Structural Impact of Missense Mutation in Ubiquitin-Associated Domain of Human Rad23 Protein: Molecular Dynamics Simulations Using Different Force-Fields

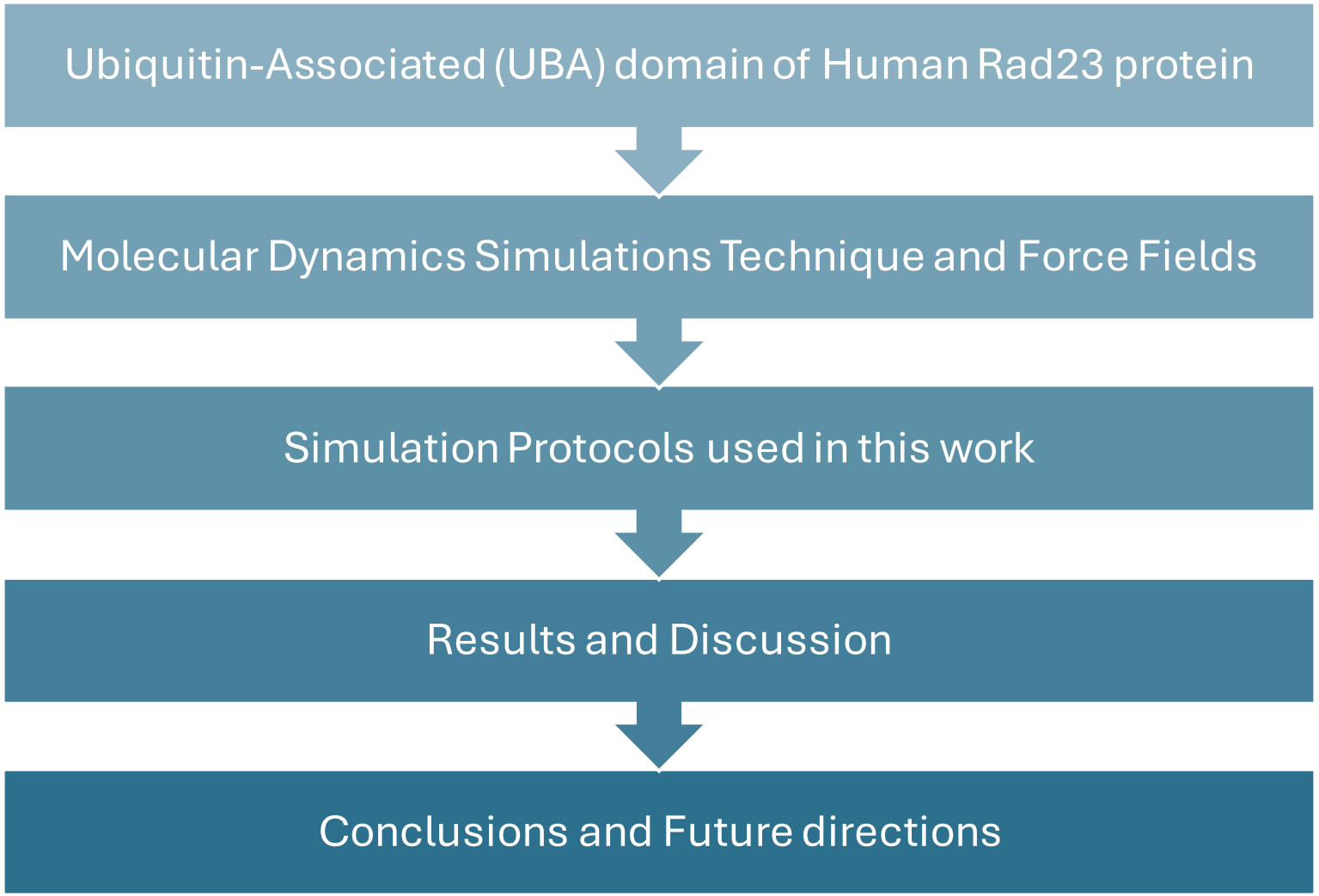
MTECH THESIS DEFENSE

DEPT OF BIOLOGICAL SCIENCES AND BIOENGINEERING

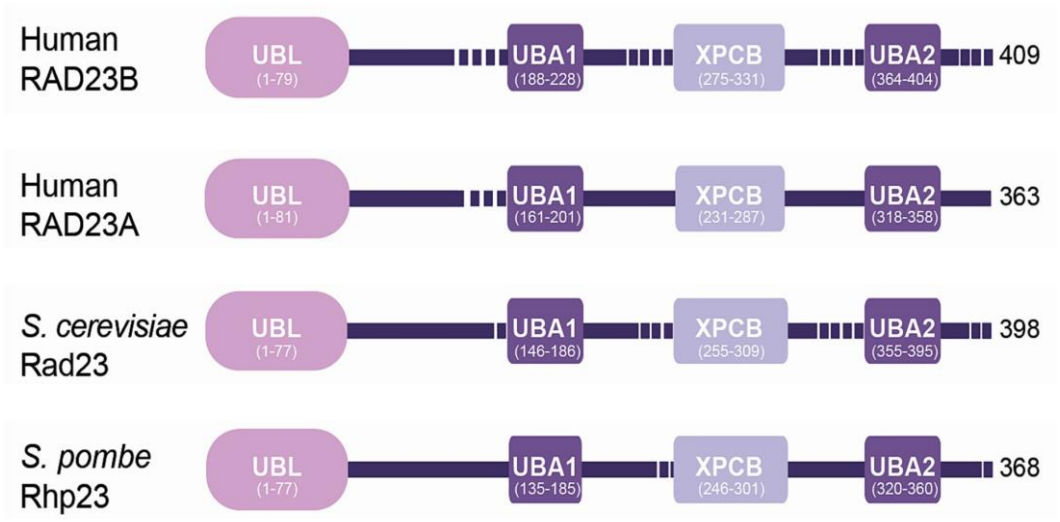
Vaibhav Anand (18807843)

Supervised by – Prof. R. Sankararamakrishnan

Outline

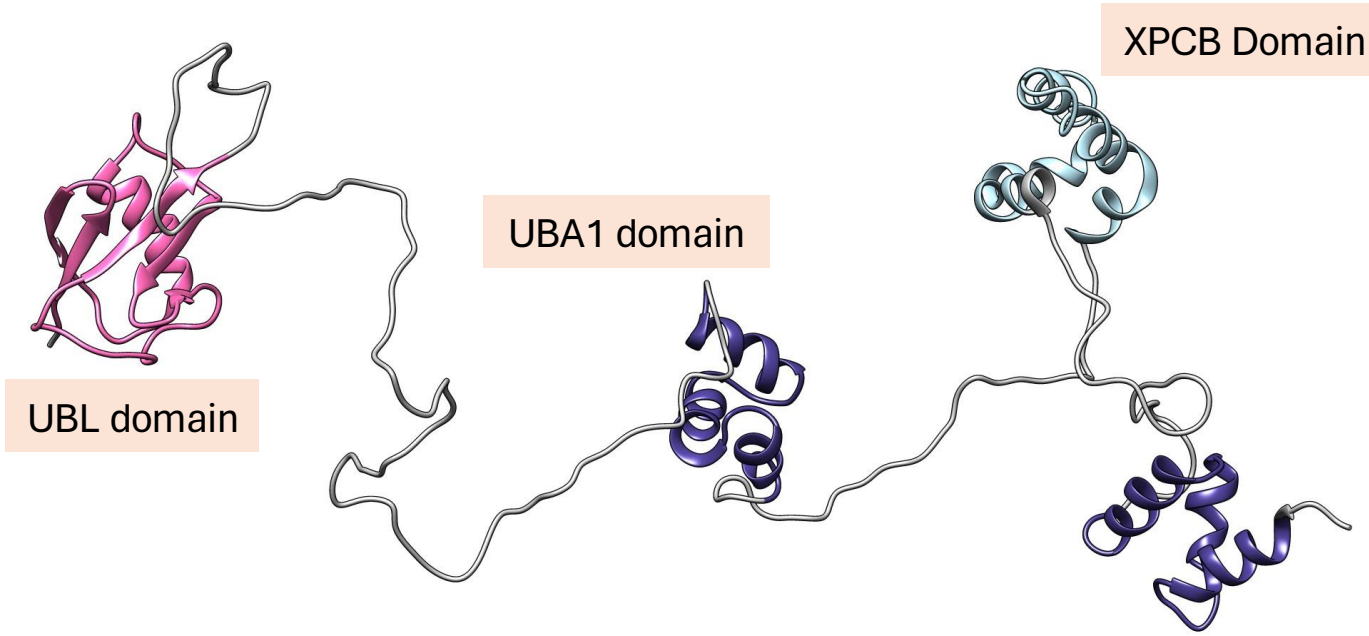

Ubiquitin-Associated (UBA) domain of Human Rad23 protein
Molecular Dynamics Simulations Technique and Force Fields
Simulation Protocols used in this work
Results and Discussion
Conclusions and Future directions

Human Rad23 protein and its Ubiquitin-Associated (UBA) domains



Domain organization of Rad23 orthologs [1]

UBL domain – Ubiquitin-like domain
UBA domain – Ubiquitin-associated domain
XPCB - xeroderma pigmentosum complementation group C binding domain

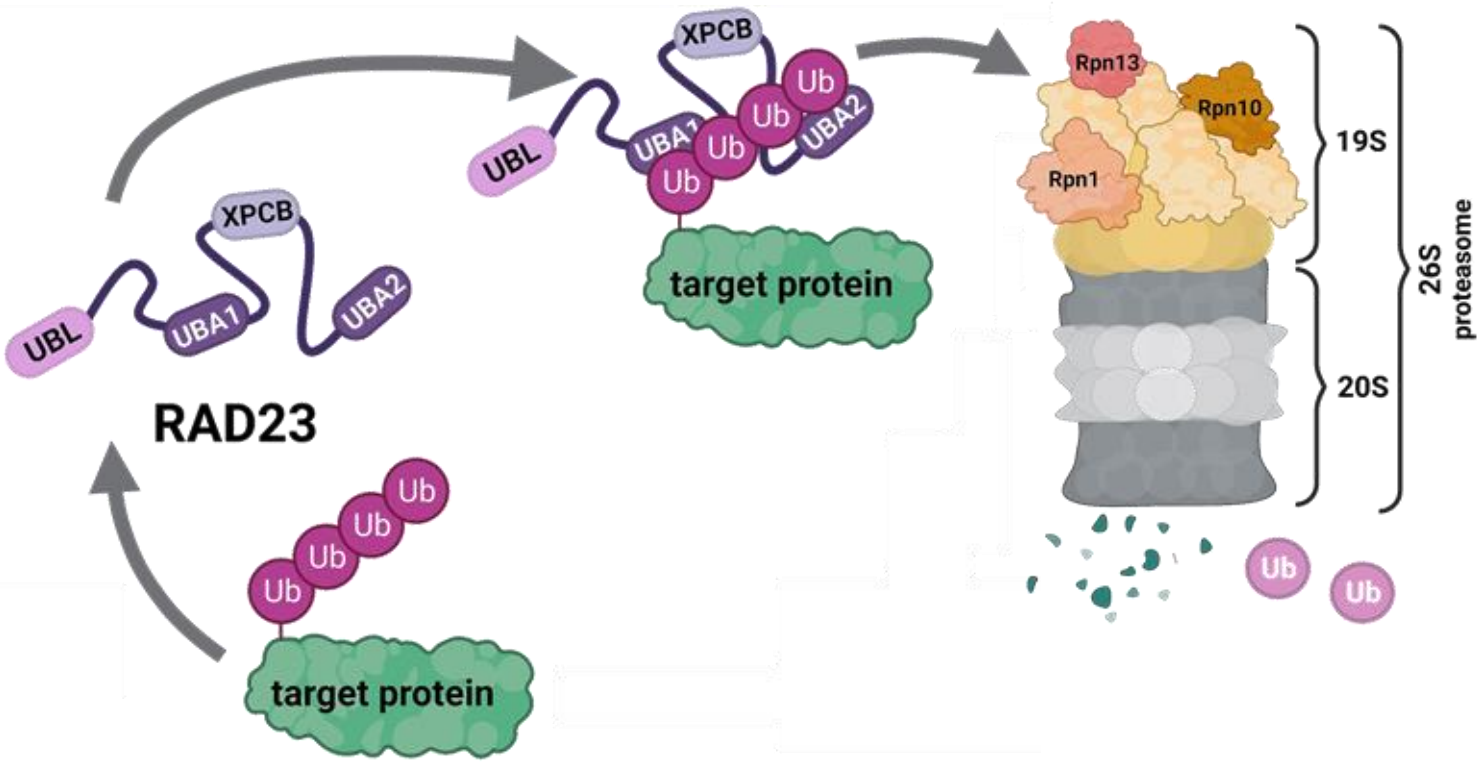


Structure of hHR23a (human RAD23A) (PDB ID: 1OQY) [2]

