

The Role of Contiguous Hydrophobicity and Non-Aliphatic Hydrophobic Residues in Coevolution.

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

We have recently presented a sequence-based algorithm for detecting intrinsic modularity in proteins ("blobulation") and demonstrated its predictive power in conformational clustering of intrinsically disordered proteins. In particular, we found that regions of contiguous hydrophobicity ("hydrophobic blobs") were likely to form tertiary interactions even in the absence of structure [1]. This result suggests an approach for "ultra coarse-graining" protein sequences using monomers that each represent a varying number of residues. However, the significance of interactions between certain "specialty" hydrophobic residues, as well as the role of the local sequence in stabilizing such interactions, remains unclear. Identification of coevolving residues allows for a sequence-based approach to detecting residue-residue interactions using aligned protein sequences and phylogenies. Here, we test whether coevolving residues are more likely to be found in hydrophobic blobs, and whether the non-aliphatic hydrophobic residues in those blobs are particularly likely to be coevolving. We find that coevolving residue pairs are enriched for the case that both residues are found in hydrophobic blobs or both in polar blobs, and for pairs of "specialty" hydrophobic residues. These results suggest that ultra-coarse-grained models of complex polymers should consider not just the overall hydrophobicity and/or net charge of each monomer, but the number of residues that participate in specific attractive interactions.

Specialty Residue Interactions

- Pairwise methionine interactions change the conformational ensemble of the BDNF prodomain in molecular dynamics simulations [1]
- Hydrophobic non-aliphatic residues are found in contact in structured proteins in the PDB [2] (Figure 1)

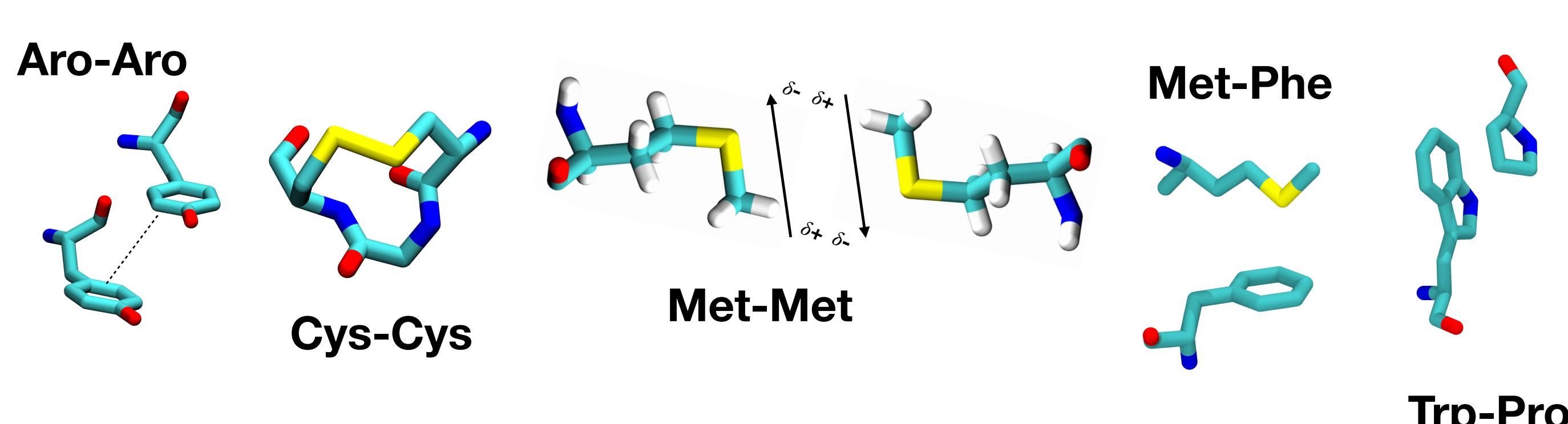


Figure 1: Examples of pairwise specialty residue interactions.

