

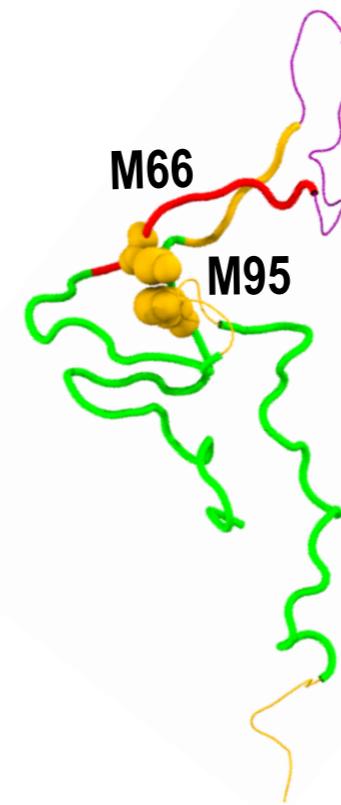
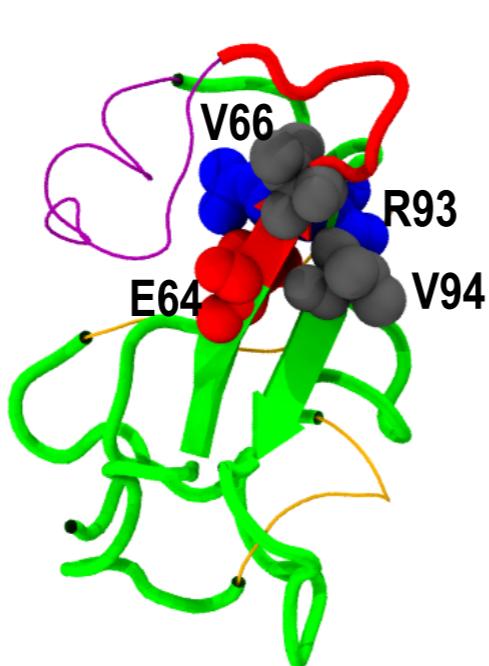


Predicting the effect of genetic variance on the sequence ensemble relationship of intrinsically disordered proteins

Ruchi Lohia

Grace Brannigan

Center for Computational and Integrative Biology (CCIB)



Thesis Theme

- A. General questions about the role of neutral residues in long intrinsically disordered proteins
- B. Specific questions about the mechanism underlying one particularly significant residue (Val66Met) in one particularly significant protein (BDNF)

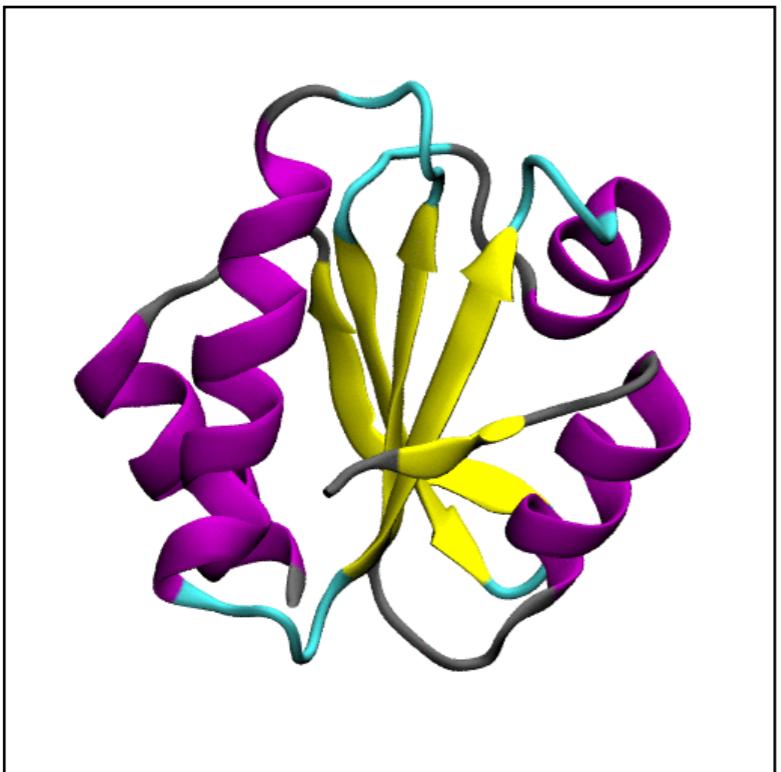
Thesis Outline

1. Introduction
 - A. Intrinsically disordered proteins (IDPs)
 - B. Particularly significant SNP (Val66Met) in one particularly significant protein (BDNF)
2. Development of sequence-decomposition based hierarchical framework on BDNF and Val66Met SNP
3. Application of hierarchical analysis to other mutations in BDNF
4. Sequence-decomposition applied to a database of Mendelian SNPs

Intrinsically Disordered Proteins (IDPs)

Structured

X-ray crystallography,
NMR

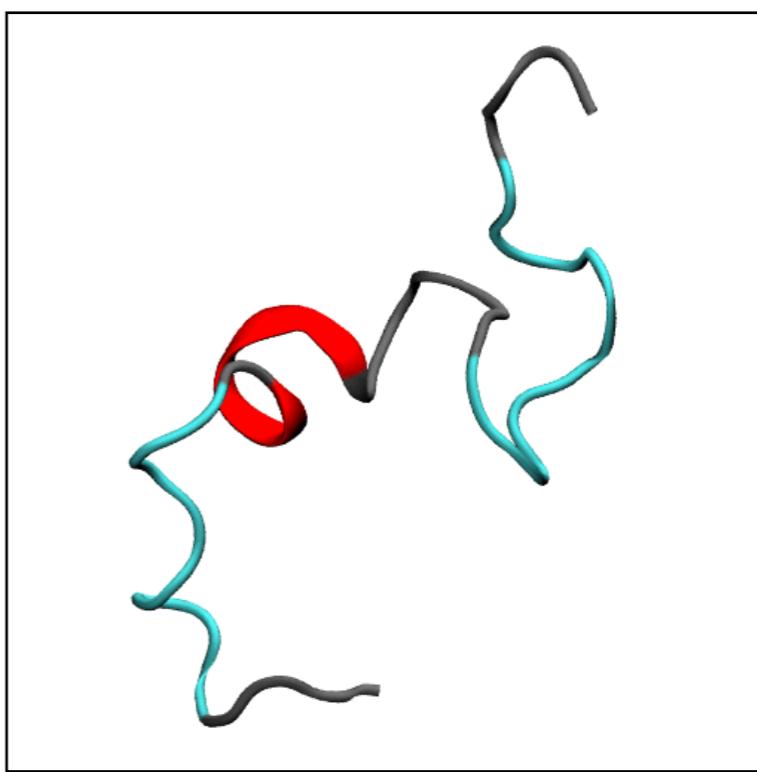


2n5a (nmr)
yeast Thioredoxin

sequence -> structure

IDPs

NMR (secondary)



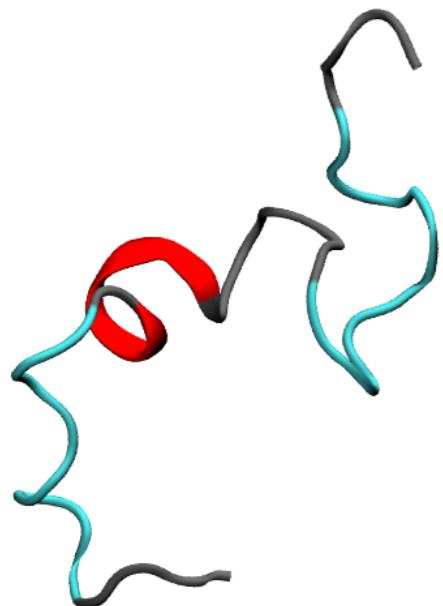
prodomain (MD)

sequence -> ensemble

Significance of IDPs

IDPs

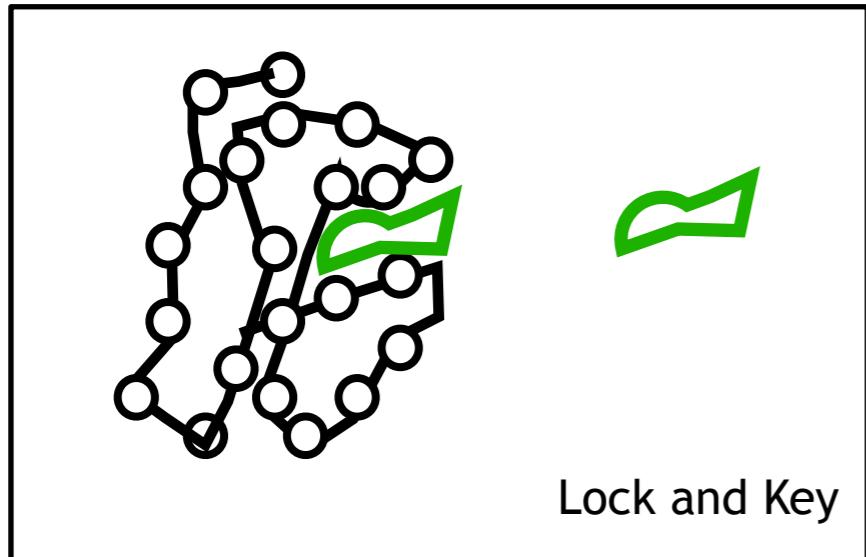
NMR (secondary)



- > 33 % of eukaryotic proteins have long (>30 residues) disordered regions.
- α -synuclein, p53, Tau are all IDPs
- Involved in critical biological functions including transcriptional activation and intracellular signaling
- > 25 % of missense disease mutations
- implicated in: cancer, cardiovascular disease, amyloidosis, neurodegenerative disease, diabetes, among others
- Potential drug targets

Potential functional mechanisms

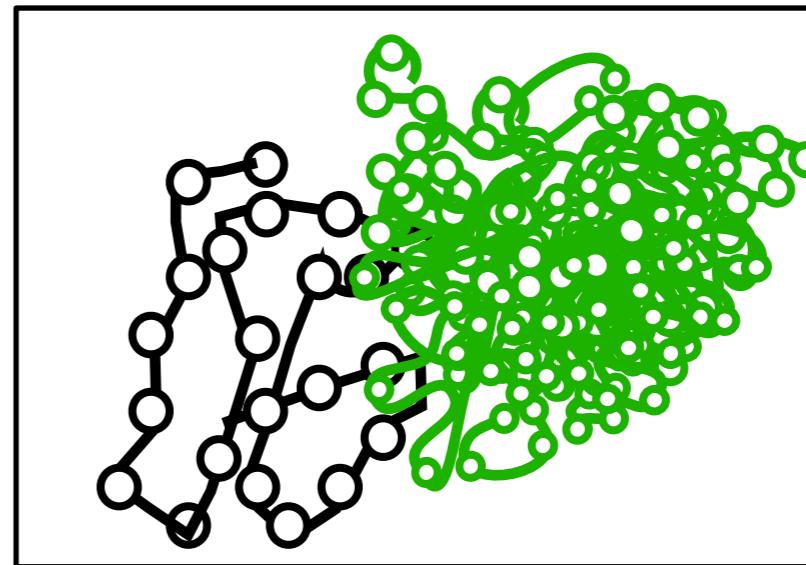
Structured



Lock and Key

(Fischer et al., 1894)

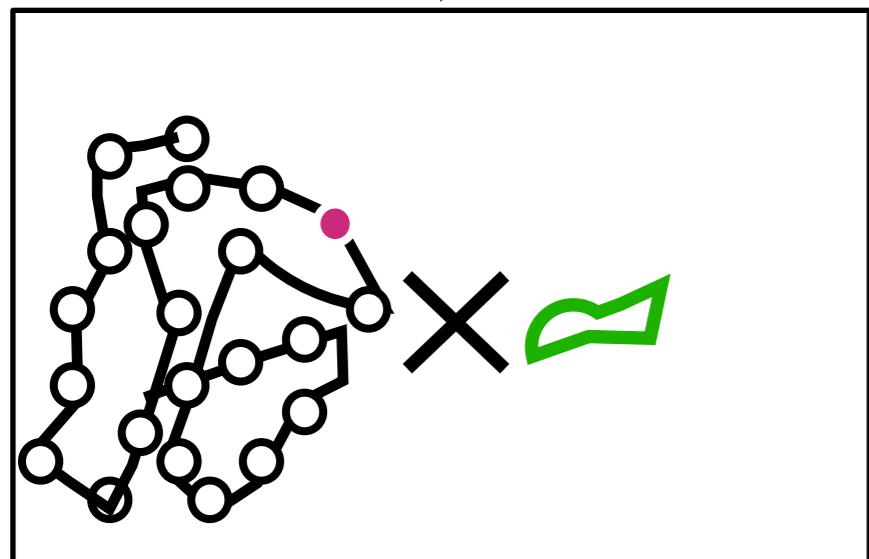
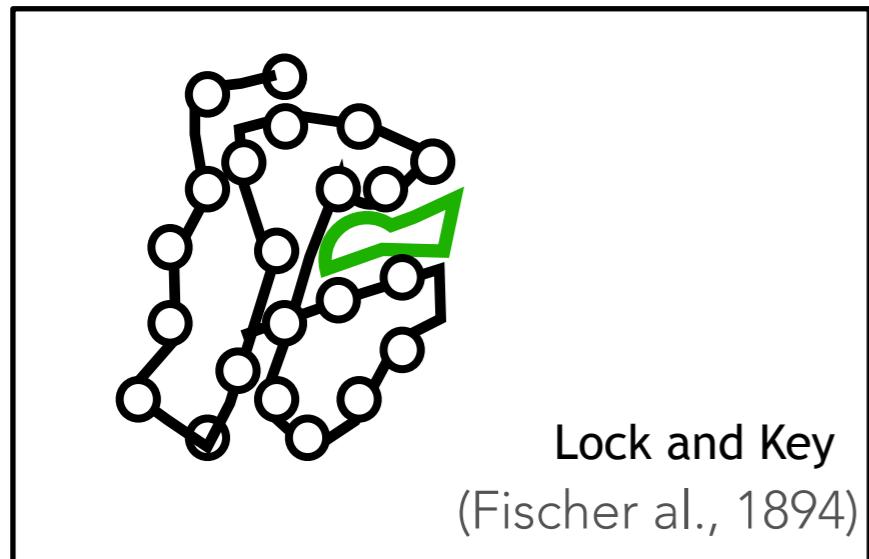
IDPs



- Unique partner
- High affinity
- sequence -> function
- Can bind multiple partners
- Possibility of high specificity/low affinity binding (binding+folding)
- sequence -> function

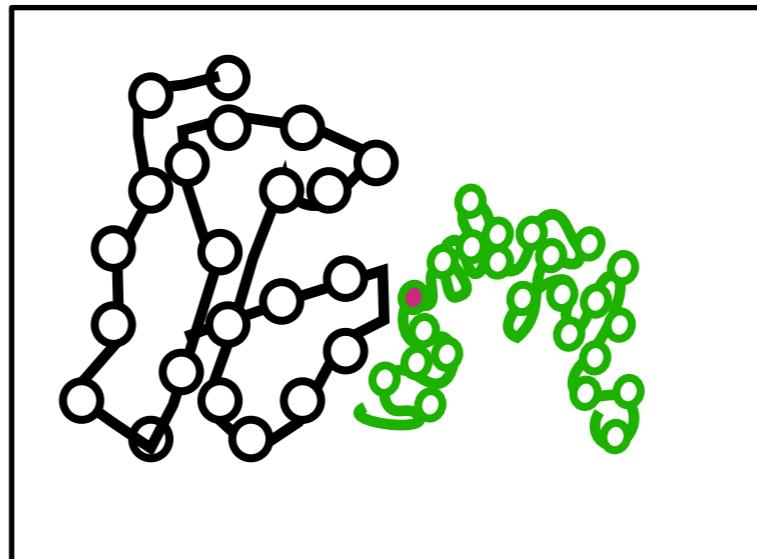
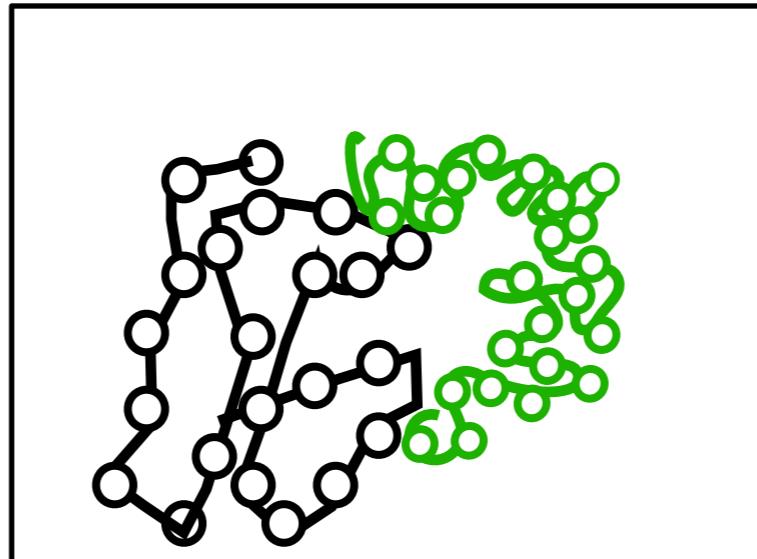
Explaining effects of disease causing mutations

Structured



sequence -> structure ->
function

IDPs

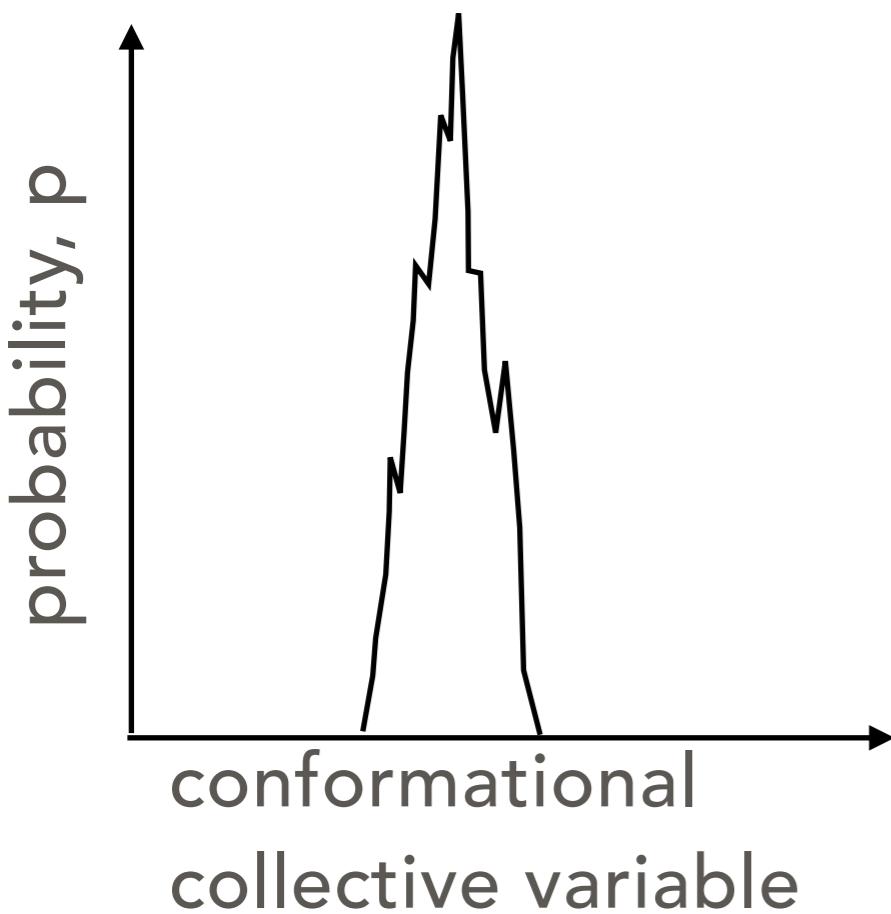


sequence -> ensemble -> function

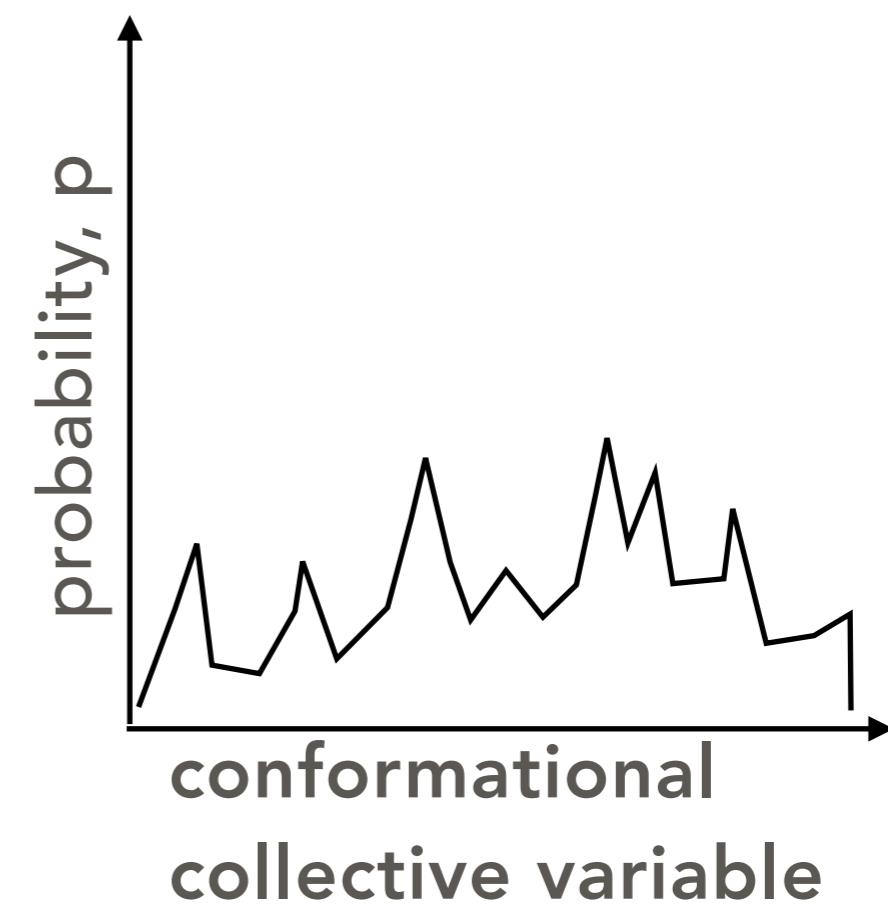
What is “ensemble” in sequence-ensemble ?

- Ensemble captures the probability distribution of conformations

Structured



IDPs

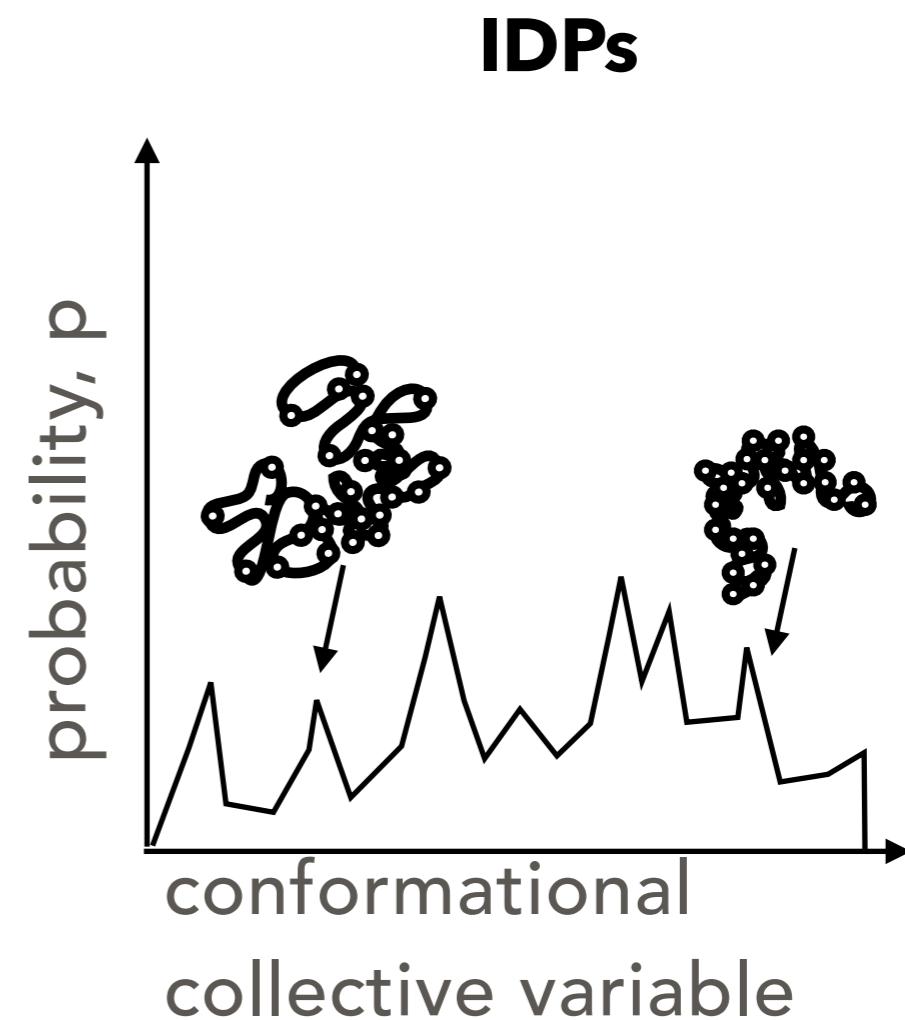
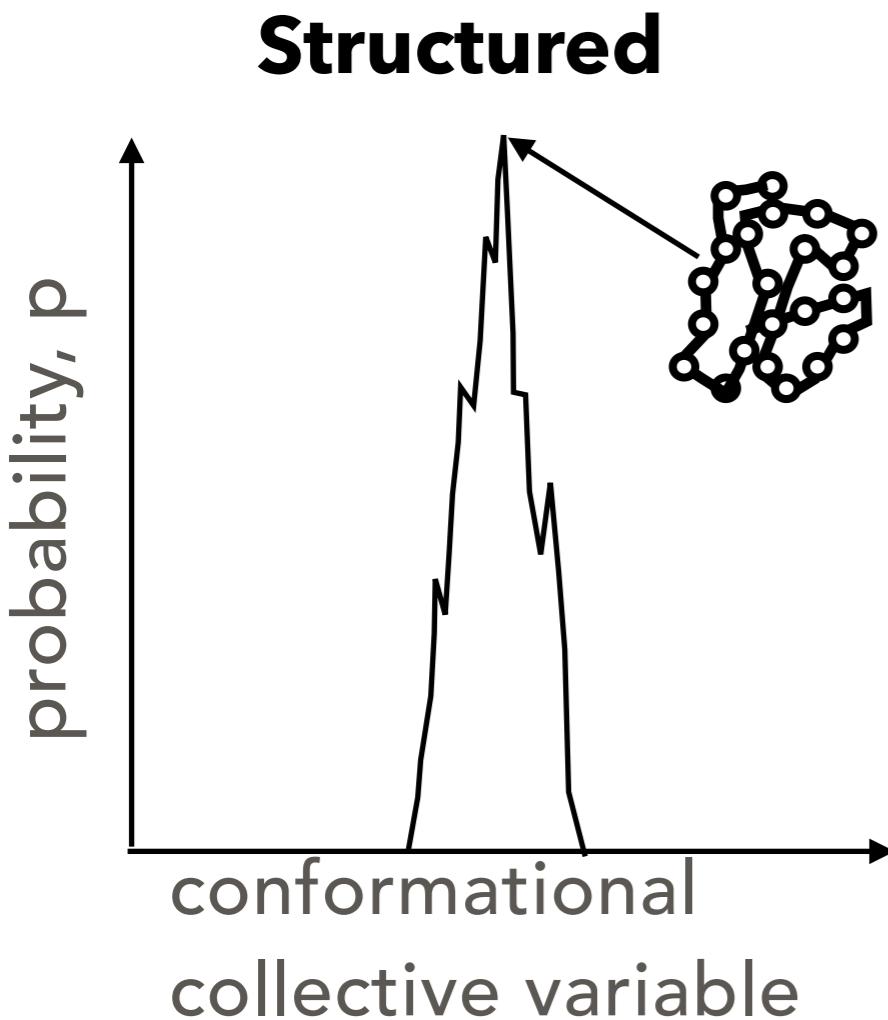


- Ensemble —> few probable conformations

- Ensemble —> large number of probable conformations
- Need to determine relative distribution of probabilities

Capturing conformational distributions

- Sample all the conformations and define their frequencies
- NMR, X-ray can capture the average ensemble properties

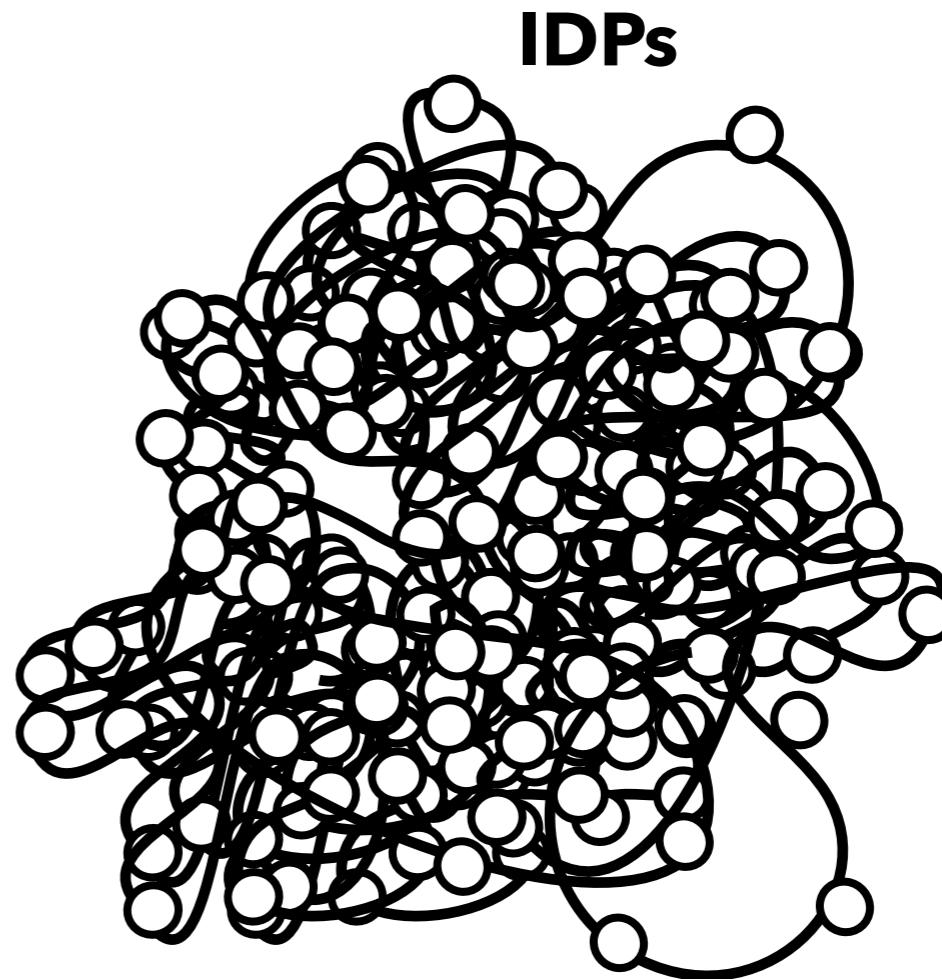
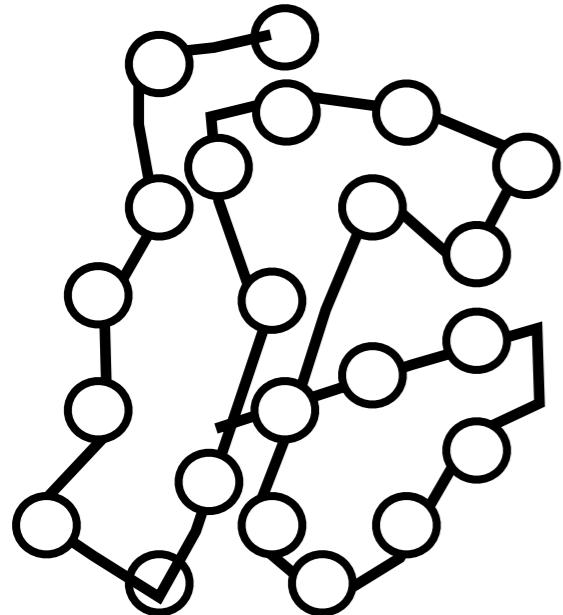


