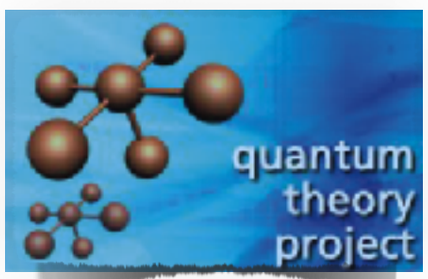
Sampling rare events: folding pathways

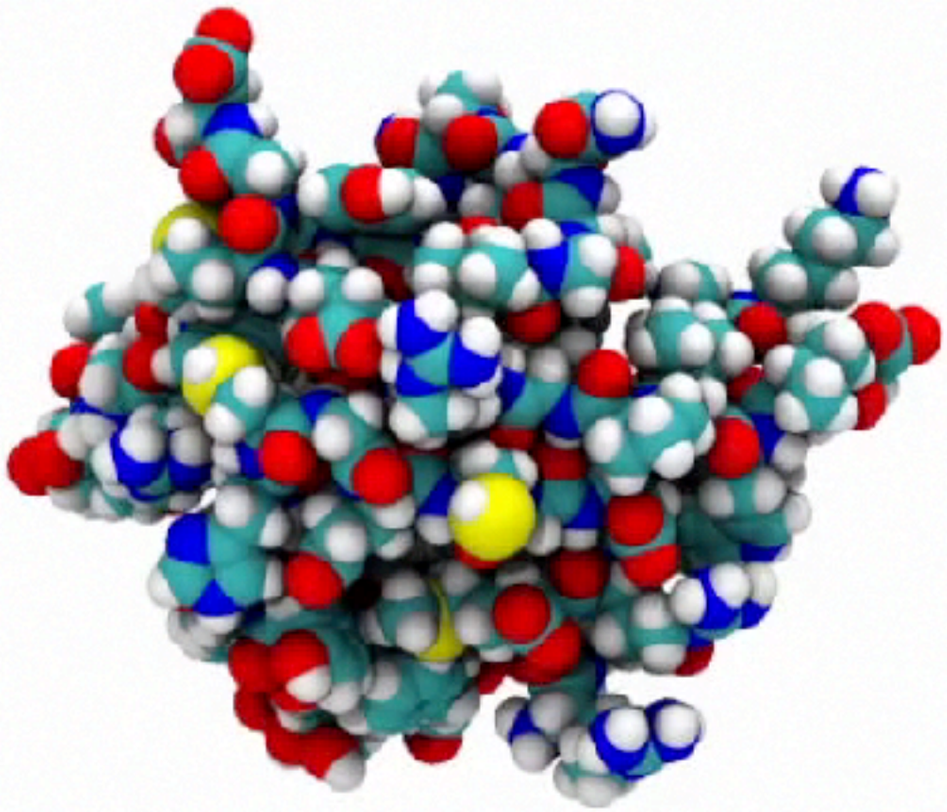
Alberto Perez

Mar 10th 2022 — OpenEye CUP XXI

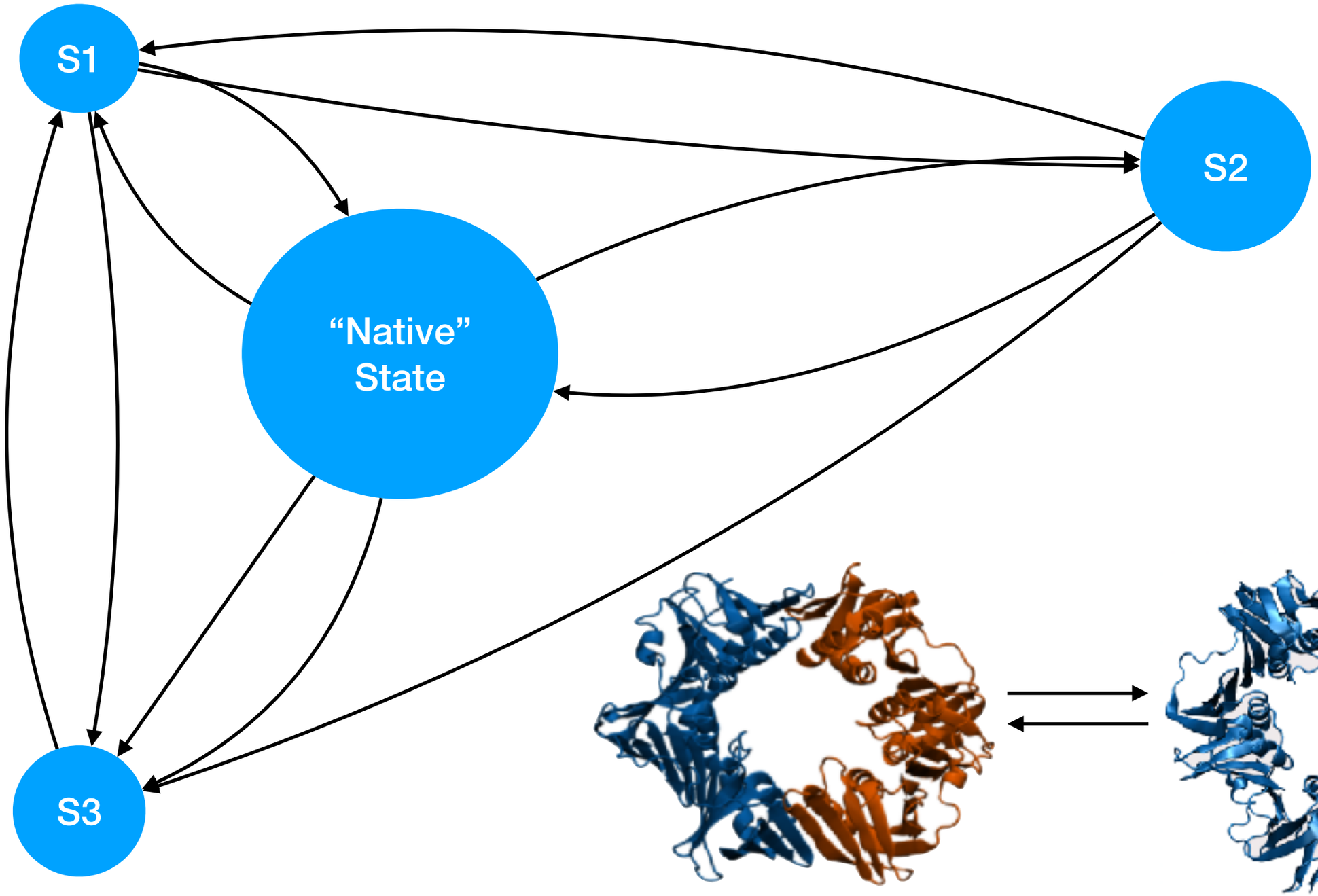


Everything that living things can do can be understood in terms of the jiggling and wigglings of atoms

- Richard Feynman

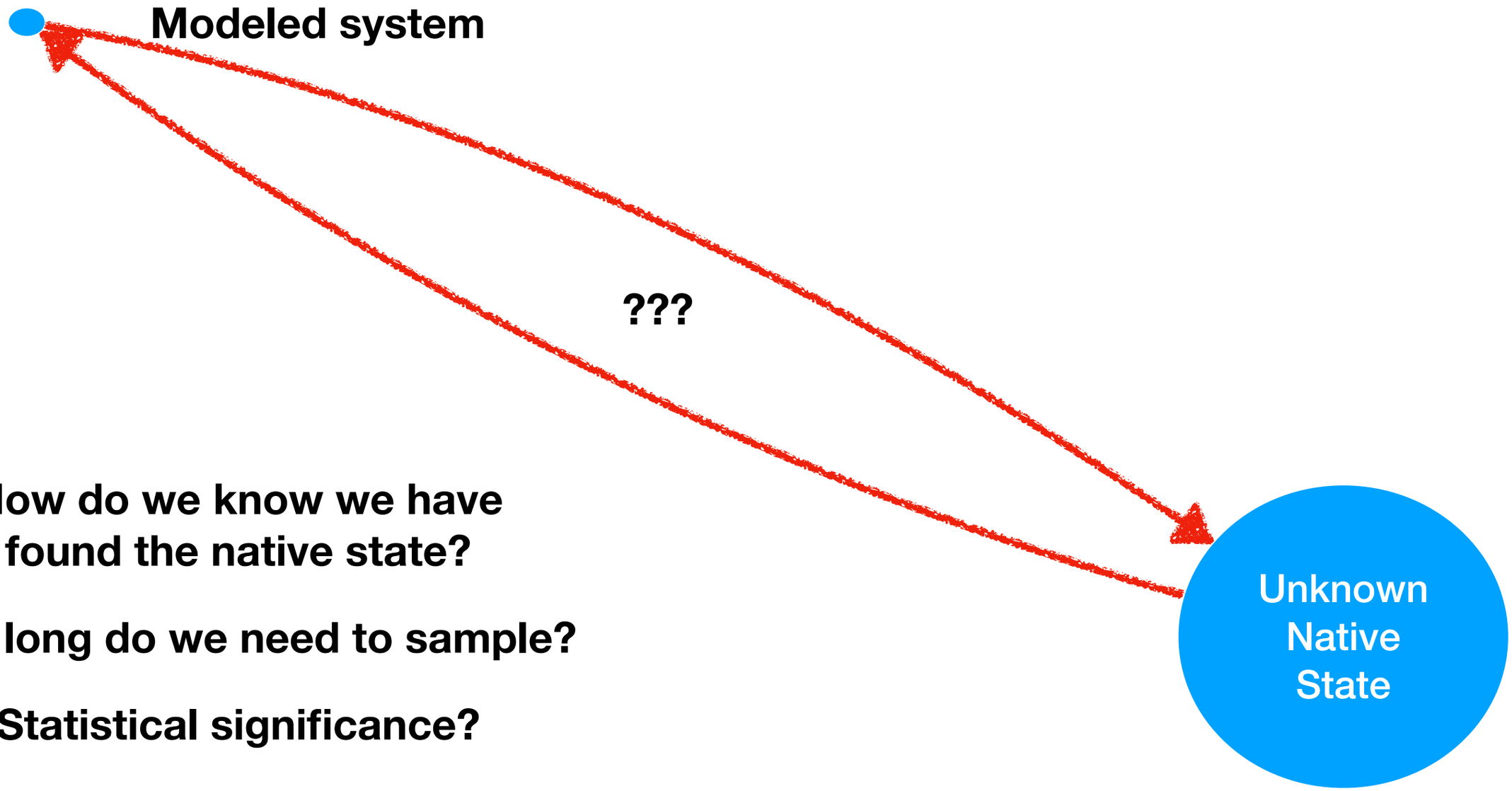


We can understand MD as jumps between states at certain timescales



1.Esmaeeli, R., Andal, B. & Perez, A. *Life* **12**, 261 (2022).

Sometimes we are interested in the native state but we don't even know what it looks like



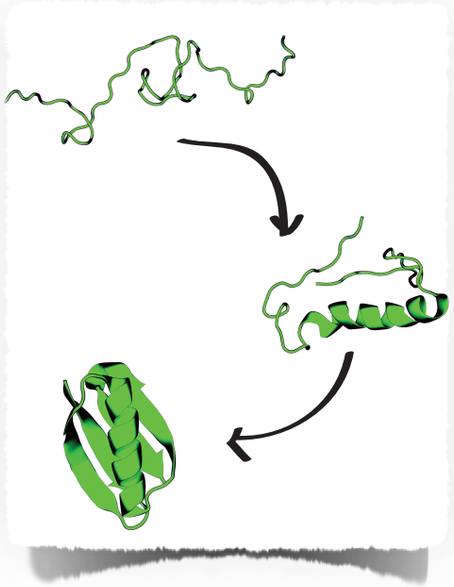
How do we know we have found the native state?

How long do we need to sample?

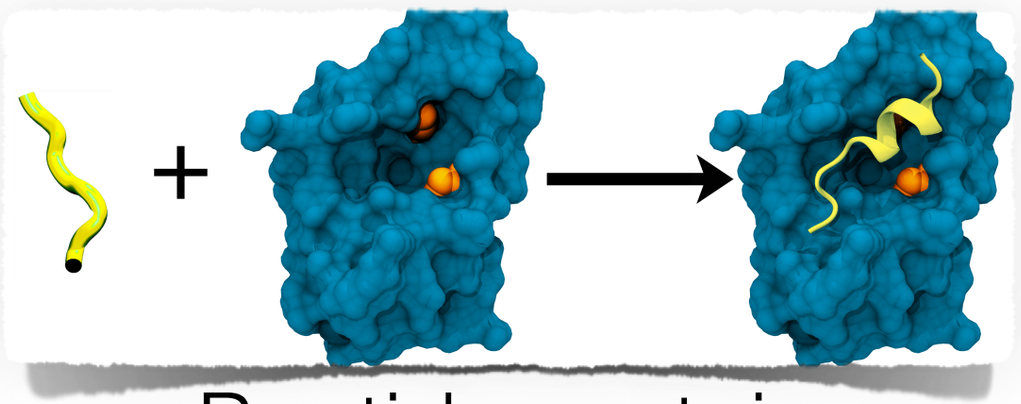
Statistical significance?

What is the physical process?

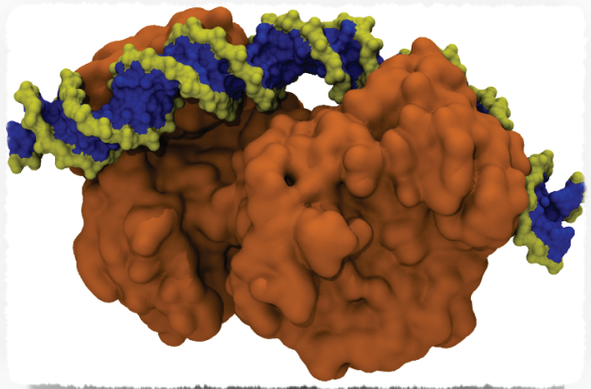
Our group works in understanding molecular interactions



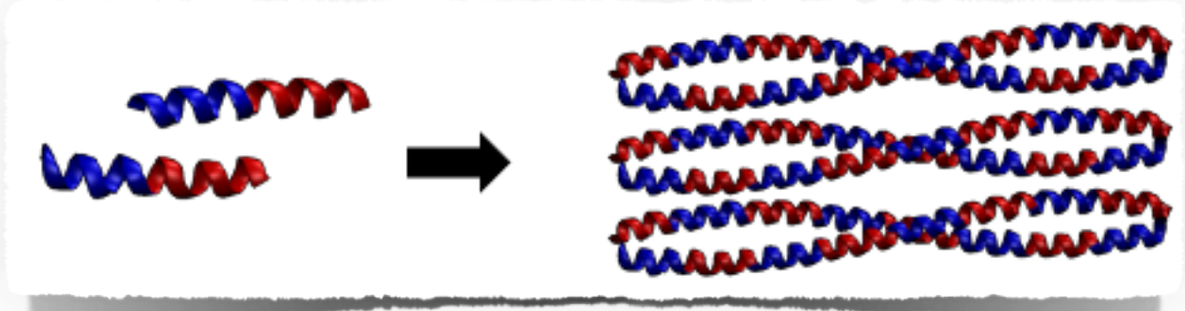
Protein Folding/ Structure determination



