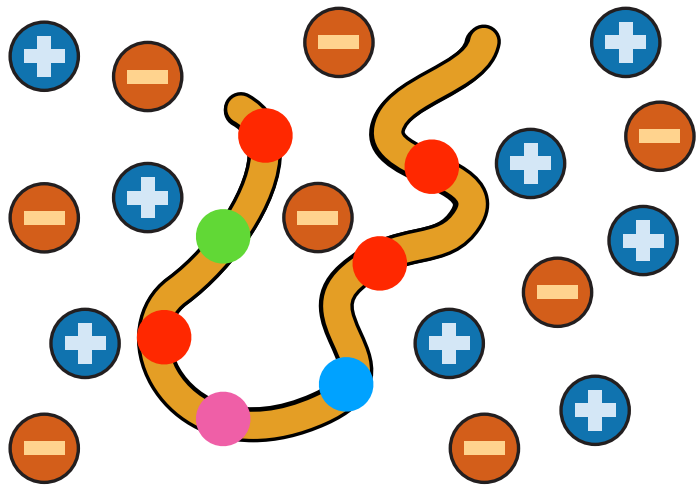
Sequence dependent regulation of IDP conformation and function

Kingshuk Ghosh

University of Denver
Department of Physics and Astronomy

IDPs have different regulators



(Salt), pH,

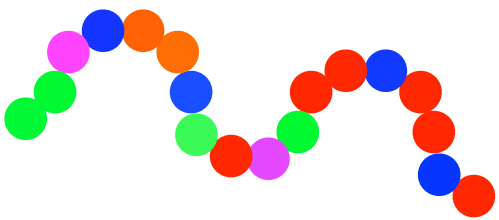
Environment

IDPs rely on electrostatics, charges are topologically correlated

IDP conformation can also be modulated by non-electrostatics

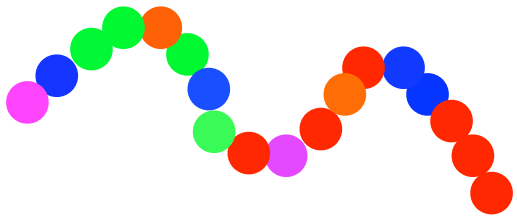