Rad23 proteins are involved in Nucleotide-excision repair and Ubiquitin-mediated proteolysis

1. Grønbæk-Thygesen M, Kampmeyer C, Hofmann K, Hartmann-Petersen R. *Biochim Biophys Acta Gene Regul Mech*. 2023
2. Walters KJ, Lech PJ, Goh AM, Wang Q, Howley PM. *Proc Natl Acad Sci U S A*. 2003

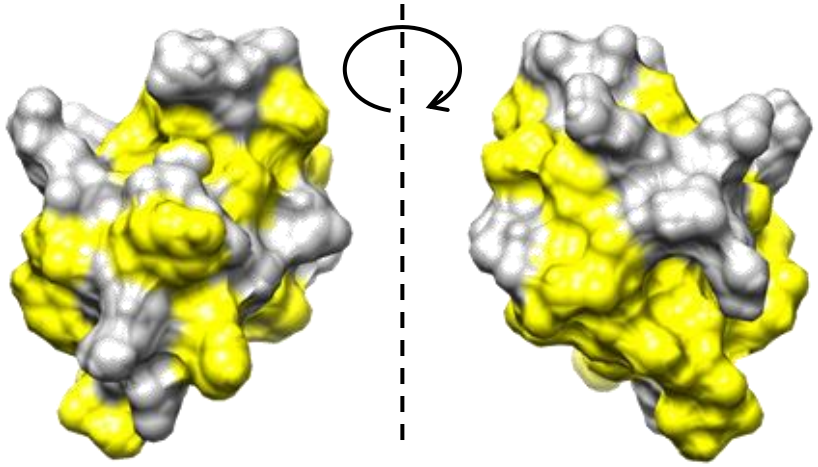
UBA domains function in Ubiquitin-mediated Proteolysis



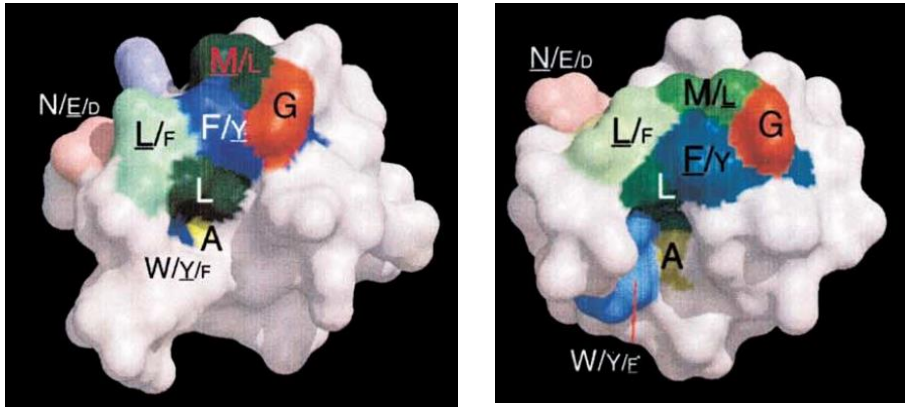
Rad23 as shuttle to direct ubiquitylated substrate to proteosome [1]

1. Grønbæk-Thygesen M, Kampmeyer C, Hofmann K, Hartmann-Petersen R. *Biochim Biophys Acta Gene Regul Mech*. 2023

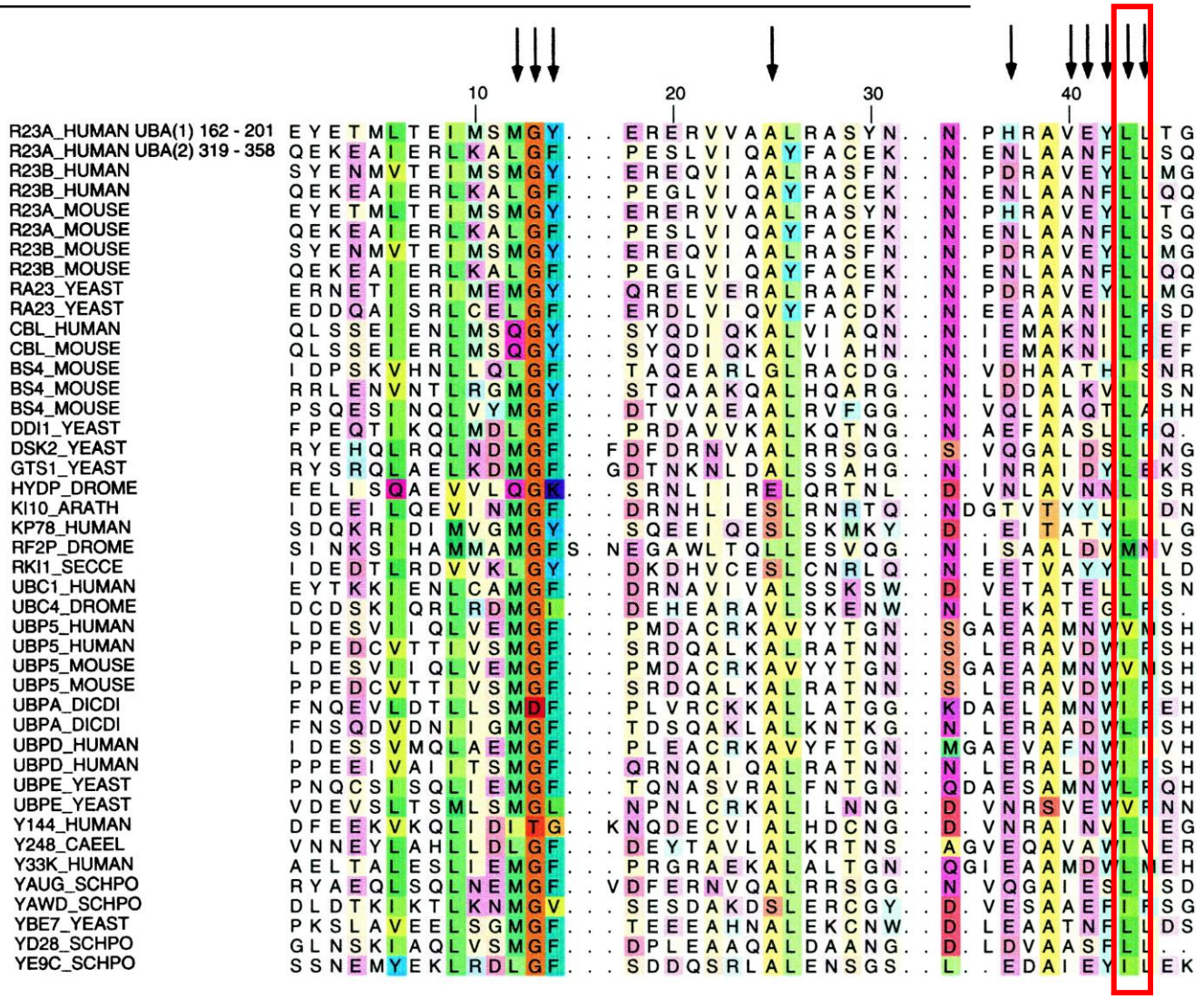
Conserved hydrophobic surface on UBA domains



Large Hydrophobic patches (yellow) on UBA1 domain of hHR23a (PDB ID: 1OQY 161-200 residues)



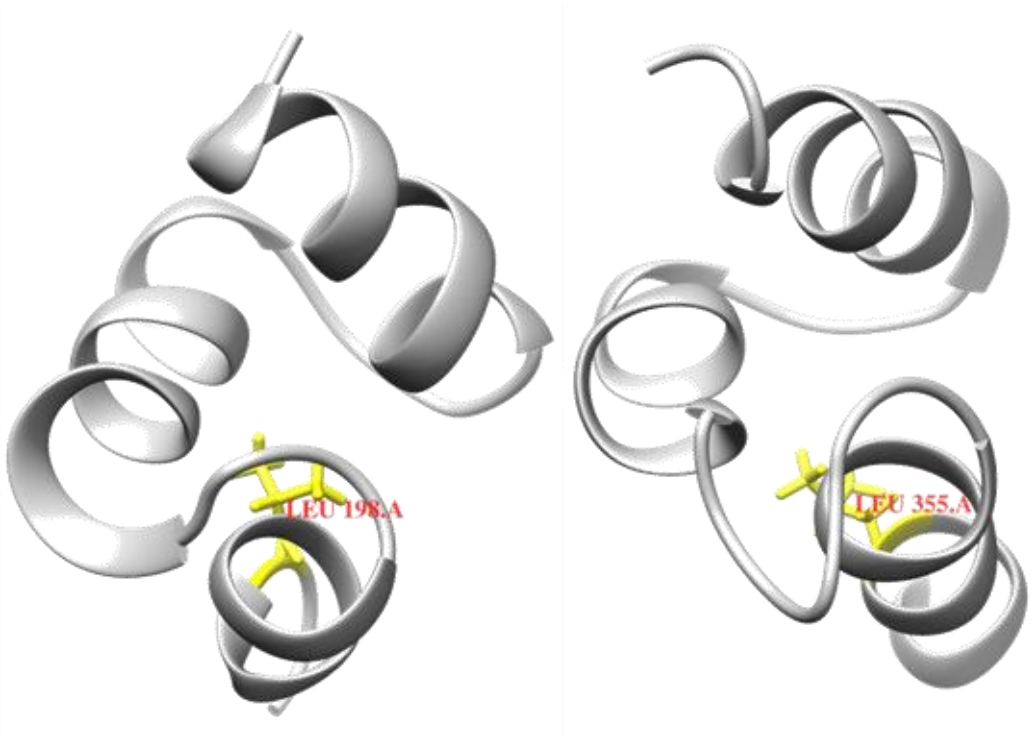
Surface representation of conserved residues (shown by pointed arrow) for UBA1 (left) and UBA2 (right) domains [1]



