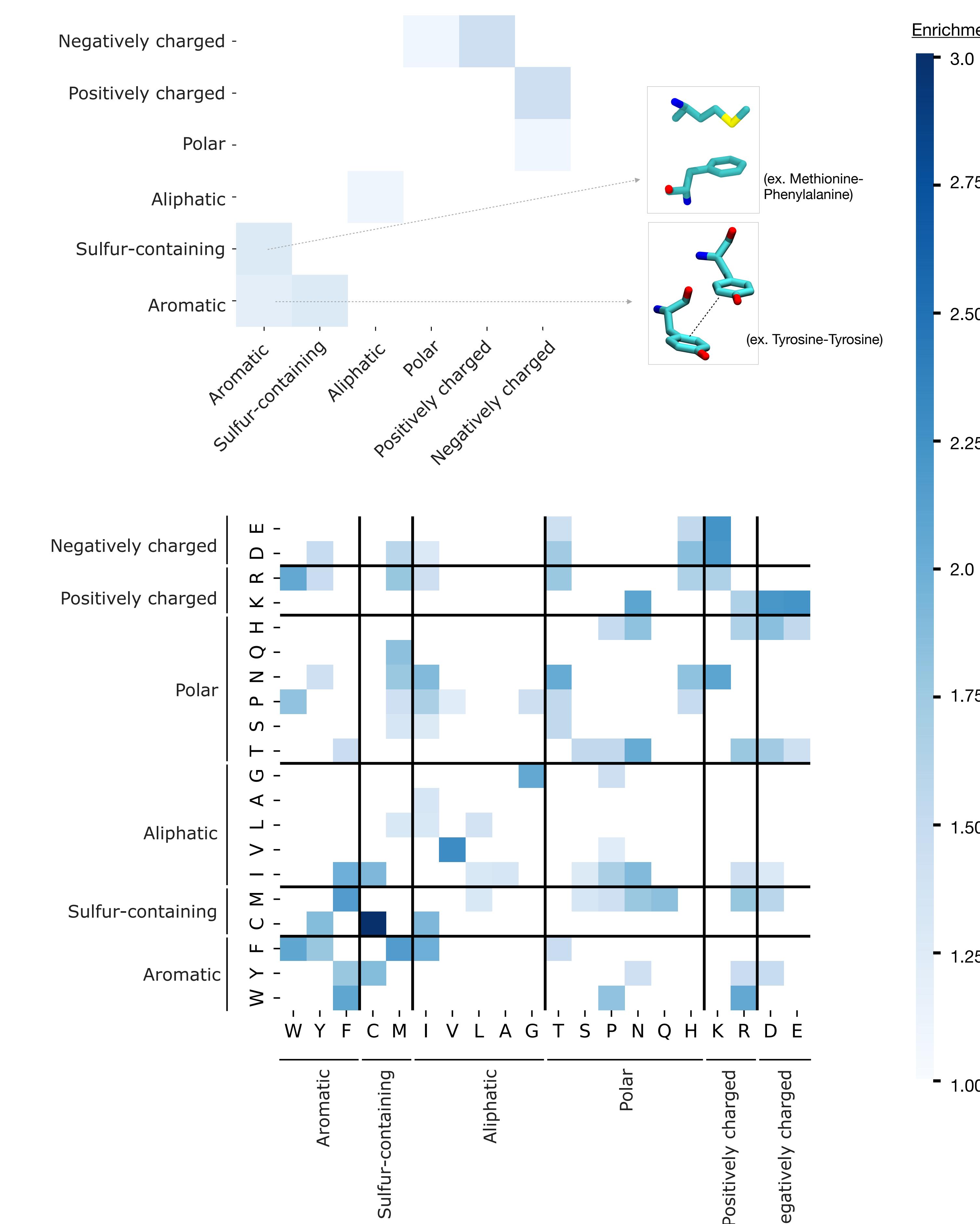


Figure 4: Enrichment of amino acid types (top) and individual amino acid pairs (bottom) for coevolving pairs. Only enrichments with an FDR < 0.05 are shown.

What is the local sequence context surrounding coevolving pairs?