IDPs may also exhibit charge modulation

IDP sequence is critical



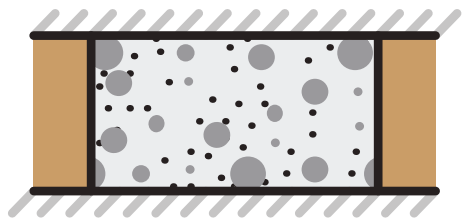
6R4G4B2B2B

≠

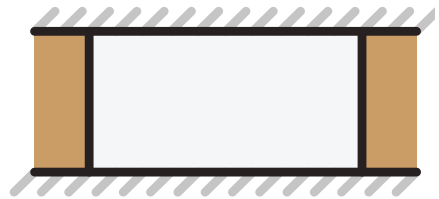


6R4G4B2B2B

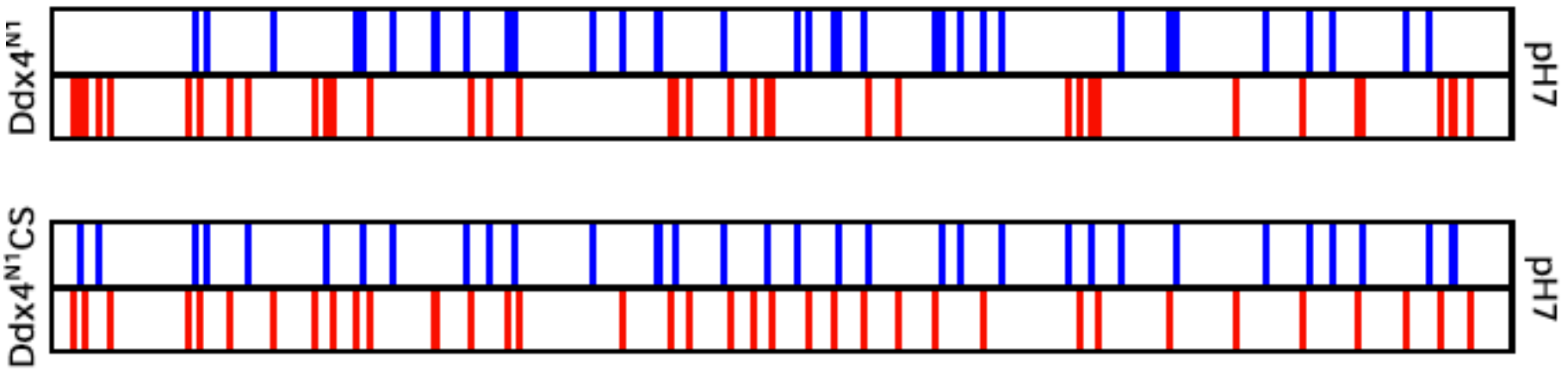
Sequence matters in IDP function



LLPS



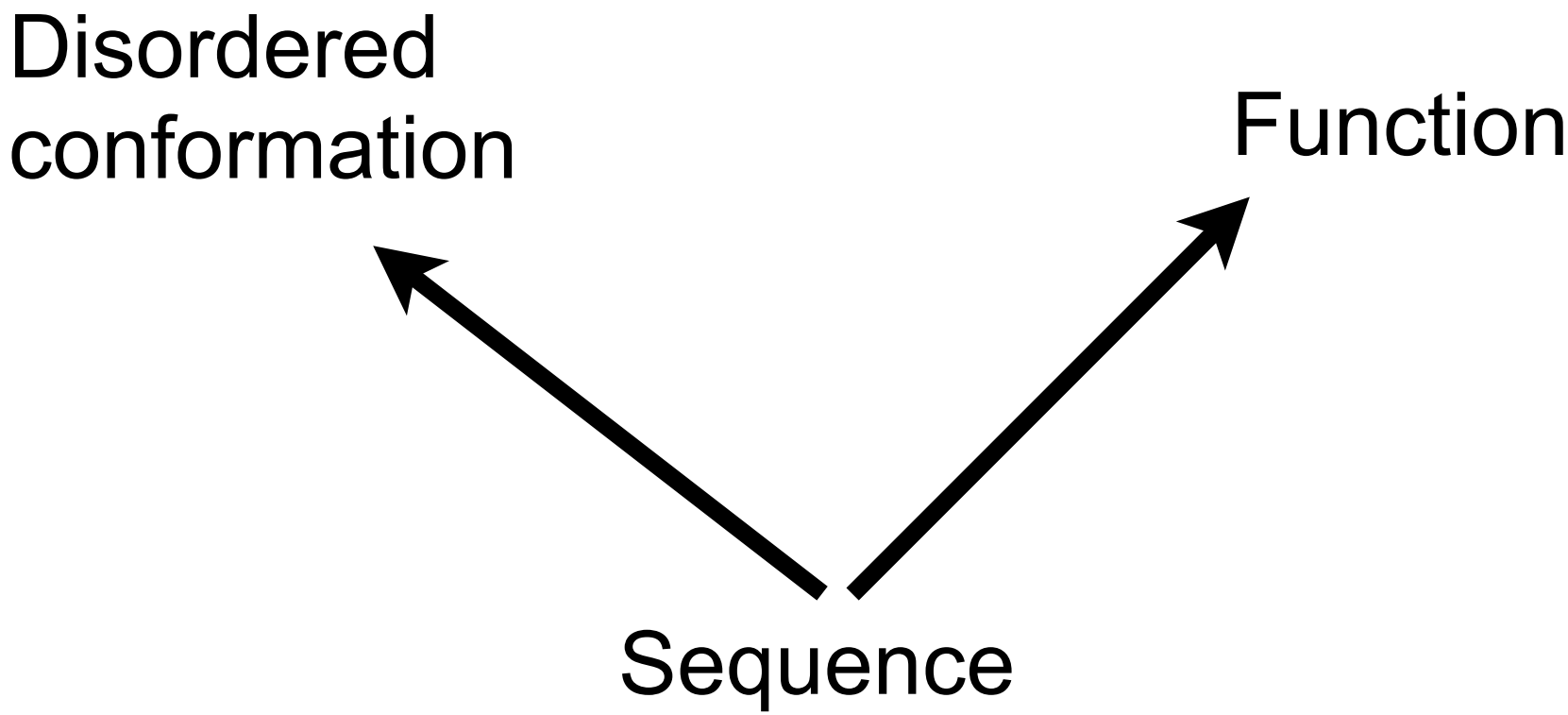
No LLPS



Ddx4N phase separates but Ddx4CS does not

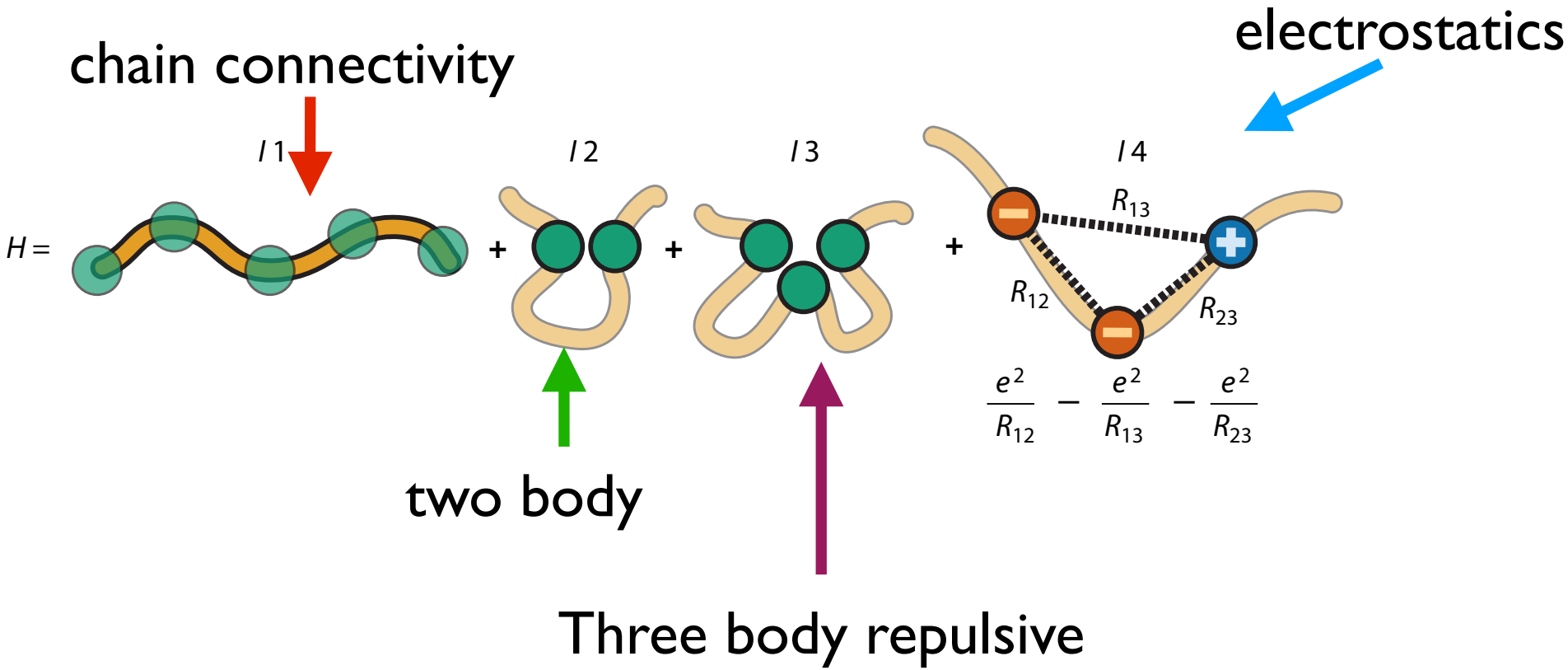
Nott et al Mol Cell (2015)

