

# A toolkit for the analysis of contiguous regions of hydrophobicity in a protein sequence

Lindsey Riggs<sup>1</sup>, Connor Pitman<sup>1</sup>, Ezry Santiago-McRae<sup>1</sup>, Ruchi Lohia<sup>3</sup>, Ryan Lamb<sup>1</sup>, Kaitlin Bassi<sup>1</sup>, Thomas T. Joseph<sup>4</sup>, Mathew E.B. Hansen<sup>5</sup>, Grace Brannigan<sup>1,2</sup>

Center for Computational and Integrative Biology, Rutgers - Camden, NJ<sup>1</sup>, Department of Physics, Rutgers - Camden, NJ<sup>2</sup>, Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, PA<sup>3</sup>, Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, PA<sup>4</sup>, Department of Genetics, Perelman School of Medicine, University of Pennsylvania, PA<sup>5</sup>  
RUTGERS-CAMDEN  
Center for Computational and Integrative Biology  
lmr294@rutgers.edu

## Abstract

Current bioinformatics tools can characterize sequences using features such as charge, sequence conservation, or sequence patterns, but they do not identify hydrophobic modules that may be relevant for functional annotation or mutation analysis. Therefore, we developed the algorithm, blobulation, to identify contiguous regions of hydrophobicity in a protein sequence, which can be used for large-scale data analysis. Applying blobulation to human genetic variation data, we discovered that disease-associated mutations are enriched in stretches of hydrophobic regions compared to non-hydrophobic regions [1]. These results highlighted the significance of hydrophobic stretches in human proteins. To make blobulation an easily accessible tool for different user types, we implemented this algorithm into three tools: a command-line interface, a Visual Molecular Dynamics plug-in, and a web tool. These tools are used for detecting, visualizing, and characterizing hydrophobic modularity and allow us to define the local context around each residue in a sequence. Here, we present examples of how a user could use the toolkit and the information revealed by using the blobulation algorithm. The blobulator web tool can be found at [www.blobulator.branniganlab.org](http://www.blobulator.branniganlab.org), and the source code with a pip installable command line tool, as well as the VMD plugin with installation instructions, can be found on GitHub at [GitHub.com/BranniganLab/blobulator](https://github.com/BranniganLab/blobulator).

