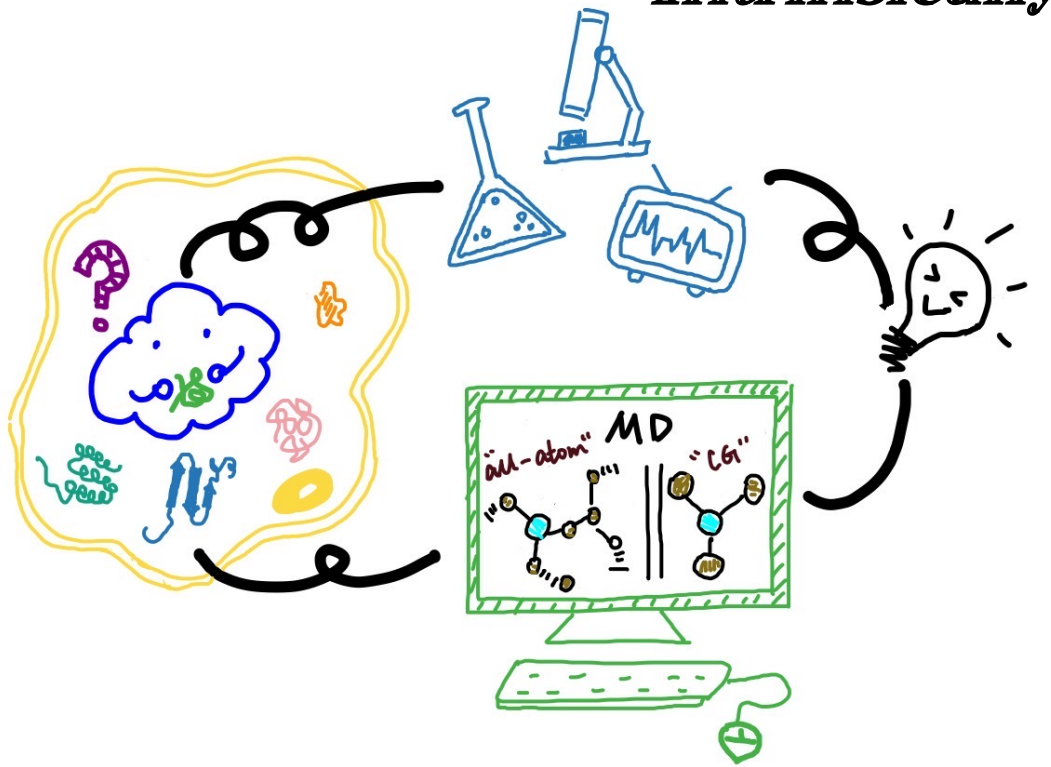
University of Massachusetts Amherst
College of Natural Sciences
Department of Chemistry

Integration of Experiments and Multiscale Simulations to Study

Intrinsically Disordered Proteins



A Prospectus Presented
By

Yumeng Zhang

Jianhan Chen Research Group

April 6, 2021

>_

[Yumeng@Prospectus]$ ls

Part1: Introduction&Significance **Part2: Specific Aims**

[Yumeng@Prospectus Part2: Specific_Aims]$ ls

Aim1. Integration of MD simulations and experiments for IDP studies

Aim2. Advanced methods for multi-scale simulations on IDPs

[Yumeng@Prospectus Aim1*]$ ls

Subaim1a. IDP specific tight interactions: SPIN-NTD/MPO

Subaim1b. IDP dynamic interactions: p53-NTD/CypD

Subaim1c. IDP enzymatic interactions: Flaviviral proteases

[Yumeng@Prospectus Aim2*]$ ls

Subaim2a. Advanced sampling method: REST3

Subaim2b. Optimized force field: HyRes*

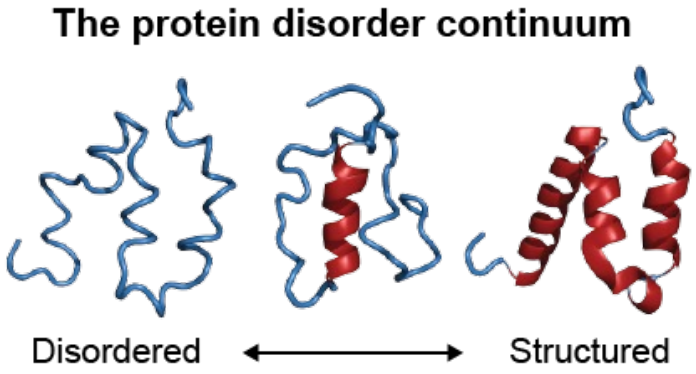
>_

[Yumeng@Prospectus]$ Part1: Introduction & Significance

[Yumeng@Prospectus]$ Part2: Specific Aims

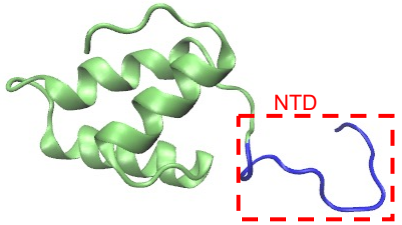
[Part1]$ Introduction

- **Intrinsically disordered proteins (IDPs)**

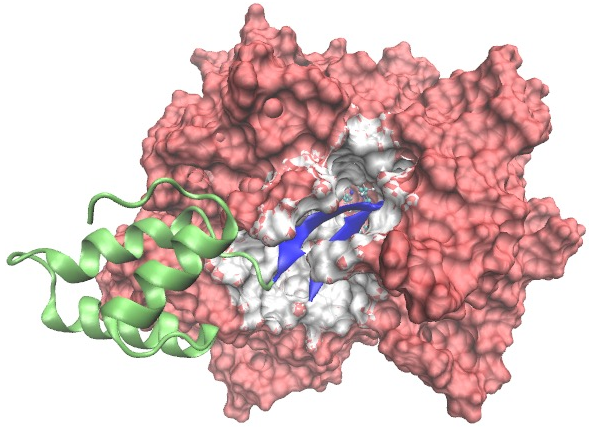


- ❖ Lack a fixed or ordered three-dimensional structures.
- ❖ Range from totally unstructured to partially structured.
- ❖ Rich in polar and charged residues.
- ❖ Large and functionally important class of proteins

- **IDPs participates in diverse cellular processes**
- Regulations
  - Specific tight bindings

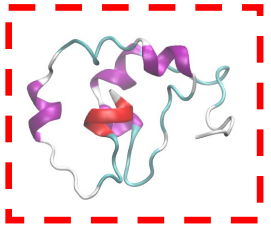


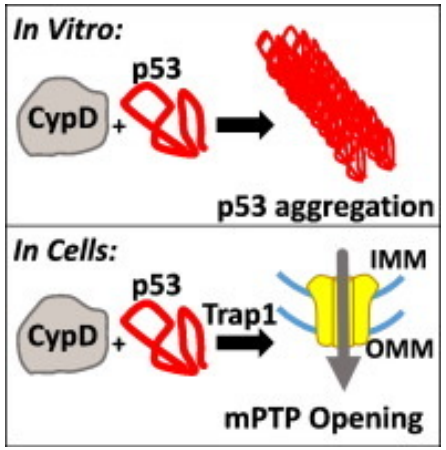
(SPIN)



(Complex: MPO/SPIN)

  - Dynamic bindings





  - Signaling, Recognition

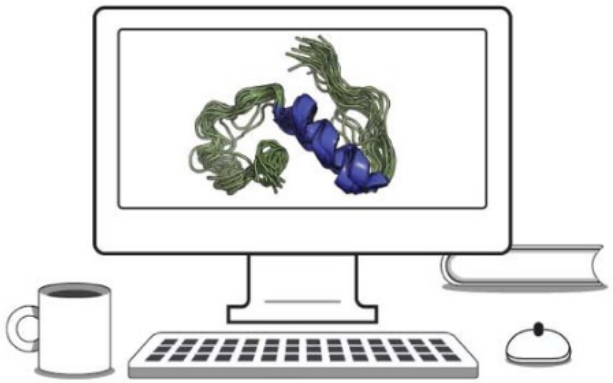
doi.org/10.1016/j.jmb.2016.08.001

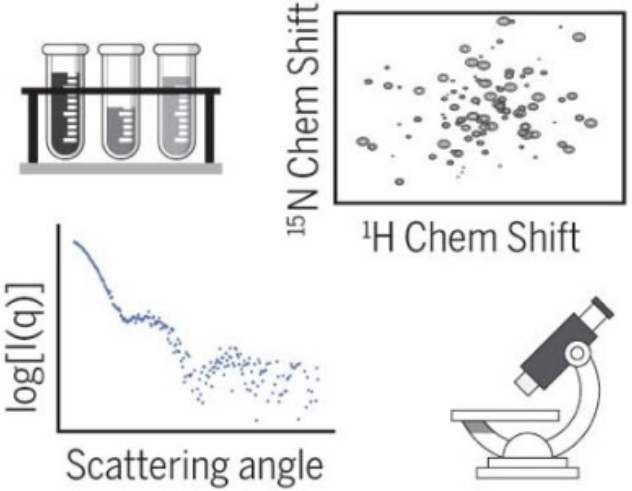
[Part1]$ Significance & Innovation

- Experiments
- Challenges
  - IDP high structural heterogeneity.
  - IDP high dynamics.
  - Timescales of IDP conformational fluctuations.

A unique opportunity for computational modeling!

- MD simulations

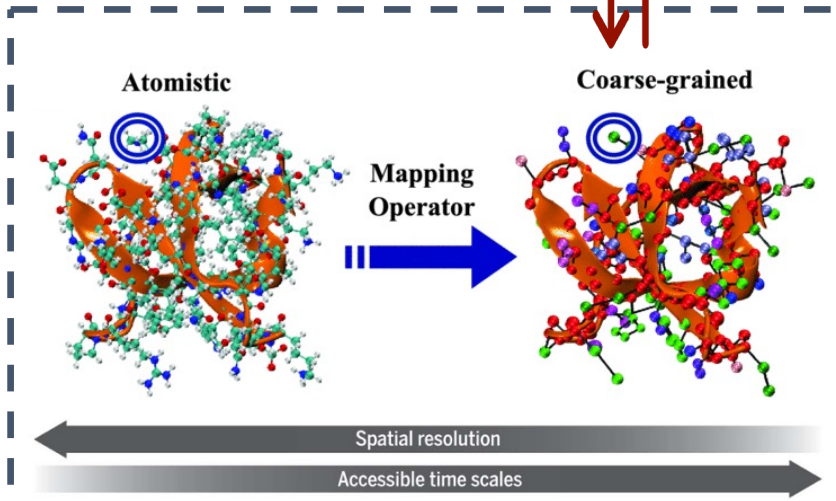




Increasing resolution

X-ray, NMR
Cryo-EM
SAXS,SANS,
EPR,FRET

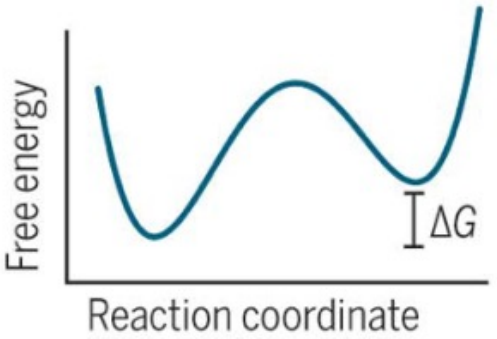
Integration

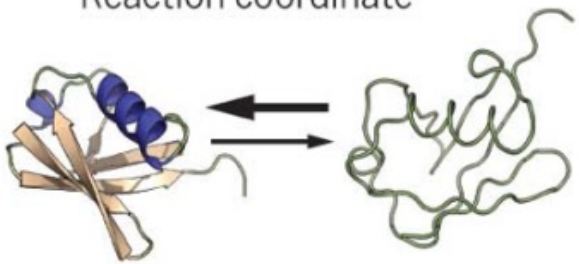


Advanced methods

- Advanced sampling methods
- Optimized force fields

Multi-scale MD simulations





- Thermodynamics
- Kinetics
- Mechanisms
- …

DOI: 10.1126/science.aat4010

>_

[Yumeng@Prospectus]$ Part2: Specific Aims

Aim1. Integration of MD simulations and experiments for IDP studies

Aim2. Advanced methods for multi-scale simulations on IDPs

[Yumeng@Prospectus Aim1]$

Subaim1a. IDP specific tight interactions: SPIN-NTD/MPO

Subaim1b. IDP dynamic interactions: p53-NTD/CypD

Subaim1c. IDP enzymatic interactions: Flaviviral proteases

METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS*

THREAT LEVEL **SERIOUS**

323,700 Estimated cases in hospitalized patients in 2017

10,600 Estimated deaths in 2017

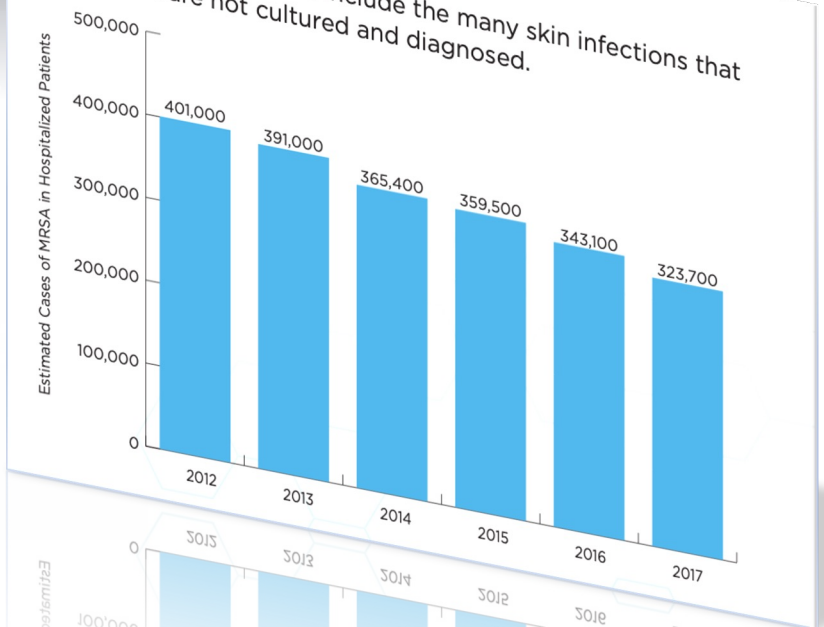
$1.7B Estimated attributable healthcare costs in 2017

Staphylococcus aureus (S. aureus) are common bacteria that spread in healthcare facilities and the community. Methicillin-resistant *S. aureus* (MRSA) can cause difficult-to-treat staph infections because of resistance to some antibiotics.

- **Although several treatments are still available, MRSA has become resistant to many first-line antibiotics.**
- **While MRSA infections overall are dropping, progress to prevent MRSA bloodstream infections in healthcare is slowing.**

CASES OVER TIME

Cases represented do not include the many skin infections that happen, but are not cultured and diagnosed.

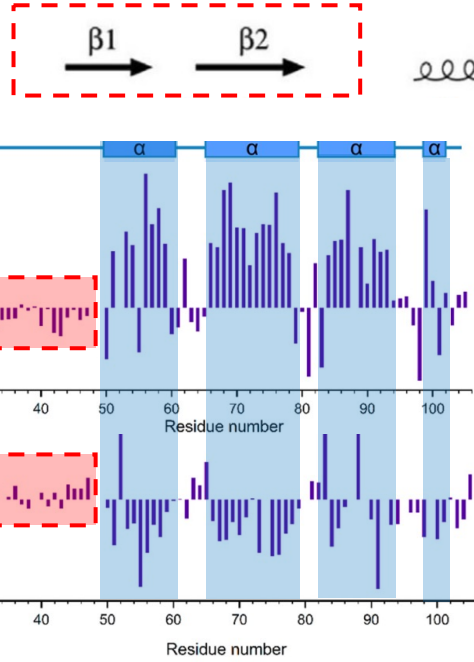


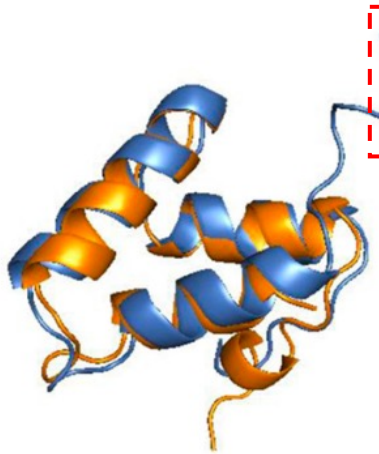
U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf

[Subaim1a]$ IDP Specific Tight Interactions: SPIN-NTD/MPO

- **SPIN** (Staphylococcal Peroxidase Inhibitor)
  - α-helical bundle
  - Intrinsic disordered N-terminal





SPIN
(Unbound State, NMR)

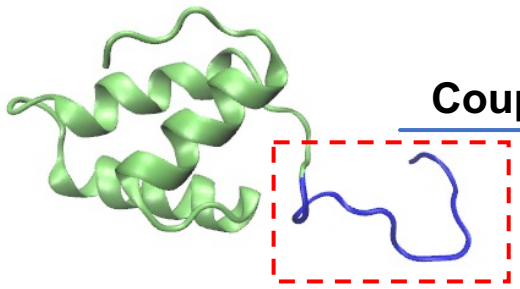
SPIN
(Co-crystal structure, X-ray)

EXPM

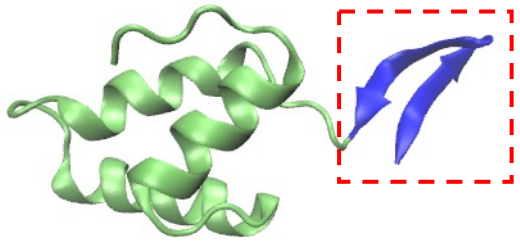
MD

Coupled binding and folding

MPO



Disordered NTD in unbound state *doi/10.1074/jbc.RA117.000134*

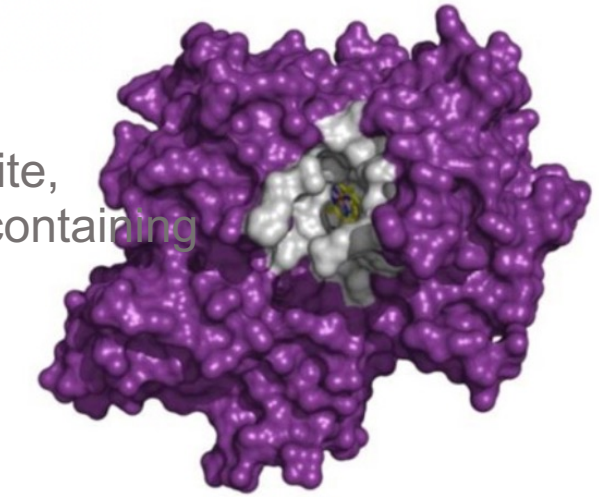


β-hairpin structured NTD in SPIN/MPO co-crystal

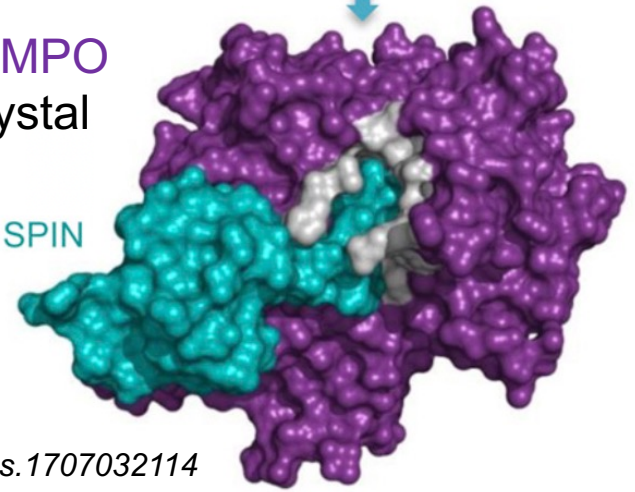
- **Specific Binding (SPIN/MPO)**

Enzyme: Myeloperoxidase (MPO)

Active site,
HEME containing



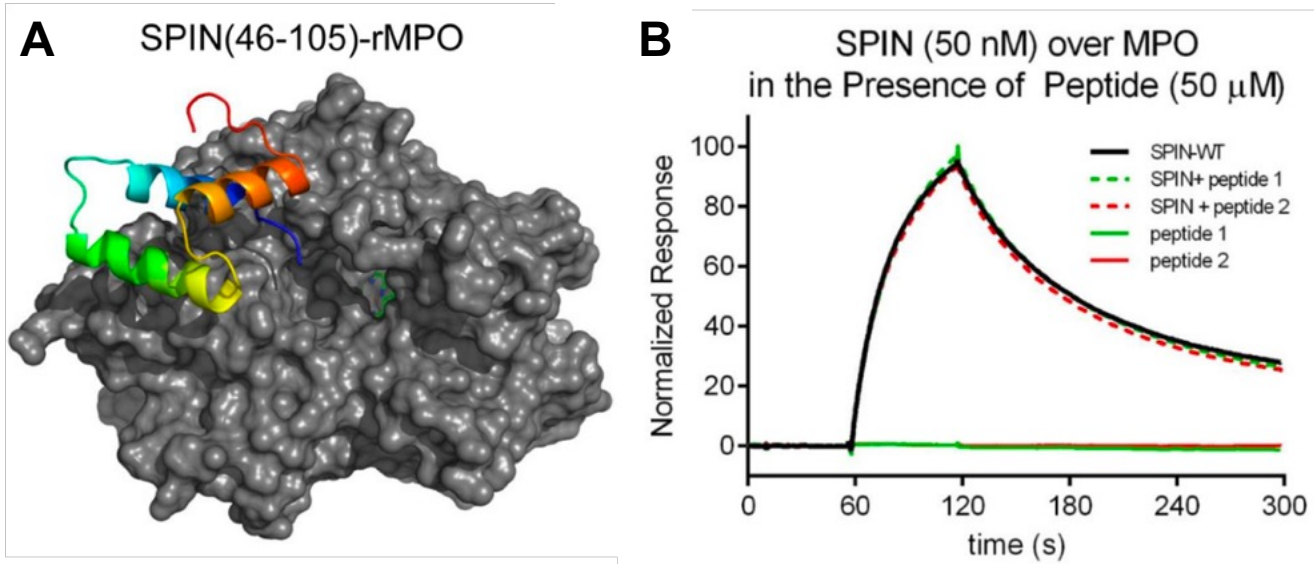
SPIN/MPO
Co-crystal



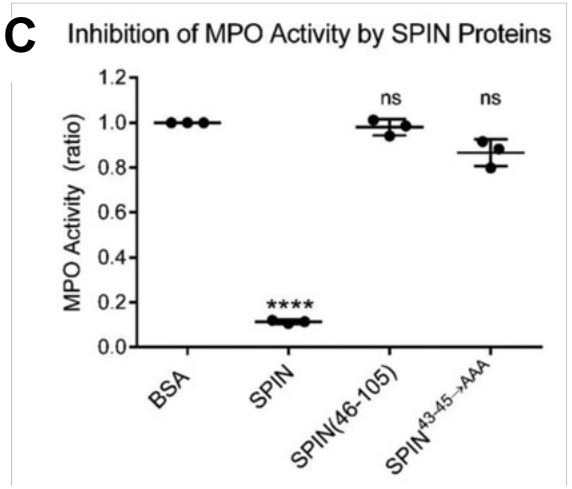
doi/10.1073/pnas.1707032114

[Subaim1a]$ SPIN Functional Regions (EXPM)

- **Binding Determinator: α-helical bundle**



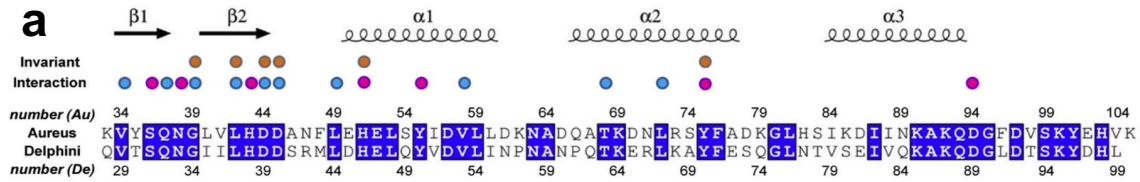
- **Inhibitory Functional Domain: NTD**



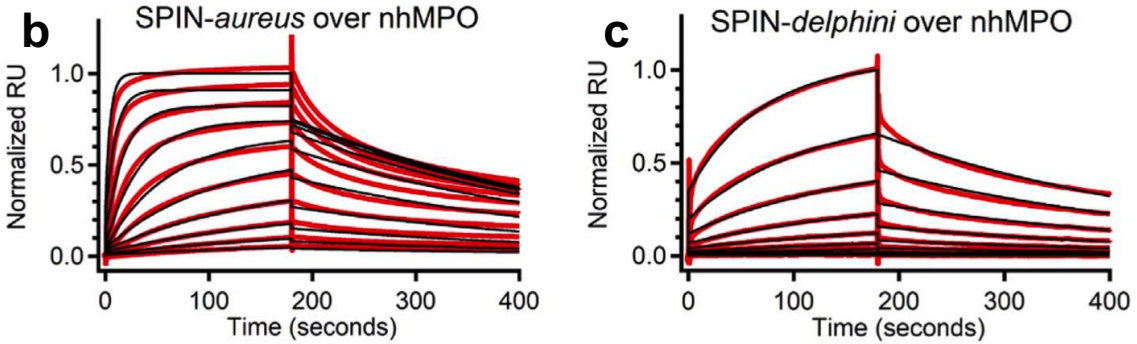
- SPIN's **binding** to MPO is controlled by **α-helix bundle** totally.
- SPIN's **inhibition** to MPO is functioned by **NTD** only.
- Mutations on SPIN-NTD will lead to significant weaker inhibitory ability on SPIN.

