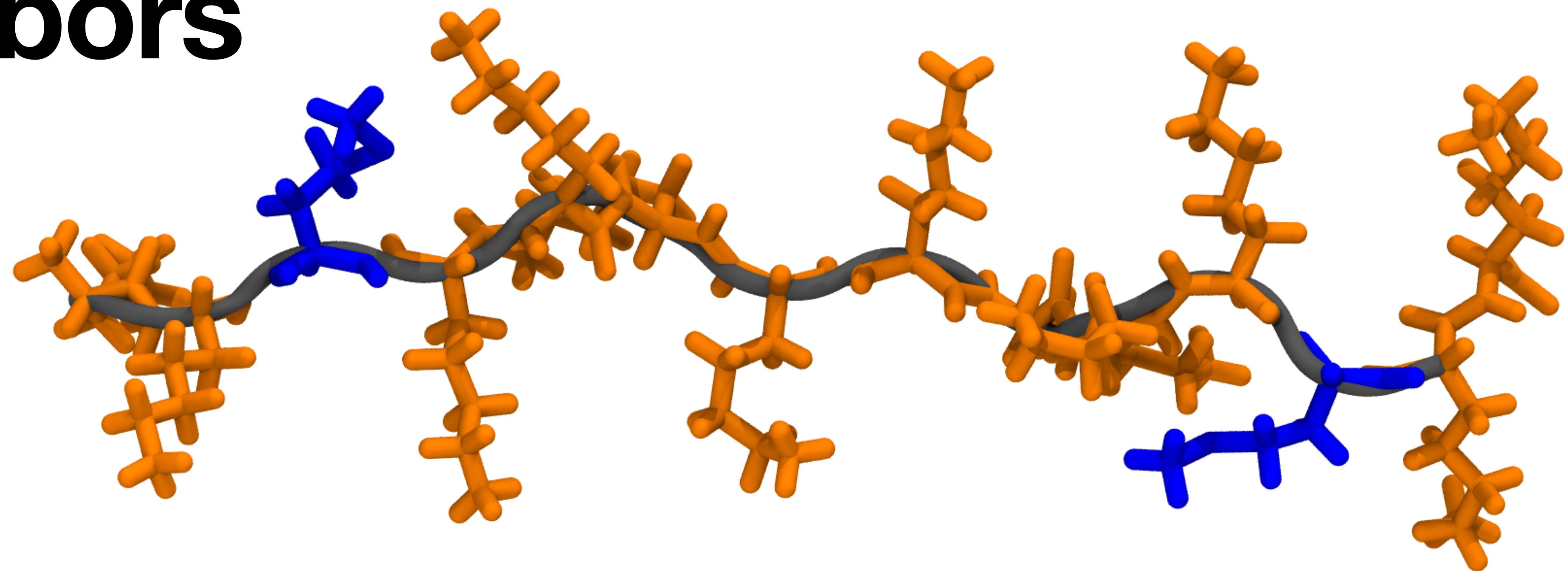


Good Fences Make Distant Neighbors



Determining the effect of local sequence context on intra-protein interactions

Connor Pitman
Brannigan Lab
CCIB Seminar - April 25, 2023

Agenda

1. Background and Preliminary Work
2. Research Questions
3. Results
 1. Methionine interactions in different blob types
 2. The blobulator GUI
 3. Example applications of blobulation
4. Summary/Conclusion

Intrinsically disordered proteins lack a well-defined structure

IDPs

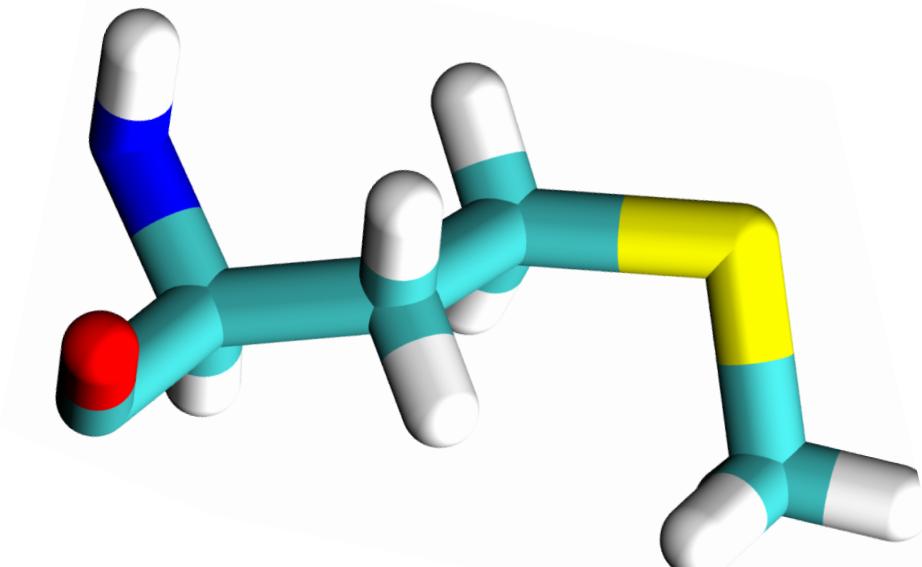
BDNF prodomain

Mutations to IDPs can cause functional effects

- 21.7% of disease-associated mutations found in IDRs (Vacic et al., PLOS, (2012))
 - BDNF prodomain: V66M
 - Tumor protein p63: R243W
 - Alpha synuclein: A53T
- ~33% of proteins in the Eukaryotic proteome contain IDRs

Methionine can have specific interactions with other Methionines

- Methionine:
 - Start codon
 - Hydrophobic
 - Sulfur containing
- Specific **hydrophobic** interactions
- Found in contact with other methionine residues in solved protein structures



Hypothesis:

Mutations involving methionine can cause intra-protein contact shifts through Methionine-Methionine (Met-Met) interactions

Challenge:

How do we measure intra-protein contact shifts?

Quantifying the extent of sequence changes, step 1: break “sentences” into “words”