- MD can isolate individual conformations
- Relative frequencies depends on the conformational collective variable chosen

Challenges in defining conformational variable

- Relative frequencies depends on the conformational collective variable chosen

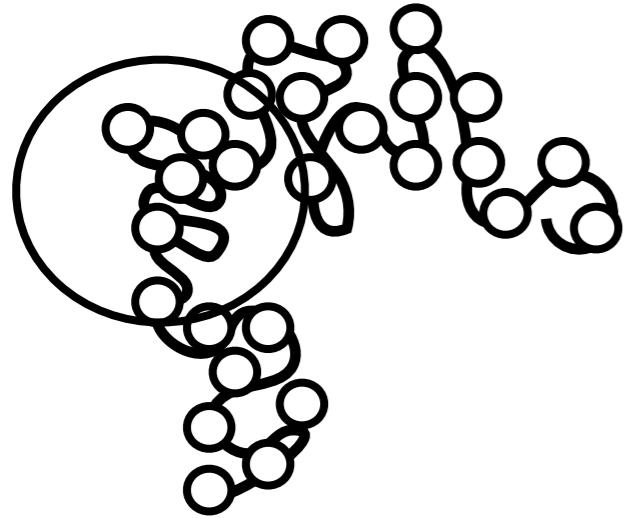
Structured



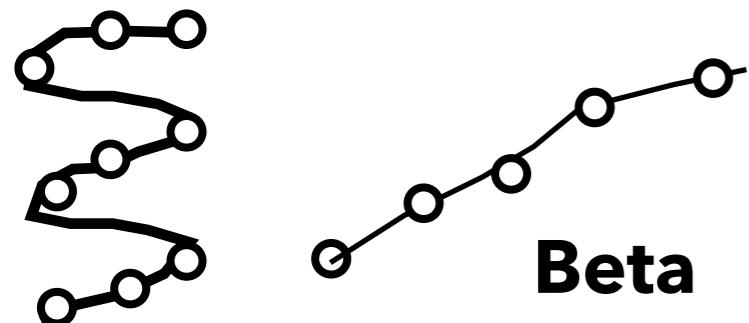
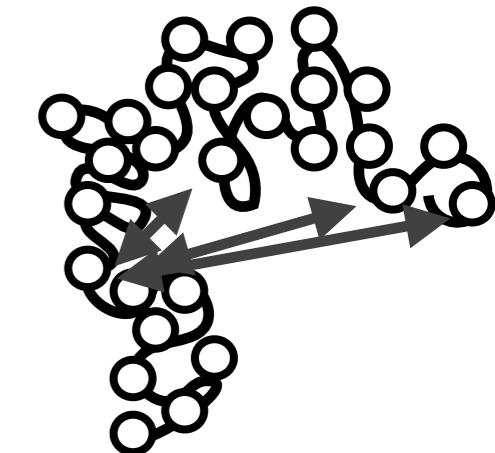
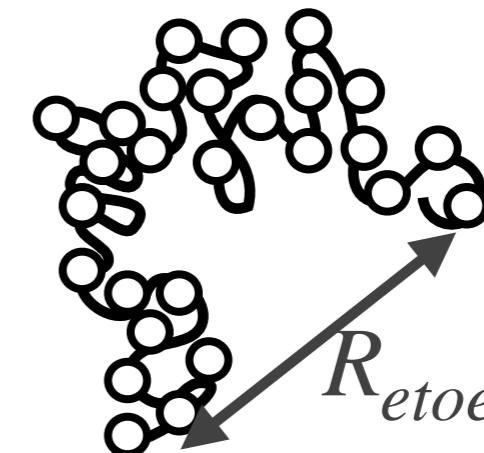
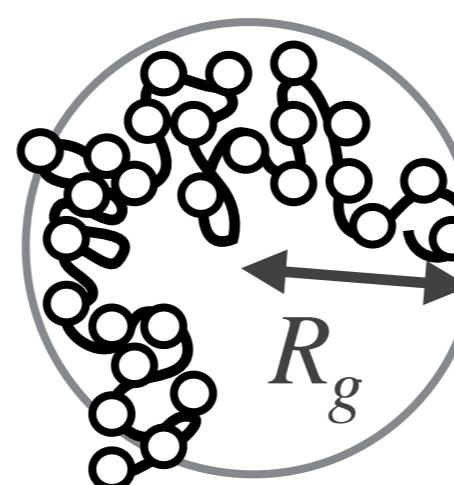
- x,y,z coordinates of each residue
- Distribution of conformations can be captured using RMSD
- x,y,z coordinate of each residue
- No longer meaningful or tractable

Challenges in defining IDPs conformational variable

Secondary structure



"Tertiary structure"

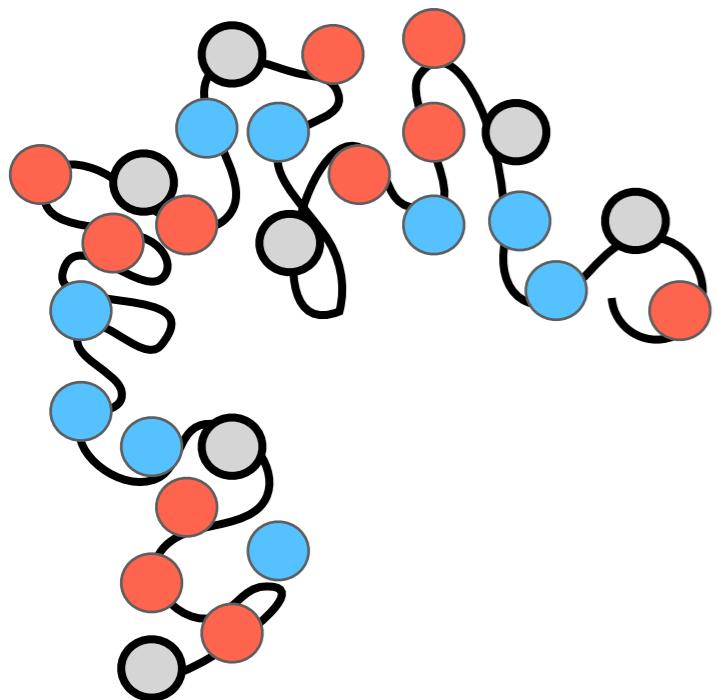


Helix

- NMR, MD

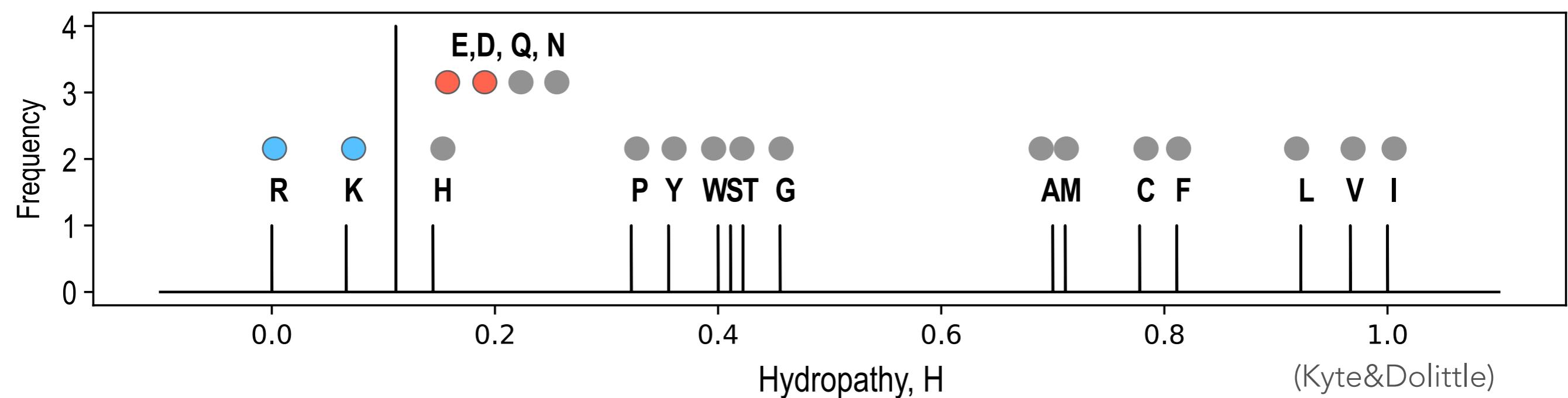
- Radius of gyration (R_g)
- End to end distance (R_{etoe})
- Tertiary interactions, where possible interactions: $O(N^2)$

Capturing conformational distribution from sequence properties



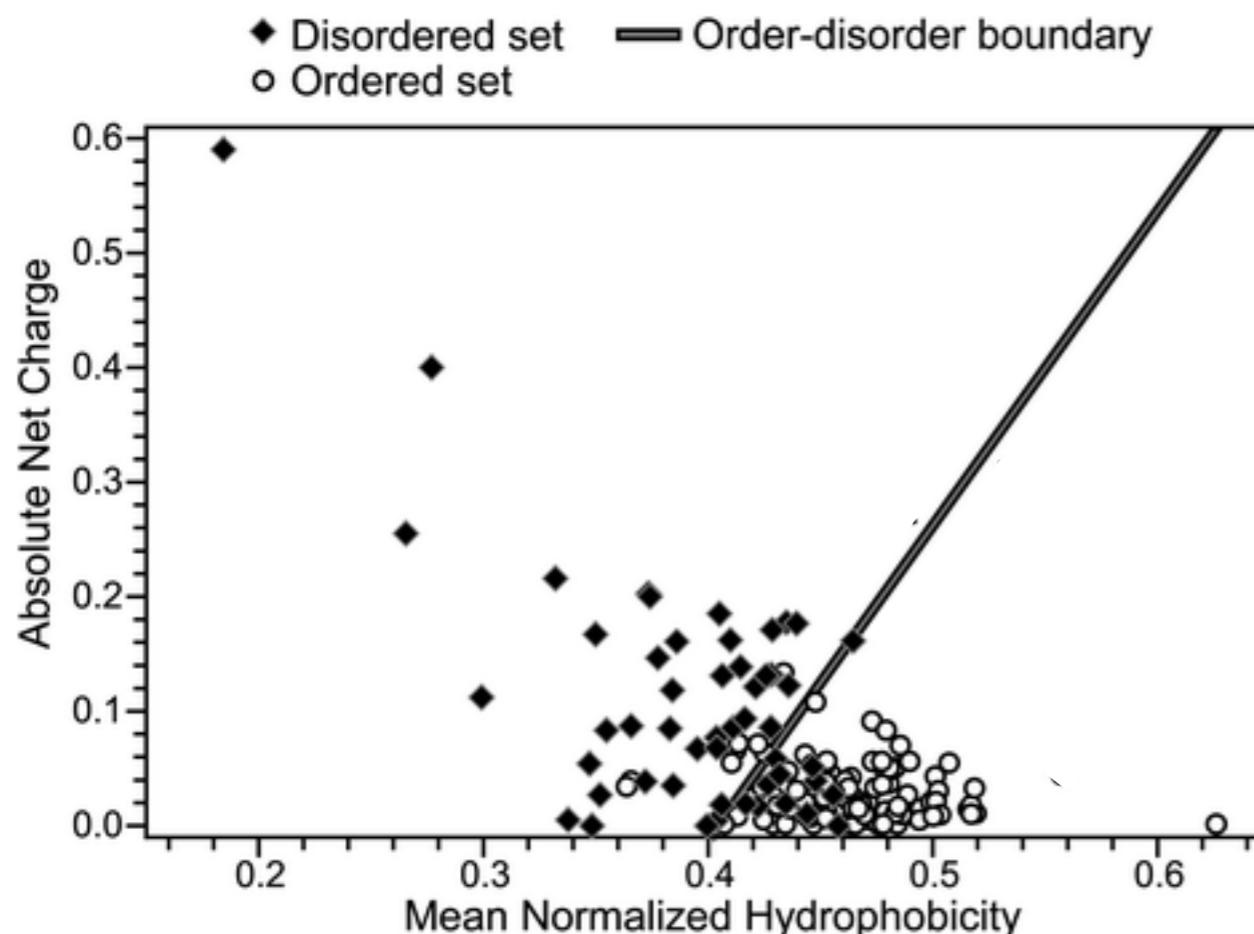
- Fraction of charged residues (f)
- Net charge per residue (NCPR)
- Hydrophobicity (H)

- Negatively charged residues
- Positively charged residues
- Charge neutral residues



Challenges in predicting order/disorder from IDPs sequence

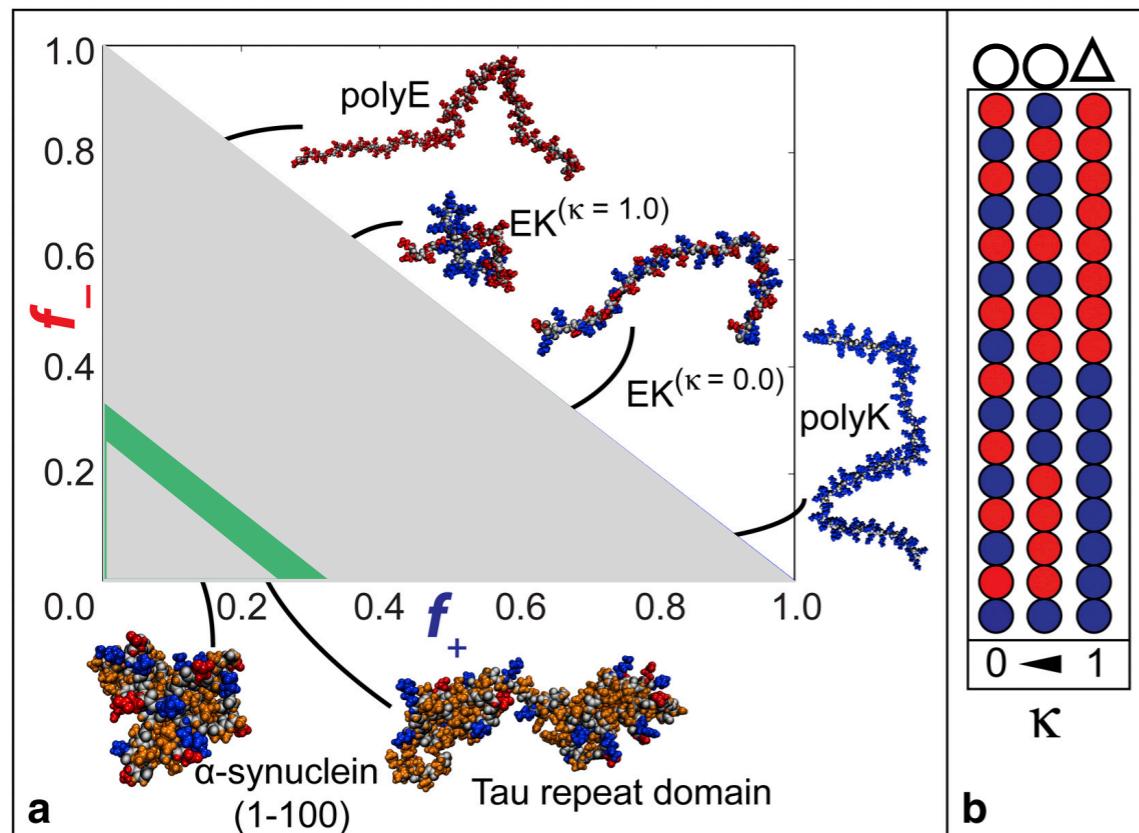
- Established theories consider ratio of sequence properties: Mean hydrophobicity (H) and mean net charge (NCPR)
- ~50% of sequences fall at the boundary.



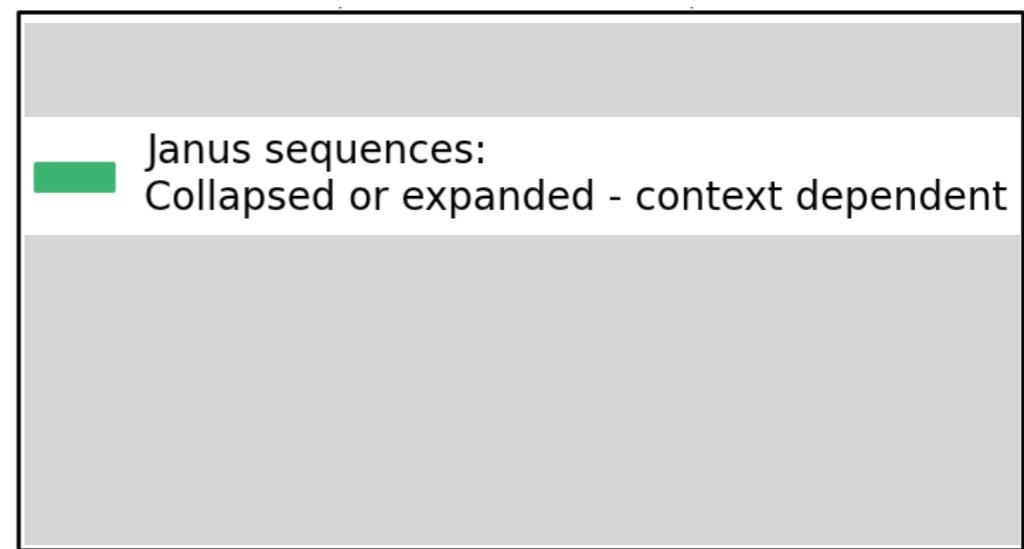
- Can predict if the conformational distribution will be more rugged or peaked

Challenges in predicting phase behavior from IDPs sequence

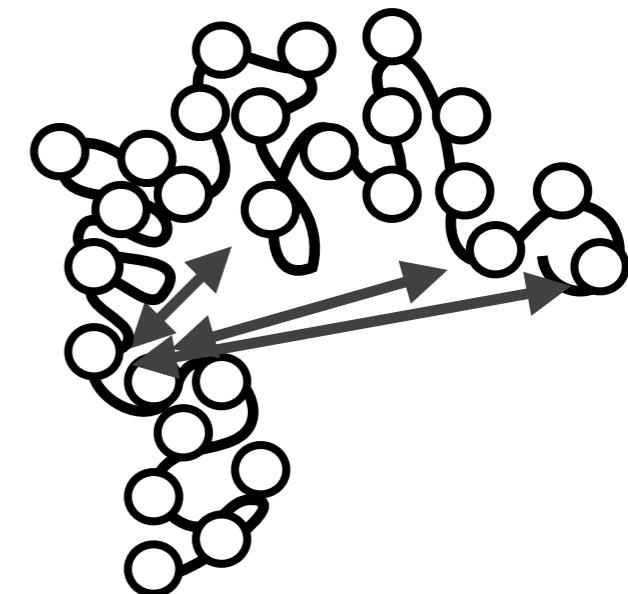
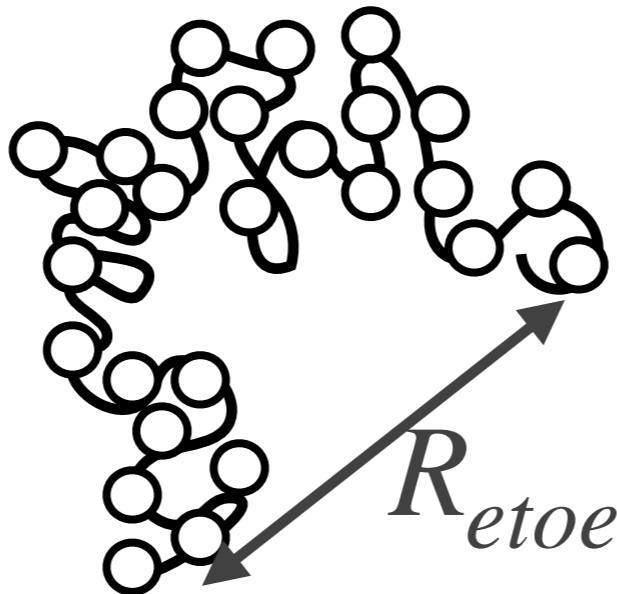
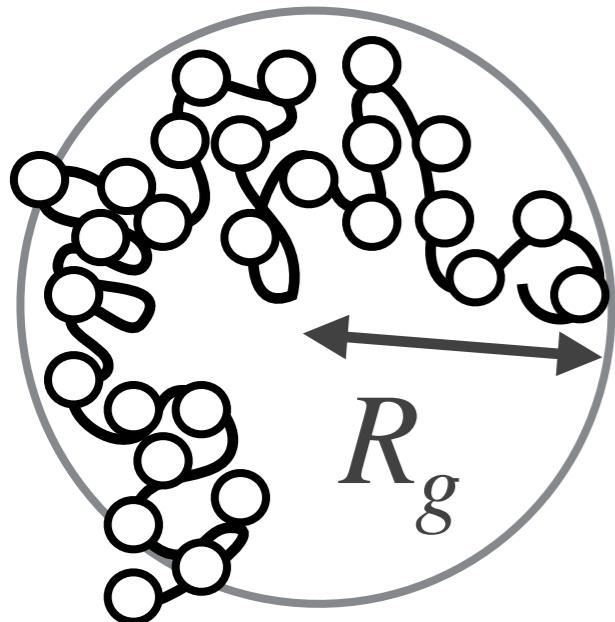
- ~40% of IDP's fall in the Janus region.
- Do not consider hydrophobicity
- Not very sensitive to single residue substitutions
- Can predict whether the distribution as a function of R_g will be far left (globular) or far right (extended)



f_+ fraction of positively charged residues
 f_- fraction of negatively charged residues
 κ sequence distribution of oppositely charged residues
● negatively charged ● positively charged $O\Delta$ uncharged residues

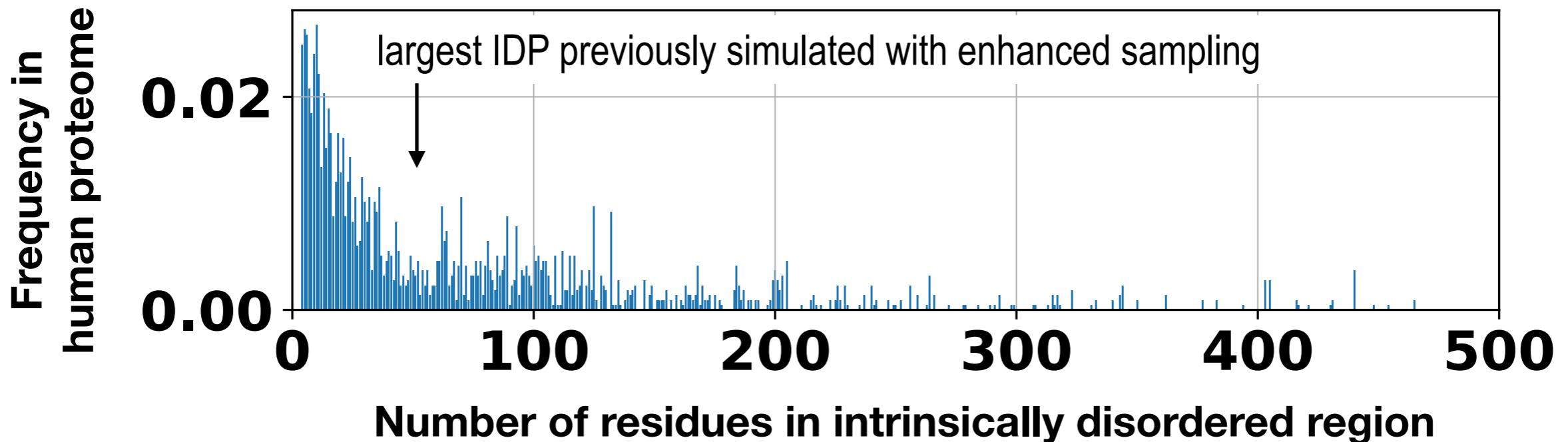


Challenges in defining IDPs “Tertiary structure” conformational variable



- Pappu and Uversky metric can predict some qualities of conformational distribution from sequence
- FRET, MD
- Possible interactions: $O(N^2)$
- No tractable way of defining tertiary contacts and their distribution in long IDPs
- MD

Challenges in simulating IDP's



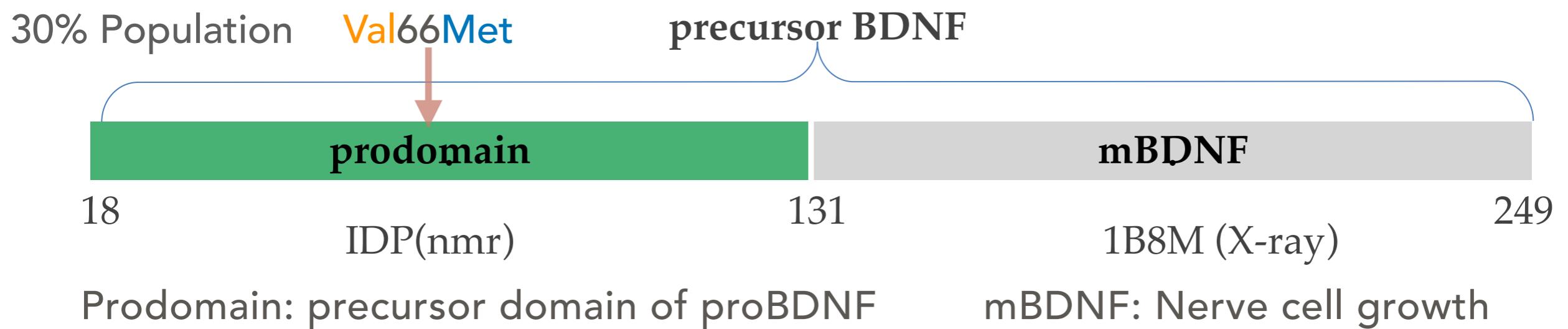
- Longest IDPs previously simulated with explicit solvent, replica-exchange accuracy are much shorter (~40) than average IDP length
- Number of contacts increases as N^2 but frequency of each contact also decreases!

General IDP Questions

1. How can we meaningfully detect tertiary interactions in a long disordered protein?
2. Can we predict the effect of amino-acid substitutions, including charge-neutral substitution on the tertiary interactions in a long disordered protein?

Functionally significant hydrophobic mutation : BDNF Val66Met

BDNF: Brain derived neurotrophic factor



Brief Communication

The Brain-Derived Neurotrophic Factor val66met Polymorphism and Variation in Human Cortical Morphology

Lukas Pezawas, Beth A. Verchinski, Venkata S. Mattay, Joseph H. Callicott, Bhaskar S. Kolachana, Richard E. Straub, Michael F. Egan, Andreas Meyer-Lindenberg, and Daniel R. Weinberger
Genes, Cognition, and Psychosis Program, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892-1379

Brain-Derived Neurotrophic Factor Val66Met and Psychiatric Disorders: Meta-Analysis of Case-Control Studies Confirm Association to Substance-Related Disorders, Eating Disorders, and Schizophrenia

Mònica Gratacòs, Juan R. González, Josep M. Mercader, Rafael de Cid, Mikel Urretavizcaya, and Xavier Estivill

Cell

CellPress

Volume 112, Issue 2, 24 January 2003, Pages 257-269

Article

The BDNF val66met Polymorphism Affects Activity-Dependent Secretion of BDNF and Human Memory and Hippocampal Function

Michael F. Egan¹, Masami Kojima^{7, 4, 5, 6}, Joseph H. Callicott^{7, 1}, Terry E. Goldberg^{7, 1}, Bhaskar S. Kolachana¹, Alessandro Bertolino¹, Eugene Zaitsev⁴, Bert Gold³, David Goldman², Michael Dean³, Bai Lu^{*, 4}, Daniel R. Weinberger^{*, 1}  

Molecular Psychiatry (2010) 15, 260–271
© 2010 Nature Publishing Group All rights reserved 1359-4184/10 \$32.00
www.nature.com/mp

ORIGINAL ARTICLE

Meta-analysis of the BDNF Val66Met polymorphism in major depressive disorder: effects of gender and ethnicity

M Verhagen¹, A van der Meij¹, PAM van Deurzen^{1,2}, JGE Janzing¹, A Arias-Vásquez^{1,3,4}, JK Buitelaar^{1,2} and B Franke^{1,3}

A Genetic Variant BDNF Polymorphism Alters Extinction Learning in Both Mouse and Human

Fatima Soliman,^{1,2*} Charles E. Glatt,² Kevin G. Bath,² Liat Levita,^{1,2} Rebecca M. Jones,^{1,2} Siobhan S. Pattwell,² Deqiang Jing,² Nim Tottenham,^{1,2} Dima Amso,^{1,2} Leah H. Somerville,^{1,2} Henning U. Voss,³ Gary Glover,⁴ Douglas J. Ballon,³ Conor Liston,^{1,2} Theresa Teslovich,^{1,2} Tracey Van Kempen,^{1,2} Francis S. Lee,^{2,*} B. J. Casey^{1,2*}

REPORTS

Genetic Variant BDNF (Val66Met) Polymorphism Alters Anxiety-Related Behavior

Zhe-Yu Chen,^{1,4†} Deqiang Jing,^{1*} Kevin G. Bath,^{1,*} Alessandro Ieraci,¹ Tanvir Khan,¹ Chia-Jen Siao,² Daniel G. Herrera,¹ Miklos Toth,³ Chingwen Yang,⁵ Bruce S. McEwen,⁶ Barbara L. Hempstead,² Francis S. Lee^{1,3†}

Molecular Psychiatry (2005) 10, 631–636
© 2005 Nature Publishing Group All rights reserved 1359-4184/05 \$30.00
www.nature.com/mp



ORIGINAL RESEARCH ARTICLE

Brain-derived neurotrophic factor val66met polymorphism and volume of the hippocampal formation

PR Szeszko^{1,2}, R Lipsky³, C Mentschel⁴, D Robinson^{1,2}, H Gunduz-Bruce⁵, S Sevy^{1,2}, M Ashtari⁶, B Napolitano¹, RM Bilder⁷, JM Kane^{1,2}, D Goldman³ and AK Malhotra^{1,2}

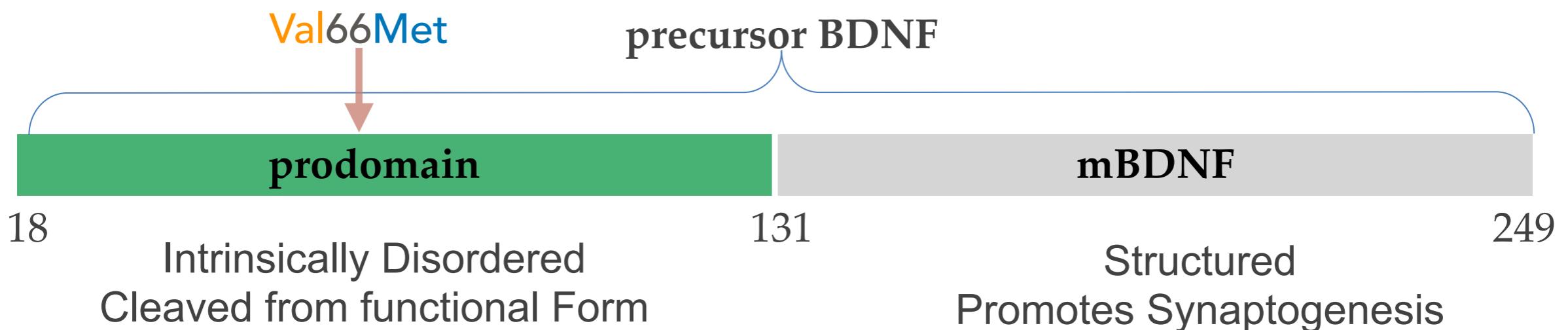
Results: 2,258

(from Web of Science Core Collection)