Linking sequence to conformation and function

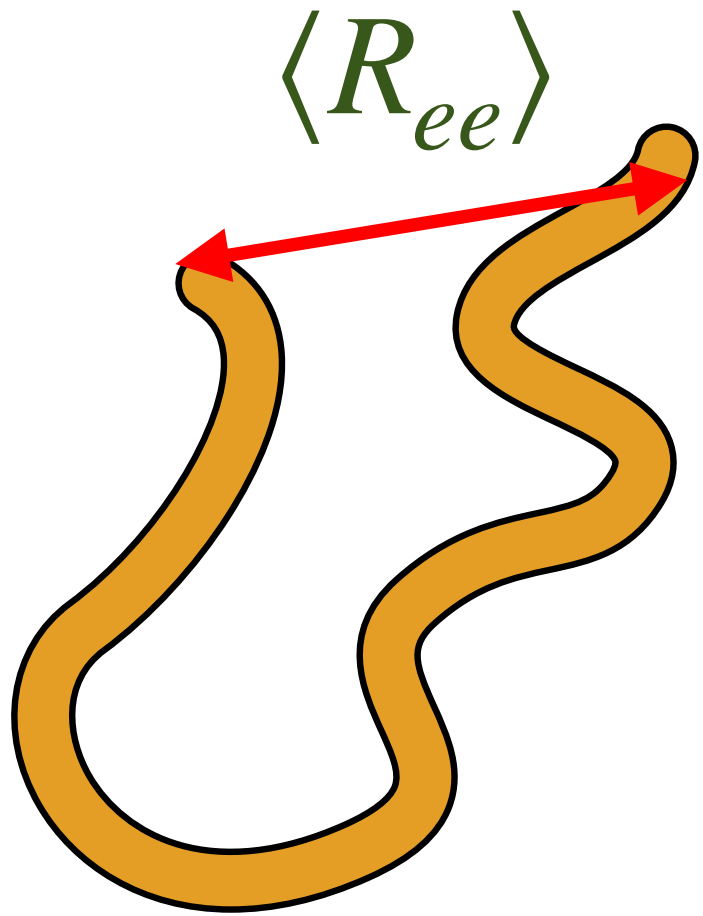


Hamiltonian (coarse grain) based theory

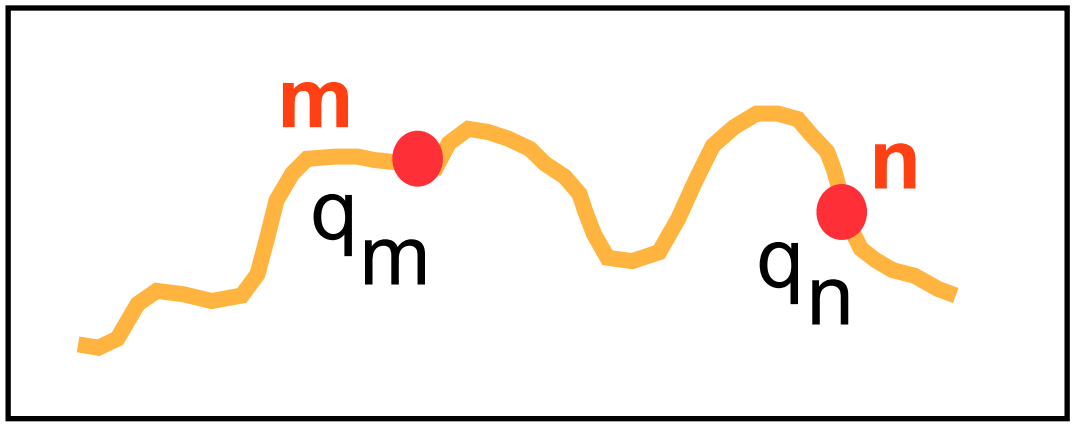
H based polymer theory can be useful



Theory can compute ensemble average end-to-end distance



H based polymer theory can describe charge correlation



electrostatics

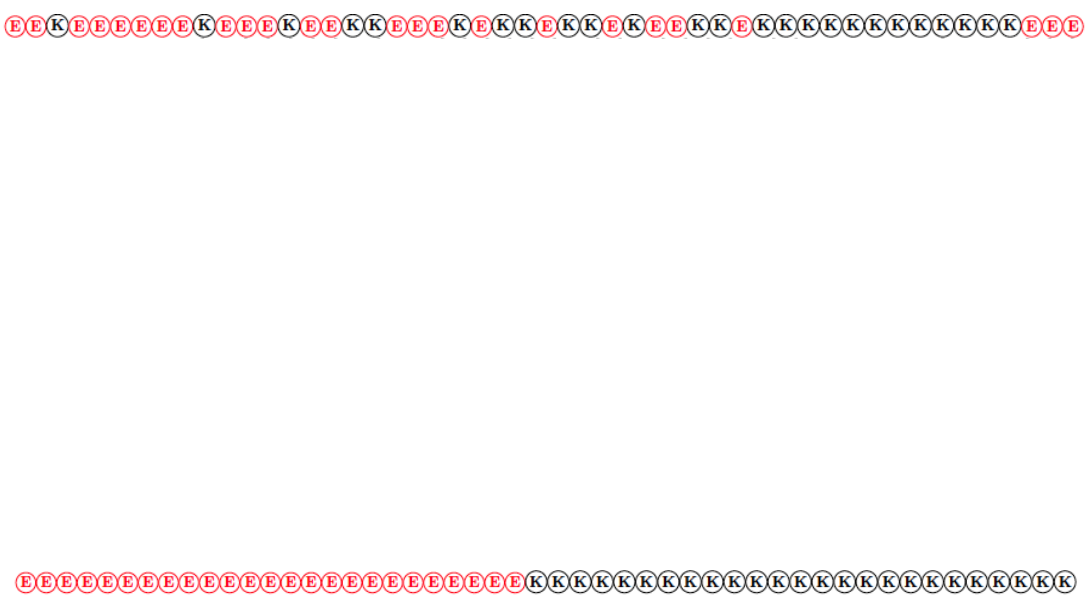
$$\beta F(R_{ee}) = -S_{ent}(R_{ee}) + \Omega_{non-elec}(R_{ee}) + Q_{elec} f(R_{ee})$$