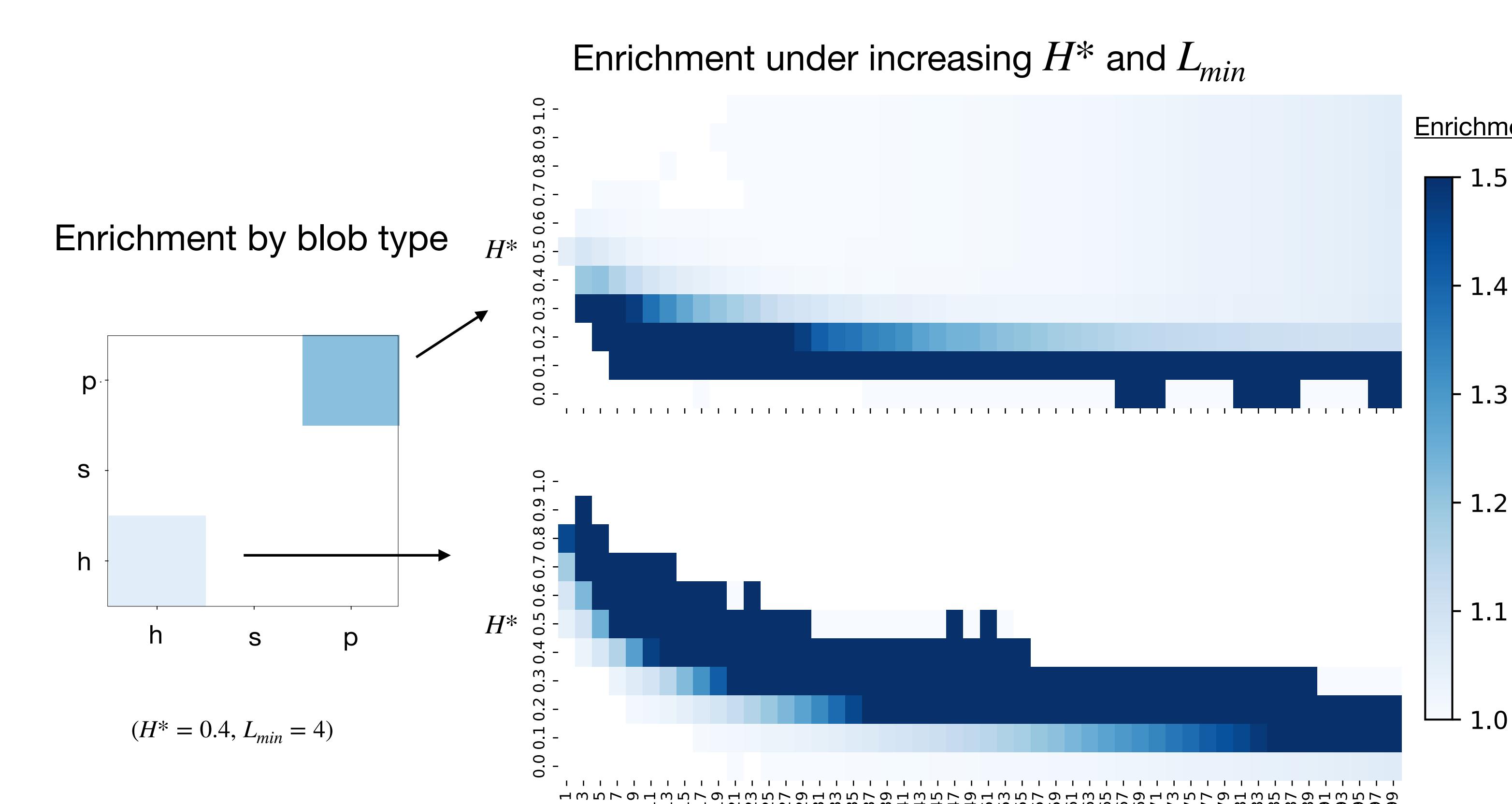


Figure 5: Enrichment of blob types (left), pairs both found in polar blobs (right, top), or both in hydrophobic blobs (right, bottom) for coevolving pairs. Only enrichments with an FDR < 0.05 are shown.