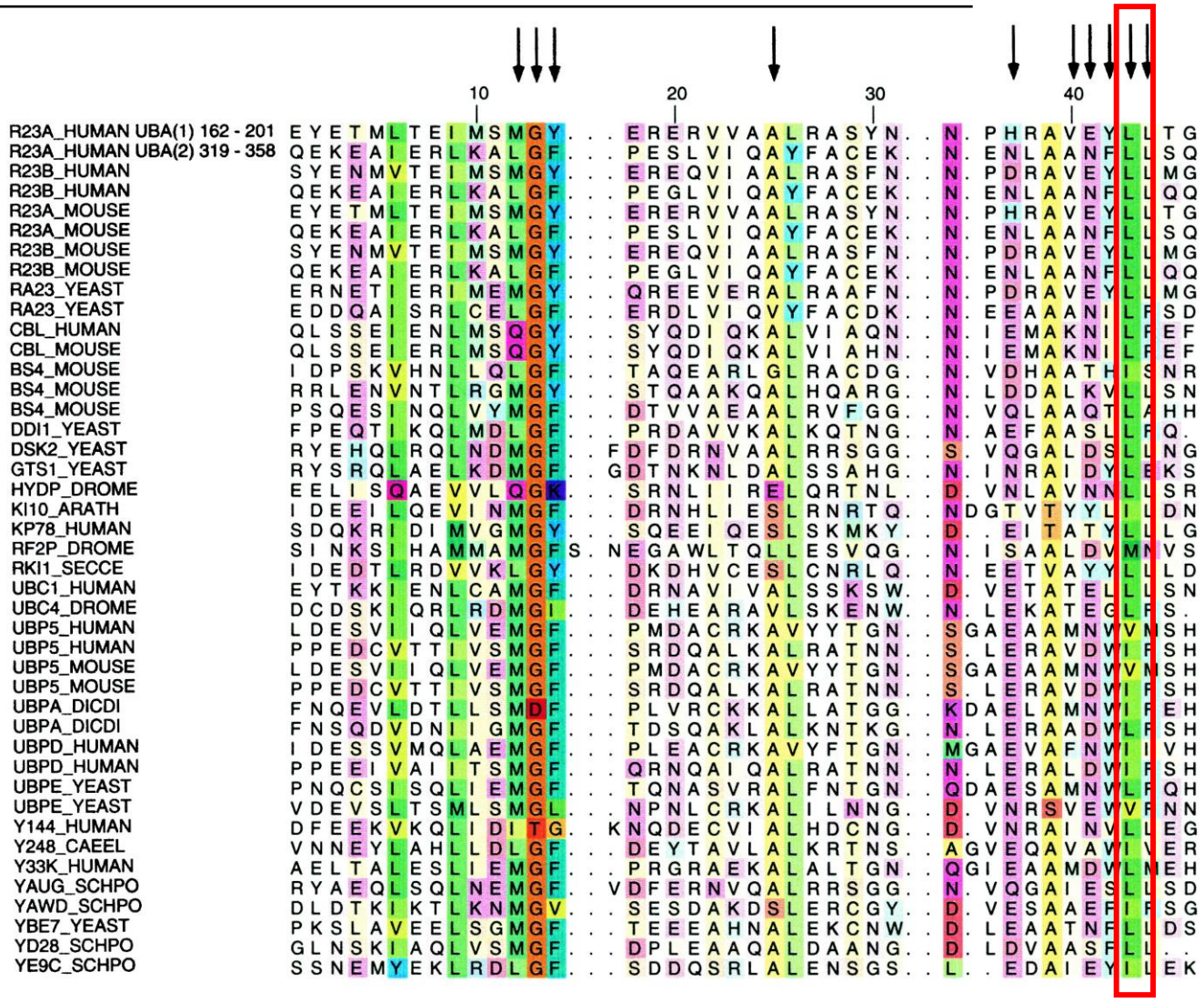
Sequence alignment UBA domains (from Pfam database) [1]

1. Mueller TD, Feigon J. *J Mol Biol*. 2002;319(5):1243-1255. doi:10.1016/S0022-2836(02)00302-9

Role of highly conserved Leucine residue in Structural Integrity



Inward-facing sidechain of Leu198 and Leu355 residues in UBA1 (left), UBA2 (right) domains of hHR23a, respectively.

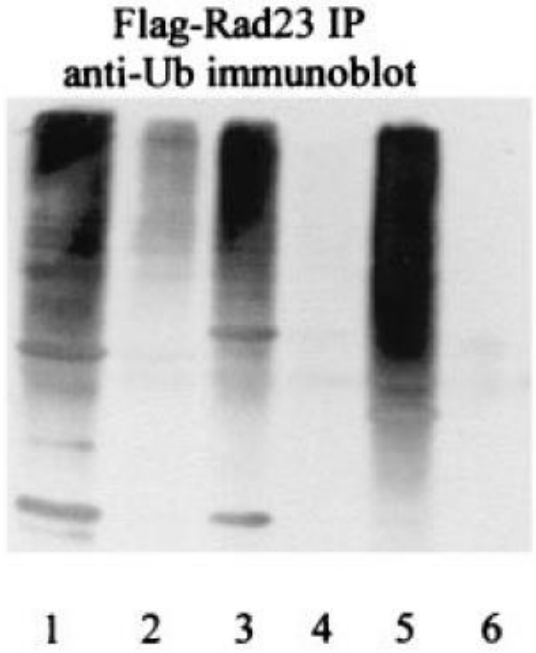


Sequence alignment UBA domains (from Pfam database) [1]

1. Mueller TD, Feigon J. *J Mol Biol*. 2002;319(5):1243-1255. doi:10.1016/S0022-2836(02)00302-9

Alanine substitution of Leucine residue disrupt the UBA domain fold

In yeast homolog of Rad23 –



- Rad23 became **deficient** in ubiquitin binding upon L183A mutation in UBA1 domain
- Rad23 showed **reduced** binding activity on L392A mutation in UBA2 domain

Lane1 – Flag-Rad23; **Lane2**- Flag-Rad23^{UBA1}(L183A); **Lane3-** Flag-Rad23^{UBA2}(L392A); **Lane4** – Flag-Rad23^{UBA1UBA2}(L183A and L392A); **Lane5** – Flag-Rad23$^{\Delta UBL}$(Deletion of UBL domain); **Lane6** – Flag-Rad23$^{\Delta UBA1}$(deletion of UBA1 domain) [1]

In human homolog of Rad23 –

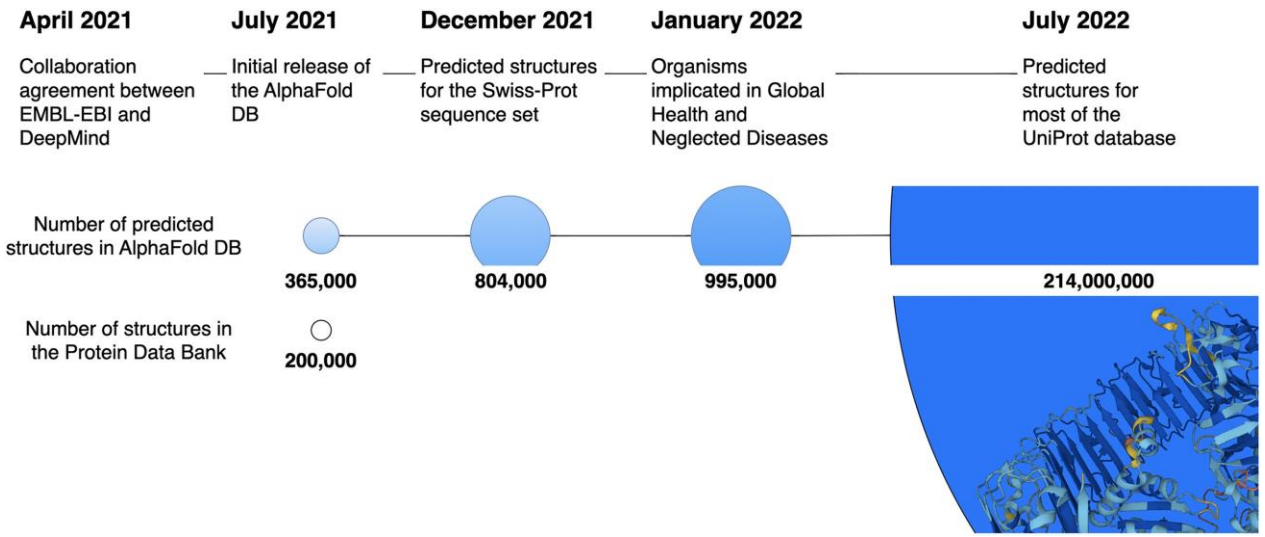
Single amino acid substitutions L198A and L335A in UBA1 and UBA2 domains of hHR23a, respectively –

- Both mutant UBA domain constructs – deficient in Ubiquitin binding
- And became disordered – loss chemical shift dispersion – a characteristic of random-coil peptides [2]

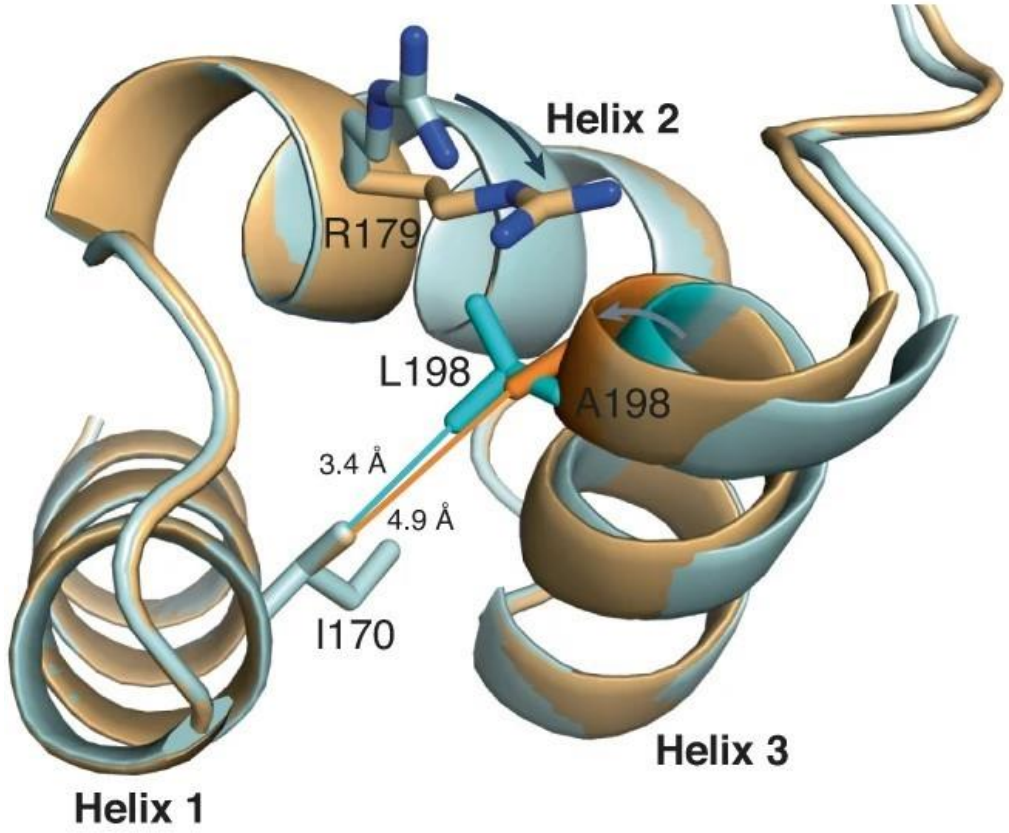
1. Chen L, Madura K. *Mol Cell Biol*. 2002;22(13):4902-4913. doi:10.1128/MCB.22.13.4902-4913.2002
2. Wang Q, Goh AM, Howley PM, Walters KJ. *Biochemistry*. 2003