QEB NRFZH YOLTK CLU GRJMP LSBO QEB IXWV ALD

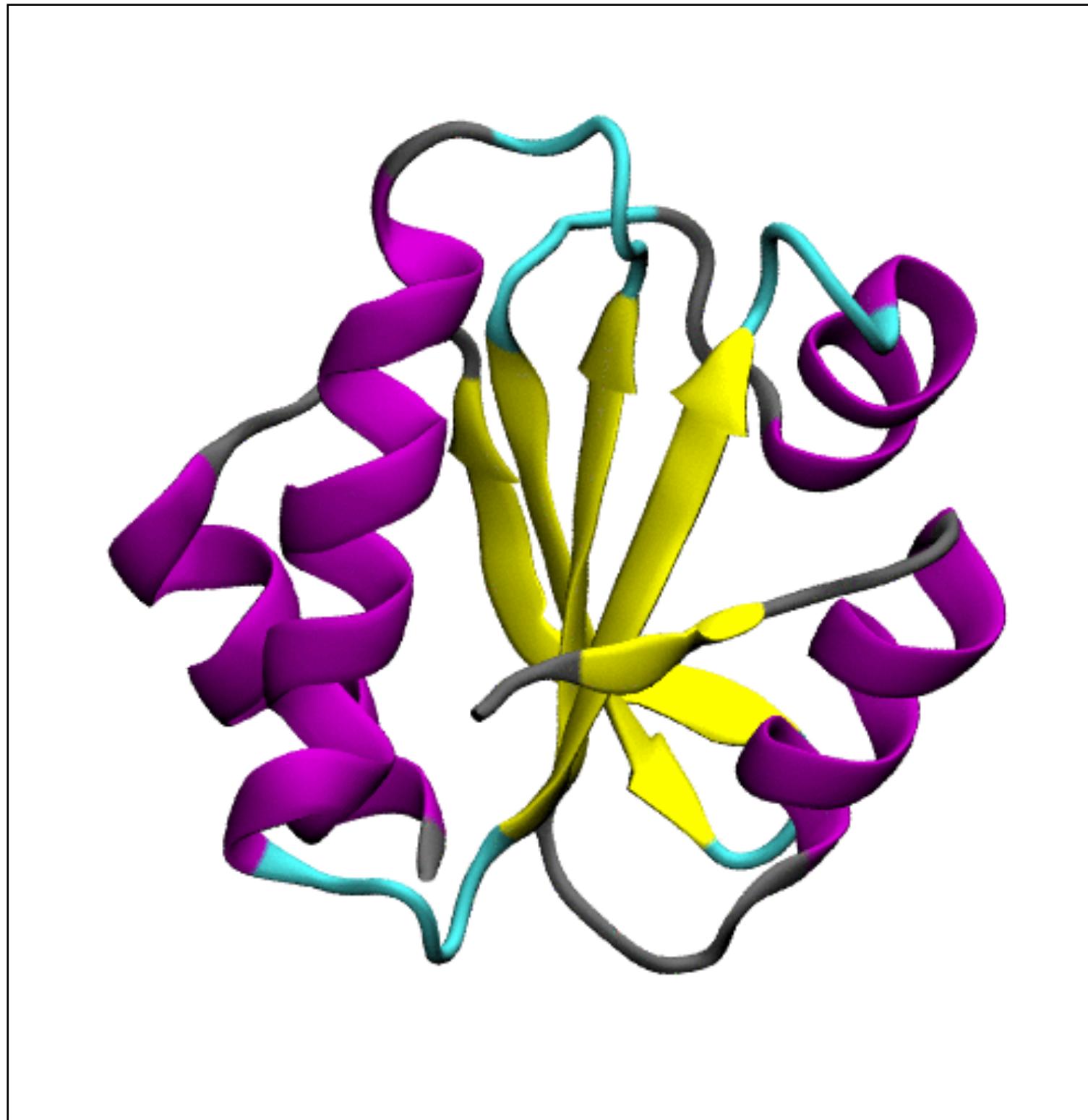
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG

QEBNRFZHYOLTKCLUGRJMPLSBOQEBIXWVALD

THEQUICKBROWNFOXJUMPEDOVERTHELAZYDOG

Classic approach: use secondary structure elements as “words”