You searched for: TOPIC: (Val66Met)
...More

Functionally significant hydrophobic mutation : BDNF Val66Met

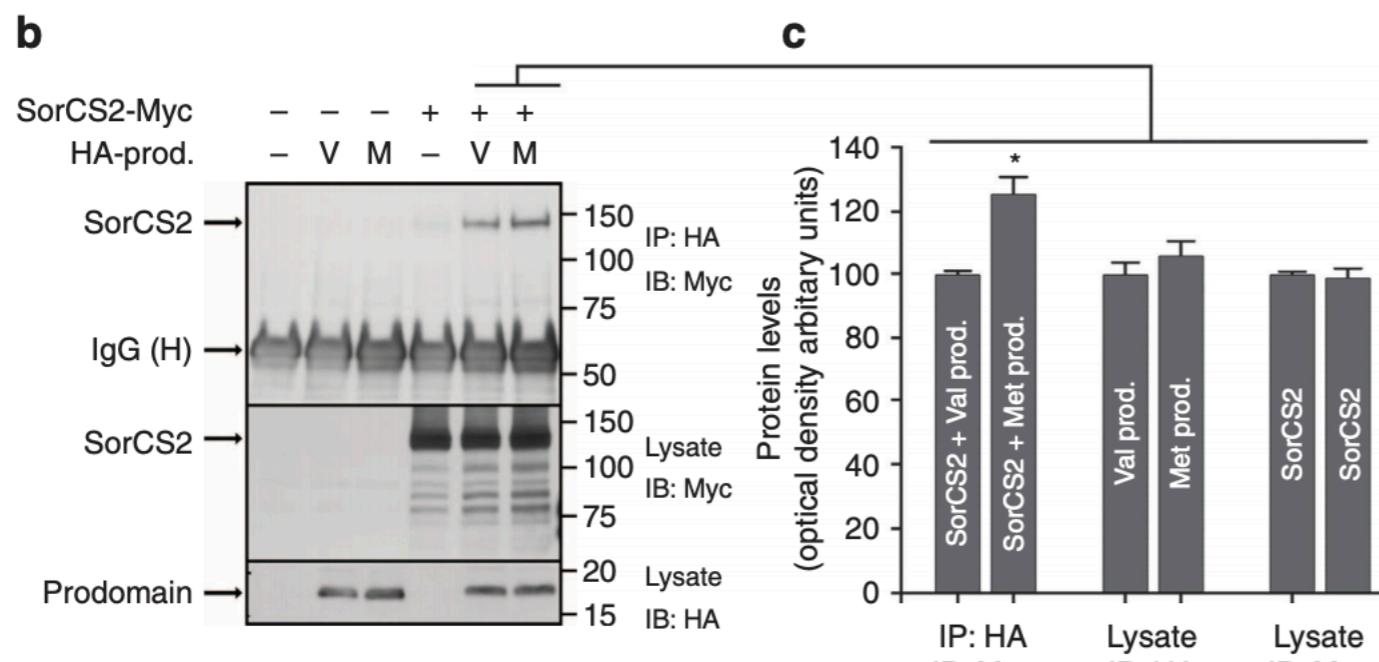
Residue 66 is not part of mature BDNF



Why is the non-functional form having physiological effects?

possible explanation

Met66 prodomain has a **new receptor** (SorCS2)



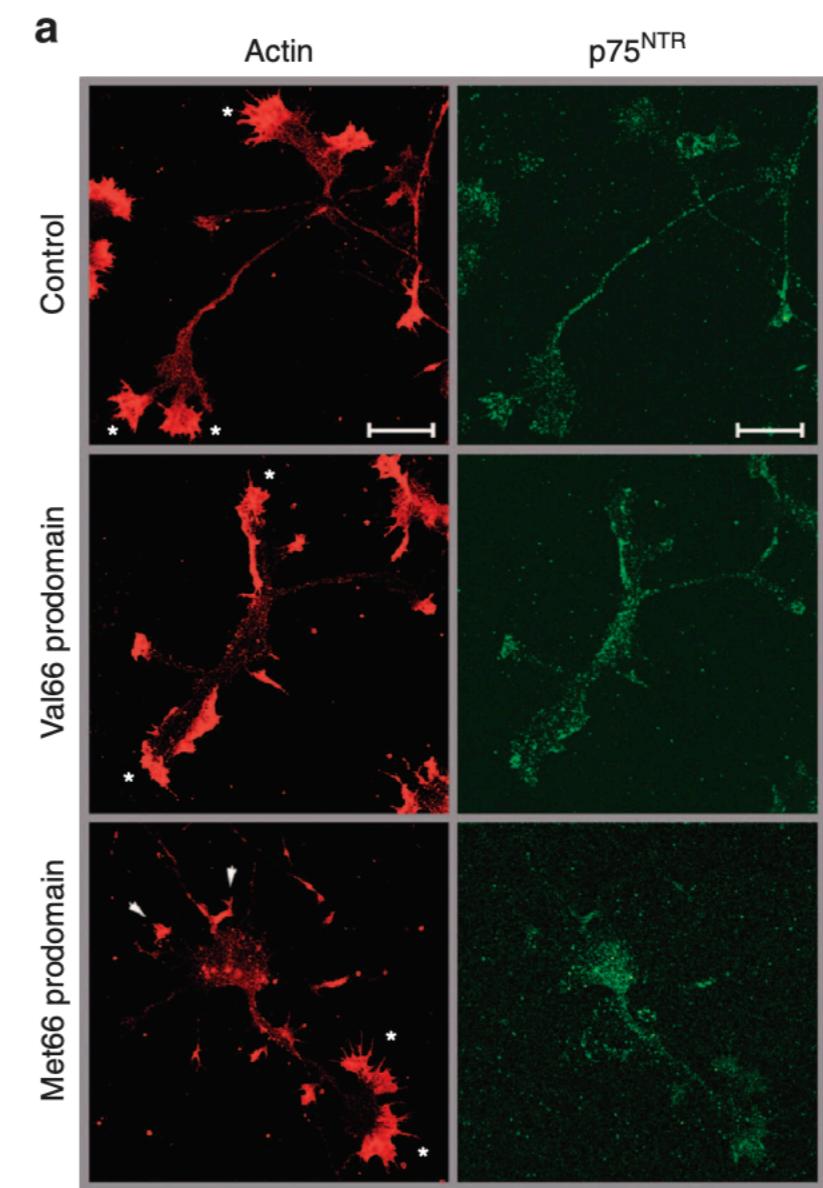
ARTICLE

Received 18 Apr 2013 | Accepted 21 Aug 2013 | Published 18 Sep 2013

DOI: 10.1038/ncomms3490

Val66Met polymorphism of BDNF alters prodomain structure to induce neuronal growth cone retraction

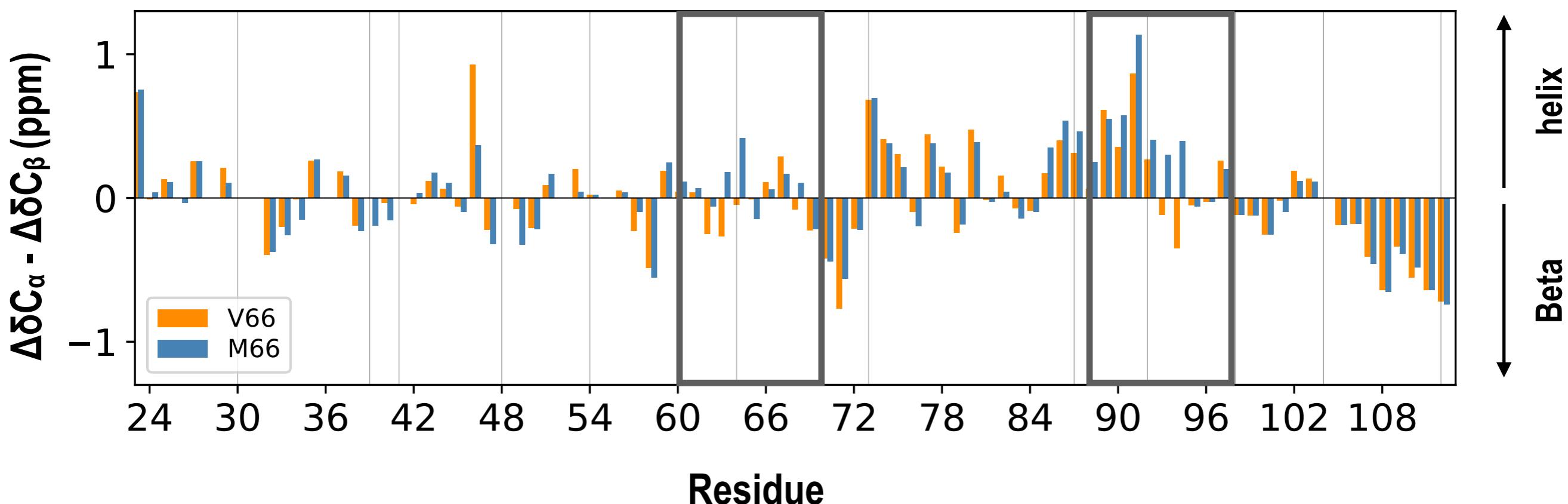
Agustin Anastasia¹, Katrin Deinhardt^{2,3}, Moses V. Chao², Nathan E. Will¹, Krithi Irmady¹, Francis S. Lee⁴, Barbara L. Hempstead¹ & Clay Bracken⁵



neuronal growth cones have retracted

Val66Met causes change in non-local secondary structure

Same study detects effects of Val66Met on residual secondary structure

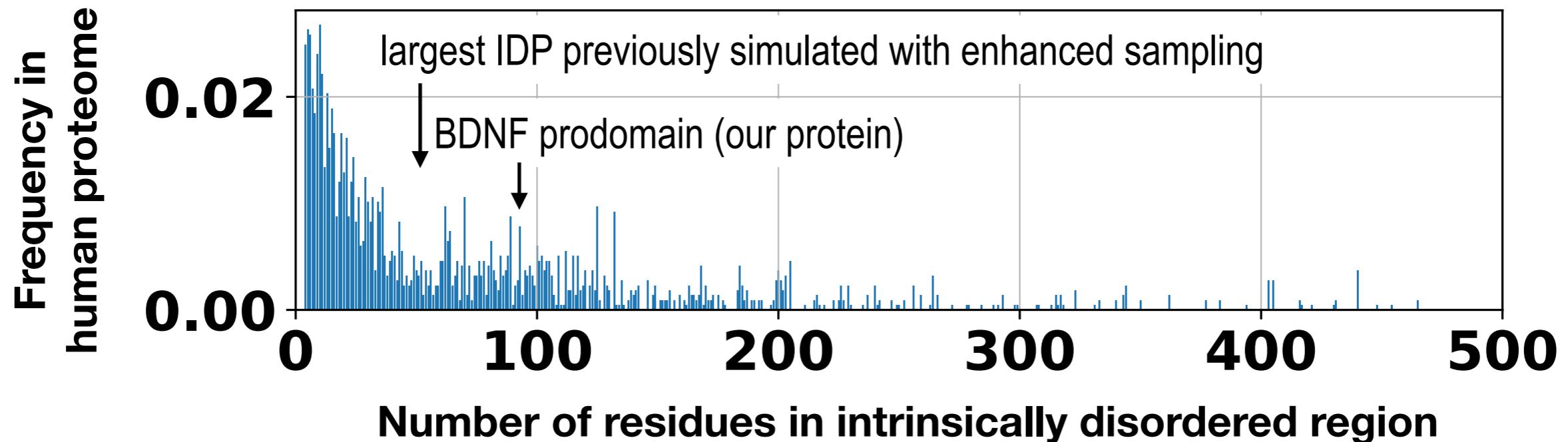


(Anastasia et al., Nat Commun., 2013)

Specific BDNF Questions

1. Why does Val66Met cause change in non-local secondary structure?
2. What is unique about methionine in BDNF prodomain at residue 66?

BDNF prodomain is long

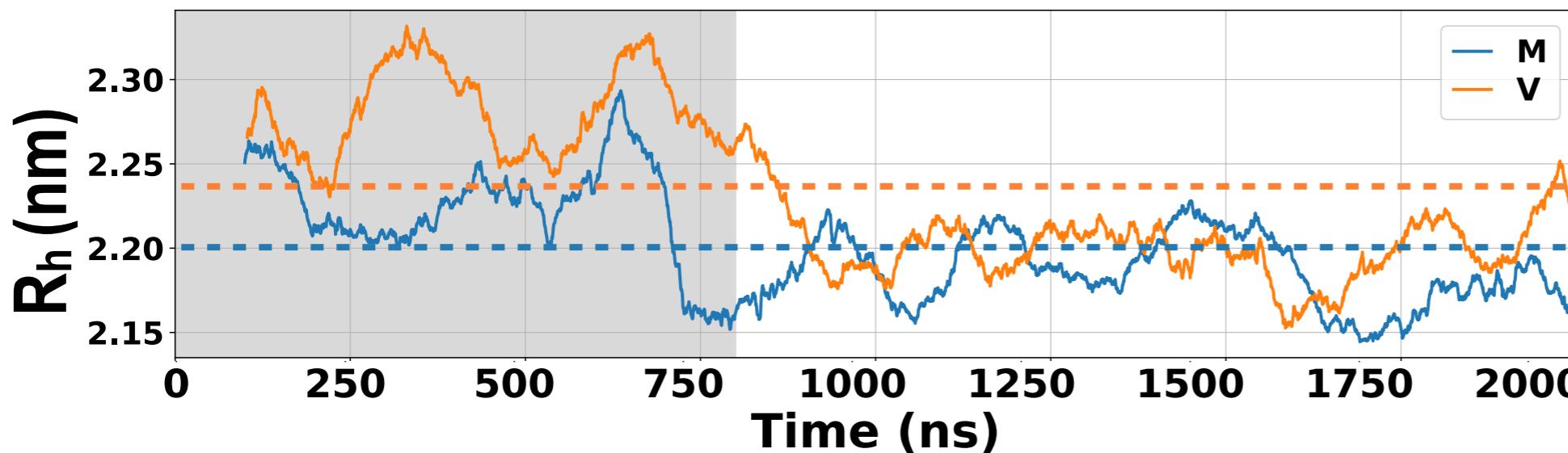


Caliburn (Rutgers Discovery Informatics Institute) (45 million SUs)
Rutgers Office of Advanced Research Computing (OARC)

Method - MD simulations

- 256 μ s (2 μ s x 64replicas) explicit solvent temperature replica exchange simulation (300K - 385K) of each prodomain.
- Amber99SB*-ILDN-q with TIP4PD water model was used.

MD is in agreement with NMR diffusion Rh



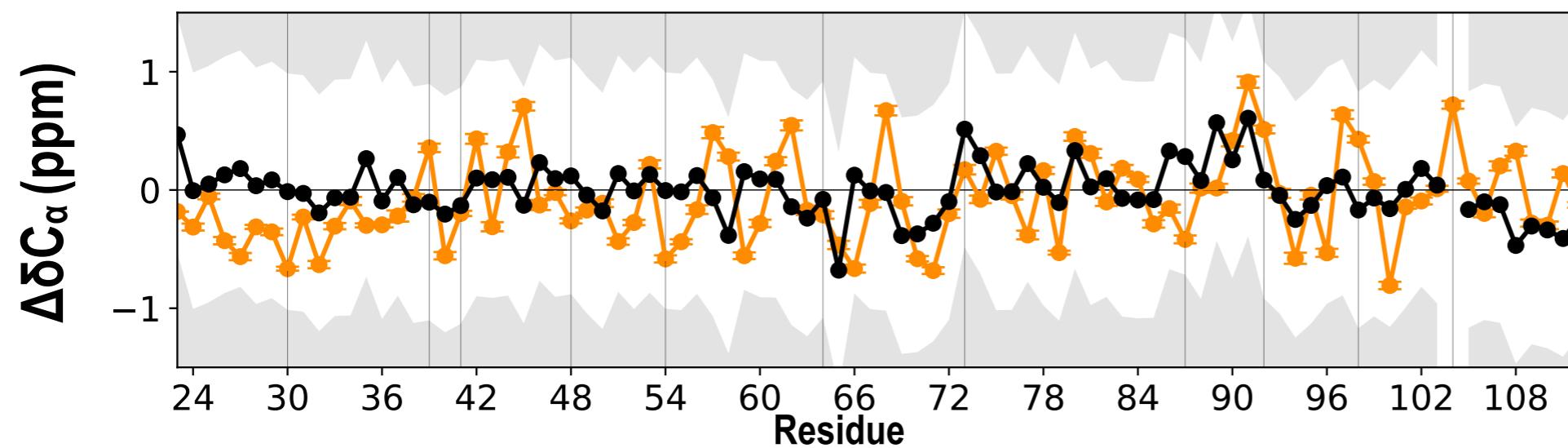
.8 μ s discarded
for equilibration

2.24nm

2.20nm

Rh from experiments
(Anastasia et al., Nat
Commun., 2013)

MD is in agreement with NMR CS

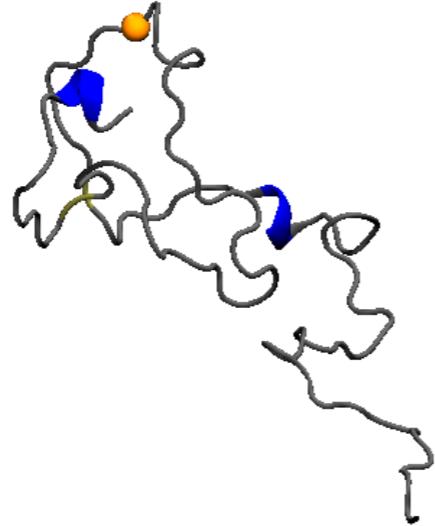


RMSD ~ .5ppm

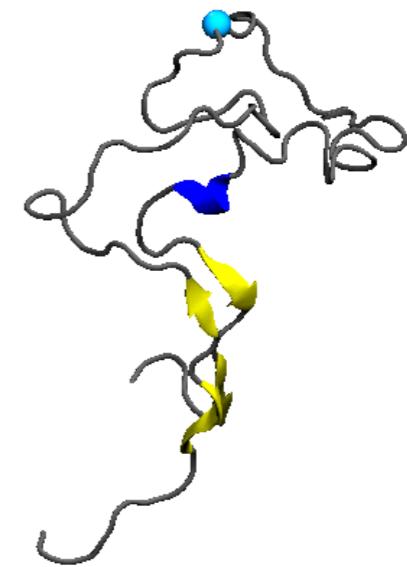
proBDNF trajectories

residues 23-113,
300K replica,
explicit solvent
not shown

V66



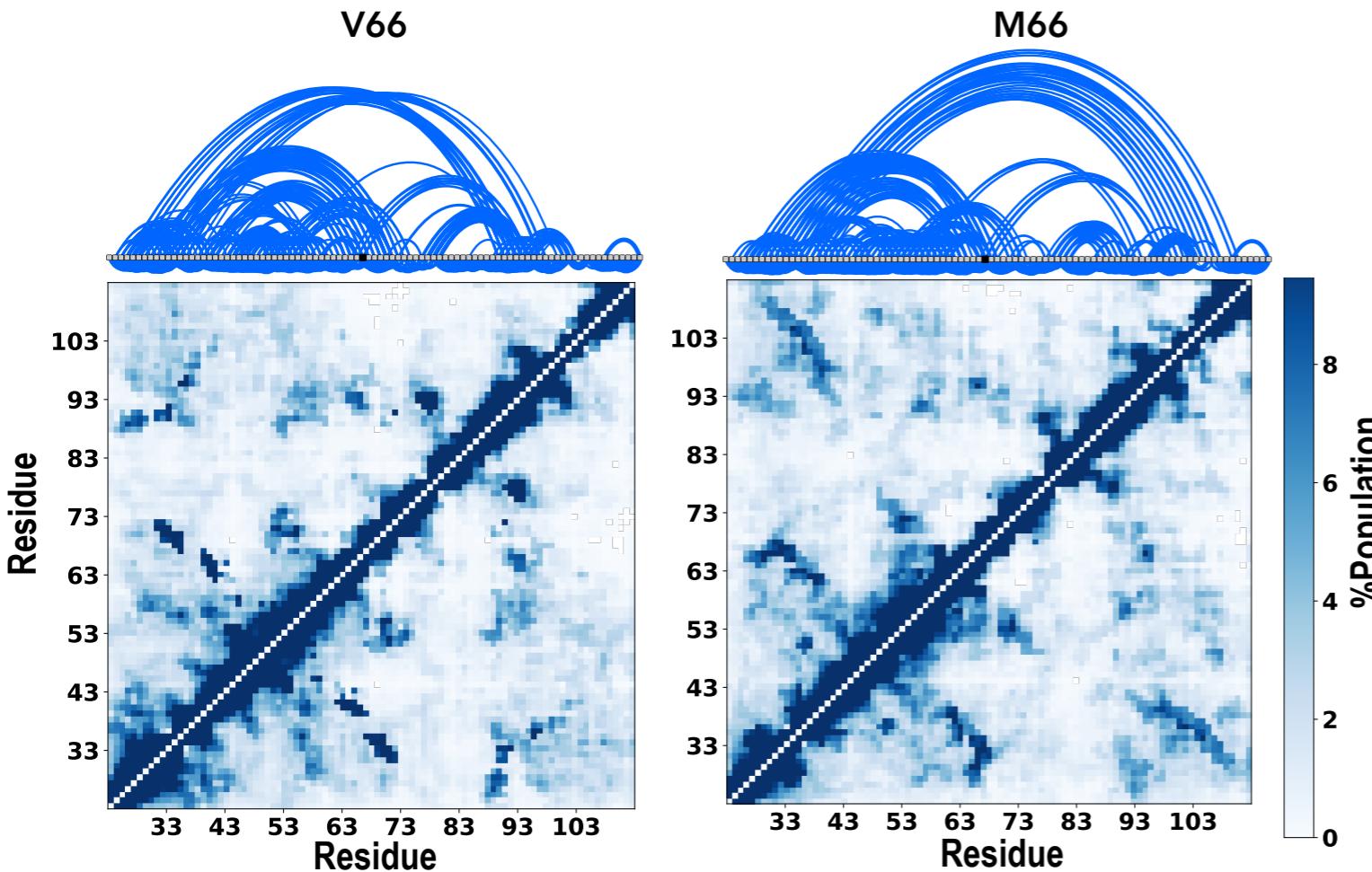
M66



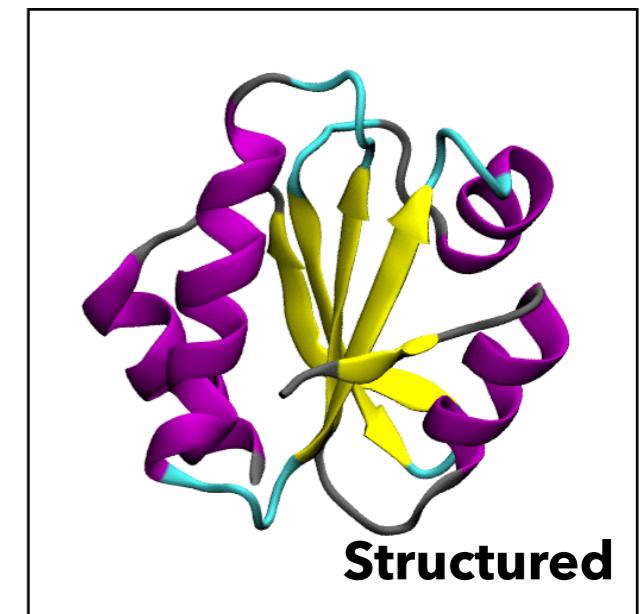
How to find specific interactions among so many non-specific interactions?

How can we possibly analyze this?

Step 1: try analyzing changes in contacts:

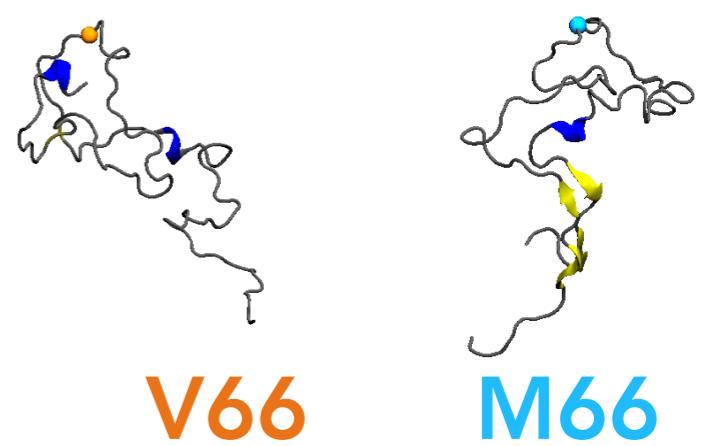


Structured : $O(N)$ Possible contacts



PDBID: 2N5A yeast Thioredoxin

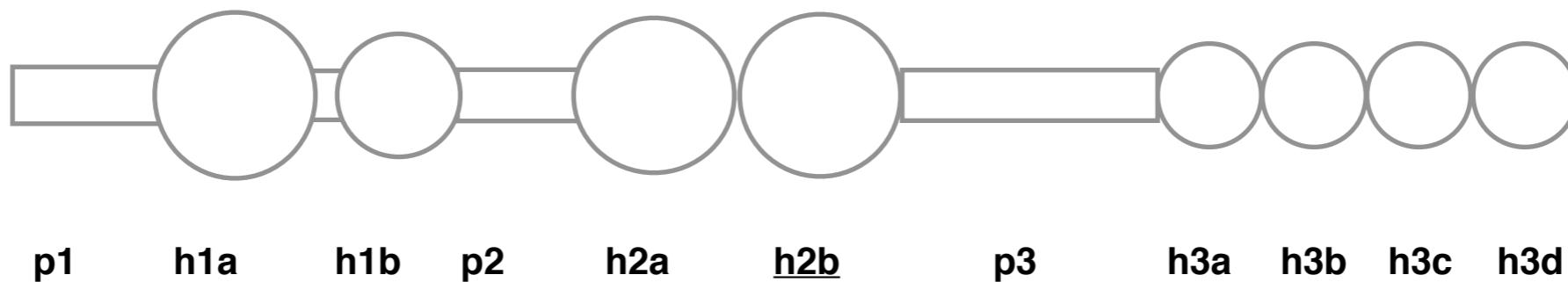
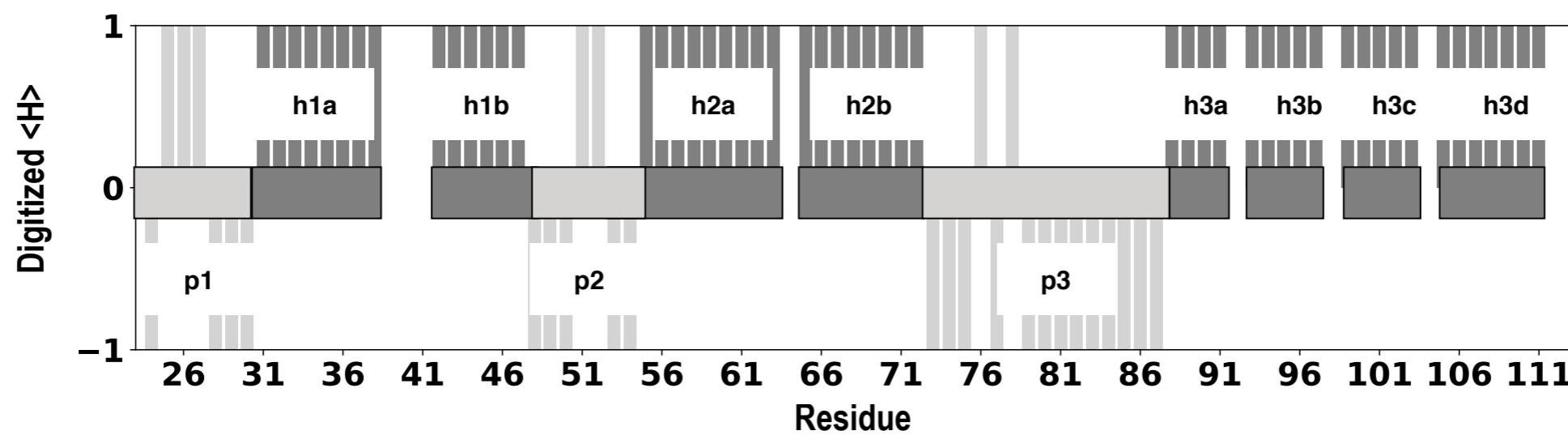
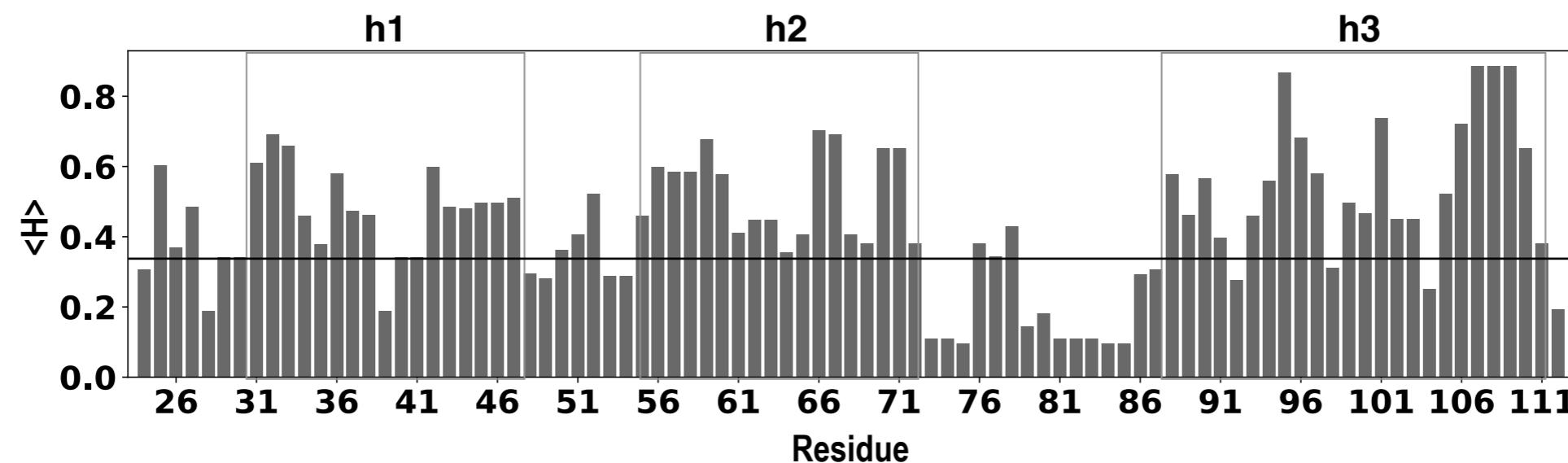
IDP: $O(N^2)$ Possible contacts



This is statistically challenging

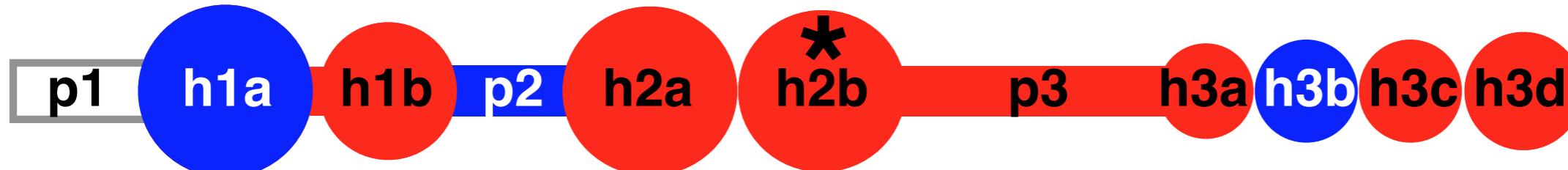
Solution : coarse-grain analysis
of atomistic simulation