doi/10.1074/jbc.RA117.000134

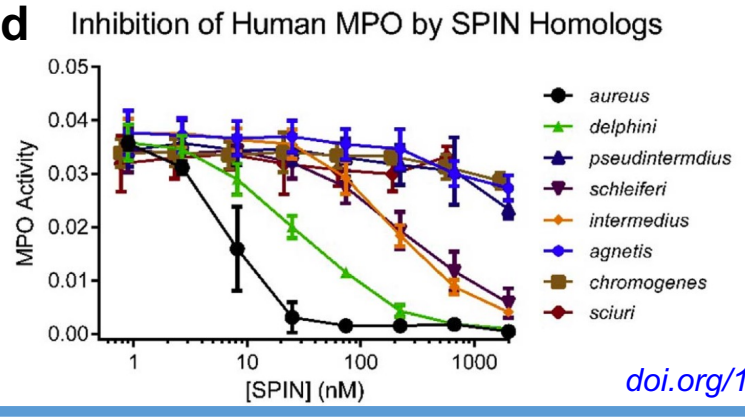
- **Inhibition discrepancies in SPIN homologs**
- **High Structural identity** of *S.aureus* and *S.delphini*



- **Similar binding affinity** of *S.aureus* and *S.delphini*



- **Weaker inhibition ability** of *S.delphini*



doi.org/10.1016/j.abb.2018.03.007

❖ Mystery:

Two SPIN have **identical structures** but **different inhibitory efficacies.**

number (Au)	34					39					44								
Aureus	K	V	Y	S	Q	N	G	L	V	L	H	D	D	A					
Delphini	Q	V	T	S	Q	N	G	I	I	L	H	D	D	S					
number (De)	29					34					39								

$N_{intra}^{S.aureus-\beta1} = 12$ $N_{inter}^{S.aureus/MPO} = 35$

$N_{intra}^{S.delphini-\beta2} = 10$ $N_{inter}^{S.delphini/MPO} = 36$

Accessible NTD surface Ares ($\mathring{A}^2$):

S.aureus = 1626.74

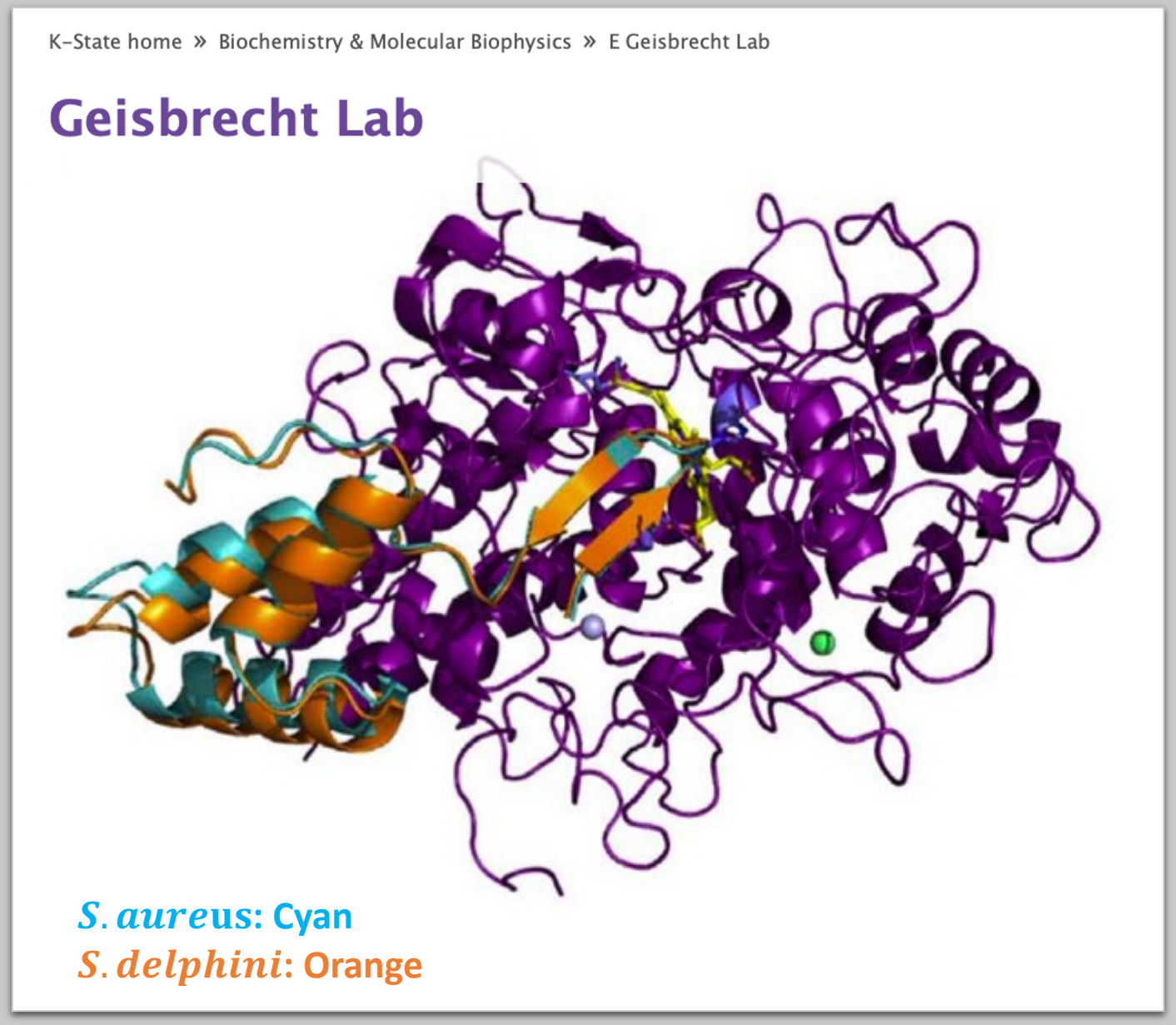
S.delphini = 1588.49

❖ We want to have crucial synergistic interaction insights...

MD Simulations Can offer!

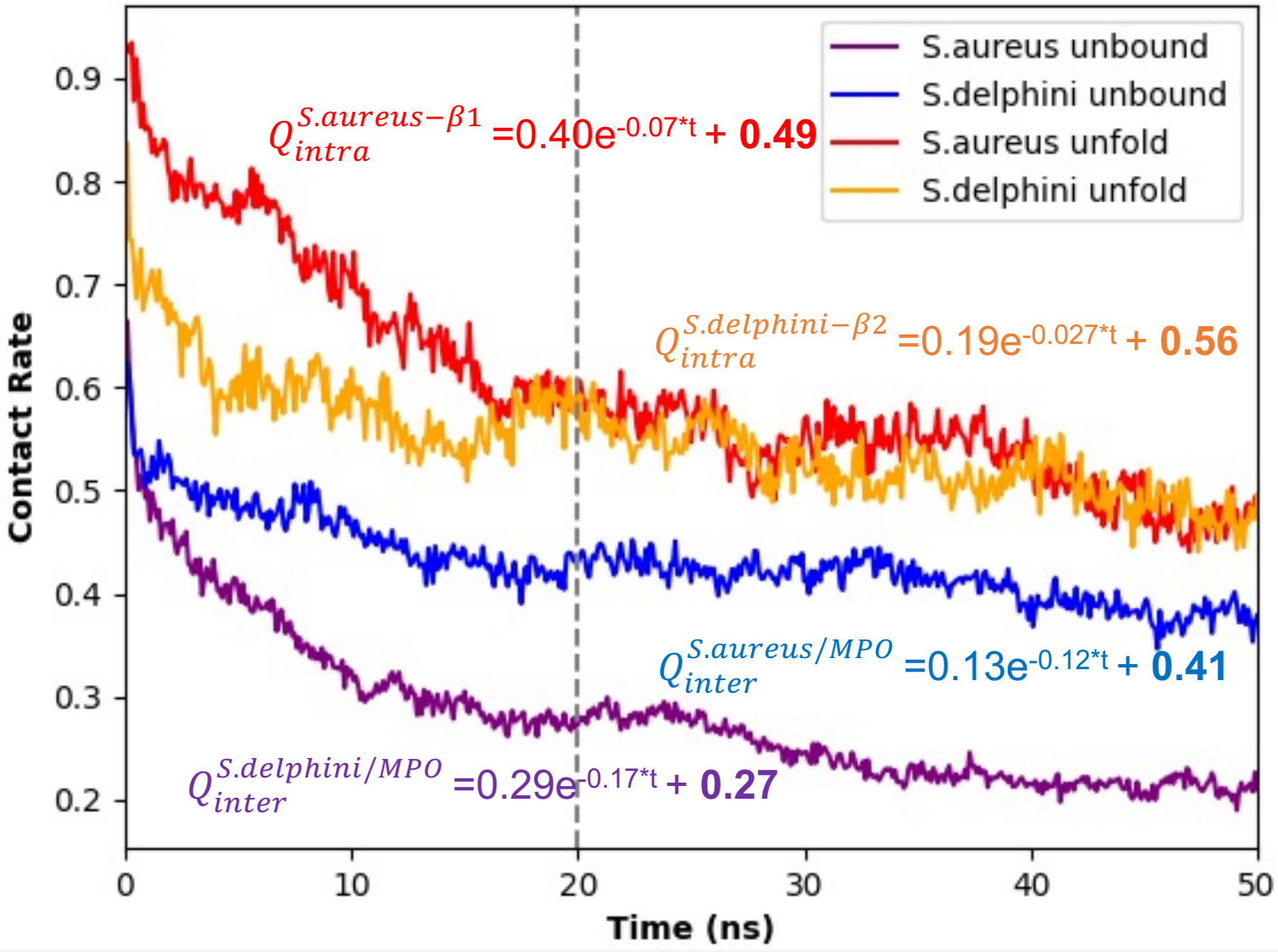
Atomistic model: High resolution.

CG model: Non-native and long-time scale details.



[Subaim1a]$ Atomistic Simulations on SPIN/MPO

- **Dissociation simulations on SPIN/MPO (450 K)**



Setup:

- Temperature: 450 K.
- NPT ensemble, atomistic model. 40 replicas simultaneously and independently.

Results:

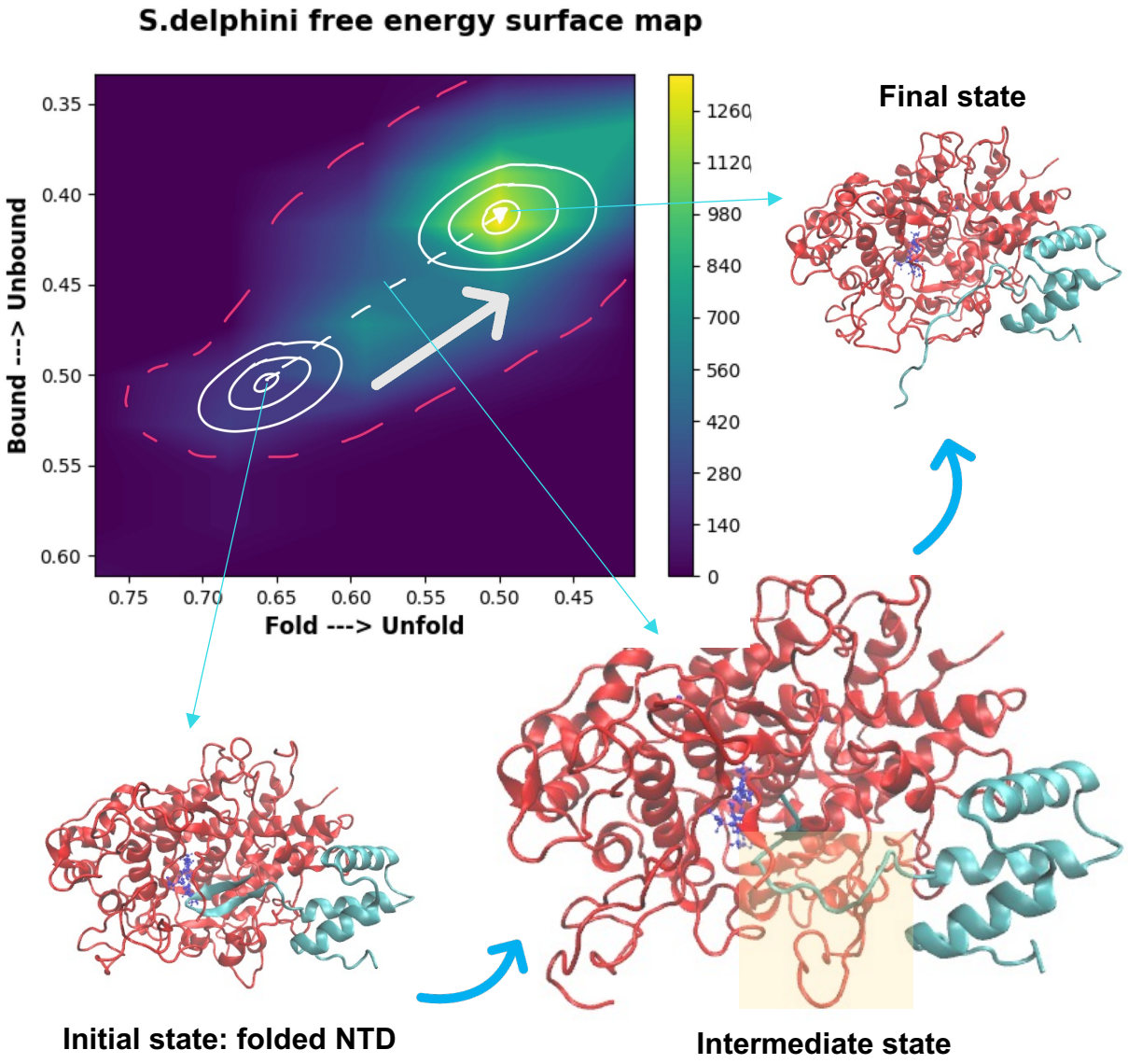
- *S.delphini* NTD unfolds faster than *S.aureus.*
- *S.delphini* unbinds faster than *S.aureus*.
- *S.delphini* NTD is less stable.
- *S.delphini* NTD takes equal time to unbind and unfold.

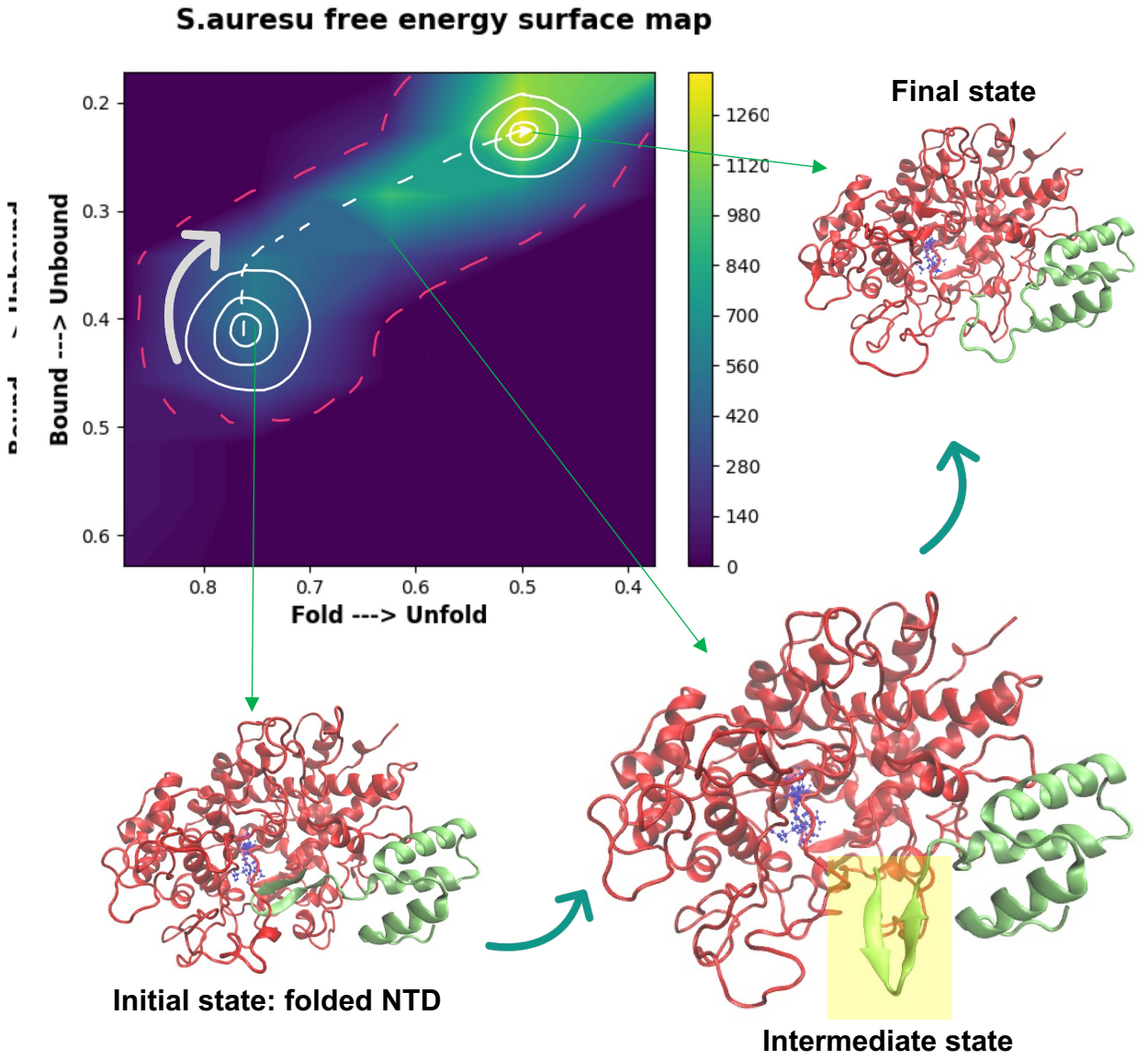
Hypothesis:

Stability $\xrightarrow{?}$ **Inhibition ability**

[Subaim1a]$ “pseudo”-Free Energy Landscape for SPIN/MPO

- **Cooperative binding: *S.delphini***

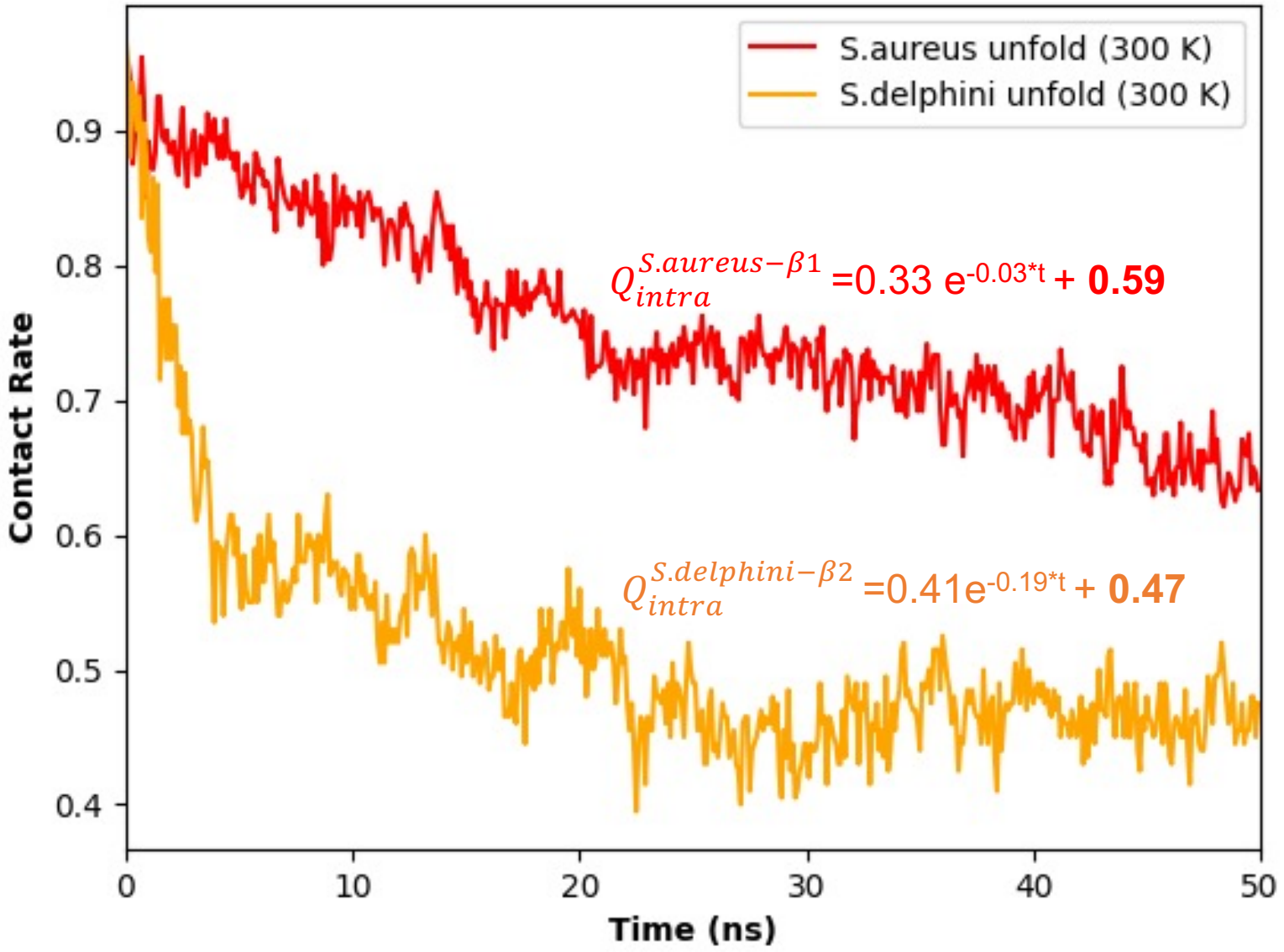




- **Conformational selection: *S.aureus*** (slightly)

[Subaim1a]$ Atomistic Simulations on SPIN-NTD

- ***S.aureus*** **NTD is more rigid and stable**



Setup:

- Temperature: 300 K.
- NPT ensemble, atomistic model. (High accuracy)
- 20 replicas simultaneously and independently.

S.aureus NTD

$T_{1/2}$ ≈ 22 (ns)

S.delphini NTD

$T_{1/2}$ ≈ 4 (ns)

[Subaim1a]$ SPIN/MPO Tight Specific Binding Investigations

Hypothesis:

- **SPIN-NTD** structured β-hairpin **stability** influences **inhibitory efficacy.**
- More **stable SPIN-NTD** is, more preference to have the **conformational selection** mechanism.

MD Results:

- ***S.aureus* NTD** is more **stable.**
- ***Binding mechanism:***

a. *S.delphini:* **cooperative binding.**

b. *S.aureus: slightly* **conformational-selection.**

Experimental Results:

- ***S.aureus*** shares high structural **identical** with ***S.delphini***.
- **Binding affinity** has no influence on SPIN **inhibitory**.
- ***S.aureus*** shows **highest inhibitory** ability to MPO.

Validation:

- **Experiments:** Mutants with pre-folded NTD stabilized by disulfide bonds.
- **MD simulations:**

(Coupled binding and folding simulations under physiological conditions.)

i. **Atomistic models** coupled with advanced sampling methods (i.e., REST, umbrella samplings…)

ii. **CG models** for quick binding and folding process kinetics calculation.

Blueprint:

- Effective therapeutic strategies targeting SPIN-NTD.