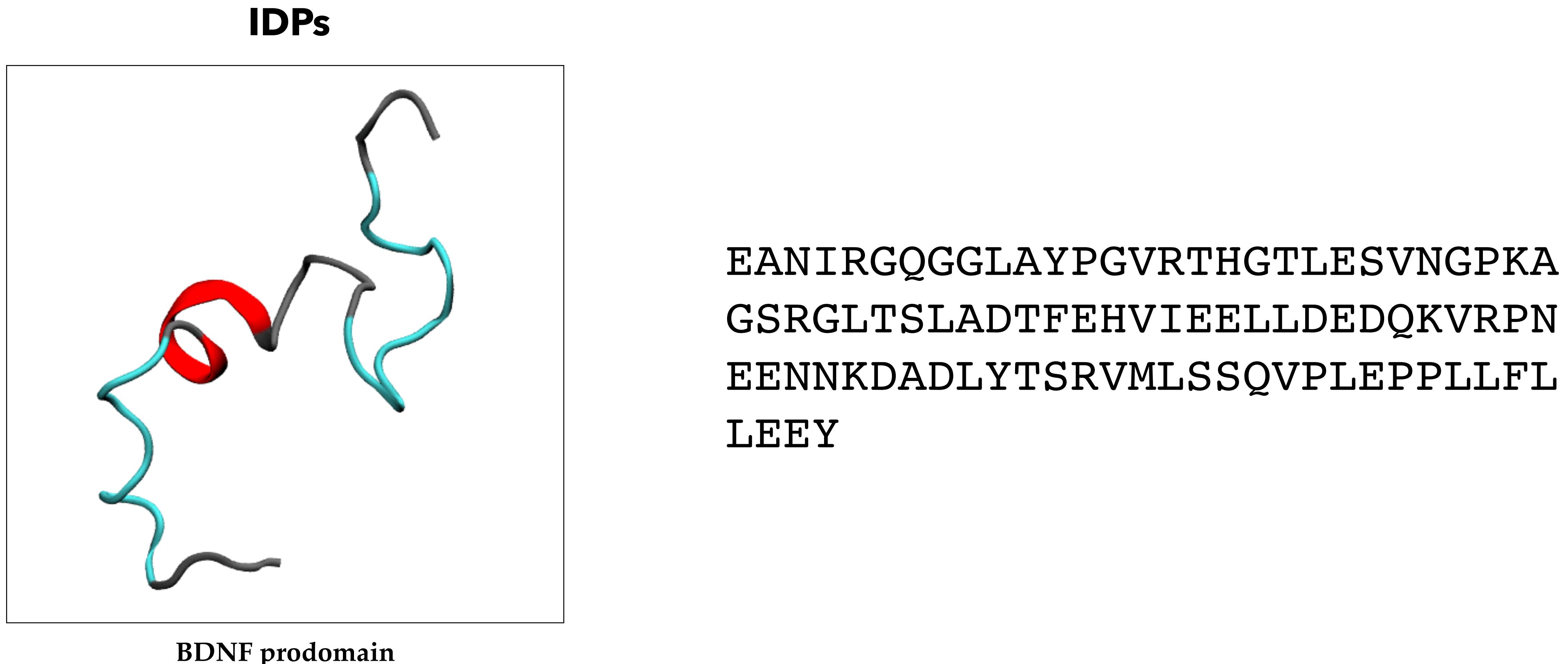
Structured



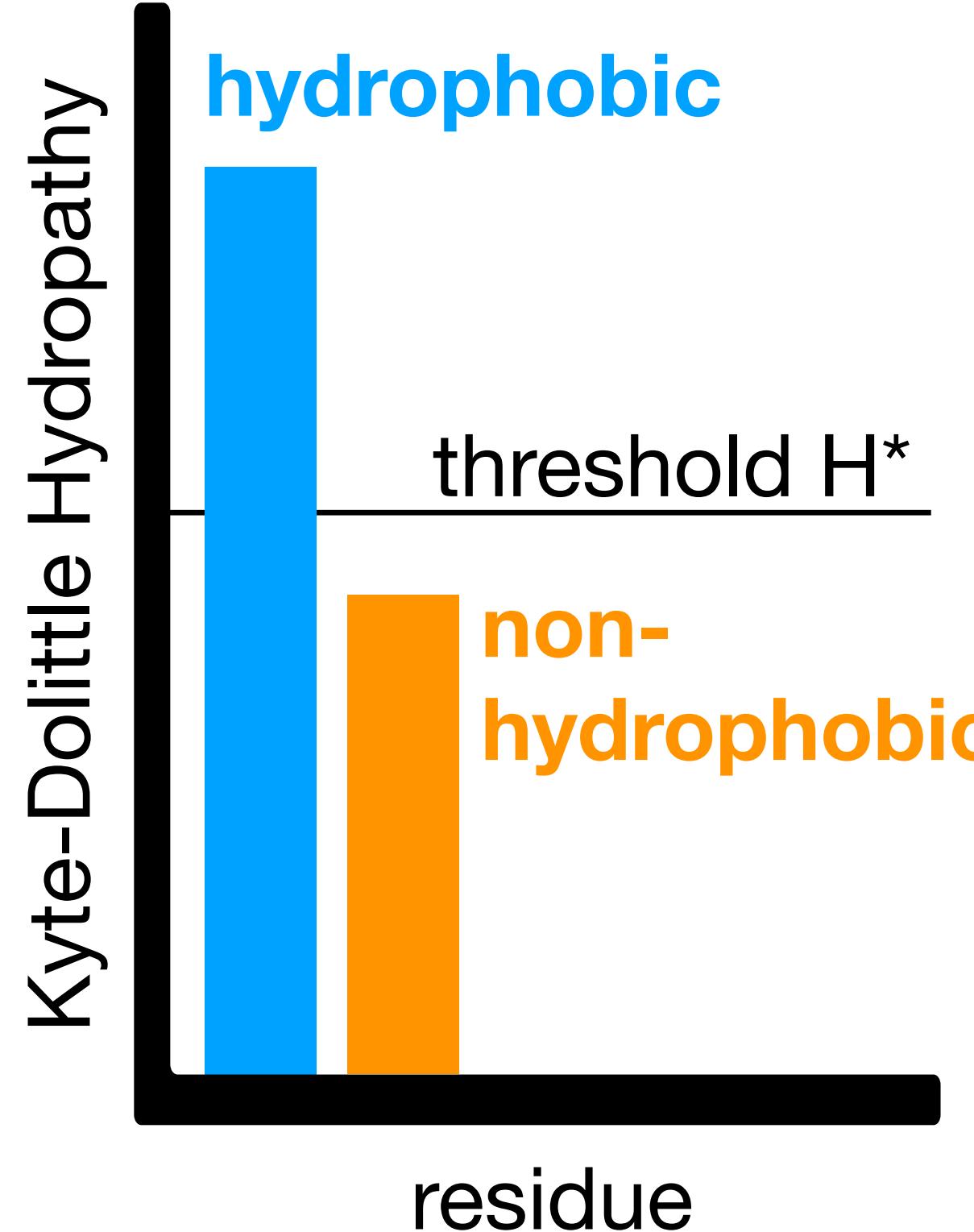
yeast Thioredoxin

MV TQL KS ASEYDSAL ASG DKLVVVDFF
ATWC GPCKMIAPMIEKFAEQ YSD AAFYKLD
VDEV SDVAQK AEVS SMPTLIFYK GG
KEVTRVVGA N PAAIKQAIA SNV

Frustrating problem: IDPs



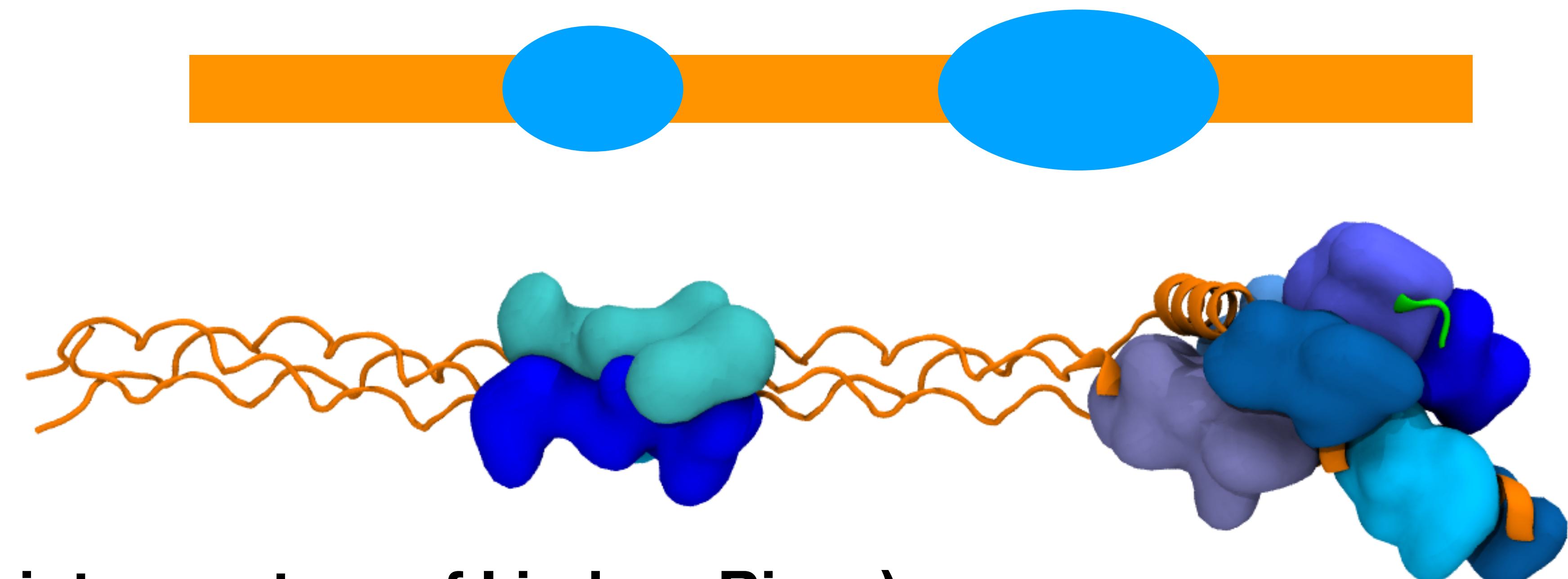
We needed a different approach for word detection: blobulation