Sequence-based blob identification

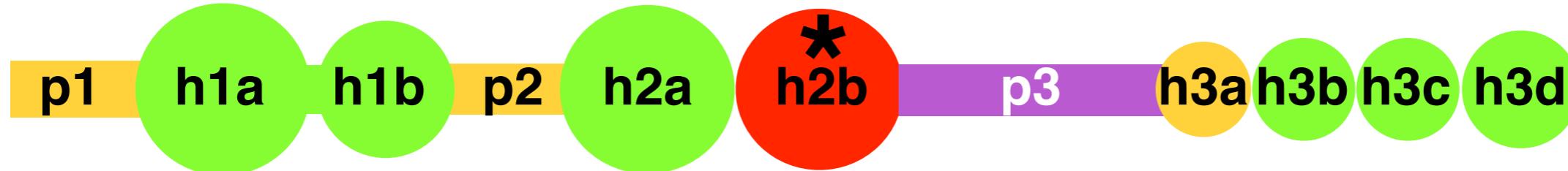
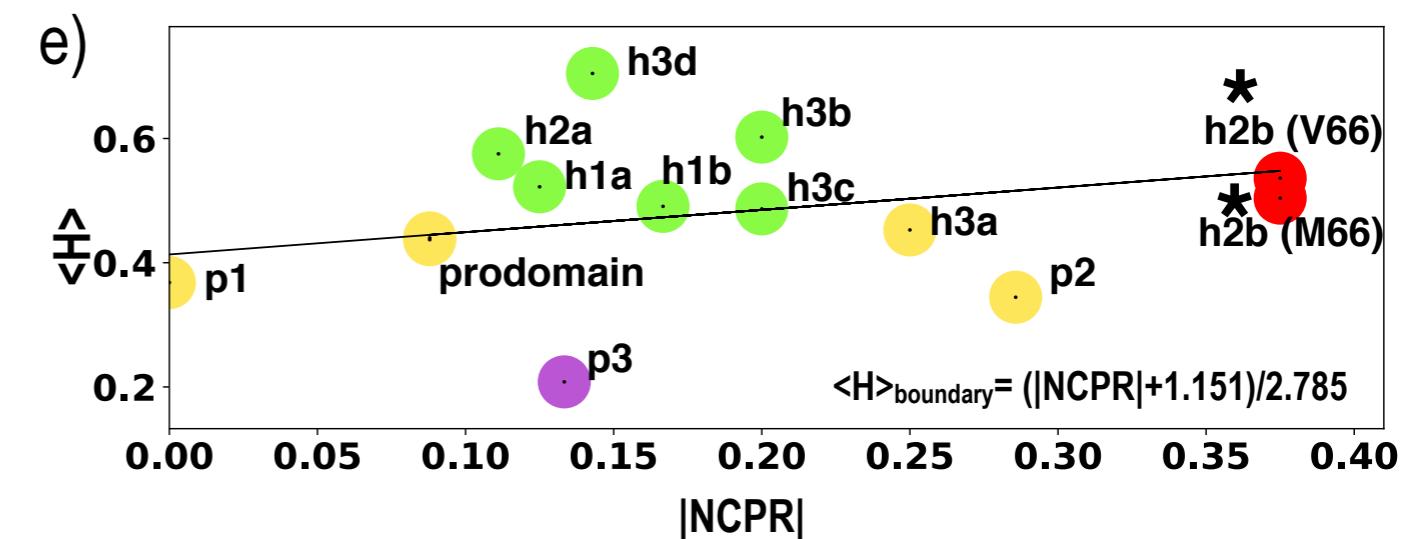
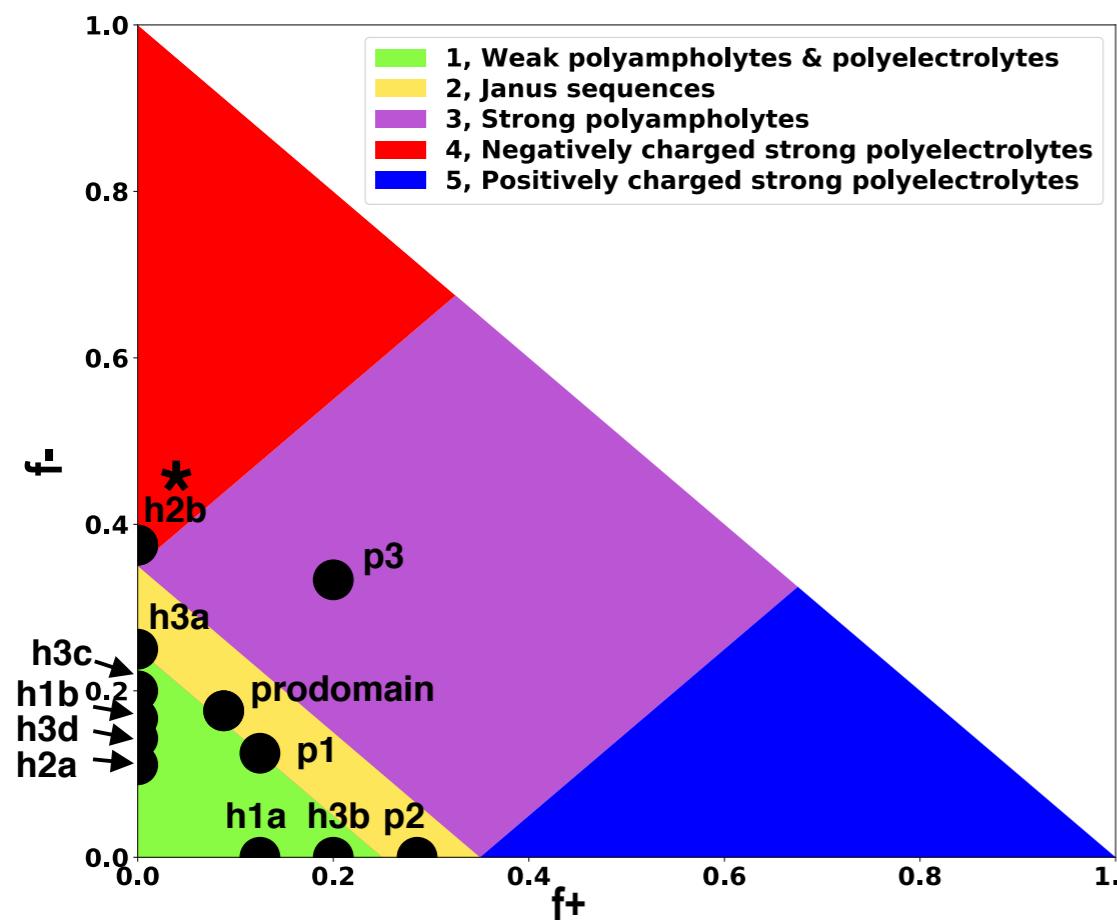


Fun new ways to conceptualize the protein

By net charge:

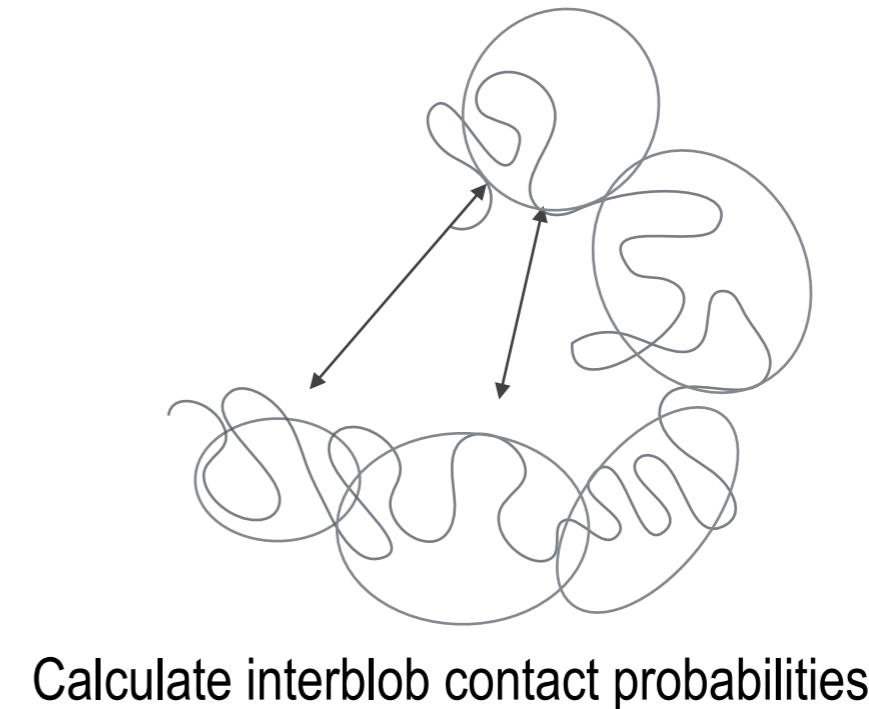
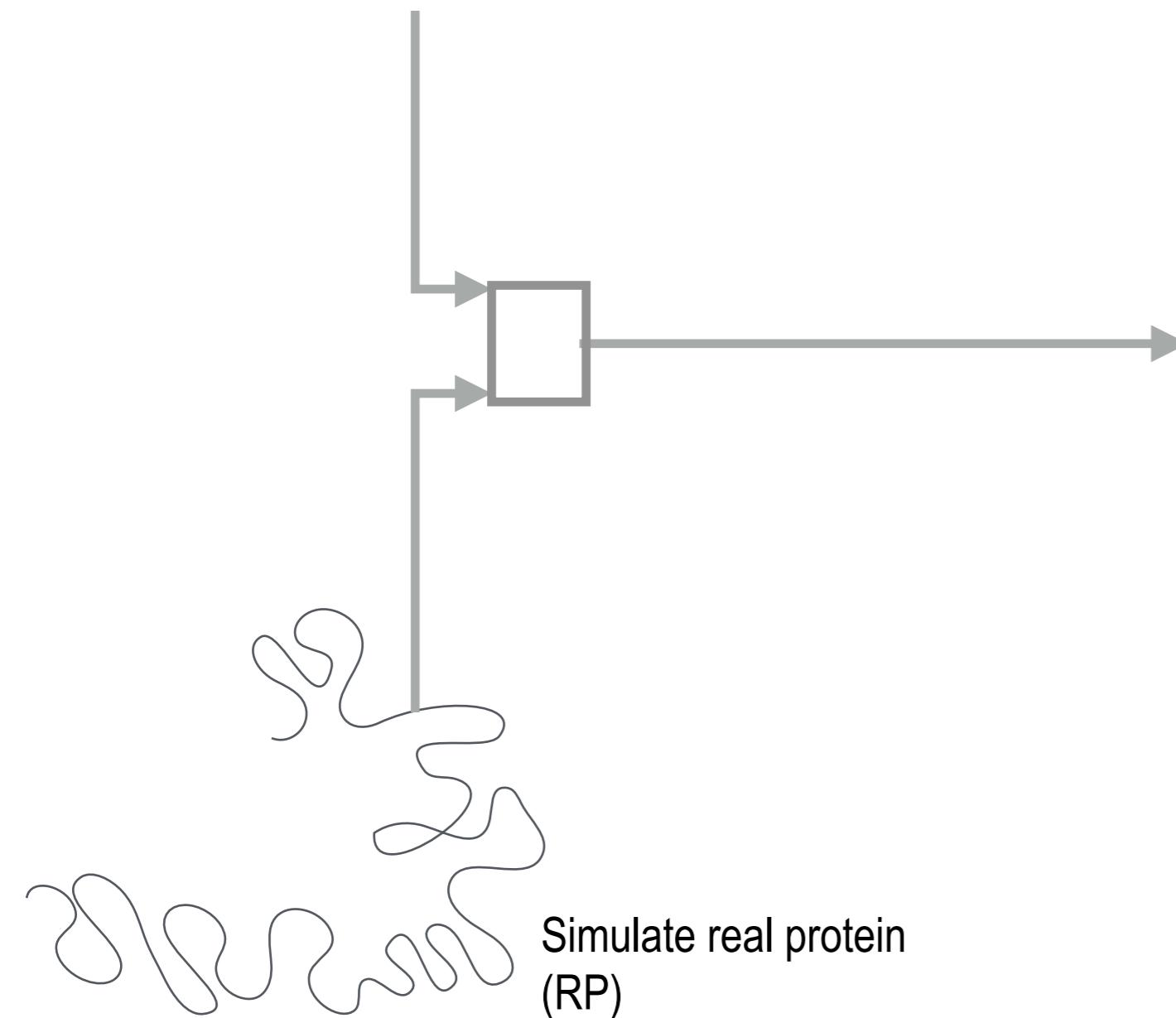
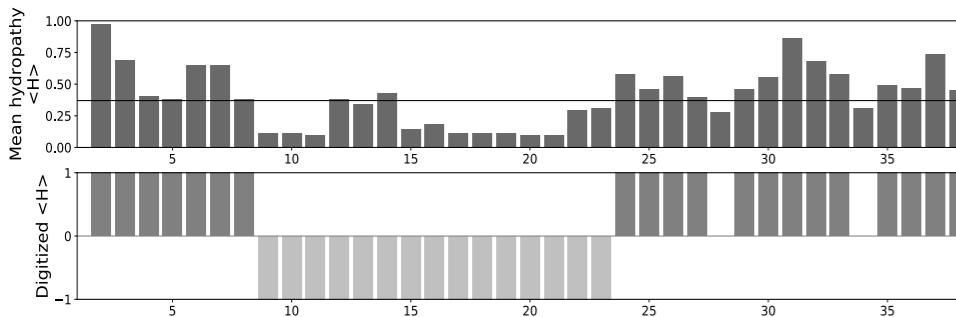


By predicted conformation:

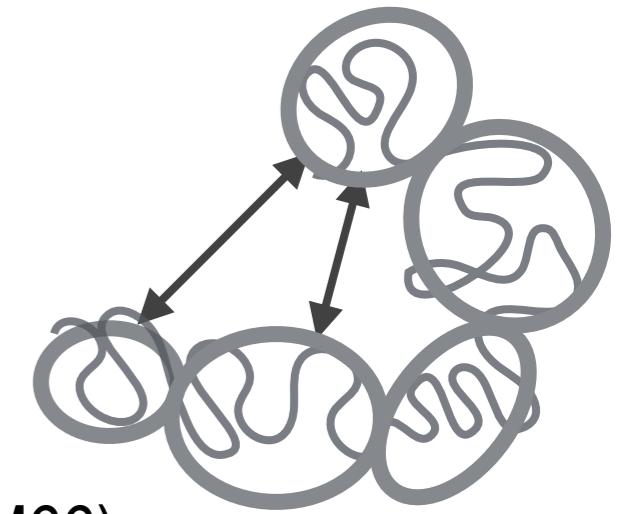


Hierarchical framework for identifying blob level tertiary contacts

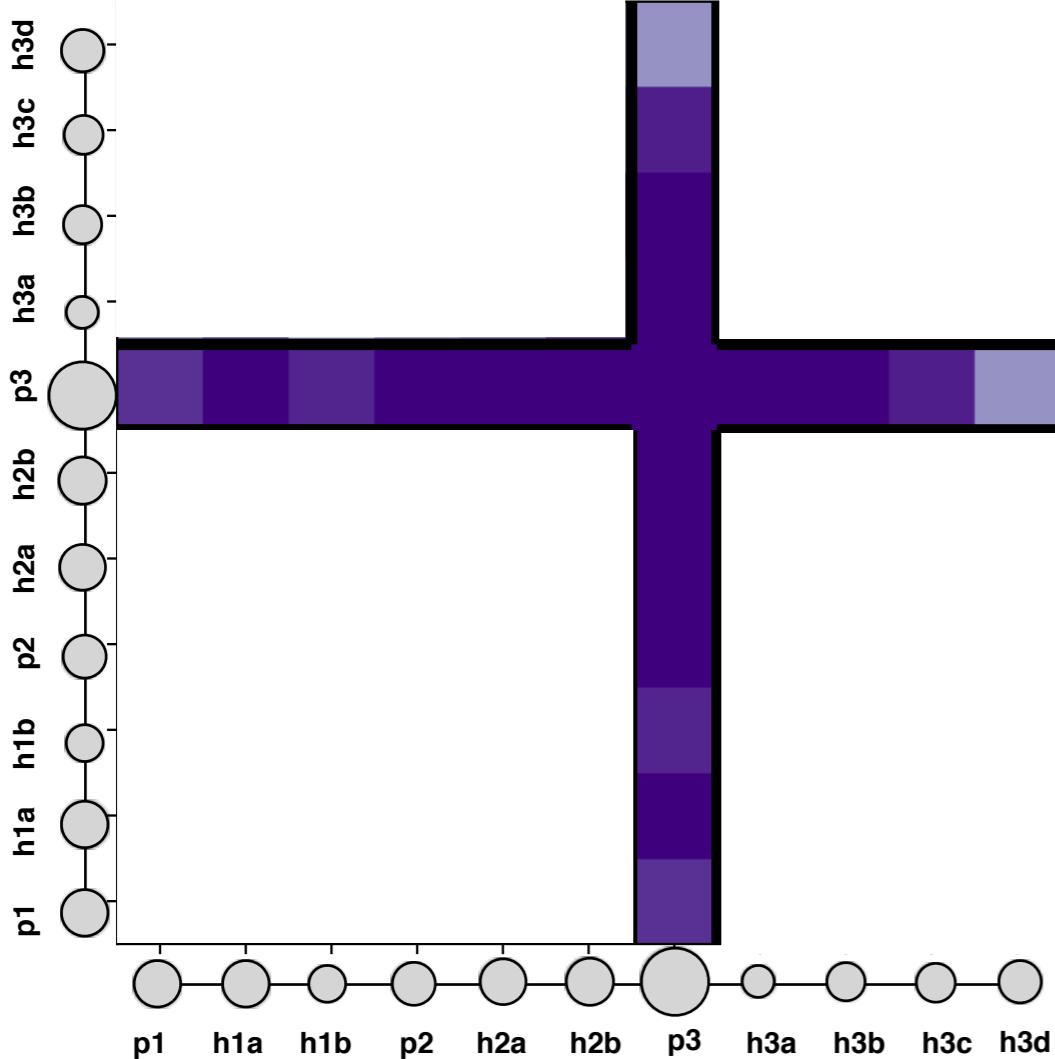
Define blobs from sequence



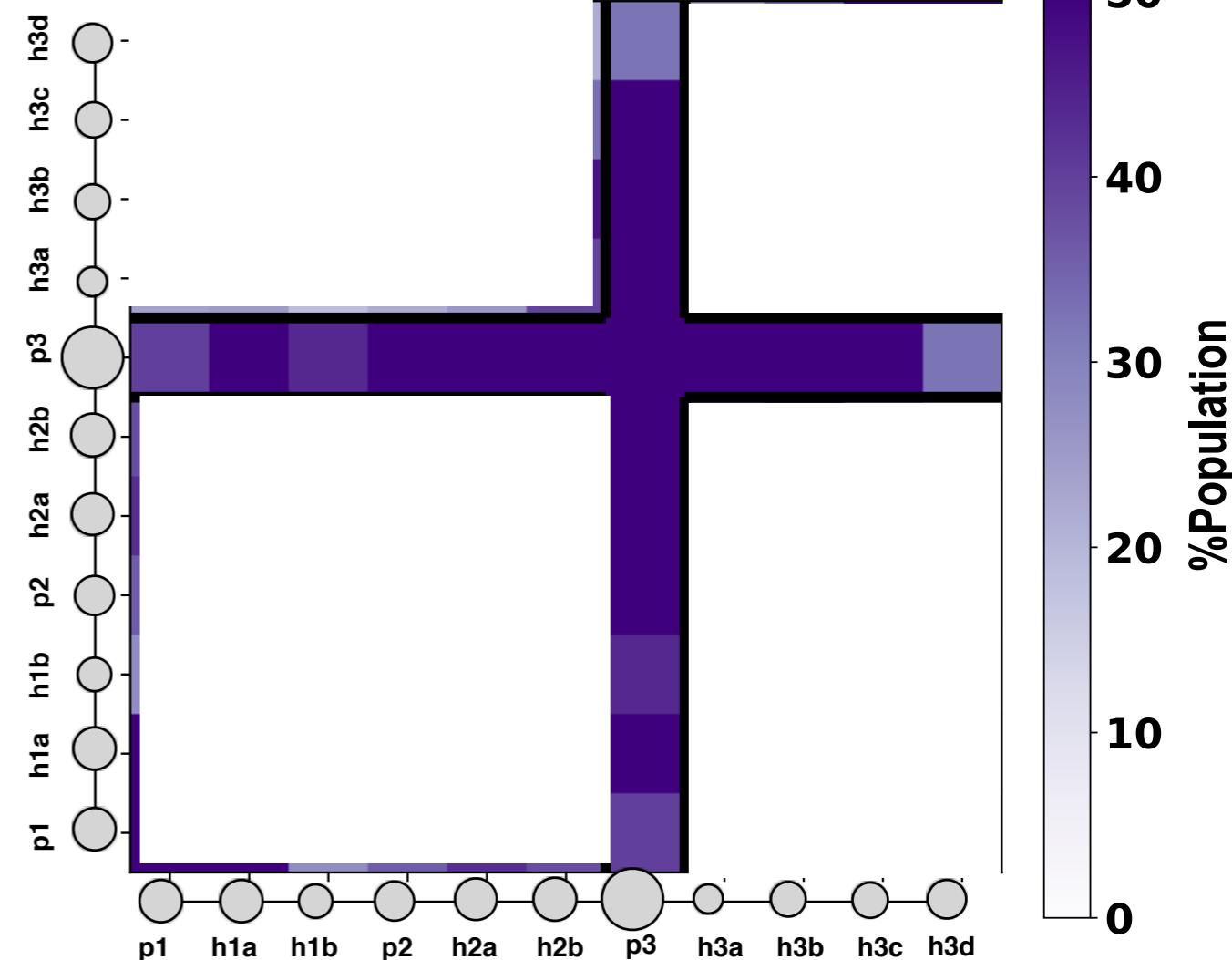
Calculate RP contact probabilities



RP (V66)

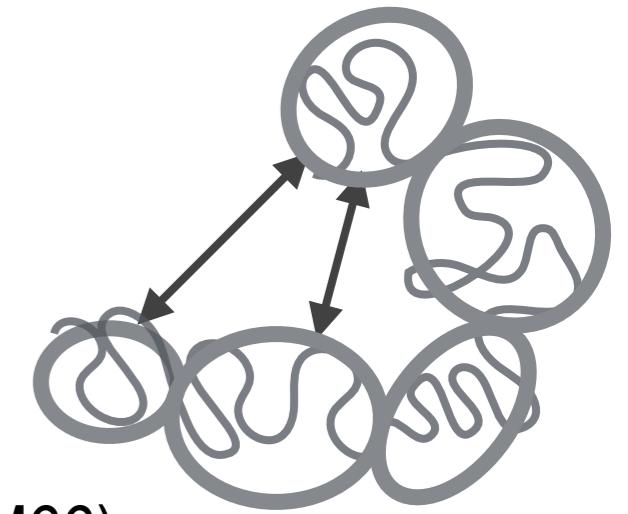


RP (M66)

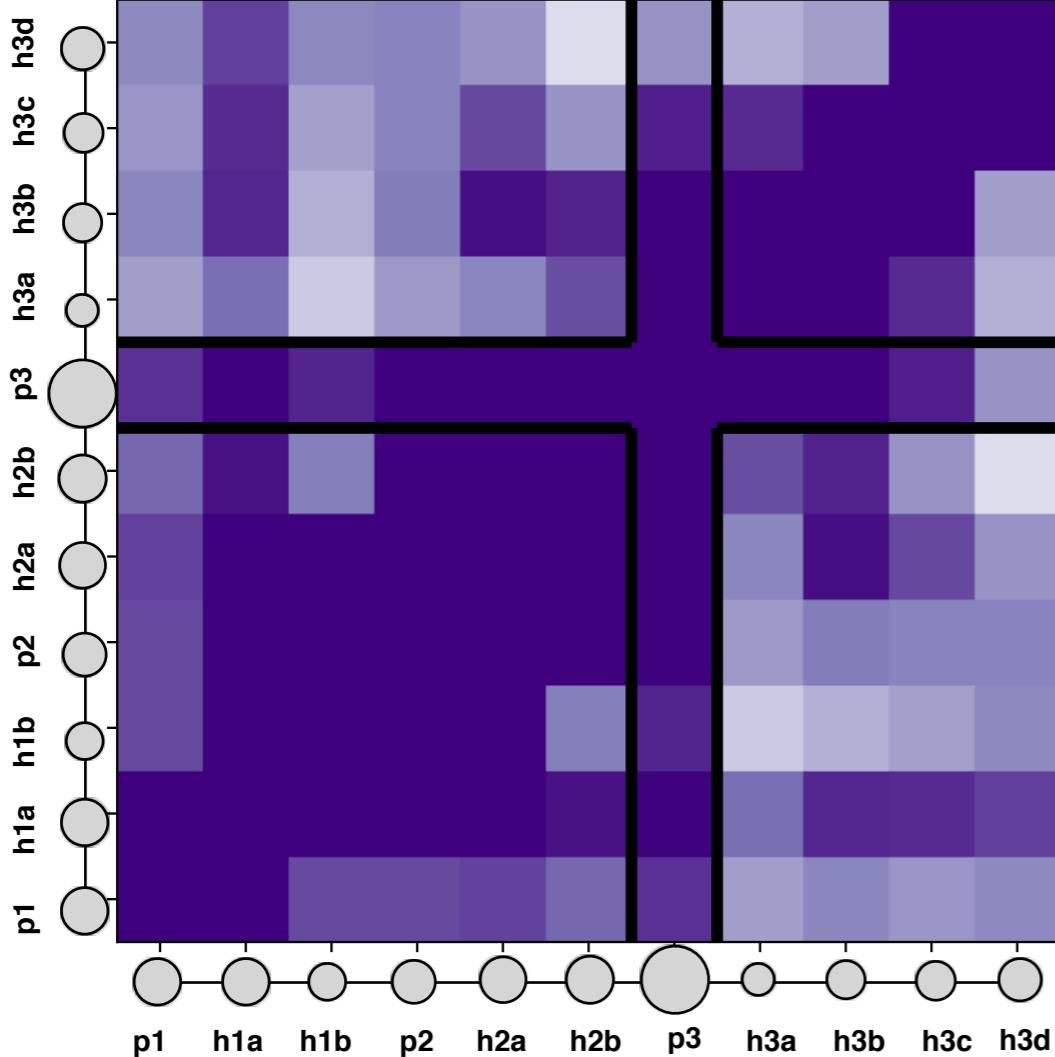


How many of these contacts result from short-range structure?

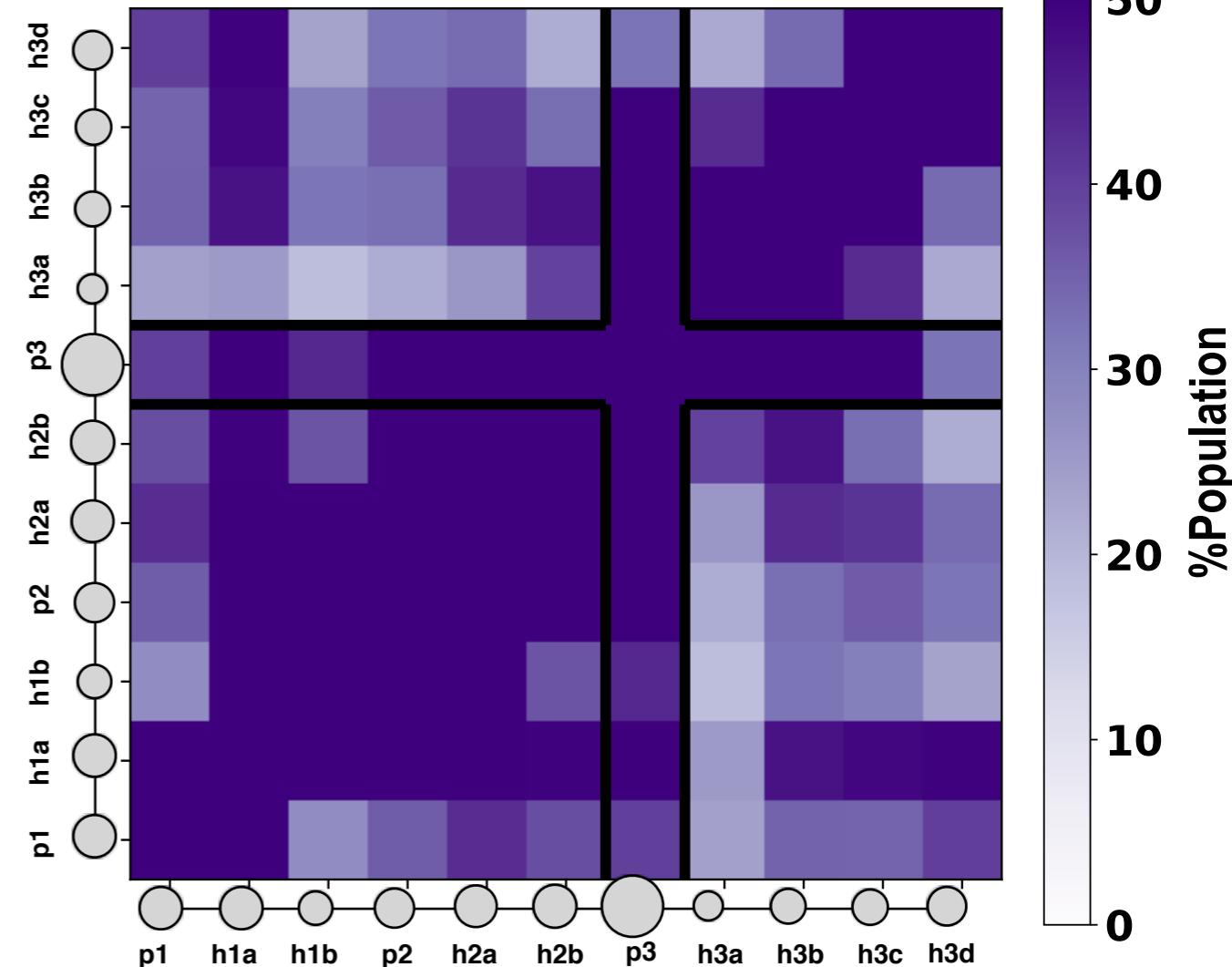
Calculate RP contact probabilities



RP (V66)

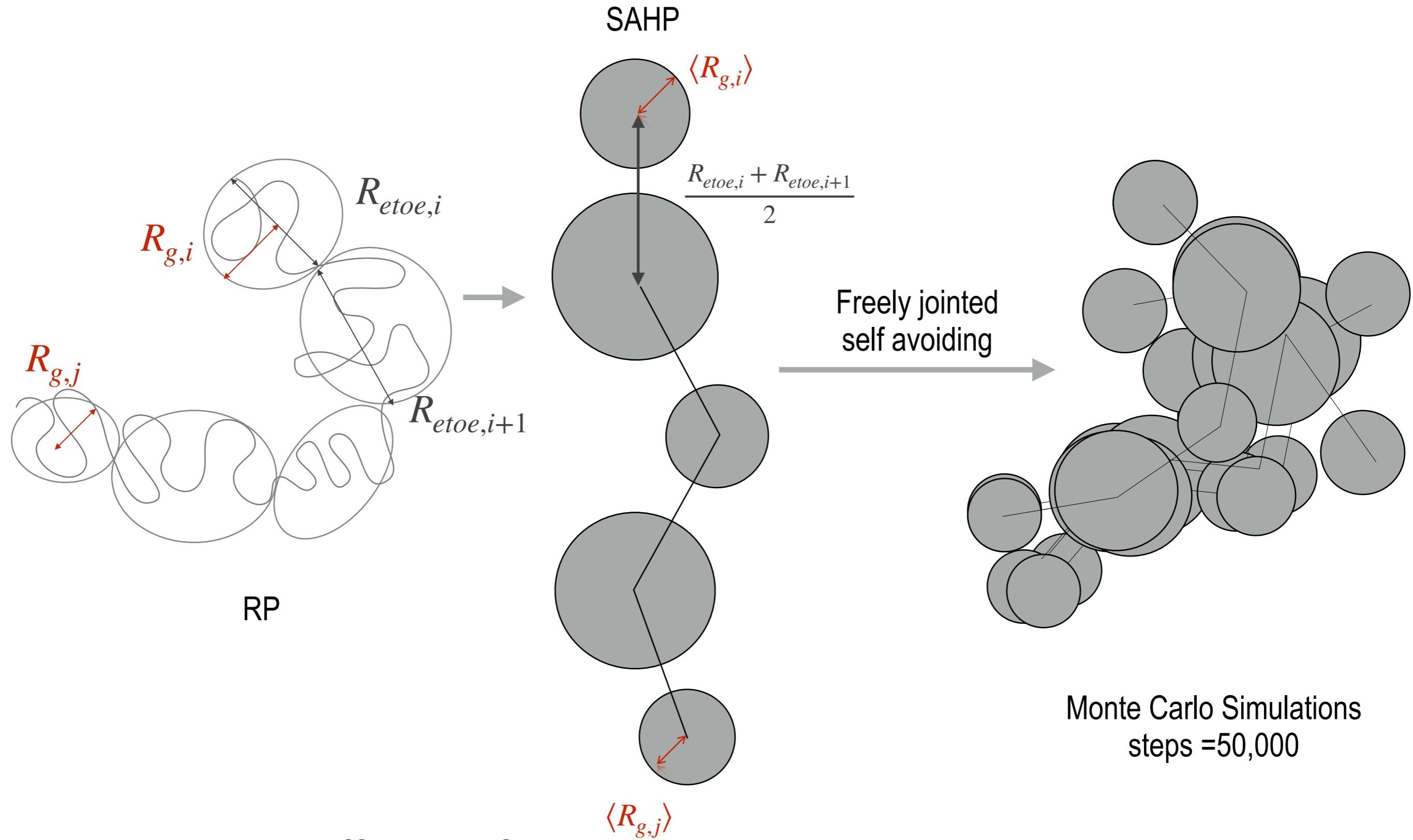


RP (M66)



How many of these contacts result from short-range structure?