Calculating χ

$$N_{ab[cc|coev]} = \text{Number of detected coevolving pairs 'ab' both in blob type 'cc'}$$

$$N_{ab[cc]} = \text{Number of pairs 'ab' in blob type 'cc'}$$

$$N_{ab[any|coev]} = \text{Number of detected coevolving pairs 'ab'}$$

$$N_{ab[any]} = \text{Number of pairs 'ab'}$$

$$\chi_{ab[cc]} = \frac{N_{ab[cc|coev]}}{\frac{N_{ab[cc]}}{N_{ab[any|coev]}}}$$

$\chi_{ab[cc]}$ captures how much more likely a pair is to coevolve in a given context (cc)

What is the effect of local sequence context (blob) on coevolving pairs?

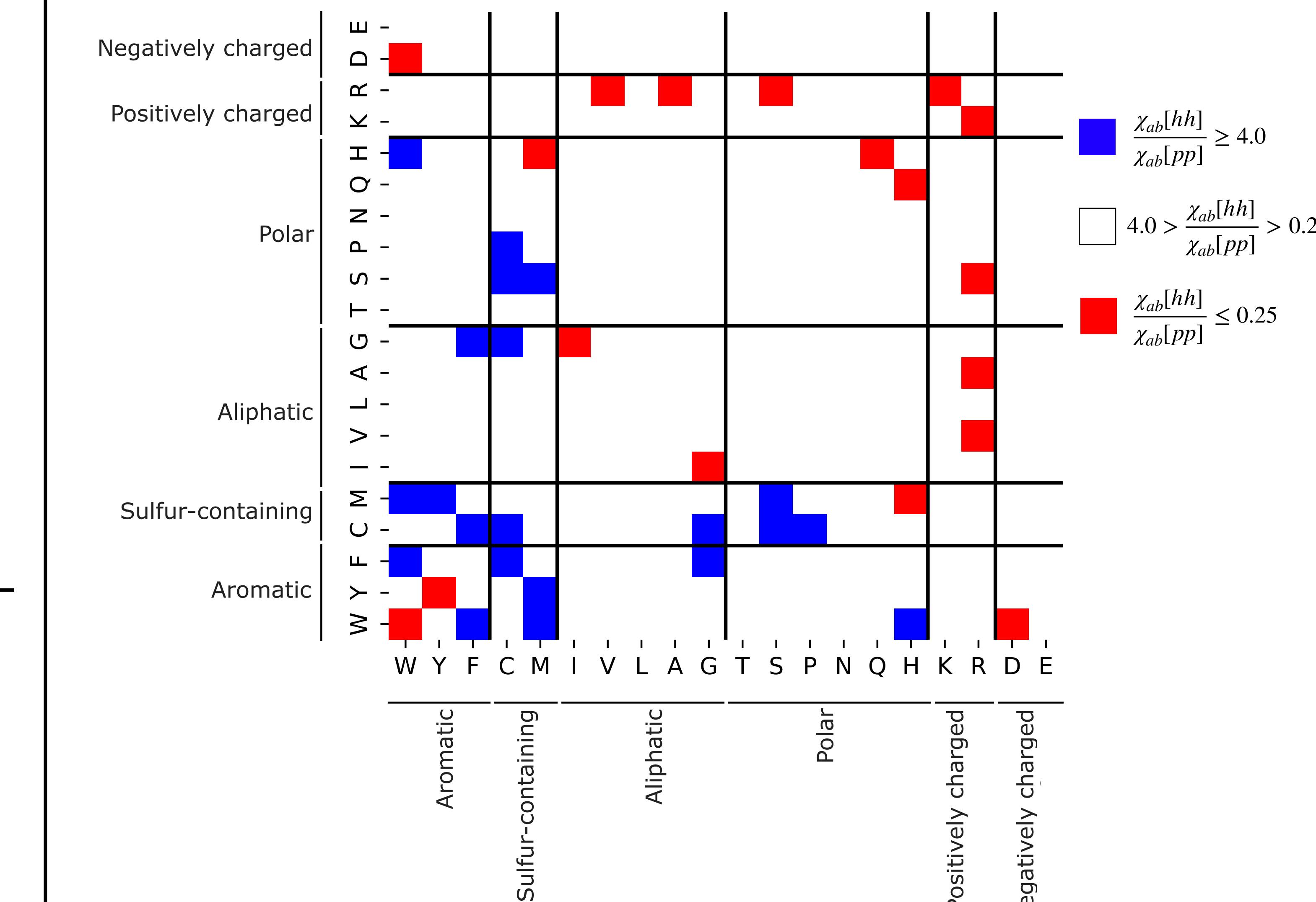
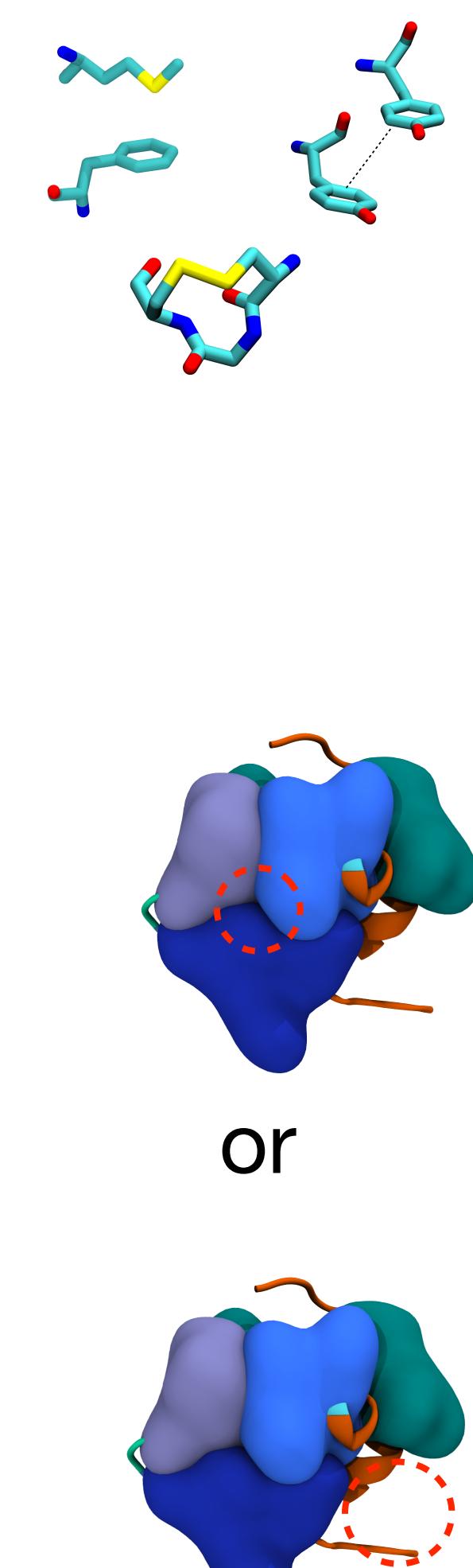