Peptide-protein binding

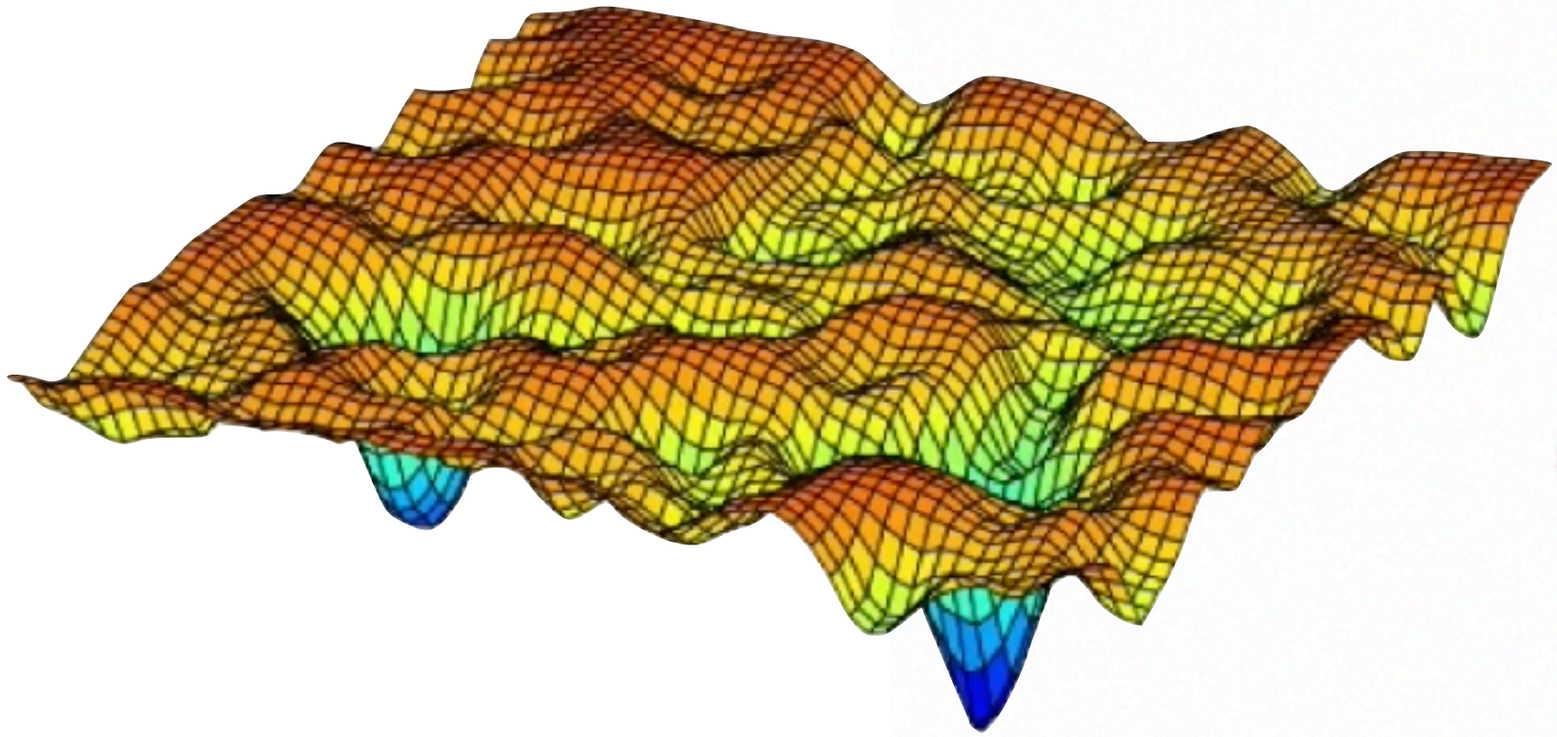


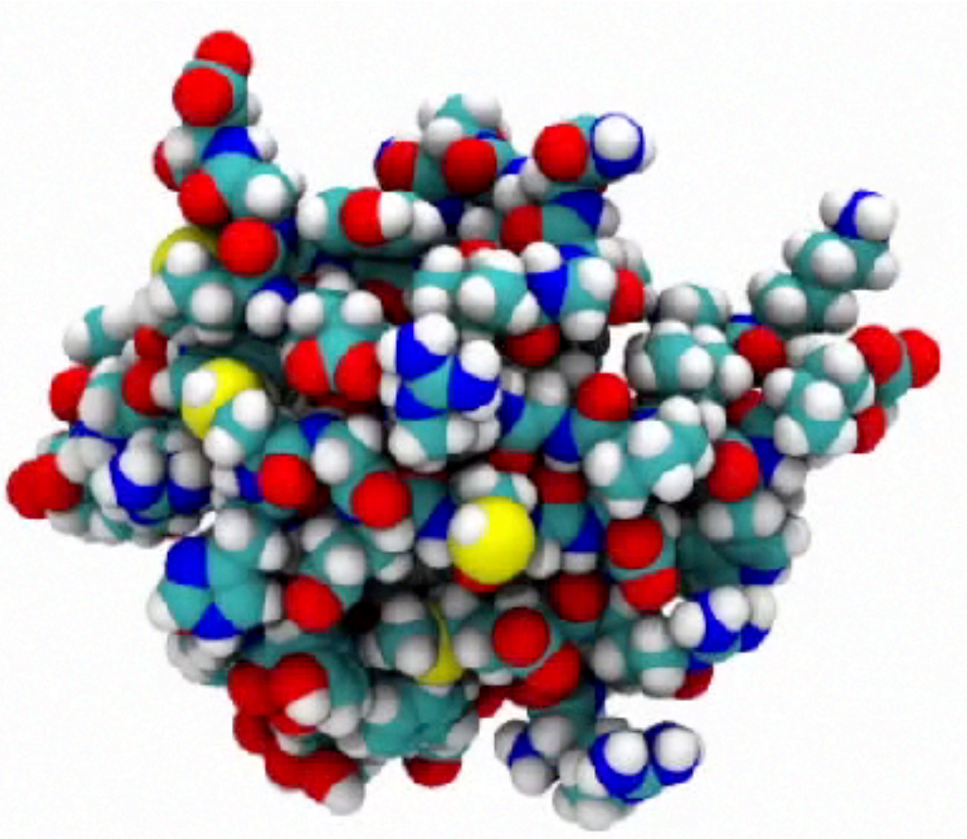
DNA-protein binding

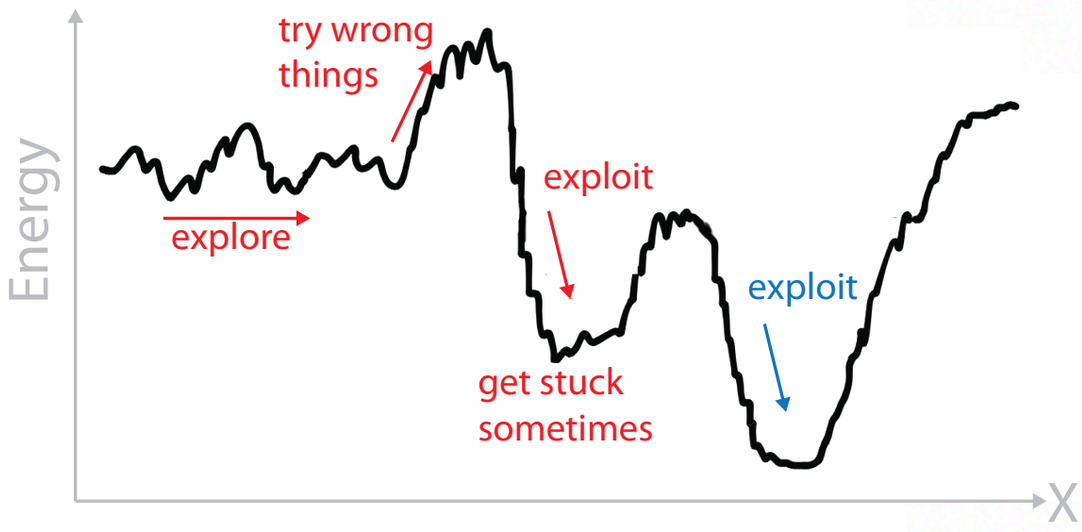


Peptide self-assembly

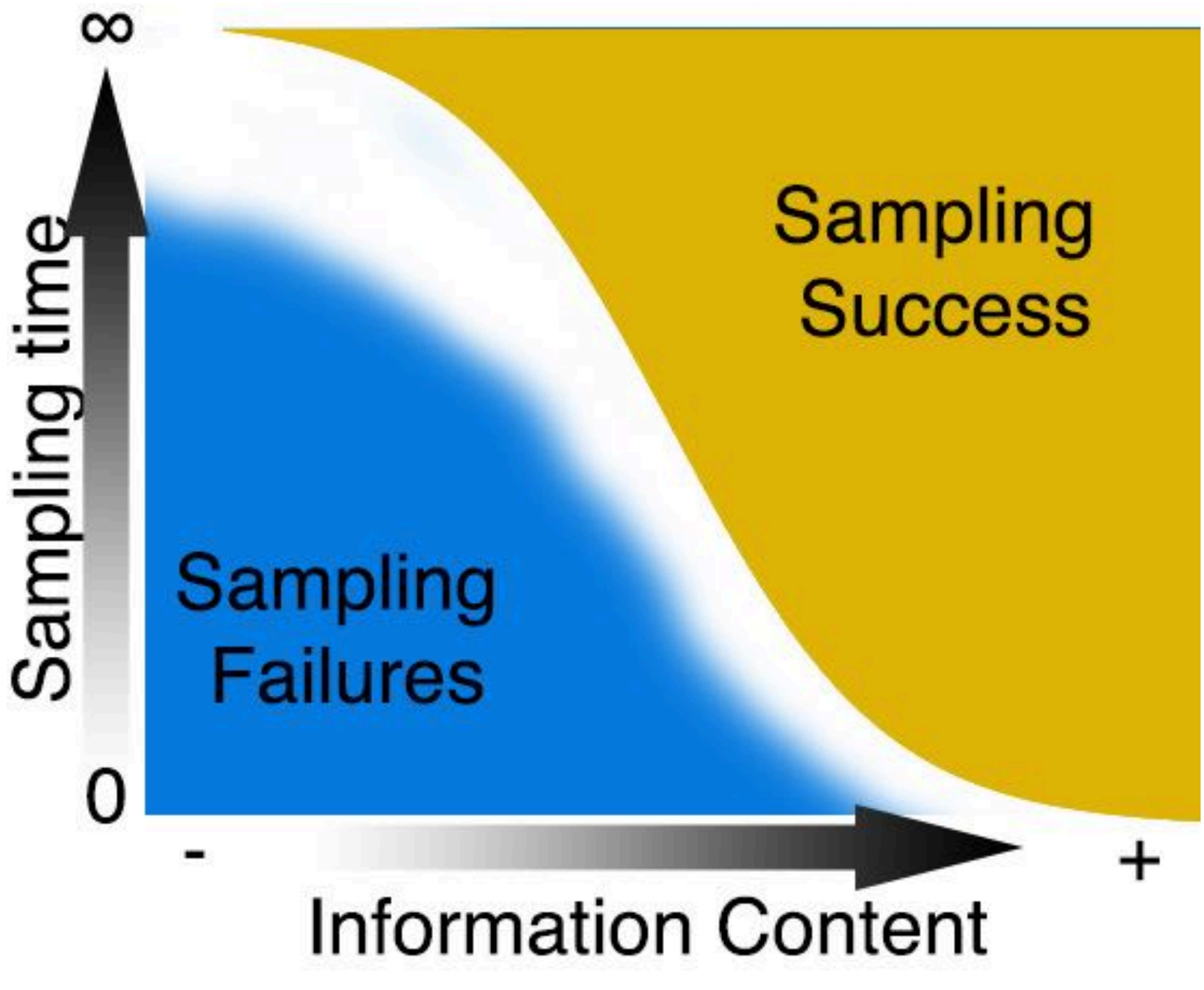
Conventional MD is inefficient at exploring the energy landscape







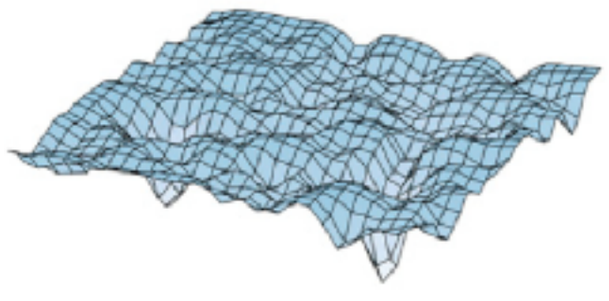
We leverage biophysical data that has been insufficient for structural determination



We incorporate data into simulations through Bayesian inference

$$\overbrace{p(\mathbf{x}|\mathbf{D})}^{\text{posterior}} = \frac{p(\mathbf{D}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{D})} \sim \overbrace{p(\mathbf{D}|\mathbf{x})}^{\text{likelihood}}\overbrace{p(\mathbf{x})}^{\text{prior}}$$

Force Field
(prior)

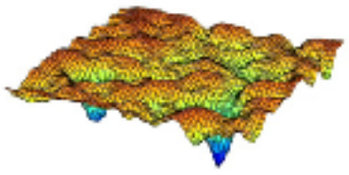


$$p(\mathbf{x}) \sim \exp[-\beta E_{\substack{\text{force}\\\text{field}}}(\mathbf{x})]$$

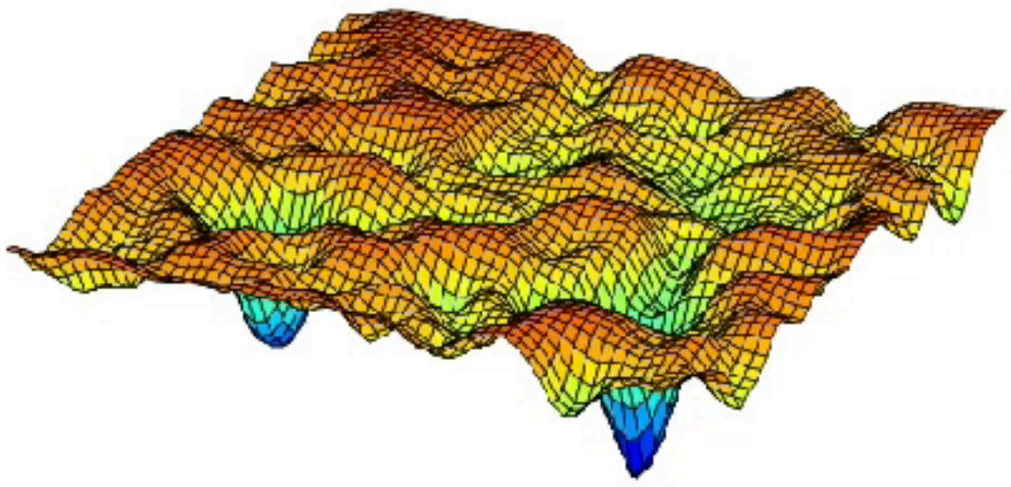
MELD

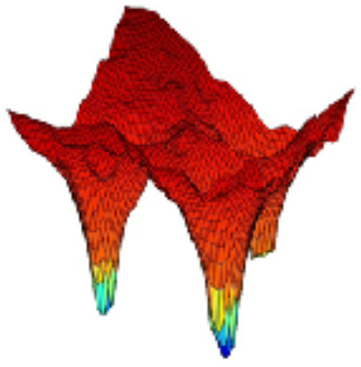
MacCallum*, Perez*, & Dill, Proc. Natl. Acad. Sci. U.S.A. 112, 6985–6990 (2015).