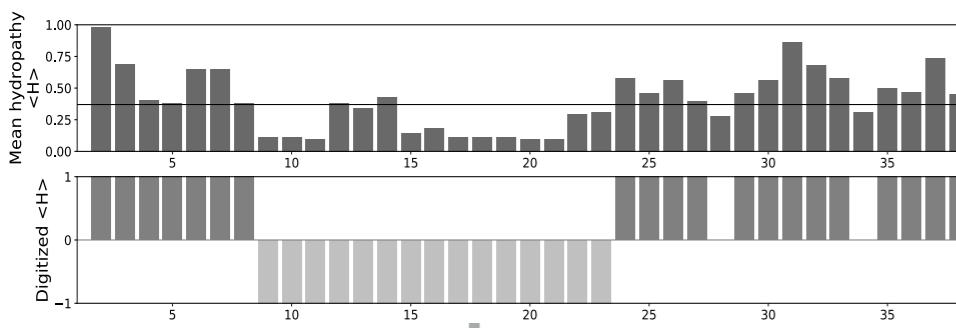
Reference Data: Self-avoiding heteropolymer



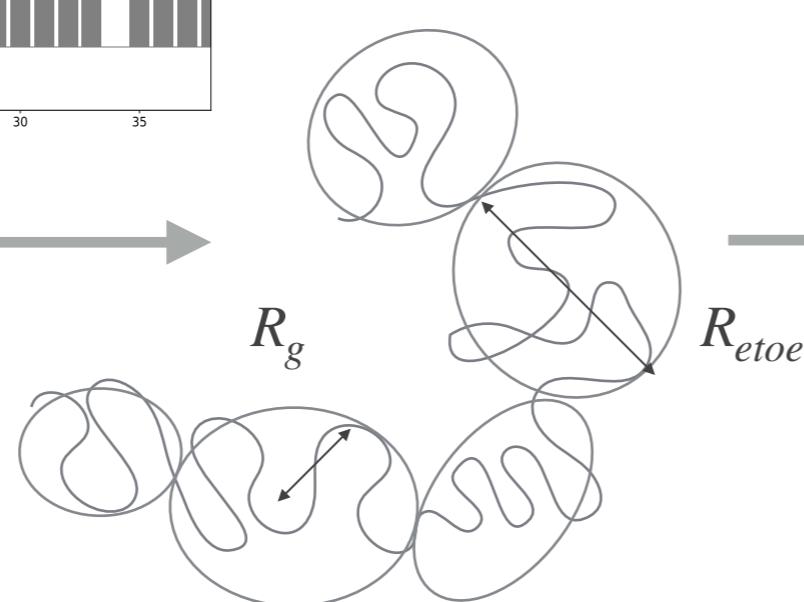
Goal: Retain effects of short-range structure without long-range interactions

Hierarchical framework for identifying short-range structure

Define blobs from sequence

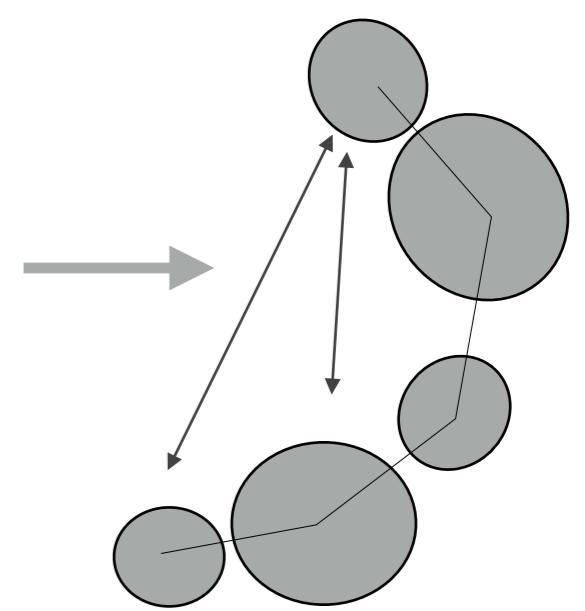


Parameterize self-avoiding heteropolymer (SAHP)



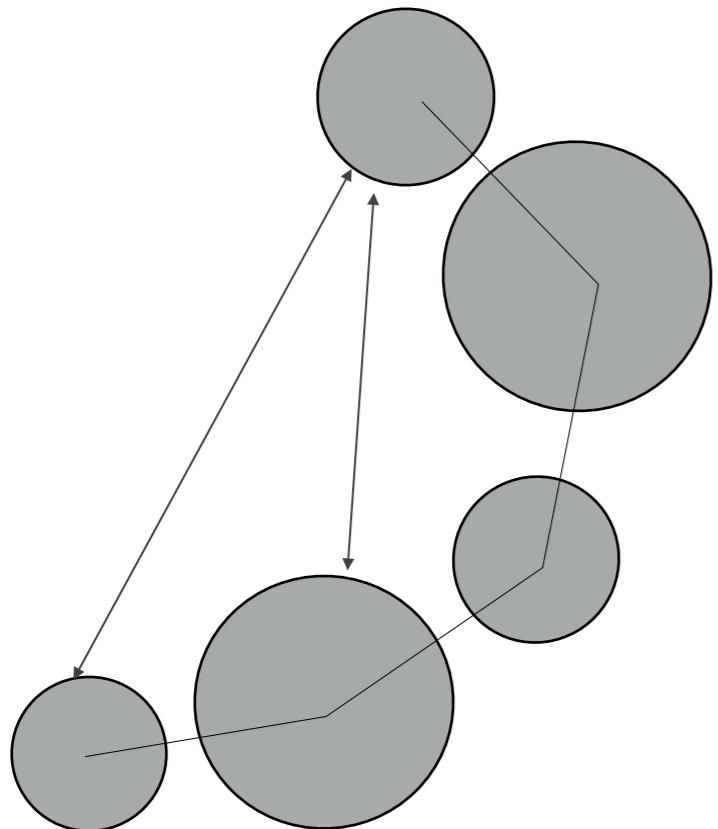
Calculate average size/ shape of each blob

Simulate SAHP and calculate contact probabilities

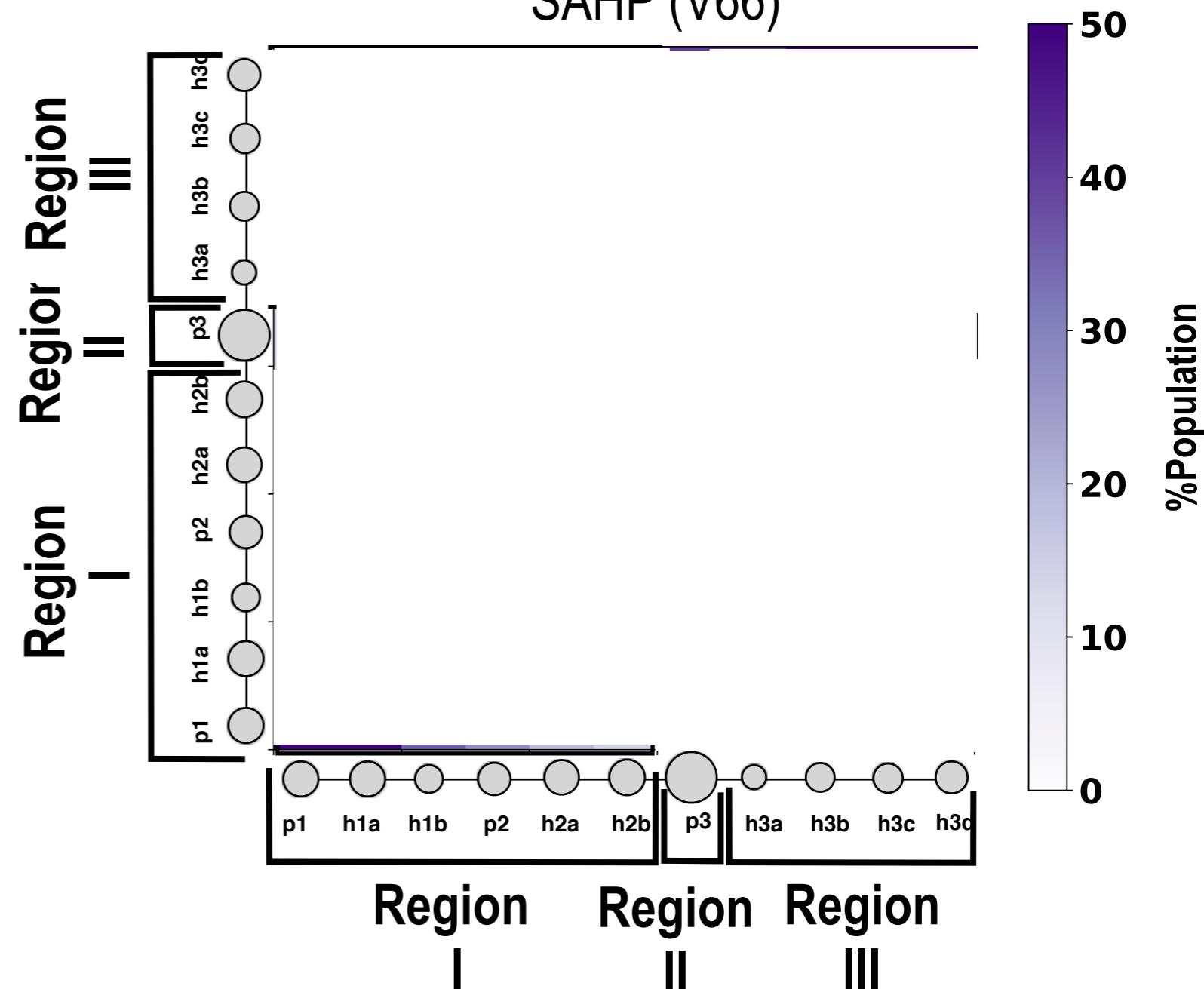


Calculate SAHP contacts

Probability of Contact

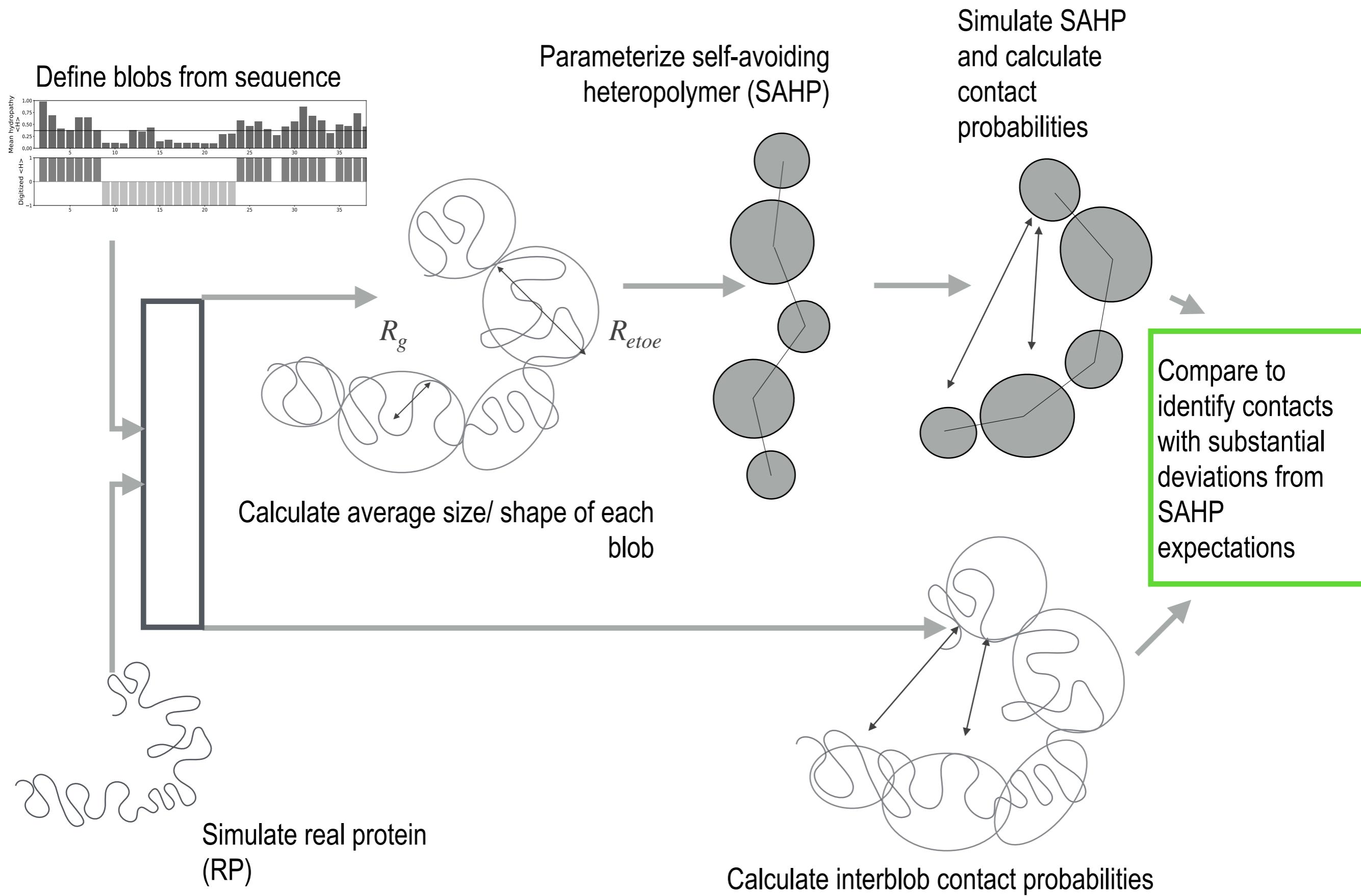


SAHP (V66)

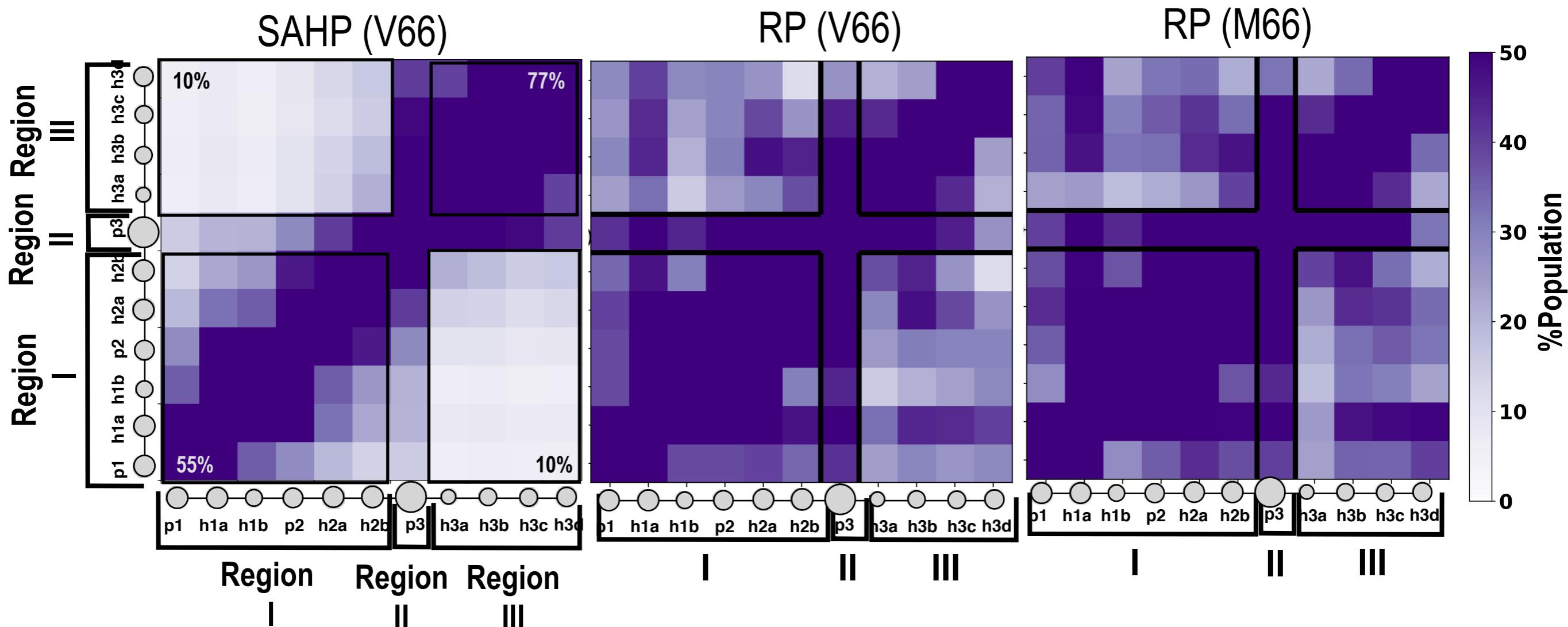
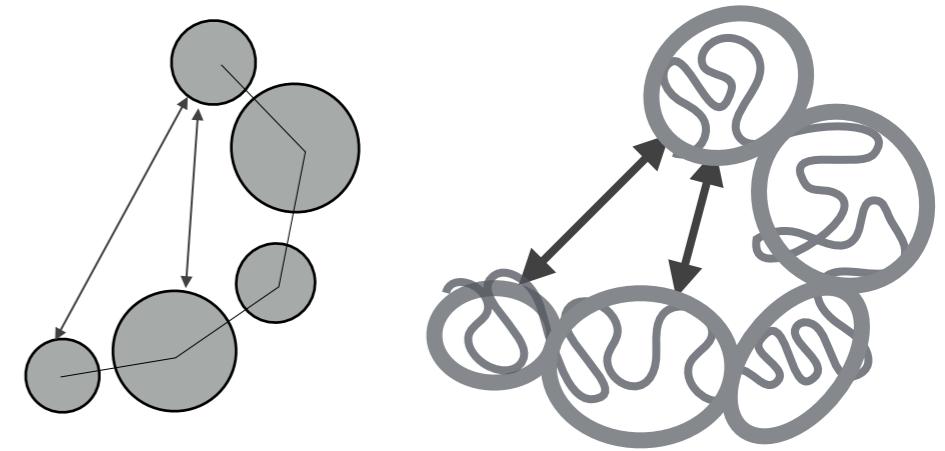


prediction : Large extended domain would effectively split the protein into two segments that don't interact

Hierarchical framework



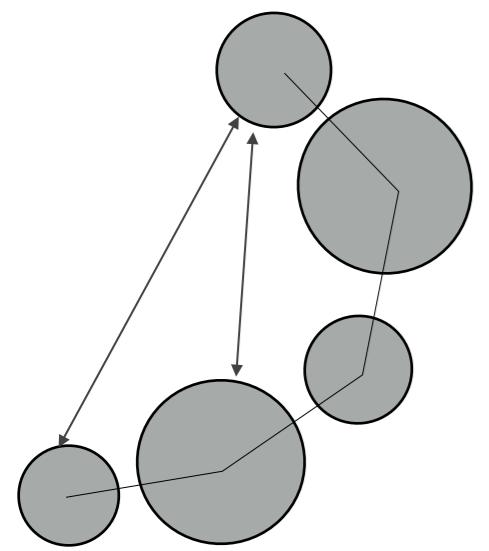
Calculate RP contact probabilities



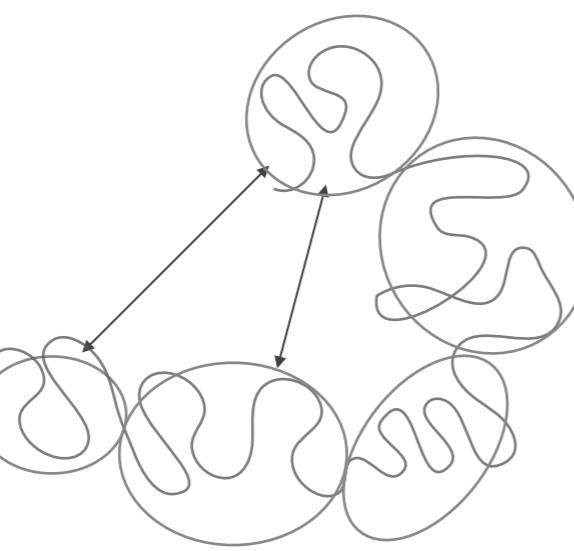
Segmentation predicted by short-range structure alone

Compare to identify contacts with substantial deviations from SAHP expectations

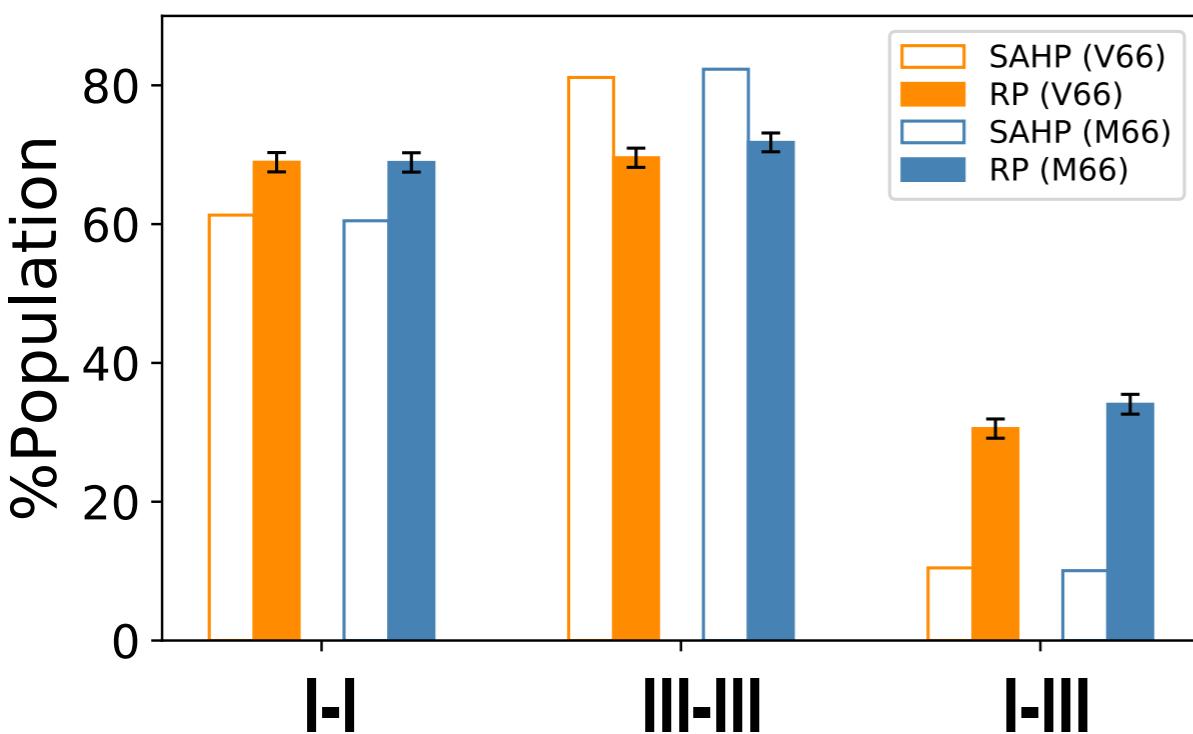
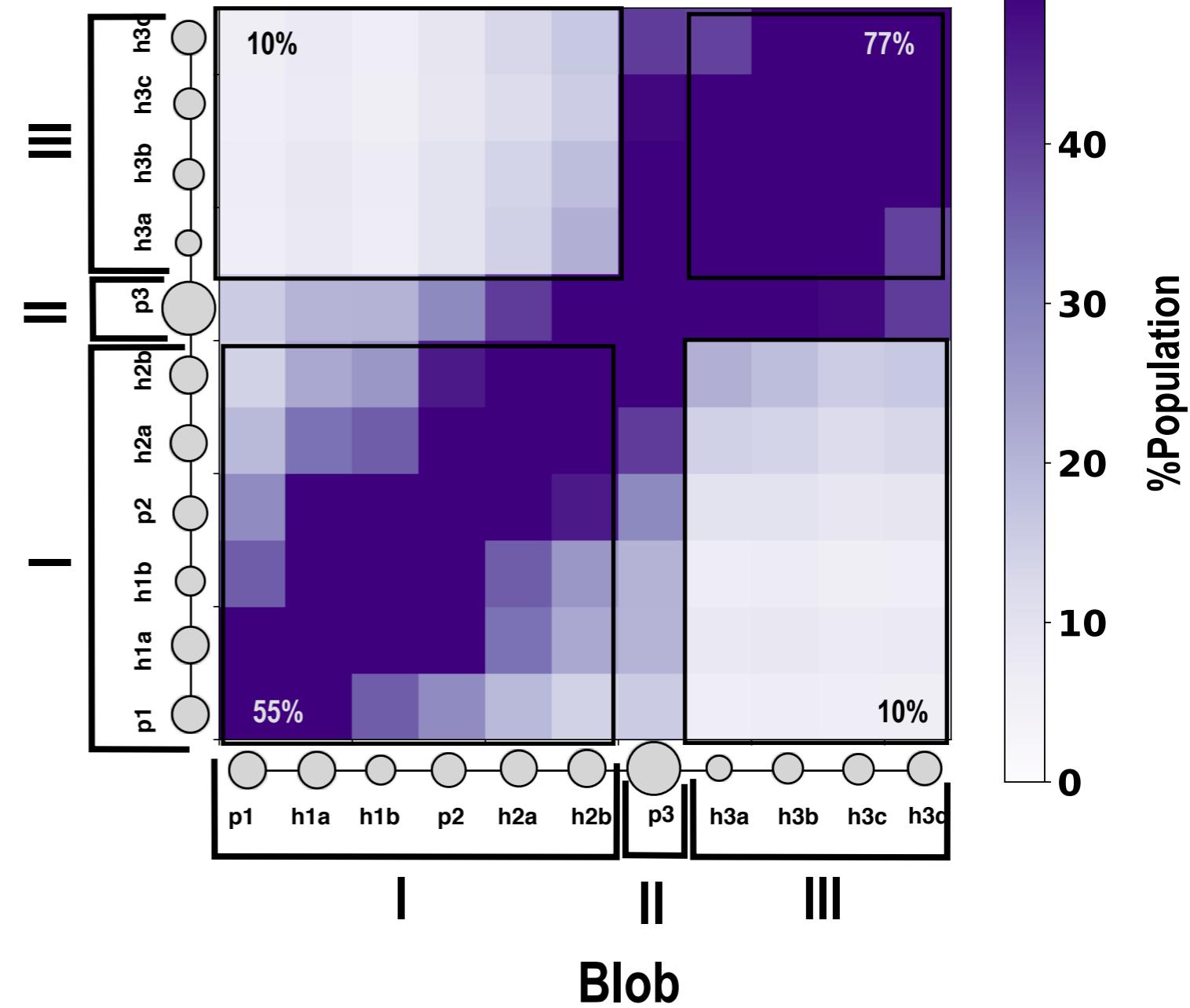
SAHP



RP

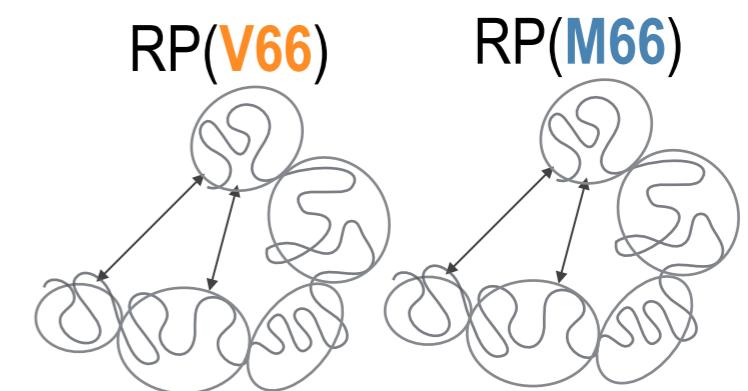
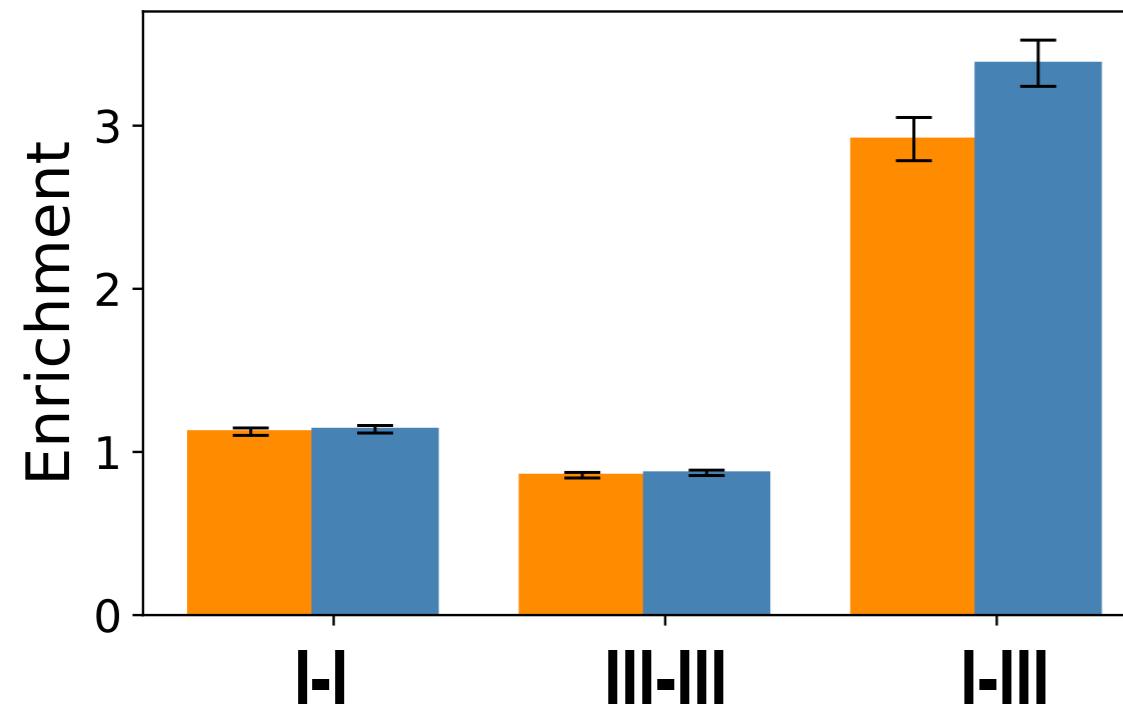