Figure 6: χ for amino acid pairs. All pairs with $\frac{\chi_{ab[hh]}}{\chi_{ab[pp]}}$ greater than 4.0 are blue, pairs with $\frac{\chi_{ab[hh]}}{\chi_{ab[pp]}}$ less than 0.25 are red, and all other pairs are white.

Summary

1. Non-aliphatic hydrophobic pairs are significantly enriched, and tend to be found in h-blobs, indicating that they play vital roles in protein structure and function.
2. Coevolving pairs tend to either both be found in hydrophobic (h-h) or both on polar (p-p) blobs.
3. Longer and moderately hydrophobic and short highly hydrophobic blobs have high enrichments, potentially corresponding to hydrophobic cores and transmembrane domains.
4. Pairs in highly polar (detected at low H^*) blobs have high enrichments, potentially corresponding to pairs found at the surface of proteins.
5. Non-aliphatic pairs tend to be in h-h pairs, except for W-W and Y-Y
6. Pairs in polar blobs tend to be composed of at least one charged pair, consistent with the finding that the local sequence context surrounding them is highly polar



h-h
F-W, W-M, Y-Y
M, F-C, C-C
p-p
W-W, Y-Y

Acknowledgements

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

References

- [1] S. Chaurasia and J. Dutheil. Molecular Biology and Evolution, 2022.
- [2] R. Lohia, M. Hansen, G. Brannigan. PNAS, 2022.
- [3] J. Dutheil and N. Galtier. BMC Biology and Evolution, 2007.