

The Blobulator: A Toolkit for Identification and Visual Exploration of Hydrophobic Modularity in Protein Sequences

R
RUTGERS

By: Connor Pitman¹, Ezry Santiago-McRae¹, Ruchi Lohia², Ryan Lamb³, Kaitlin Bassi³, Lindsey M. Riggs³, Thomas Joseph², Matthew Hansen², Grace H. Brannigan⁴

¹CCIB, Rutgers University -- Camden, Camden, NJ, USA, ²University of Pennsylvania, Philadelphia, PA, USA, ³Rutgers University — Camden, Camden, NJ, USA, ⁴Rutgers University Dept Physics, Camden, NJ, US

Abstract

Clusters of hydrophobic residues are known to promote structured protein stability and drive protein aggregation. In recent work, we have shown that identifying contiguous hydrophobic residue clusters within protein sequences (termed “blobs”) has proven useful in interpreting intra-protein contacts in a series of intrinsically disordered protein simulations and defining the “local context” around disease-associated mutations across the human proteome. However, an accessible toolkit that identifies these clusters was unavailable, and the role that blobs play across the structural context of a variety of protein families remained unclear. Here, we present the blobulator toolkit: consisting of a webtool, a command line interface, and a VMD plugin. We demonstrate, in three example applications, how one might use the toolkit to identify blobs that reveal useful information about a globular protein, two orthologous membrane proteins, and an IDP. Finally, we present new features for the webtool, including the ability to view blobs on a protein structure via an interface to upload PDB files and integration of alphafold structures when using a UniProt ID. The blobulator webtool can be found at [www.blobulator.branniganlab.org](http://blobulator.branniganlab.org), and the source code with pip installable command line tool, as well as the VMD plugin with installation instructions, can be found on GitHub at <https://github.com/BranniganLab/blobulator>.

Background

- Clusters of hydrophobic residues form structural elements of proteins, such as the hydrophobic core and transmembrane domains
- Residue hydrophobicity has been critical in identifying intrinsically disordered proteins and regions from sequence (IDPs and IDR)s[1, 2]
- Yet there was no method for segmenting proteins by hydrophobicity
- We developed an algorithm called “blobulation” and used it to:
 - Detect shifts in intra-protein interactions in a long IDP simulation [3]
 - Define the local sequence context surrounding deleterious mutations across the human proteome [4]
- Here we present the blobulator toolkit: a webtool, a VMD plugin, and a pip-installable command line interface

Methodological Goals

1. Identify modules that correspond to known structural/functional features in proteins (blobs)
2. Quantify the impact of mutations to the “local sequence environment”
3. Provide user-friendly interfaces to adaptively detect blobs in protein sequences and structures

The Blobulation Algorithm

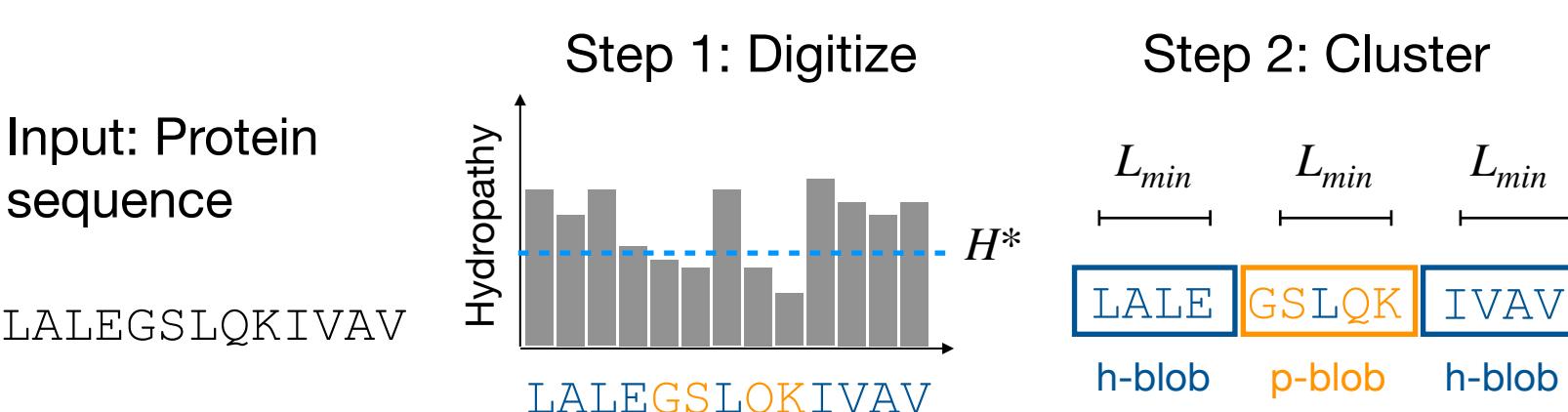


Figure 1: The blobulation algorithm. Blobulation consists of two steps: digitization (each residue is classified as hydrophobic (blue) or non-hydrophobic (orange) if its hydrophathy is greater than or less than the user-defined hydrophathy cutoff H^*) and clustering (contiguous stretches of hydrophobic residues longer than the user-defined length minimum L_{min} are clustered into h-blobs, all other segments are classified as p-blobs if they are $> L_{min}$ or s-blobs if they are $< L_{min}$).

