

# Clustering the ensemble of an intrinsically disordered protein by contacts between specific hydrophobic blobs reveals distinguishable conformational states

**RUTGERS**  
THE STATE UNIVERSITY  
OF NEW JERSEY



Lindsey Riggs<sup>1</sup>, Ruchi Lohia<sup>1</sup>, Grace Brannigan<sup>1,2</sup>

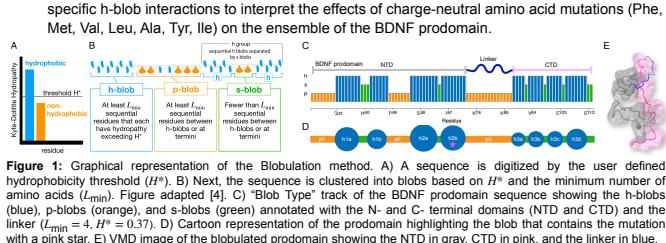
Center for Computational and Integrative Biology<sup>1</sup>, Rutgers - Camden, NJ, Department of Physics, Rutgers - Camden, NJ<sup>2</sup>

## Abstract

Intrinsically disordered proteins (IDPs) sample a broad distribution of conformational states rather than adopting a single, fixed conformation, representing the conformational ensemble. The lack of defined structure makes it challenging to interpret how a missense mutation in a sequence of an IDP affects its ensemble. Current methods for interpreting the effects of a mutation include clustering the ensemble by degrees of compactness or structural similarity but they fail to capture broader conformational trends. Here, we developed an approach that clusters an ensemble by the presence of contacts between contiguous hydrophobic regions, or h-blobs, in the long disordered brain derived neurotropic factor (BDNF) prodomain sequence. Using 2  $\mu$ s all-atomistic enhanced molecular dynamic simulations of the wild-type BDNF prodomain (V66) and mutants of the BDNF prodomain that contain a charge-neutral, hydrophobic mutation at position 66 of the amino acid sequence (A66, I66, L66, F66, Y66, and M66), we investigated how a charge-neutral missense mutation affects the ensemble of the prodomain. We observed that each mutation affects conformational characteristics of the prodomain such as compactness and intra-interactions. To interpret these results, we clustered the ensemble based on the presence of a few specific h-blob contacts and identified four soft-folded clusters within the ensemble of the prodomain: Curled, Zipped, Looped, Flared, and any states that do not display any of the h-blob contacts are classified as the Elongated cluster. We found that each mutation shifts the population of clusters in the ensemble of the prodomain. In addition, we demonstrate a high correlation between the population of soft-folded clusters in the ensemble and the total ensemble compactness. These results highlight that our cluster definition effectively captures mutation-induced effects of the conformational ensemble, providing a meaningful approach for interpreting the subtle effects of a missense mutation on the ensemble of an IDP.

## Background

- IDPs participate in transient interactions, contributing to their flexibility which enables them to adopt many heterogeneous conformations known as the conformational ensemble.
- The Val66Met mutation of the BDNF prodomain, a long IDP, has been shown to affect the prodomain ensemble and is associated with neuropsychiatric disorders [1].
- We used our purely sequence-based method, Blobulation (Figure 1), to identify which contiguous hydrophobic clusters of the prodomain, termed "blobs", stabilize transient tertiary interactions that shift the ensemble [2].
- Clustering methods can be used to decompose the ensemble into representative states to interpret the subtle effects of mutations but current approaches do not capture broad conformational trends, or meaningful dominant states [3].
- We will develop a computational approach that clusters states based on the presence of a few specific h-blob interactions to interpret the effects of charge-neutral amino acid mutations (Phe, Met, Val, Leu, Ala, Tyr, Ile) on the ensemble of the BDNF prodomain.



## Research Questions

- Can the conformations of BDNF prodomain be meaningfully clustered by a few contacts between blobs?
- What are the dominant states within the conformational ensemble of each mutant of the BDNF prodomain?
- How does each mutation at site 66 stabilize conformational states within the ensemble?

## Approach

- Simulate charge-neutral mutations (F, M, V, L, A, Y, I) using temperature-replica exchange molecular dynamics for 2  $\mu$ s each.
- Analyze conformational characteristics by measuring the ensemble compactness (radius of gyration) and contact frequency between blobs.
- Cluster the prodomain ensemble based on the presence of interactions between a few selected h-blob pairs.
- Determine cluster stabilization by measuring residue-level contacts.