- Clusters of hydrophobic residues (figure 2) form tertiary interactions in IDPs [1] and are enriched for disease-associated mutations (per length) [3]
- Coevolving residues (consistently co-occurring across evolutionary history) are often found in contact [4]

Coevolution of Amino Acid Pairs

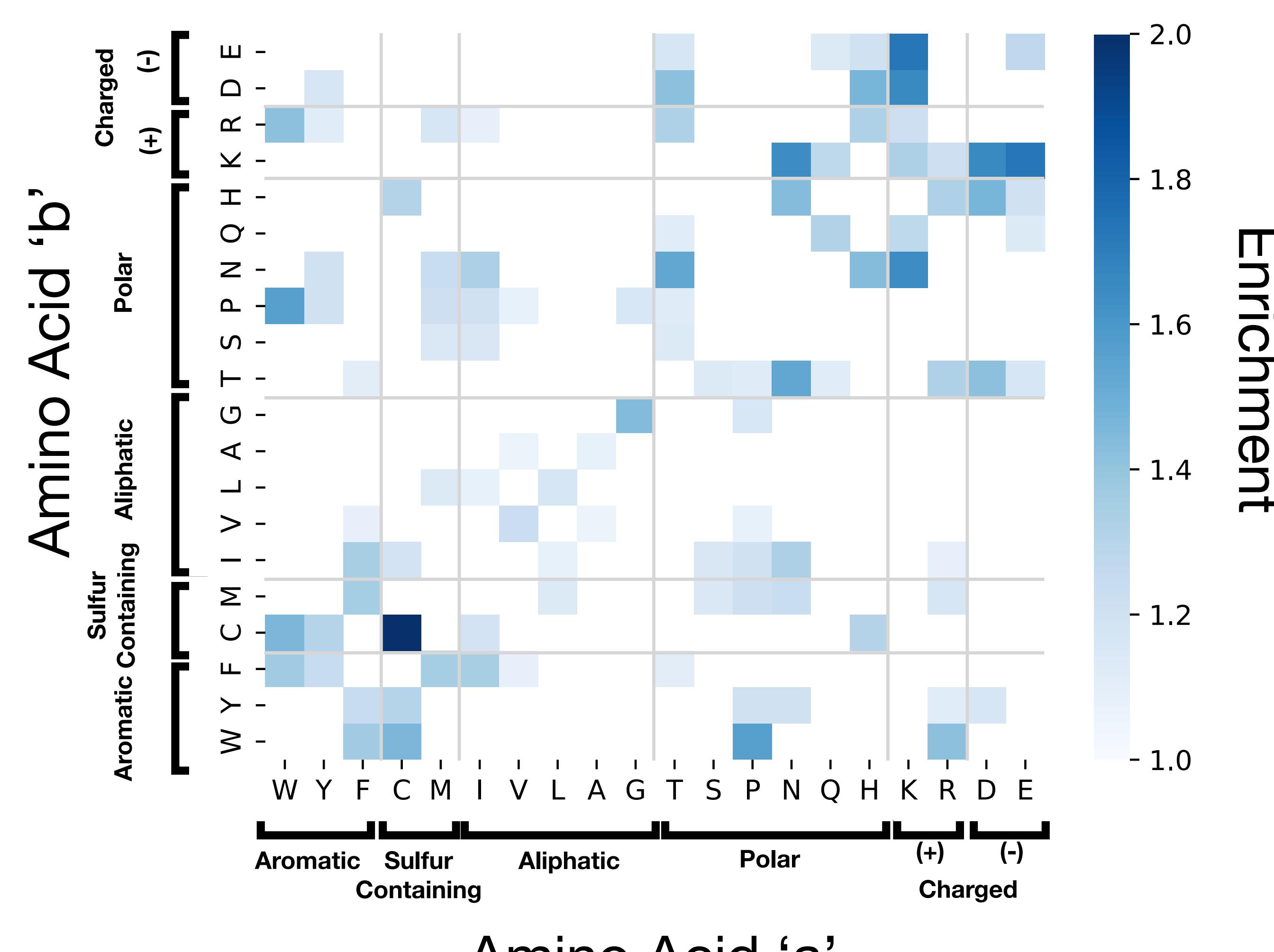


Figure 4: Enrichment of coevolving residue pairs 'ab'. Residues are labeled by their biochemical properties.

Approach

1. Detected coevolving residue pairs in a large Bacterial protein dataset (1630 protein families, with ~229 orthologs per family - previously used to investigate the role of structure in coevolution [4]) using CoMap [5]
2. Blobulated all protein sequences (as in Figure 2)
3. Calculated enrichment for coevolving residues of given amino acid pair types. Null expectation was generated using a permutation test. All detected coevolving sites were shuffled, with the amino acids found at each remaining unshuffled.