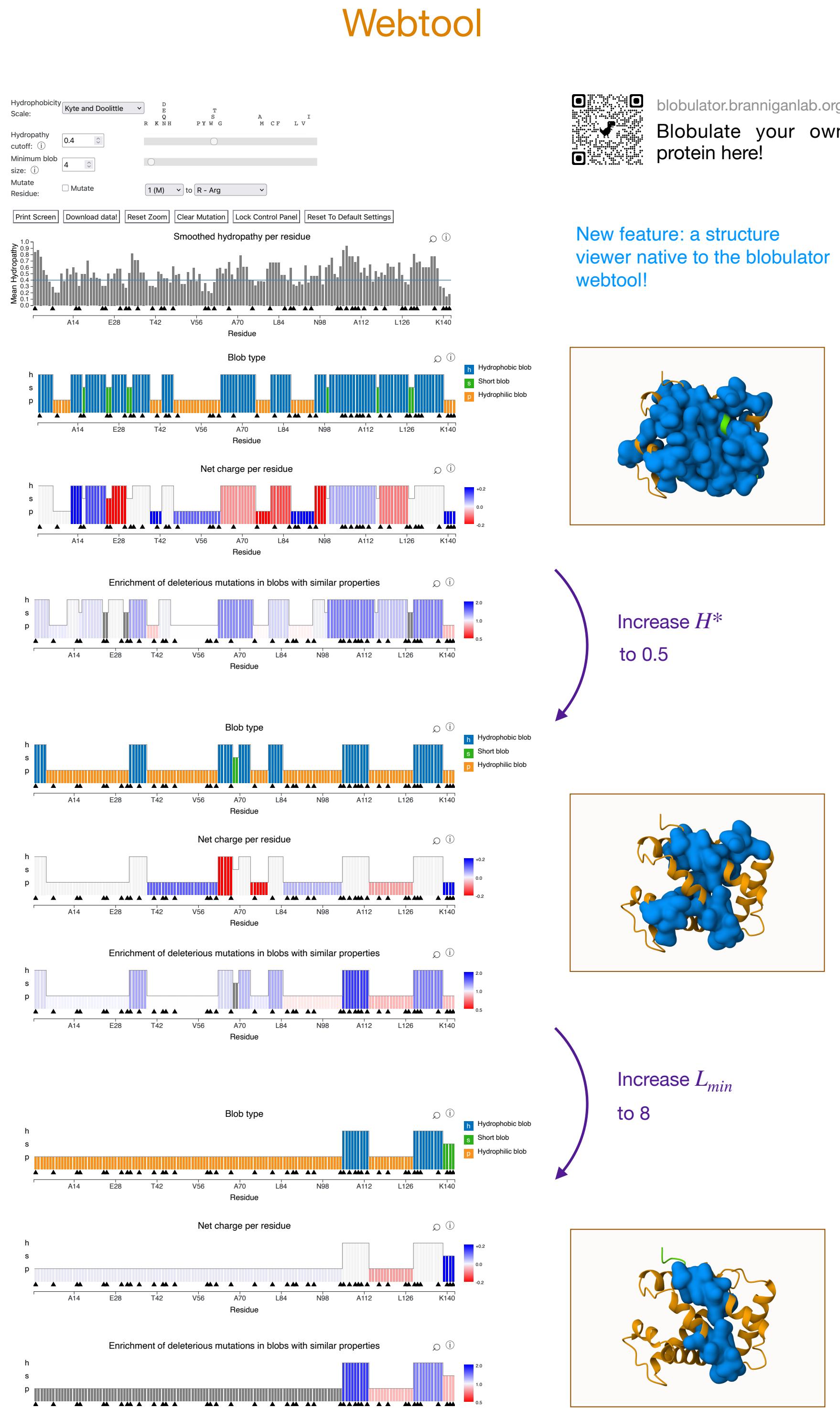


Figure 2: Adjusting parameters using the blobulator webtool. By increasing the hydrophathy and length thresholds, users can view the placement of longer blobs composed of more hydrophobic residues in both sequence (left) and structure (right). Example shown here is Hemoglobin subunit alpha (P69905). The structure shown was predicted by AlphaFold [5].