## Prodomain is sensitive to a charge-neutral mutation at site 66

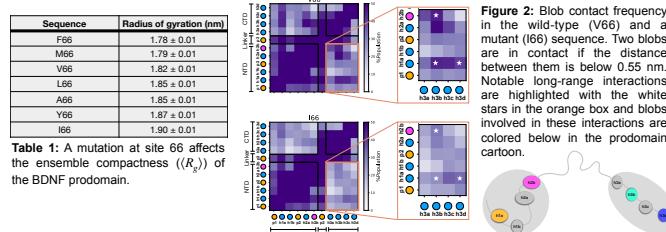
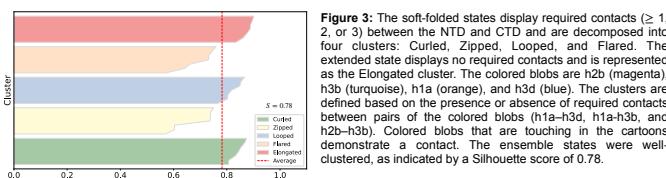
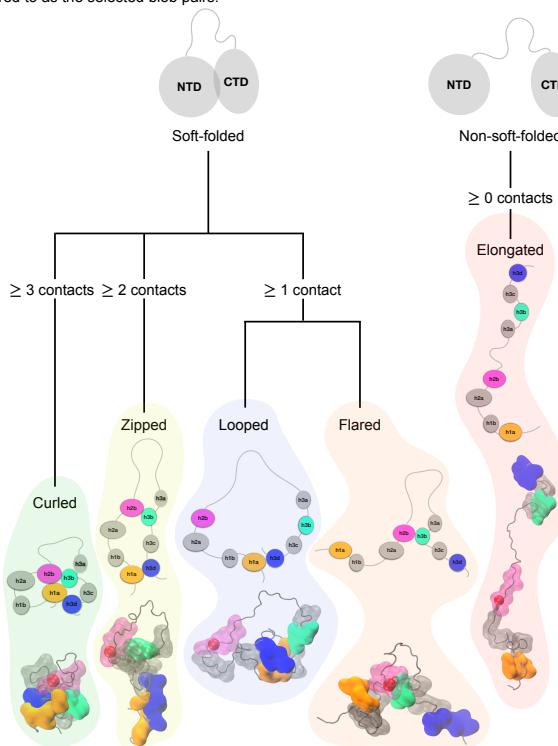


Table 1: A mutation at site 66 affects the ensemble compactness ( $\langle R_g \rangle$ ) of the BDNF prodomain.

## Blob-based clustering reveals meaningful and distinct states

We defined five conformational clusters based on the presence or absence of a few interactions (contacts) between two non-adjacent h-blobs that span the p3 linker and are highlighted in Figure 2: h1a-h3d, h1a-h3b, and h2b-h3b. These specific h-blob pairs are referred to as the selected blob pairs.



## Cluster compactness correlates with the number of contacts between selected blob pairs

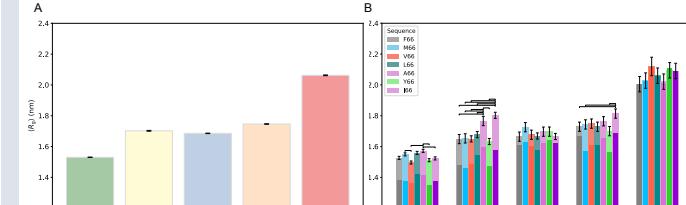


Figure 4: A) The cluster compactness ( $\langle R_g \rangle$ ) is correlated to the number of required contacts for a cluster ( $r = -0.88$ ). B)  $\langle R_g \rangle$  of each sequence by cluster. Darker colored bars represent  $\langle R_g \rangle$  of all the h-blobs in each sequence while the lighter colored bars represent  $\langle R_g \rangle$  of each simulated residue in the sequence. Generally,  $\langle R_g \rangle$  is similar across all sequences in each cluster with a few exceptions. Within the Curled, Zipped, and Flared clusters, there are one or two sequences that stand out: the most significant differences involves either I66 or A66; The one exception is M66 vs V66 in the Curled cluster. Data on the x axis are arranged by the sequence with the lowest ensemble  $\langle R_g \rangle$  (F66) to the highest (I66).

## Ensemble compactness correlates with % of soft-folded clusters

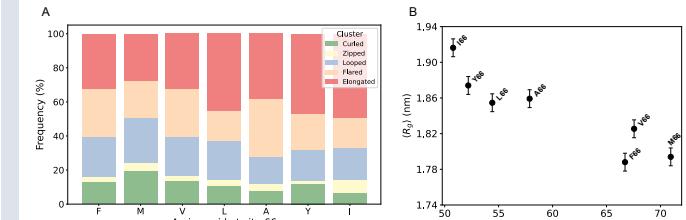


Figure 5: A) The mutation at site 66 affects the frequency of clusters in the ensemble of the BDNF prodomain. Data on the x axis are arranged by the amino acid that influences the lowest ensemble compactness ( $\langle R_g \rangle$ , F) to the highest (I), as shown in Figure 3A. B) The ensemble ( $\langle R_g \rangle$ ) is correlated to the frequency of soft-folded clusters in the prodomain ensemble ( $r = -0.92$ ). Soft-folded clusters include the Curled, Zipped, Looped, and Flared clusters.

## Summary

- Each missense mutation affects ensemble compactness of the BDNF prodomain.
  - Clustering the ensemble by the presence of interactions between the selected h-blob pairs revealed distinguishable soft-folded states indicated by a Silhouette score of 0.78.
  - The ensemble compactness is correlated to the population of soft-folded states present in the ensemble.
  - These results provide a meaningful clustering approach that enables interpretation of conformational effects due to a missense mutation.
- Future work: Which residue interactions or transient secondary structural elements stabilize each cluster?

## References

- M. Egan et al. *Cell*. 2003.
- J. Zhu, X. Salvatella, and P. Robustelli. *Nat. Commun.*, 2022.
- R. Lohia, R. Salari, and G. Brannigan. *PLoS Comput. Biol.*, 2019.
- R. Lohia, M.E.B. Hansen, and G. Brannigan. *PNAS*, 2022.

## Acknowledgements

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