## Blobulation algorithm

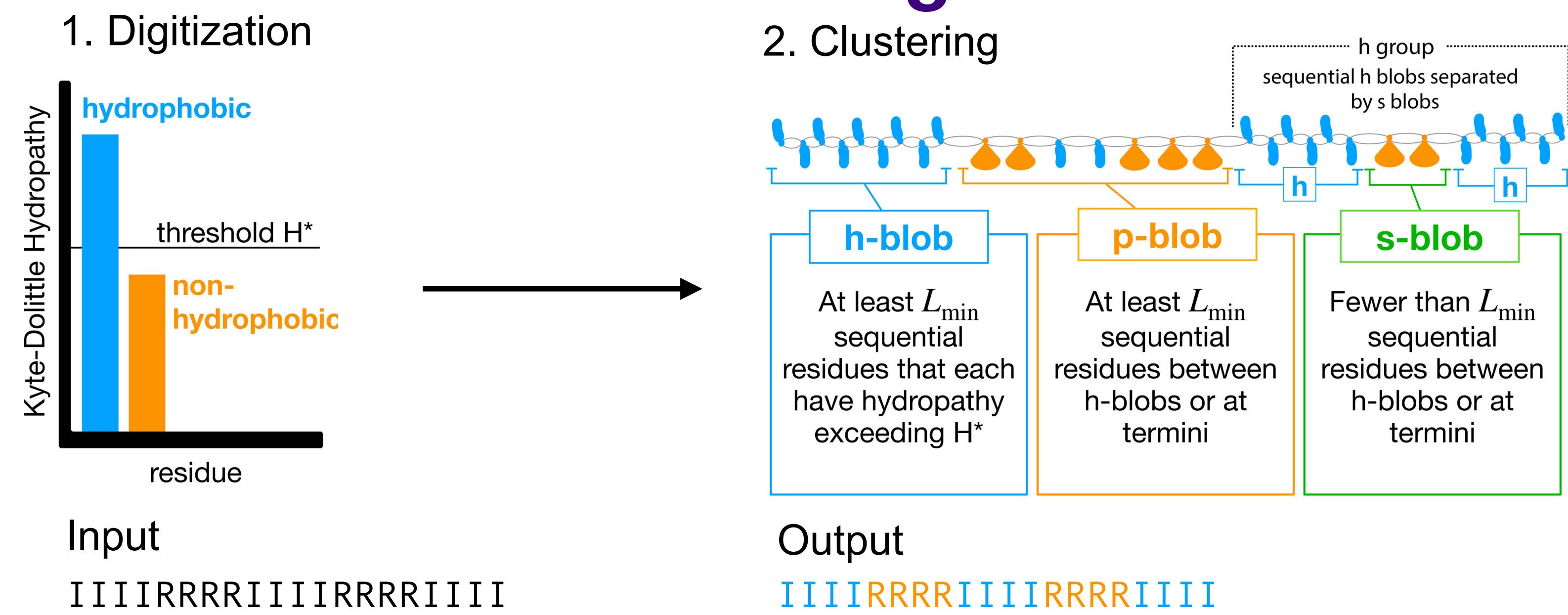


Figure 1: Graphical representation of the Blobulation algorithm. (A) The amino acid sequence is digitized by the user defined hydrophobicity threshold ( $H^*$ ). (B) Next, the sequence is clustered into 3 different blobs based on  $H^*$  and the minimum number of amino acids ( $L_{min}$ ). Figure adapted from [1].