>_

[Yumeng@Prospectus]$ Part2: Specific Aims

Aim1. Integration of MD simulations and experiments for IDP studies

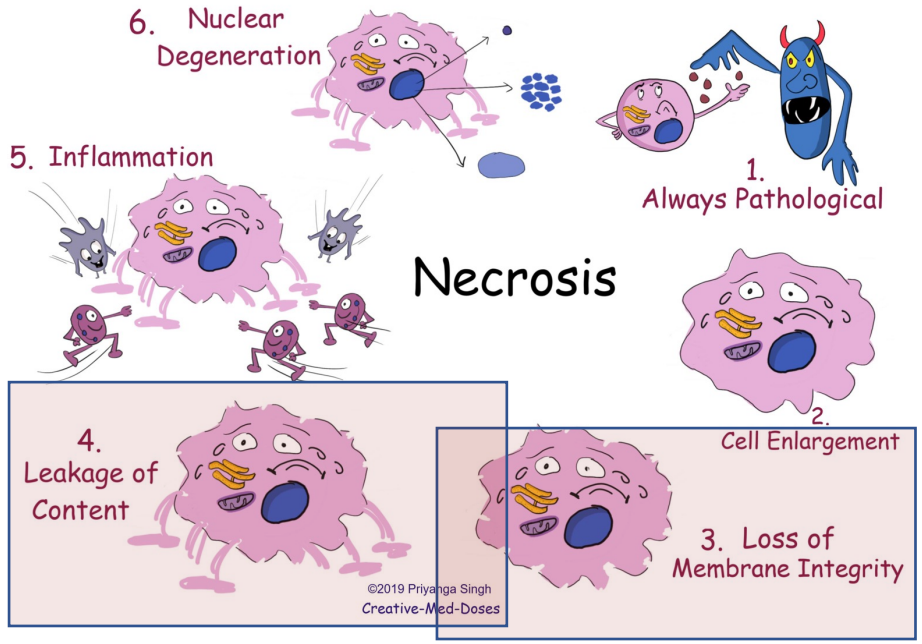
Aim2. Advanced methods for multi-scale simulations on IDPs

[Yumeng@Prospectus Aim1]$

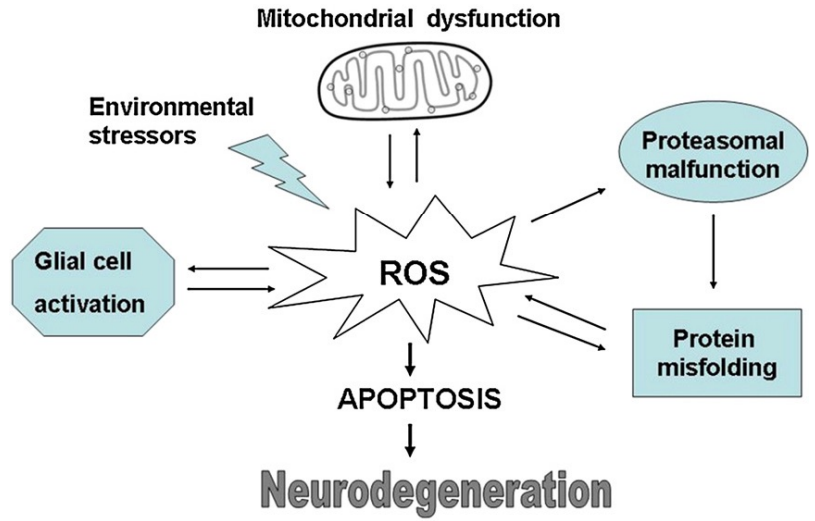
Subaim1a. IDP specific tight interactions: SPIN-NTD/MPO

Subaim1b. IDP dynamic interactions: p53-NTD/CypD

Subaim1c. IDP enzymatic interactions: Flaviviral proteases

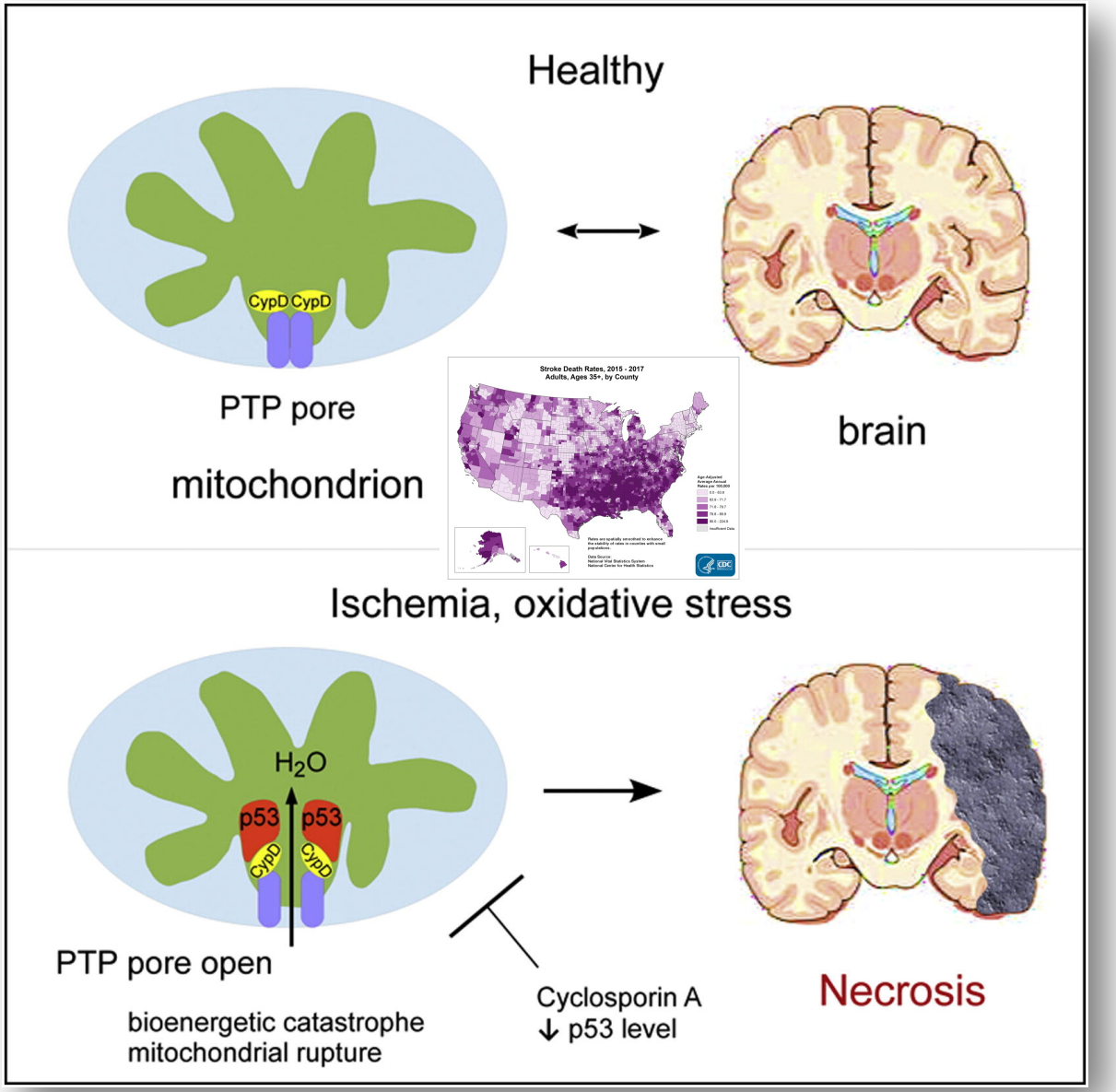


https://www.diffen.com/difference/Apoptosis_vs_Necrosis



doi.org/10.1155/2013/942916

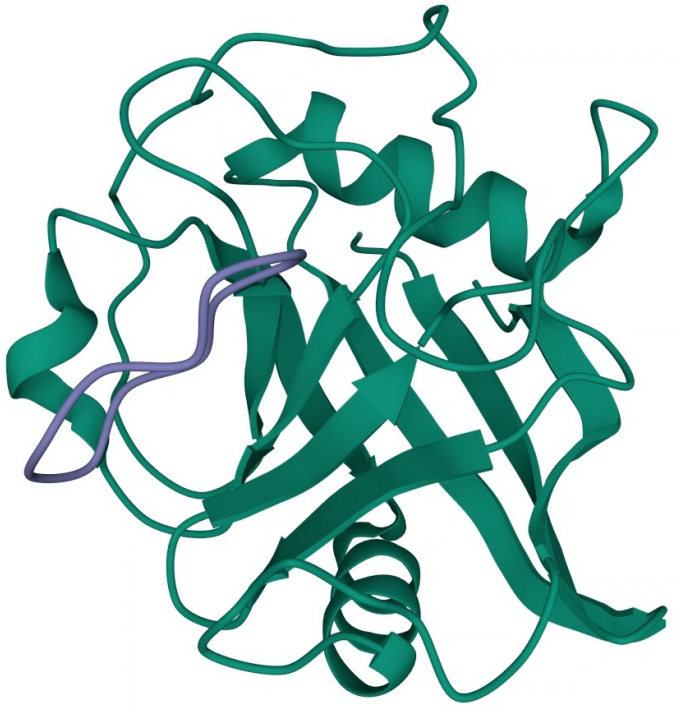
Oxidative stress, mitochondrion and necrosis



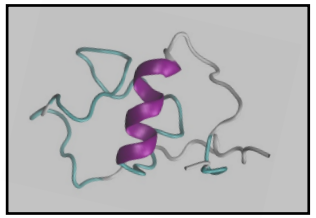
https://www.cdc.gov/stroke/maps_data.htm

[Subaim1b]$ Regulative Interactions: p53/CypD

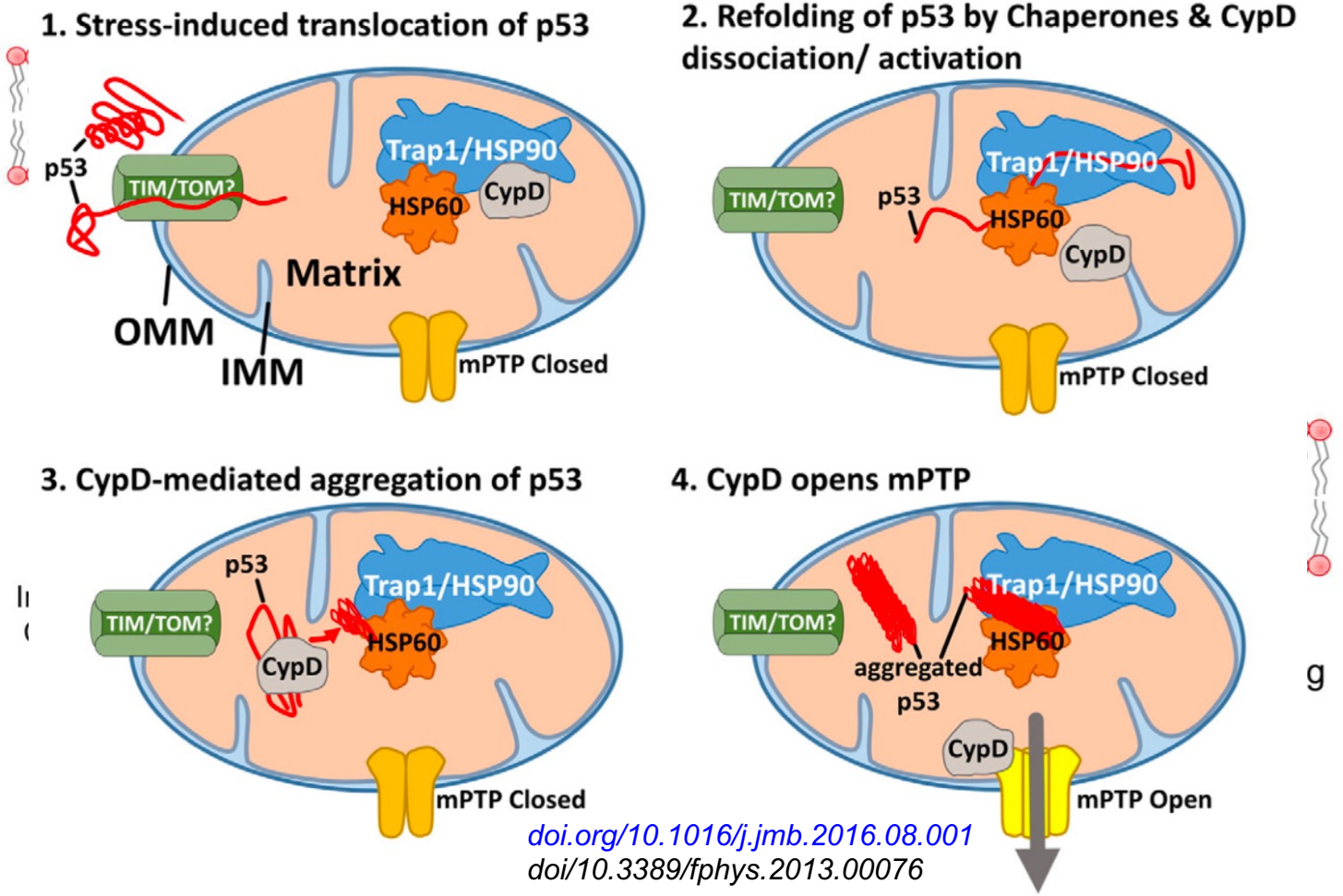
- Cyclophilin D (CypD)



- p53



- FLP53/CypD Cellular Regulation Cycle (Full-Length p53)

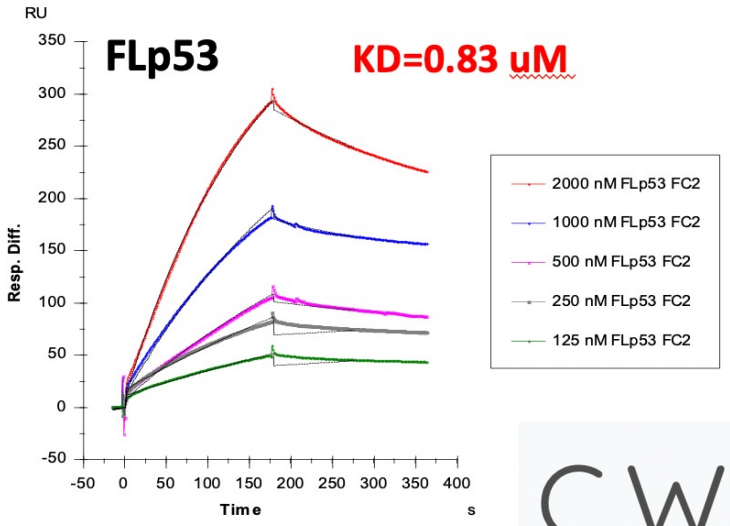


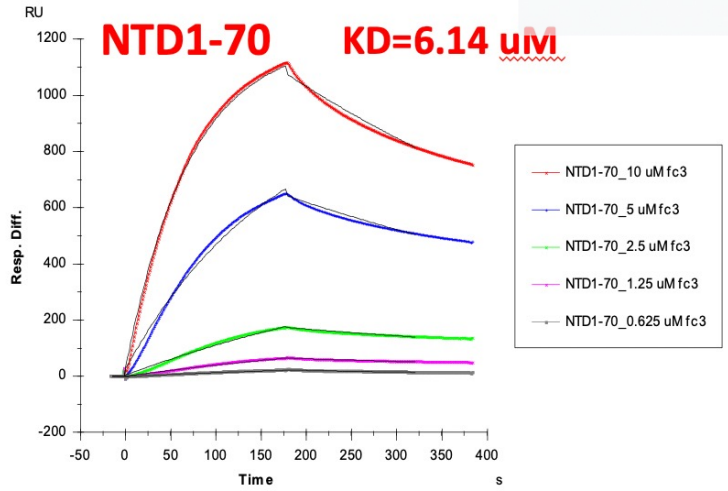
doi.org/10.1016/j.jmb.2016.08.001
doi/10.3389/fphys.2013.00076

- p53/CypD triggering pore opening is Ca^{2+} independent.
- Interactions are dynamic without stable complex forming.

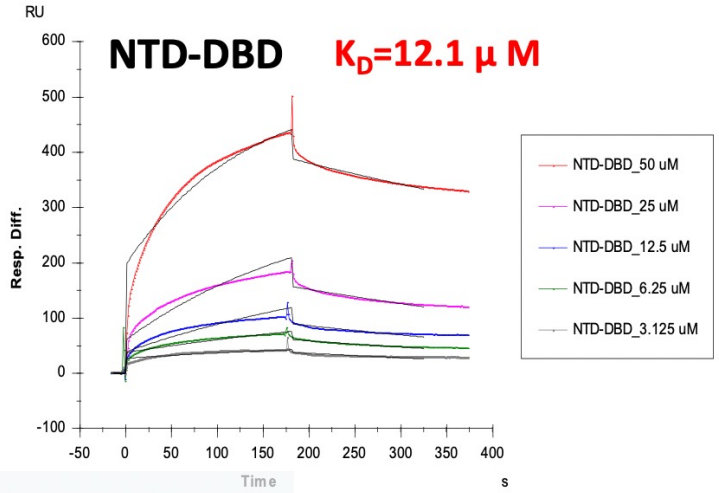
[Subaim1b]$ p53-NTD dominates p53/CypD interactions

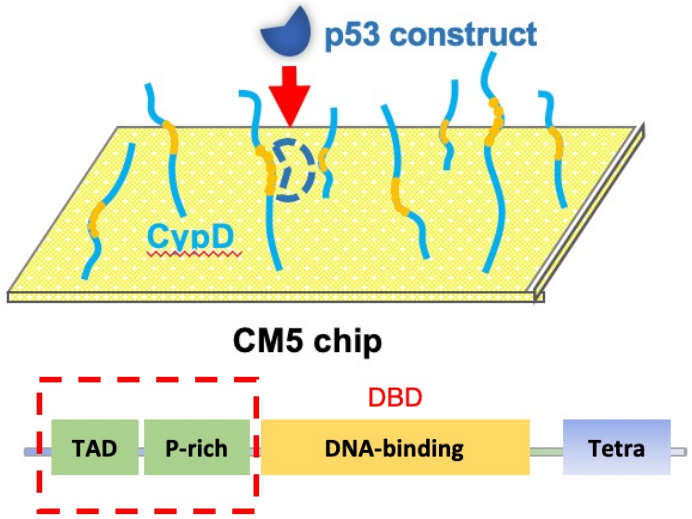
- **p53-NTD is the smallest binding region to CypD**





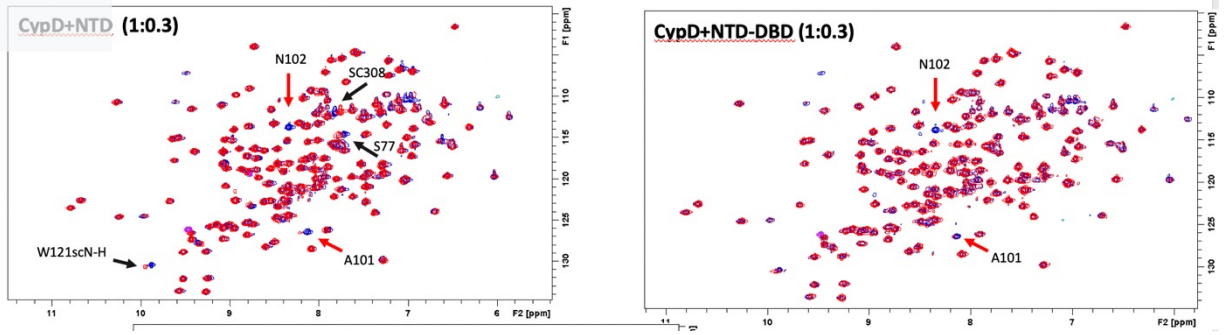
- **NTD-DBD has lower binding affinity**





NMR titration of p53 construct into ^{15}N CypD

Blue: CypD apo
Red: Complex

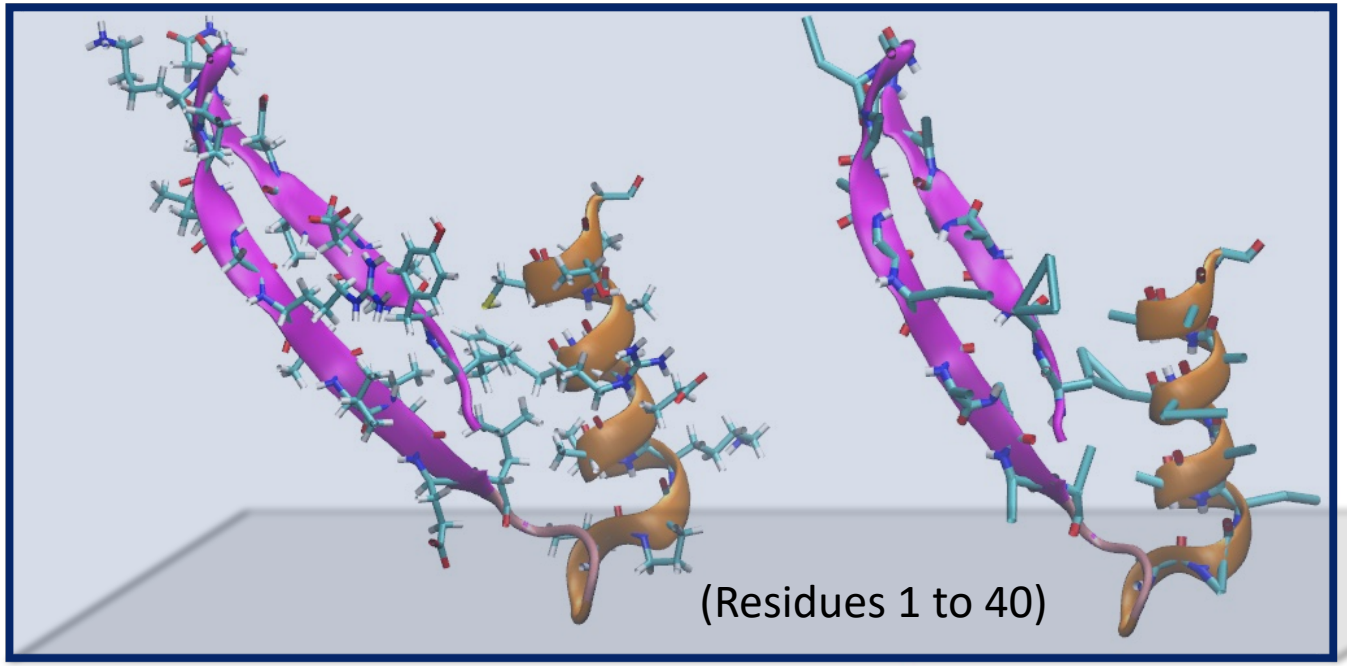


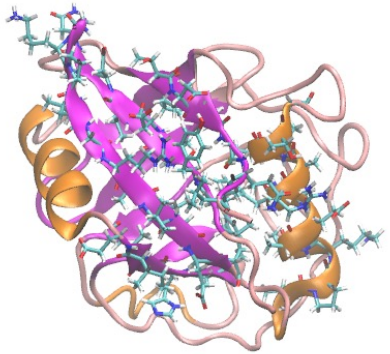
- **How NTD dominates p53/CypD binding?**

We NEED a computational model!