EANIRGQGG**LAYPGVRTHGTLESVN**NGPKAGSR

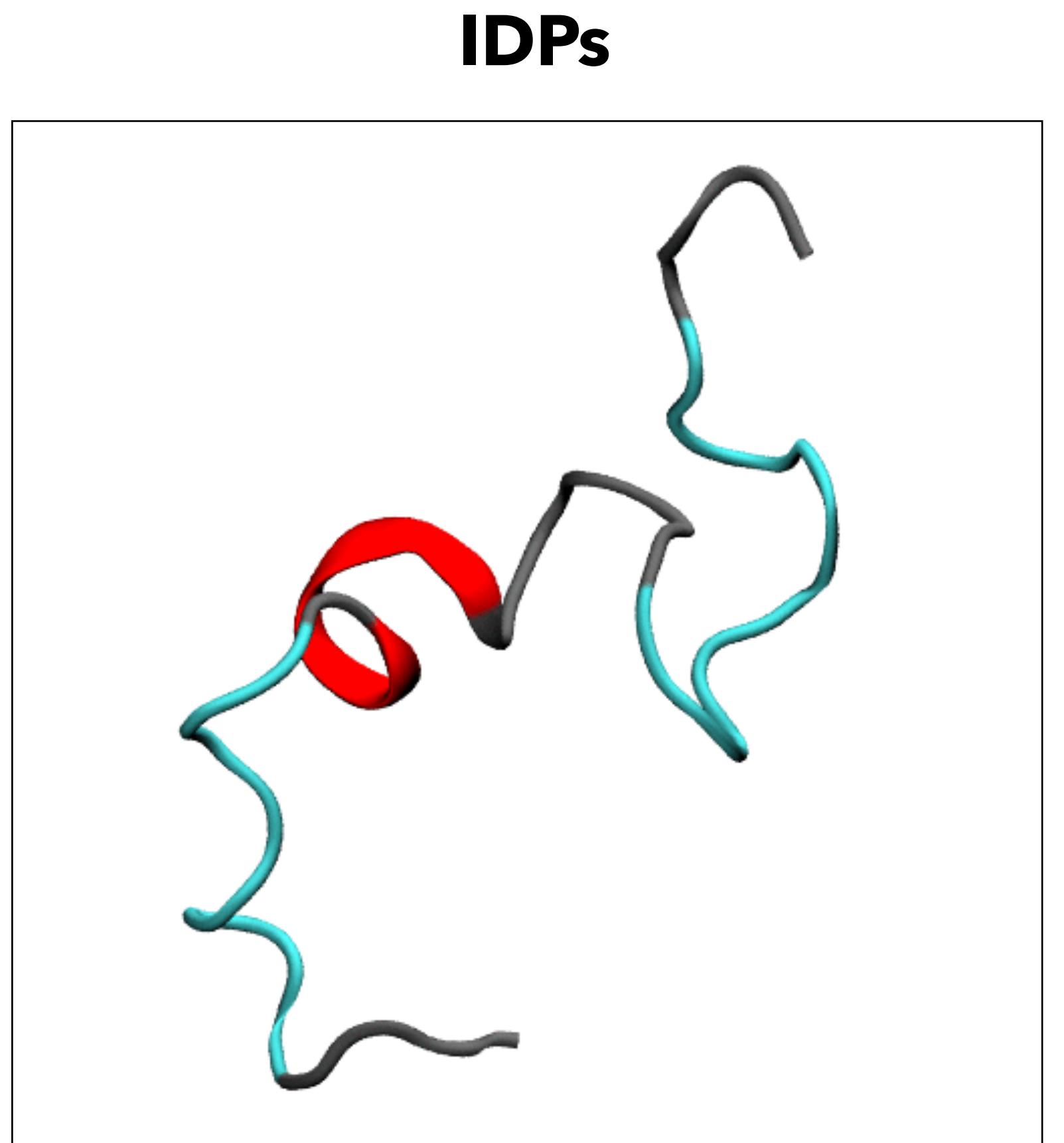
─ L_{min} ─

EANIRGQG**GLAYPGVRTHGTLESVN**NGPKAGSR



(VMD Blob coloring scripts courtesy of Lindsey Riggs)

Blobulation first use-case: the BDNF prodomain



EANIRGQG GLAY PGVRTHG TLESVN
GPKAGSR GLTSLADTF E HVIE
ELLDEDQKVRPNEENNKDADLYTS
RVMLS S QVPLE P PLLFLL EEY

h-blobs are functionally necessary

PNAS

RESEARCH ARTICLE

BIOPHYSICS AND COMPUTATIONAL BIOLOGY



Contiguously hydrophobic sequences are functionally significant throughout the human exome

Ruchi Lohia^{a,b} , Matthew E. B. Hansen^{c,1} , and Grace Brannigan^{a,d,1,2}

Edited by Ken Dill, Stony Brook University, Stony Brook, NY; received September 12, 2021; accepted February 2, 2022

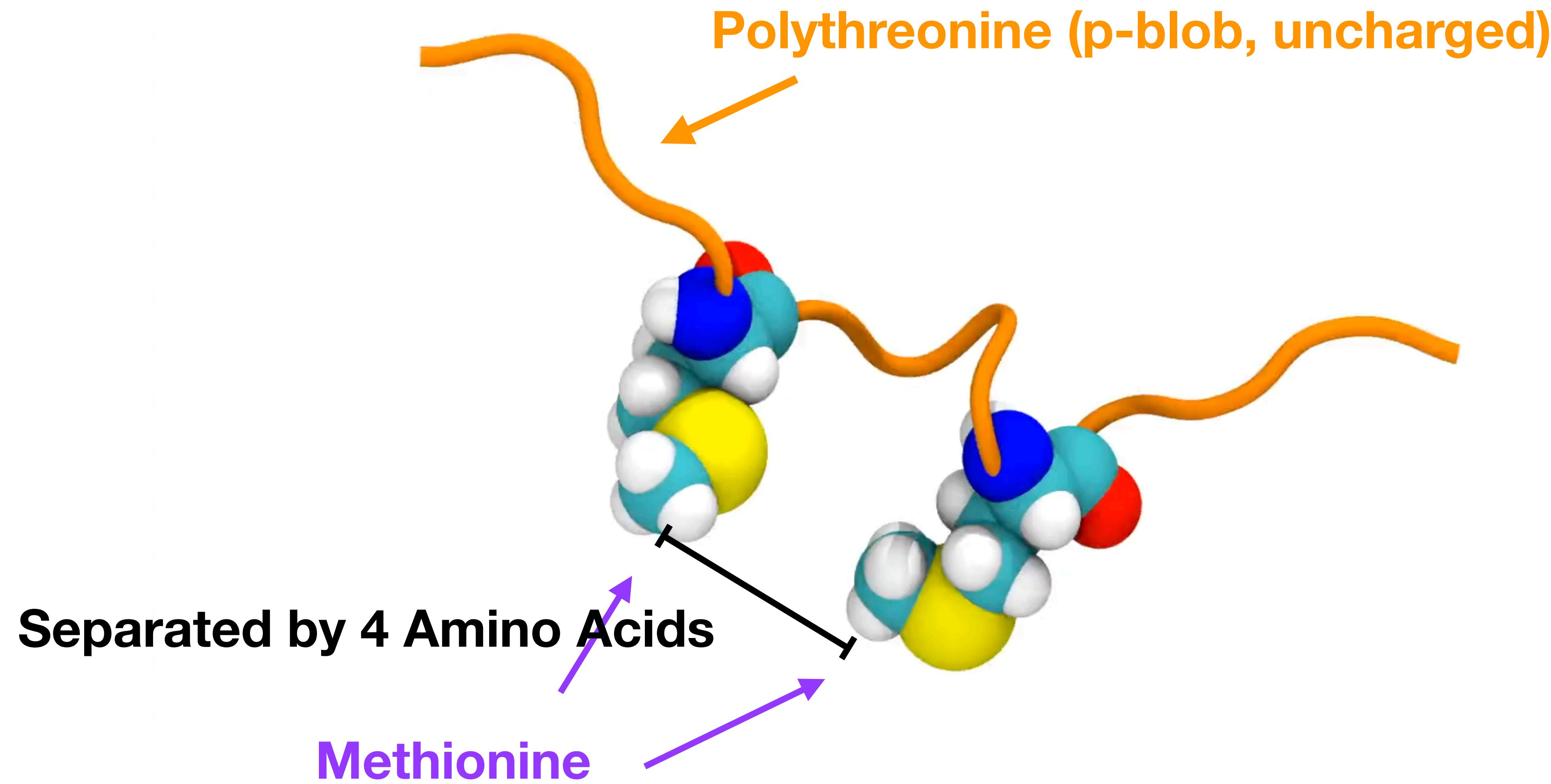
Research Questions:

- 1. Are Met-Met interactions dependent on whether the Methionines are in a h or p blob?**
- 2. How can blobulation give us code-breaking information about other proteins?**