AlphaFold2 Prediction for L198A UBA1 domain

AlphaFold is a state-of-the-art **machine-learning-based method** that uses physical and biological information, including multiple-sequence alignments, to predict the three-dimensional structure of proteins [1]



The growth of AlphaFold DB [2]

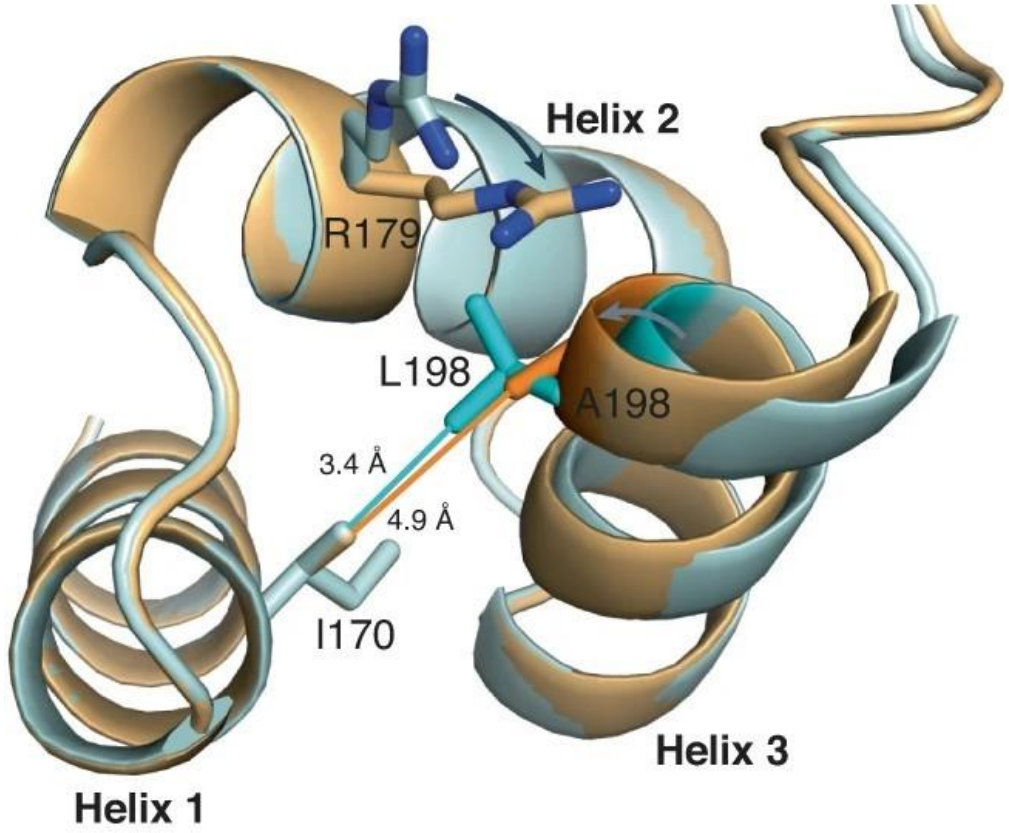


Superimposed structures of AF2 predicted UBA1 wild-type (shown in light blue) and mutated UBA1 (L198A) (shown in orange) domains. **RMSD (C-alpha atoms) – 0.1Å** [3]

1. Jumper, J., Evans, R., Pritzel, A. *et al. Nature* (2021)
2. Varadi M, Bertoni D, Magana P, et al. *Nucleic Acids Res*. 2024
3. Buel, G.R., Walters, K.J. *Nat Struct Mol Biol* (2022)
4. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Research, 2000*

AlphaFold2 Prediction for L198A UBA1 domain

- Highlighted the **deficiency** of AlphaFold in predicting the structural impact of missense mutations
- Because – AlphaFold mainly uses the structural data of proteins available in the Protein Data Bank (PDB) [4]

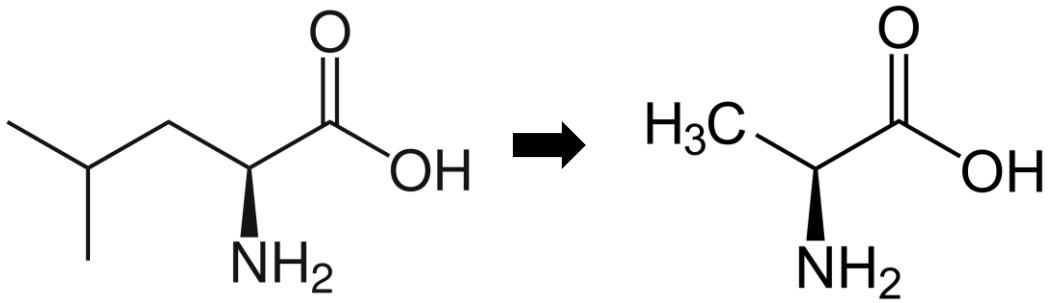


Superimposed structures of AF2 predicted UBA1 wild-type (shown in light blue) and mutated UBA1 (L198A) (shown in orange) domains. **RMSD (C-alpha atoms) – 0.1Å** [3]

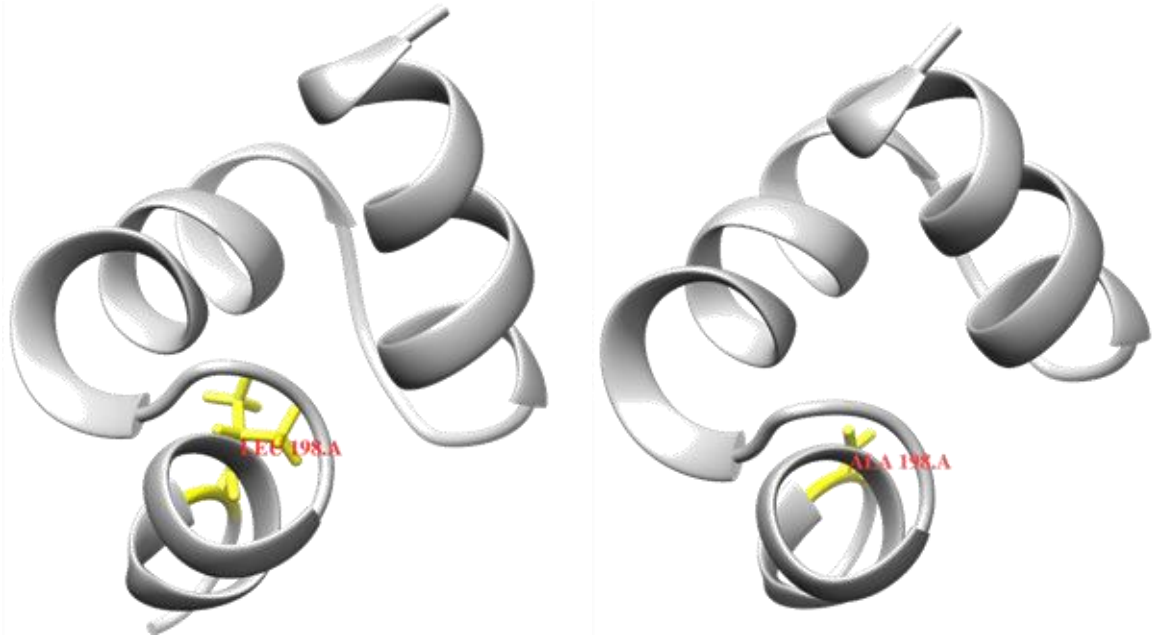
1. Jumper, J., Evans, R., Pritzel, A. *et al. Nature* (2021)
2. Varadi M, Bertoni D, Magana P, et al. *Nucleic Acids Res*. 2024
3. Buel, G.R., Walters, K.J. *Nat Struct Mol Biol* (2022)
4. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Research, 2000*

Motivation for the study

The proposed explanation – for structure to become disordered – **reduced packing** of hydrophobic core leading to destabilization of the structure



Leucine residue Alanine residue



Structure of UBA1 domain wild-type (left), mutant (right); Sidechain of residue 198 is shown (in yellow)

1. Buel, G.R., Walters, K.J. *Nat Struct Mol Biol* (2022)

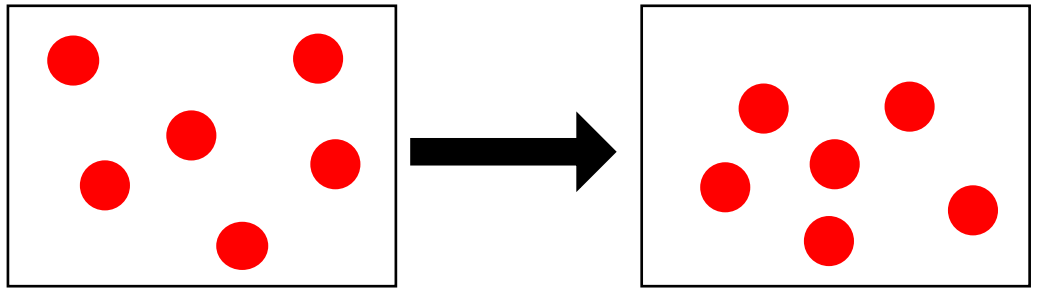
Objectives of the Thesis

- What makes this mutation to destabilize the protein structure?
- Use Molecular Dynamics Simulations to see if we can see the structural destabilization
- Are results Force Field dependent? – Comparison of Force Fields
- Can the result provide an indicator of which of the existing force fields can be used in evaluating the mutational effects in proteins?