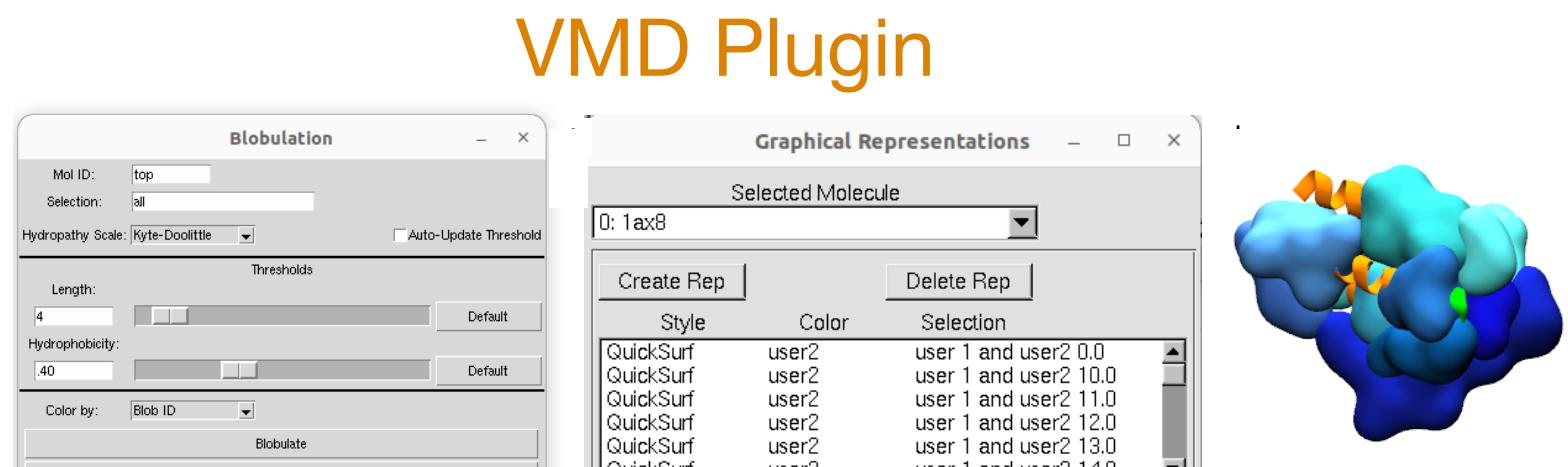


Figure 3: VMD plugin interface. Users can blobulate a selected protein using the GUI (left), which creates representations (middle) that populate in the VMD viewer window (right). Figure adapted from [6].

Example Application: α -synuclein

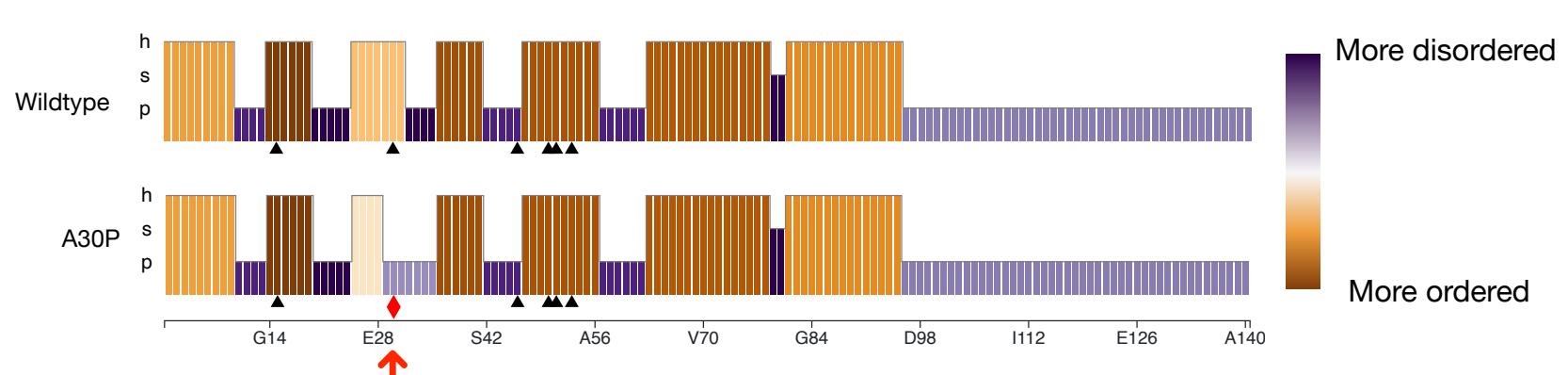


Figure 4: The effect of the A30P mutation (which is disease-associated), on α -synuclein’s blobs. A30P both shortens the third h-blob and causes it to become more disordered, showing how the effect of a mutation on the surrounding residues can be detected by the blobulator. $H^* = 0.4$, $L_{min} = 4$. Figure adapted from [6].