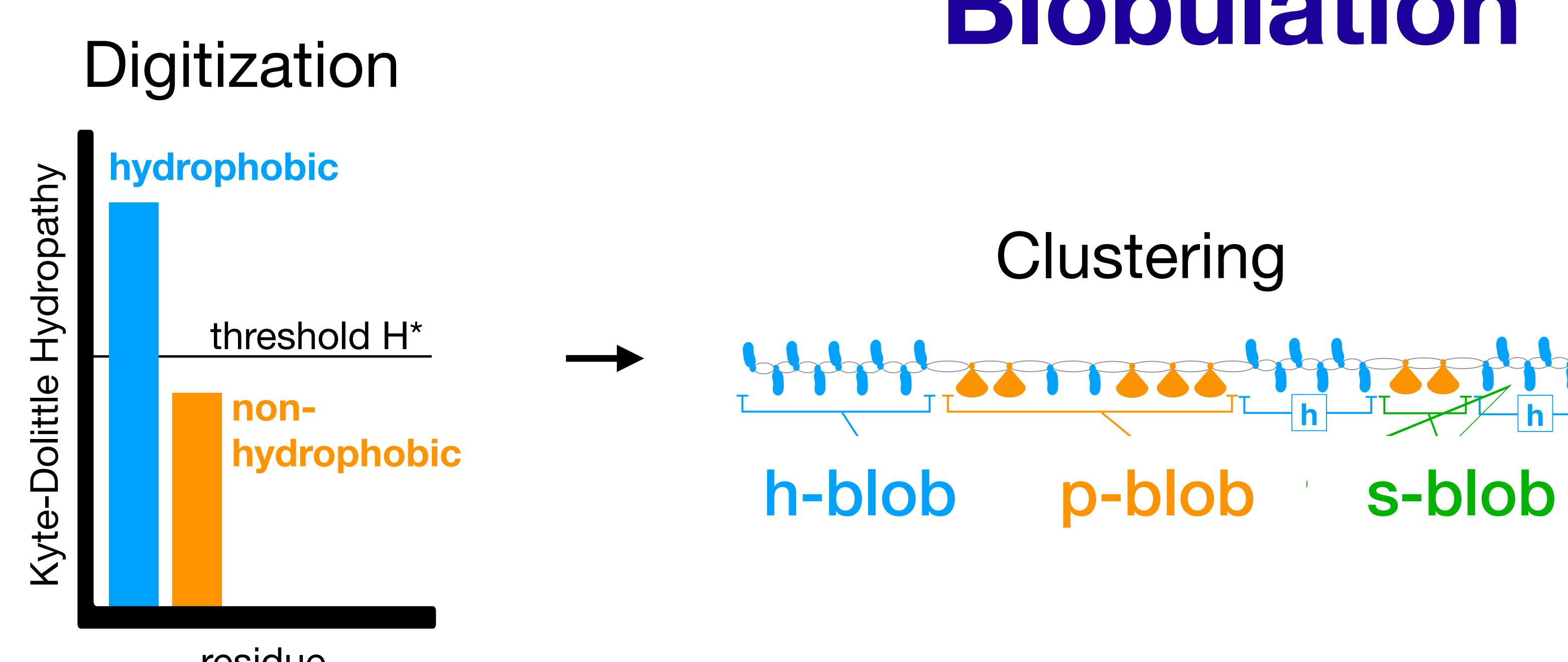


Figure 2: Blobulation, our algorithm for detecting intrinsic modularity in protein sequences based on hydrophobicity. The algorithm involves two steps: digitization using hydrophobicity threshold H^* (left), and clustering (middle). Figure adapted from [5]. Example representation made in VMD of Lysozyme blobs (right, Uniprot: P00720, PDB:2LZM).