SAHP (V66)



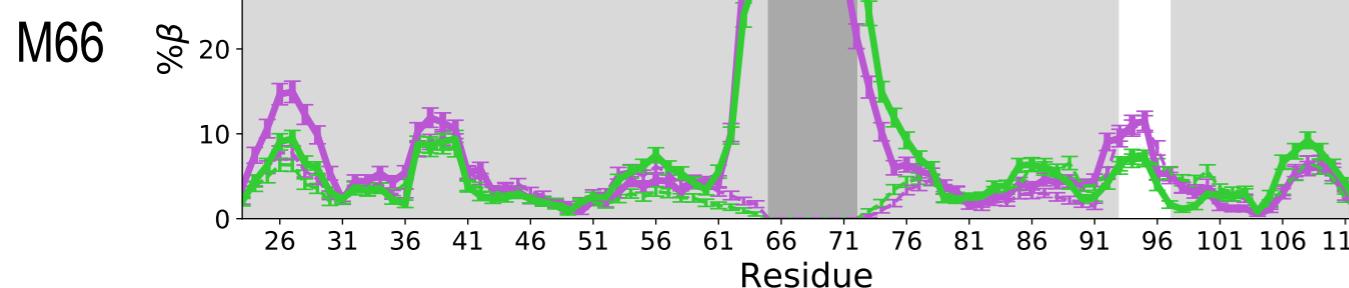
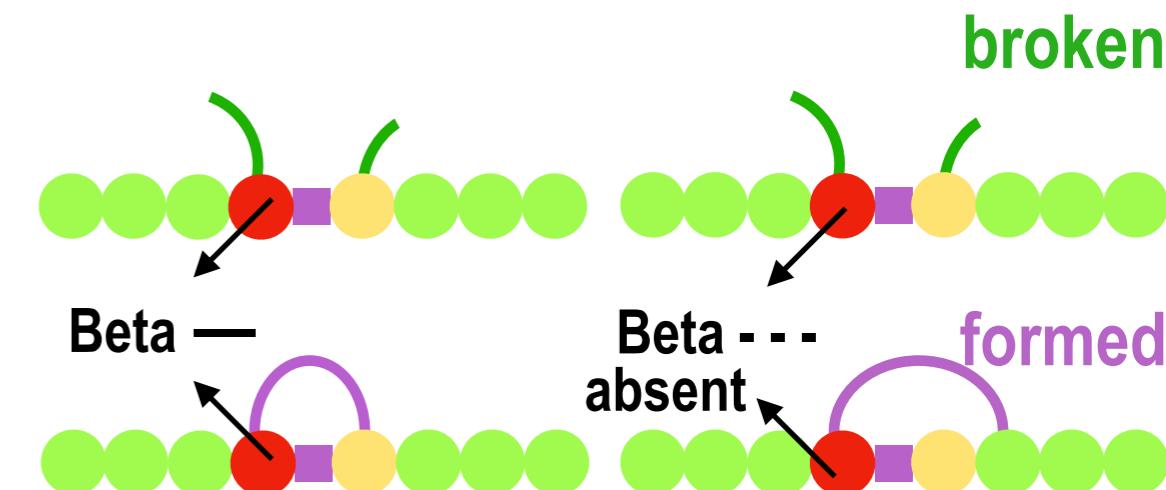
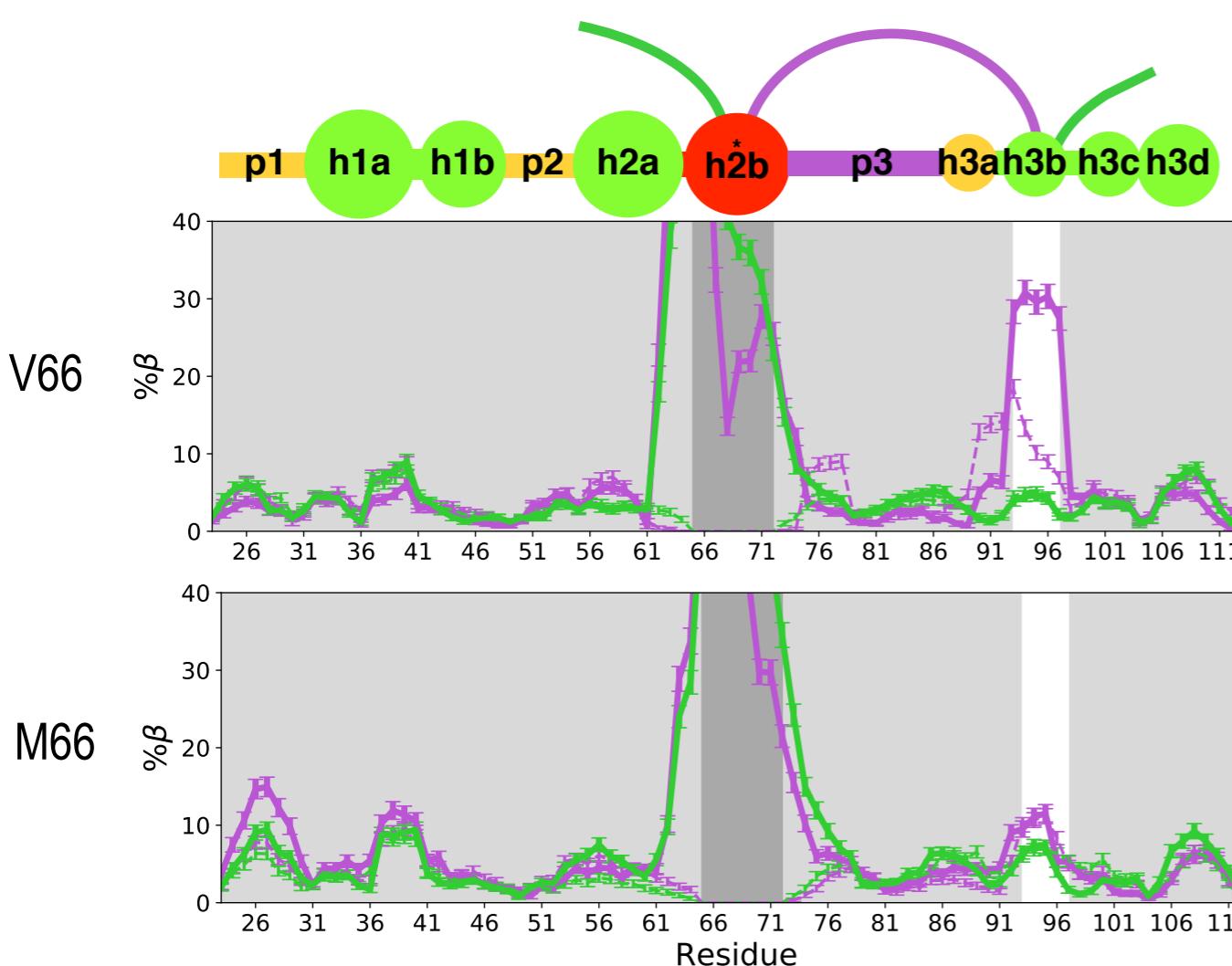
RP : two sides of the real protein (RP) interact more with each other!

Val66M increases contact frequency between domains on either side of linker domain p3



Two sides of the protein interact more with each other!

h2b-h3b interactions stabilize beta formation only in V66



Blob h3b and h2b are stabilized with beta pairing only in V66

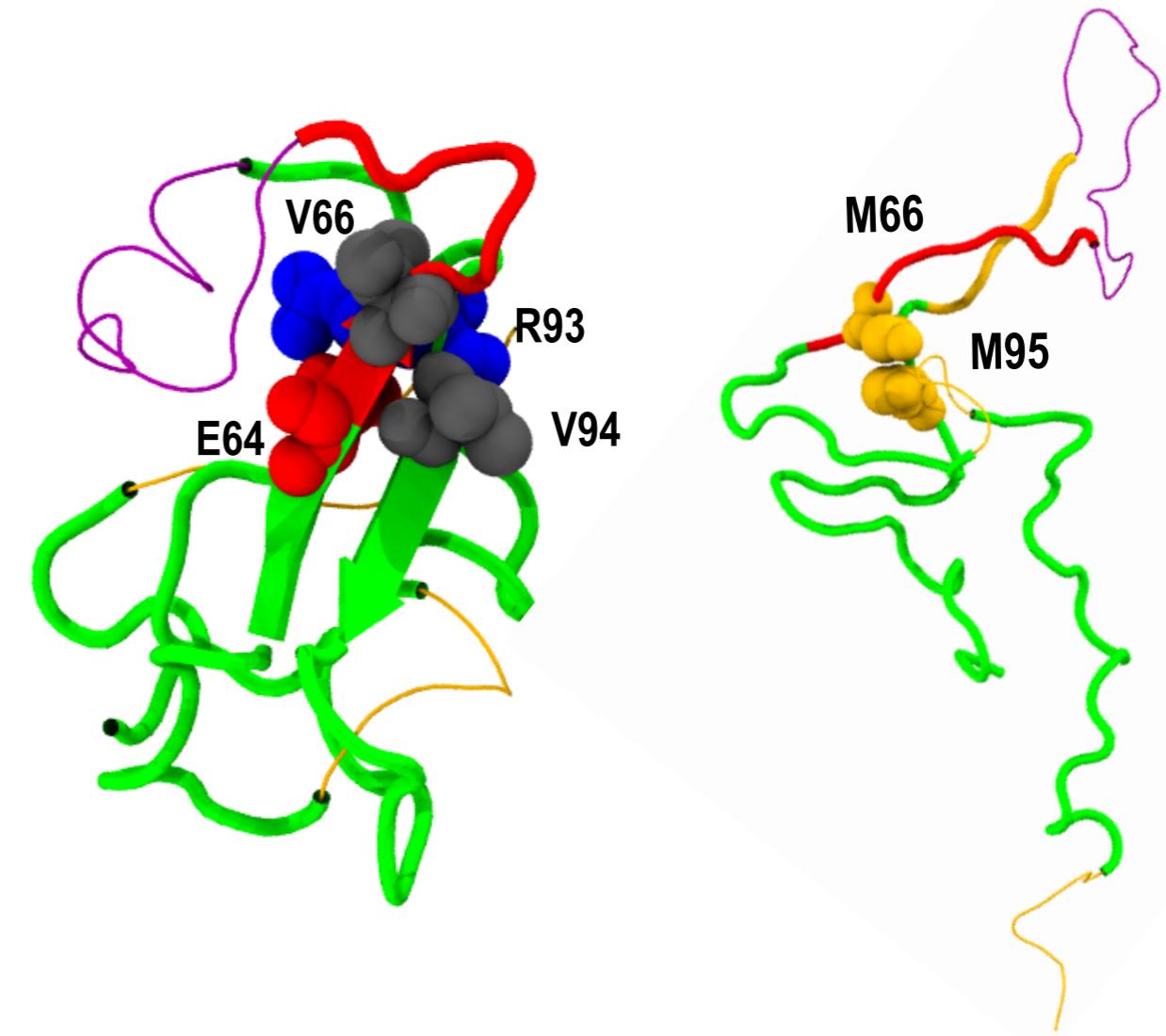
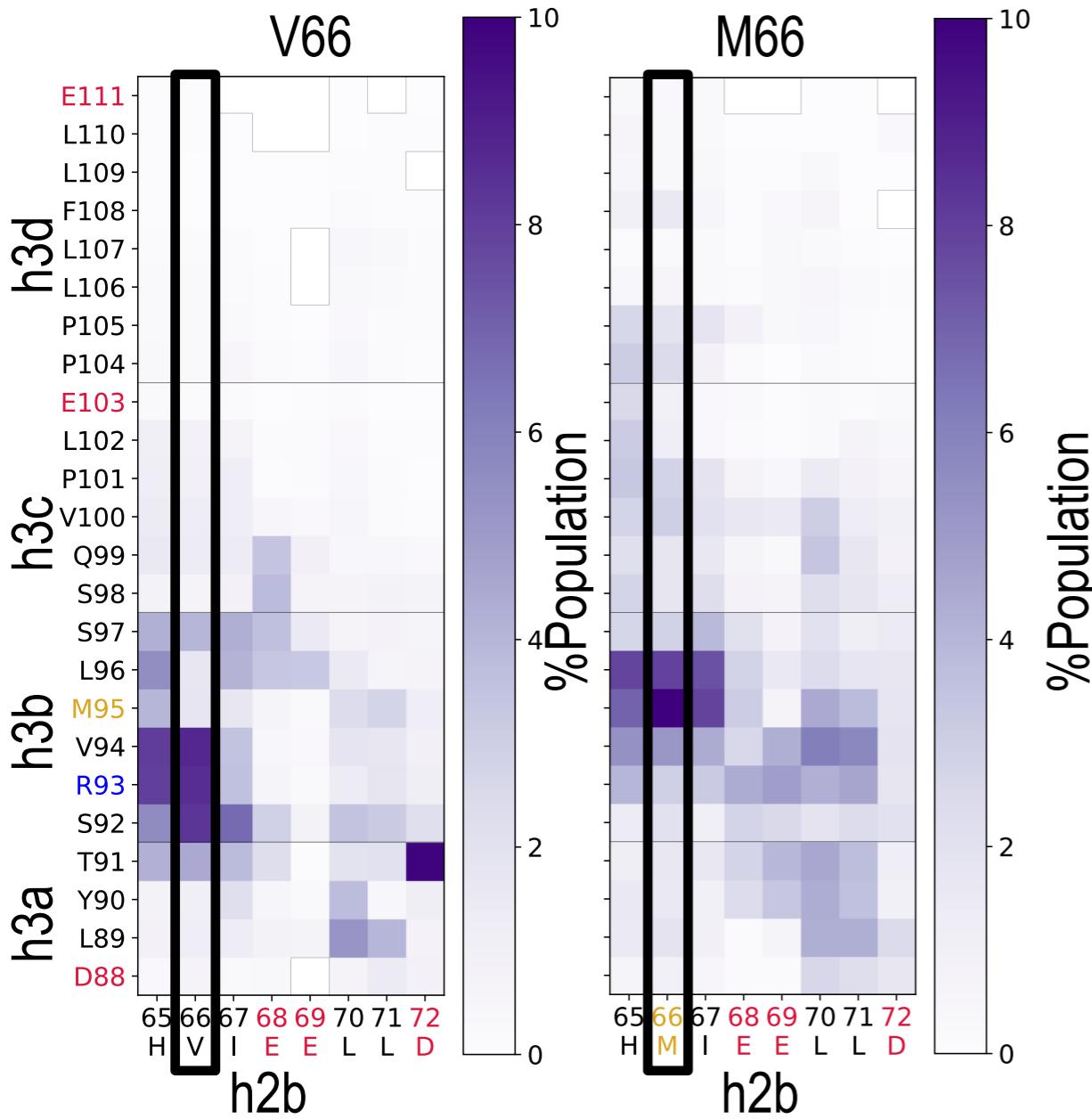
high hydrophobicity



low hydrophobicity



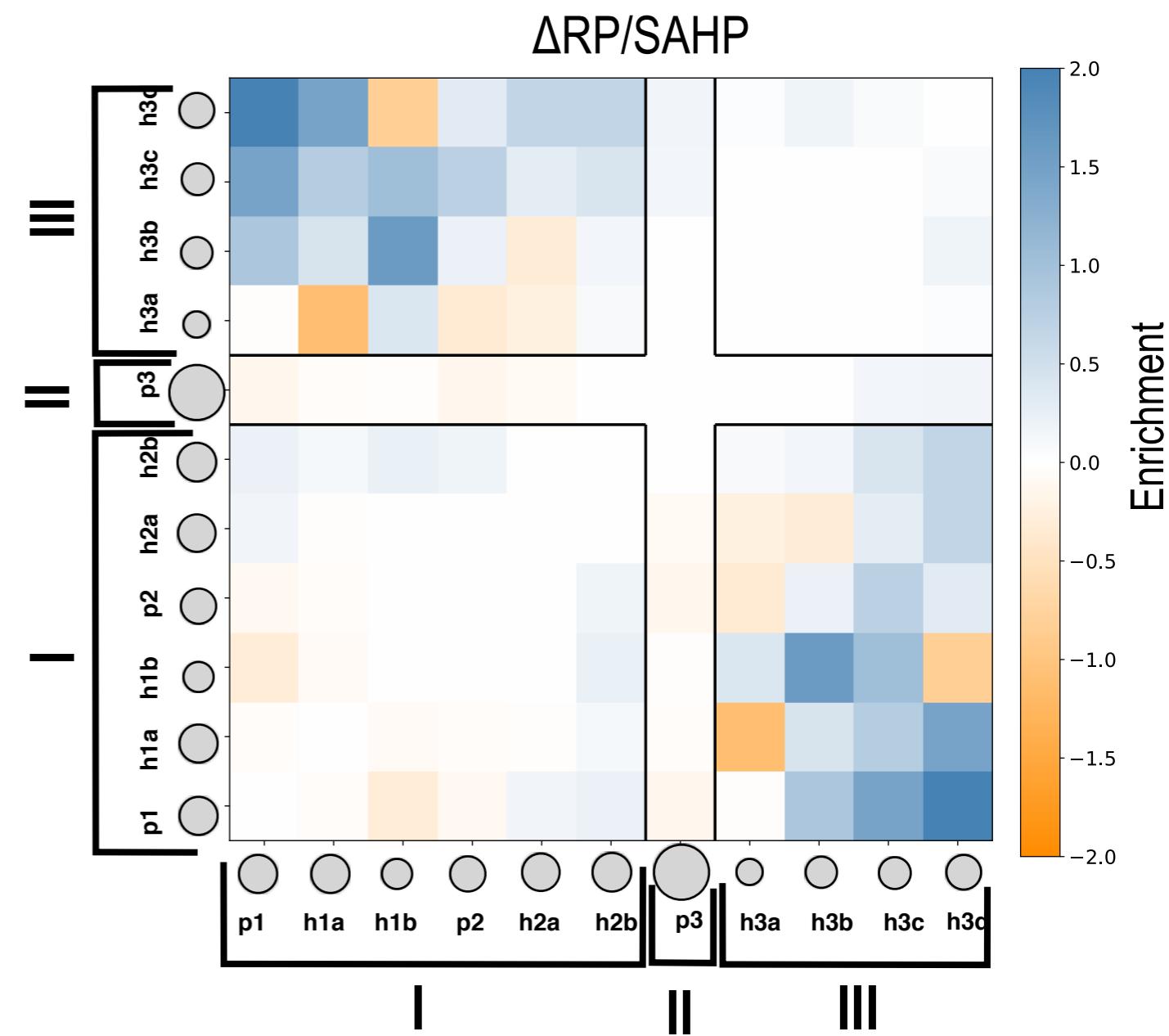
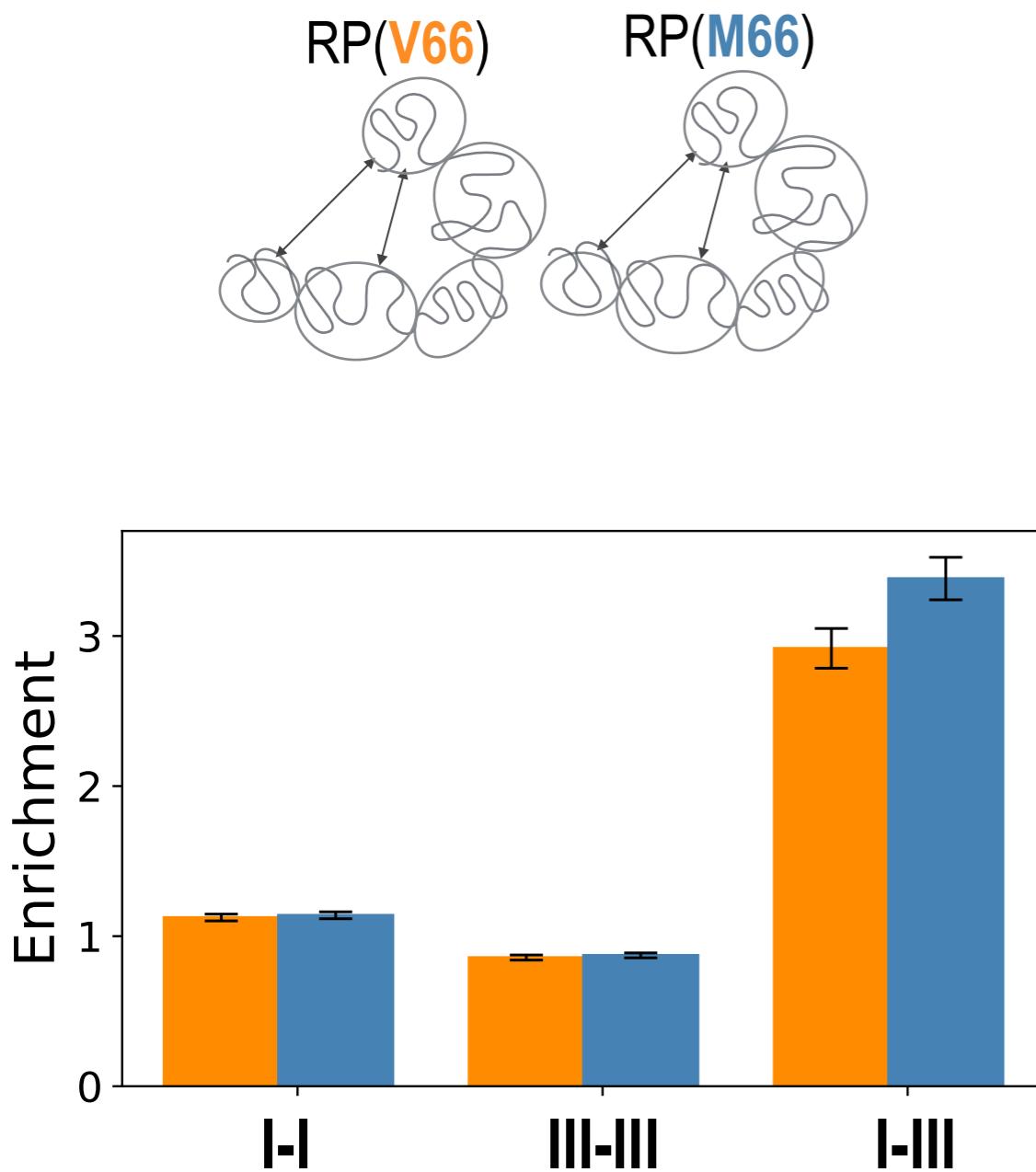
Residue level interactions



Met-Met interactions are particularly strong among hp residues:

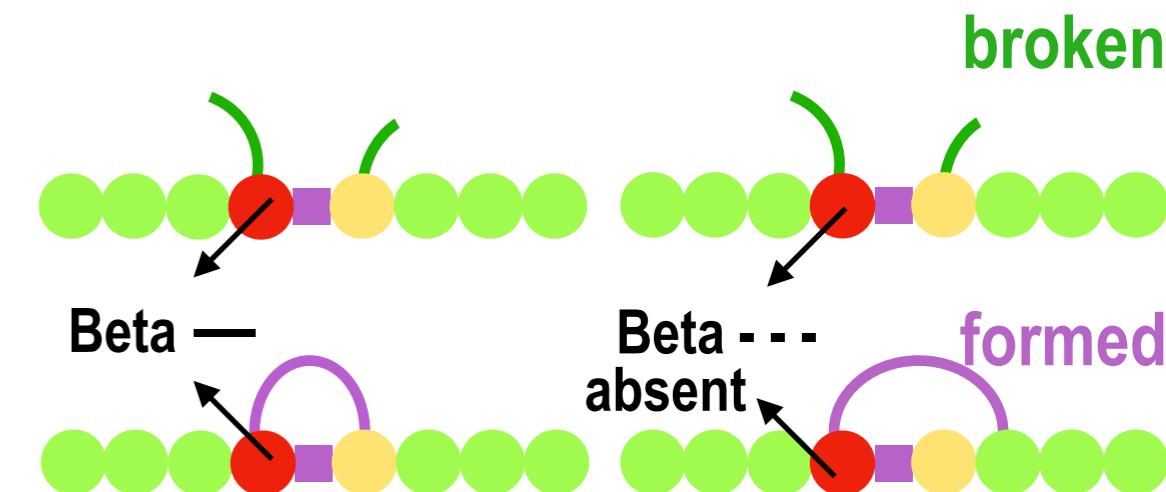
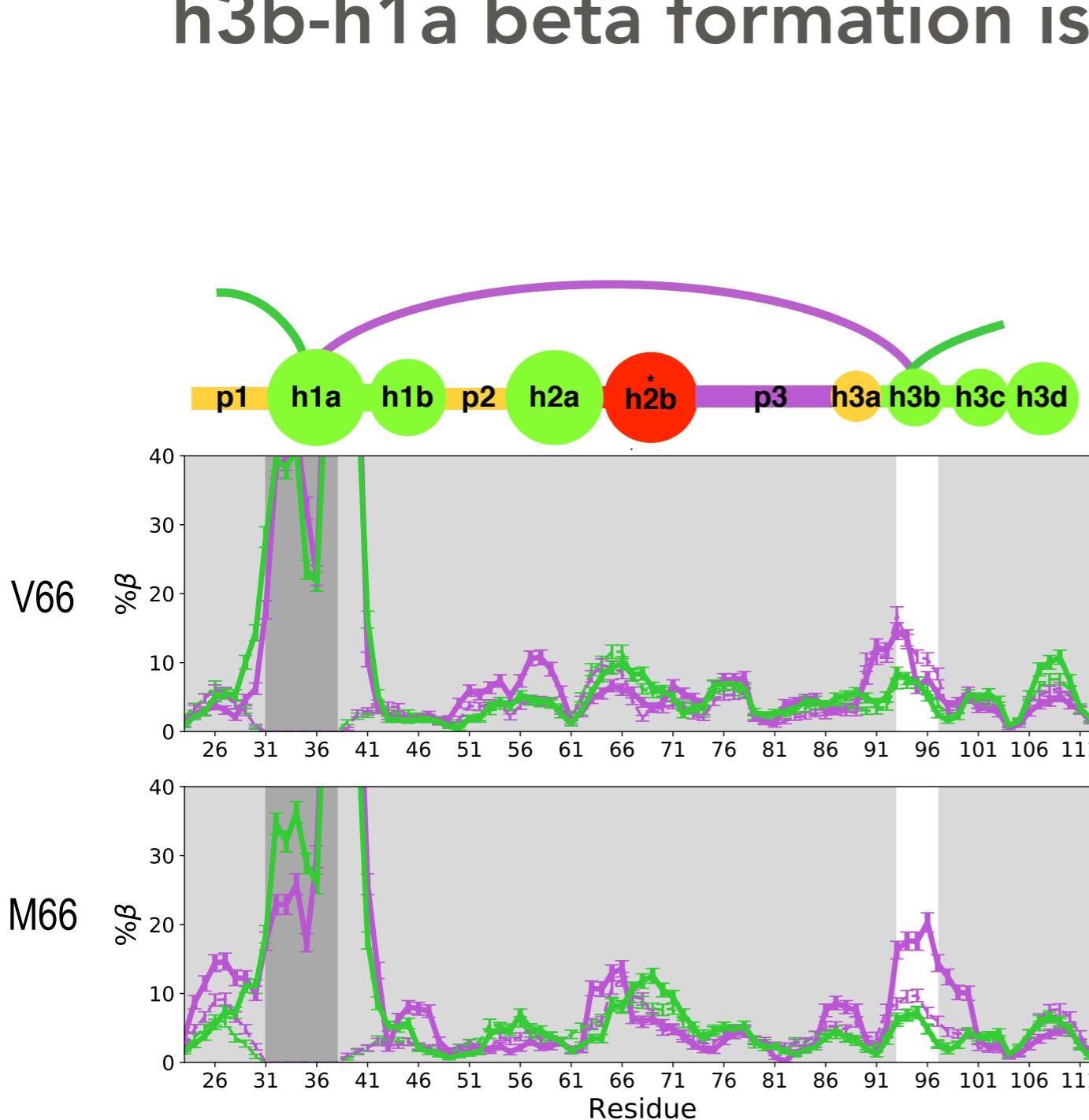
Met—Met > Met—Val or Val—Val, Met—Phe or Aromatic—Aromatic interactions

Val66M increases contact frequency between domains on either side of linker domain p3



M66 : two sides of the protein interact more with each other!

h3b-h1a beta formation is stabilized only in M66



Domain h3b and h1a are stabilized with beta pairing

high hydrophobicity

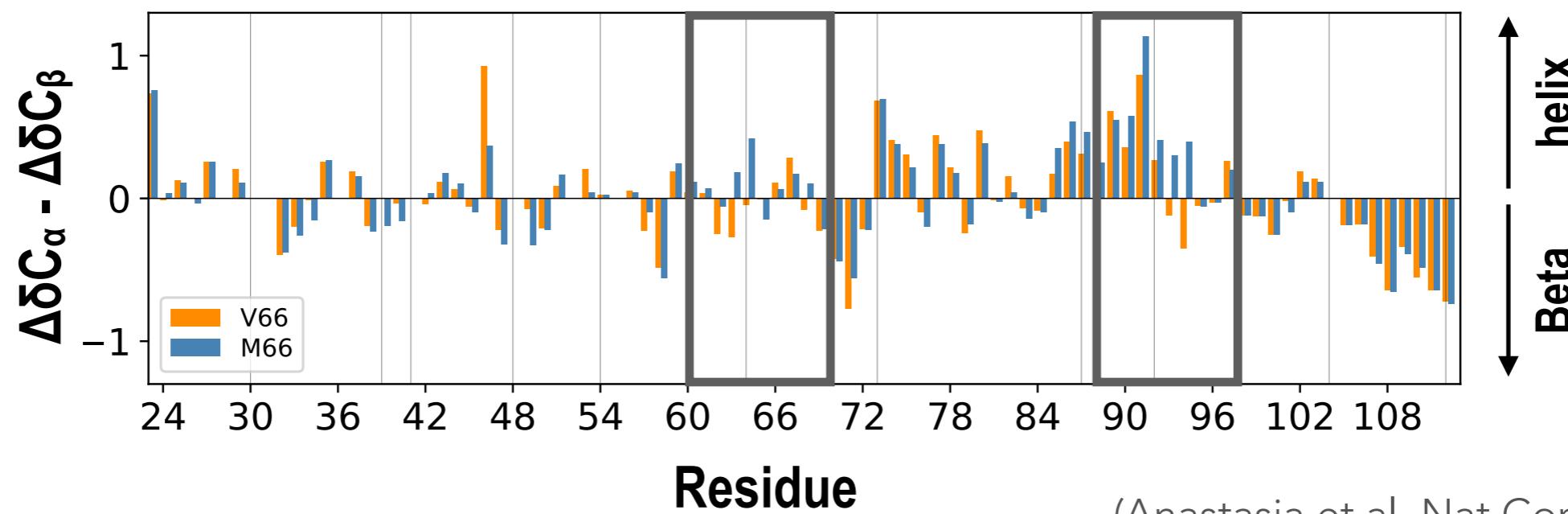


low hydrophobicity

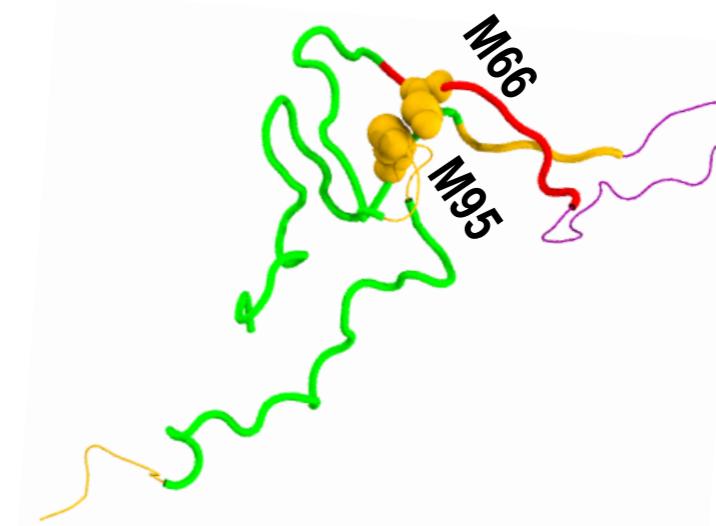
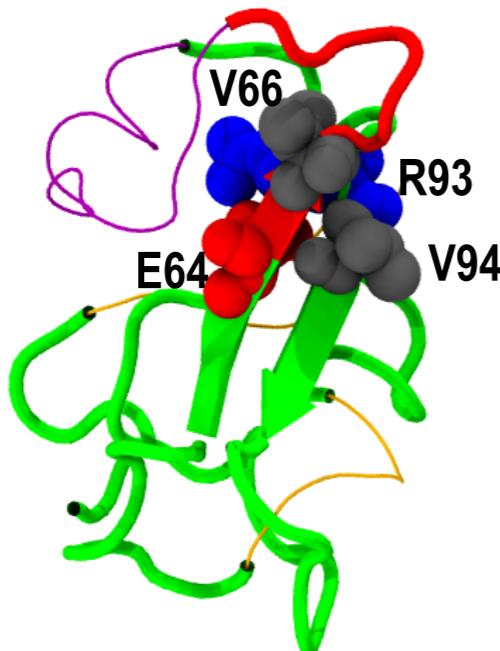