Molecular Dynamics Simulation Technique

Newton's equation of motion –

$$a_i = \frac{dv_i}{dt} = \frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$



Conformation at time t Conformation at time t+Δt

Initial coordinates – Starting structure

Prepare system (define simulation box, add water, ions, and missing atoms)

Run simulation (thermostat, barostat)

Analysis

Force Fields

$$F_i = -\frac{d}{dr_i} V(r_1, r_2 \ \ldots r_N)$$ i = 1.....$N_{particles}$

$$V(r) = \underbrace{\frac{1}{2}\sum_{i=1}^{N_{bonds}} K_b(r - r_{eq})^2 + \frac{1}{2}\sum_{i=1}^{N_{angles}} K_\theta(\theta - \theta_{eq})^2 + \frac{1}{2}\sum_{i=1}^{N_{dihedral}} K_\phi[1 + \cos(n\phi - \delta))]}_{\text{Bonded Interactions}} + \underbrace{\sum_{i>j}^{nbpairs} \frac{q_i q_j}{4\pi\epsilon r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]}_{\text{Non-Bonded Interactions}}$$

- **GROMACS**[1], **AMBER**[2], **LAMPS**[3], **NAMD**[4] are a few popular software to perform MD simulations
- **Amber ff94**[5]**, CHARMM22**[6]**, and OPLS-AA**[7] are the pioneering force fields developed for biomolecular simulations

1. H. J. C. Berendsen et al. *Comput Phys Commun*, 1995
2. D. A. Case et al. *J Comput Chem*, 2005
3. S. Plimpton et al. 1995
4. J. C. Phillips et al. 2005
5. C. I. Bayly et al. *J Am Chem Soc*, 1995
6. A. D. MacKerell et al. *Journal of Physical Chemistry B*, 1998
7. W. L. Jorgensen et al. *Journal of the American Chemical Society* 1996

Intrinsically Disordered Proteins and Regions (IDPs/IDRs)

- Lacks stable three-dimensional structure under physiological conditions
- Exhibit biased amino-acid composition – rich in charged and polar residues and deficient in hydrophobic residues
- Highly abundant - 33% of eukaryotic proteins contains disordered regions (>30) residue long [1]

Challenges with IDPs simulations –

- Transition from disorder to order (induced folding)
- Fluctuating conformational states
- Comprise both folded and disordered regions

✓ Ideal force field should adhere to different chemical environment in IDPs simulations

1. Dunker AK, Lawson JD, Brown CJ, et al. *J Mol Graph Model*. 2001
2. C. Ferreon, A. Chris, M. Ferreon, R. Trivedi, and H. A. Nagarajaram, *International Journal of Molecular Sciences,* 2022

Force Fields for IDPs Simulations

- ❑ Overestimation of secondary structures
- ❑ Generates too compact conformations
  - Imbalance between the strengths of protein-protein and protein-water interactions

- ✓ Reparameterization of dihedral parameters using data of coil fragments
  - Residue-specific dihedral parameters
- ✓ Refinement of protein-water Van der Waals parameters

❑ **Amber ff99IDPs**, **ff03w**, **Charmm36m**, and **Charmm22*** are some of the upgraded forces fields for IDPs simulations

Charmm36m force field –

- Improved dihedral parameters
- Improved protein-water Van der Waals parameters
- Balanced force fields for both folded and IDPs

1. J. Mu, H. Liu, J. Zhang, R. Luo, and H. F. Chen. *J Chem Inf Model*, 2021
2. J. Huang and A. D. MacKerell, *Curr Opin Struct Biol*, 2018
3. J. Huang et al. *Nature Methods* 2016
4. M. U. Rahman, A. U. Rehman, H. Liu, and H. F. Chen, *J Chem Inf Model*, 2020

Simulation Protocols used in this work

Protein system	Force Fields	Water model
UBA1 wild-type	Amber ff99SB-ILDN	TIP3P
	CHARMM36	TIP3P
	OPLS-AA/L	SPC/E
	CHARMM36m	TIP3P
UBA1 mutant (L198A)	Amber ff99SB-ILDN	TIP3P
	CHARMM36	TIP3P
	OPLS-AA/L	SPC/E
	CHARMM36m	TIP3P

Structure of UBA1 domain – (PDB ID: 1OQY, residues 161-200)

↓

Prepared systems in cubic box, added water, ions, and missing atoms

↓

Equilibrated the systems at the temperature of 300K and pressure of 1bar

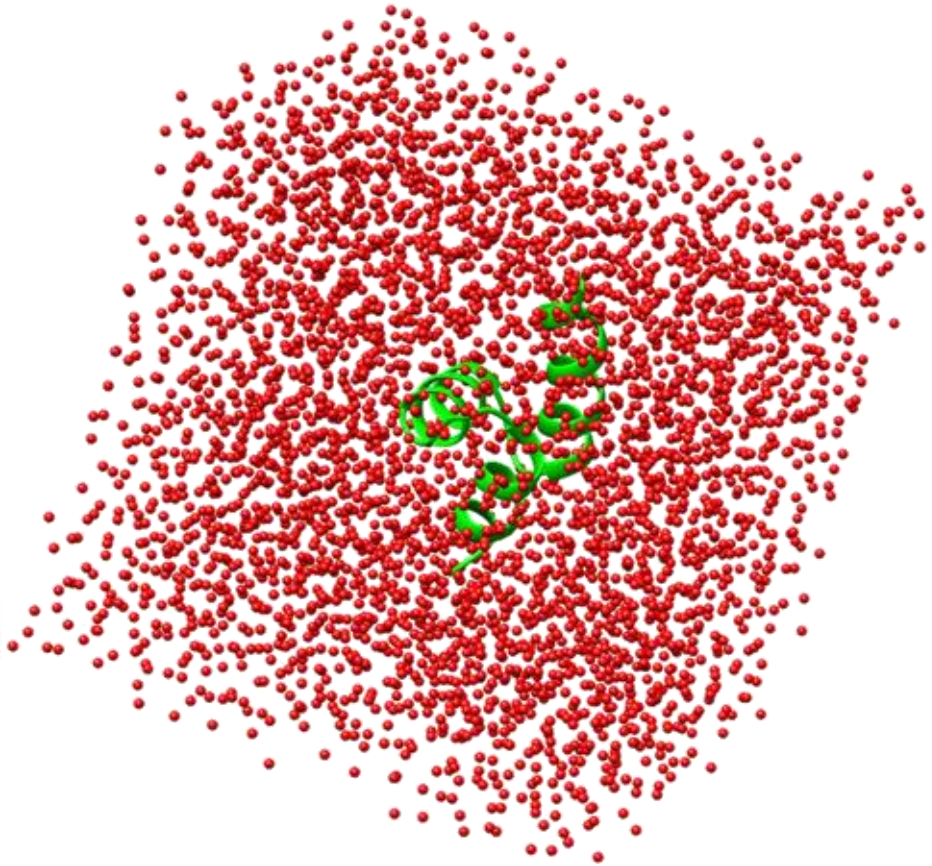
↓

Production run for 1000 ns

↓

Analysis

Simulation Protocols used in this work



Simulation box (12233 atoms); UBA1 domain at center is shown in green and only the oxygen atoms of water molecules is shown in red

Structure of UBA1 domain – (PDB ID: 1OQY, residues 161-200)