We use generalized ensemble methods for enhanced sampling

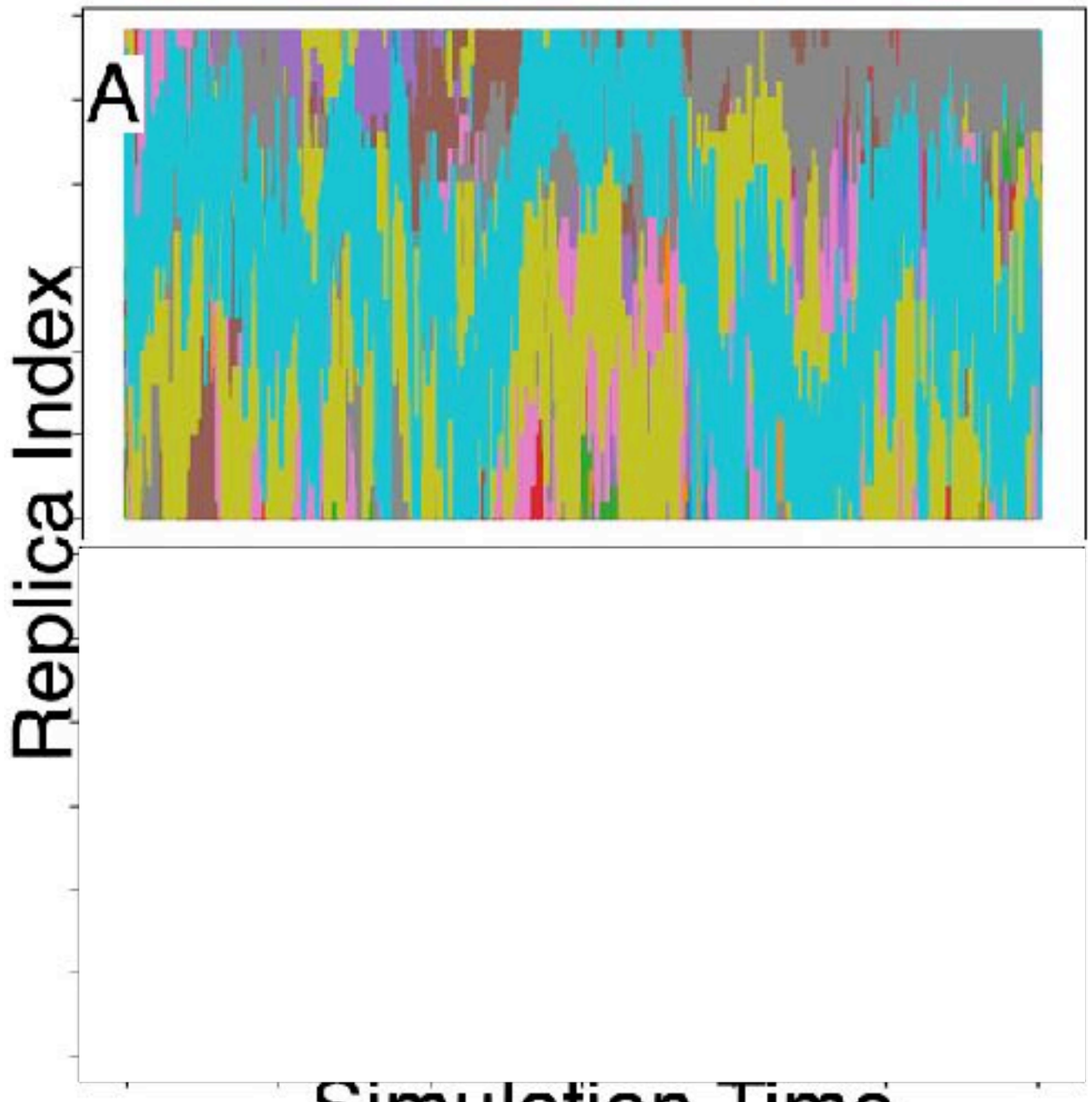


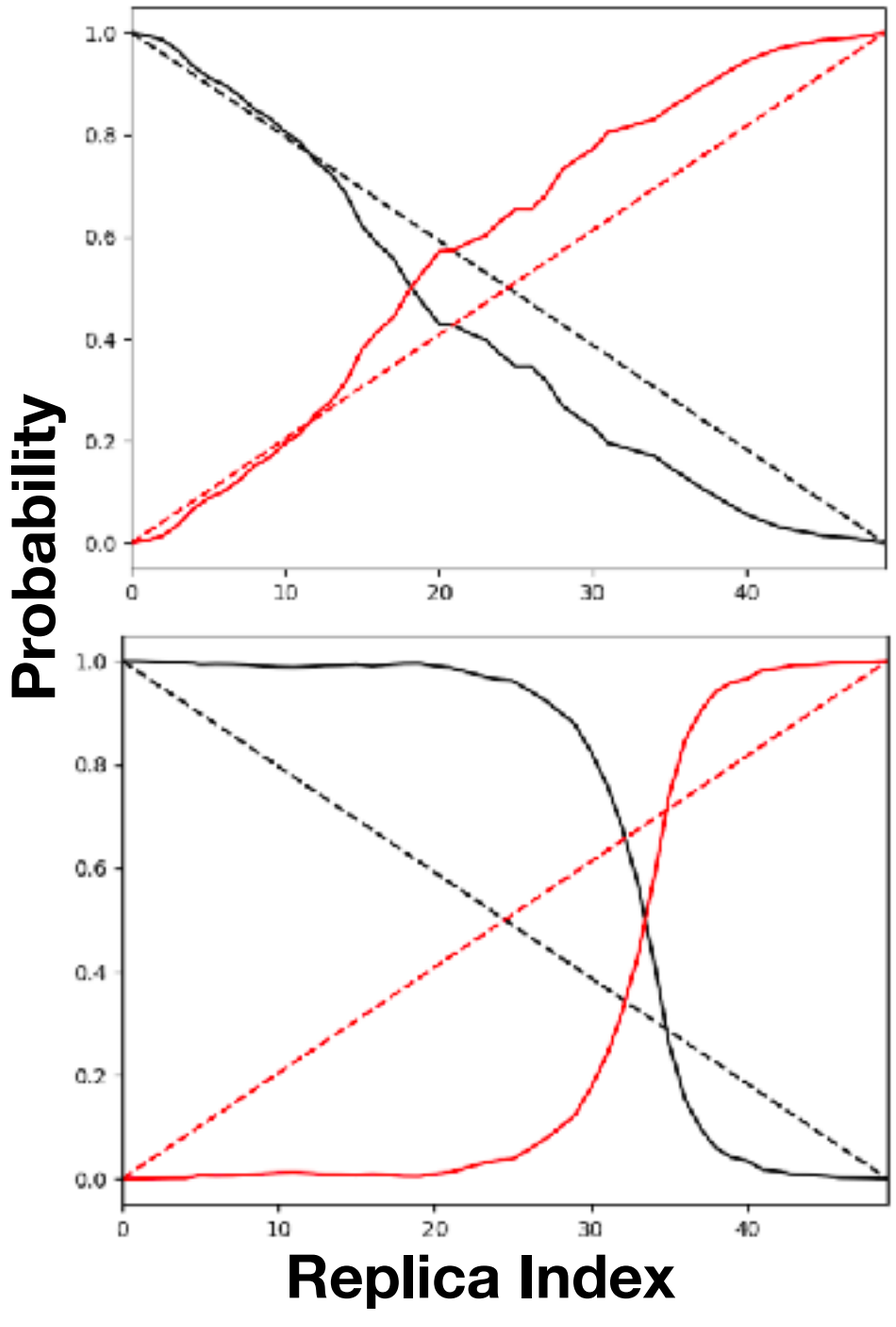
- High Temperature / Weak Restraints
-
-
-
-
-
-
- Low Temperature / **Strong Restraints**





Phase transitions lead to local exchanges and limit sampling efficiency





What types of data?

- Sparsely labeled NMR
- Cryo-EM (*CryoFold*)
- Φ-value analysis
- Chemical Shift Perturbation NMR
- Paramagnetic relaxation enhancement (PRE)
- General knowledge
- …

Perez, A., Gaalswyk, K., Jaroniec, C. P. & MacCallum, J. L.. *Angewandte Chemie Int Ed* **58**, 6564–6568 (2019)

Shekhar, M. *et al*. *Matter* **4**, 3195–3216 (2021)

Lawson, C. L. *et al*. *Nat Methods* **18**, 156–164 (2021)

Mondal, A. & Perez, A. *Frontiers Mol Biosci* **8**, 774394 (2021)

Mondal, A. *et al*. *Biorxiv* 2021.12.31.474671 (2022)

What types of data?

- Sparsely labeled NMR
- Cryo-EM (CryoFold)
- Φ-value analysis
- Chemical Shift Perturbation NMR
- Paramagnetic relaxation enhancement (PRE)
- General knowledge
- …

Perez, A., Gaalswyk, K., Jaroniec, C. P. & MacCallum, J. L.. *Angewandte Chemie Int Ed* **58**, 6564–6568 (2019)

Shekhar, M. *et al*. *Matter* **4**, 3195–3216 (2021)

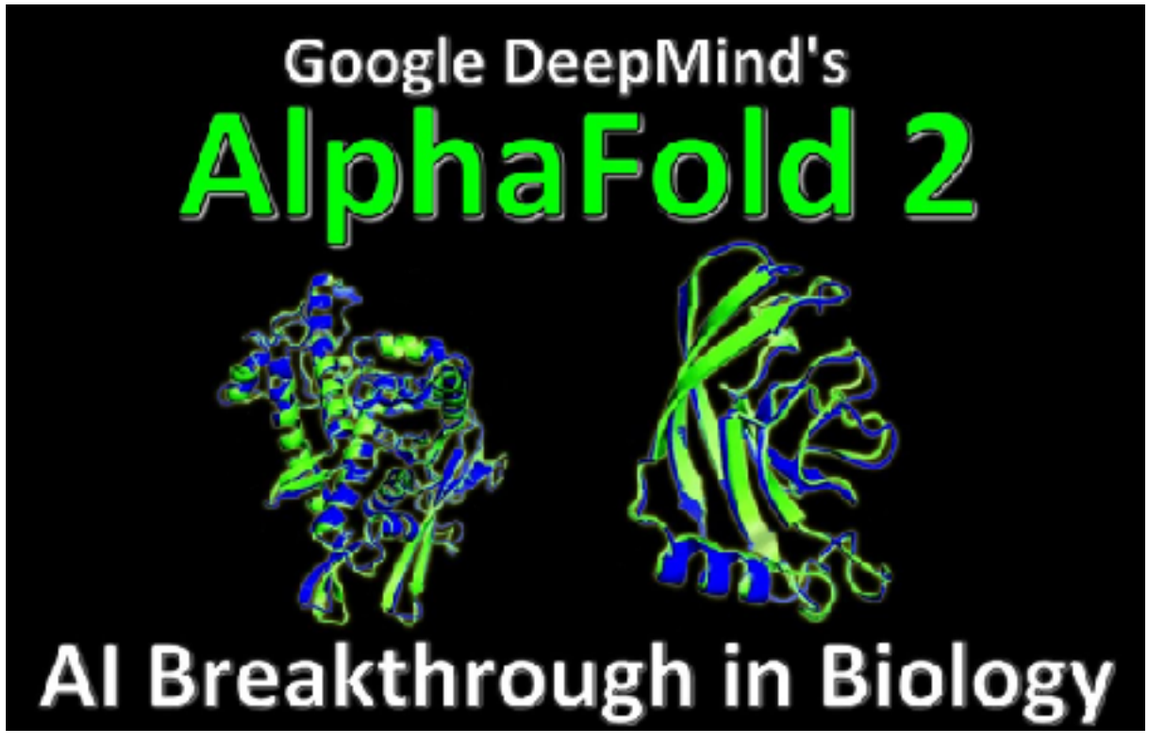
Lawson, C. L. *et al*. *Nat Methods* **18**, 156–164 (2021)

Mondal, A. & Perez, A. *Frontiers Mol Biosci* **8**, 774394 (2021)

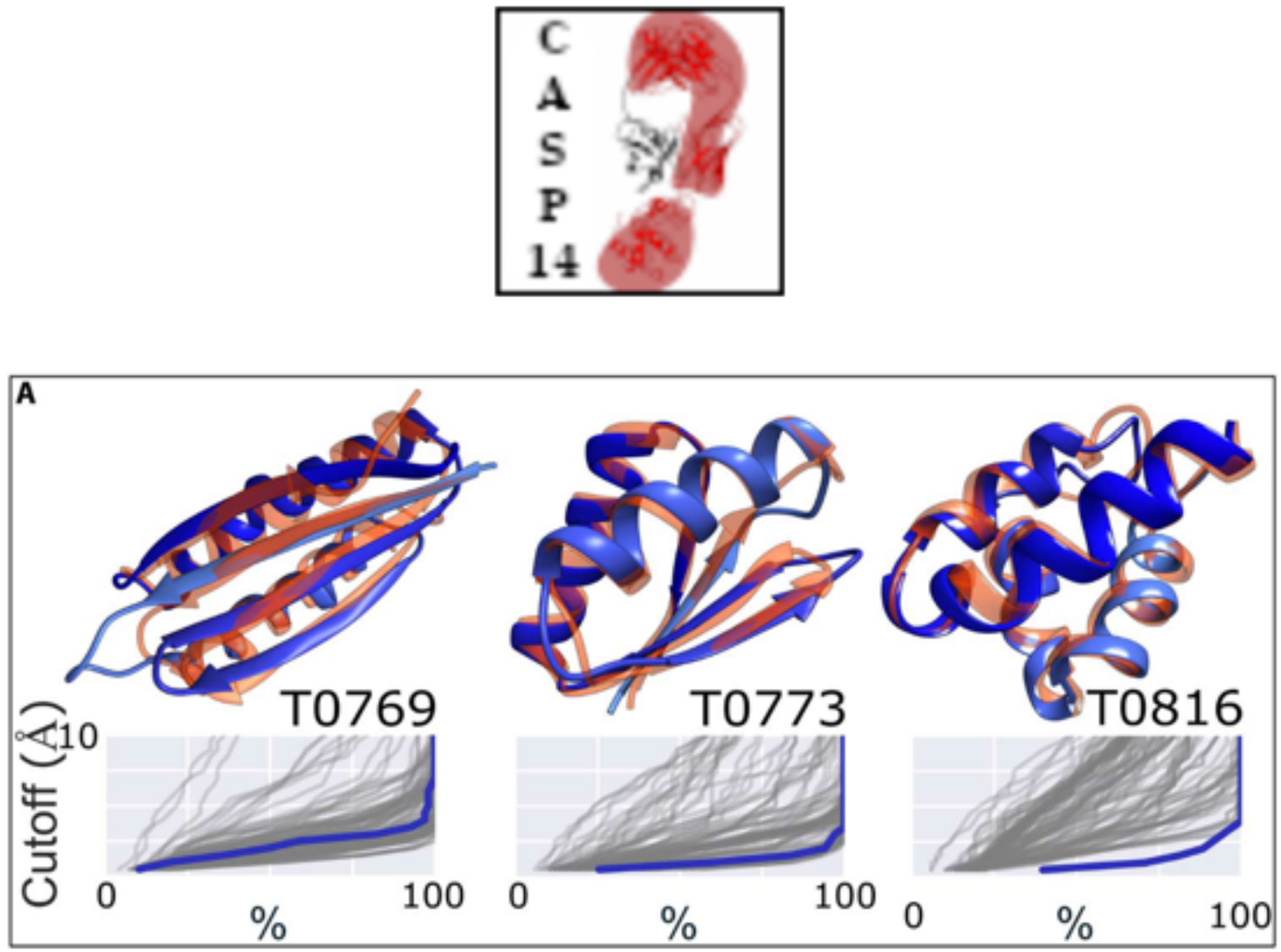
Mondal, A. *et al*. *Biorxiv* 2021.12.31.474671 (2022)

Protein Folding problem

- (1) What structure encoded by a sequence
- (2) How do proteins fold that fast (pathways)
- (3) Can we design new proteins



Blind competition events are a great way to validate methodologies

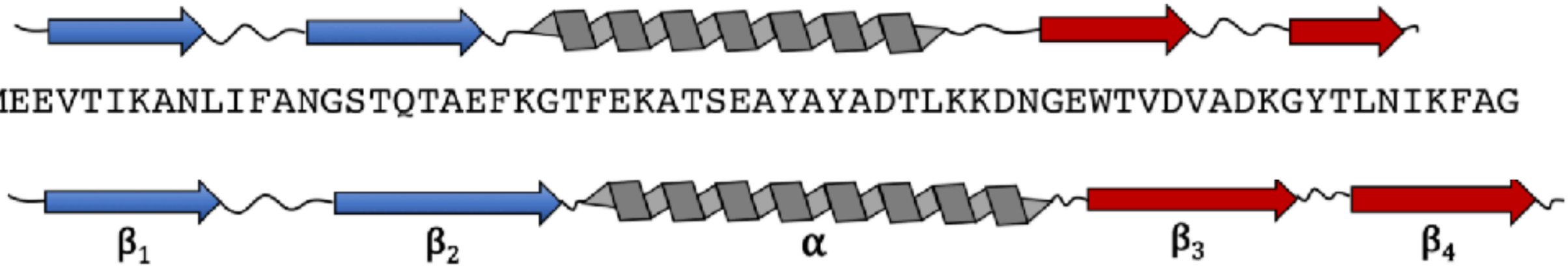


1.Perez, A., Morrone, J. A., Brini, E., MacCallum, J. L. & Dill, K. A. *Sci Adv* **2**, e1601274 (2016).