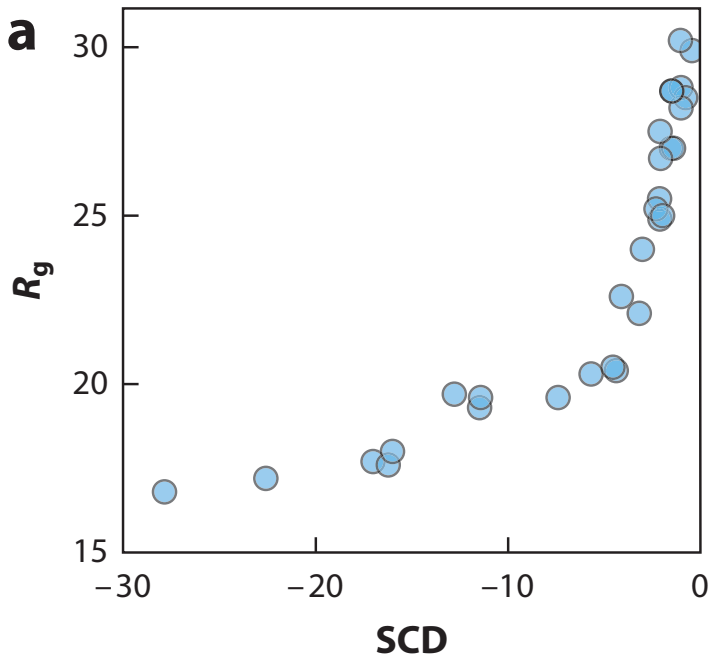
$$\text{SCD} = Q = \frac{1}{N} \sum_{m=2}^{N} \sum_{n=1}^{m-1} q_m q_n (m-n)^{1/2}$$

sequence specificity

Theory captures all-atom simulation of toy sequences

$$\text{SCD} = Q = \frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1} q_m q_n (m-n)^{1/2}$$





Sawle and Ghosh (JCP 2015)

Ghosh, Huihui, Phillips, Haider Annual Reviews of Biophysics (2022)

Modeling the Disordered Proteins (IDPs)

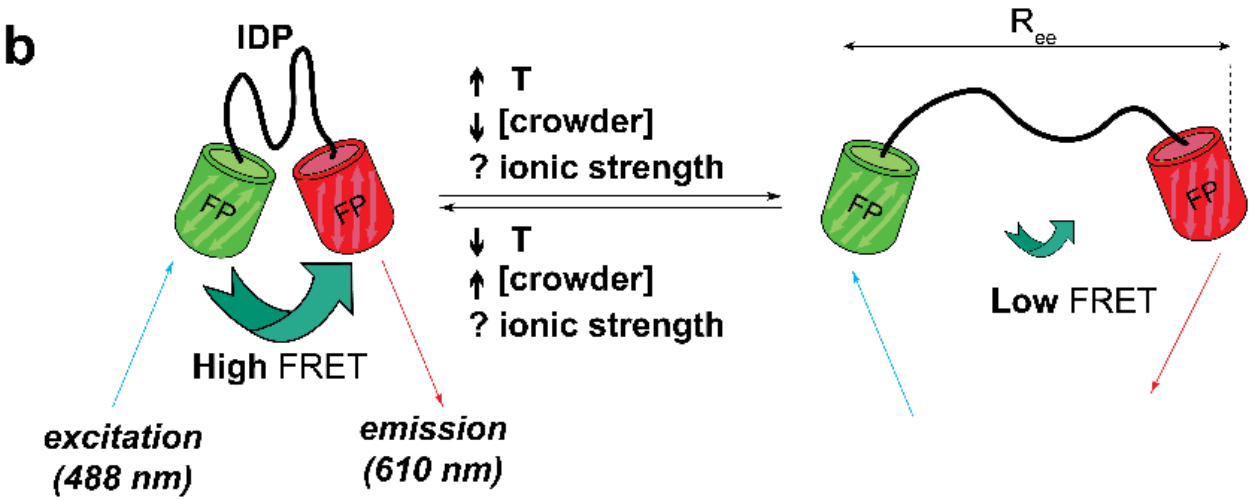
Experiment confirms predicted mutational hot spots