↓

Prepared systems in cubic box, added water, ions, and missing atoms

↓

Equilibrated the systems at the temperature of 300K and pressure of 1bar

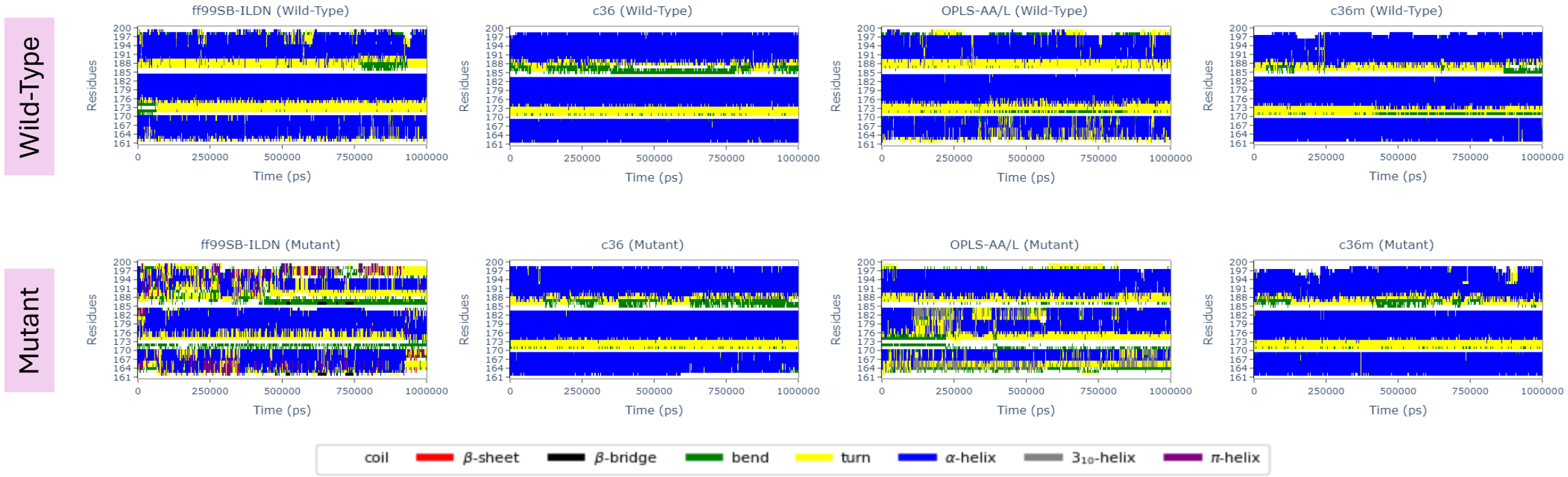
↓

Production run for 1000 ns

↓

Analysis

Secondary Structure Analysis

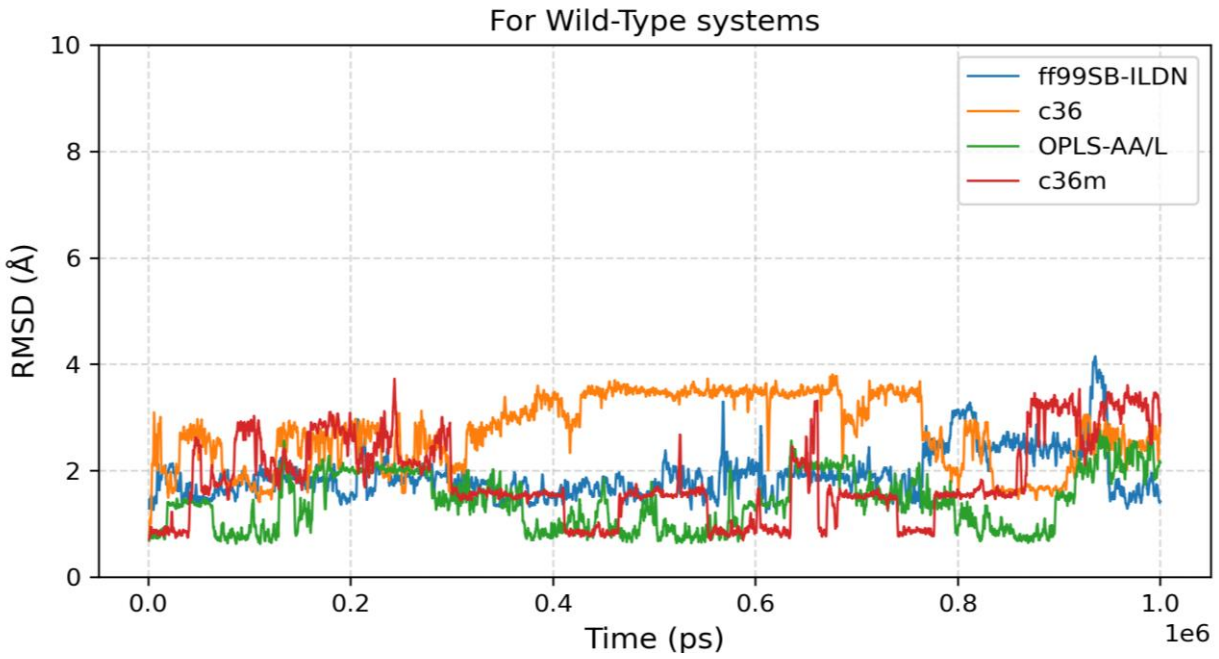


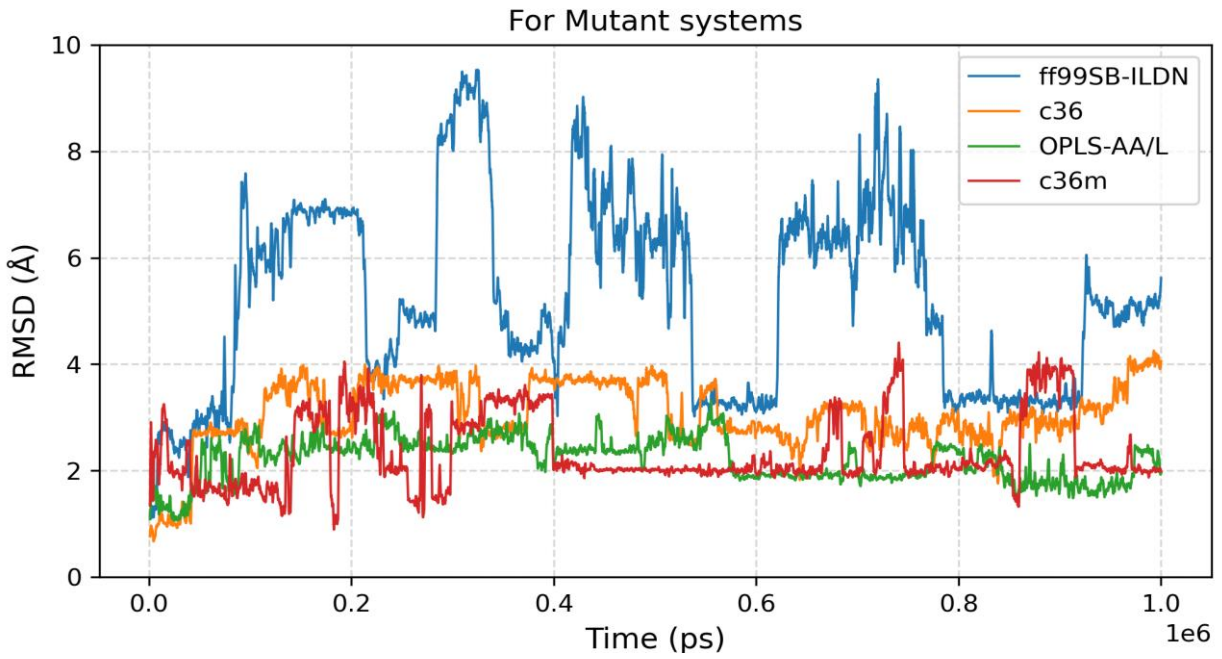
DSSP [1] plots – Evolution of secondary structures with respect to time

- Significant loss of Alpha helical content [mutant-Amber ff99SB-ILDN followed by OPLS-AA/L]
  - Alpha-helical to Turn – **Unwinding of helices**
- Mutant-Charmm36 and mutant-Charmm36m showed minimal changes
- Negligible change in structures from wild-type systems

1. Kabsch W, Sander C. *Biopolymers*. 1983

Root Mean Square Deviation (RMSD)

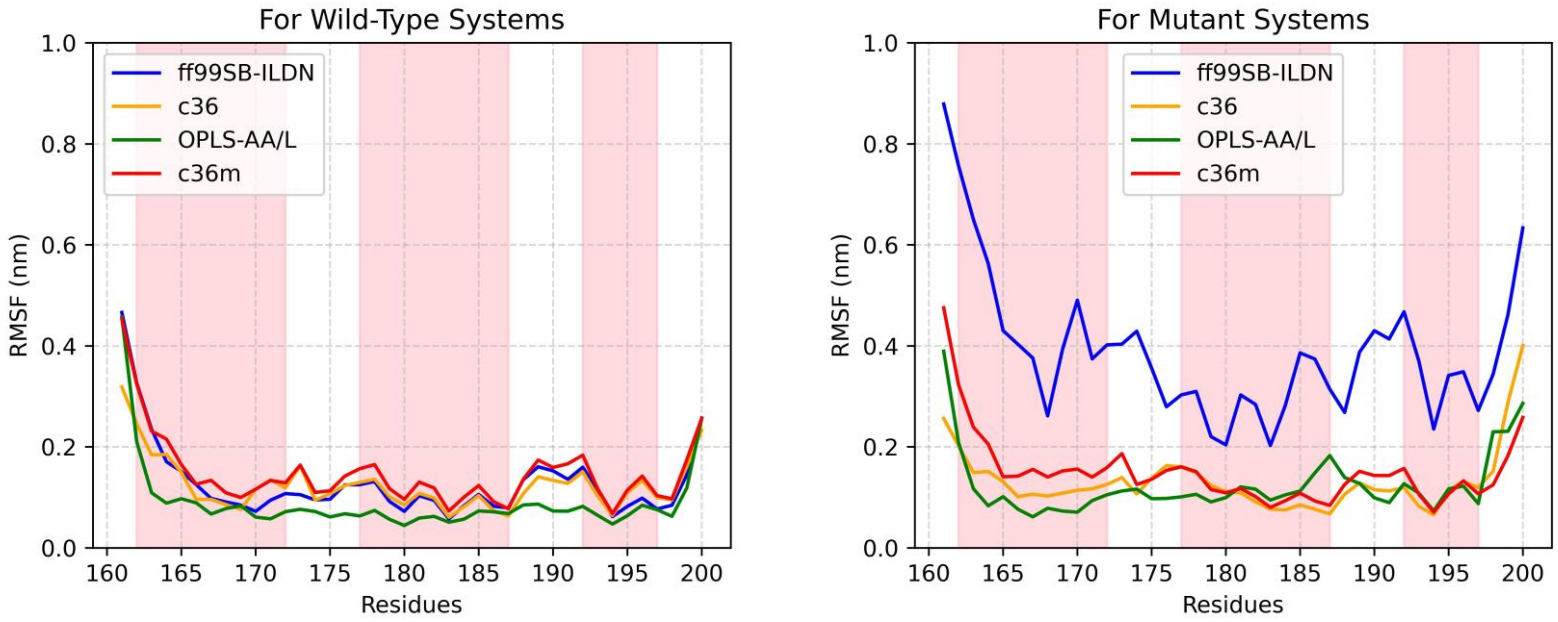




RMSD values for C-alpha atoms of helix residues; wild-type systems (left), mutant systems (right)

- RMSD values from wild-type systems remained stable throughout the simulation period
- Mutant-Amber ff99SB-ILDN – very high RMSD values with high fluctuations - **large deviation from the initial structure**
- All other mutant systems – stable RMSD values and showed minimal changes

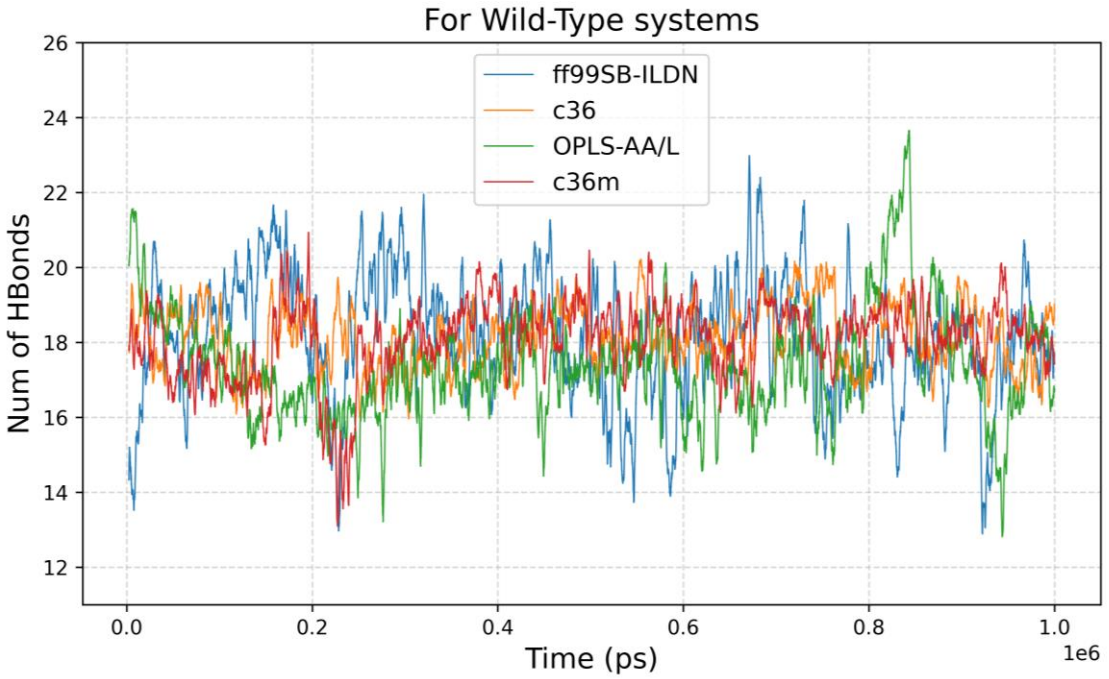
Root Mean Square Fluctuations (RMSF)

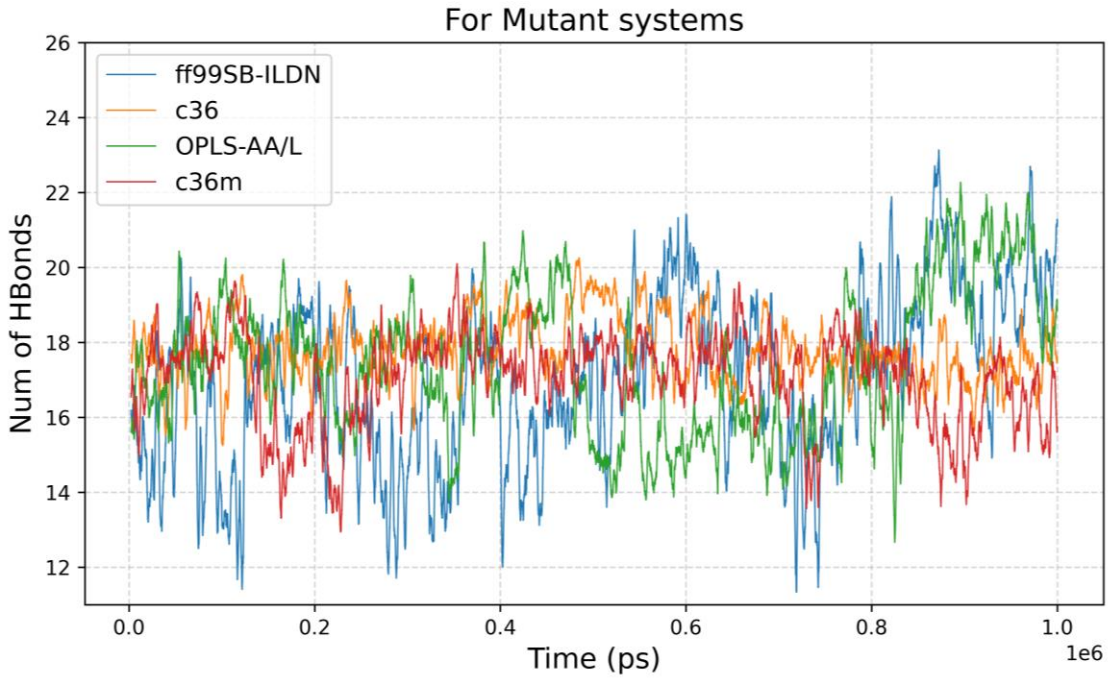


RMSF values for C-alpha atoms: wild-type (left), mutant (right) systems

- Mutant-Amber ff99SB-ILDN – high RMSF values for all residues – **Large displacement** of atoms from their starting structure
- Small fluctuations values for all other systems – **small local changes in the structures**