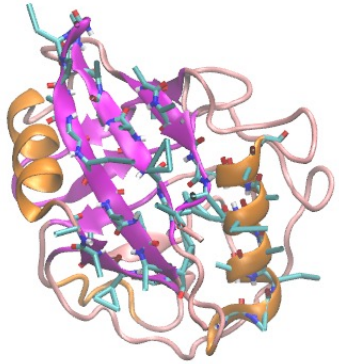
[Subaim1b]$ HyRes Model Simulations on p53-NTD/CypD

- **CypD in HyRes model**



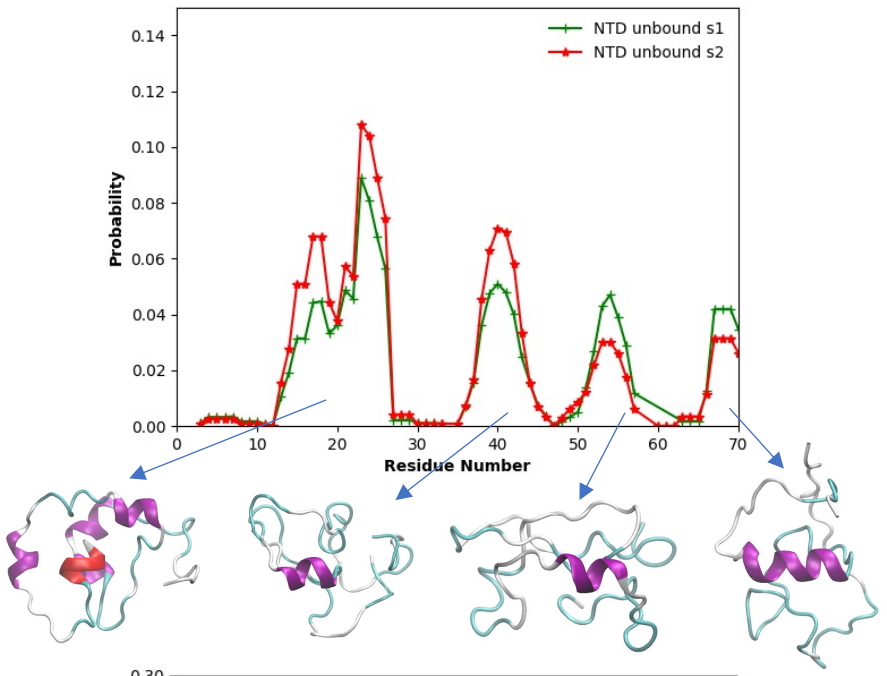


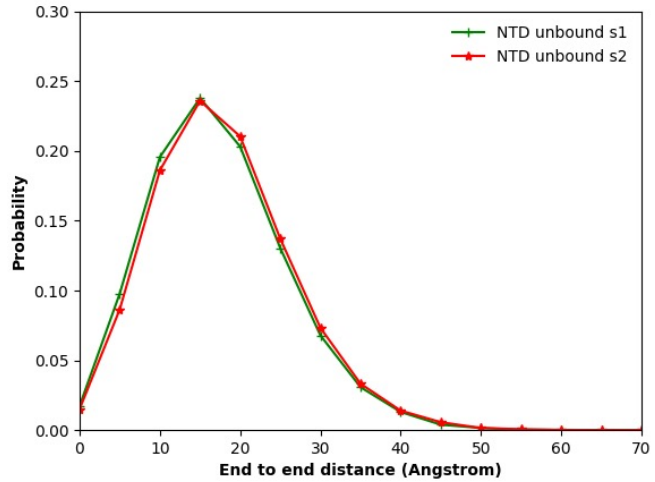
All-atom



HyRes

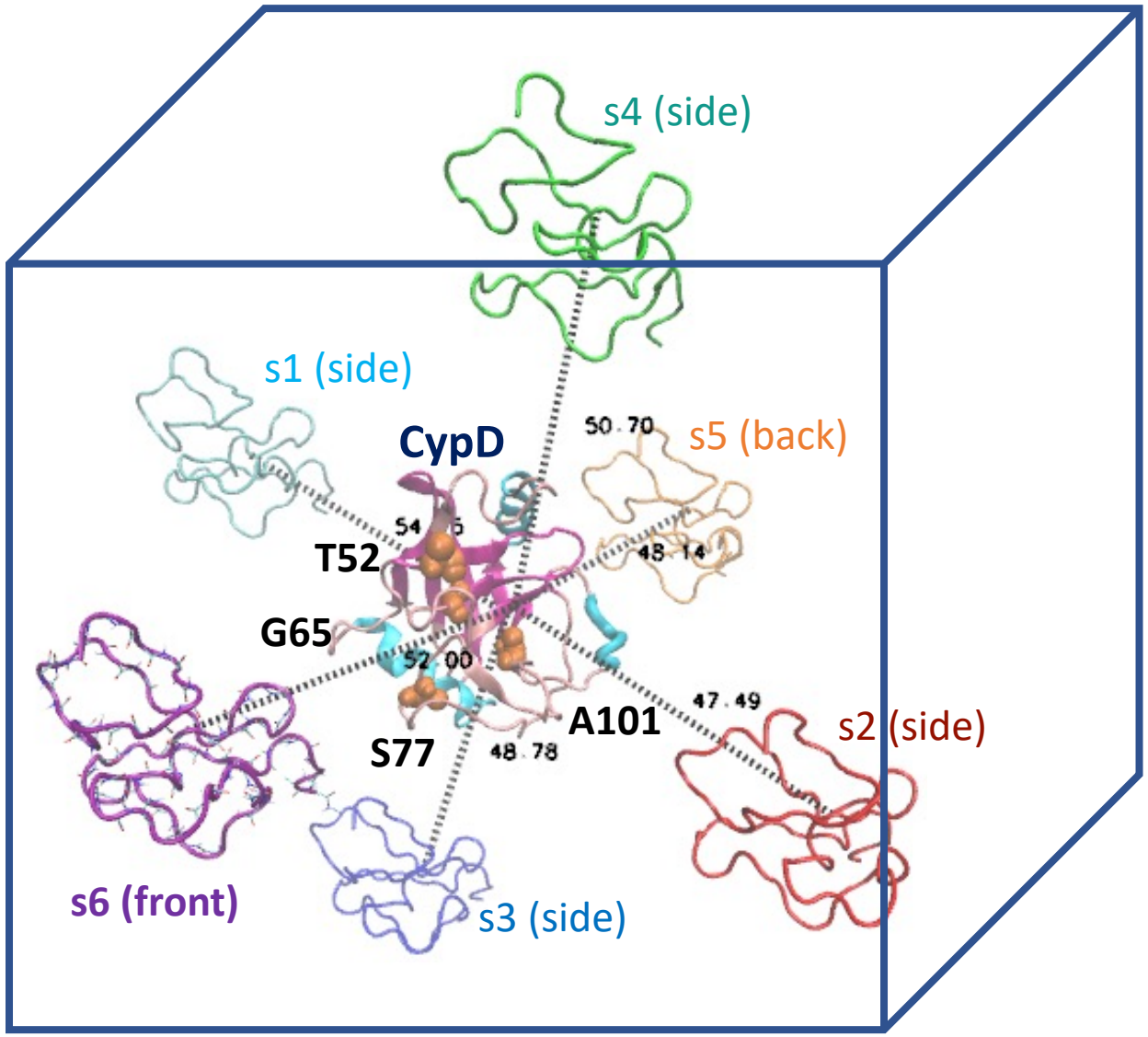
- **p53-NTD HyRes simulations (1 *μs*)**



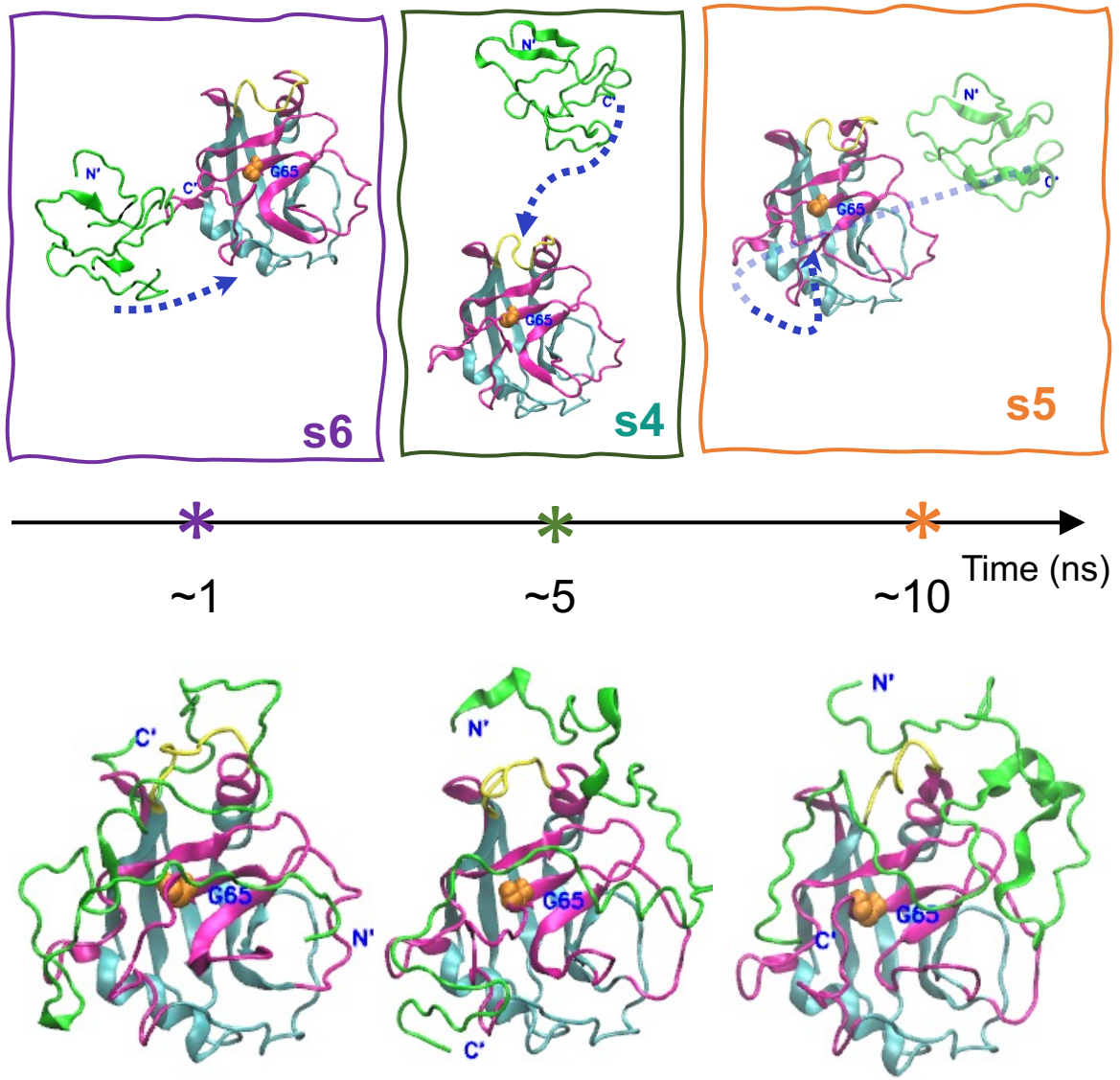


[Subaim1b]$ CG MD Simulations on p53-NTD/CypD

- **Initial simulation setup**

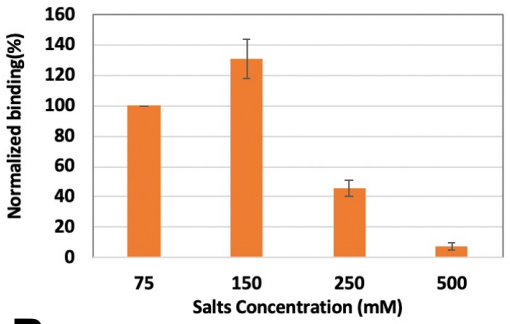


- **Simulation Trajectories**



[Subaim1b]$ Simulation & Experiment Preliminary Results

A



- **Experiments**
  1. Electrostatic driven. (Figure A)
  2. Broad binding interface found on CypD. (Figure B,C)

- **Simulations**
  1. Electrostatic driven. (Fig.a)
  2. Broad binding interface found on CypD (Fig.b)
  3. **Can capture dynamic interactions accurately. (CypD residues ~15)**
  4. **Broader and dynamically binding on NTD.**

a

CypD Binding Surface	Analysis Results
Total contact sites	14
Charged:	9
Non-polar:	5
Polar:	0

Total contact sites: 14

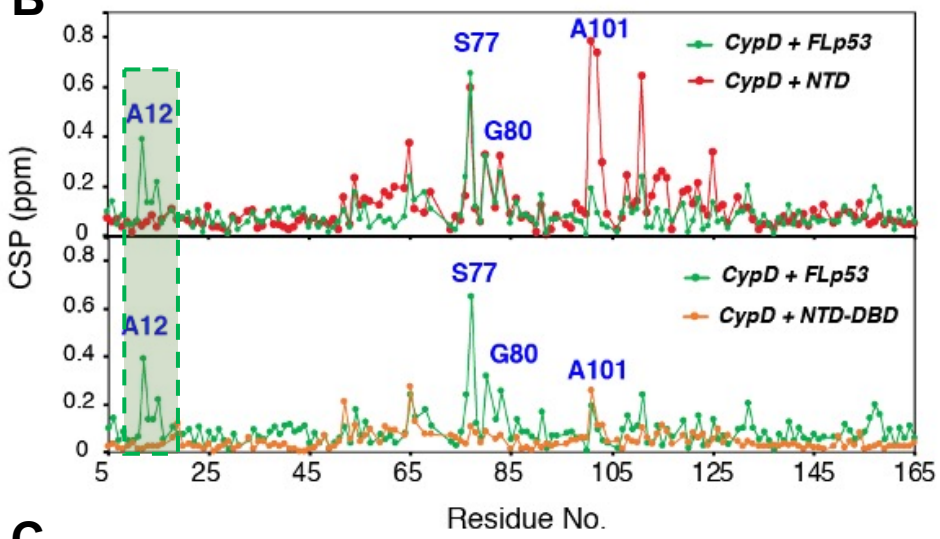
Positive charged(**9**):

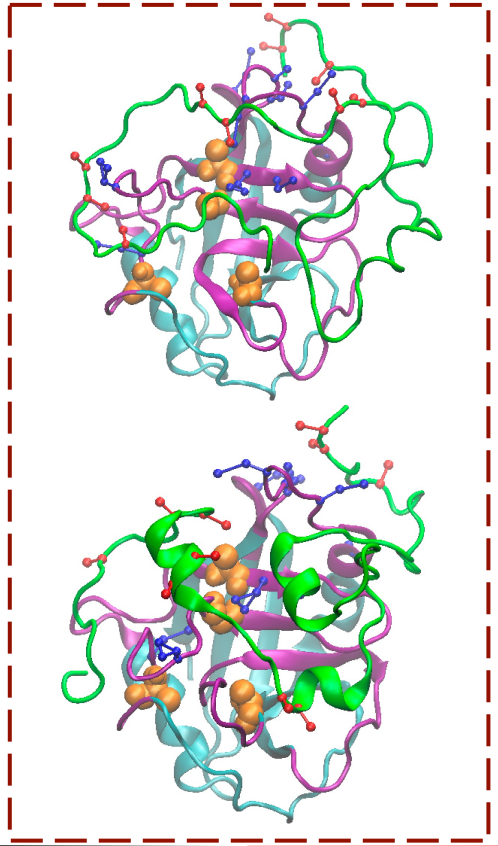
Lys: 6; Arg: 2; His(H69): 1.

Non-polar (**5**):

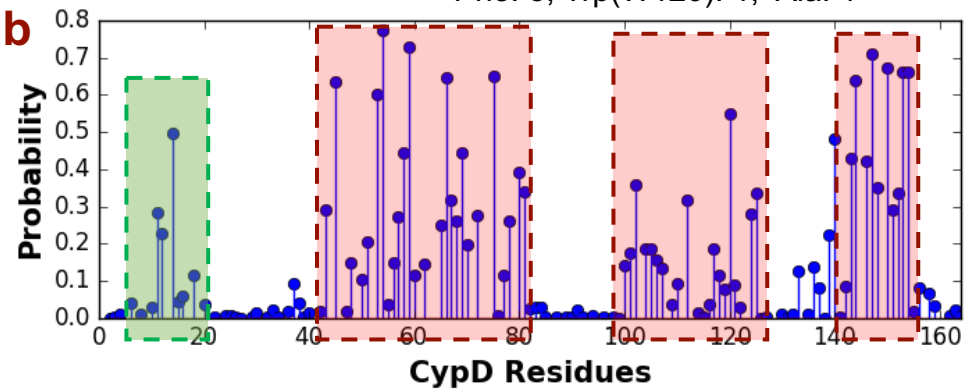
Phe: 3; Trp(W120): 1; Ala: 1

B

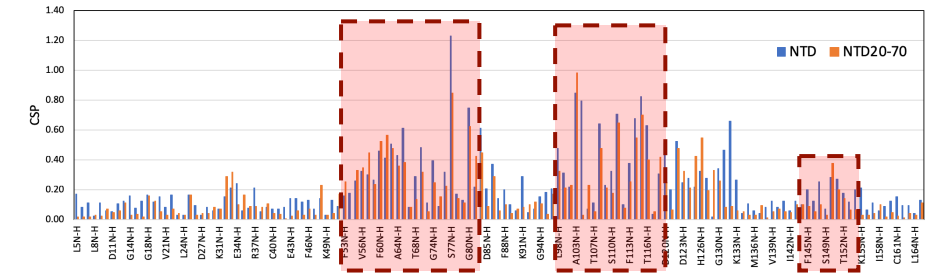


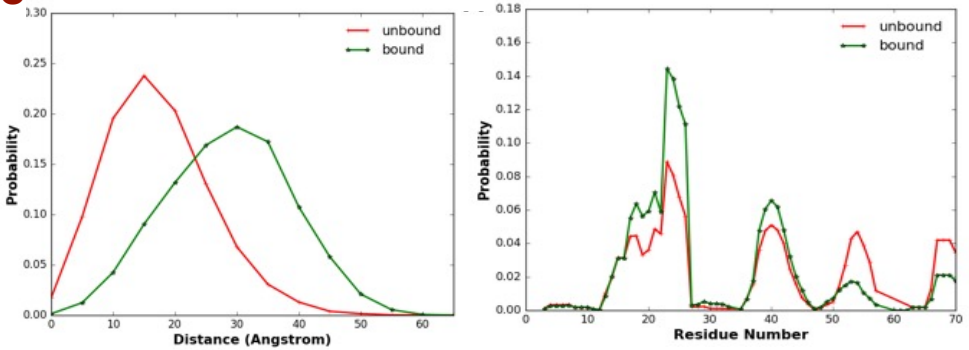


b

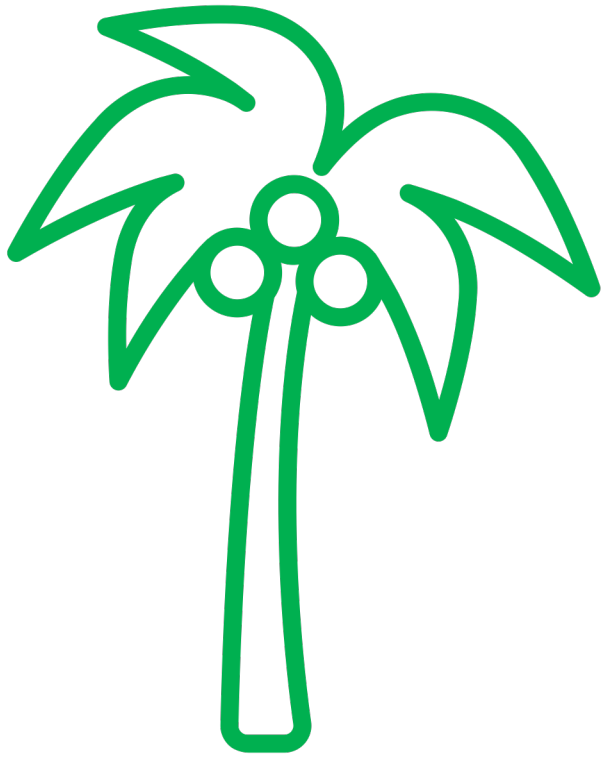


c





[Subaim1b]$ p53-NTD/CypD Investigations



MD Simulations:

- **MD simulation** results are highly **consistent** with **experimental** measurements.
- **HyRes model** is very powerful and can accurately describe **IDP 2nd structural** profiles and **long-range intermolecular interactions**.
- **MD simulations** can picture **deeper insights** and capture more comprehensive **interaction dynamics**.

Potential Problem:

- **HyRes protein model is a little over-compaction.**

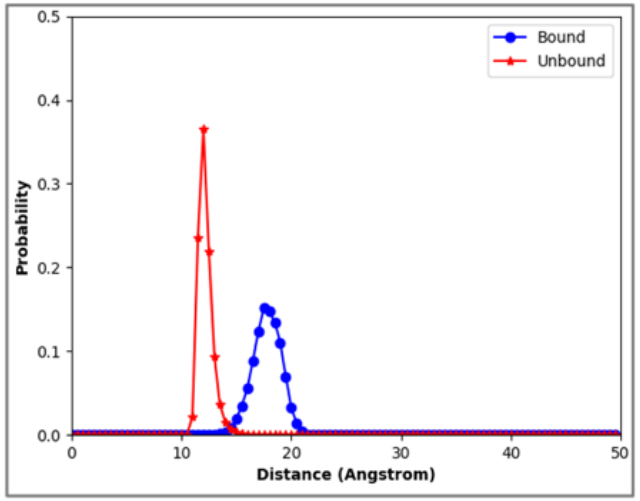


Fig. Radius of gyration results of NTD calculated in HyRes protein model in bound and unbound state.

Future plan:

- **Optimize HyRes protein model (subaim2b)**
- **NTD-DBD studies.**
- **Therapeutic target: CypD, to protect in models of diseases.**

>_

[Yumeng@Prospectus]$ Part2: Specific Aims

Aim1. Integration of MD simulations and experiments for IDP studies

Aim2. Advanced methods for multi-scale simulations on IDPs

[Yumeng@Prospectus Aim1]$

Subaim1a. IDP specific tight interactions: SPIN-NTD/MPO

Subaim1b. IDP dynamic interactions: p53-NTD/CypD

Subaim1c. IDP enzymatic interactions: Flaviviral proteases

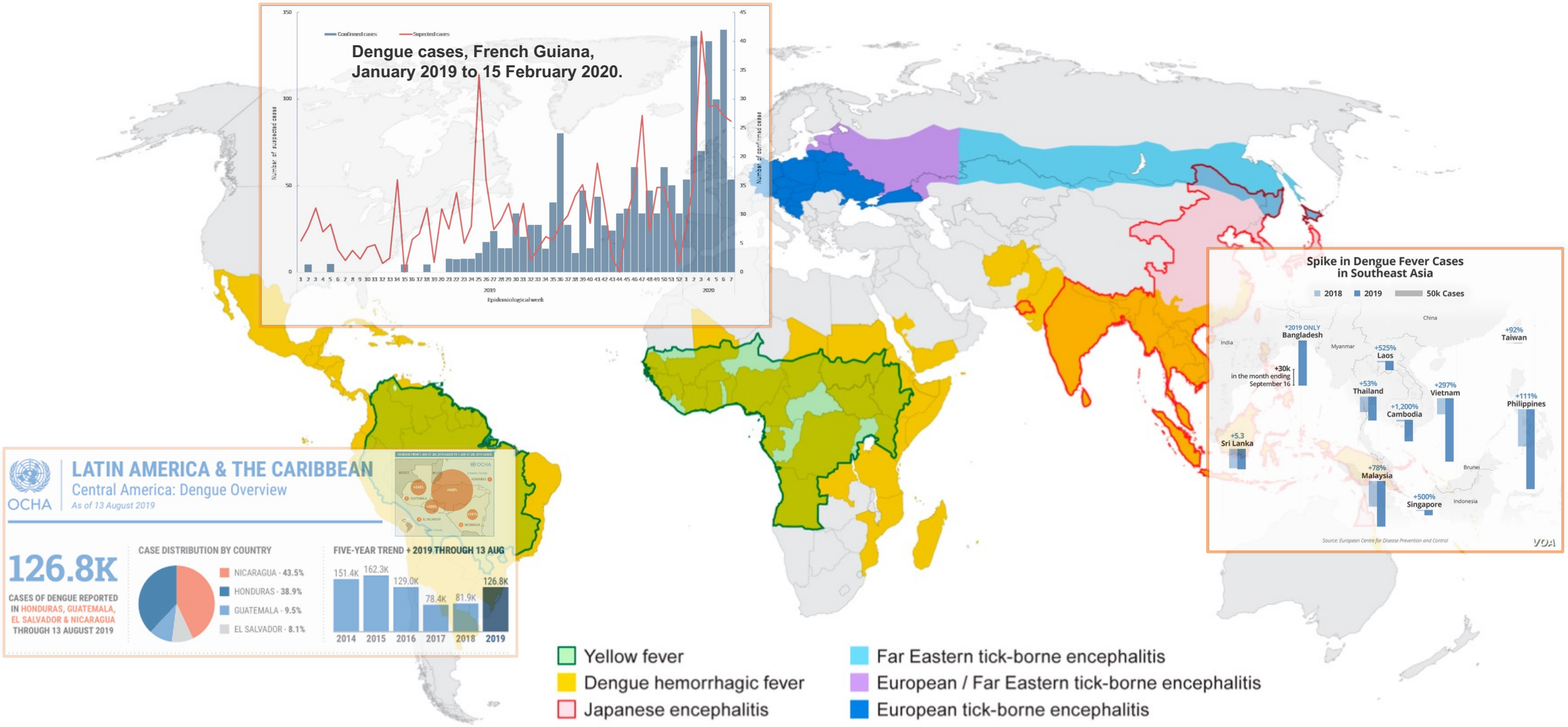
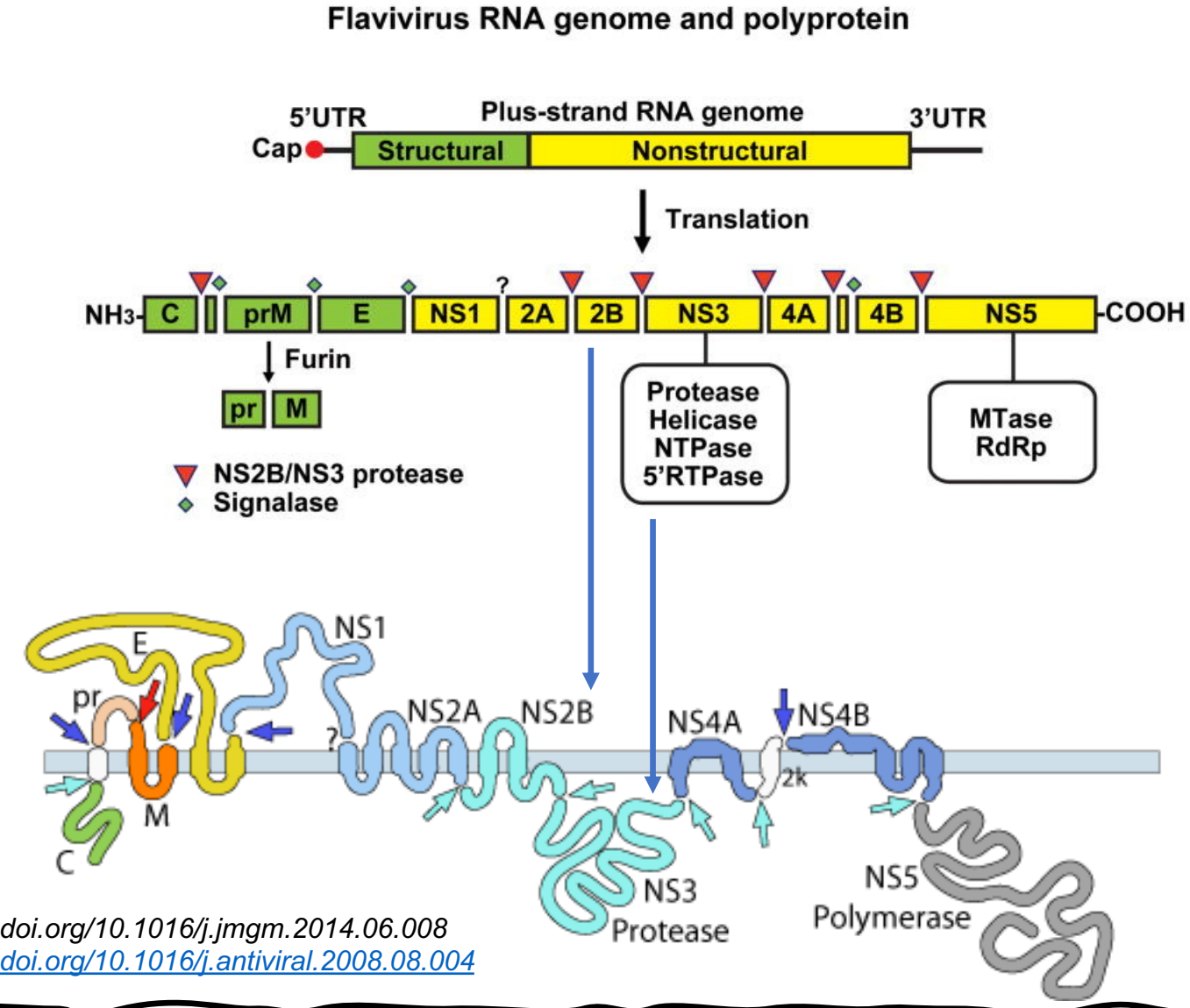


Figure 1. Distribution of major flaviviruses discussed in this article. Information was adapted from data and figures provided on Centers for Disease Control and Prevention (CDC) and World Health Organization (WHO) websites.