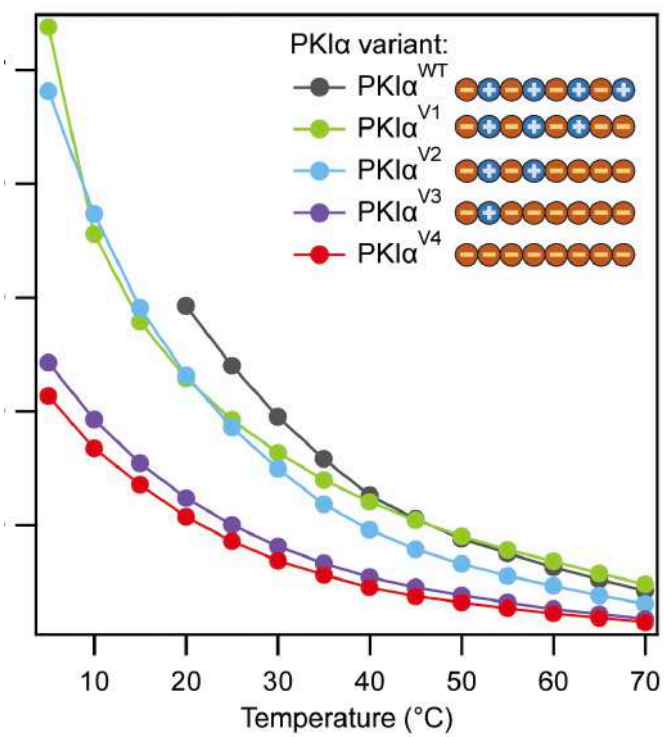
PKI-alpha—WT (7 positive and 14 negative charges)

PKI-alpha—v4 (3 positive and 18 negative charges)

Prediction → $R_{ee, WT} < R_{ee, v4}$

Experiment with Gruebele





Over 35 possibilities

PNAS 121, e2316408121 (2024)

Experiment confirms predicted mutational hot spots

HYPK-WT (22 positive and 31 negative charges)

Prediction → $R_{ee,\,v1} < R_{ee,\,v2}$

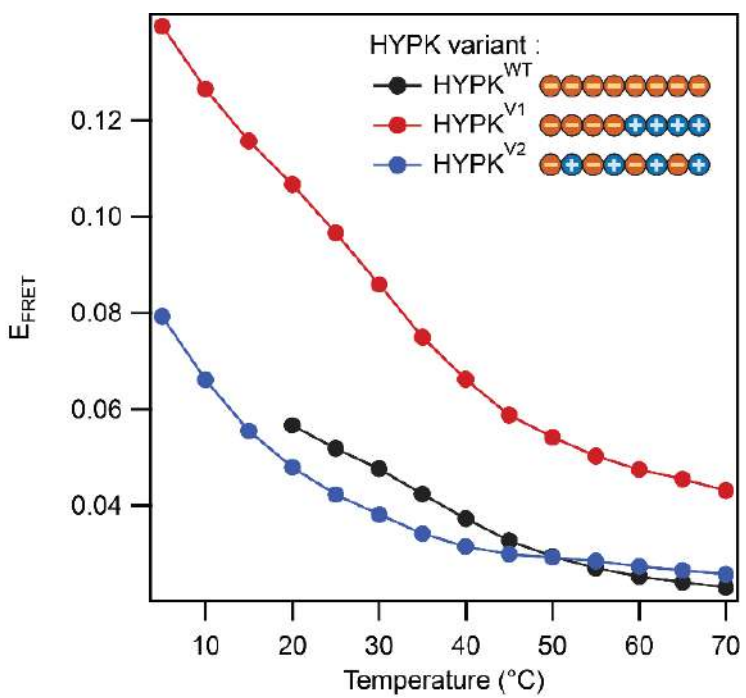
HYPK-4K v1

MRRRGEIDMATEGDVELELETETSGPERPPEKPRKHDSGAADLERVTDYAEEKEIQSSNLETAMSVIGDRRSREQKAKQEREKELAKVTIK
KEDLELIMTEM**K**ISRAAA**K**RSLR**K**HMGNVV**K**ALIALTN

HYPK-4K v2

MRRRGEIDMATEGDV**K**LEL**K**TETSGPERPPEKPRKHDSGAADLERVTDYA**K**EKEIQSSNLETAMSVIGDRRSREQKAKQEREKELAKVTIK
KEDLELIMT**K**MEISRAAAERSLREHMGNVVEALIALTN

Over 31000 possibilities



PNAS 121, e2316408121 (2024)

Discovery of a marginal IDP and its sequence dependence

PNAS 121, e2316408121 (2024)

Salt as another regulator of conformation

How to model salt dependence ?

$$\beta F(R_{ee}) = -S_{ent}(R_{ee}) + \Omega_{non-elec}(R_{ee}) + Q'_{elec}(R_{ee}, c_s)$$

$$Q'_{elec}(R_{ee}, c_s) \approx A_1 \sum_m \sum_n q_m q_n (m-n)^{1/2} - A_2 c_s^{1/2} \sum_m \sum_n q_m q_n (m-n) + \ldots$$