Protein G and L are two proteins with same topology and different folding pathways

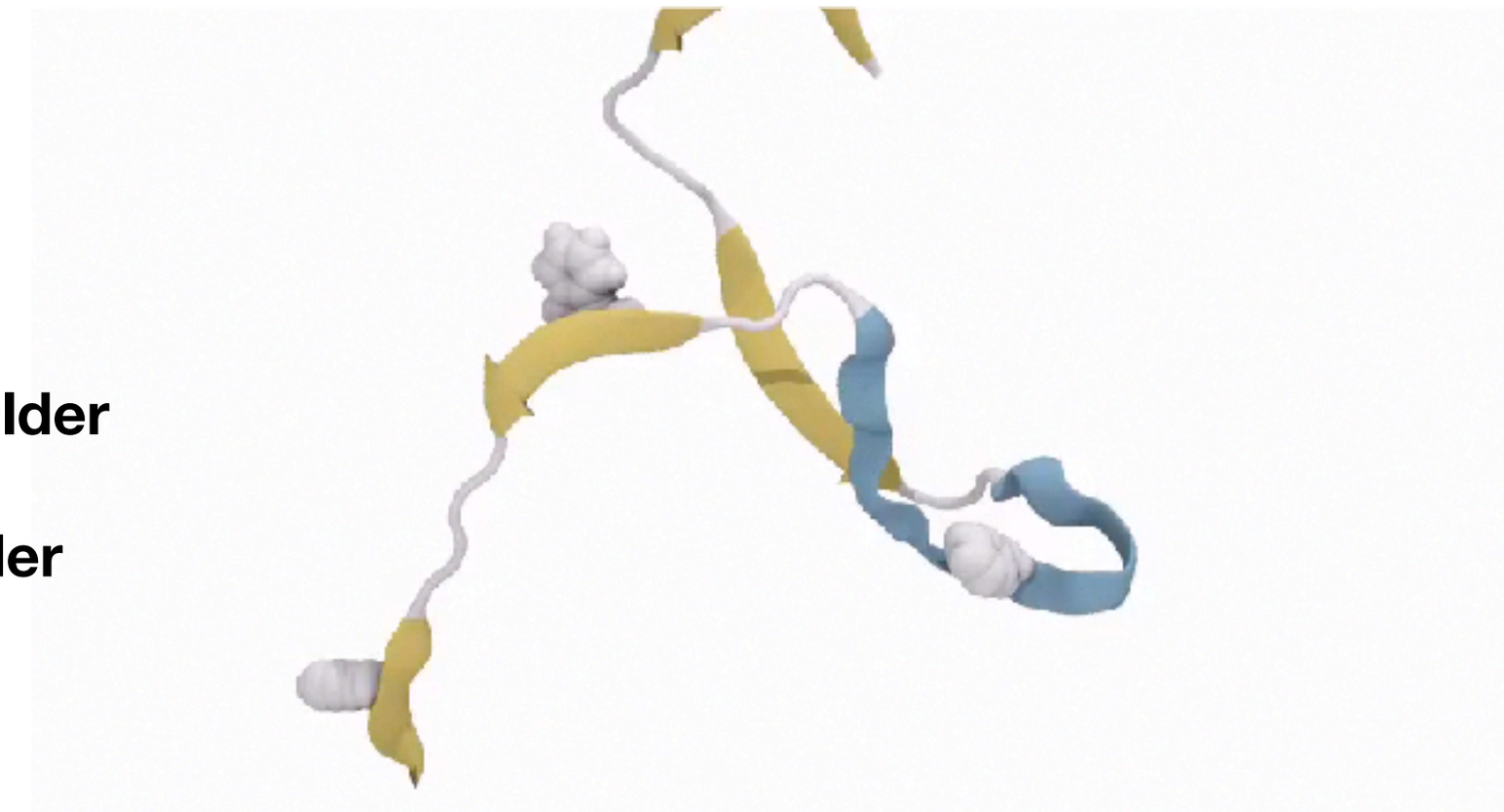
G MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

L MEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKFAG

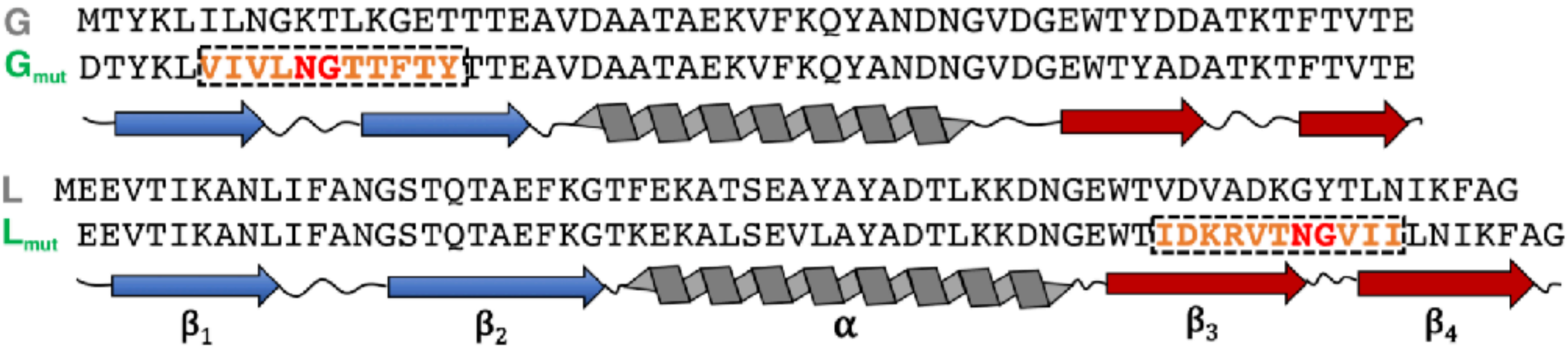


G: ms folder

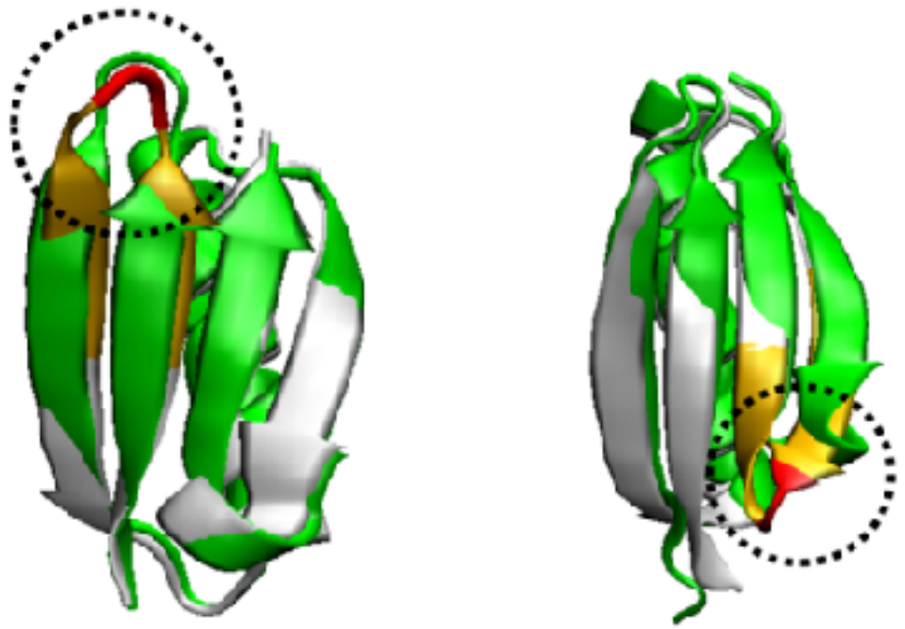
L: s folder



Protein G and L are two proteins with same topology and different folding pathways

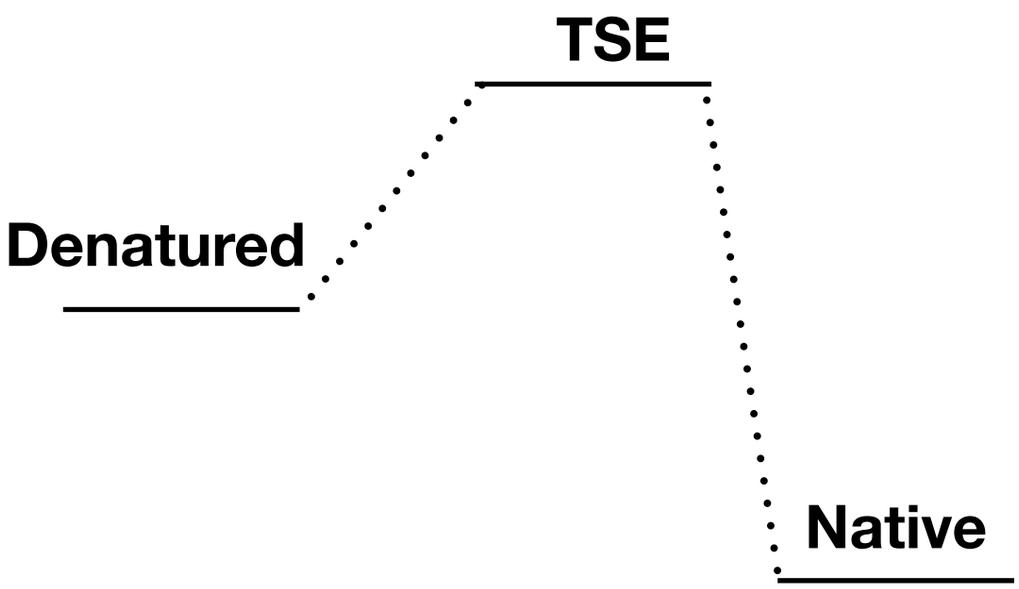


Gmut: μs folder
G: ms folder
Lmut: ms-s folder
L: s folder



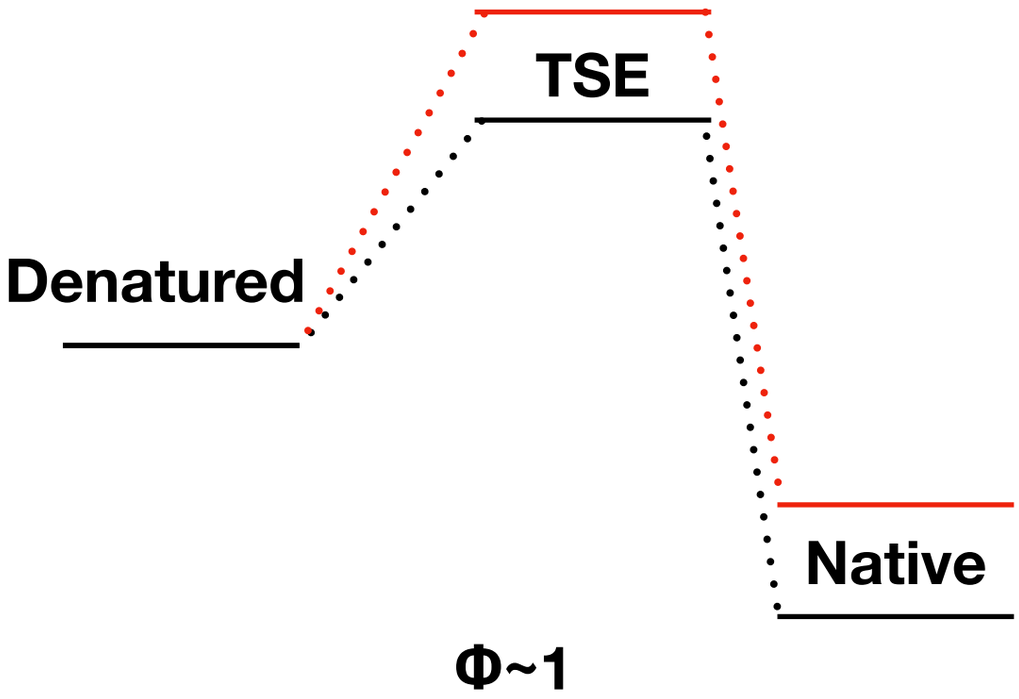
Φ-value analysis informs us of residues likely to be making native interactions in the transition state ensemble

$$\phi = \frac{(\Delta G_W^{TS \to D} - \Delta G_M^{TS \to D})}{(\Delta G_W^{N \to D} - \Delta G_M^{N \to D})} = \frac{\Delta\Delta G^{TS \to D}}{\Delta\Delta G^{N \to D}}$$



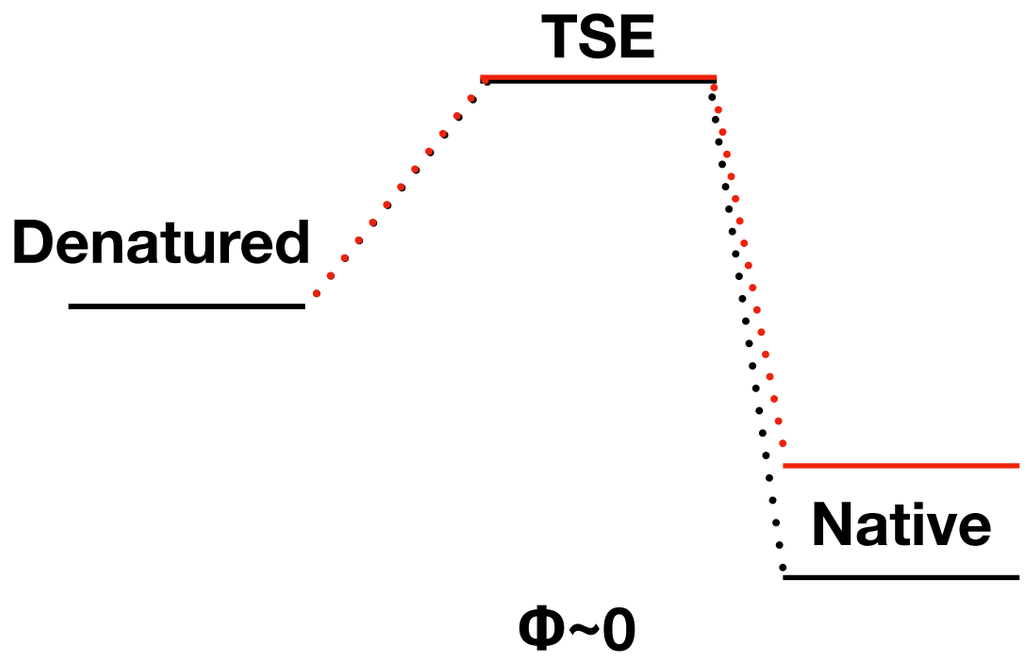
Φ-value analysis informs us of residues likely to be making native interactions in the transition state

$$\phi = \frac{(\Delta G_W^{TS\rightarrow D} - \Delta G_M^{TS\rightarrow D})}{(\Delta G_W^{N\rightarrow D} - \Delta G_M^{N\rightarrow D})} = \frac{\Delta\Delta G^{TS\rightarrow D}}{\Delta\Delta G^{N\rightarrow D}}$$