Approach: Short peptide simulations

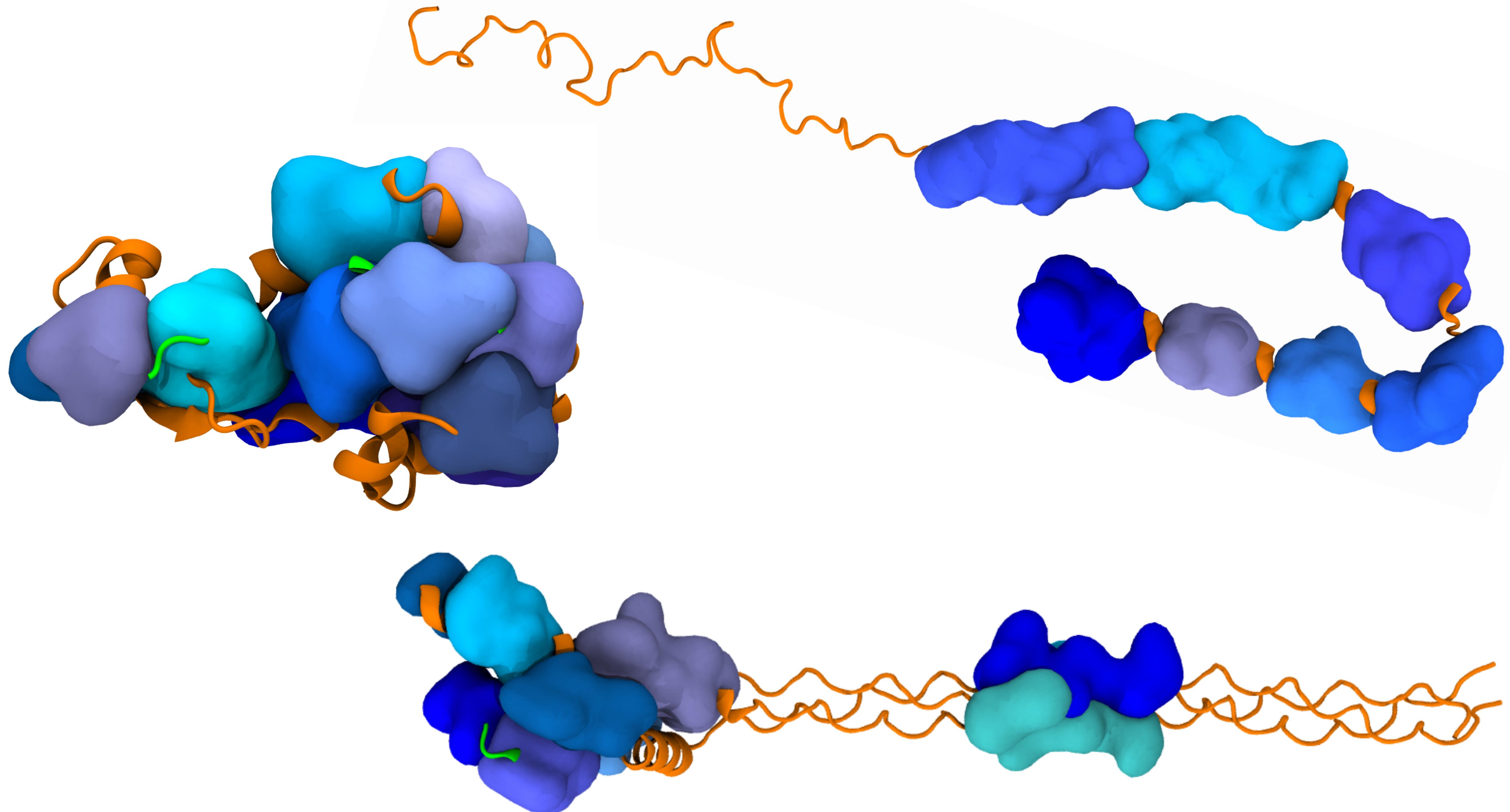
Methionines separated by...

Approach: Peptide simulations



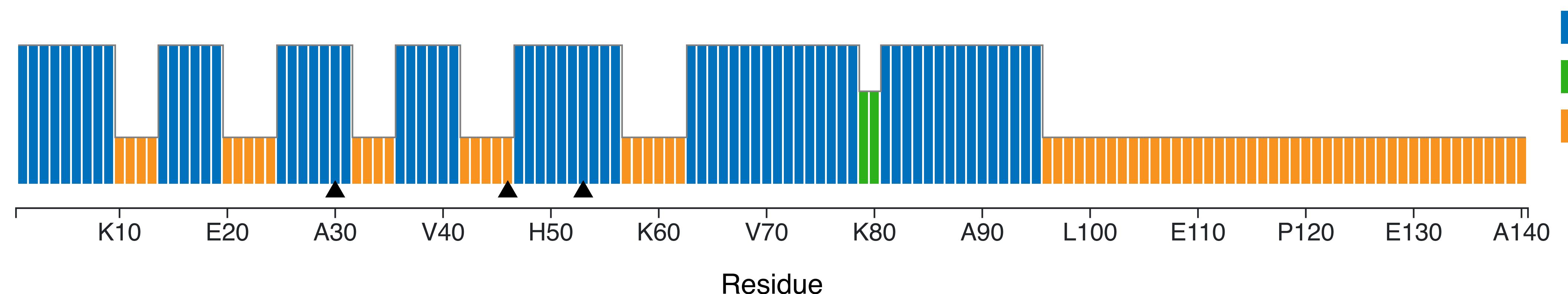
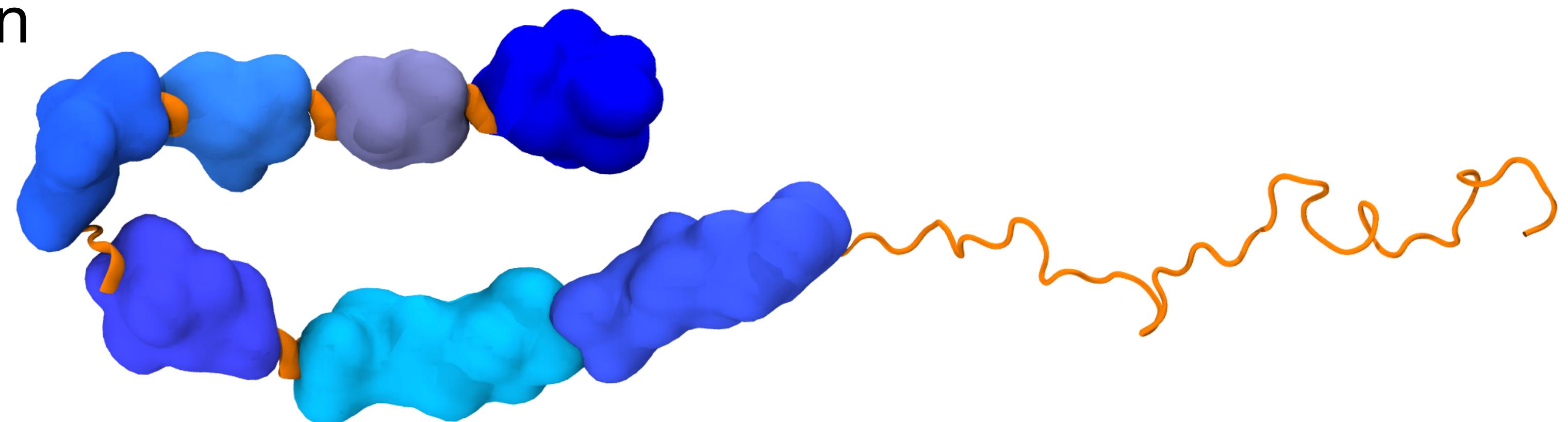
Met-Met interactions differ between blob types

What about other proteins?



Alpha-synuclein

- Present in plaques formed in Parkinsons' patients
- Aggregation
- Intrinsically disordered
- Studies involving disease-associated SNPs