## Visualization of hydrophobic modules

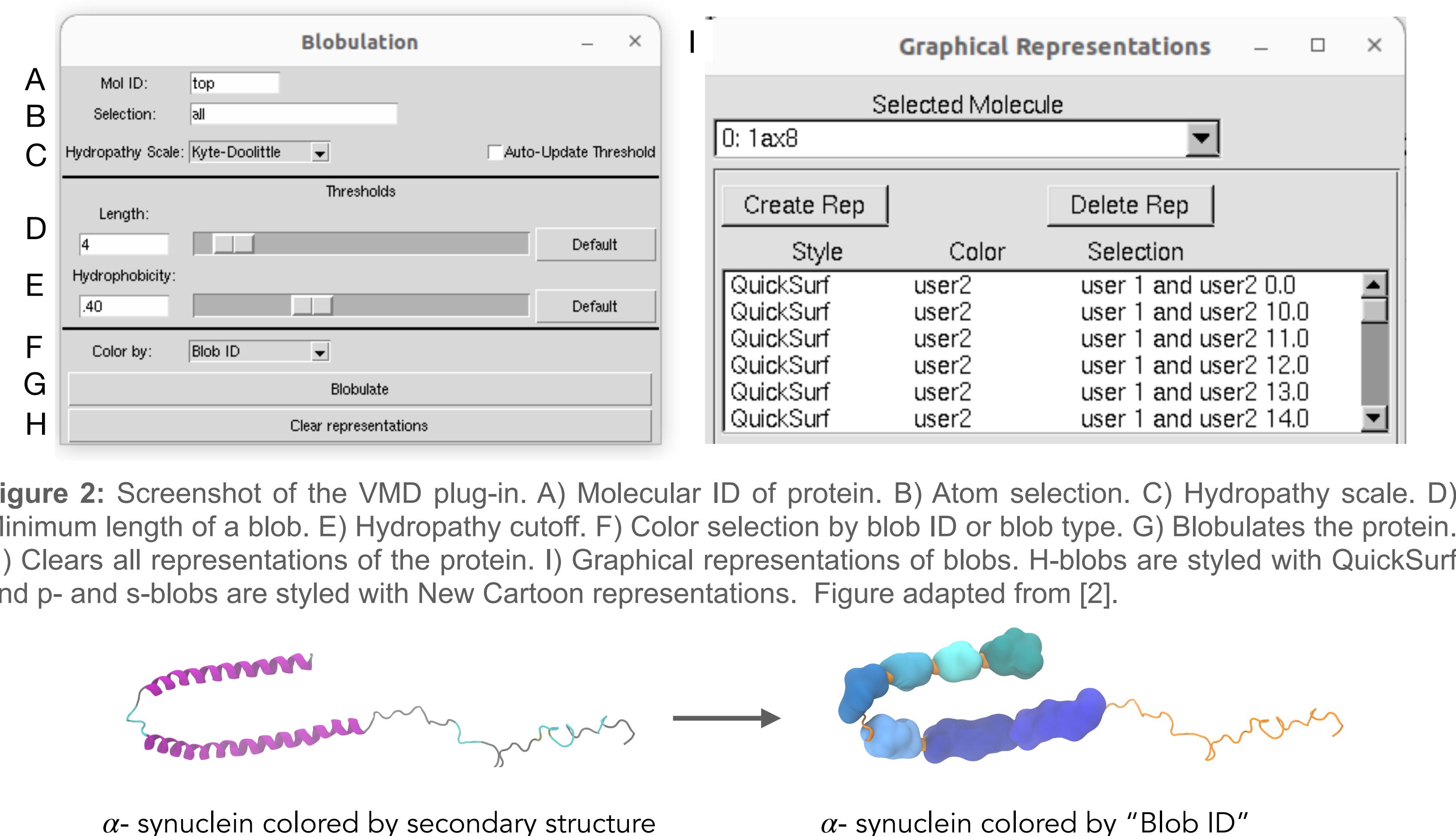


Figure 2: Screenshot of the VMD plug-in. A) Molecular ID of protein. B) Atom selection. C) Hydrophobicity scale. D) Minimum length of a blob. E) Hydrophobicity cutoff. F) Color selection by blob ID or blob type. G) Blobulates the protein. H) Clears all representations of the protein. I) Graphical representations of blobs. H-blobs are styled with QuickSurf and p- and s-blobs are styled with New Cartoon representations. Figure adapted from [2].