Φ~1

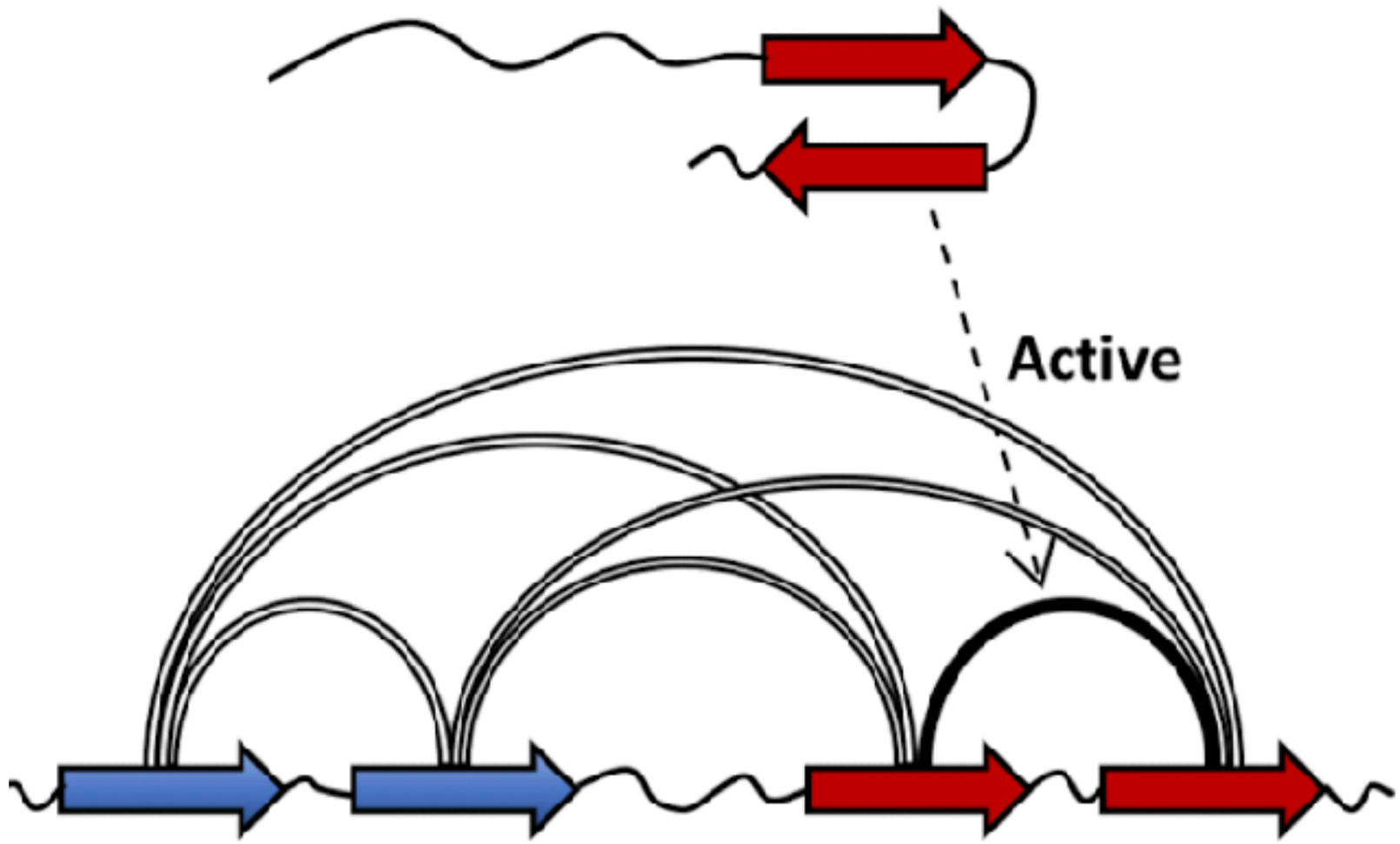
Mutant residue ordered in TS



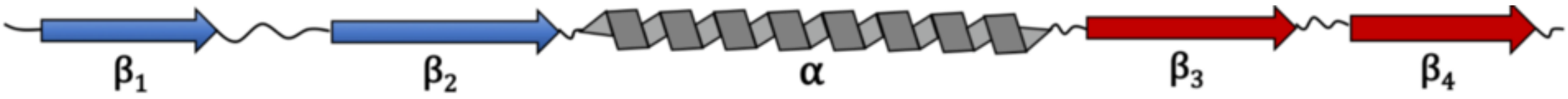
Φ~0

Mutant residue disordered in TS

What's the minimal amount of information that will focus sampling and identify folding pathways?