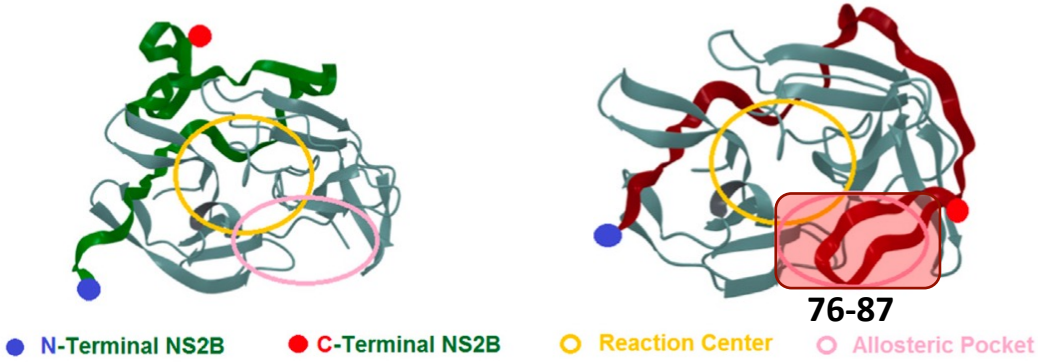
[Subaim1c]$ Flavivirus NS2B/NS3 proteases

- **Function of NS2B/NS3 proteases**



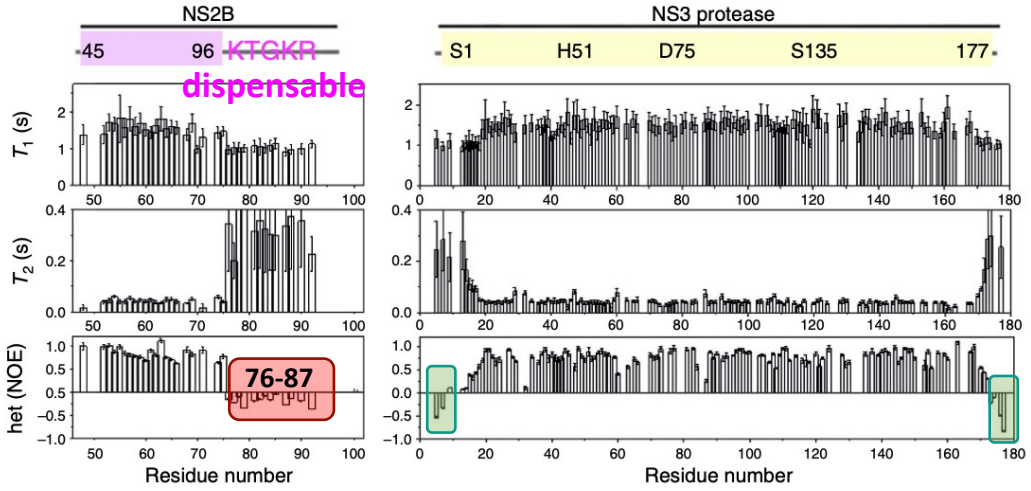
doi.org/10.1016/j.jmgm.2014.06.008
doi.org/10.1016/j.antiviral.2008.08.004

- **Structure of NS2B/NS3 proteases**



doi.org/10.1016/j.jmgm.2014.06.008

- **Dynamic of NS2B/NS3 proteases**



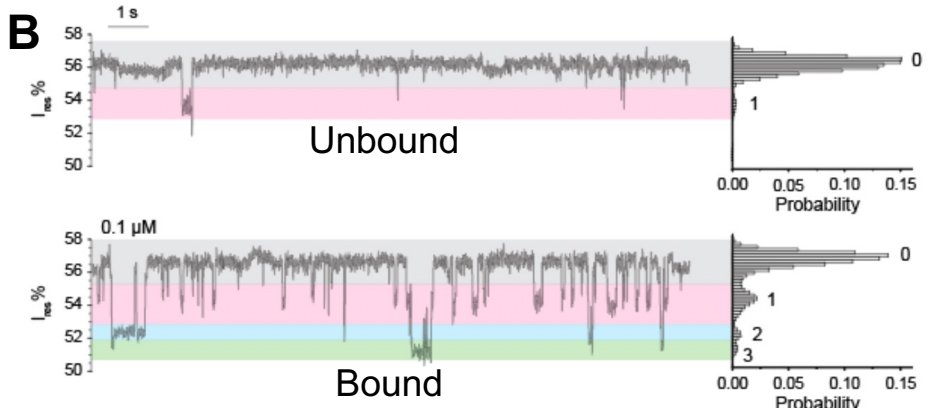
doi.org/10.1038/ncomms13410

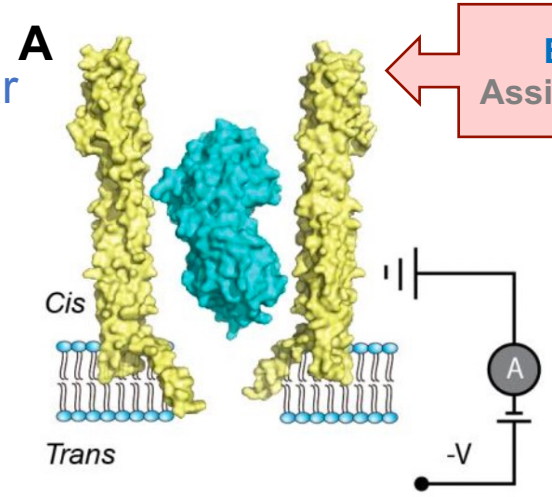
[Subaim1c]$ Flavivirus NS2B/NS3 Proposed Research Plan

Chen Research Group

University of Massachusetts Amherst

- **ClyA nanopore tweezers tool for probing NS2B/NS3 proteases**





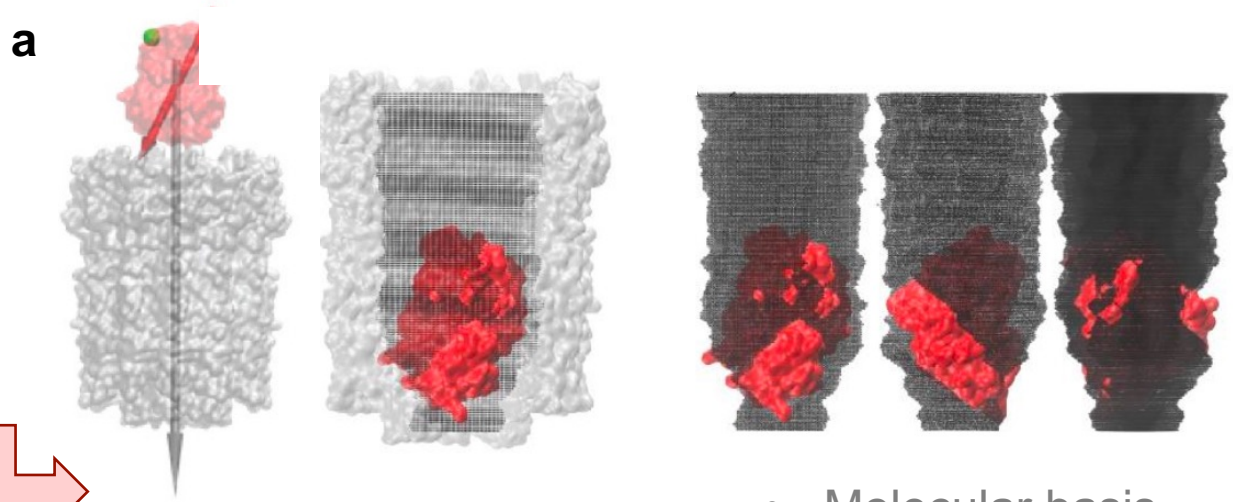
- Single Molecular
- High sensitivity
- …

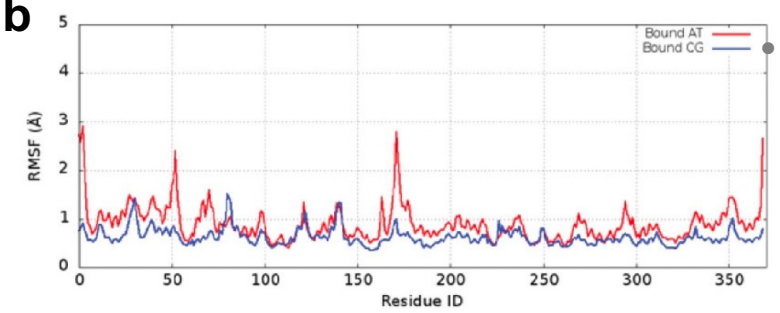
EXPM signals
Assign current states

Jianhan Chen Research Group

Computational Biophysics and Biomaterials

- **Advanced multi-scale samplings to investigate NS2B/NS3 proteases**





- Molecular basis
- Binding insights
- …

doi.org/10.1101/727503

>_

[Yumeng@Prospectus]$ Part2: Specific Aims

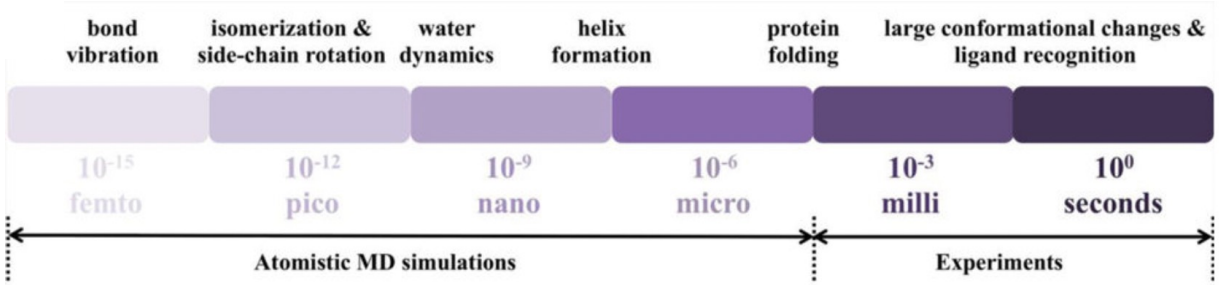
Aim1. Integration of MD simulations and experiments for IDP studies

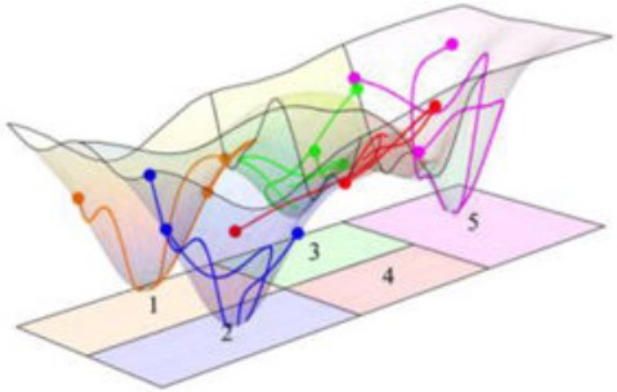
Aim2. Advanced methods for multi-scale simulations on IDPs

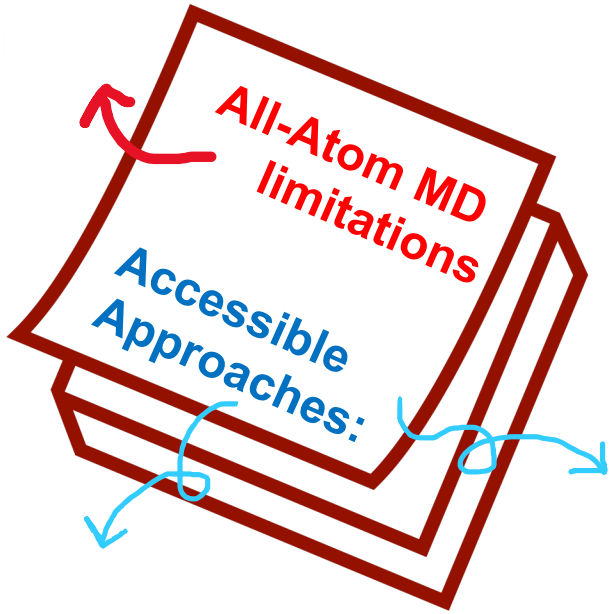
[Yumeng@Prospectus Aim2]$

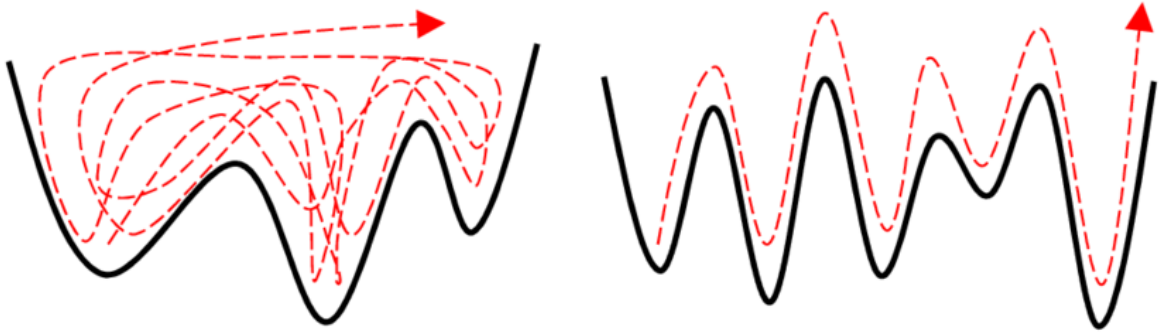
Subaim2a. Enhanced sampling method: REST3

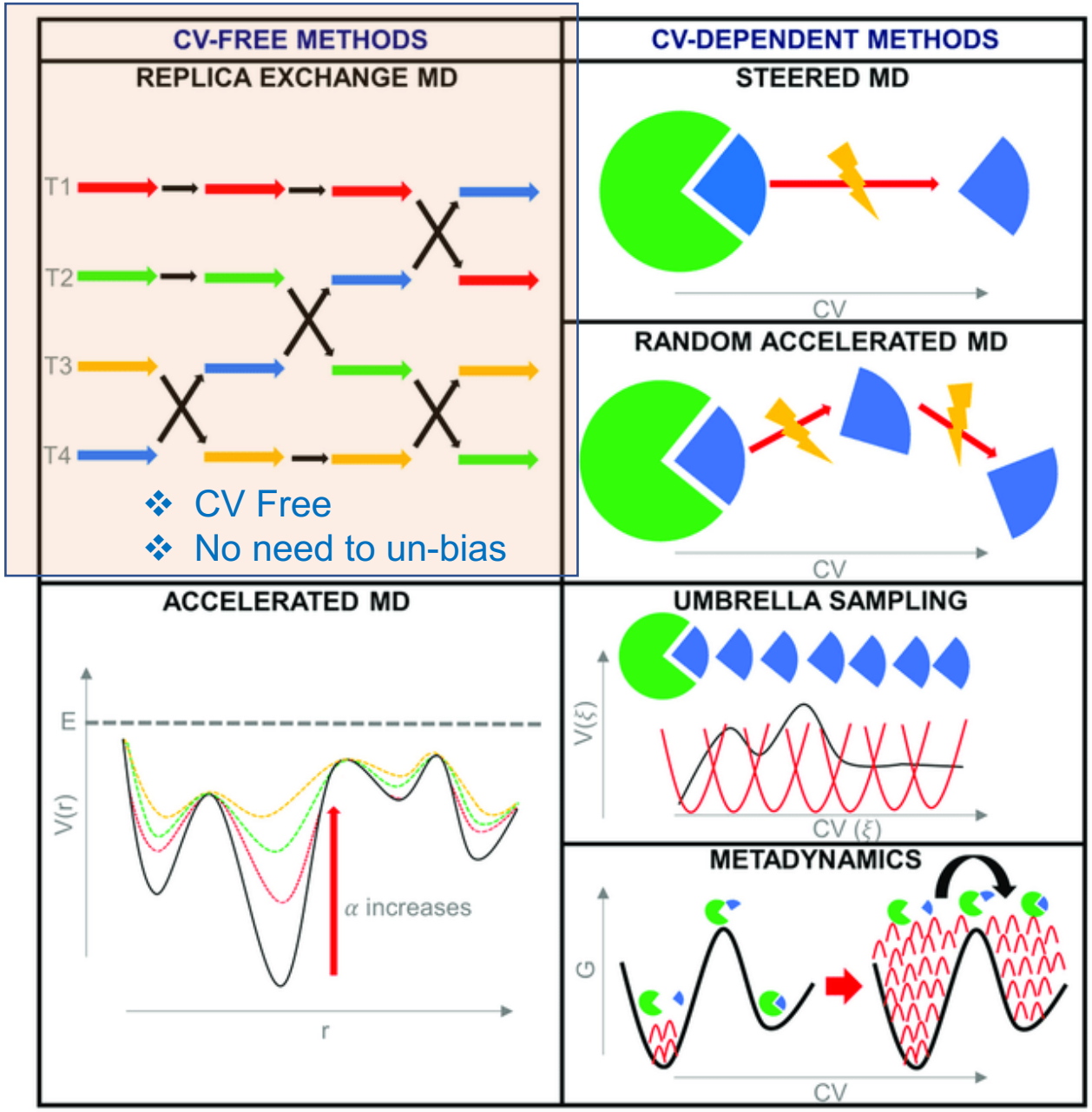
Subaim2b. Optimized force field: HyRes*









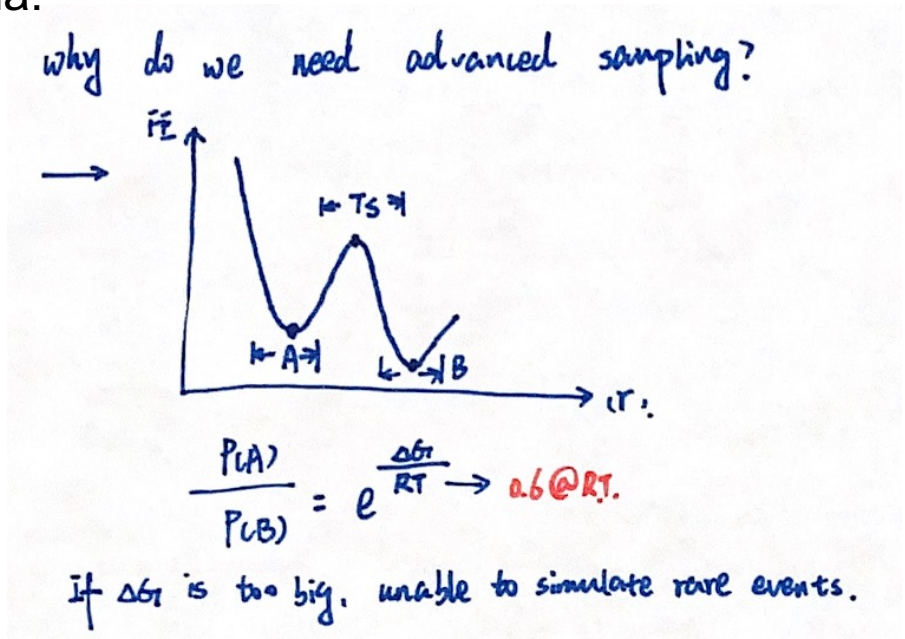


doi: 10.3389/fphar.2018.00923
doi: 10.1021/acs.accounts.5b00536

[Subaim2a]$ Enhanced Sampling Methods for Atomistic Models

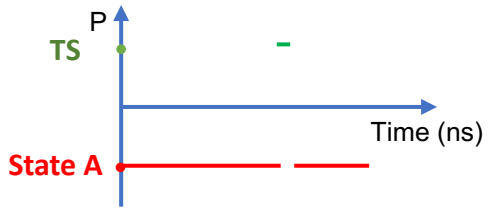
- **Atomistic Simulations limitation:**

Complex system is frequently **trapped** in local minima.



$\Delta G_{A->TS} = 6$ kcal/mol

$\frac{P_A}{P_{TS}} = e^{10} \sim 10^5$



1 μs simulation: only 10 ps may come to transition state.

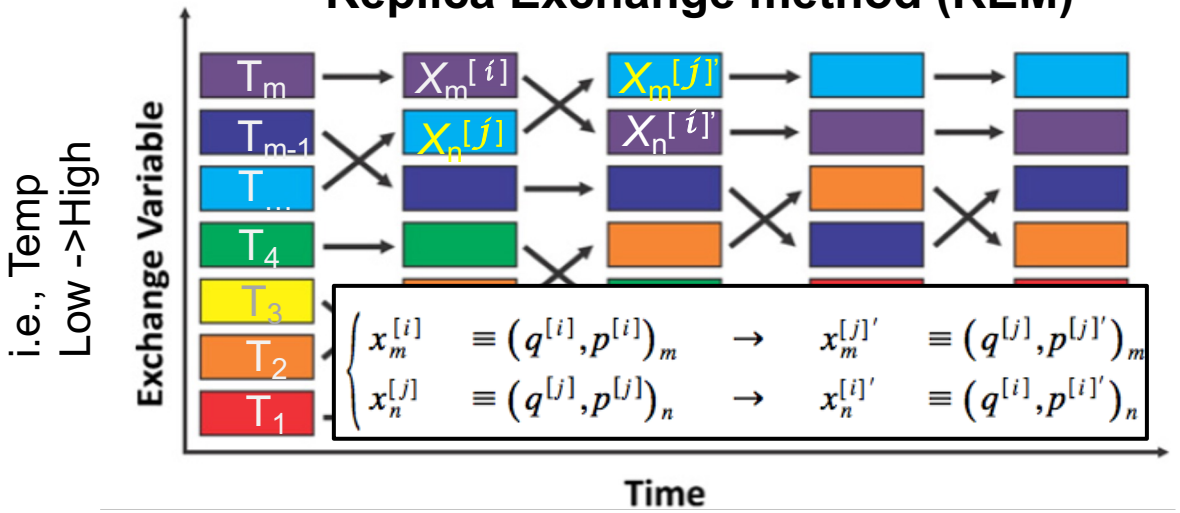
We can merely capture it!

doi.org/10.1016/S0009-2614(00)00999-4

- **Approaches:**

Realize random walking on energy surface.