Val66Met causes change in non-local secondary structure

Same study detects effects of Val66Met on residual secondary structure



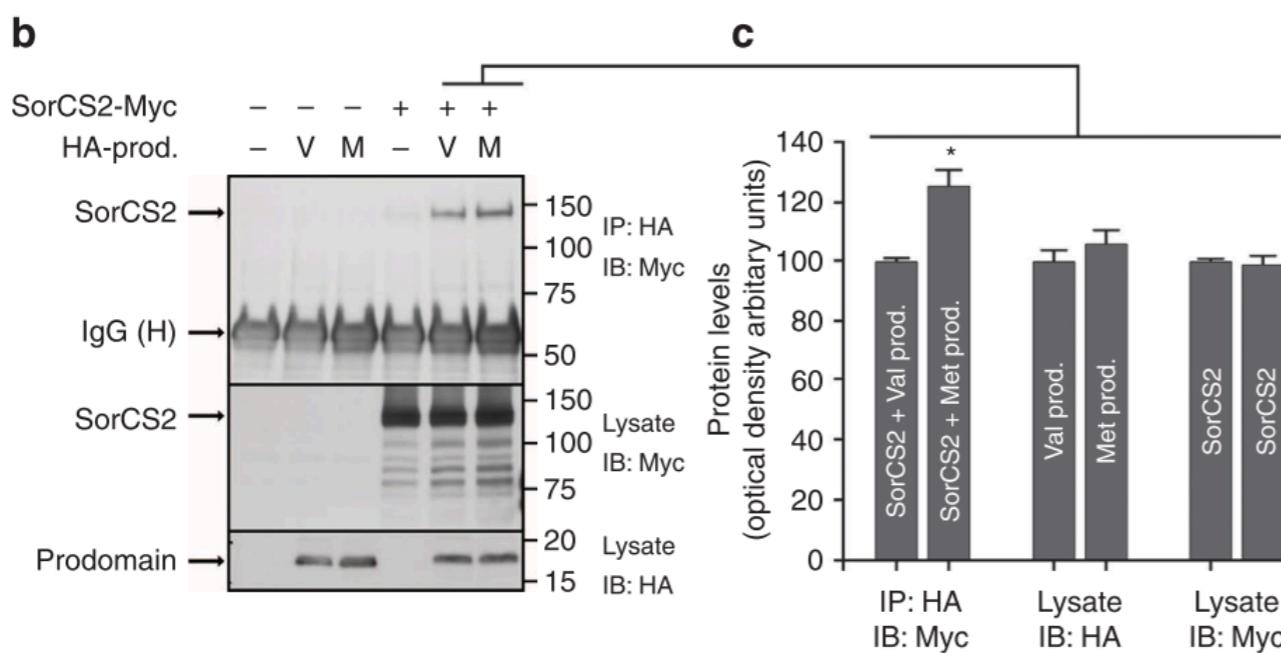
(Anastasia et al., Nat Commun., 2013)



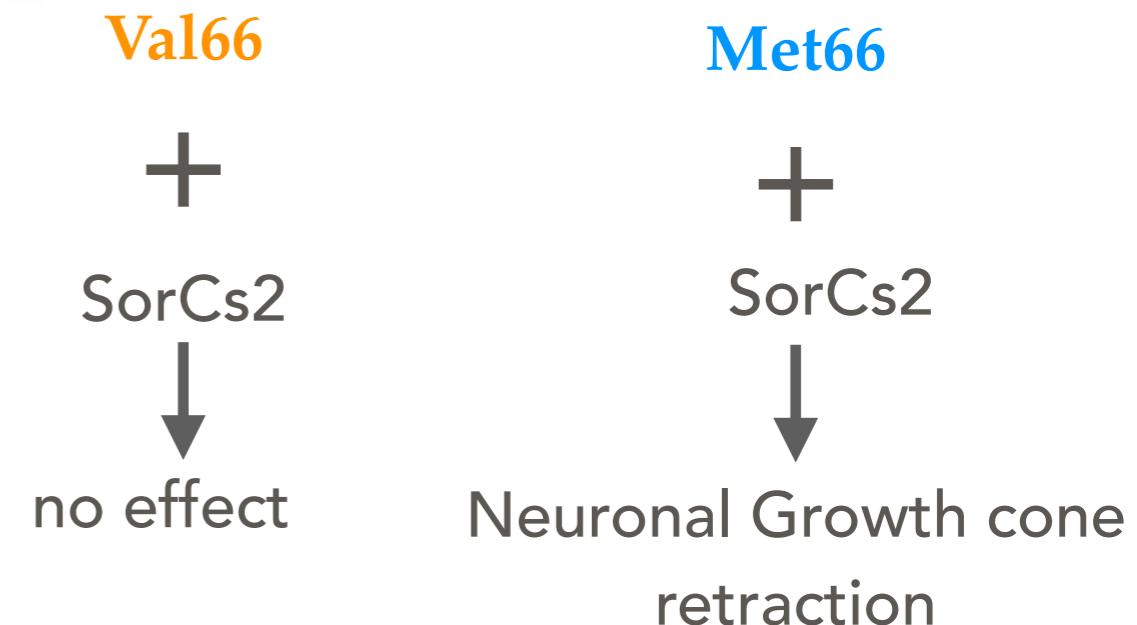
SorCS2 — M66 prodomain interaction



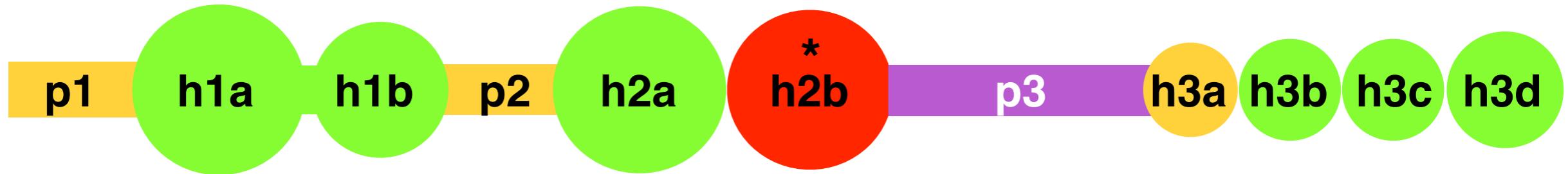
Potential Mechanism : binding to SorCs2



(Anastasia et al., Nat Commun., 2013)



Summary



- Over 250 μ s of atomistic explicit solvent T-REMD simulation of a 90 residue IDP; secondary structure/Rh agrees with NMR
- Hierarchical analysis involving sequence-based coarse-graining of hydrophobic domains
- Val66Met SNP lies in a unique strong polyelectrolyte/hydrophobic domain of proBDNF
- SNP increases the interactions between the two sides of long linker domain p3, stabilized by Met-Met interactions

(In Press - Lohia, Salari, Brannigan *PLoS Comput Biol* 2019)

high hydrophobicity

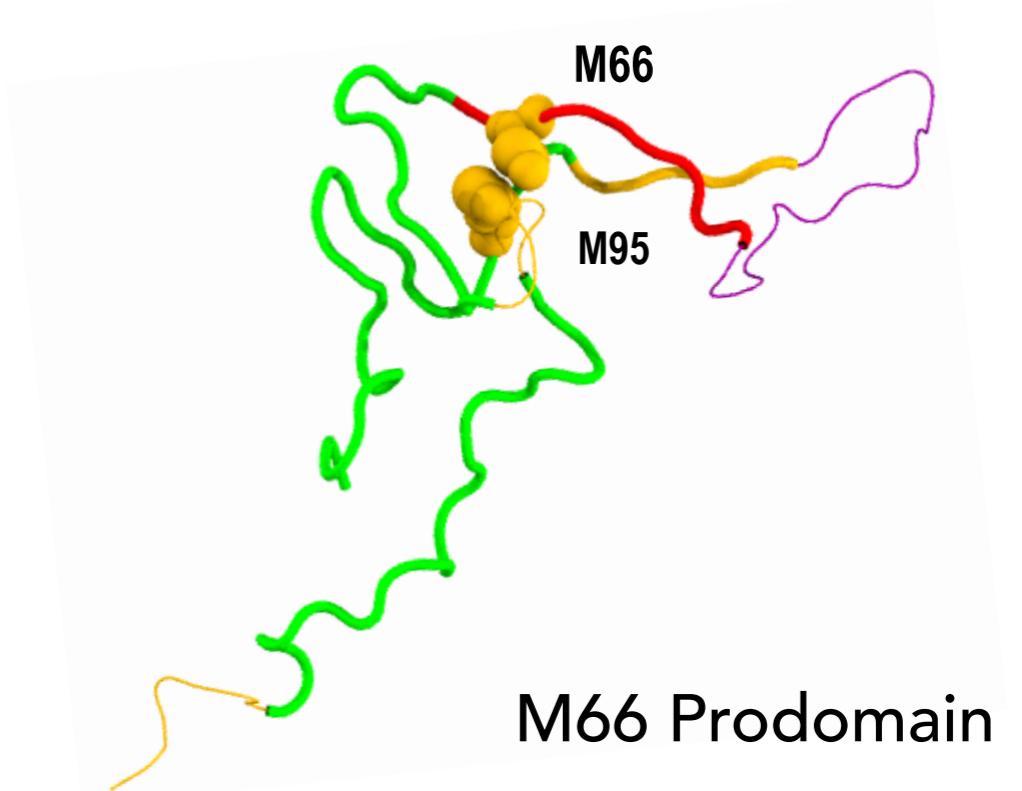
globular Janus strong - charged polyelectrolyte (& SNP)

low hydrophobicity

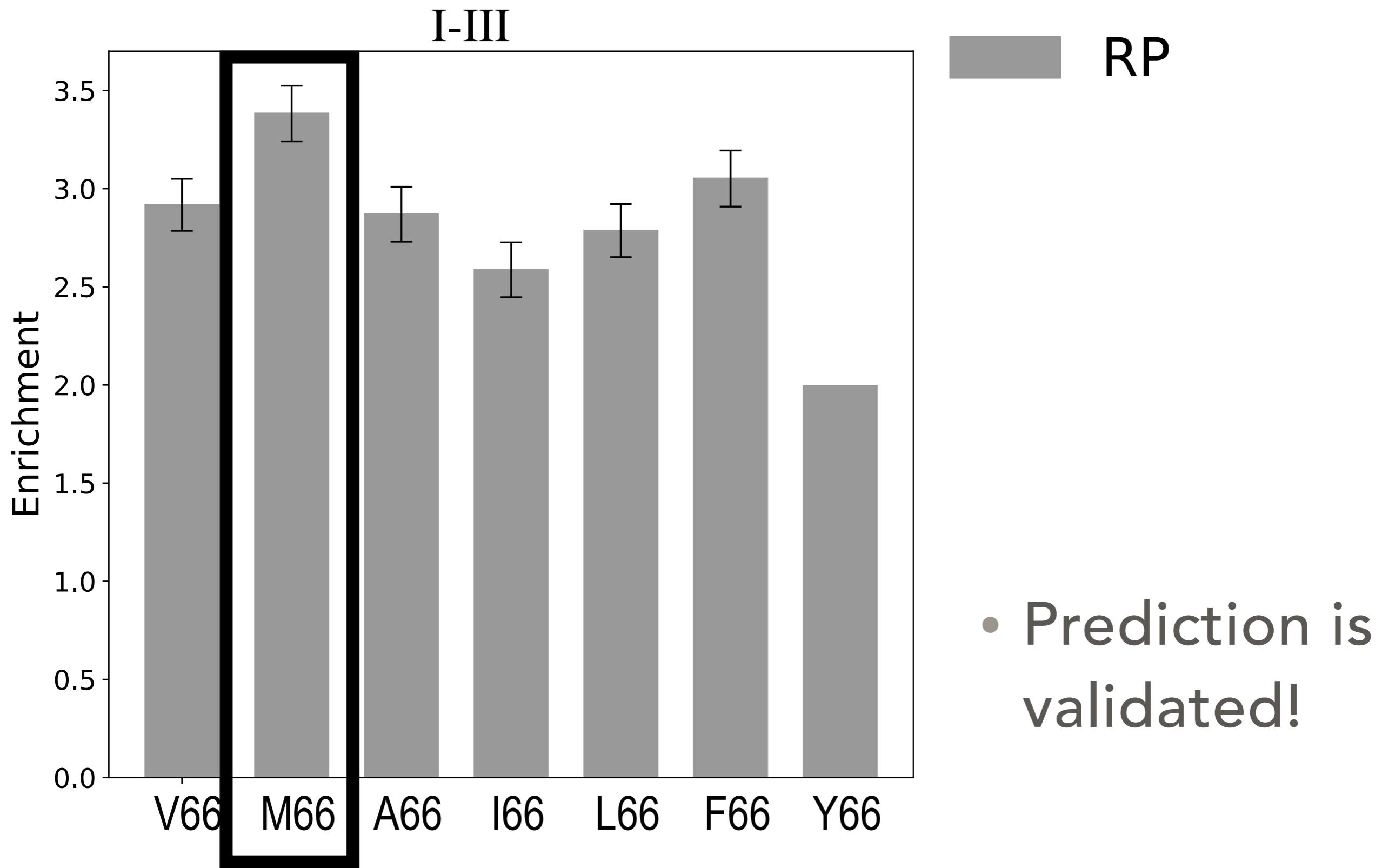
strong polyampholyte

#3a Does our hypothesis hold true for other hydrophobic mutations in prodomain BDNF ?

- Hypothesis: Increased contacts across the long linker are due to Met-Met contacts.
- Prediction: no other hydrophobic mutations at residue 66 will cause the same conformational effects.



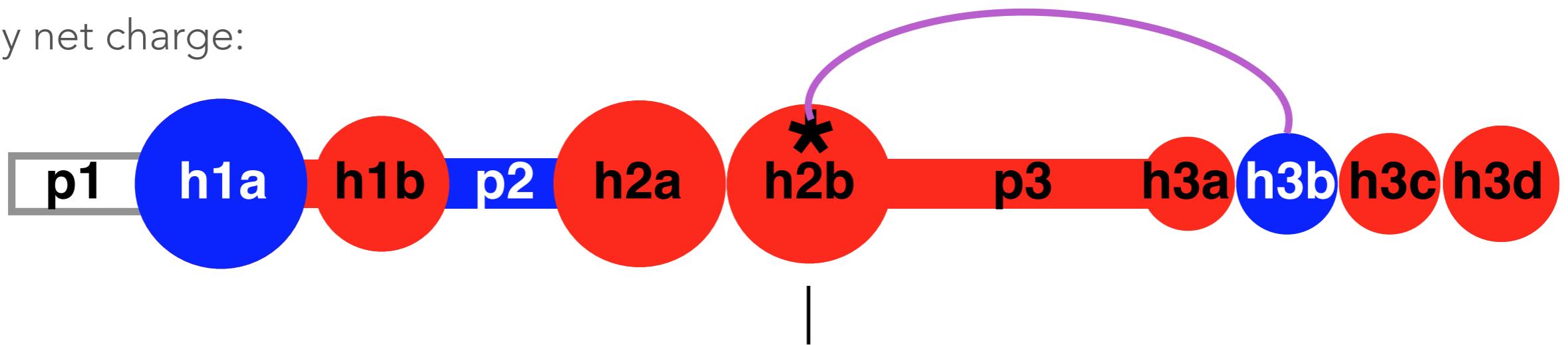
Effect of all other mutations on interblob contacts



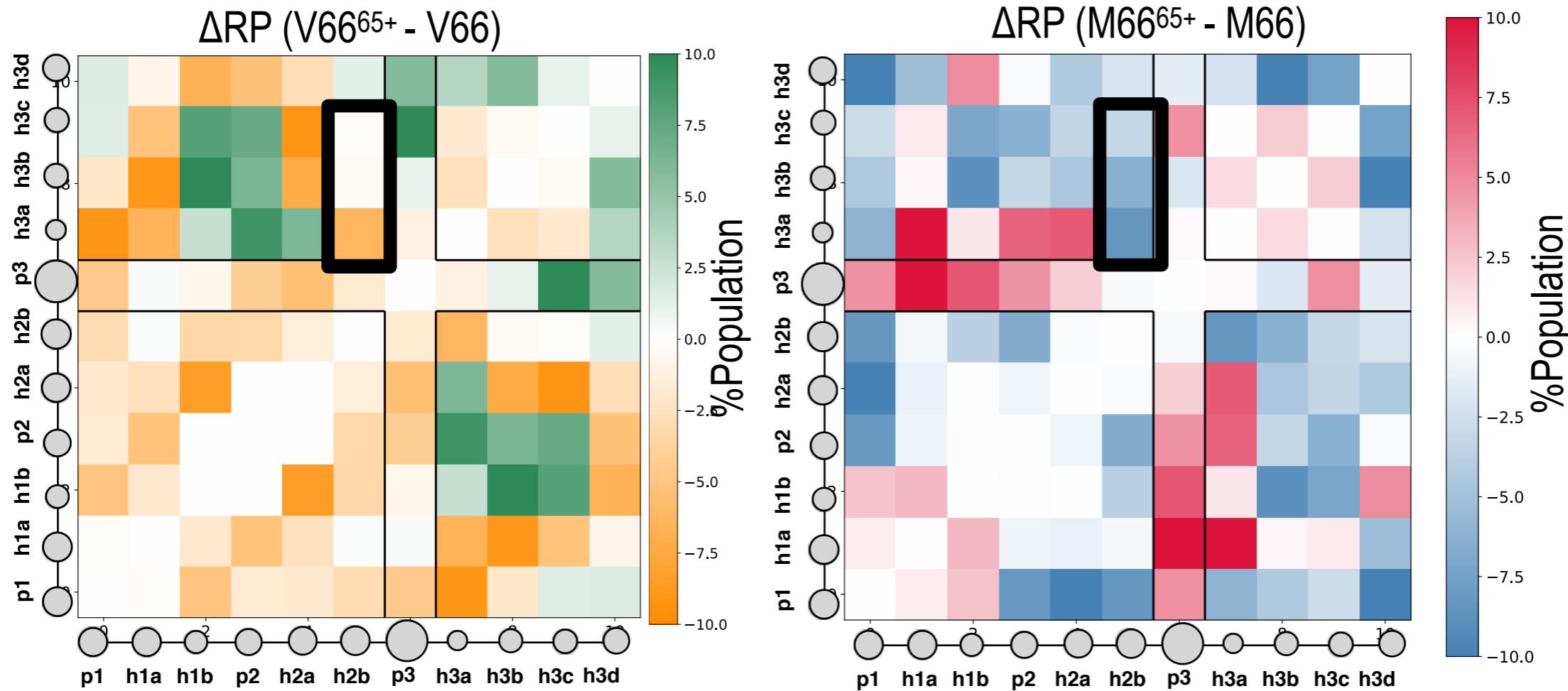
#3b Does our hypothesis hold true for other charged residue mutations in prodomain BDNF ?

- hypothesis: reduce interactions between h2b:h3b, and thus reduce cross-linker interactions
- prediction: reducing the charge on h2b will decrease the interaction between h2b:h3b

By net charge:



Effect of changing charge on interblob contacts



$V66$ $V66^{65+}$

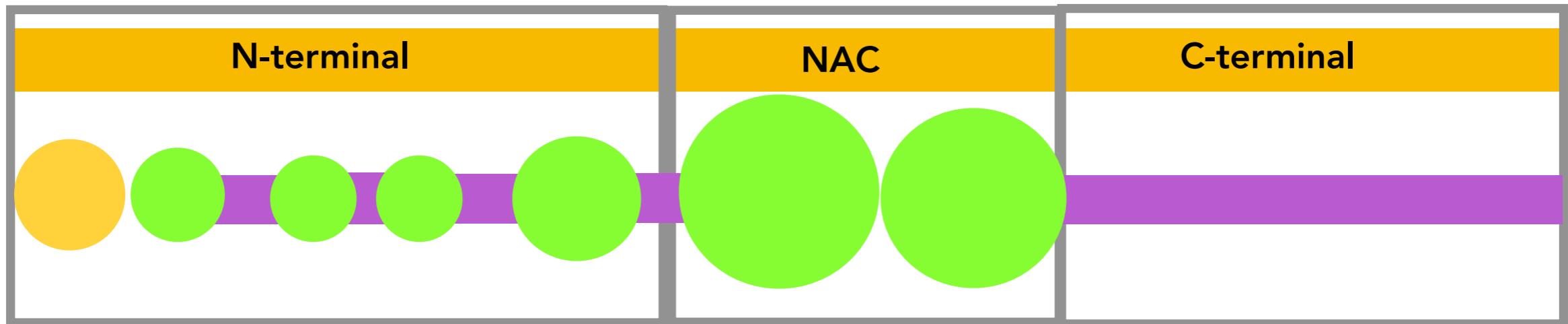
$M66$ $M66^{65+}$

- Our prediction holds

Blob identification might be meaningful for other IDPs

- Blob identifications works for BDNF prodomain to identify contacts.
- Blob identification may be useful for other disorder proteins.

Example: α -synuclein



- NAC domain and C-terminal are separated by a long polyampholyte region
- Interactions between the C-terminal domain and the NAC region of α -syn is thought to be responsible for inhibition of α -syn aggregation

Emamzadeh FN. J Res Med Sci. 2016

high hydrophobicity



low hydrophobicity



Visualizing blob in other proteins with the web tool

[Home](#) [About](#) [Analysis](#) [Brannigan Lab](#)

Uniprot ID:

P37840

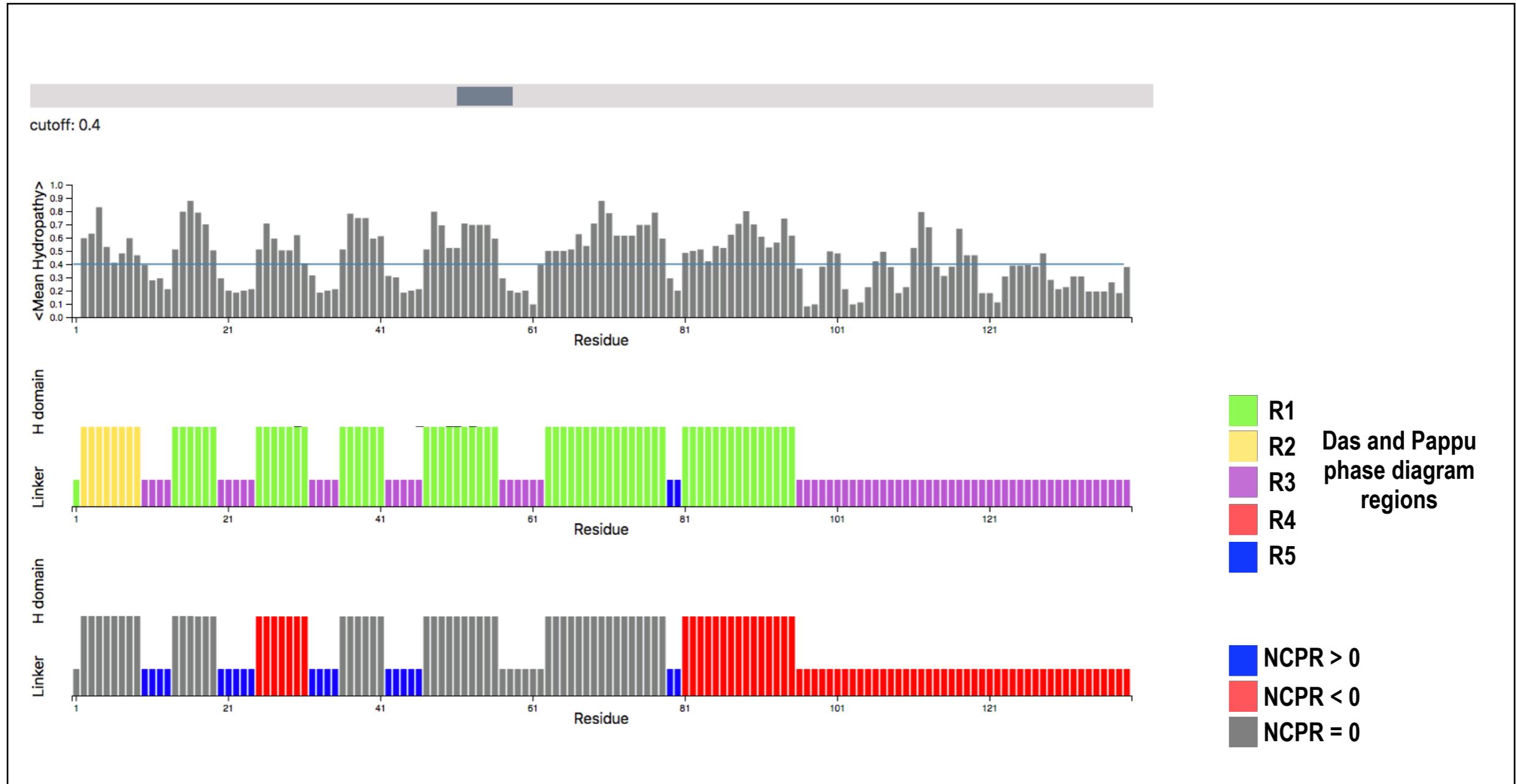
Cutoff:

0

Compute

Visualizing blob in other proteins with the web tool

Alpha-synuclein

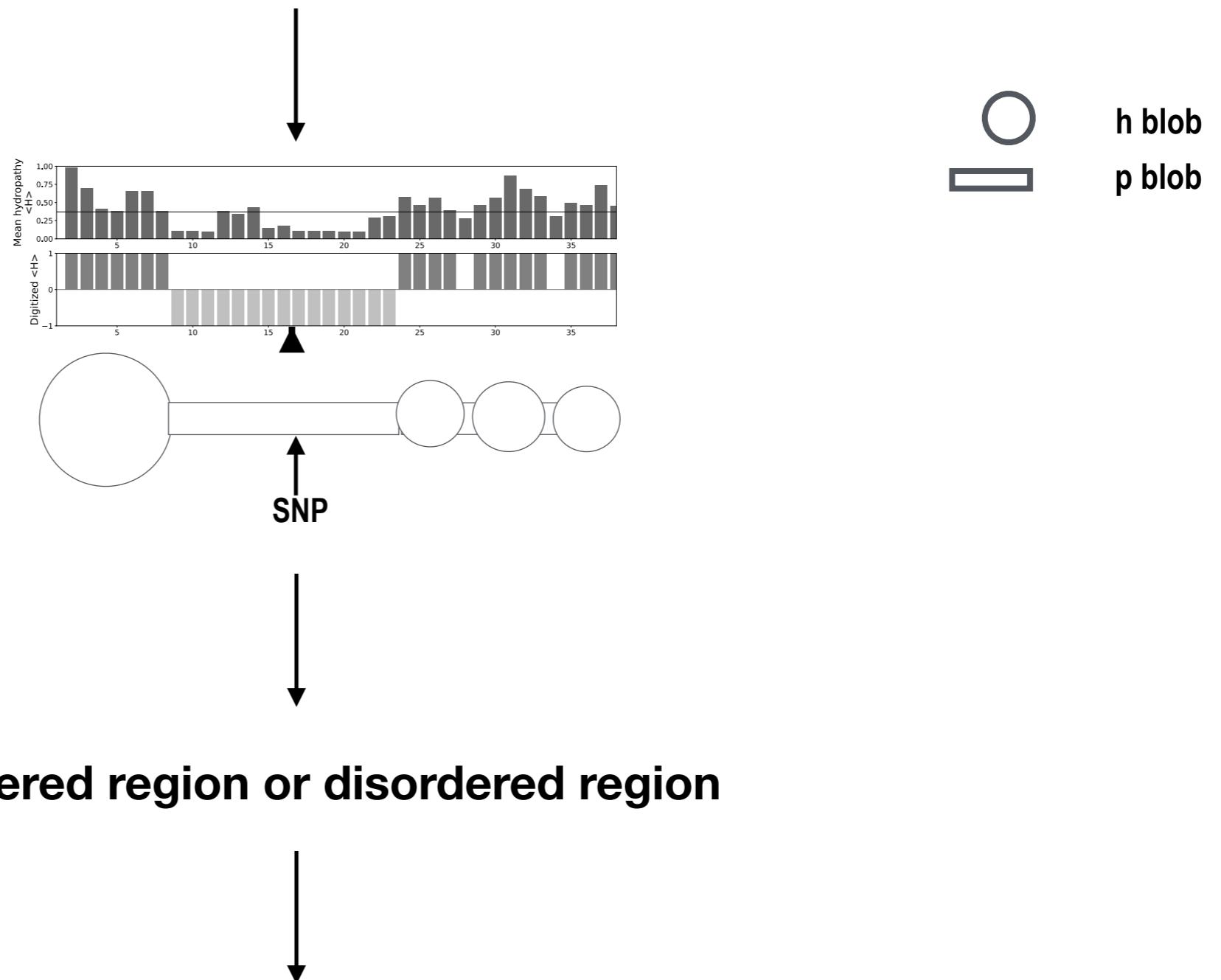


#4: Sequence-decomposition applied to a database of Mendelian SNPs

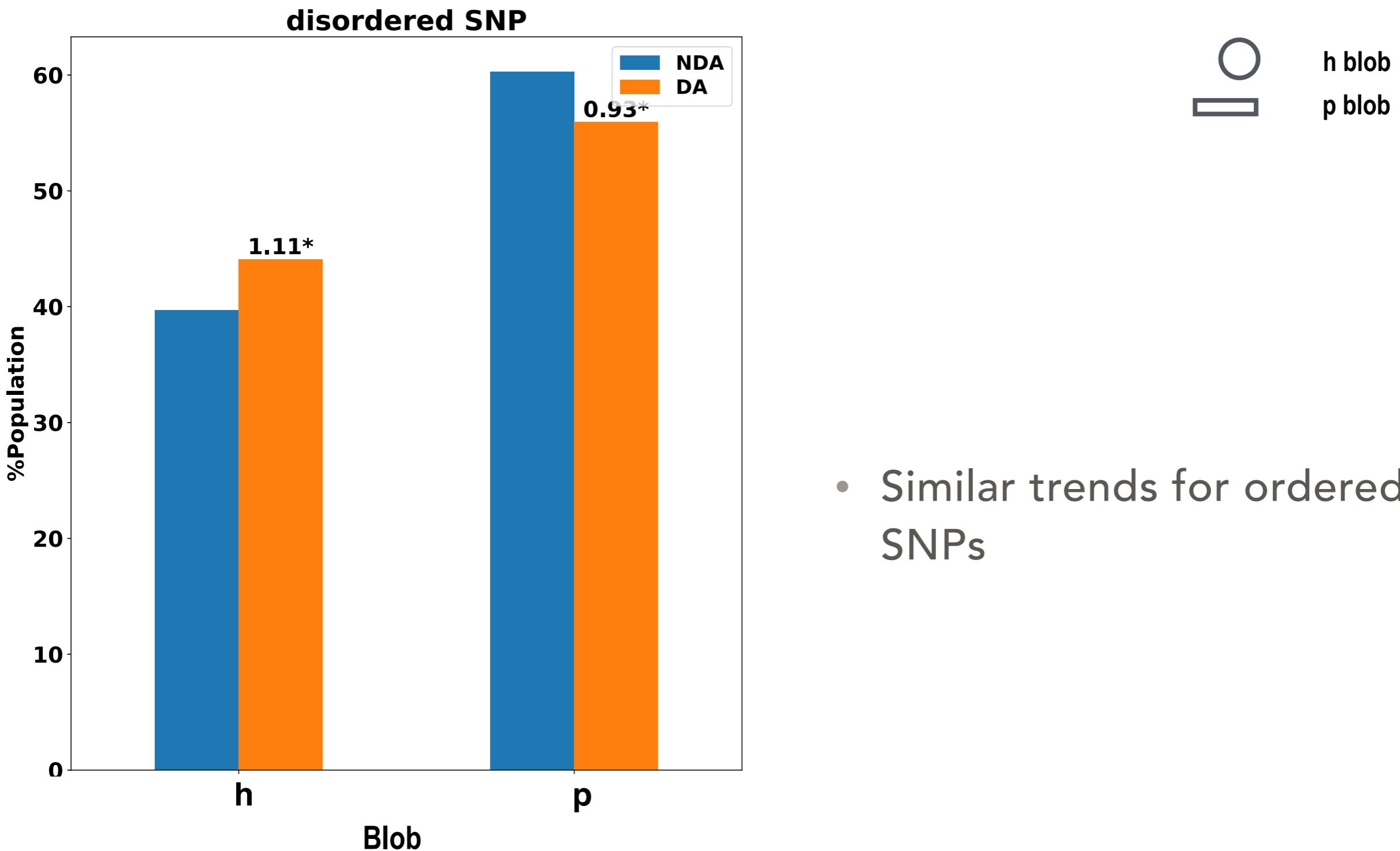
- Analyzed all the SNPs reported in Uniprot database to test the domains they occur in and their likelihood of being associated with diseases