Protein G folds through the second hairpin, through an intermediate



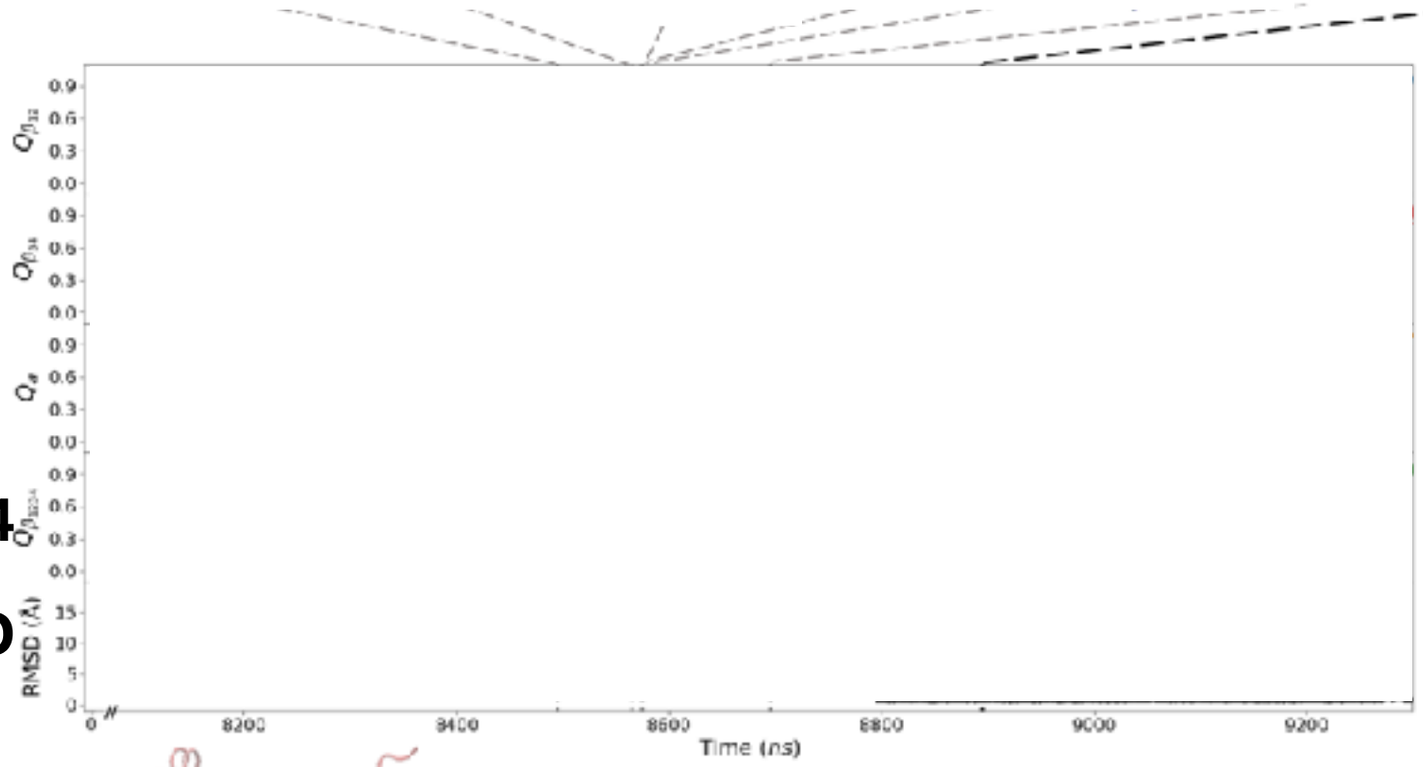
Q12

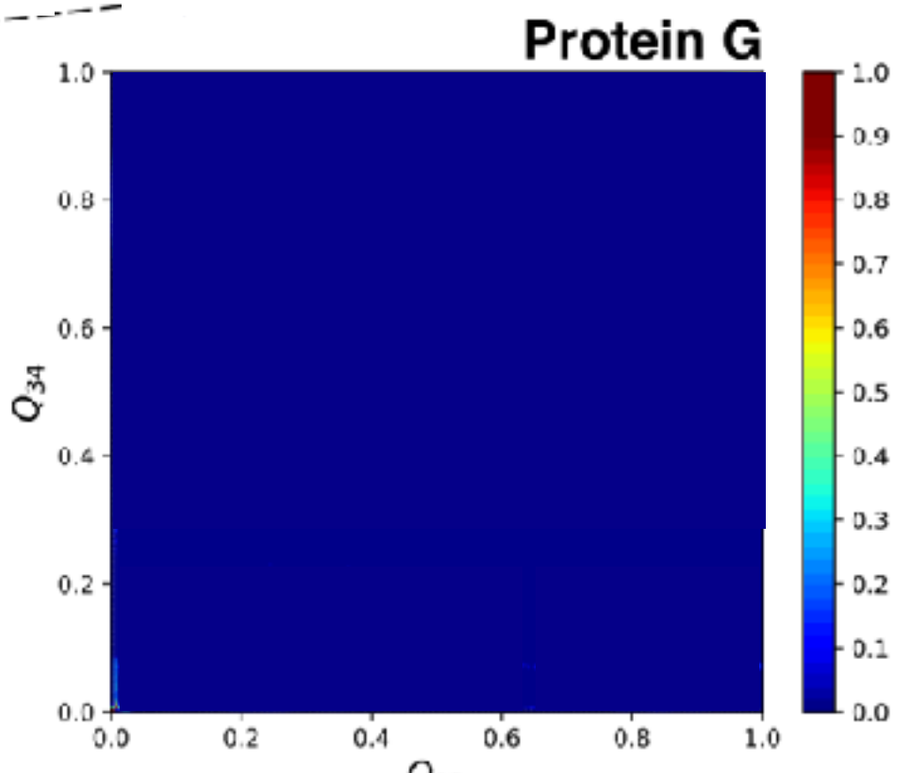
Q34

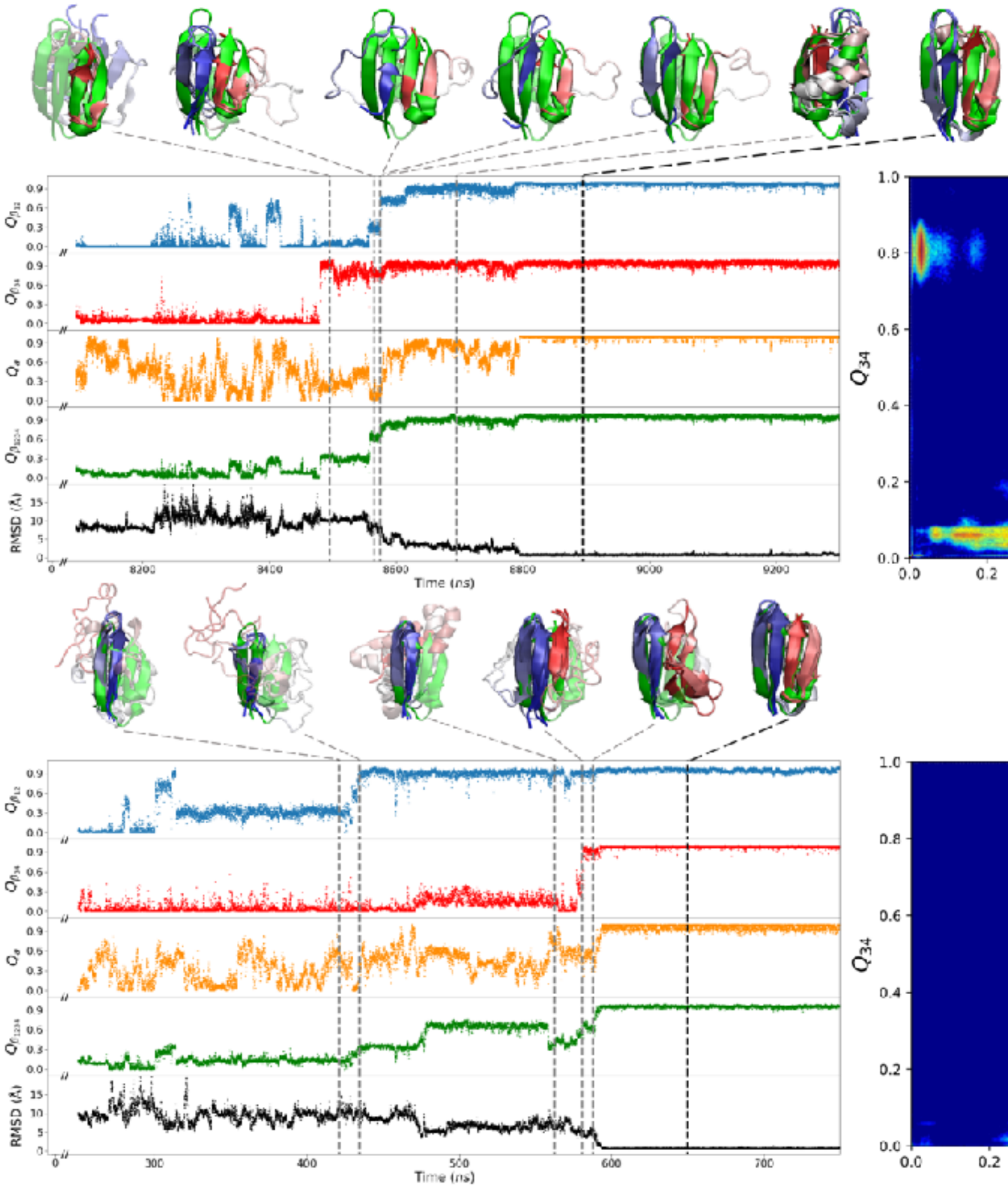
Qα

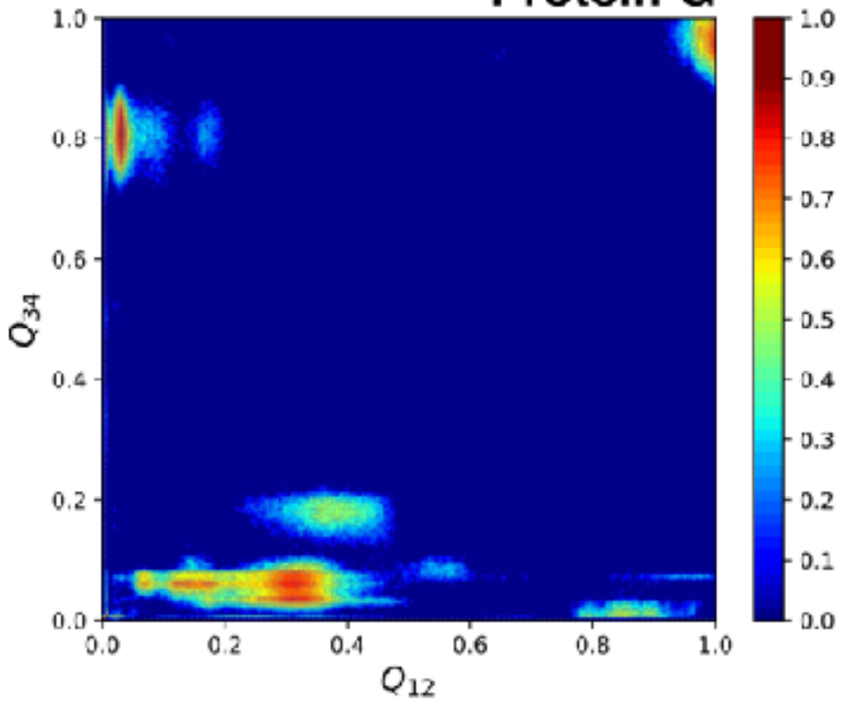
Q1234

RMSD

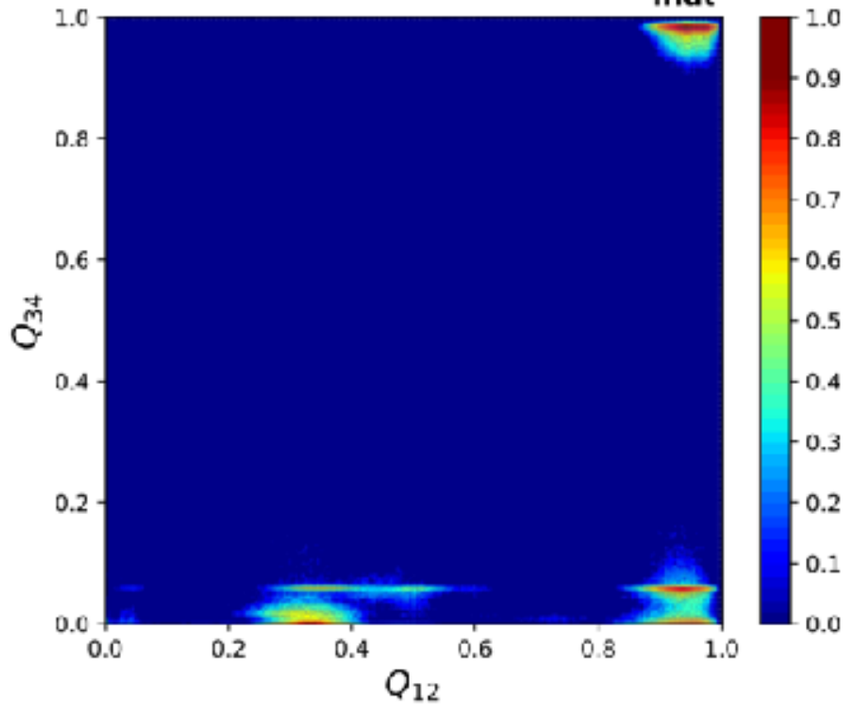




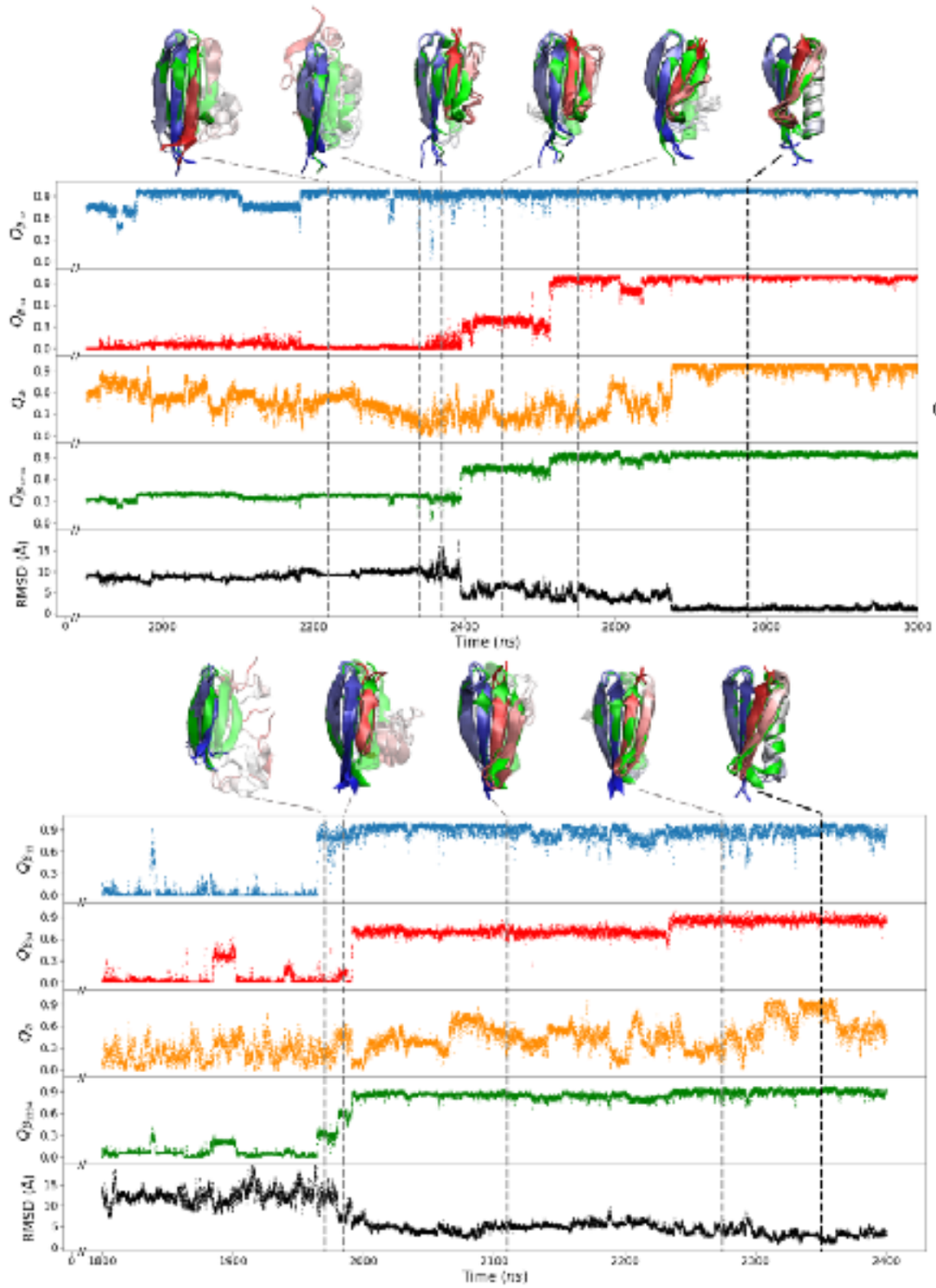

$Q_{\beta_{12}}$
$Q_{\beta_{34}}$
Q_{α}
$Q_{\beta_{1234}}$
RMSD (Å)
Time (ns)

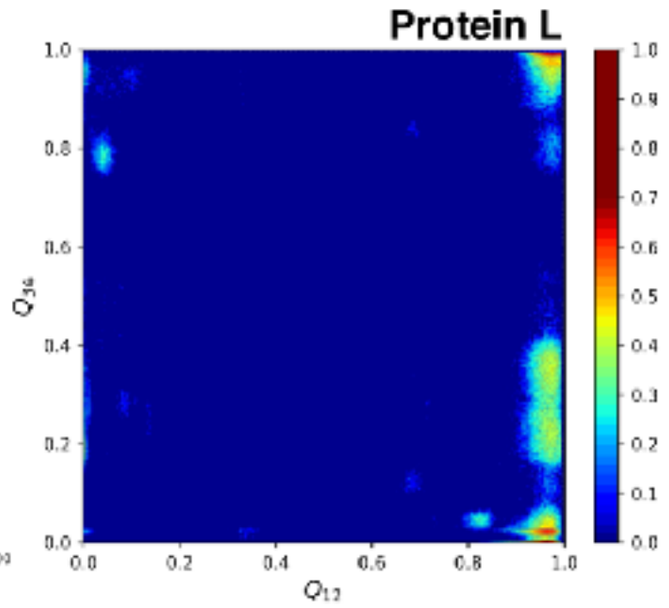

Protein G
Q_{34}
Q_{12}

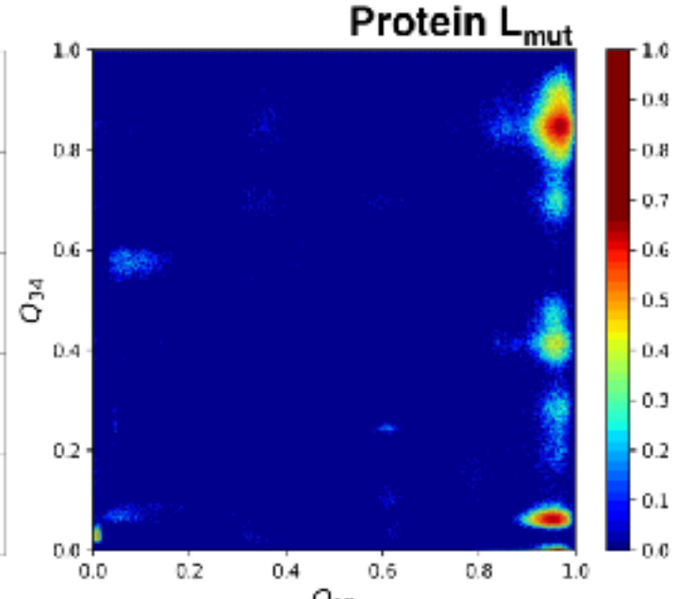
$Q_{\beta_{12}}$
$Q_{\beta_{34}}$
Q_{α}
$Q_{\beta_{1234}}$
RMSD (Å)
Time (ns)


Protein G_{mut}
Q_{34}
Q_{12}

Protein L and its mutant fold through the first hairpin, with no intermediates







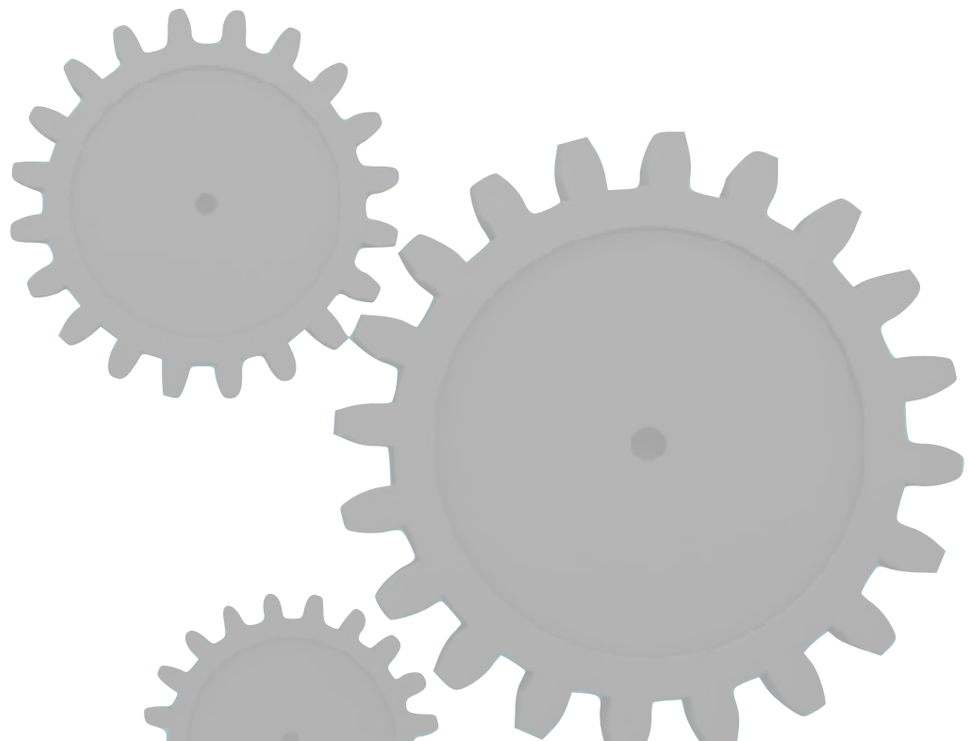
MELD captures TSE and intermediates. Is this systematic?

For protein G and G_{mut} we have few folding events

Markov State Models
Using Adaptive sampling capture the folding kinetics of Protein G mutant

Peptide Binding

Chemical Shift Perturbation Data



Intrinsically disordered peptides fold upon binding

SRLTWRVQRSQNPLKIRLTREAP

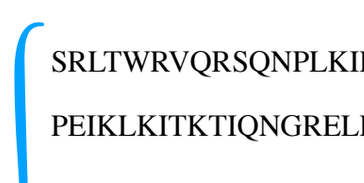
PEIKLKITKTIQNGRELFESSLCGDLLNEVQASE

NLQSSIVKFKKPLPLTQPG

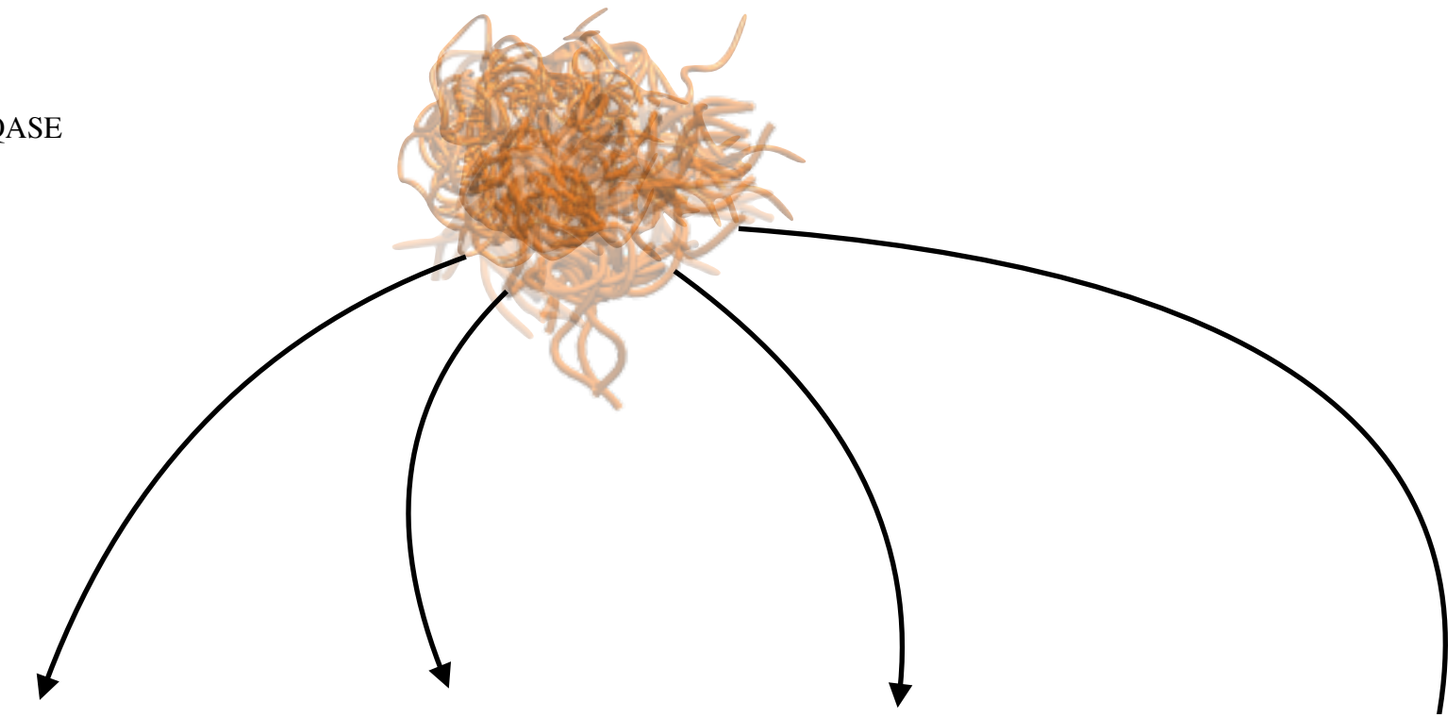
KWTLERLKRKYRN

KVAPLKIKLGGF

RSVKVKIKLGRK



ET domain of BRD3



Mondal, A. *et al*. *Biorxiv* 2021.12.31.474671 (2022)

