



The Blobulator: Edge-Detection for Protein Sequences

Connor Pitman¹, Ezry Santiago-McRae¹, Ruchi Lohia, Kaitlin Bassi, Matthew E.B. Hansen², Thomas T. Joseph³, and Grace Brannigan^{1, 4}

¹Center for Computational and Integrative Biology, Rutgers–Camden, NJ, 08102, USA ²Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA ³Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, PA, 19104, USA

⁴Department of Physics, Rutgers–Camden, NJ, 08102, USA

Background

- Hydrophobic interactions are critical for protein structure
- Contiguous hydrophobic residues act cooperatively to create stable globular regions
- This is true even in disordered proteins and amphipathic polymers
- Many mutation prediction methods incorporate an arbitrarily defined local sequence
- Blobulation adaptively determines local sequence by clustering residues based on their hydrophobicity
- Blobulation can be used to generate hypotheses about the functional effects of mutations, including human disease-causing mutations [1]
- The Blobulator is an interactive GUI for Blobulation

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

Graphical User Interface

Given a sequence and two parameters

Minimum blob length

Hydrophobicity

Hydrophobic/hydrophilic cutoff

Introduce a mutation

Measure

Blob identity

Electrostatics

Local interactions

Likelihood of deleterious mutations

Local order

Predict

Blobulation Reveals Intrinsic Modularity

ELIC
 $H^* = 0.33$
 $L_{\min} = 19$

Lysozyme
 $H^* = 0.5$
 $L_{\min} = 8$

Alpha-Synuclein
 $H^* = 0.4$
 $L_{\min} = 4$

Figure 2: Structural representation of blobs from three different protein types: membrane (*Erwinia* ligand-gated ion channel, ELIC), intrinsically disordered (alpha-synuclein), and soluble enzyme (lysozyme). The blobs shown here were detected using parameters aimed at detecting different types of residue clusters

Blobulate your own protein!
blobulator.branniganlab.org

Algorithm

Digitization

Kyte-Doolittle Hydrophobicity

threshold H^*

Residue

Clustering

h group: sequential h blobs separated by s blobs

h-blob: At least L_{\min} sequential residues that each have hydrophathy exceeding H^*

p-blob: At least L_{\min} sequential residues between h -blobs or at termini

s-blob: Fewer than L_{\min} sequential residues between h -blobs or at termini

Figure 1: Blobulation, our sequence-based algorithm for identifying contiguous hydrophobic regions, involves two steps: digitization (left), which classifies individual residues as either hydrophobic or hydrophilic based on a cutoff H^* ; and identification of clusters (right) at least L_{\min} residues long. “H,” “P,” and “S” refer to “hydrophobic,” “polar,” and “short” blobs, respectively. Figure adapted from [1].

Hydrophobic Blobs Simplify Analysis of Intrinsically Disordered Protein Simulations

WT

V12M

A21M

Figure 3: Contacts between blobs in amyloid beta. Dashed lines represent the percentage of frames where blobs are in contact. An asterisk (*) indicates that there is a methionine in the blob

Disrupting Hydrophobic Blobs Leads to Functional Consequences

nSNPs dSNPs

Fraction enrichment of dSNPs

hydrophobicity threshold H^*

blob length L

Figure 4: Enrichment of disease-associated SNPs (dSNPs) in hydrophobic-blobs of length L , calculated with the threshold H^* for 70,000 disease- and non-disease-associated human SNPs. Reproduced from [1].

Ref Alt

Formed Split Dissolved Merged

**

Figure 5: Fractions of SNPs that modify blob topology. Fraction of nSNPs (non-disease-associated) or dSNPs (disease-associated) that change the blobular topology by either forming or dissolving, splitting, or merging a hydrophobic blobs. Based on analysis of 70,000 human exome SNPs. Reproduced from [1].

References

1. Lohia, R., Hansen, M. E., & Brannigan, G. (2022). Contiguously hydrophobic sequences are functionally significant throughout the human exome. PNAS, 119(12).
2. Lohia, R., Salari R, Brannigan G. (2019) Sequence specificity despite intrinsic disorder: How a disease-associated Val/Met polymorphism rearranges tertiary interactions in a long disordered protein. PLoS Comp. Bio.

Acknowledgements

- NRT, NSF DGE 2152059
- Busch Biomedical Foundation
- NIH 1R35GM134957
- NIH R01AR076241