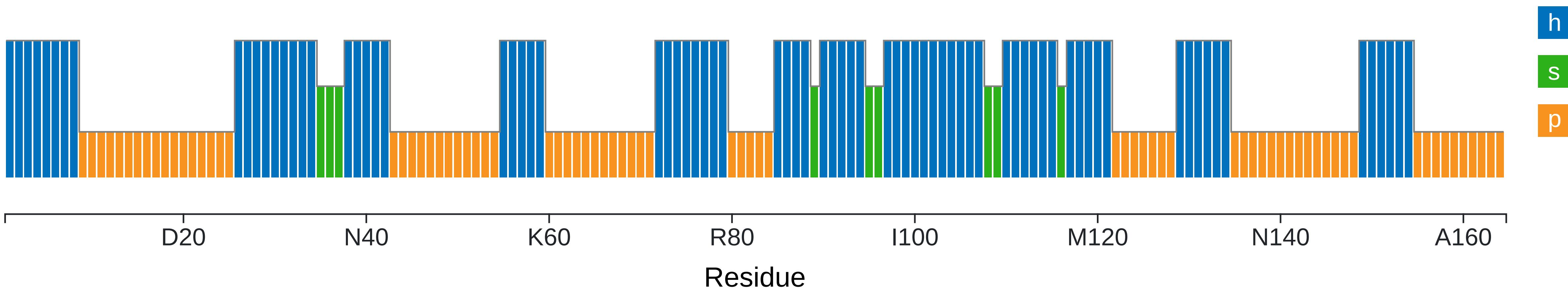
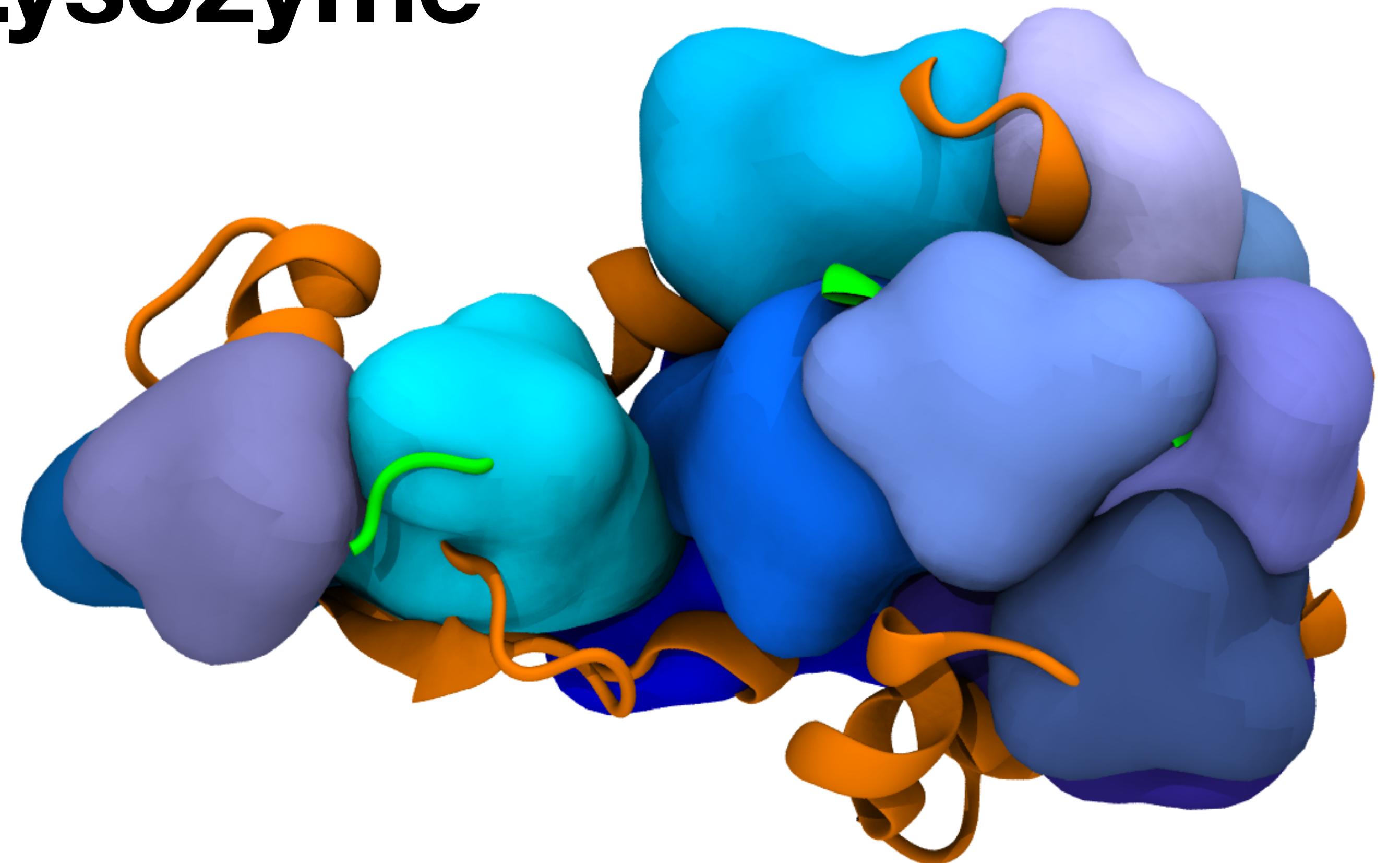


The Blobulator: A Graphical User Interface for blobulating any protein

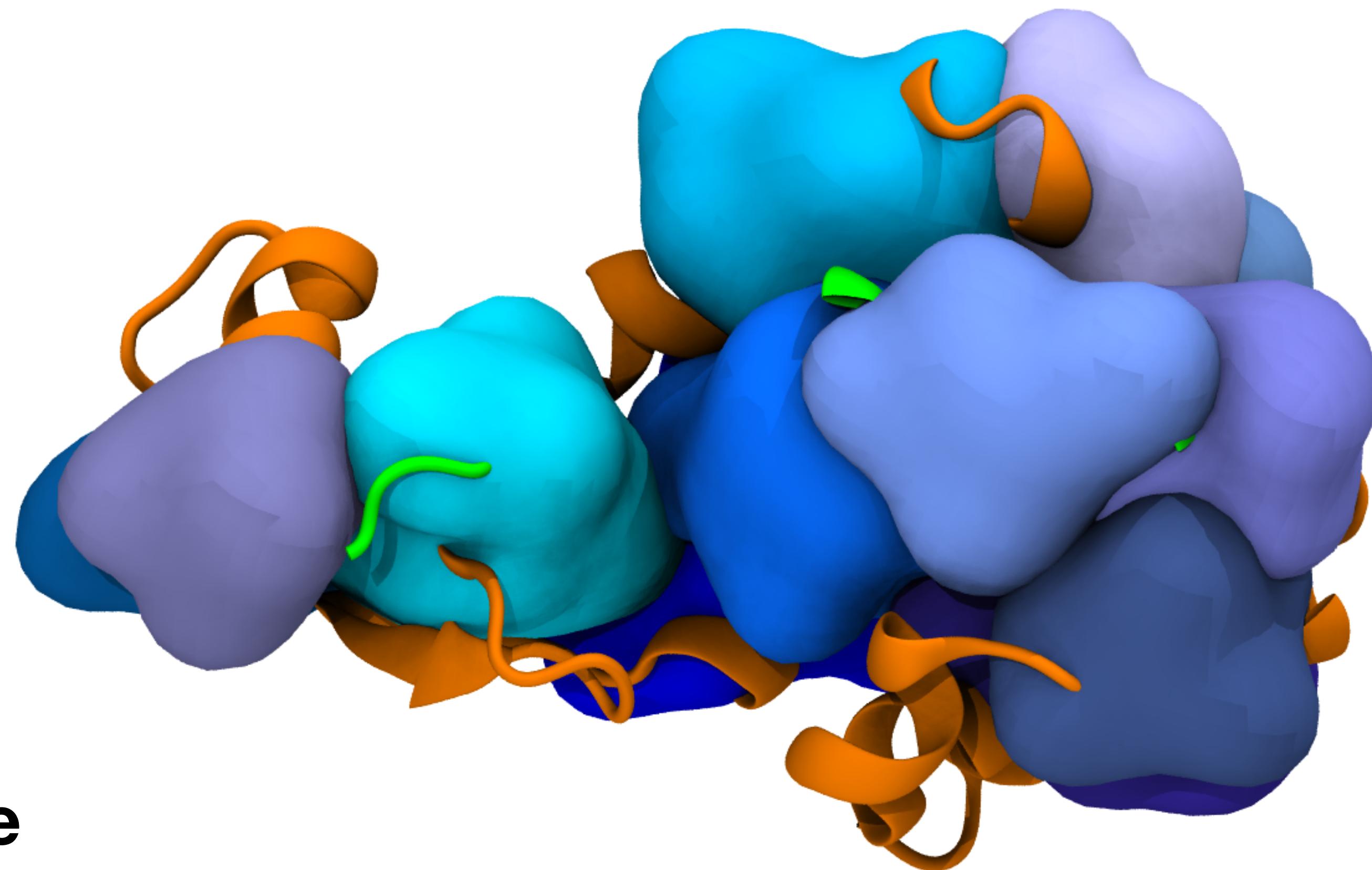
<https://www.blobulator.branniganlab.org/>