## Local context from sequence

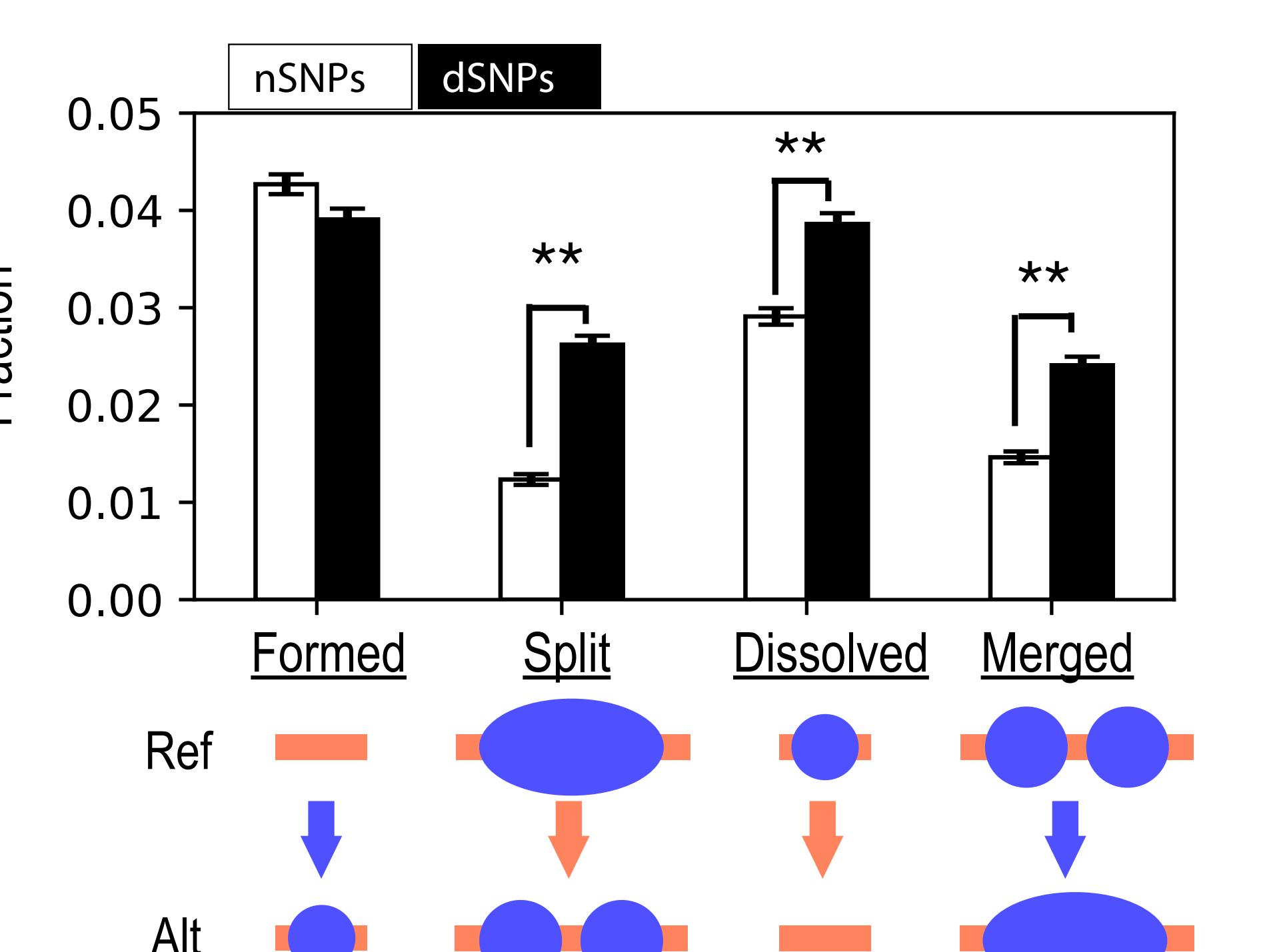


Figure 4: Disease-associated single nucleotide polymorphisms (dSNPs) can disrupt hydrophobic modularity by splitting them, dissolving them into a p-blob, or merging them into a nearby h-blob compared to non-disease-associated SNPs. Figure adapted from [1].

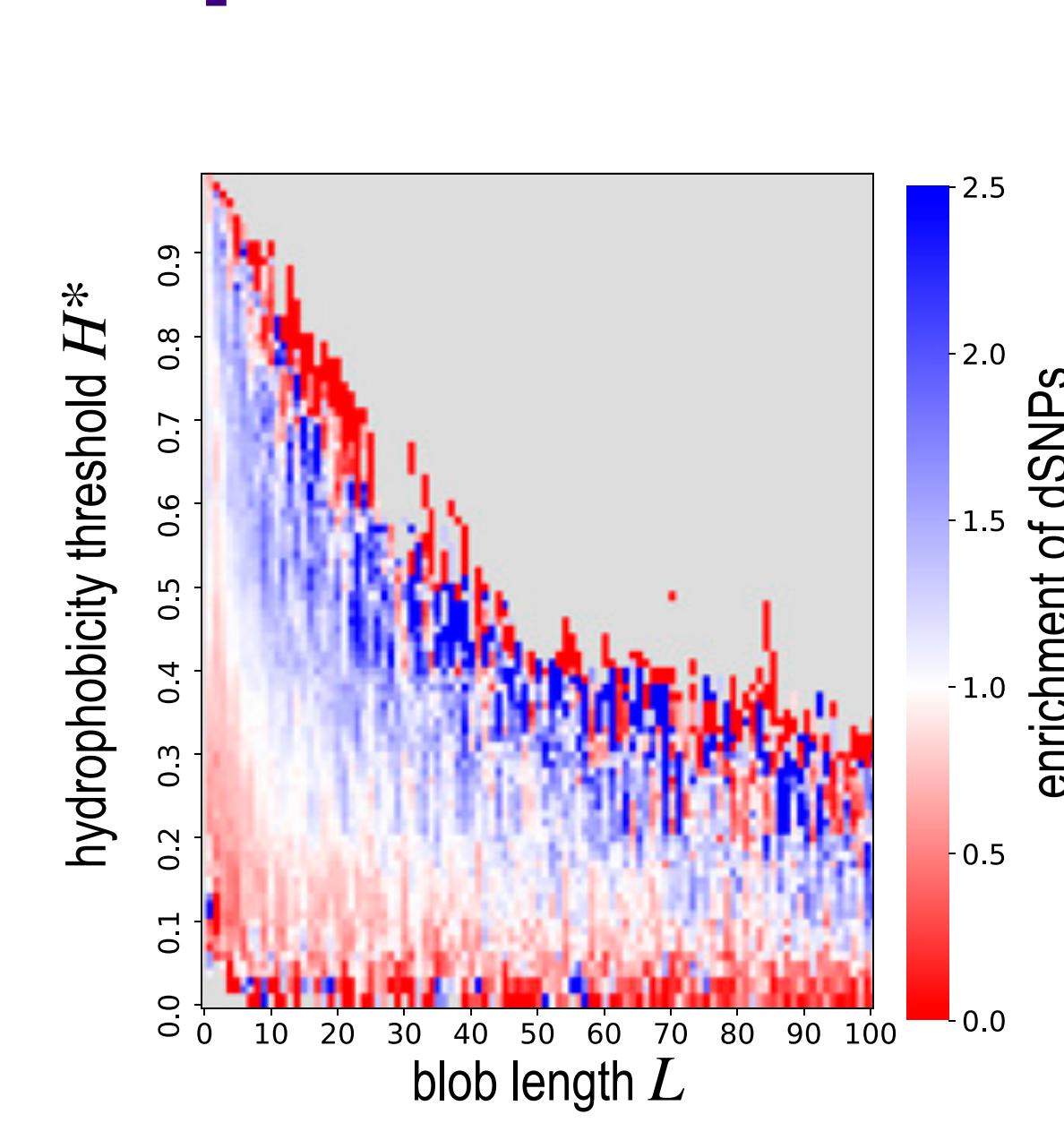


Figure 5: Enrichment of dSNPs in hydrophobic modules with varying lengths  $L$  and  $H^*$ . Enrichment depends on the local context, or the length and hydrophobicity of h-blobs in the sequence. Figure adapted from [1].

