

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 [1] as well as bioinformatics analyses of human disease [2]. In particular, we found that regions of contiguous hydrophobicity (“hydrophobic blobs”) were likely to form tertiary interactions even in the absence of structure, and were more likely to contain disease-associated mutations across the human exome. However, the functional significance of residue-residue interactions within hydrophobic blobs is unclear. Identification of coevolving residues allows for an alternative 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. Furthermore, we characterize how the physicochemical properties of individual blobs and blob pairs affect the likelihood that they contain coevolving residues.

Background

- 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)

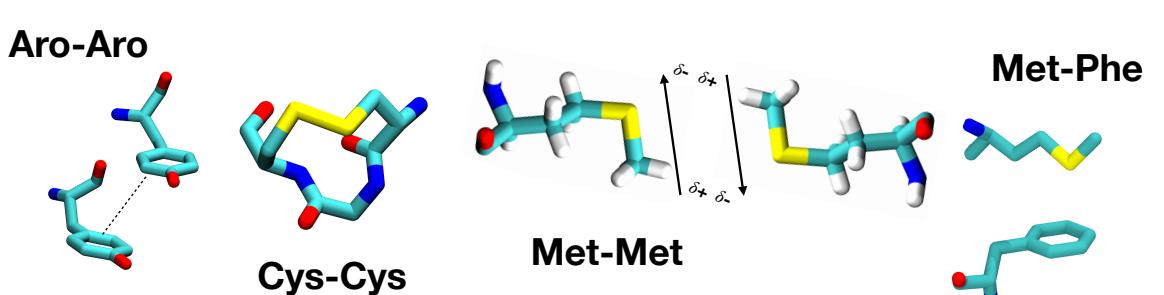


Figure 1: Examples of pairwise specialty hydrophobic residue interactions.

- 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]