Coevolution of Amino Acids by Blob Type

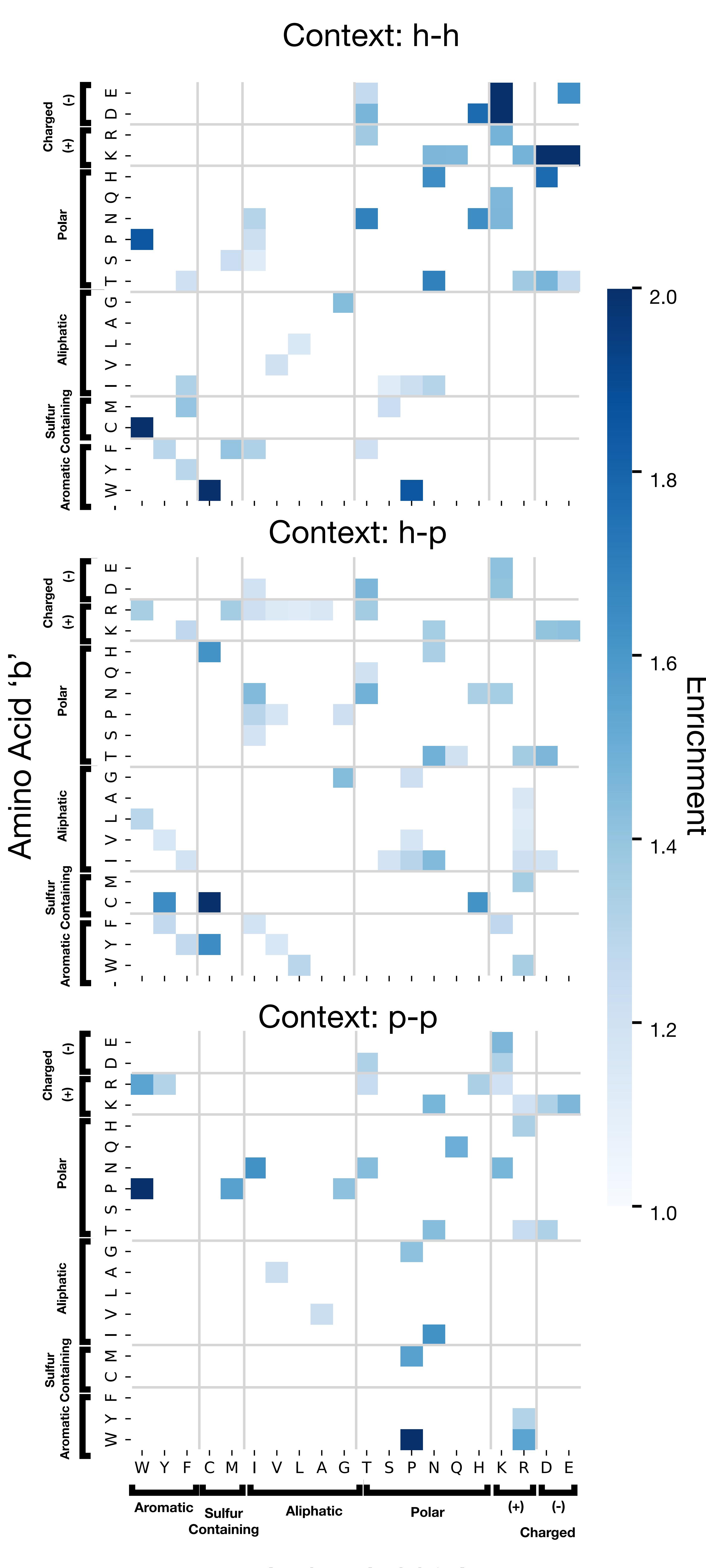


Figure 5 : Enrichment of coevolving amino acid residue pairs as in 4. Residue pairs are further broken down by the blob that contains each.

Research Questions

1. Do Residues in the same blob type tend to coevolve?
2. Do specialty residue interaction pairs (Aro-Aro, Met-Met, Cys-Cys, Met-Phe, etc.) tend to coevolve? Does blob type play a role?



Blobulate your own protein here!

Blobulator.branniganlab.org

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

N_{ab}^{perm} = Null frequency of pair 'ab', generated by random permutation of sites (as in Approach)

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

Summary

- Coevolving sites are enriched for aromatic and sulfur-containing residue pairs only in h-h and h-p contexts, and tryptophan-proline pairs are enriched only in p-p contexts
- Coevolving sites are enriched for cystine-cystine pairs only in h-p contexts and these pairs may play a role in bringing together parts of proteins that would normally not interact
- Coevolving sites are enriched for lysine-charged pairs regardless of context

Acknowledgements

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

References

1. R. Lohia, R. Salari, and G. Brannigan. *PLOS Computational Biology*, 2019.
2. G. Faure, A. Bornot, A.G. de Brevern. *Biochimie*, 2008.
3. R. Lohia, M.E.B. Hansen, and G. Brannigan. *PNAS*, 2022.
4. S. Chaurasia and J.Y. Dutheil. *Molecular Biology and Evolution*, 2022.
5. J. Dutheil, N. Galtier. *BMC Evol Biol*, 2007.