Bacteriophage T4 Lysozyme

- Enzyme, structured protein
- Breaks down peptidoglycan in bacteria cell walls
- Several studies on mutants which affect temperature sensitivity
- Looking for important residue interactions

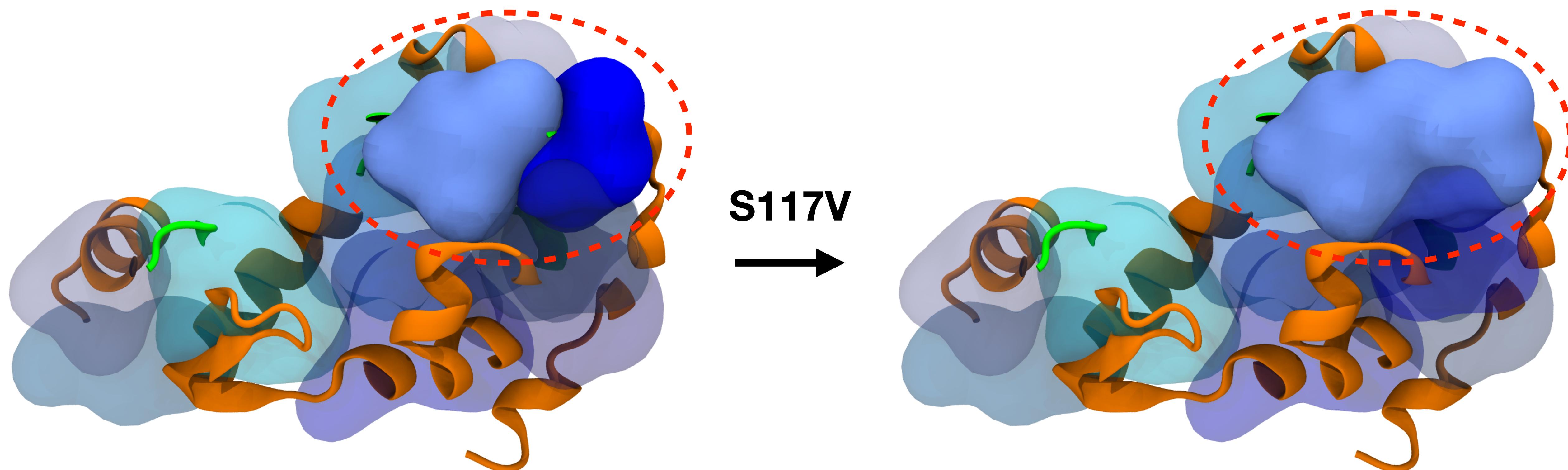


Blob-altering changes alter the temperature sensitivity of T4 Lysozyme



Wildtype

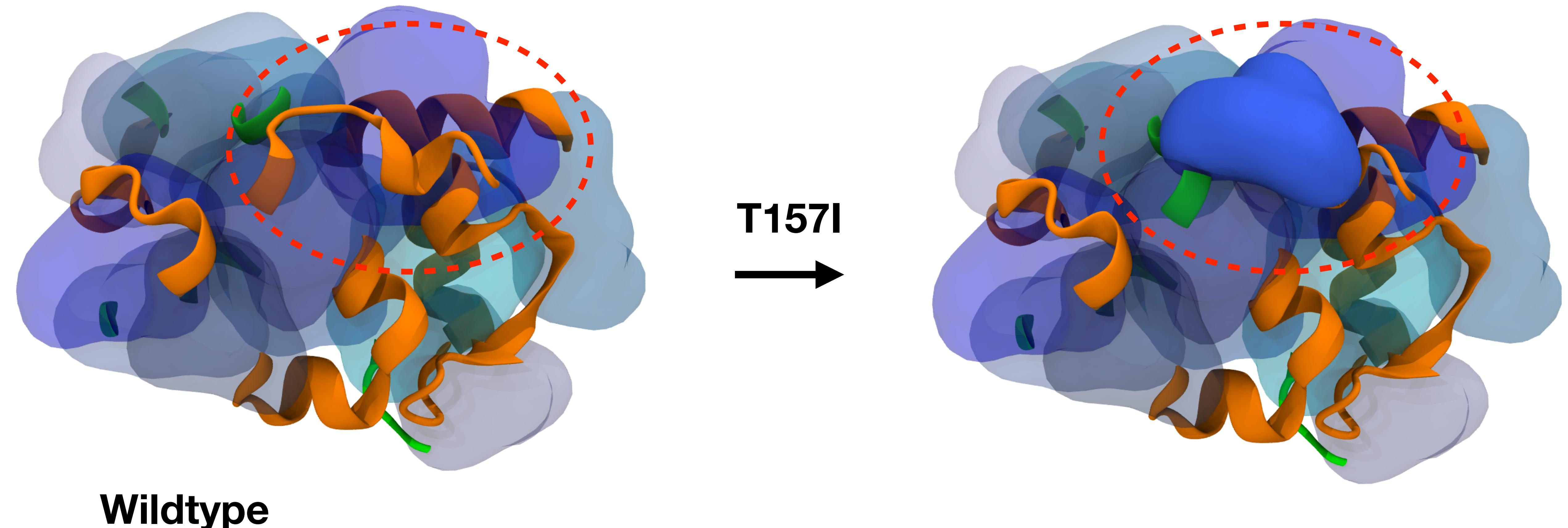
Blob-altering changes alter the temperature sensitivity of T4 Lysozyme



Wildtype

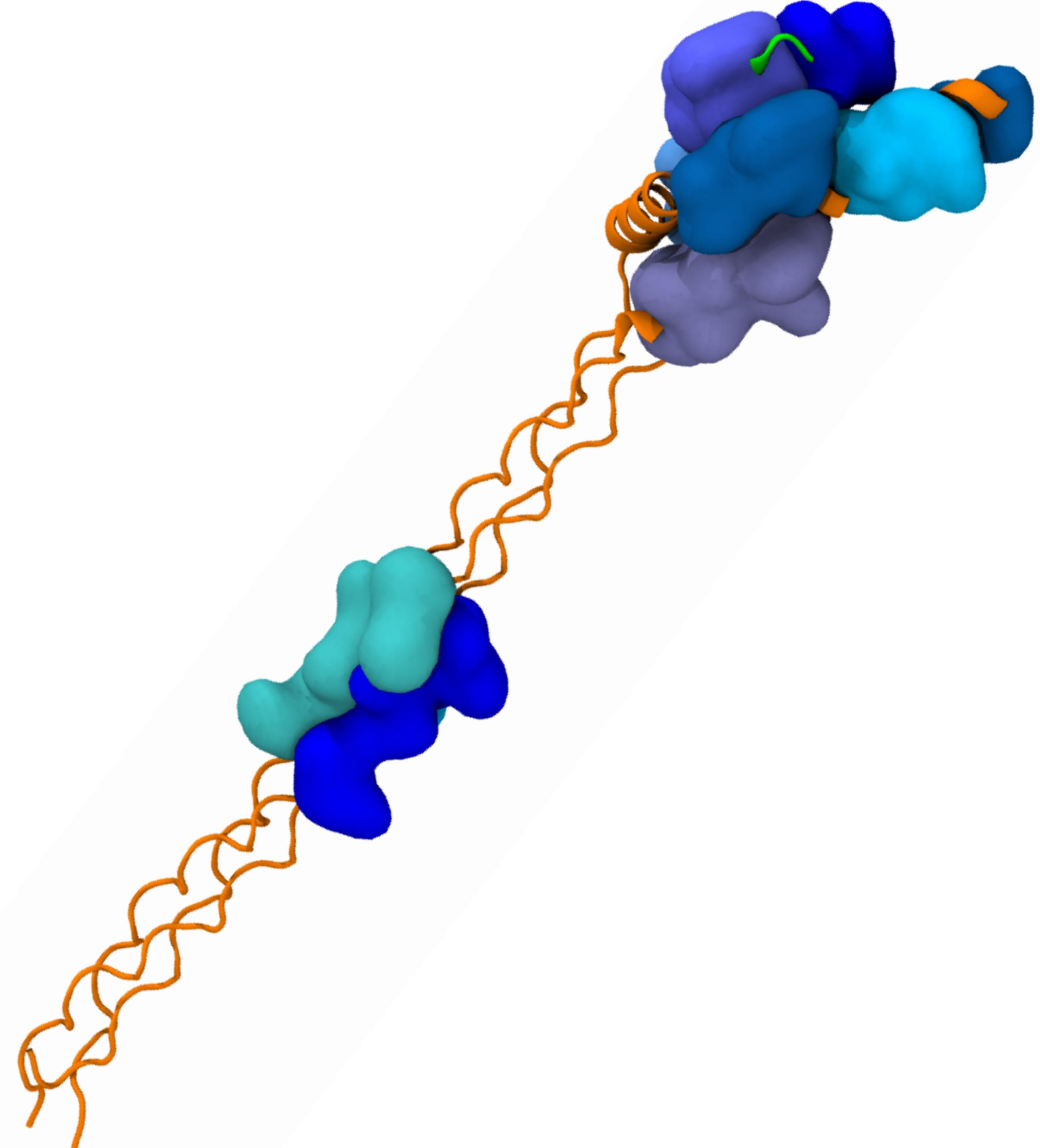
Increased stability at higher temperatures