Replica Exchange method (REM)



$$\begin{cases} x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \rightarrow x_m^{[j]'} \equiv (q^{[j]}, p^{[j]'})_m \\ x_n^{[j]} \equiv (q^{[j]}, p^{[j]})_n \rightarrow x_n^{[i]'} \equiv (q^{[i]}, p^{[i]'})_n \end{cases}$$

(Exchange neighbor replicas from condition m to n)

- **Transition probability:**

$$w(X \rightarrow X') \equiv w\left(x_m^{[i]} \mid x_n^{[j]}\right) = \begin{cases} 1, & \text{for } \Delta \leqslant 0, \\ \exp(-\Delta), & \text{for } \Delta > 0, \end{cases}$$

$$\Delta \equiv [\beta_n - \beta_m]\left(E(q^{[i]}) - E(q^{[j]})\right)$$ *(Acceptance rate)*

- **Cons:** $N_{rep} \propto O(f^{1/2})$ $A = \min\{1, \exp(-\Delta\beta\Delta E)\}$

e.g., One hairpin protein system: molecule + water 4342 atoms -> 64 Reps

$\Delta E \uparrow, A_{min} \downarrow$

$N_{rep} \uparrow$

[Subaim2a]$ Enhanced Sampling Method: REST1

- **REST** (**R**eplica **E**xchange **S**olute **T**empering)

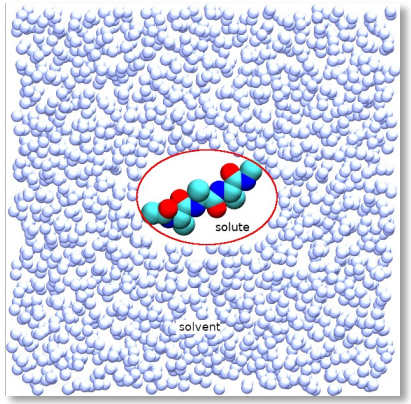
  a. **REST 1:**

$$\Delta_{nm} = -\beta_m[E_m(X_n) - E_m(X_m)] - \beta_n[E_n(X_m) - E_n(X_n)]$$

REM: ΔE

$$E_0(X) = E_p(X) + E_{pw}(X) + E_{ww}(X)$$

(E_{ww} Kept)



$$E_m(X) = E_p(X) + \left[\frac{\beta_0}{\beta_m}\right]E_{ww}(X) + \left[\frac{\beta_0 + \beta_m}{2\beta_m}\right]E_{pw}(X)$$

$$\Delta_{nm} = (\beta_n - \beta_m)[(E_p(X_m) + \tfrac{1}{2}E_{pw}(X_m)) - (E_p(X_n) + \tfrac{1}{2}E_{pw}(X_n))]$$

Δβ

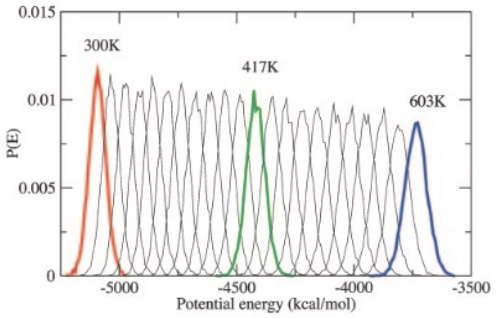
REST1: ΔE

Why efficient? $|E_p + (1/2)E_{pw}| \ll |E_p + E_{pw} + E_{ww}|$

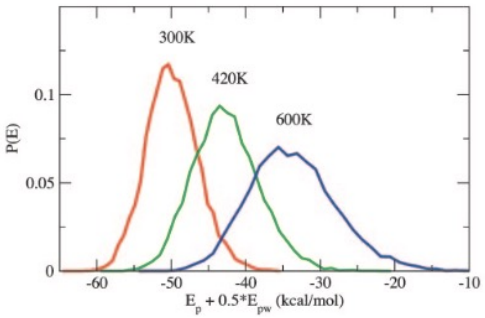
doi.org/10.1021/jp068826w
doi 10.1073 pnas.0506346102

- **REST1: Pros & Cons**
- Pros: high efficiency, accuracy.

i.e., System: Alanine Dipeptide + 512 water
Conditions: 300 K – 600 K



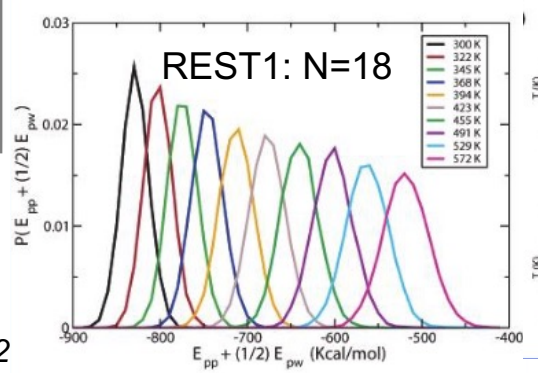
REM: N=22

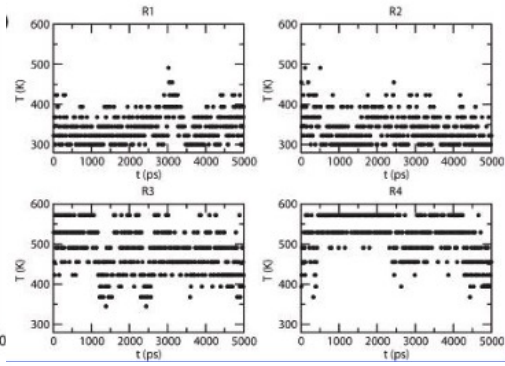


REST1: N=3

- Cons: Low exchange rate for complex system with big conformational changes.

i.e., β-hairpin system





[Subaim2a]$ Enhanced Sampling Method: REST2

- **REST** (**R**eplica **E**xchange **S**olute **T**empering)

b. REST 2:

REST1: $$E_m(X) = E_{\mathrm{p}}(X) + \left[\frac{\beta_0}{\beta_m}\right] E_{\mathrm{ww}}(X) + \left[\frac{\beta_0 + \beta_m}{2\beta_m}\right] E_{\mathrm{pw}}(X)$$

REST2: $$E_m^{\mathrm{REST2}}(X) = \frac{\beta_m}{\beta_0} E_{\mathrm{pp}}(X) + \sqrt{\frac{\beta_m}{\beta_0}} E_{\mathrm{pw}}(X) + E_{\mathrm{ww}}(X)$$

$$\Delta_{nm}(\mathrm{REST1}) = (\beta_n - \beta_m)[(E_{\mathrm{p}}(X_m) + \tfrac{1}{2}E_{\mathrm{pw}}(X_m)) - (E_{\mathrm{p}}(X_n) + \tfrac{1}{2}E_{\mathrm{pw}}(X_n))]$$

$$\Delta_{mn}(\mathrm{REST2}) = (\beta_m - \beta_n)\left[(E_{\mathrm{pp}}(X_n) - E_{\mathrm{pp}}(X_m)) + \frac{\sqrt{\beta_0}}{\sqrt{\beta_m} + \sqrt{\beta_n}}(E_{\mathrm{pw}}(X_n) - E_{\mathrm{pw}}(X_m))\right]$$

Why efficient? $\Delta_{mn}E_{REST2} < \Delta_{mn}E_{REST1}$

$$\left|\sqrt{\beta_0 * \beta_m}\, E_{pw}\right| < \left|\frac{\beta_0+\beta_m}{2}\, E_{pw}\right| \ (\beta_m < \beta_0)$$

Initial (Folding):

i. $\beta_f E_{pp} + \beta_f E_{pw} + \beta_f E_{ww}$

Final (Unfolding):

- **REX:**

ii. $\beta_u E_{pp} + \beta_u E_{pw} + \beta_u E_{ww}$

❖ E_{ww} is excluded from exchange.

- **REST1:**

iii. $\beta_u E_{pp} + \frac{(\beta_u+\beta_f)}{2} E_{pw} + \beta_f E_{ww}$

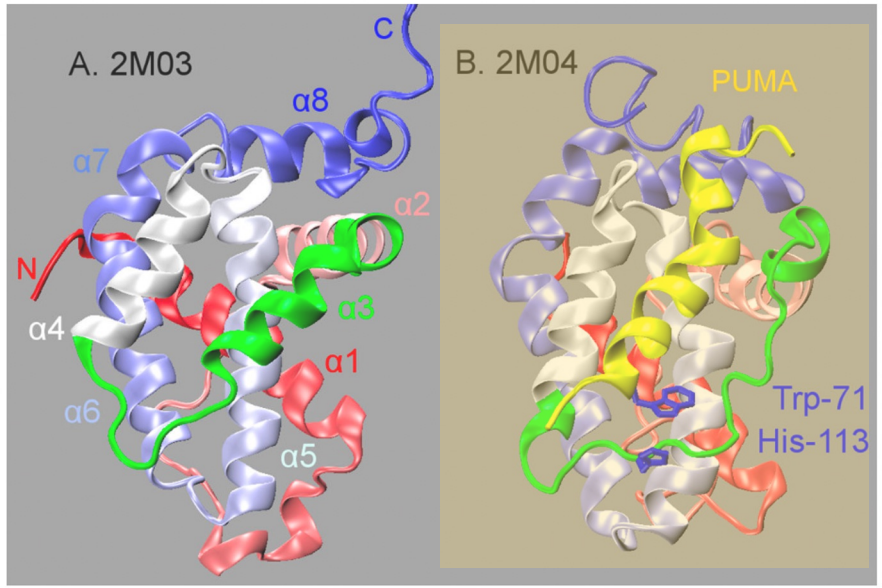
❖ E_{pw} is scaled down for higher acceptance rate.

- **REST2:**

$\beta_u E_{pp} + \sqrt{\beta_f \beta_u} E_{pw} + \beta_f E_{ww}$

doi.org/10.1021/jp204407d

[Subaim2a]$ REST2 Simulations on Bcl-xL



❖ **PDB:**

2M03: unbound

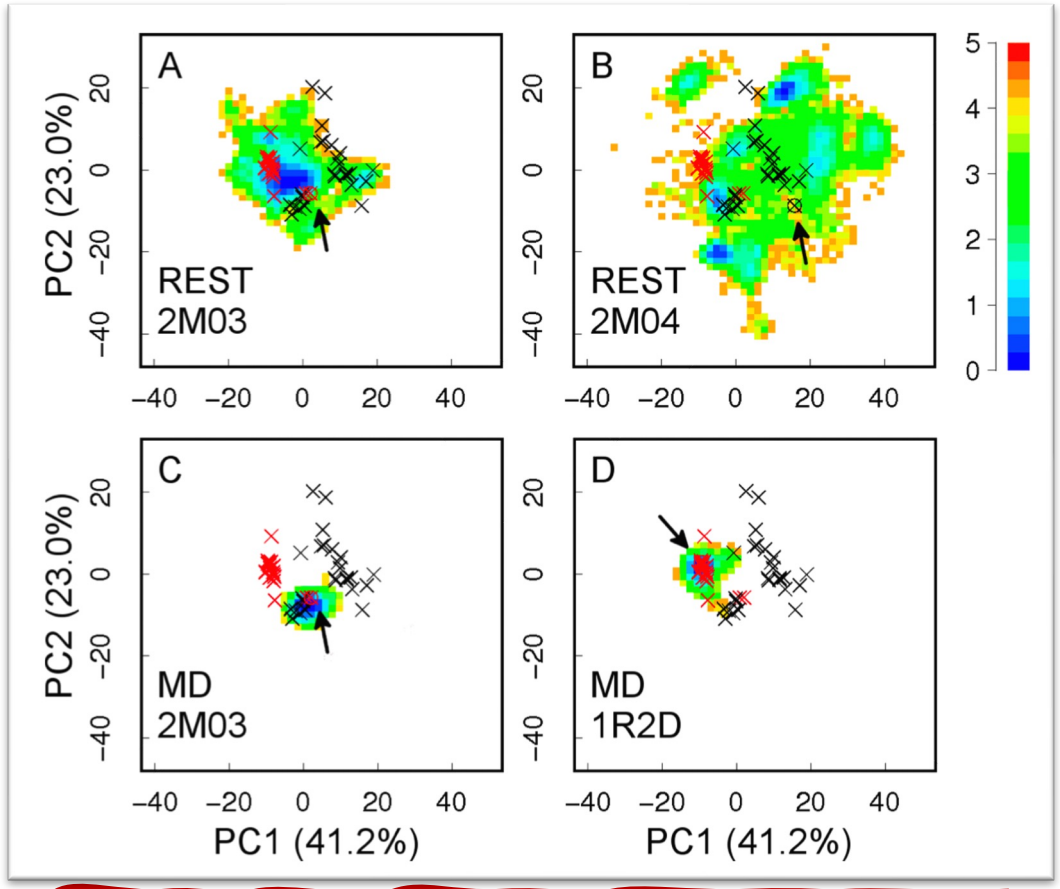
2M04: bound

❖ **Green residues:**

Bcl-xL-BH3 binding interface.

(unbound: **helix**)

(Bound: **unfolded**)



❖ **Clusters:**

(representative conformation ensembles)

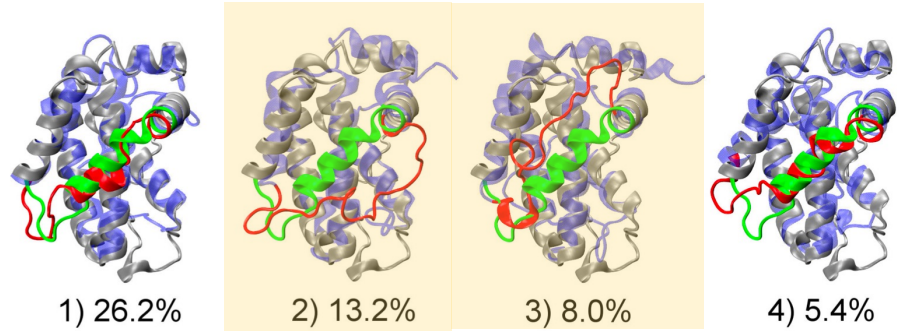
Traj from **REST2:**

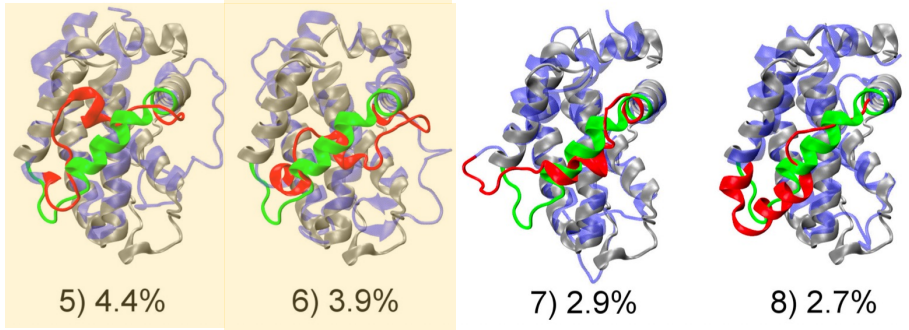
2M03 initiated:

1) 4) 7) 8)

2M04 initiated:

2) 3) 5) 6)



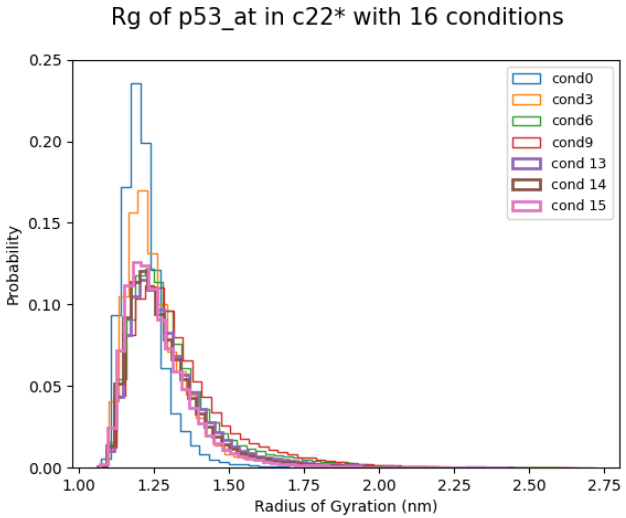


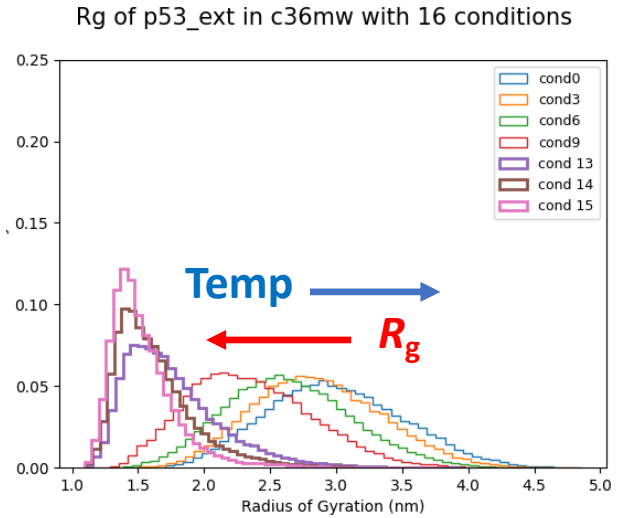
- **Much broader conformational space sampled by REST2 protocol.**
- Arrow: initial state.
- Red: starting from unbound state.
- Black: starting from bound state.

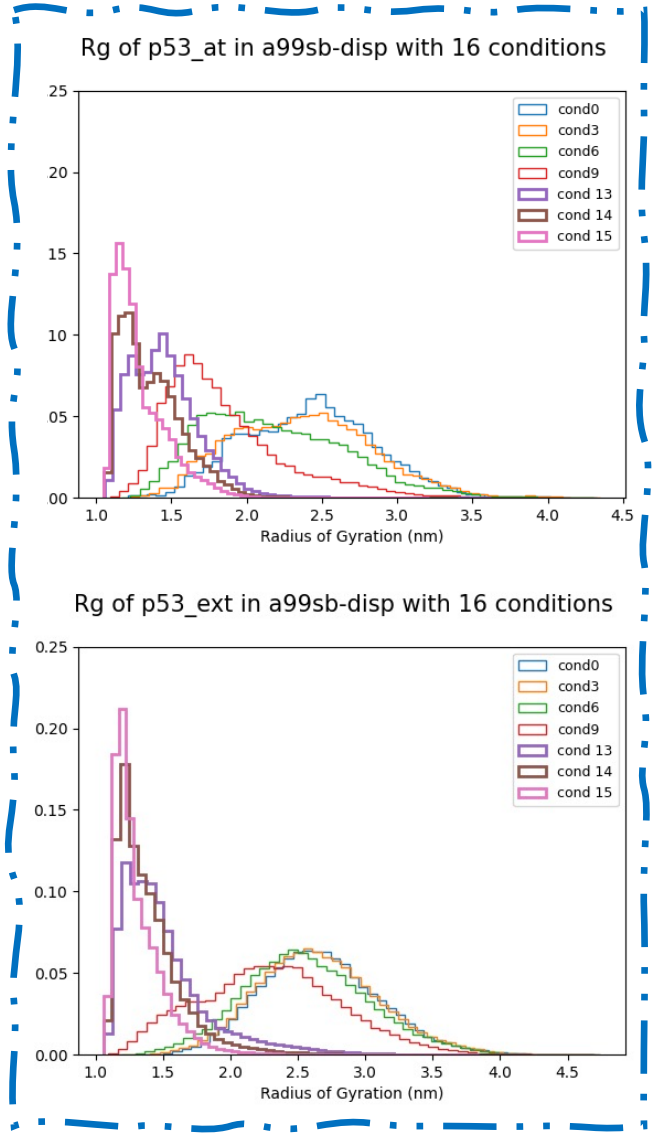
doi:10.1021/acs.jpcb.7b06768

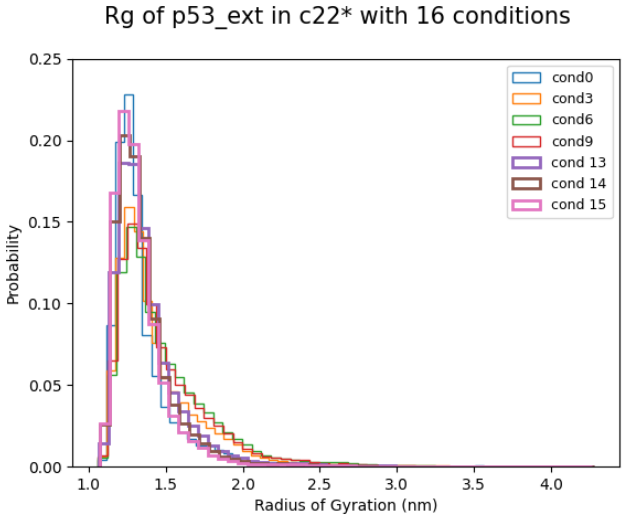
[Subaim2a]$ REST2 Incorrect IDP Ensembles Under High Temp

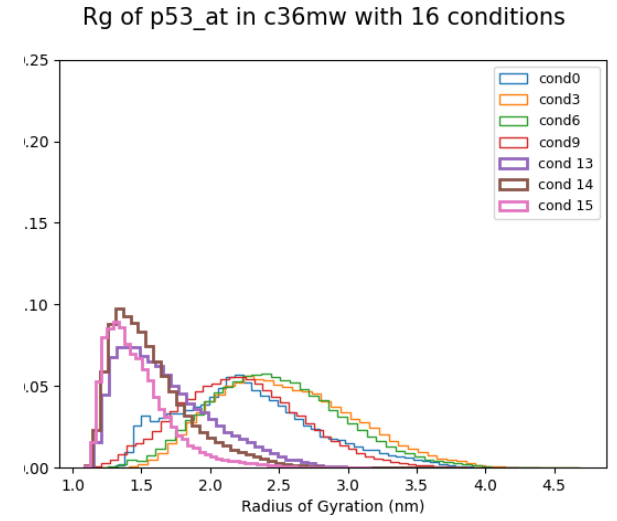
- **p53-NTD in multiple atomistic models coupled with REST2 protocol**



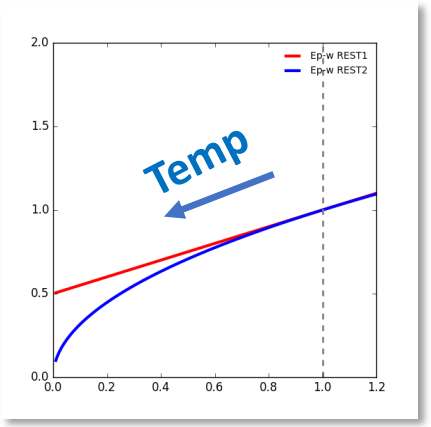








- **Severe Compact under high T:**
- Over down-scaled Protein-Water interactions.