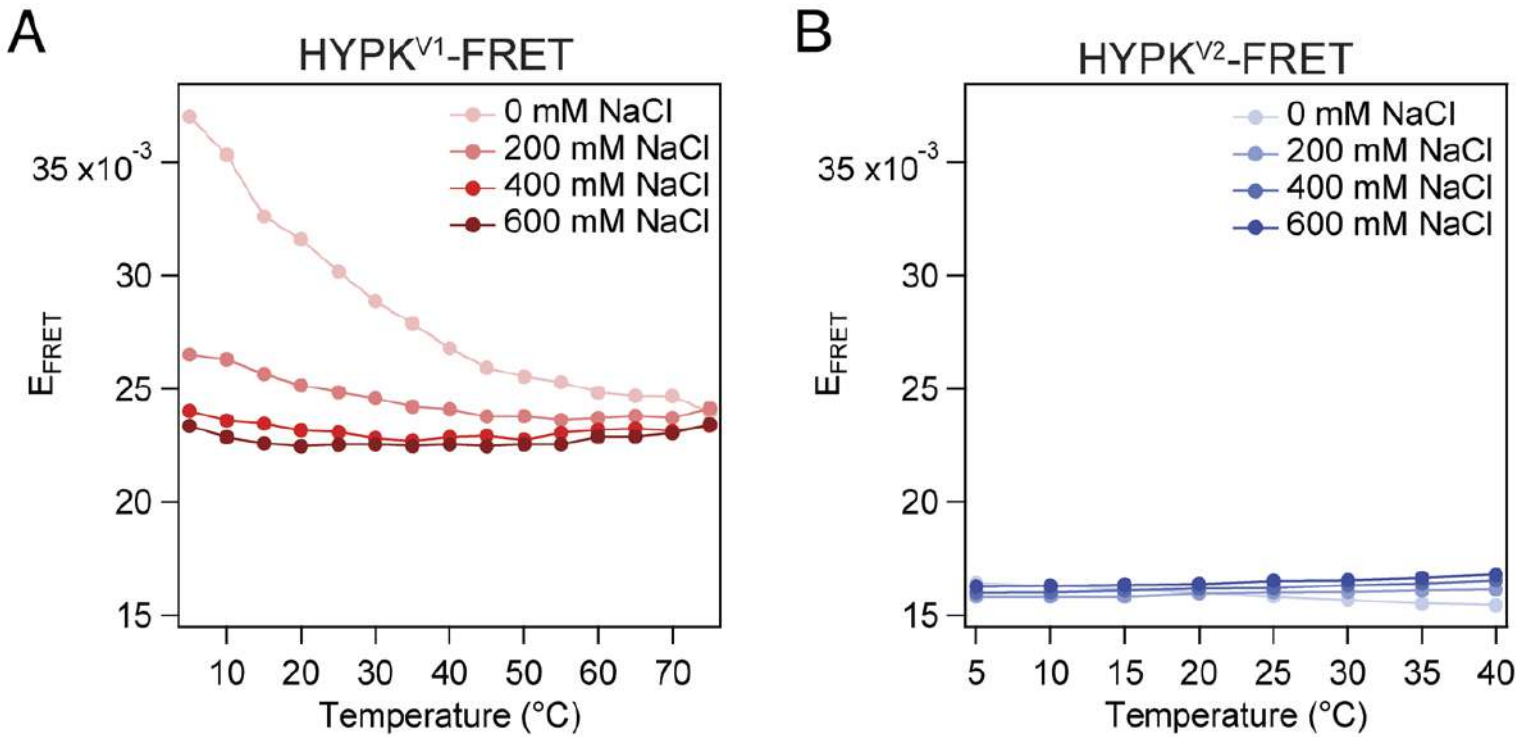
SCD

$SCD_{lowsalt}$

$SCD_{lowsalt}$ **> 0 shrink**

$SCD_{lowsalt}$ **< 0 expand**

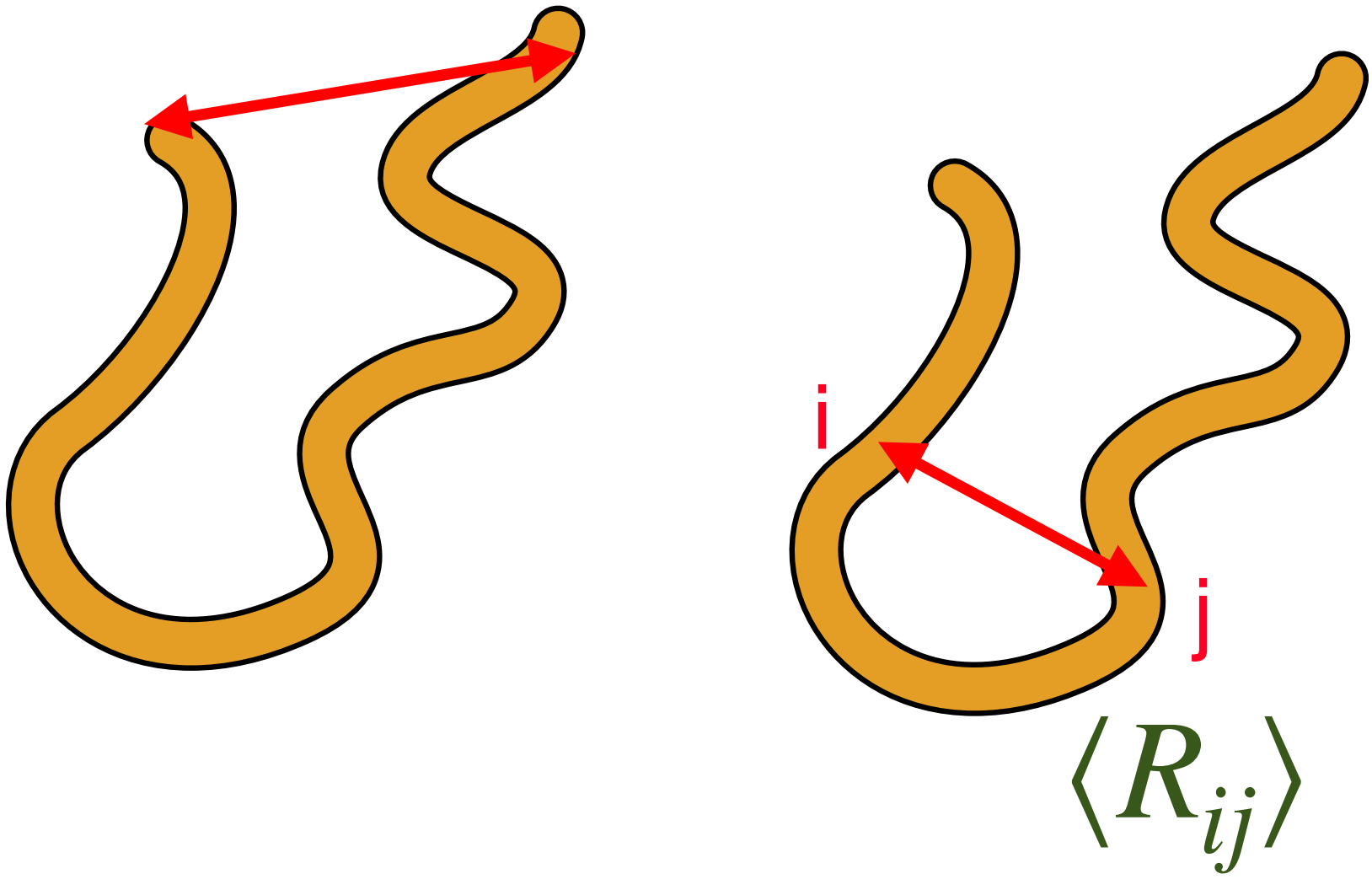
Experiment confirms the predicted trend



$$SCD_{lowsalt} = -27 \qquad SCD_{lowsalt} = -1.3$$

Ionic strength coupled to patterning modulates conformation

More patterning metrics arise describing intra-chain sizes



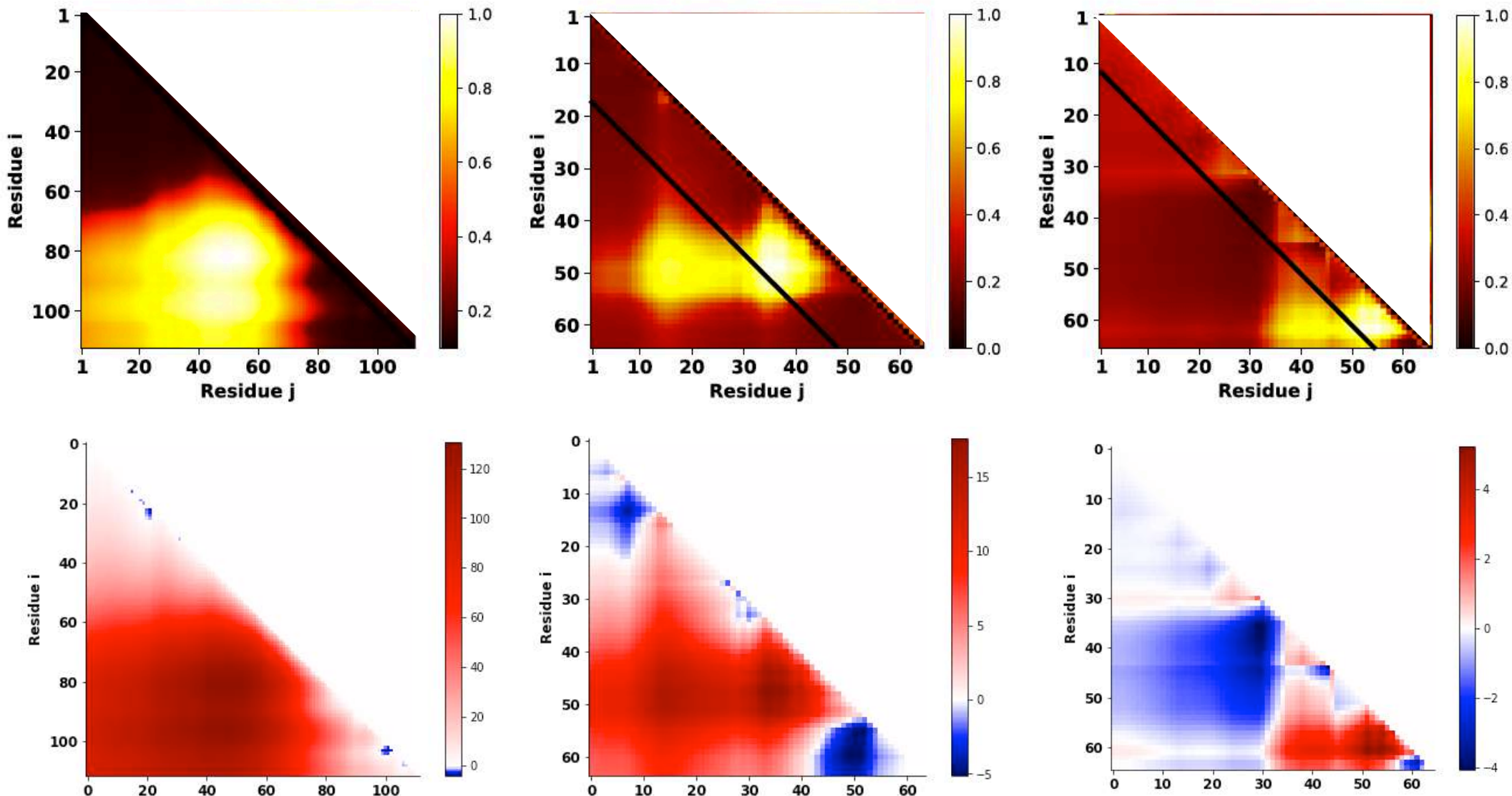
More patterning metrics arise describing intra-chain sizes

$$\beta F(R_{ij}) = \ldots\ldots\ldots + Q_{elec,ij} f(R_{ij})$$

$$SCDM_{ij} = \frac{1}{(i-j)} \left[\sum_{m=j}^{i} \sum_{n=1}^{j-1} q_m q_n \frac{(m-j)^2}{(m-n)^{3/2}} + \sum_{m=j+1}^{i} \sum_{n=j}^{m-1} q_m q_n (m-n)^{1/2} \right.$$
$$\left. + \sum_{m=i+1}^{N} \sum_{n=1}^{j-1} q_m q_n \frac{(i-j)^2}{(m-n)^{3/2}} + \sum_{m=i+1}^{N} \sum_{n=j}^{i} q_m q_n \frac{(i-n)^2}{(m-n)^{3/2}} \right]$$

IDPs have sequence specific distance profiles

SCDM maps reveal molecular blue print



Sequence Charge Decoration Matrix is IDP's molecular blueprint
Role in IDP function