Example Application: GluCl

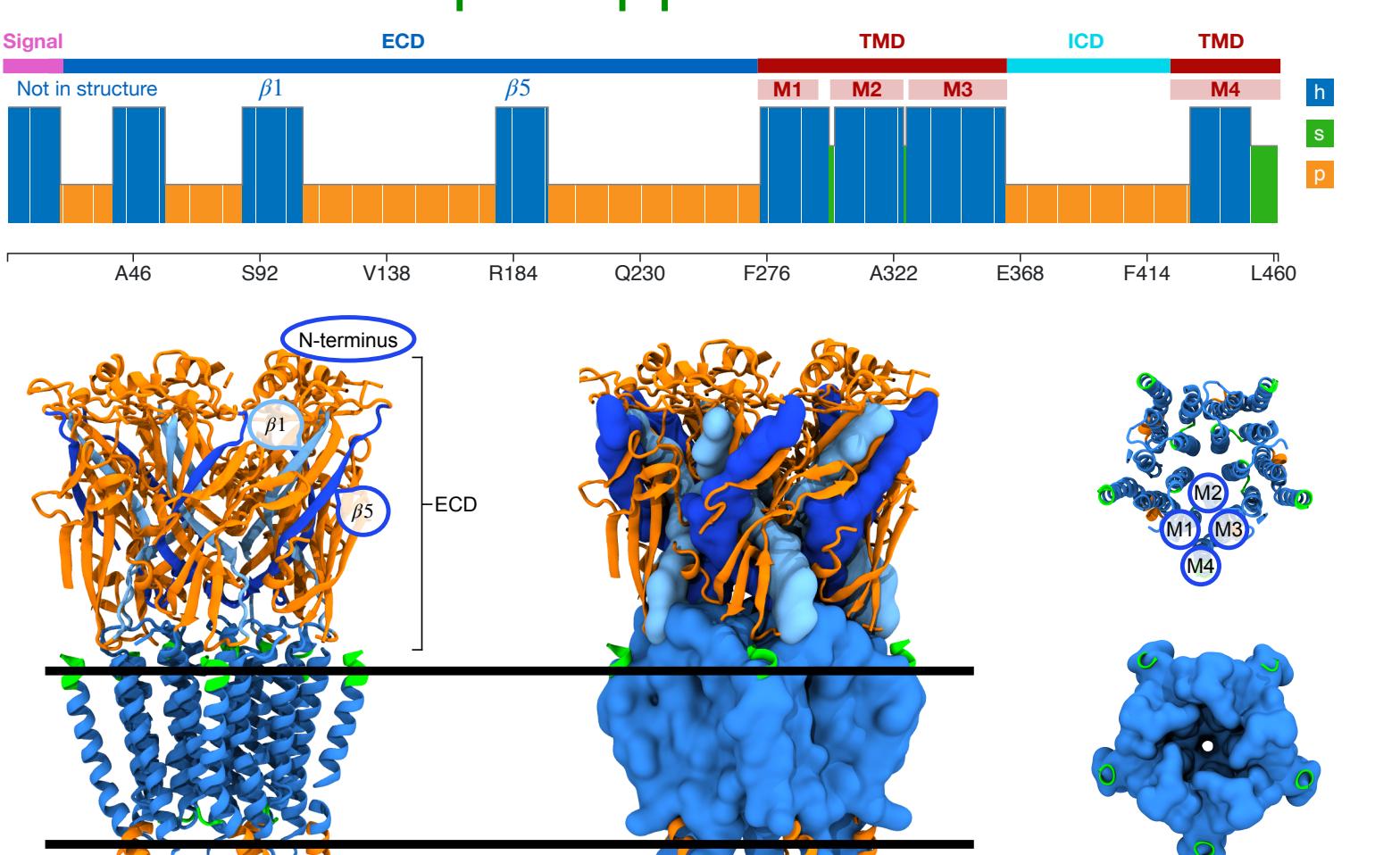


Figure 5: Blobulation of a glutamate-gated chloride ion channel (PDB: 3RHW). The transmembrane helices (M1-M4) all correspond to h-blobs (red), along with $\beta 1$ and $\beta 4$, providing an example of how structural information can be captured from sequence using the blobulator. $H^* = 0.33$, $L_{min} = 19$. The ICD is not included in the structure representation. Figure adapted from [6].

Acknowledgements

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

Citations

1. R.K. Das and R.V. Papu. *PNAS* 2013.
2. V. Uversky, J. Gillespie, A. Fink. *Proteins* 2000.
3. R. Lohia, R. Salari, and G. Brannigan. *PLOS Comp. Bio* 2019.
4. R. Lohia, M.E.B. Hansen, and G. Brannigan. *PNAS* 2022.
5. Jumper, J. et al. *Nature* 2021.
6. C. Pitman, et al. *JCIM* 2026.