Hydrogen Bond Analysis

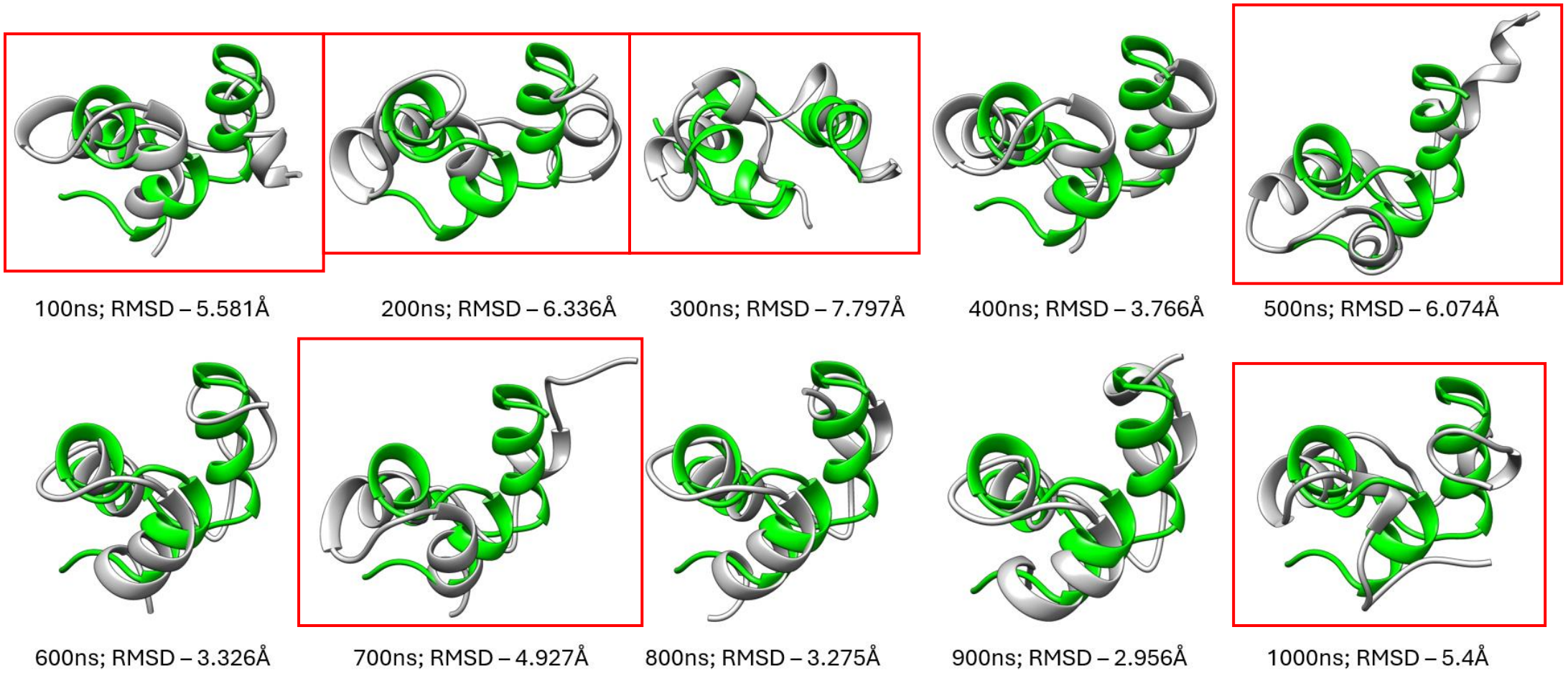




Number of hydrogen bonds associated with helix residues: wild-type (left), mutant (right) systems

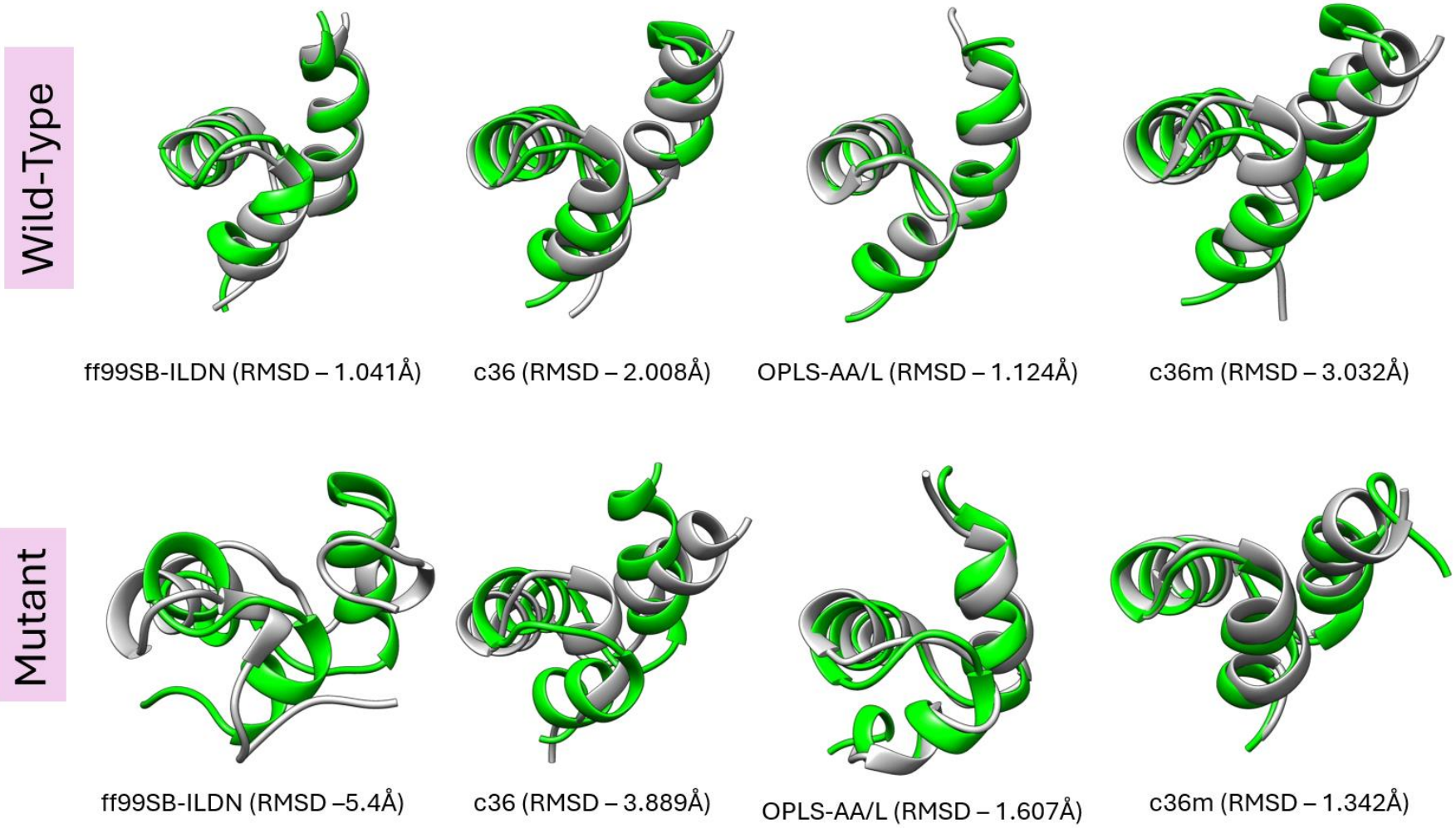
- Minimal change in the values from wild-type and mutant systems
- Mutant-Amber ff99SB-ILDN followed by OPLS-AA/L – slightly lower values with relatively high fluctuations – **small decrease indicate that, structure is not becoming fully disordered**

Superimposed structures from mutant-Amberff99SB-ILDN system



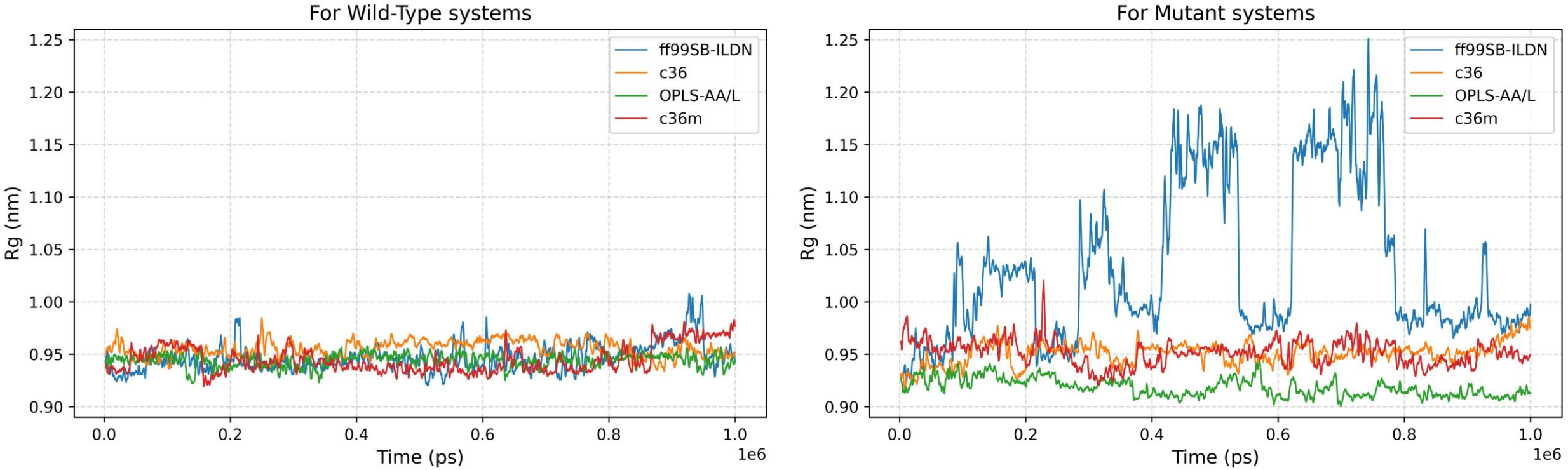
Structures at different timesteps (gray) of simulation superimposed with the structure of starting frame (green) (0ns). The RMSD values are computed for C-alpha atoms of helix residues.