Blob-altering changes alter the temperature sensitivity of T4 Lysozyme



Collagen

- Structural Proteins
- Most abundant protein in mammals
- Notable for their connecting roles in tissues such as cartilage and muscle
- Can form fibrils



Human and Mouse Collagen type 9

Sequence logo showing the alignment of Human and Mouse Collagen type 9 amino acid sequences. The x-axis represents amino acid positions from 1 to 900. The y-axis shows the frequency of each amino acid at each position.

The sequences are color-coded by residue:

- A (green)
- C (blue)
- D (red)
- E (orange)
- F (yellow)
- G (light green)
- H (purple)
- I (pink)
- K (brown)
- L (tan)
- M (grey)
- N (light blue)
- P (light pink)
- Q (light orange)
- R (dark purple)
- S (light green)
- T (light pink)
- V (light green)
- W (yellow)
- Y (orange)

Key features of the sequence logo include:

- Extracellular domain:** Positions 1-150, colored primarily in shades of green, yellow, and orange.
- Transmembrane domain:** Positions 151-300, colored in shades of grey and brown.
- Intracellular domain:** Positions 301-900, colored in shades of green, yellow, and orange.
- Conservation:** Amino acids at positions 151-300 are highly conserved across both species, with most positions being grey or brown.
- Termination:** Both sequences end at position 900, indicated by a vertical bar at the bottom.

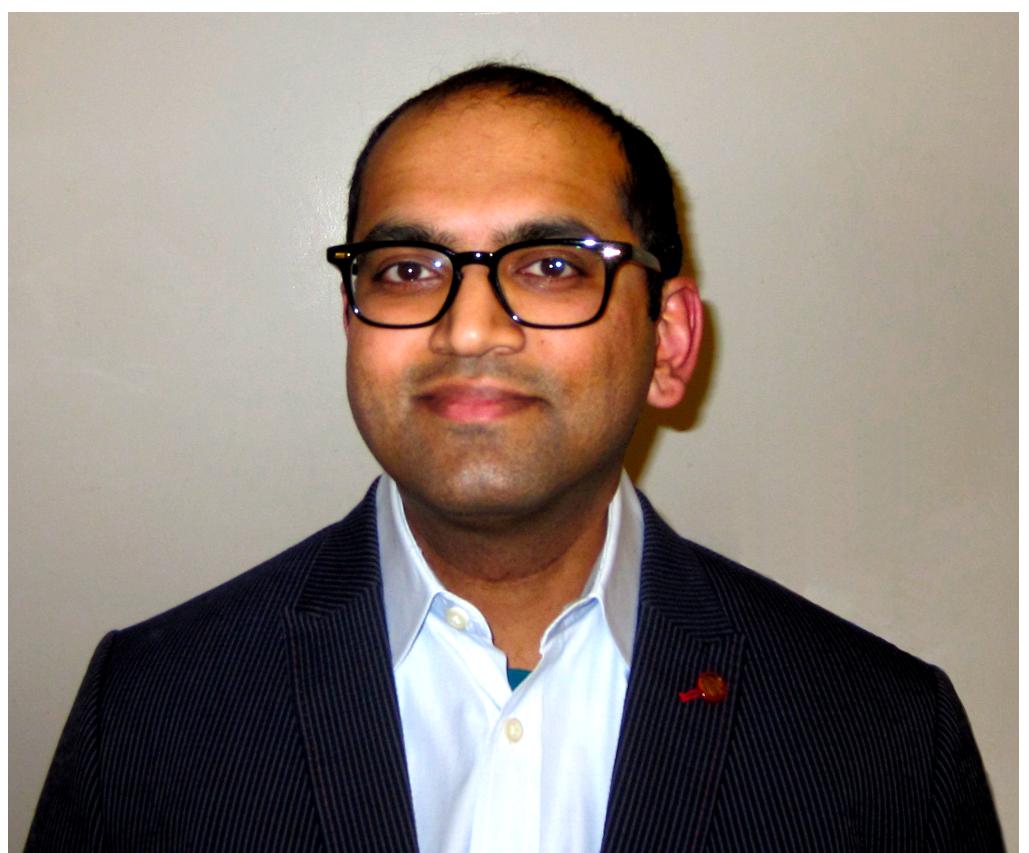
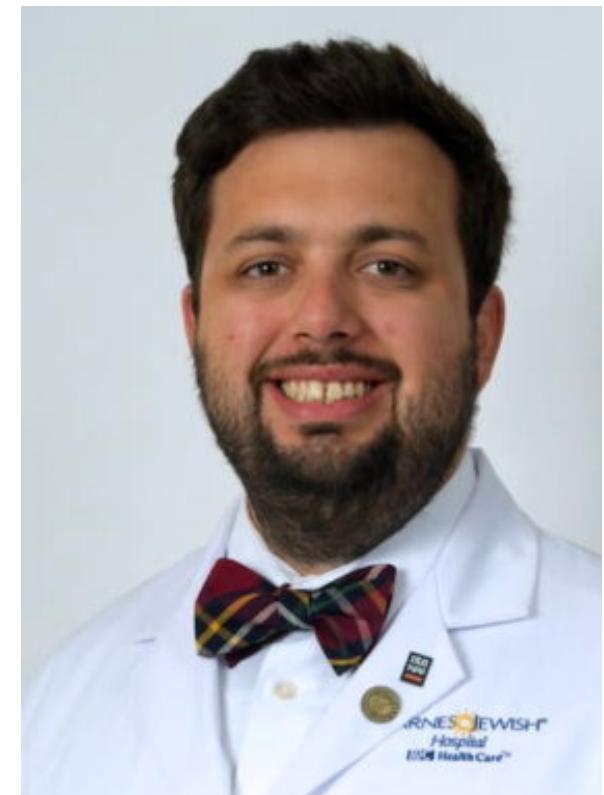
Summary

- Met-Met interactions are context dependent
- Blobulation can give us clues to how changes to proteins will affect them

Future Directions

- Blobulator improvements:
 - Genomics-focused features
 - Easier to complete batch-processes
- Using genomic clues to detect specialty interactions in IDPs

Acknowledgements



RUTGERS

Office of Advanced Research
Computing

Questions?