ET is an interaction hub involved in gene regulation and virus entry

SRLTWRVQRSQNPLKIRLTREAP

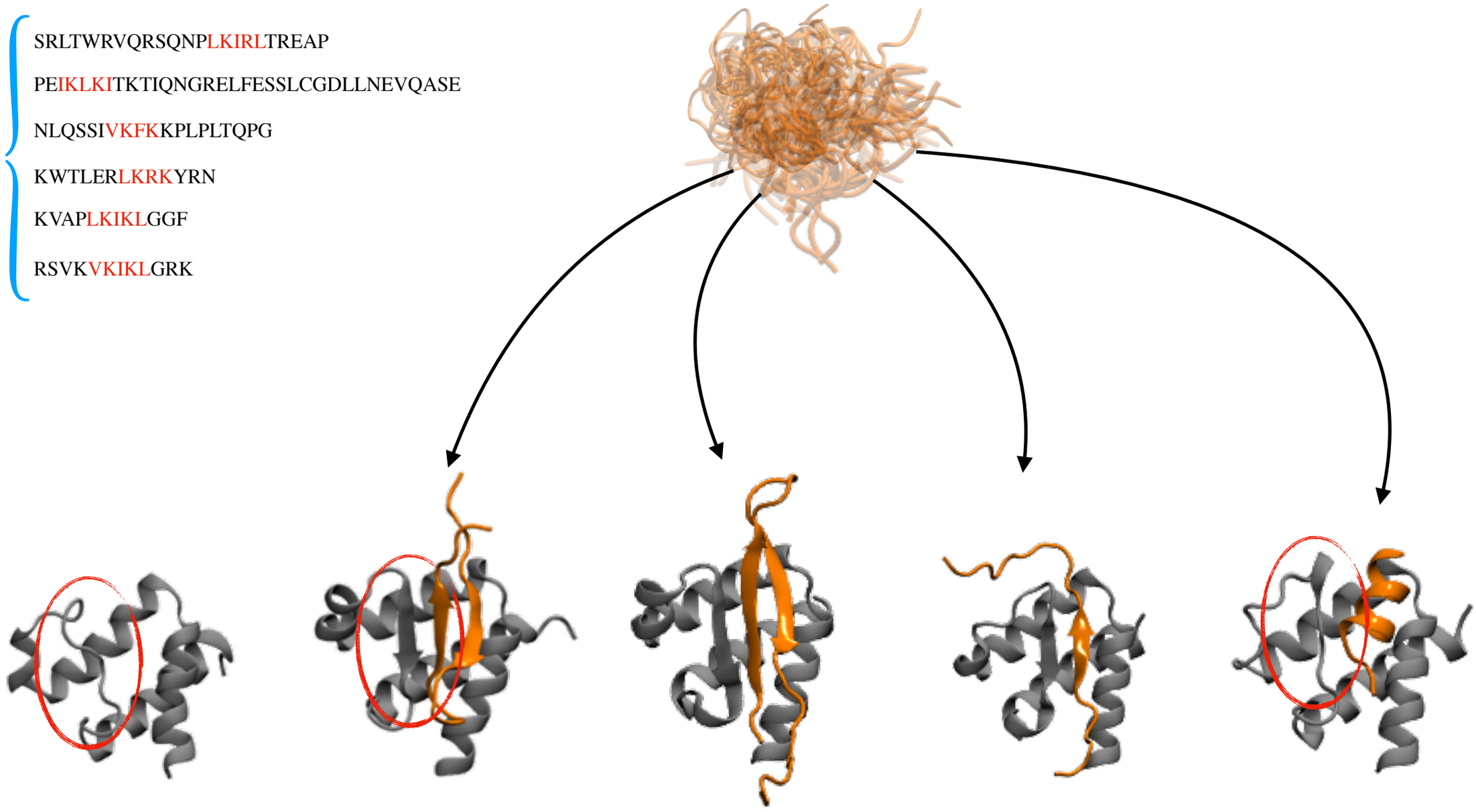
PEIKLKITKTIQNGRELFESSLCGDLLNEVQASE

NLQSSIVKFKKPLPLTQPG

KWTLERLKRKYRN

KVAPLKIKLGGF

RSVKVKIKLGRK

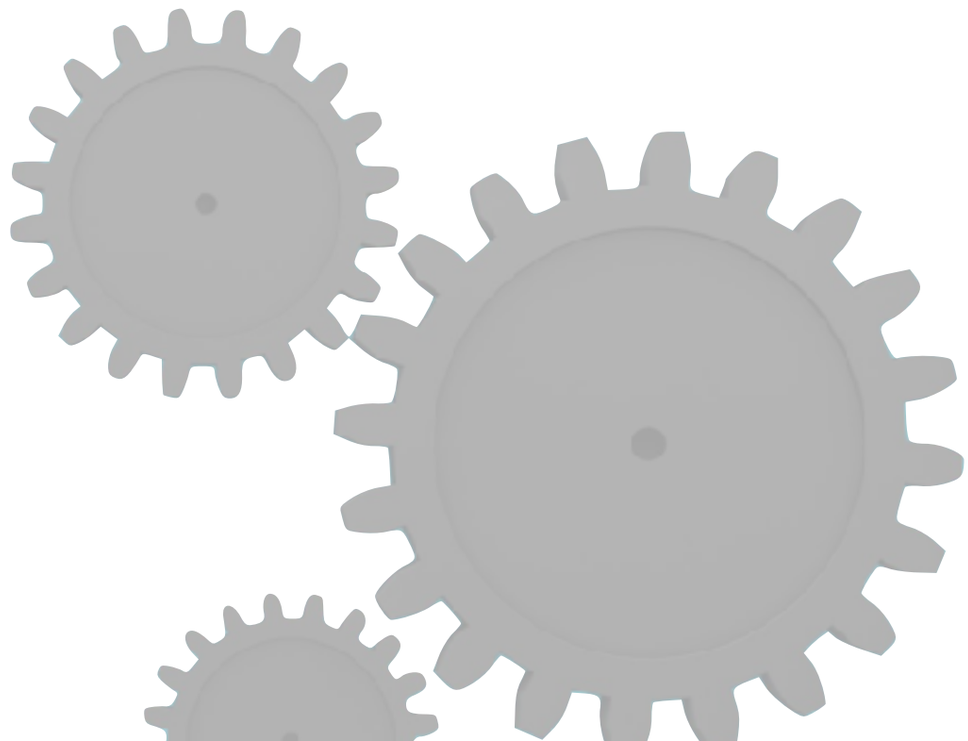


ET domain of BRD3 | BRD3-TP | BRD3-NSD3 | BRD4-LANA | BRD4-JMJD6

Mondal, A. *et al*. *Biorxiv* 2021.12.31.474671 (2022)

Peptide Binding

Chemical Shift Perturbation Data



Intrinsically disordered peptides fold upon binding

SRLTWRVQRSQNPLKIRLTREAP

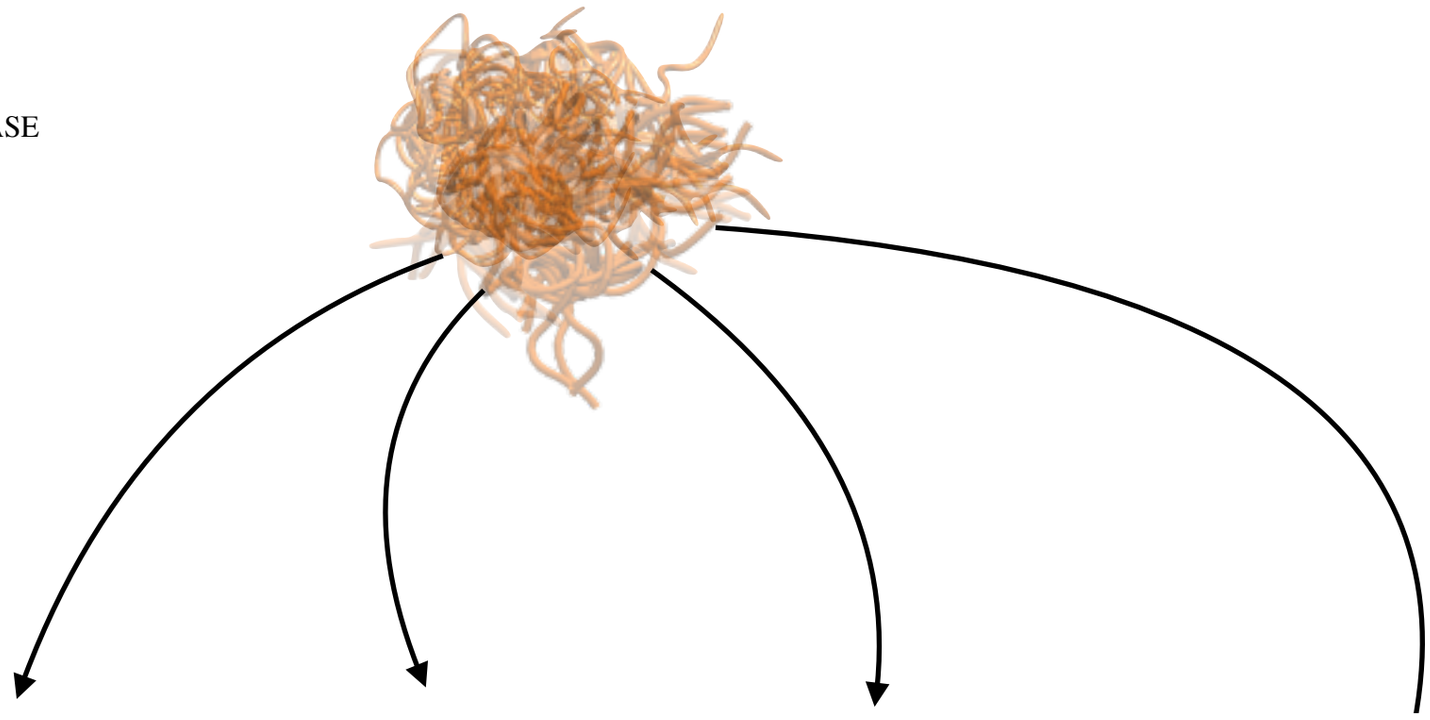
PEIKLKITKTIQNGRELFESSLCGDLLNEVQASE

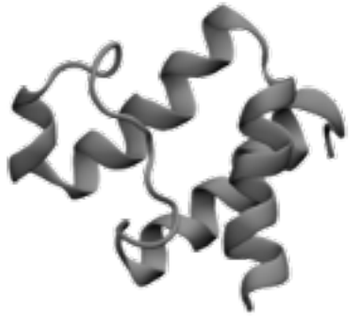
NLQSSIVKFKKPLPLTQPG

KWTLERLKRKYRN

KVAPLKIKLGGF

RSVKVKIKLGRK





ET domain of BRD3

ET is an interaction hub involved in gene regulation and virus entry

SRLTWRVQRSQNPLKIRLTREAP

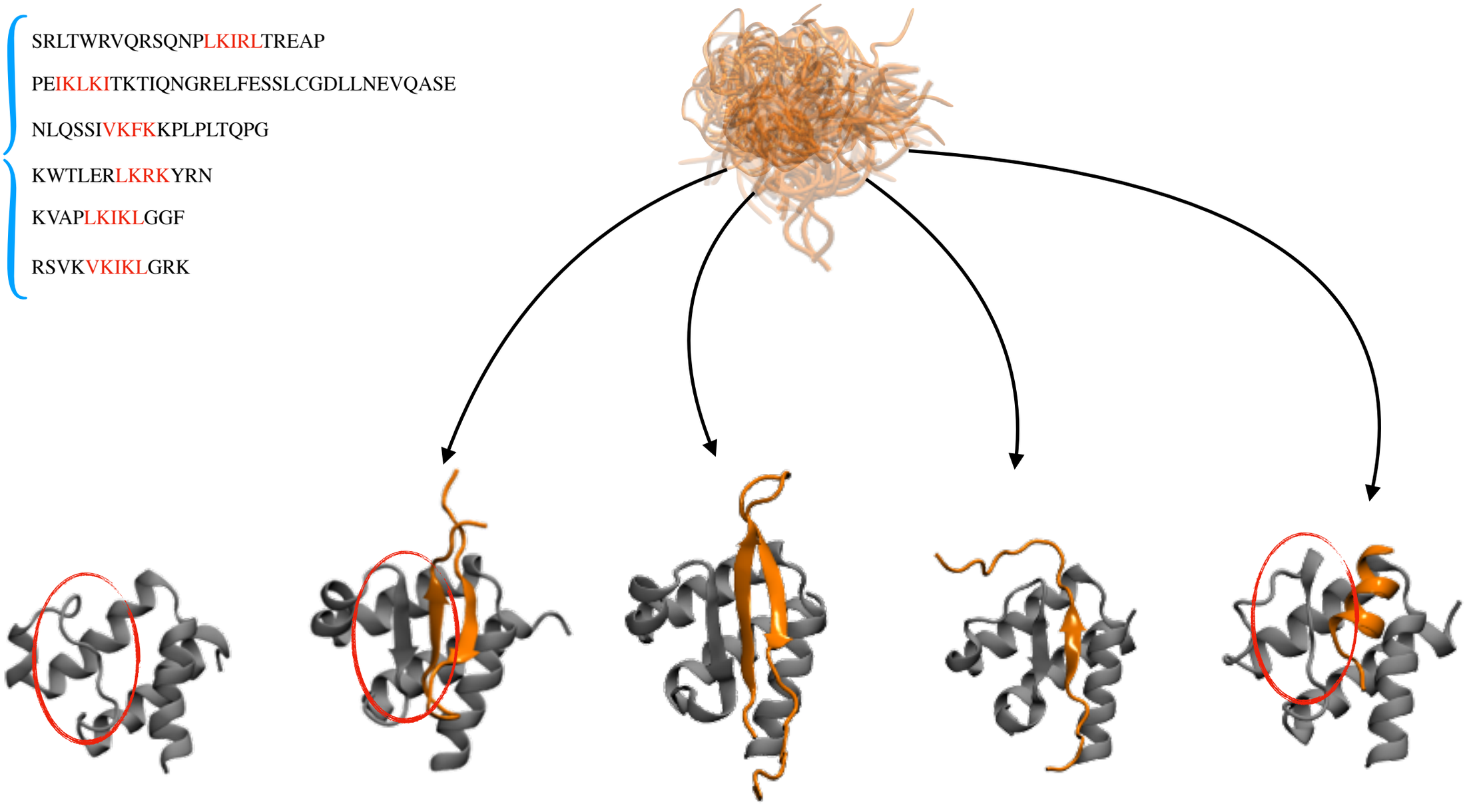
PEIKLKITKTIQNGRELFESSLCGDLLNEVQASE

NLQSSIVKFKKPLPLTQPG

KWTLERLKRKYRN

KVAPLKIKLGGF

RSVKVKIKLGRK



ET domain of BRD3

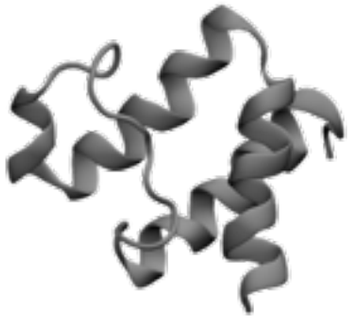
BRD3-TP

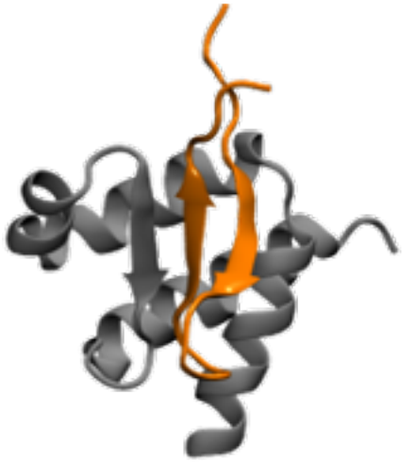
BRD3-NSD3

BRD4-LANA

BRD4-JMJD6

Peptides bind ET with a wide range of binding affinities





NMR

NOESY (protein-peptide)

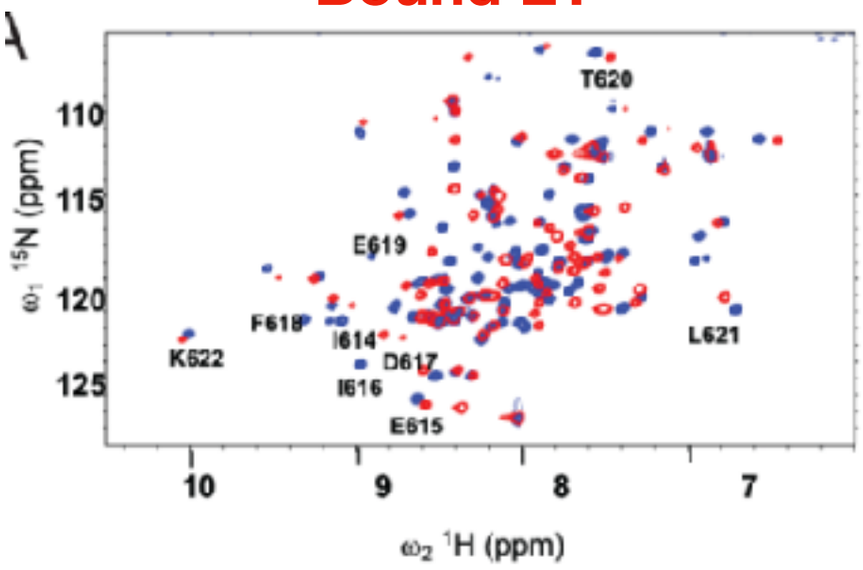
Chemical Shift Perturbation (CSP) (Protein)

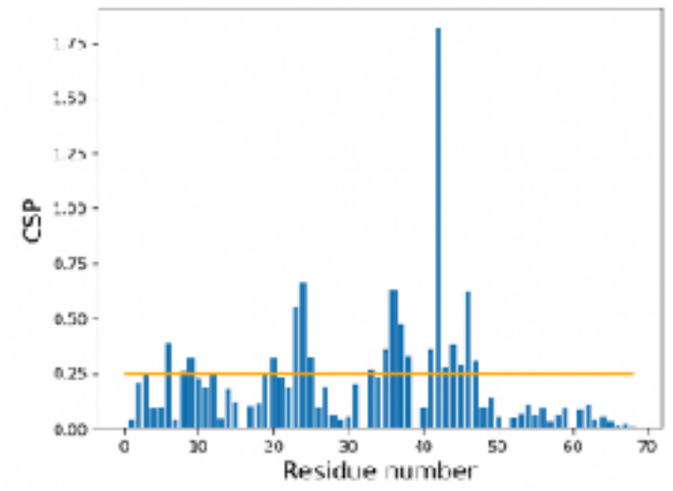
$\mathrm{Kd} \sim 10\ \mathrm{nM}$ ⟶ ~2-3 months