## Background

- Current approaches that use protein sequences to define the local context around a single residue or mutation overlook hydrophobic modularity.
- To address this, we developed the **Blobulation algorithm**, which clusters residues into contiguous regions based on their hydrophobicity.
- The **Blobulator toolkit** consists of a command-line interface, a graphical user interface, and a web tool for detecting, visualizing, and characterizing protein sequences.

**Vision statement:** Provide researchers with an interactive and intuitive interface to detect intrinsic modularity in any protein sequence based on hydrophobicity

## Characterization

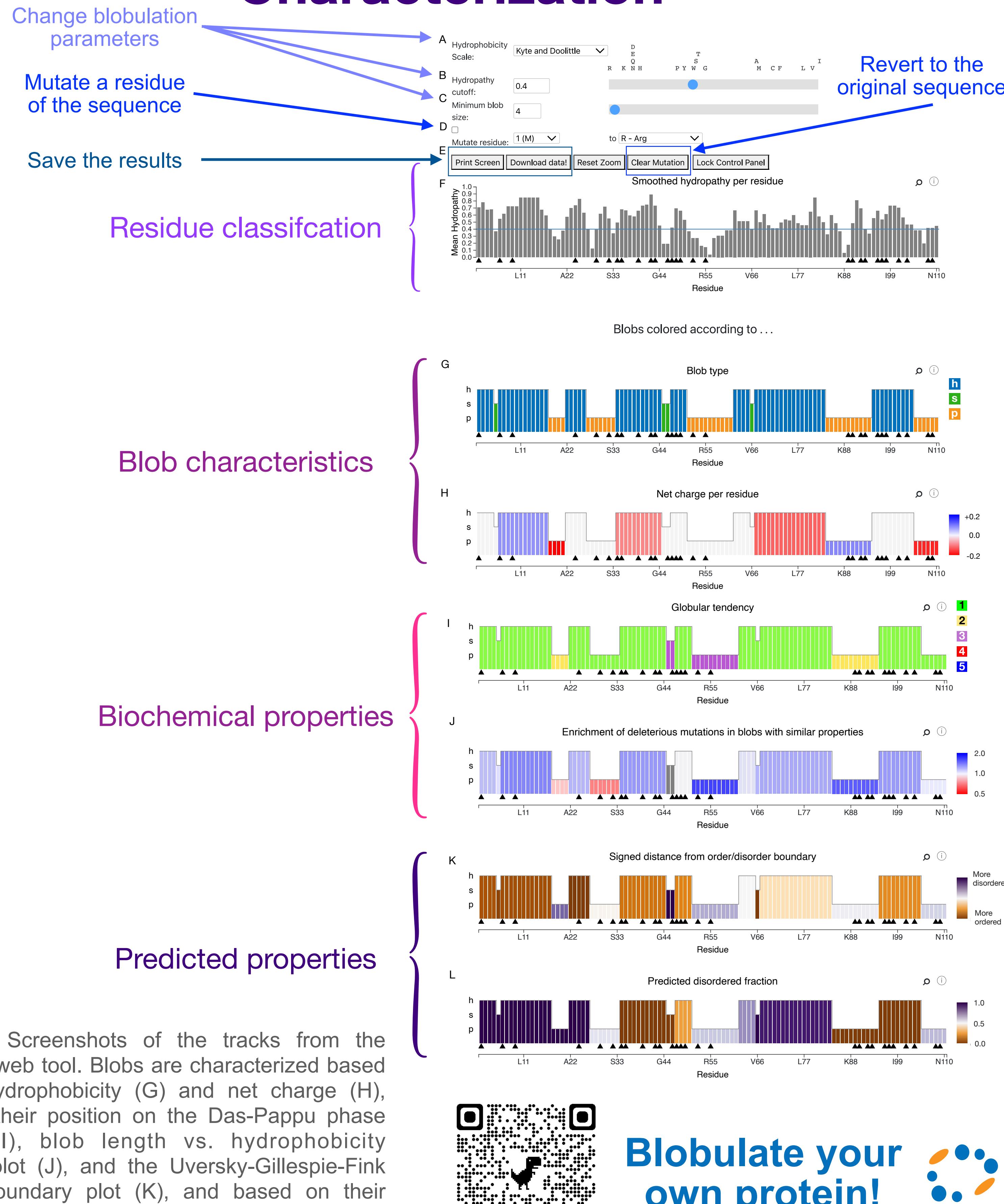


Figure 3: Screenshots of the tracks from the Blobulator web tool. Blobs are characterized based on their hydrophobicity (G) and net charge (H), based