Workflow

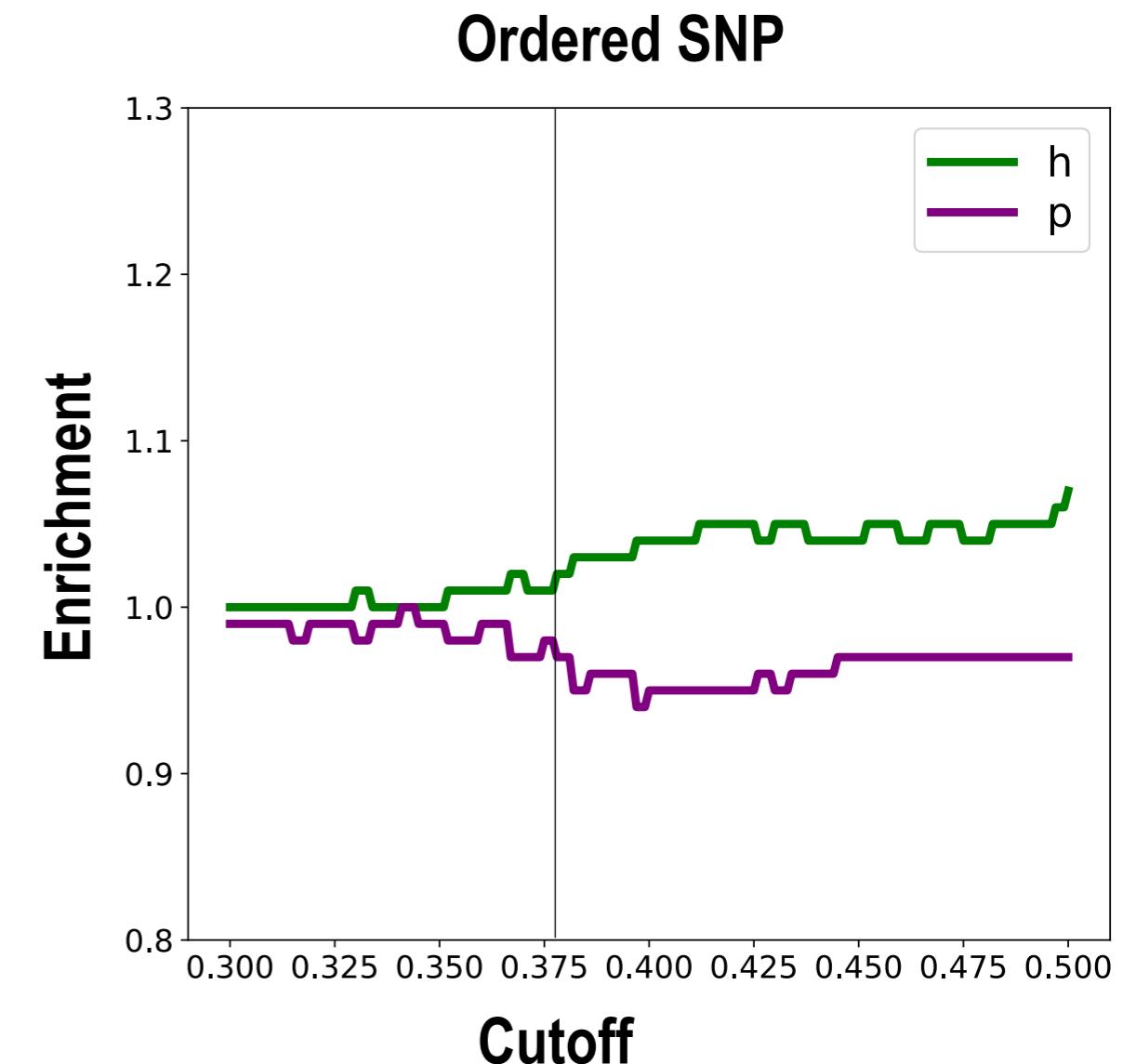
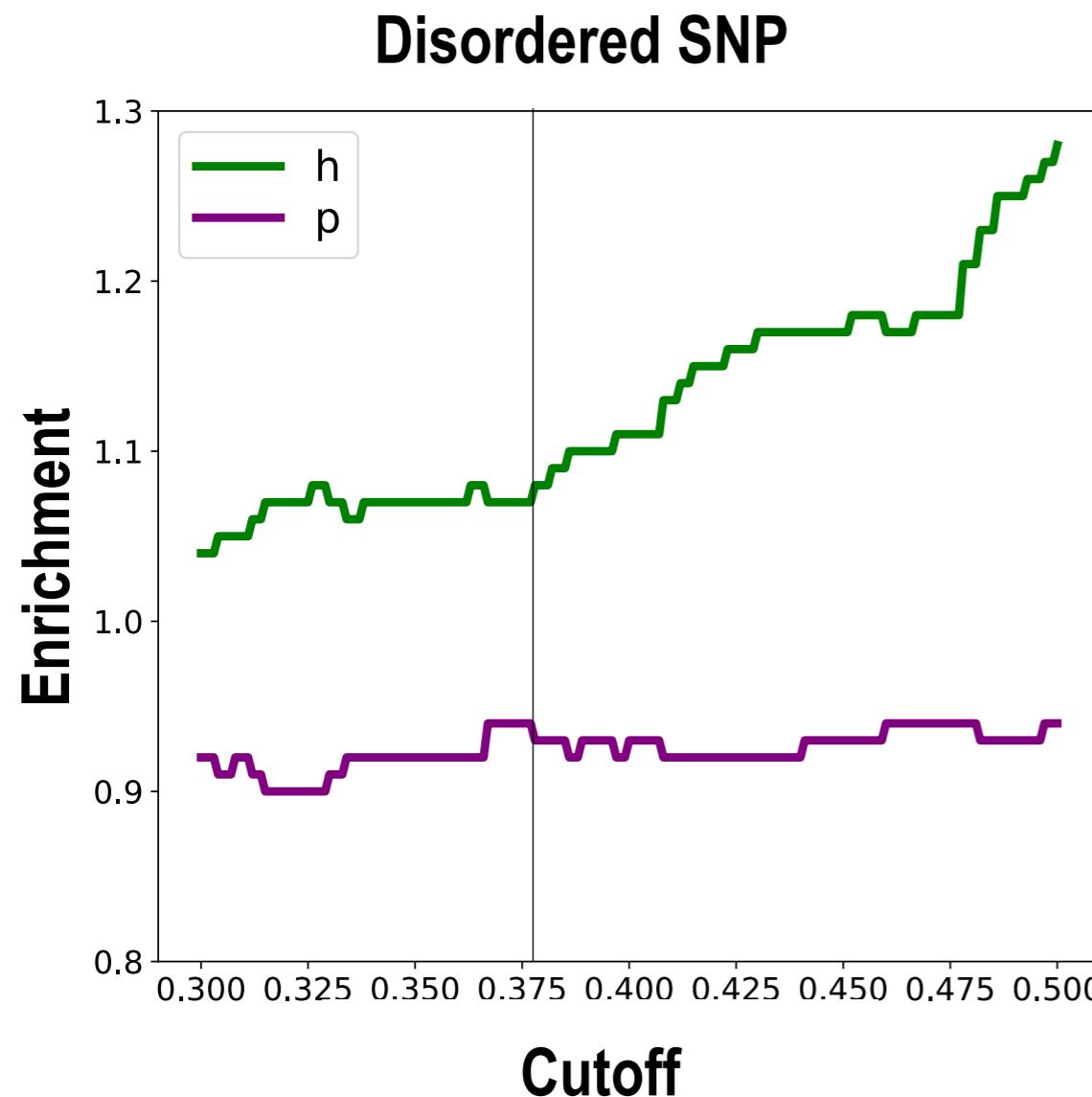
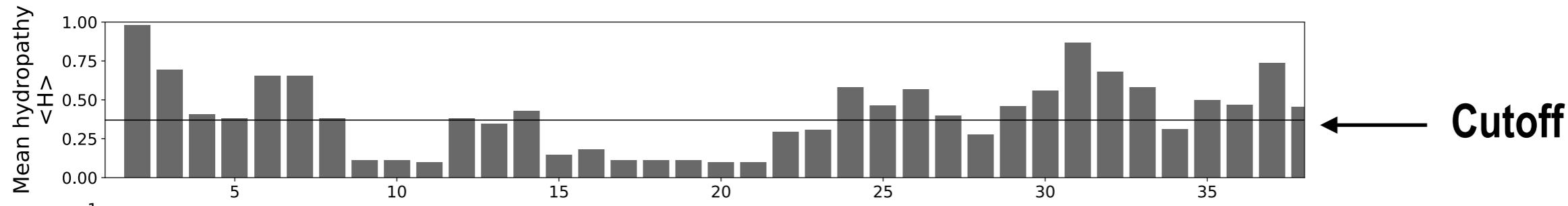
70680 SNPs (Uniprot database)
Disease-associated SNP or Non disease-associated SNP



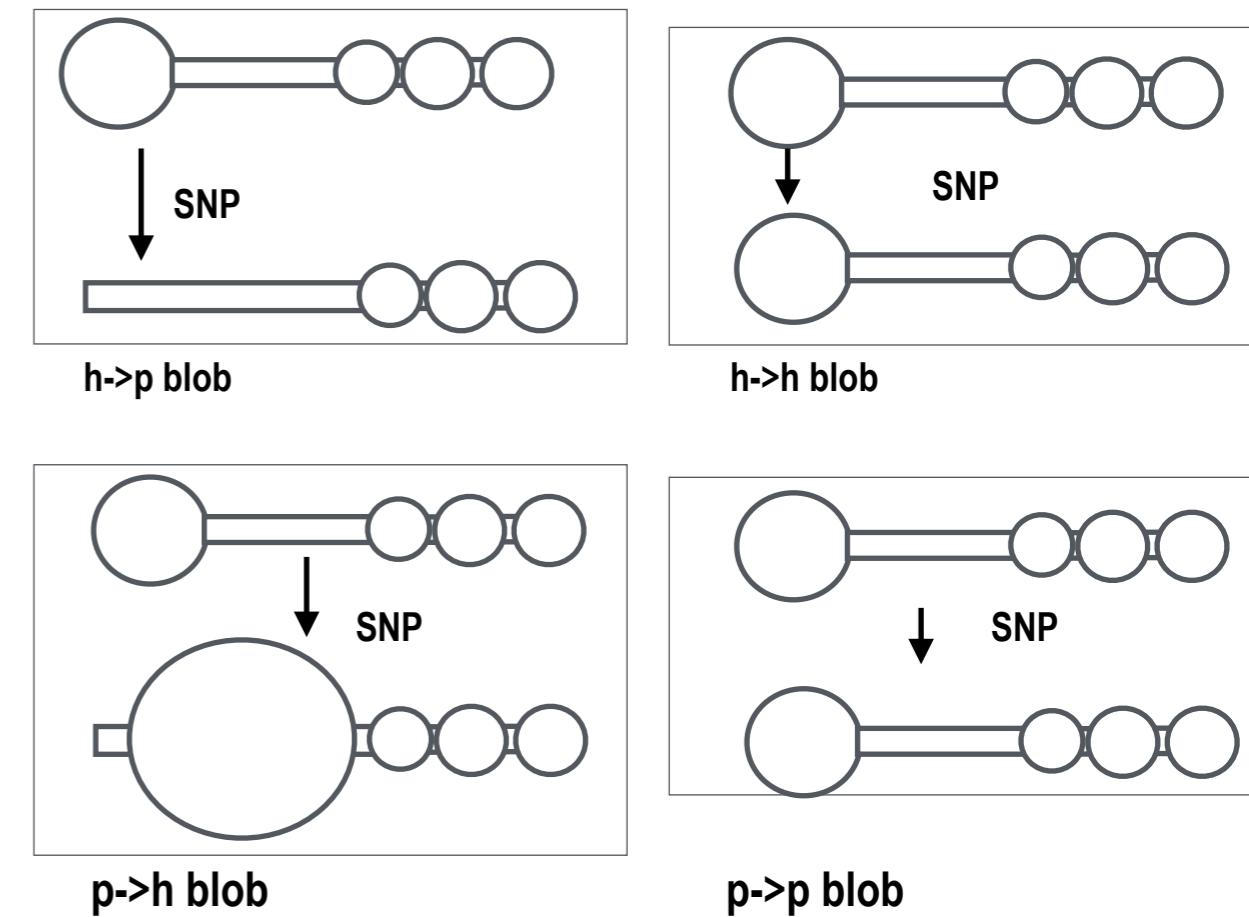
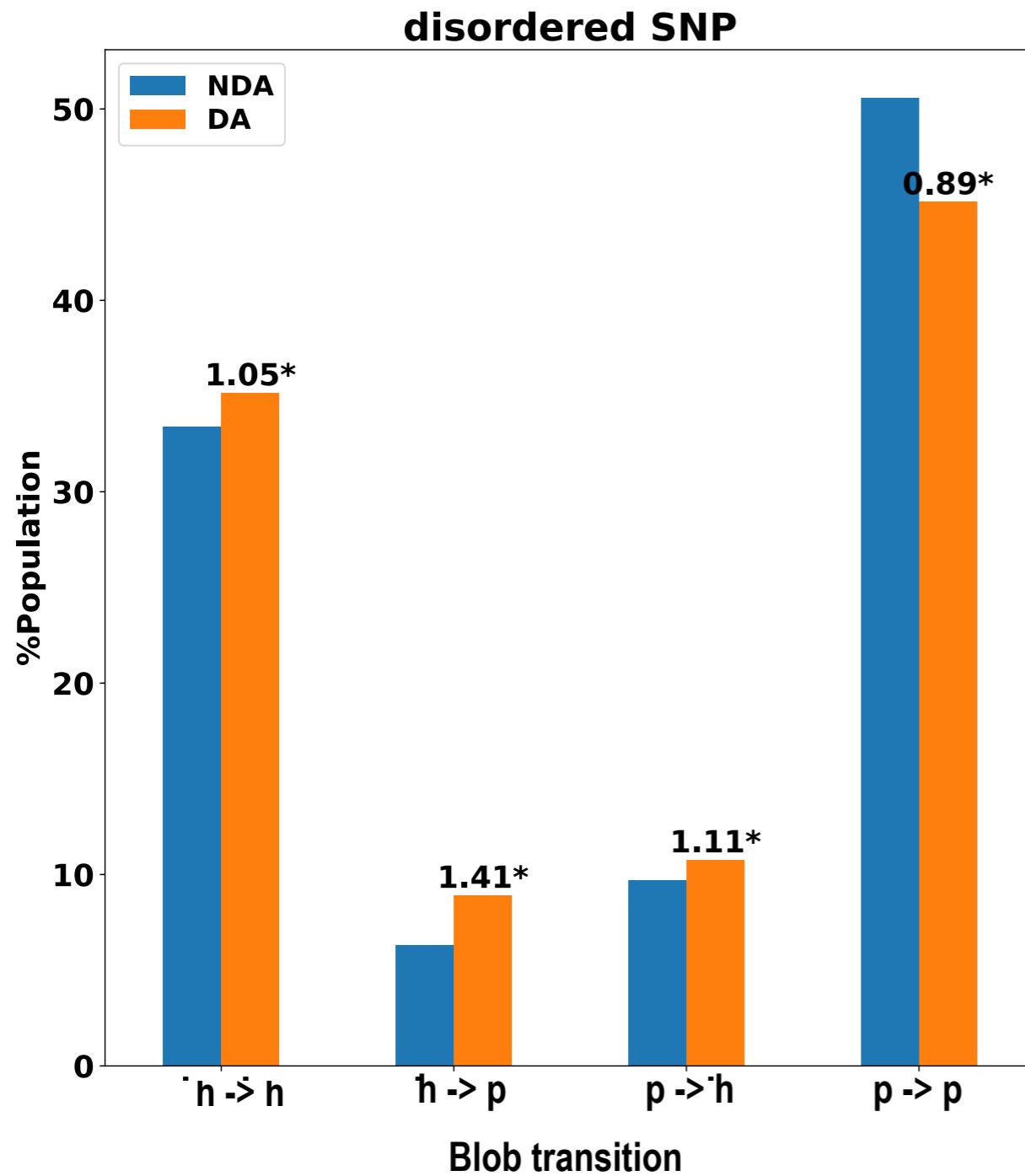
Disease associated SNP are enriched in h blobs?



Role of cutoff in blob identification



Disease associated SNP are enriched in switches between blob types



- Similar trends for ordered SNPs

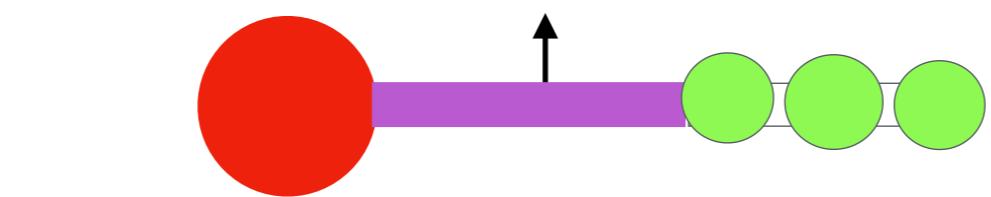
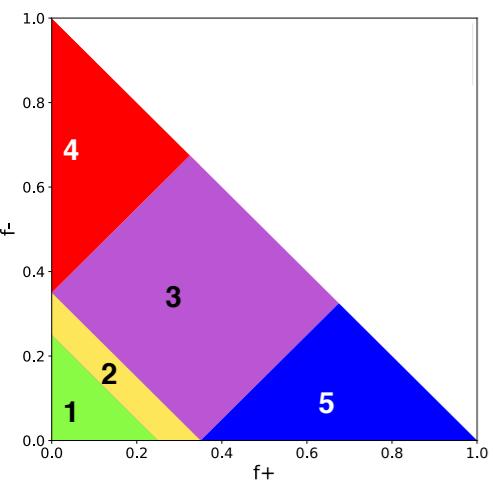


Workflow

70680 SNPs (Uniprot database)

Disease-associated SNP or Non disease-associated SNP

- 1: Weak polyampholyte and electrolyte
- 2: Janus sequences
- 3: Strong polyampholytes
- 4: Negatively charged strong polyelectrolyte
- 5: Positively charged strong polyelectrolyte



Ordered region or disordered region

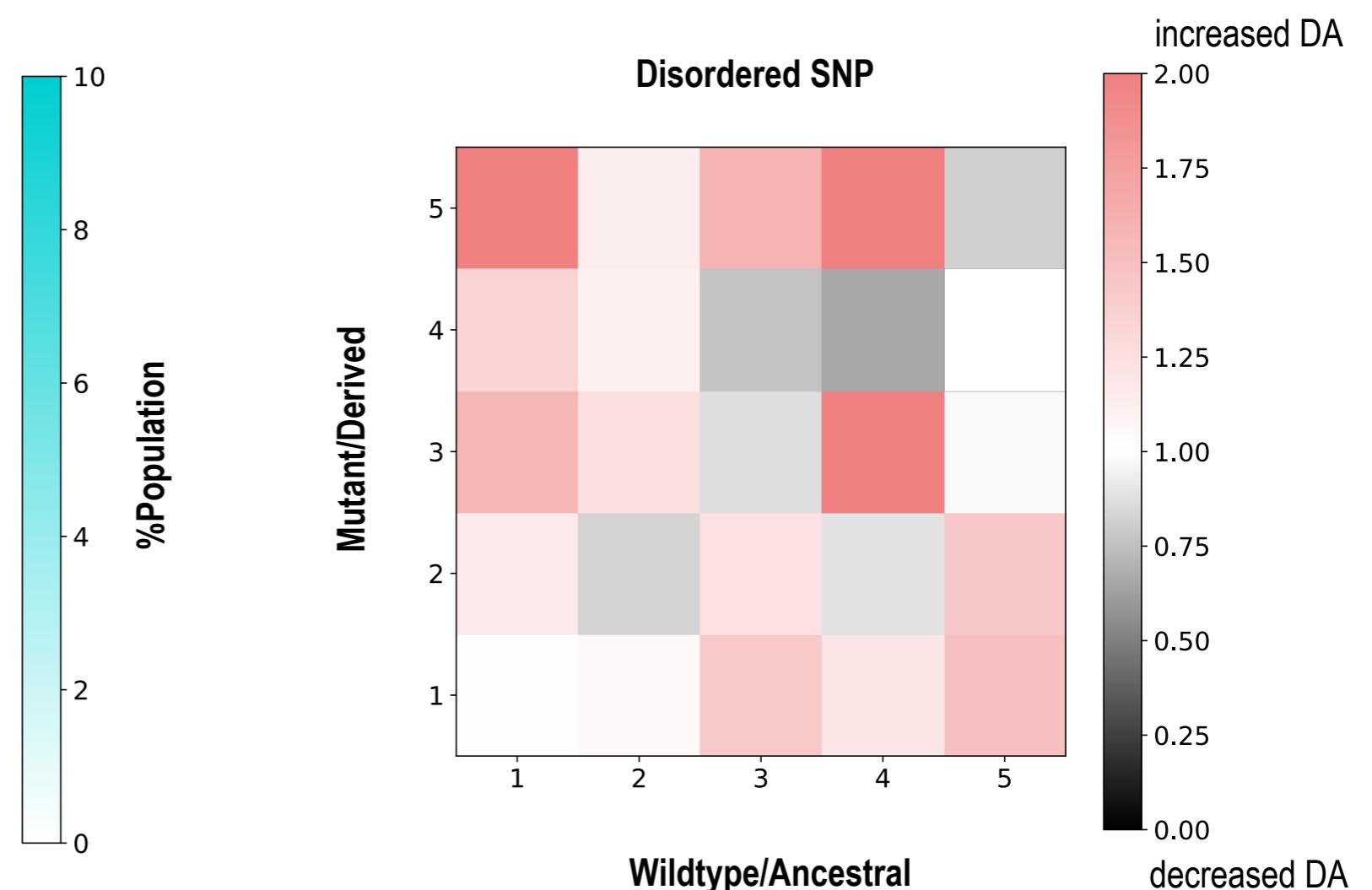
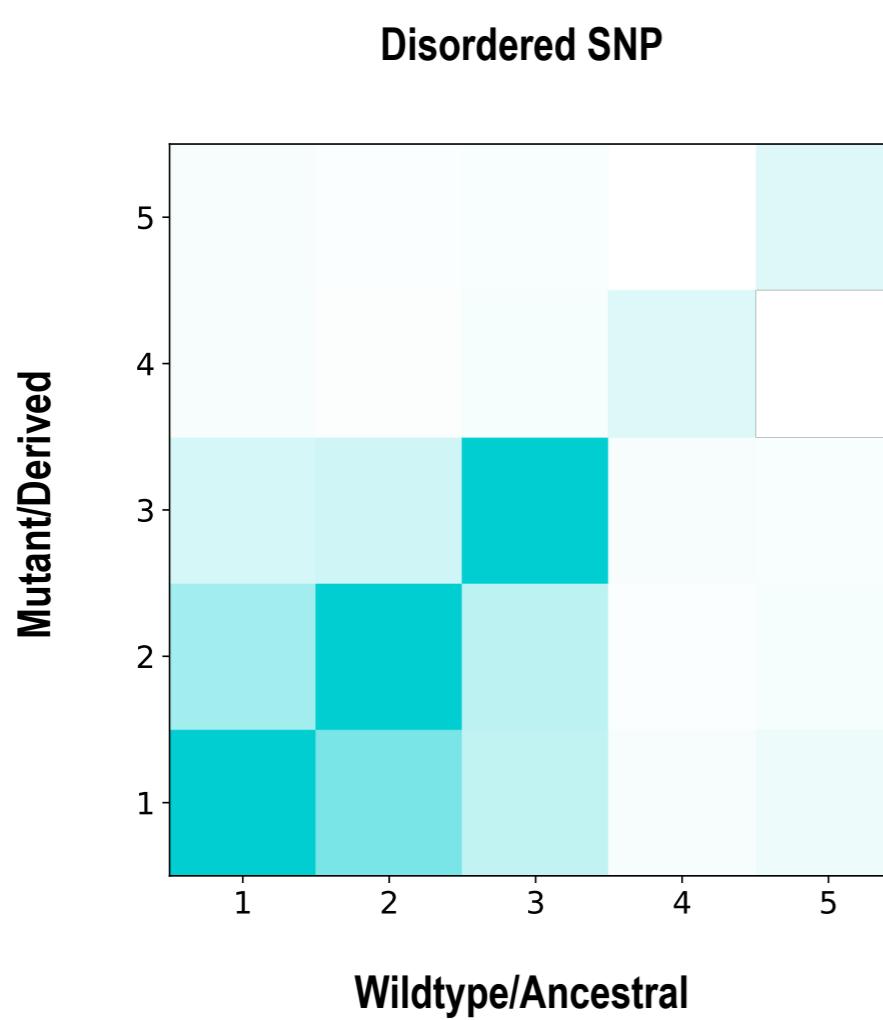
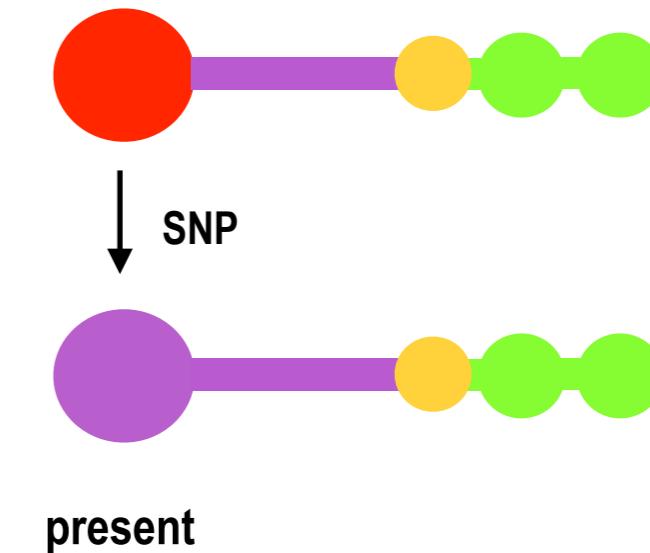
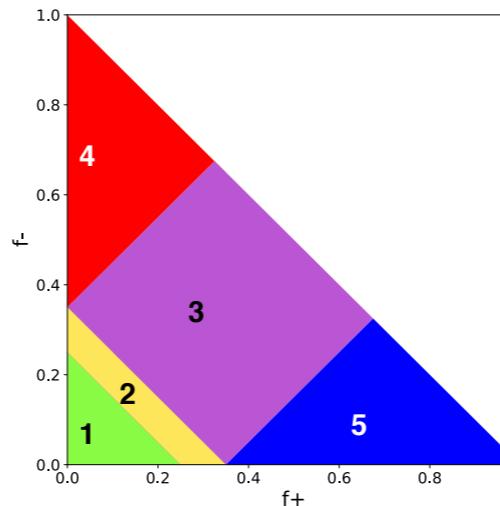


Enrichment in disease-associated SNP relative to Non disease-associated SNP

h blob
p blob

Disease associated SNP are enriched in predicted phase changes

- 1: Weak polyampholyte and electrolyte
- 2: Janus sequences
- 3: Strong polyampholytes
- 4: Negatively charged strong polyelectrolyte
- 5: Positively charged strong polyelectrolyte



Summary

- Disease associated SNPs are likely to be found in hydrophobic blobs of disordered proteins.
- Disease associated SNPs are likely to cause blob transitions and phase behavior transitions.

Overall Summary

1. **Sequence decomposition** helped us identify residue-residue interactions that cause conformational effects of Val66Met SNP for proBDNF
2. **Intraprotein interaction** effects point to **interprotein interaction** effects that **could** be important
3. Understanding Val66Met effects with sequence decomposition mechanism **allowed us to make several a priori predictions** for effects of different mutations, all of which were validated.
4. **Blob-decomposition** approach seems to holds promise for understanding of **disordered proteins in general**, and we're just beginning to apply this

Acknowledgements

Dr. Grace Brannigan

Dr. Matt Hansen (University of Pennsylvania)

Brannigan research group:

Dr. Reza Salari
Dr. Tom Joseph
Dr. Sruthi Murlidaran
Liam Sharp
Kristen Woods
Rulong Ma
Shashank Chavali

Family & Friends

Committee members:

Dr. Cameron Abrams (Drexel University)

Dr. Jinglin Fu
Dr. Eric Klein

Tool development:

Kaitlin Bassi

Dan Russo

Dr. Sunil Shende
Chris Till
David Bushta



Caliburn (Rutgers Discovery Informatics Institute) (45 million SUs)

Rutgers Office of Advanced Research Computing

XSEDE

Thank you for listening!