Blobulation

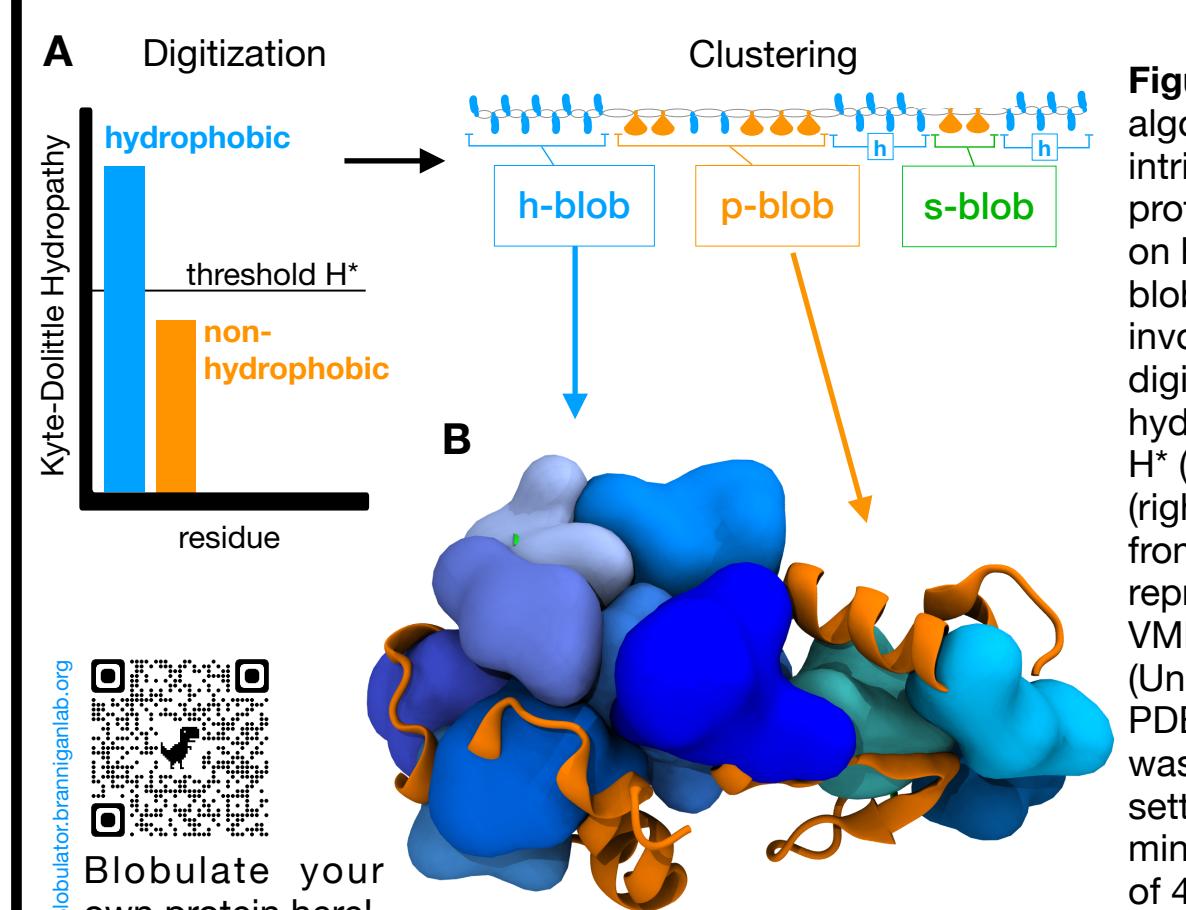


Figure 2: Blobulation, our algorithm for detecting intrinsic modularity in protein sequences based on hydrophobicity A) The blobulation algorithm involves two steps: digitization using hydrophobicity threshold H^* (left), and clustering (right) Figure adapted from [5]. B) Example representation made in VMD of Lysozyme blobs (Uniprot: P00720, PDB:2LZM). Lysozyme was blobulated using the settings $H^* = 0.4$ and a minimum length threshold of 4.

Research Questions

- Do Residues in the same blob type tend to coevolve?
- Does the net charge of a given blob have any effect?
- Do specialty residue interaction pairs (Aro-Aro, Met-Met, Cys-Cys, Met-Phe, etc.) tend to coevolve?

Approach

- Use the software CoMap to detect coevolving residue pairs in a large Bacterial protein dataset (1630 protein families, with ~229 orthologs per family) [4]
- Calculate enrichment for coevolving residues of given amino acid pair types that are found in blobs with varying properties using the following equations:

Equations for Enrichment Calculation

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

N_{ab} : number of pair ‘ab’ in set
 f_{ab} : fraction of pair ‘ab’ in set relative to other pairs

$$\text{Enrichment} = \frac{f_{ab}^*}{f_{ab}}$$

Enrichment: ratio between significant f_{ab}^* and total f_{ab}

References

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

Acknowledgments

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

Coevolution of Blobs

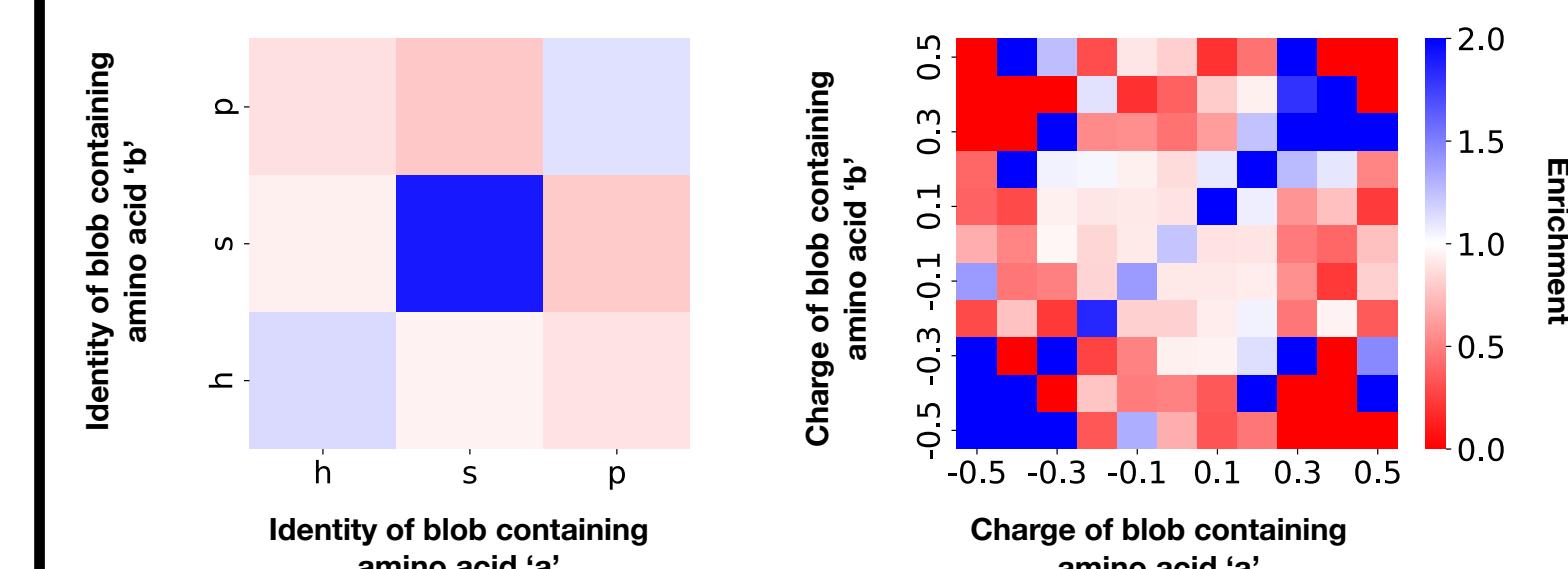


Figure 3: Enrichment of coevolving residues by blob type (left) and blob net charge (right). Enrichment was calculated as in Equation 3, with each residue being classified by the blob that contains it.