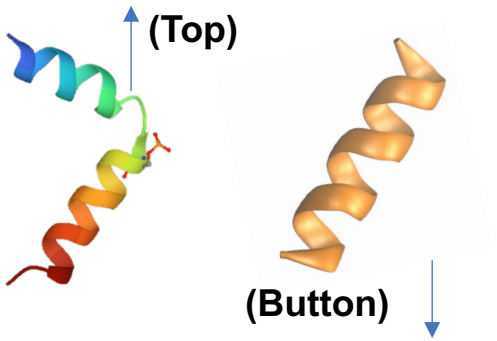
- **Unbalanced intra-/inter-molecular interactions.**

[Subaim2a]$ REST2 Unbalanced Intra/Inter molecular interactions

- **Multiple IDPs tested with REST2 protocol in a99sb-*disp* force field**

- **Templates:**

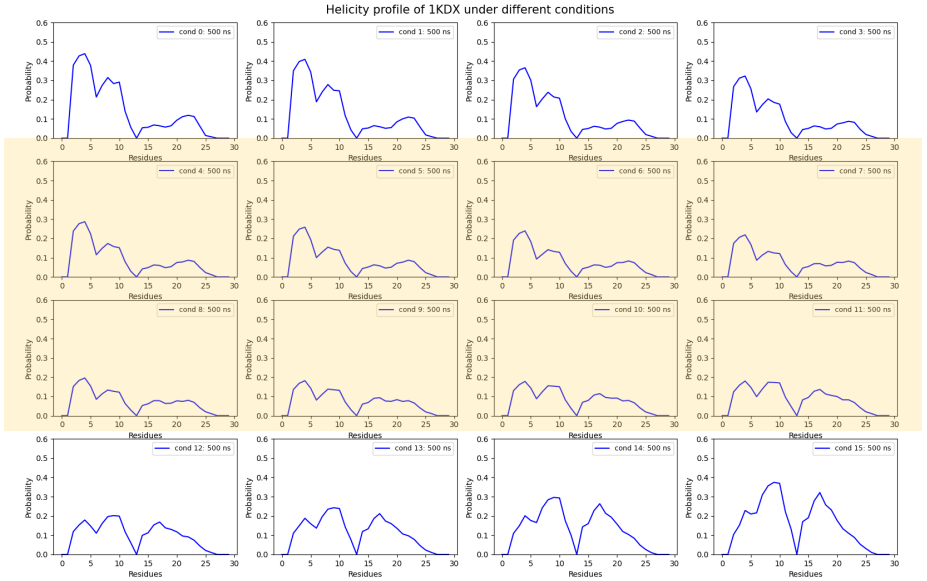
KIX (28 residues)

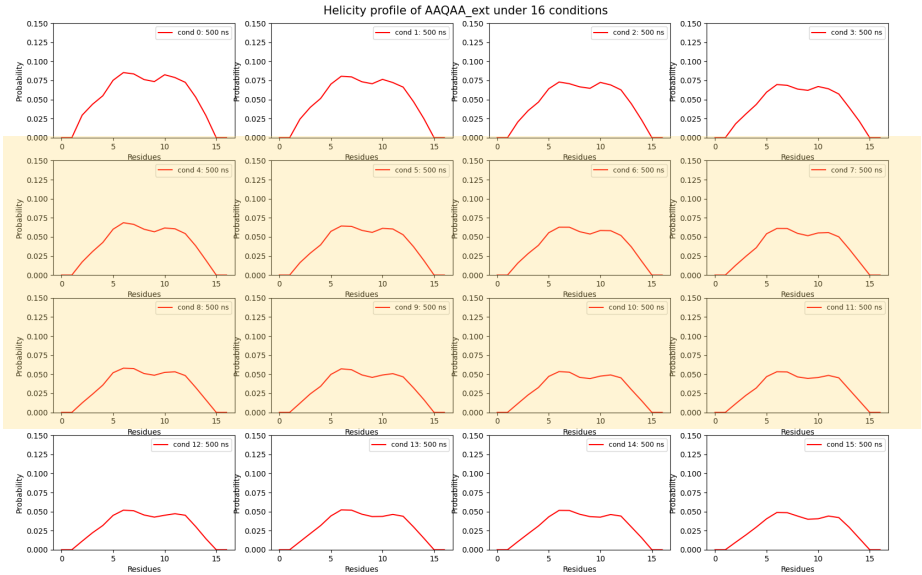


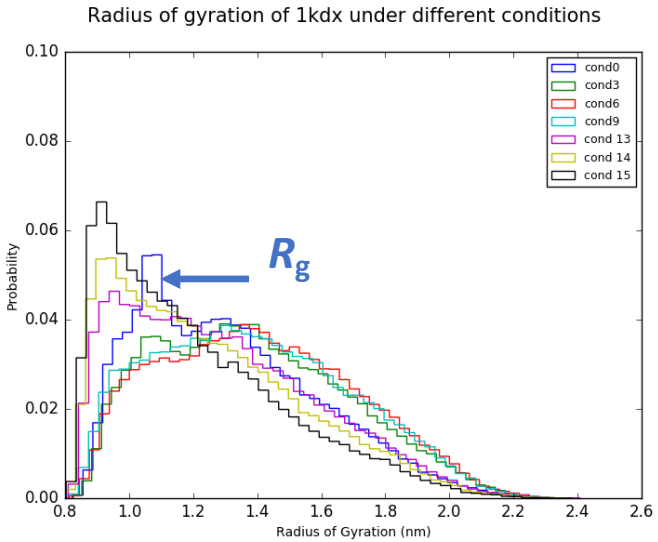
$(AAQAA)_3$ (15 residues)

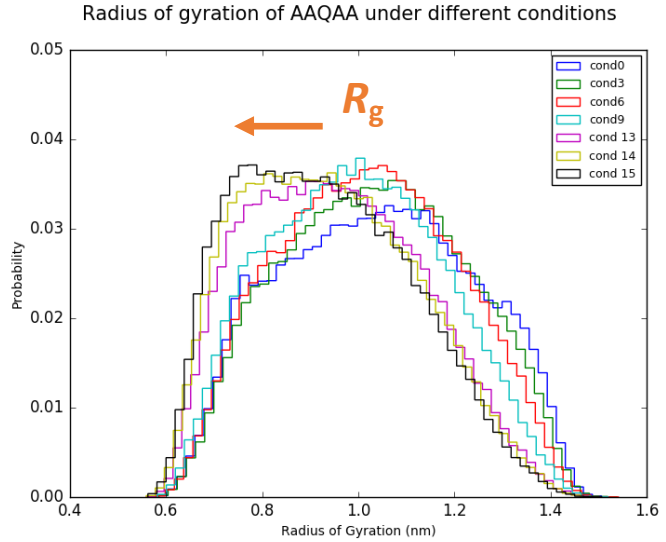
- **REST2 limitations:**

1. Insufficient **conformational sampling** between middle replicas.
2. Incorrect IDP over-compact ensembles under high temperature.









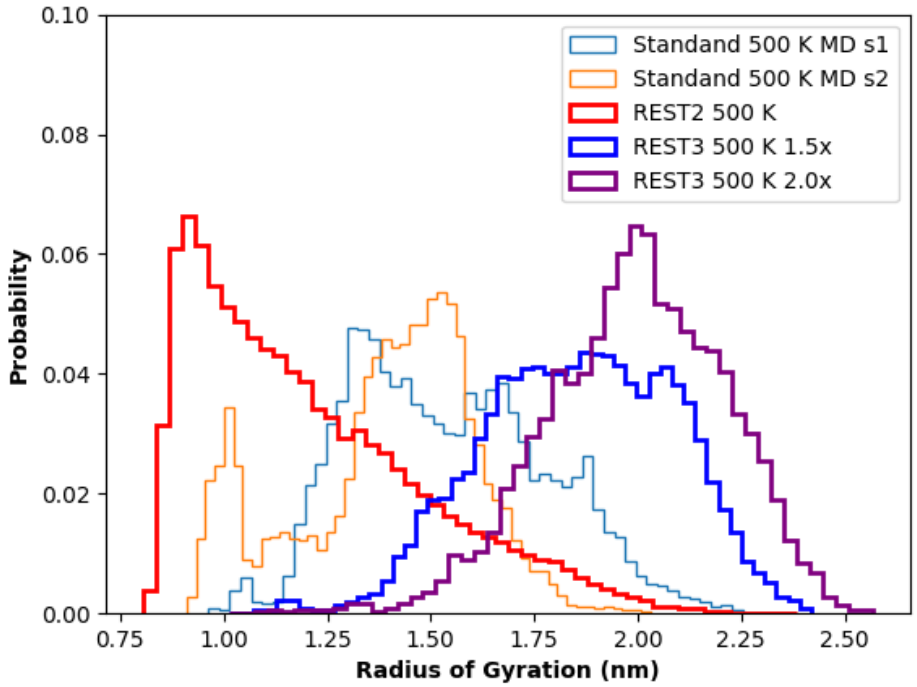
[Subaim2a]$ Enhanced Sampling Method: REST3

- **Strategy**

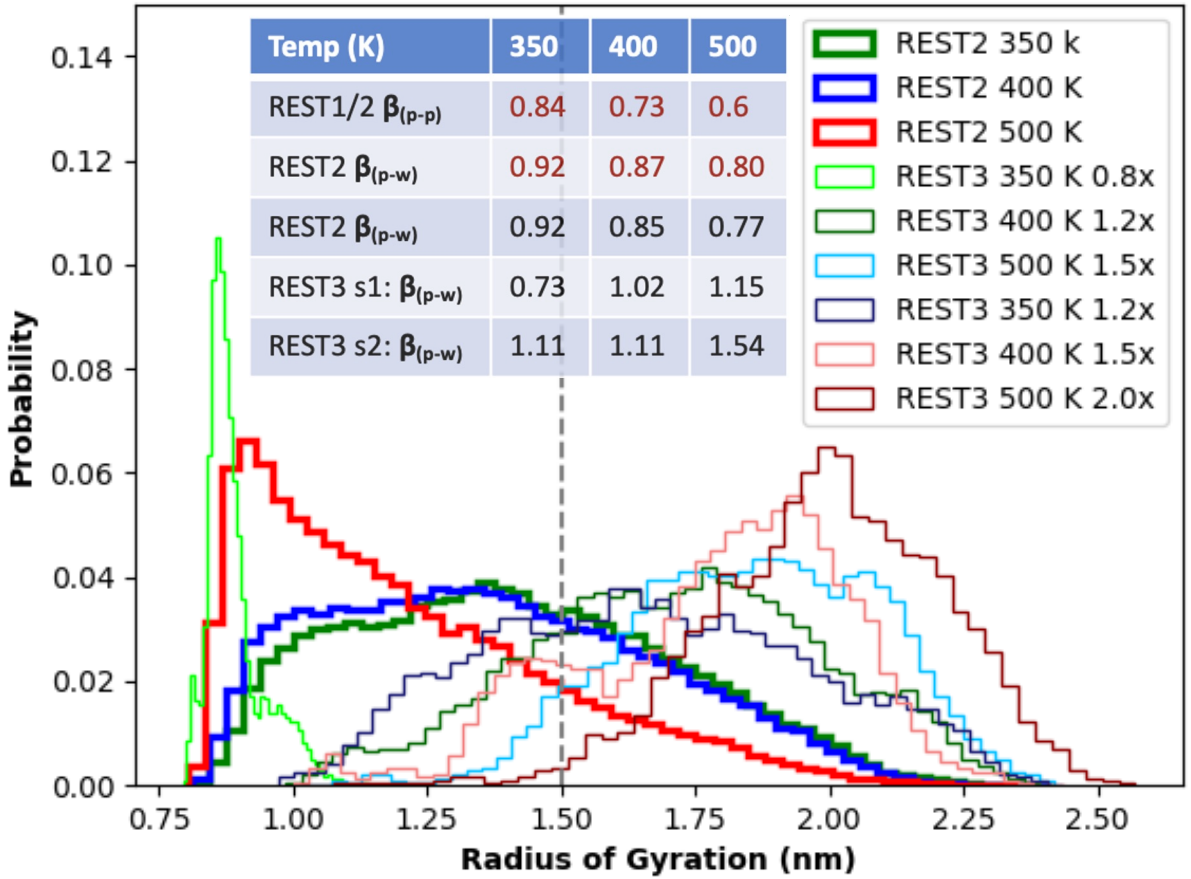
Folding: $\beta_f E_{pp} + \beta'_f E_{pw} + \beta_f E_{ww}$

Unfolding: $\beta_u E_{pp} + \beta_f' E_{pw} + \beta_f E_{ww}$

- **Step 1: Standard High Temperature MD simulations as guidance**



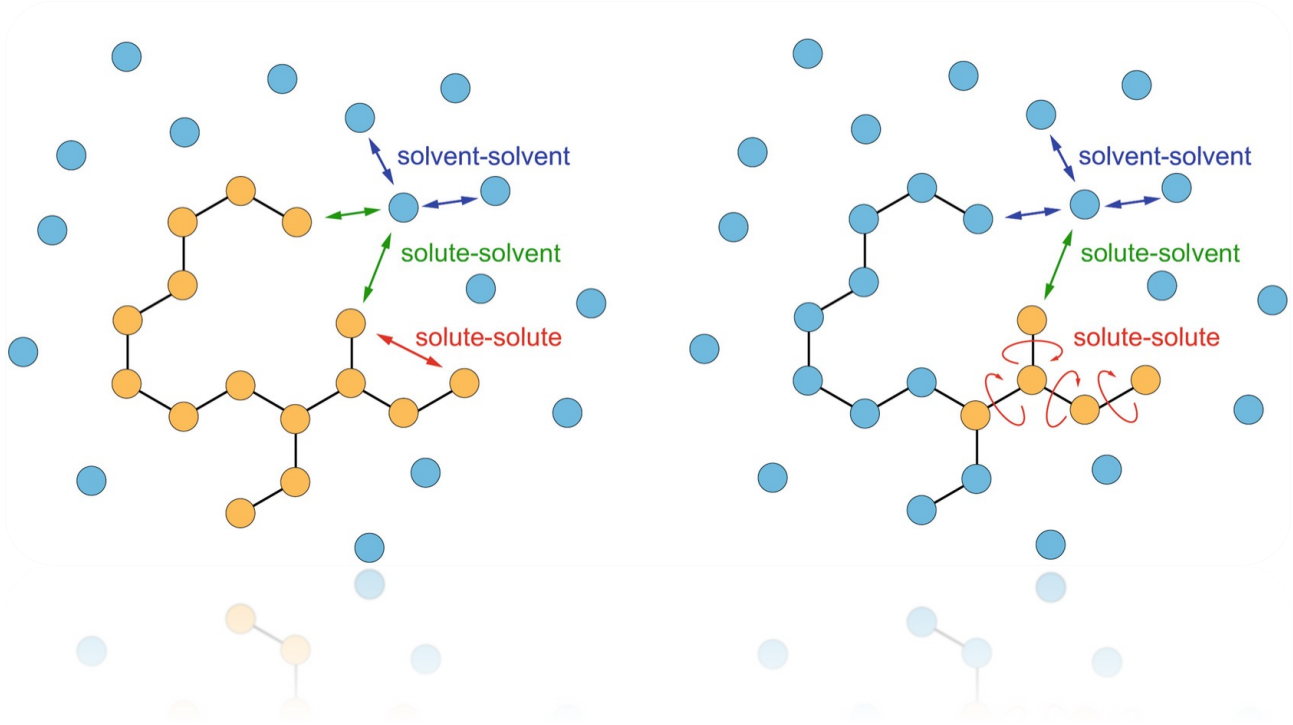
- **Step 2: Rescale P-W interactions to find the balance via vdW term**



- **Step 3: Set up replica exchange simulations with optimized REST3 p-p/p-w values.**

[Subaim2a]$ Potential Problems

❖ Sufficient conformational sampling?



❖ Exchange Rate?

❖ Appropriate for smaller and larger IDPs as well?

>_

[Yumeng@Prospectus]$ Part2: Specific Aims

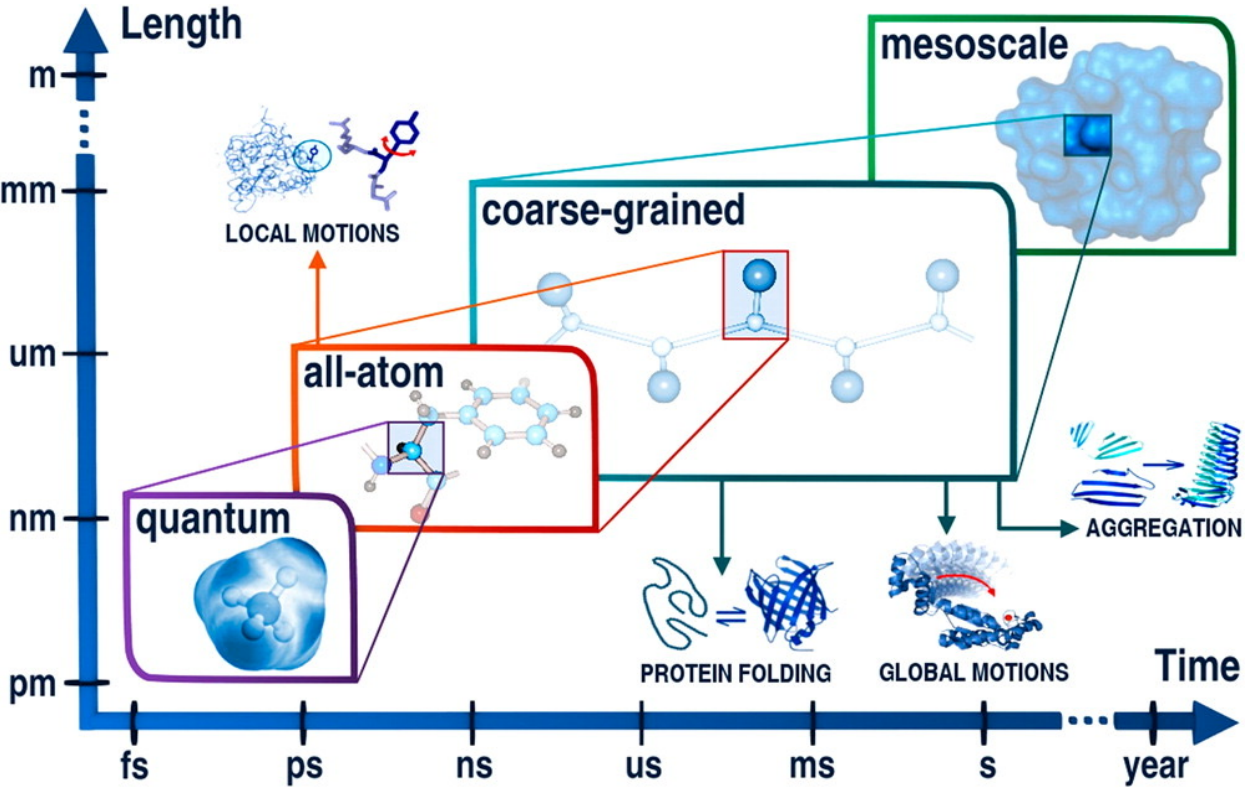
Aim1. Integration of MD simulations and experiments for IDP studies

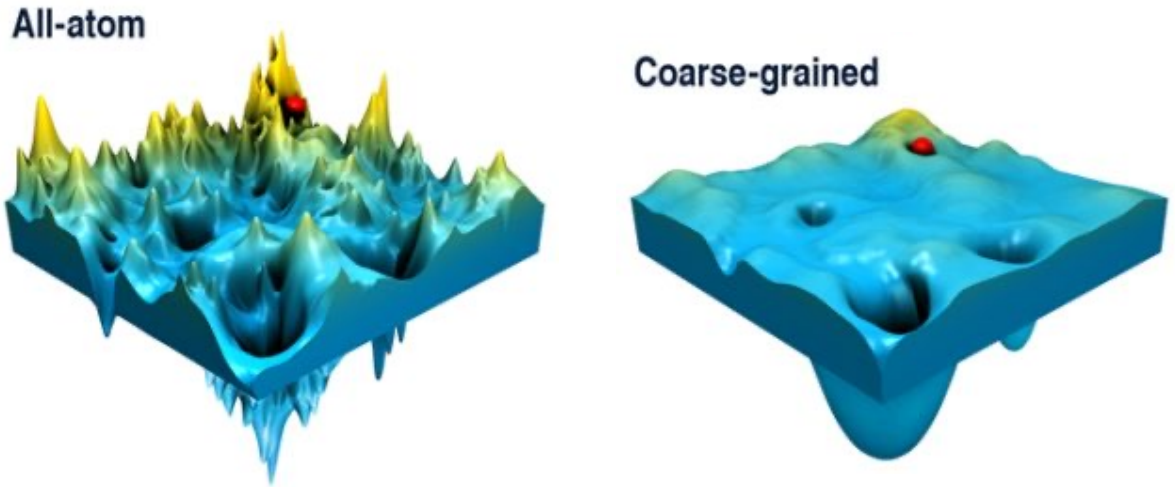
Aim2. Advanced methods for multi-scale simulations on IDPs

[Yumeng@Prospectus Aim2]$

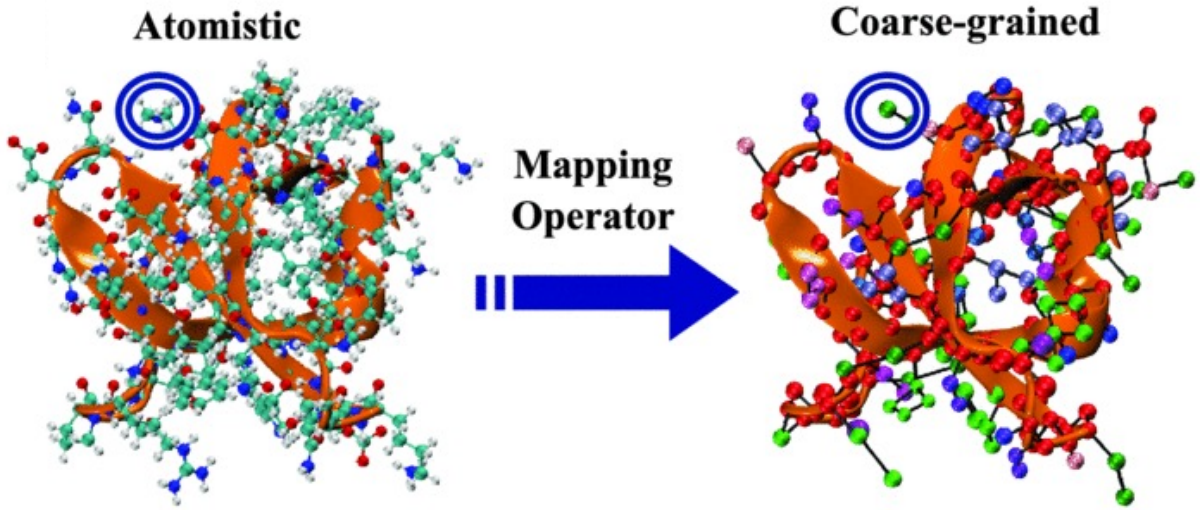
Subaim2a. Advanced sampling method: REST3

Subaim2b. Optimized force field: HyRes*





Neglect Non-native potentials



Reduced representation of interaction sites per residue.

To **Coarse Grain** a system**:**

- ❖ Energy based
- ❖ Force matching
- ❖ Structure based

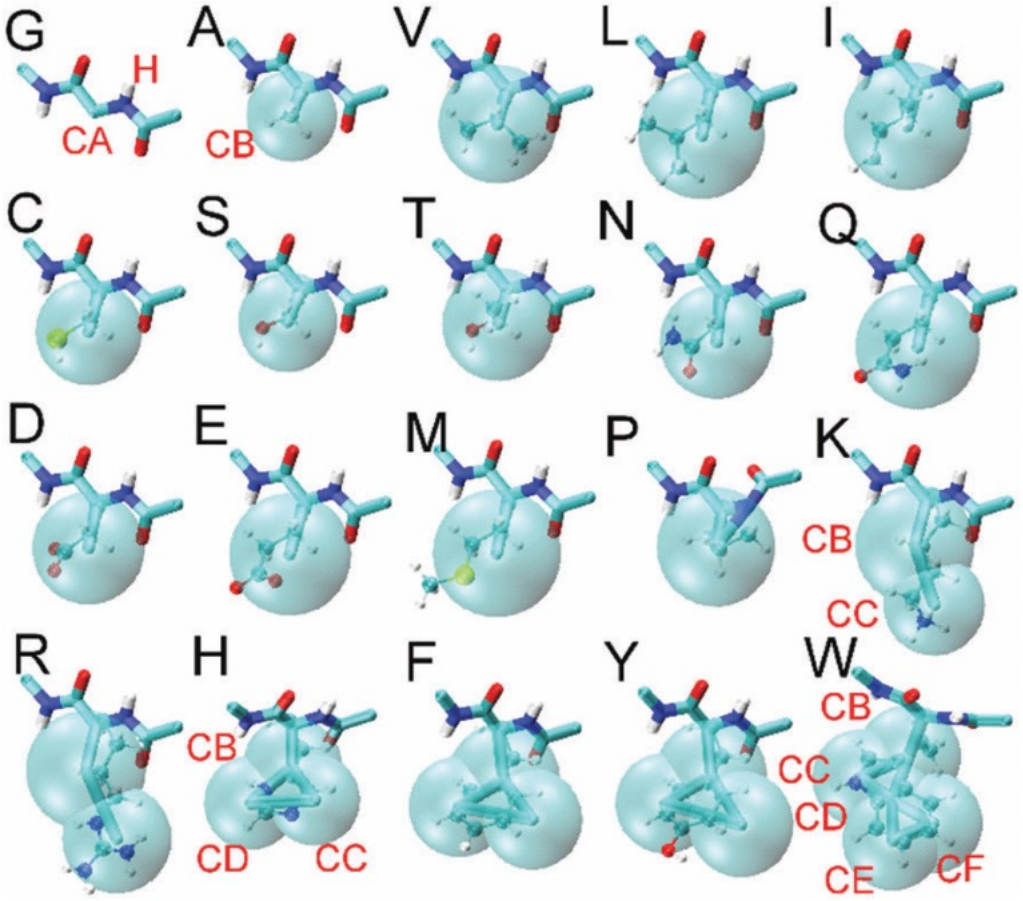
Most computational efficiency while maintaining adequate degree of details.

doi: 10.1063/1.4818908
doi: 10.1021/acs.chemrev.6b00163

[Subaim2b]$ HyRes Protein Model

- **Foundations**

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}} + U_{\text{CMAP}} + U_{\text{Hbond}} + U_{\text{LJ}} + U_{\text{elec}}$$