on their position on the Das-Pappu phase diagram (I), blob length vs. hydrophobicity threshold plot (J), and the Uversky-Gillespie-Fink disorder boundary plot (K), and based on their predicted disordered fraction (L). Black triangles represent known disease-associated mutations in humans. Figure adapted from [2].

## Quantifying the effects of mutations

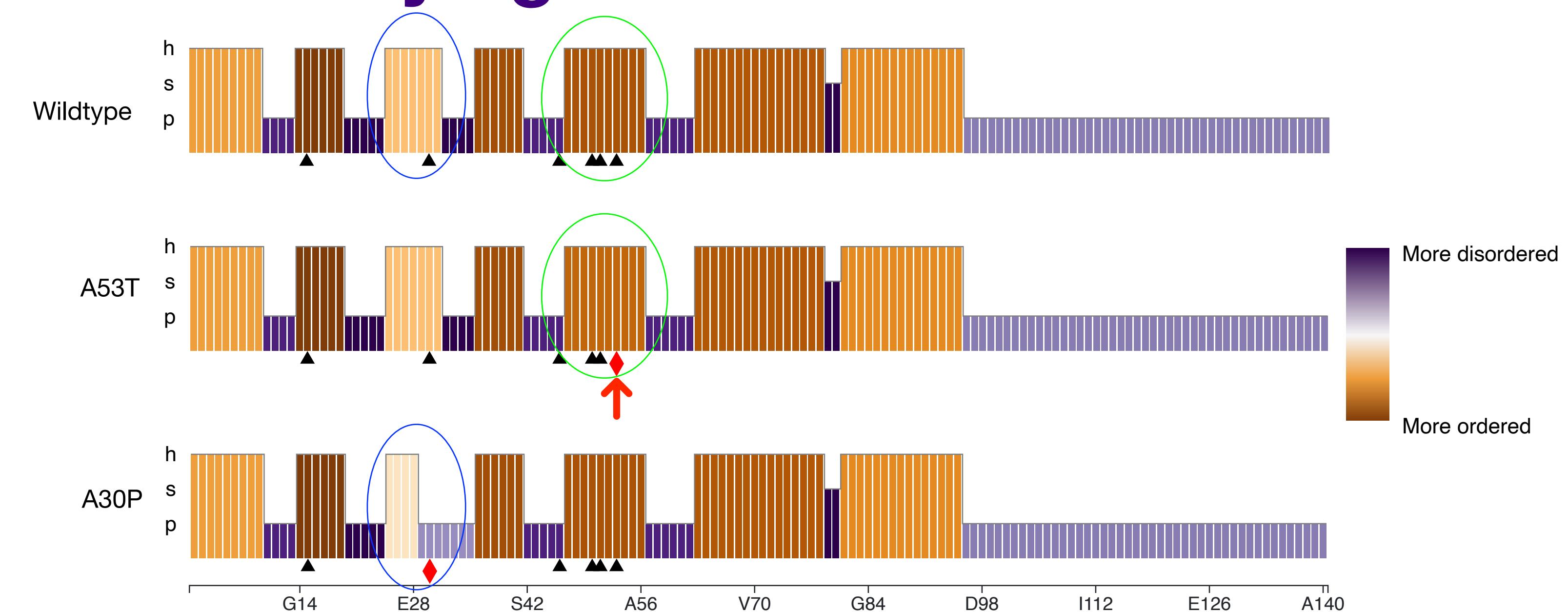


Figure 6: While the A53T mutation does not have a large effect on the intrinsically disordered protein, α-synuclein, the A30P mutation shortens the h-blob that contains the residue and decreases its predicted order property. Figure adapted from [2].

## Acknowledgements

- Rutgers Office of Advanced Research and Computing (OARC)
- NRT, NSF DGE 2152059
- 1. R. Lohia, M.E.B. Hansen, and G. Brannigan. *PNAS*, 2022.
- 2. C. Pitman, et al. *bioRxiv*. 2025.