Superimposed structures of last frame from all systems



Structures of last frame (gray) from all simulation systems superimposed with their respective starting structures (green). RMSD values are computed for C-alpha atoms of helix residues

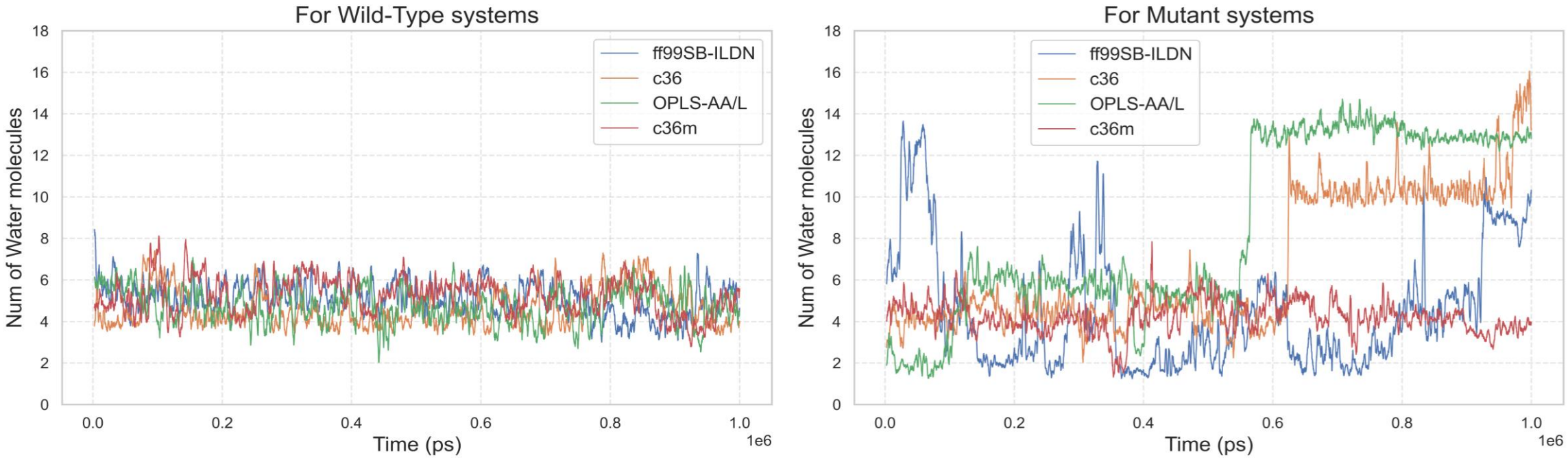
Radius of Gyration (Rg)



Rg values about the center of mass; wild-type (left), mutant (right) systems

- Stable and low Rg values for wild-type systems – indicating the **stability of fold** of wild-type structure
- Mutant-Amber ff99SB-ILDN – very high Rg values – indicating extensive spread of the atoms
- In contrast, mutant-OPLS-AA/L systems – lower Rg values than its wild-type counterpart – **more compact conformations**

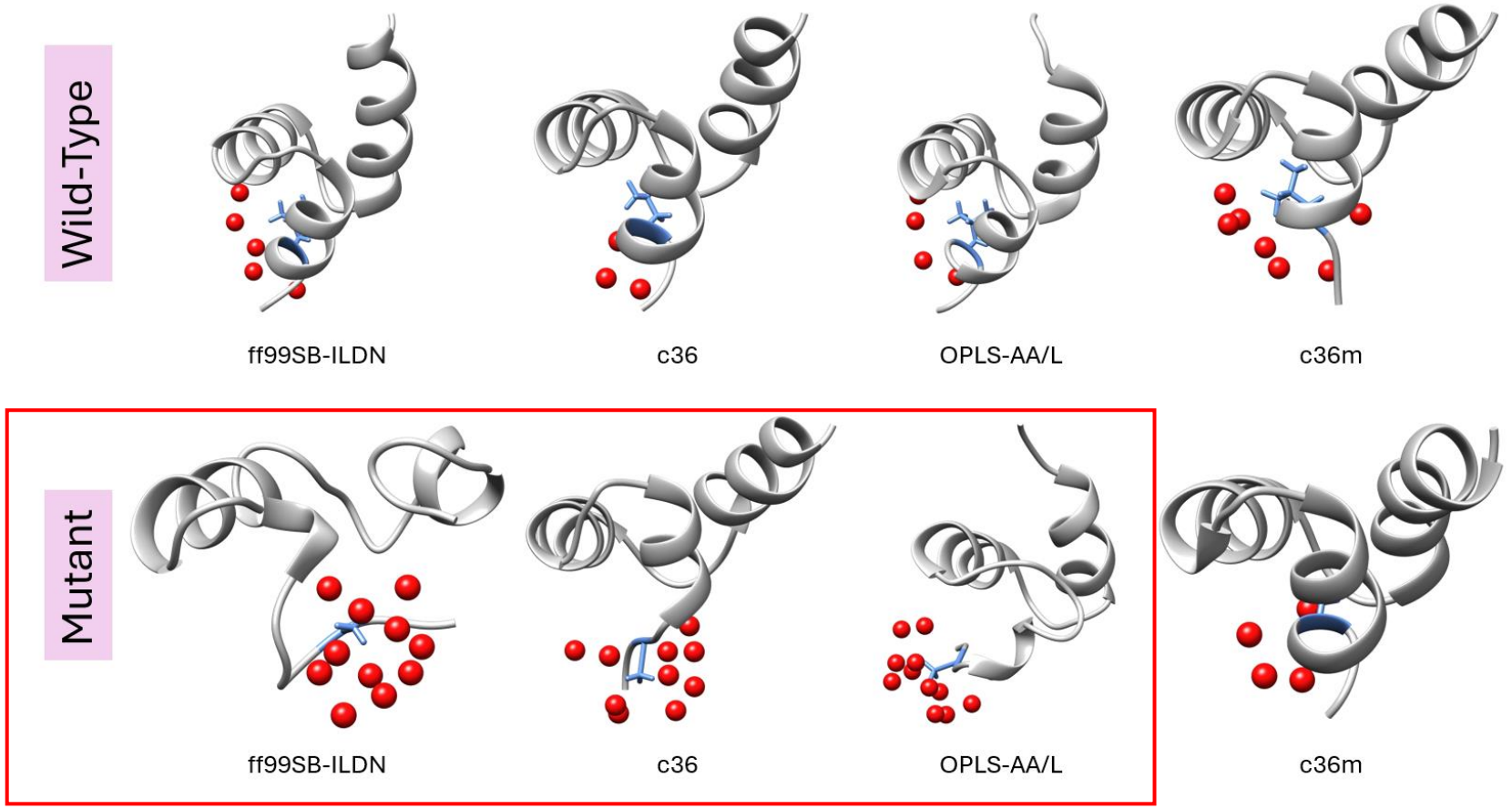
Number of Water Molecules around Residue Position 198



The number of water molecules around residue position 198 that falls within a cutoff sphere of 4Å; wild-type (left), mutant (right)

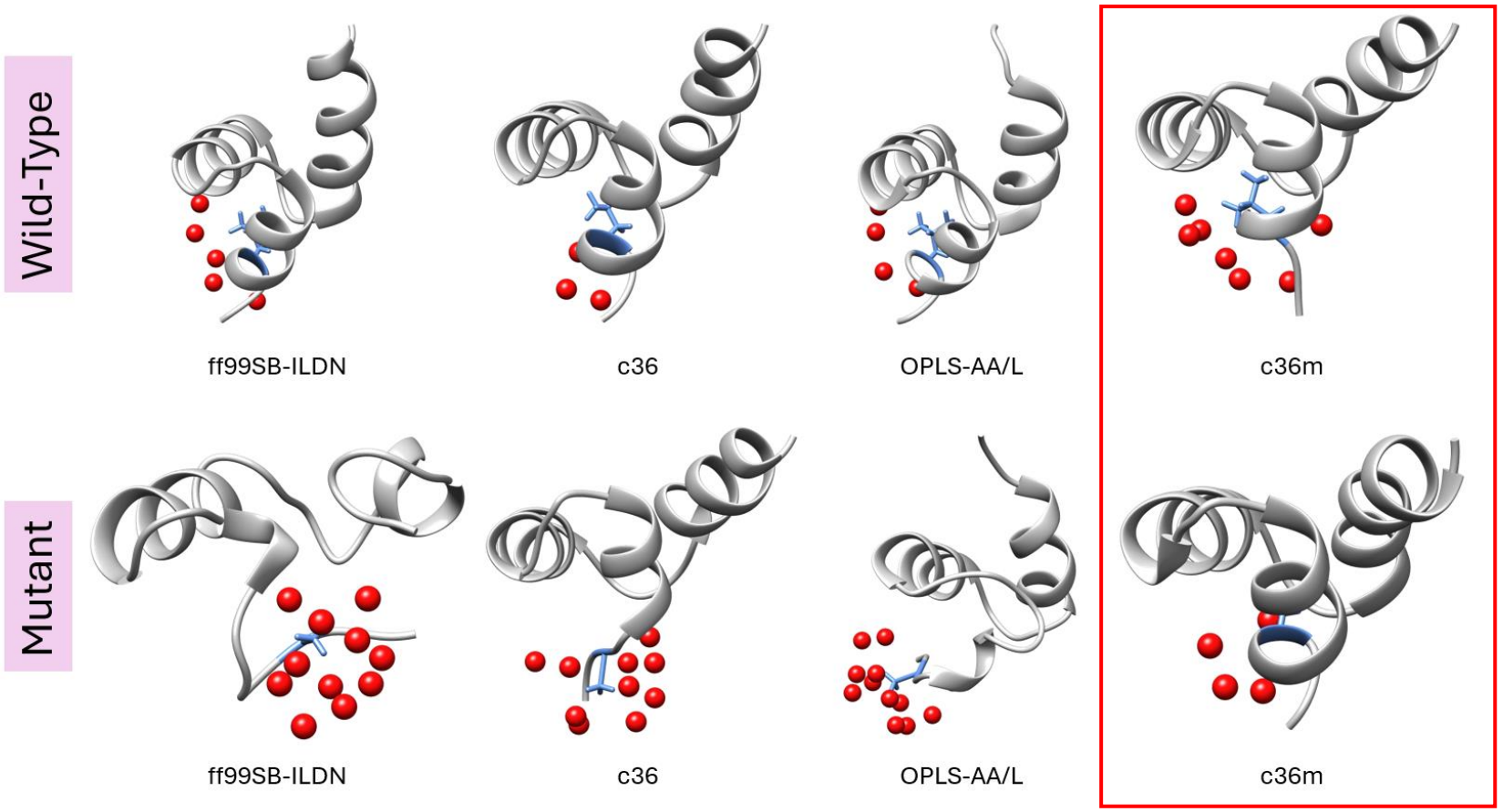
- Values from wild-type systems – stable and low number of water molecules
- High values for mutant – Amber ff99SB-ILDN, Charmm36, and OPLS-AA/L systems
- Mutant-Amber ff99SB-ILDN – high fluctuations throughout the simulation period
- Mutant-OPLS-AA/L and Charmm36 showed sudden jump in values from 550ns

Structures of the last frame with water molecules around residue 198



Structures of the last frame (1000ns) from eight simulations with surrounding water molecules (only oxygen atoms in red) within the cutoff of 4Å. The side chain of residue 198 is shown in a light blue color.

Structures of the last frame with water molecules around residue 198



Structures of the last frame (1000ns) from eight simulations with surrounding water molecules (only oxygen atoms in red) within the cutoff of 4Å. The side chain of residue 198 is shown in a light blue color.

Conclusions

- The highly exposed side chain of the alanine residue leads to reduced packing of the hydrophobic core, which likely destabilizes intramolecular interactions.
- Amber ff99SB-ILDN sampled the most disordered conformations, followed by OPLS-AA/L
- CHARMM36 and CHARMM36m both showed high propensity to reproduce helical content
- The current study demonstrates that Amber ff99SB-ILDN, followed by OPLS-AA/L, may be suitable to study the structural destabilization of missense mutations in proteins

Future Directions

To investigate the extent of the disorder caused by the L198A mutation, other improved force fields for IDPs could be explored

An experimental study to investigate the presence of this mutation in vivo and to examine its potential functional consequences

Further simulation studies on diverse proteins are required to convincingly conclude the suitability of force fields to study the structural effect of missense mutations

Acknowledgments

- Dr. R. Sankararamakrishnan
- All BSBE Faculties and Staff
- HPC2013 and Param Sanganak Facility at IITK
- My Lab Members
- My friends at IITK
- My family and friends
- IIT Kanpur