Coevolution of Amino Acids

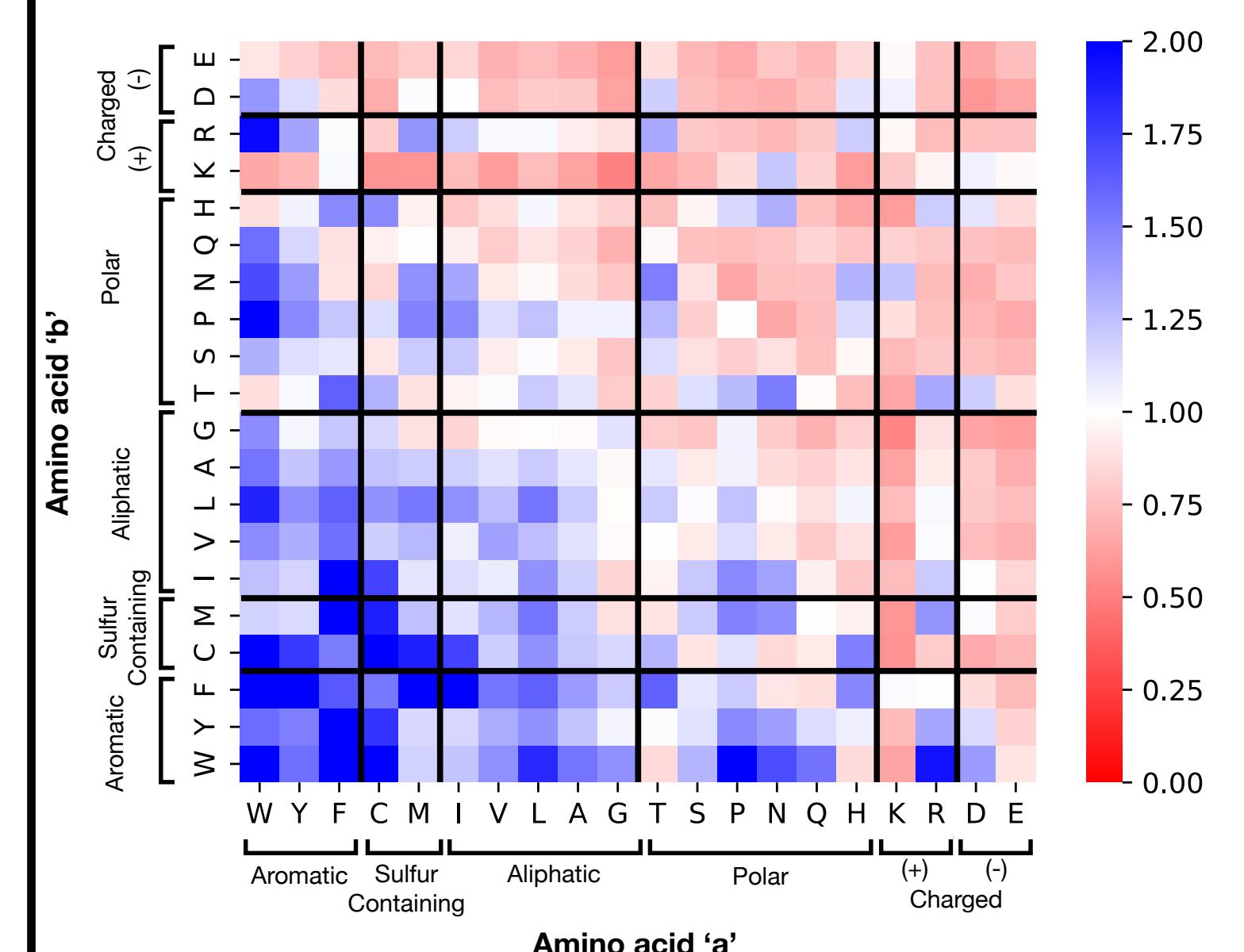


Figure 4: Enrichment of coevolving amino acid residue pairs. Residues are labeled by their biochemical properties. Enrichment was calculated as in Equation 3.

Summary

- Coevolving pairs tend to be found in blobs of the same type
- Like-charged blobs tend to be enriched for coevolving pairs, and uncharged blobs are moderately enriched
- Coevolving pairs are enriched for aromatic and sulfur-containing non-aliphatic hydrophobic residues