- **Non-bonded term**

$$U_{LJ} = \sum_{i,j} \boldsymbol{\varepsilon_{i,j}} \left[\left(\frac{r_{i,j}^{min}}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{i,j}^{min}}{r_{i,j}} \right)^{6} \right]$$

$$r_{i,j}^{min} = \frac{r_i^{min} + r_j^{min}}{2} \qquad \varepsilon_{i,j} = \sqrt{\varepsilon_i \varepsilon_j}$$

$$\boldsymbol{\varepsilon_i = 70\ \%\ \varepsilon_{i,\text{CHARMM19}}}$$

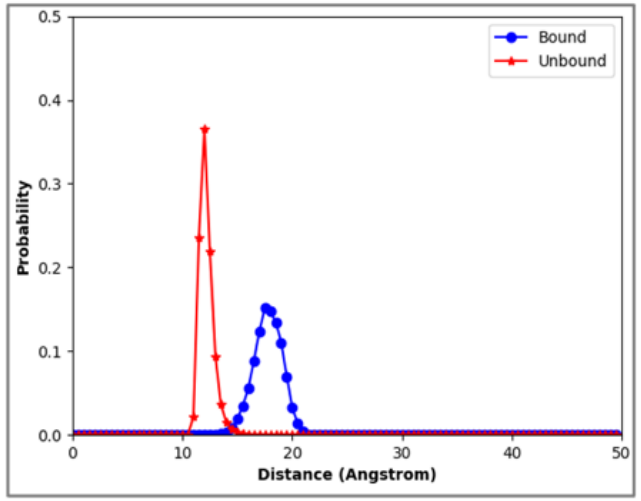
$$U_{elec} = \sum_{chrg.\,pairs} \frac{q_i q_j}{4\pi \boldsymbol{\varepsilon_r} \varepsilon_0 r_{i,j}} e^{-\frac{r_{i,j}}{D}}$$

$$\boldsymbol{\varepsilon_r = 20}$$

- **Over-compaction**

Lack solvation term.

$$W(\mathbf{r}) = V_{\text{solute}}(\mathbf{r}) + V_{\text{solvation}}(\mathbf{r})$$



DOI: 10.1039/c7cp06736d

[Subaim2b]$ 1st Adjustment: Weaken Intramolecular Interactions

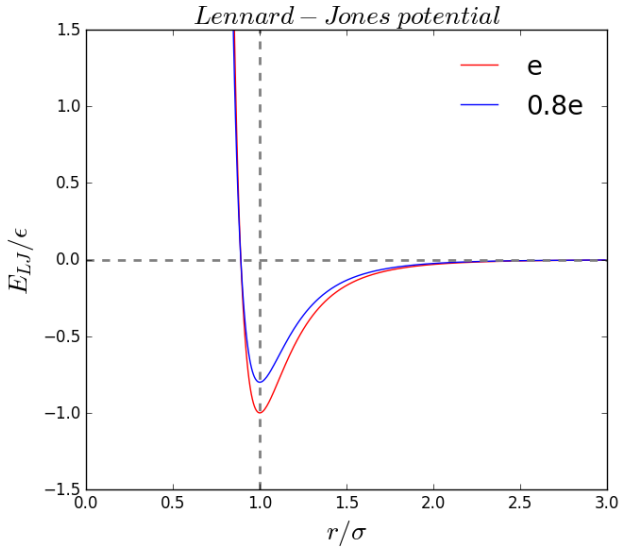
Table 3 Lennard-Jones parameters for all side chain beads

Residue	CG bead	ε_i (kcal mol^{-1})	$r_i^{min}/2$ (Å)	$r_i^{min}/2(1-4)$ (Å)
Ala	CB	−0.308	2.12	2.12
Val	CB	−0.62	2.75	2.75
Leu	CB	−0.9	2.96	2.96
Ile	CB	−0.772	2.97	2.97
Met	CB	−0.636	2.98	3.68
Asn	CB	−0.18	2.65	3.25
Asp	CB	−0.148	2.61	3.11
Gln	CB	−0.204	2.89	3.89
Glu	CB	−0.14	2.85	3.95
Cys	CB	−0.532	2.47	2.77
Ser	CB	−0.188	2.32	2.72
Thr	CB	−0.228	2.62	2.62
Pro	CB	−0.212	2.77	2.77
Lys	CB	−0.05	2.78	3.48
Lys	CC	−0.05	2.36	3.06
Arg	CB	−0.135	2.78	3.18
Arg	CC	−0.135	2.54	2.94
His	CB	−0.108	2.34	2.64
His	CC	−0.081	2.18	2.48
His	CD	−0.081	2.11	2.41
Phe	CB	−0.22	2.64	2.94
Phe	CC	−0.22	2.33	2.63
Phe	CD	−0.22	2.33	2.63
Tyr	CB	−0.197	2.64	2.94
Tyr	CC	−0.197	2.33	2.63
Tyr	CD	−0.0984	2.45	2.75
Trp	CB	−0.168	2.42	2.72
Trp	CC	−0.084	2.24	2.54
Trp	CD	−0.168	2.09	2.39
Trp	CE	−0.168	2.33	2.63
Trp	CF	−0.168	2.33	2.63

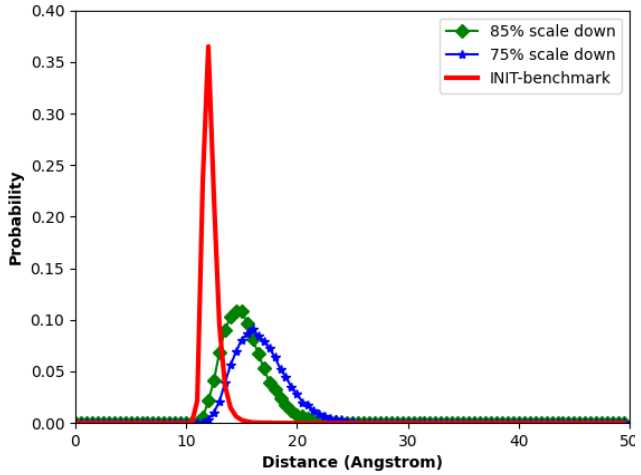
- Scale down sidechain vdW interaction strength. (intramolecular interactions)

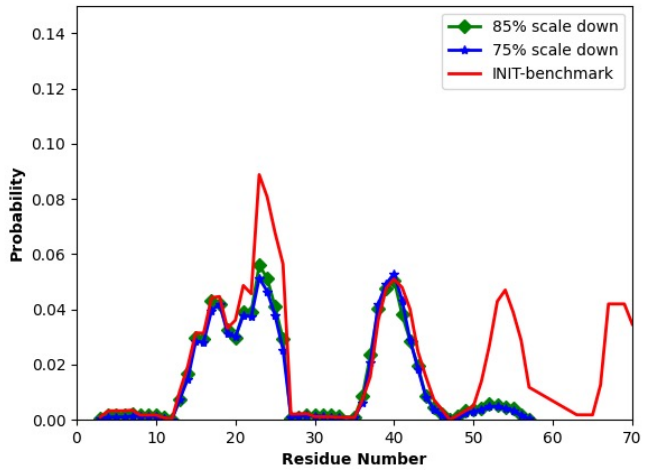
$$U_{LJ} = \sum_{i,j} \varepsilon_{i,j} \left[\left(\frac{r_{i,j}^{min}}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{i,j}^{min}}{r_{i,j}} \right)^{6} \right]$$

DOI: 10.1039/c7cp06736d



- Very limited influence in compensating the loss of dispersion (template: p53-NTD).





[Subaim2b]$ 2nd Adjustment: Introduce Solvation Term

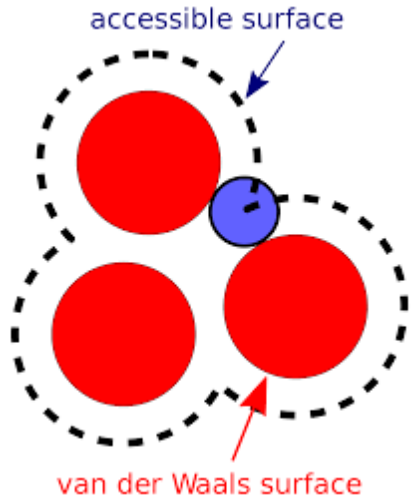
- **Solvent-accessible surface area (SASA) :**

$$W(\mathbf{r}) = V_{\text{solute}}(\mathbf{r}) + V_{\text{solvation}}(\mathbf{r})$$

$$V_{\text{solvation}}(\mathbf{r}) = \sum \sigma_i A_i(\mathbf{r})$$

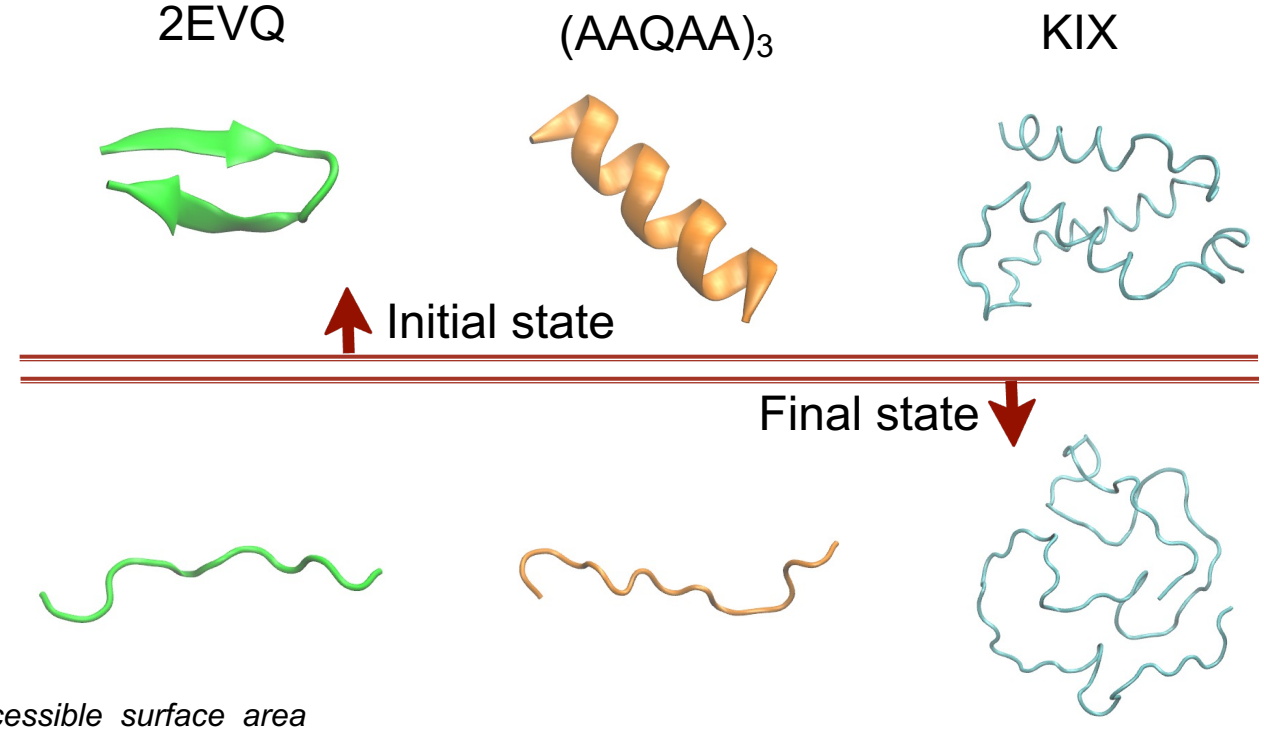
$$A_i(\mathbf{r}) = S_i \prod_{j \neq i}^{M} [1 - p_i p_{ij} b_{ij}(r_{ij})/S_i]$$

$$S_i = 4\pi(R_i + R_{\text{probe}})^2$$



- **Three templates for testing**
  - 2EVQ (beta sheet)
  - $(AAQAA)_3$ (helix)
  - KIX (helix)

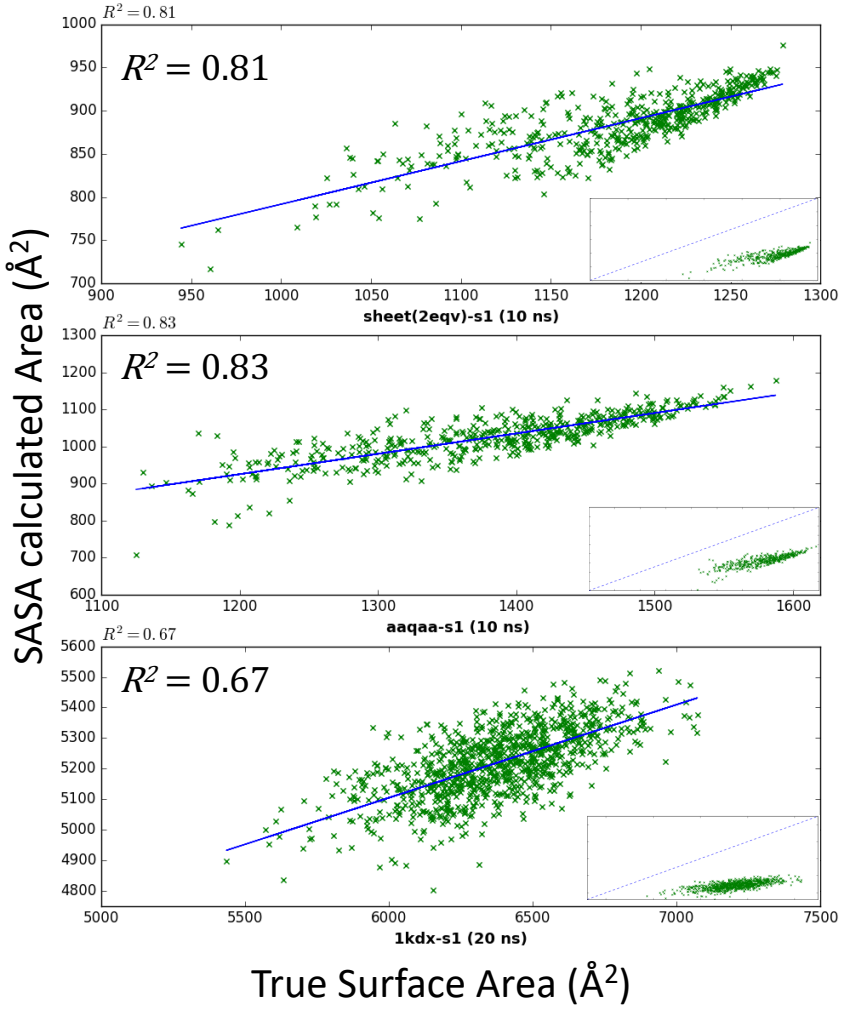
Simulations: under 450 K, from folded to extended conformations.



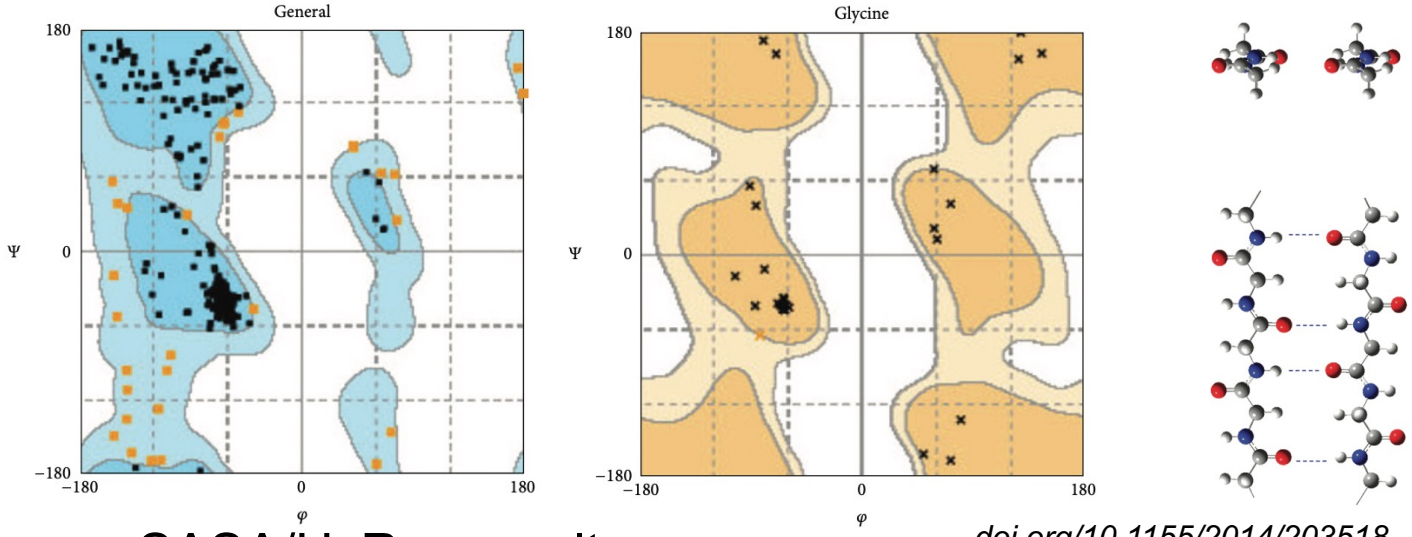
DOI: 10.1002/prot.10001
https://en.wikipedia.org/wiki/Accessible_surface_area

[Subaim2b]$ Implicit Solvent Model Help to Decrease Compaction

- SASA correlates well with ture surface calculated value

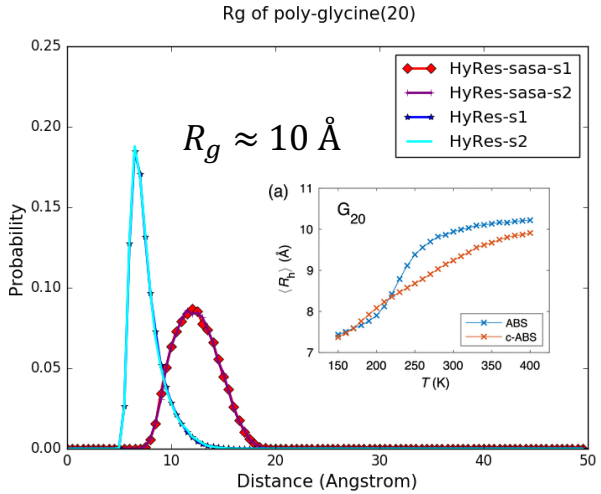


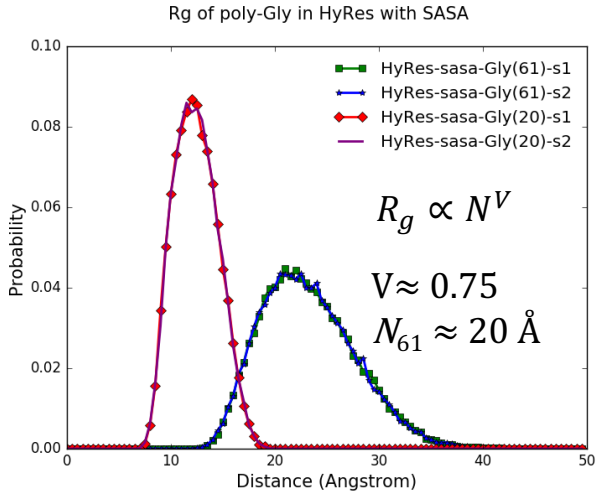
- Testing template: poly-Glycine
  - Ramachandran Plot



doi.org/10.1155/2014/203518

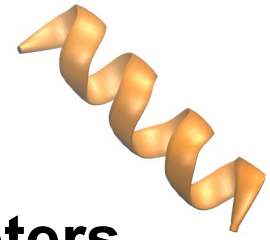
- SASA/HyRes results





[Subaim2b]$ Potential Problems

- Poly-Alanine for secondary structures tuning.



Force field parameters

$$U_{\text{dihedral}} = \sum_{\text{dihedrals}} k_\chi[1+\cos(n\chi-\delta)] \quad (5)$$

$$U_{\text{improper}} = \sum_{\text{impropers}} k_\psi(\psi-\psi_0)^2 \quad (6)$$

$$U_{\text{Hbond}} = \sum_{\text{Hbonds}} \varepsilon_{\text{HB}}\left[5\left(\frac{\sigma}{r}\right)^{12}-6\left(\frac{\sigma}{r}\right)^{10}\right]\cos^4\theta_{\text{AHD}} \quad (7)$$

$$U_{\text{CMAP}} = \sum_{\text{non-Gly,non-Pro residues}} U_{\text{CMAP}}(\phi,\psi) \quad (8)$$

Implicit solvent model

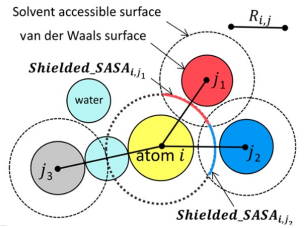
Atom type	$R_{\min}^{\text{vdW}}$ [a] (Å)	R_i [b] (Å)	p_i [b]	σ_i [c] (kcal mol^{-1} Å^{-2})	Description
C	2.1	1.72	1.554	0.012	Carbonyl carbon
CH1E	2.365	1.80	1.276	0.012	Extended aliphatic carbon with 1 hydrogen
CH2E	2.235	1.90	1.045	0.012	Extended aliphatic carbon with 2 hydrogens
CH3E	2.165	2.00	0.880	0.012	Extended aliphatic carbon with 3 hydrogens
CR1E	2.1	1.80	1.073	0.012	Extended aromatic carbon with 1 hydrogen
NH1	1.6	1.55	1.028	−0.060	Amide nitrogen
NR	1.6	1.55	1.028	−0.060	Aromatic nitrogen with no hydrogens
NH2	1.6	1.60	1.215	−0.060	Nitrogen bound to two hydrogens
NH3	1.6	1.60	1.215	−0.060	Quaternary ammonium nitrogen
NC2	1.6	1.55	1.028	−0.060	Proline nitrogen
N	1.6	1.55	1.028	−0.060	Hydroxyl oxygen
OH1	1.6	1.52	1.080	−0.060	Carbonyl oxygen
O	1.6	1.50	0.926	−0.060	Carbonyl oxygen
OC	1.6	1.70	0.922	−0.060	Aliphatic/aromatic sulfur
S	1.89	1.80	1.121	0.012	Extended aliphatic sulfur with 1 hydrogen
SH1E	1.89	1.80	1.121	0.012	Polar hydrogen
H	0.8	1.10	1.128	0.000	Polar hydrogen (in Arg, Lys and N-term)
HC	0.6	1.10	1.128	0.000	

[a]The CHARMM PARAM19 van der Waals radii are given as a basis of comparison but are not used in the solvation term.

```
20    1
MASS    1 H     1.008   ! amide H in backbone
MASS   31 N     14.007  ! backbone N (Pro)
MASS   38 NH1   14.007  ! backbone N (non-Pro)
MASS   11 C     12.011  ! carbonyl C in backbone
MASS   51 O     15.999  ! carbonyl O in backbone
MASS   12 C1E   13.019  ! CA in backbone (non-Gly)
MASS   13 C2E   14.027  ! CA in backbone (Gly)
MASS   14 C3E   15.035  ! CH3 in Ala, and capping group
MASS   10 CT    12.01100 ! aliphatic carbon
```

- ❖ **Sidechain shielding**
- ❖ **Smaller calculated surface value**
- ❖ **…**

DOI: 10.1002/prot.10001

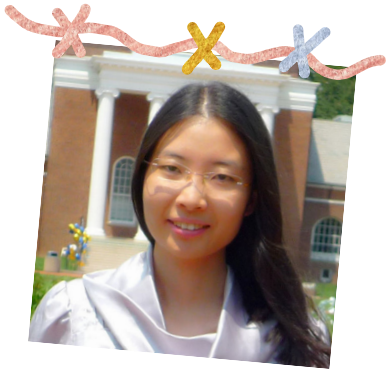


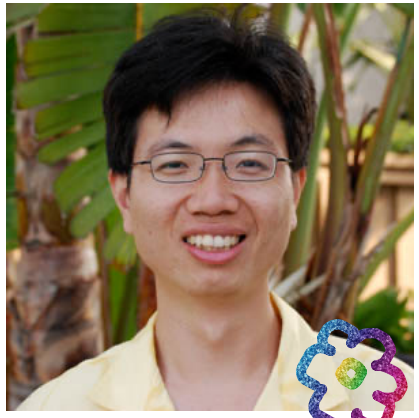
Acknowledgment

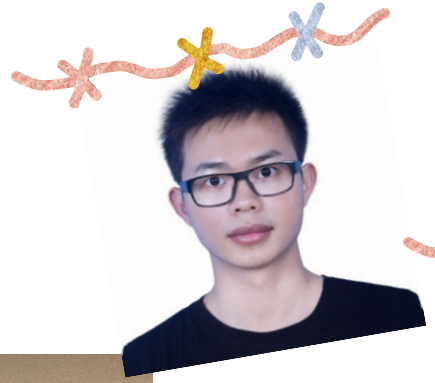
- Prof. Jianhan Chen

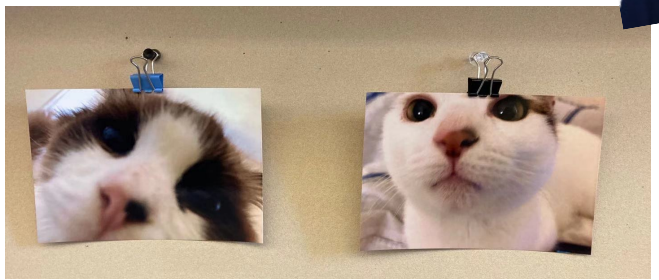
- Prof. Min Chen
- Prof. Scott Auerbach
- Prof. Greg Grason

- Chen Lab Members:
  - Dr. Xiaorong Liu
  - Dr. Zhiguang Jia
  - Dr. Mahdieh Yazdani
  - Xiping Gong,
  - Erik Nordquist
  - Qianlan Jia
  - Jian Huang









Thank you !