$\mathrm{Kd} \sim 10\ \mu\mathrm{M}$ ⟶ ~2-3 years

Chemical shift perturbation provides indirect data about where the peptide might bind

Free ET
Bound ET

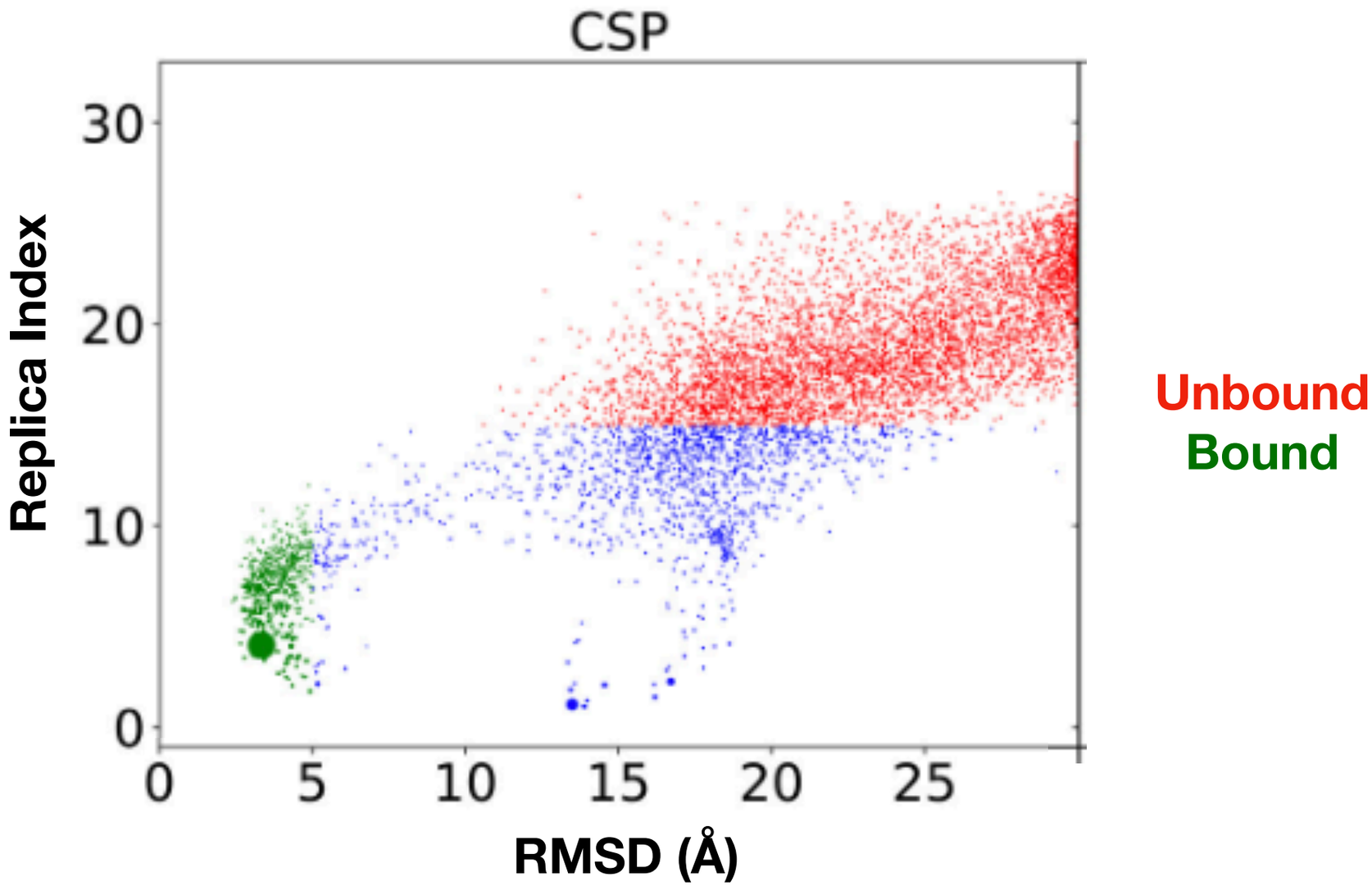




Allosteric changes?

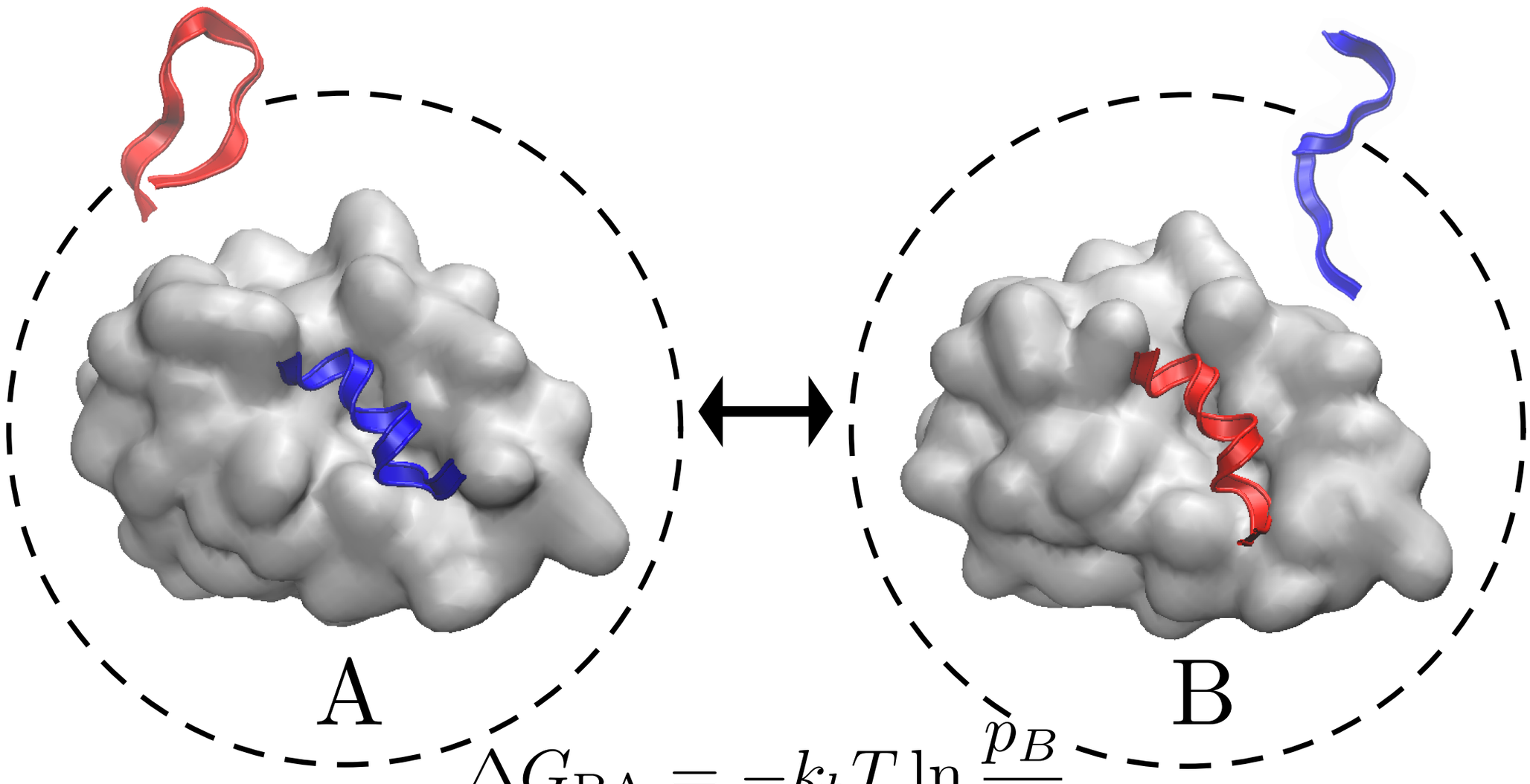
Direct contacts?

Our ensembles sample multiple bound/ misbound states and identify the native bound structure



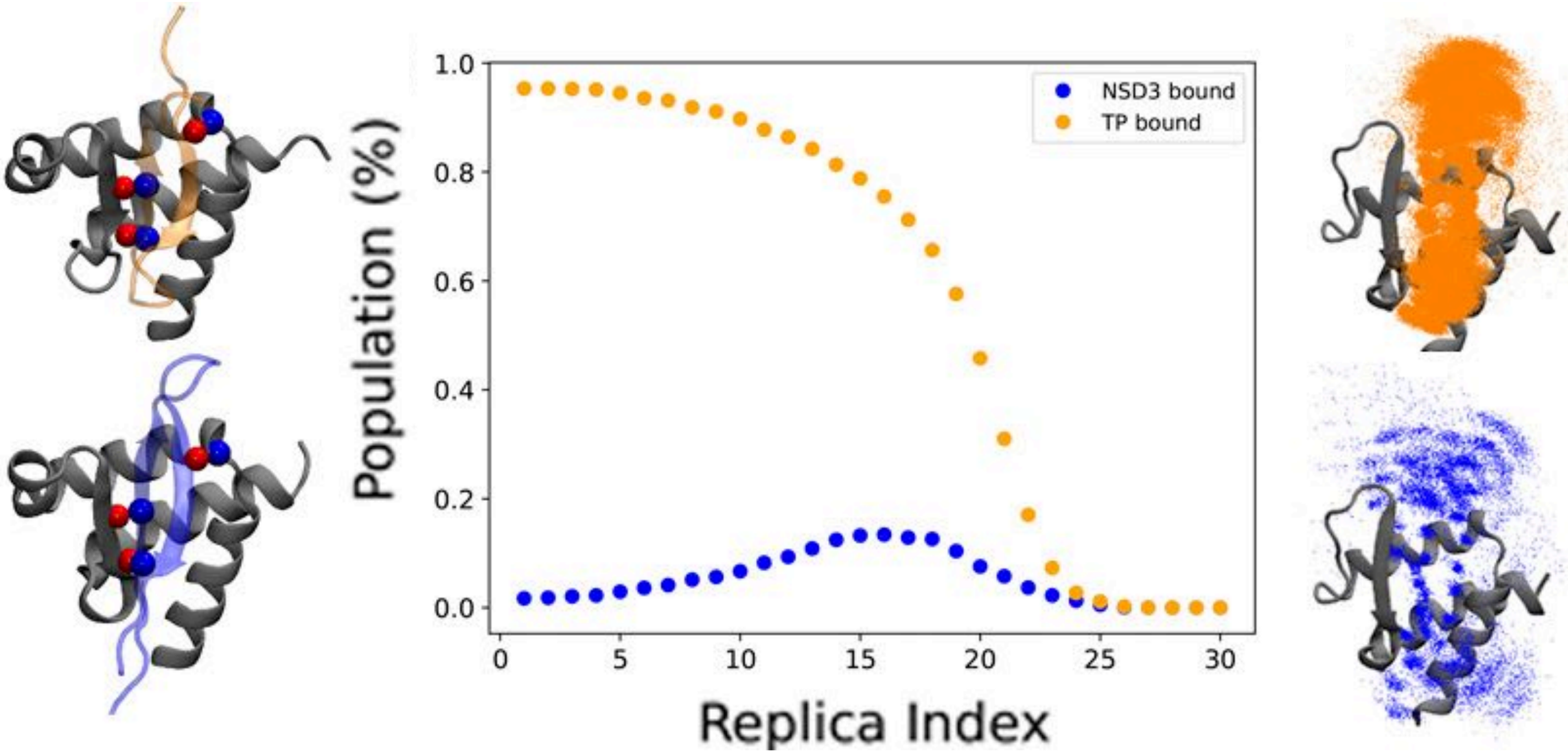
Chang, L., Perez, A*. & Miranda-Quintana, R. A*.. *Phys Chem Chem Phys* **24**, 444–451 (2021).

Competitive binding simulations help us determine binding affinity



$$\Delta G_{\mathrm{BA}} = -k_b T \ln \frac{p_B}{p_A}$$

Computed binding free energies are in agreement with ITC data



ΔΔGbind ~ 2.4 kcal/mol. **Predicted**

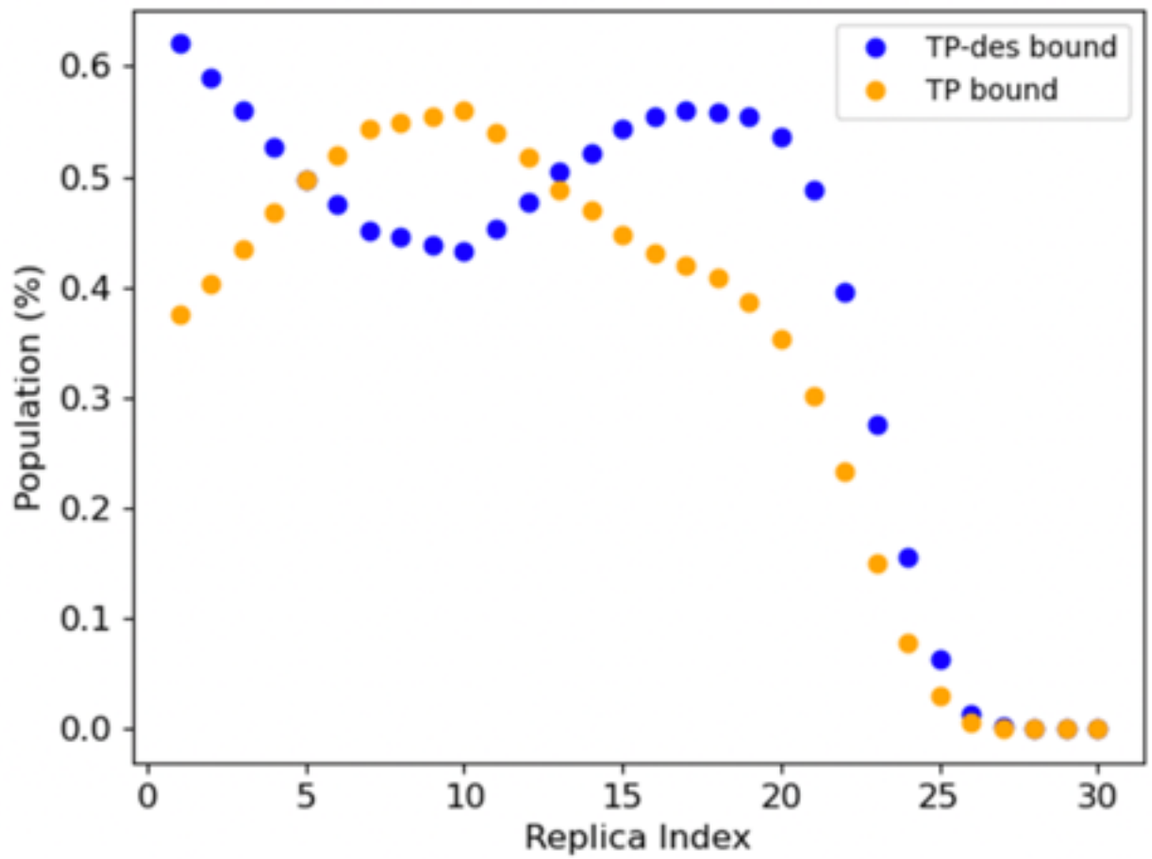
ΔΔGbind ~ 2.6 kcal/mol. **Experiment**

Mondal, A. *et al*. *Biorxiv* 2021.12.31.474671 (2022) doi:10.1101/2021.12.31.474671.

We used the lessons learnt from folding routes to design a novel peptide inhibitor

Design

Murine Leukemia Virus



Structural Biology needs physical modeling

- Learn about why/how
- Lead to design principles (smooth landscapes)
- Transferable to new materials
- Simpler and slower folding pathways for protein L over G
- Bridges experiments and atomistic structures

Thanks!

XSEDE

HiPerGator 3.0

https://perez.chem.ufl.edu