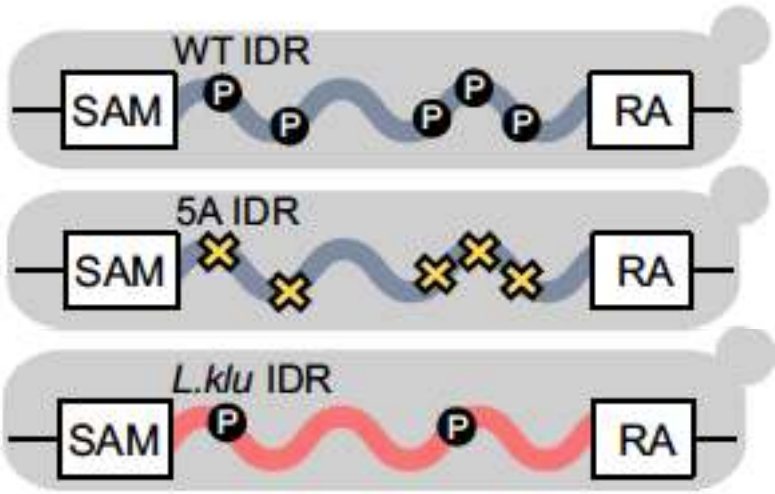
Challenges of modeling IDP function

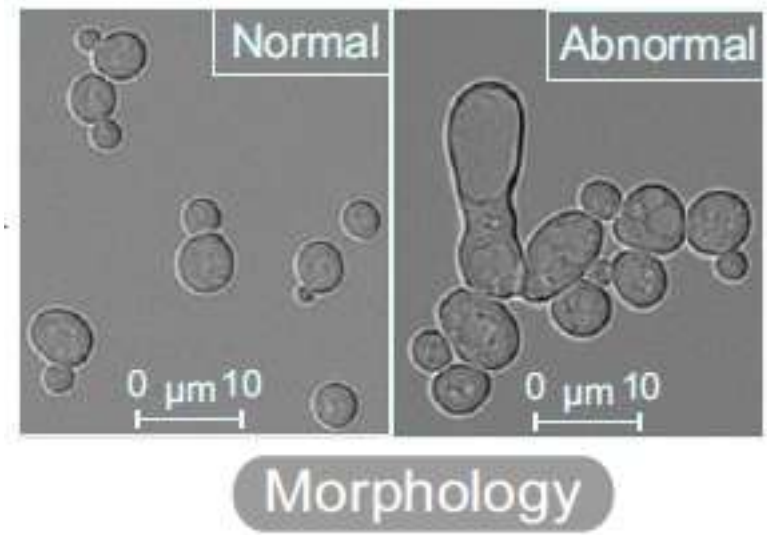
- Sequence alignment does not work
- Structure alignment does not work

Functionally similar IDPs lack sequence similarity

Zarin, Tsai, Ba and Moses PNAS 2017

Functionally similar IDPs lack sequence similarity

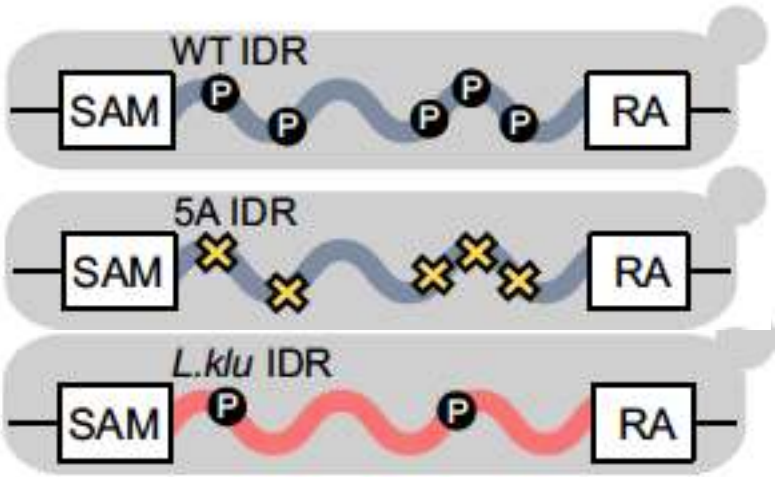




Ste50 wt Normal, Ste50 5A Abnormal, L klu normal

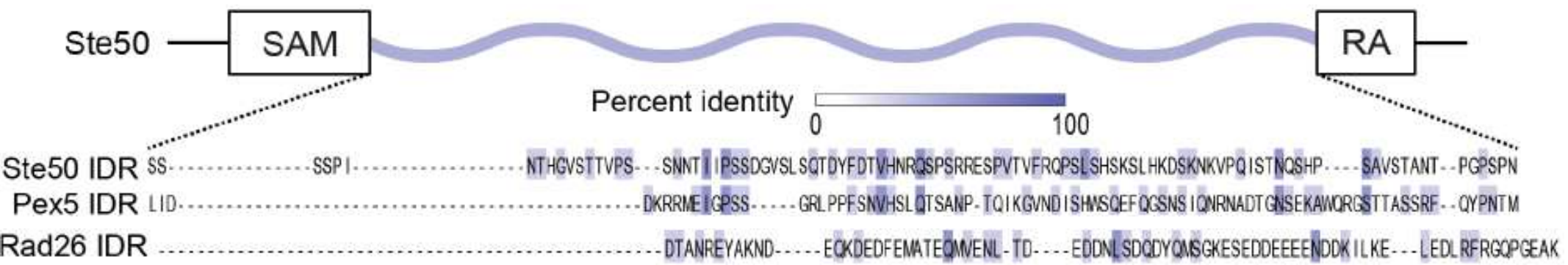
Zarin, Tsai, Ba and Moses PNAS 2017

Functionally similar IDPs lack sequence similarity



SCCharge
SC5A
LKCharge

PEX5
RAD26



Zarin, Strome, Ba, Alberti, Forman-Kay, Moses eLife 2019

Challenges of modeling IDP function

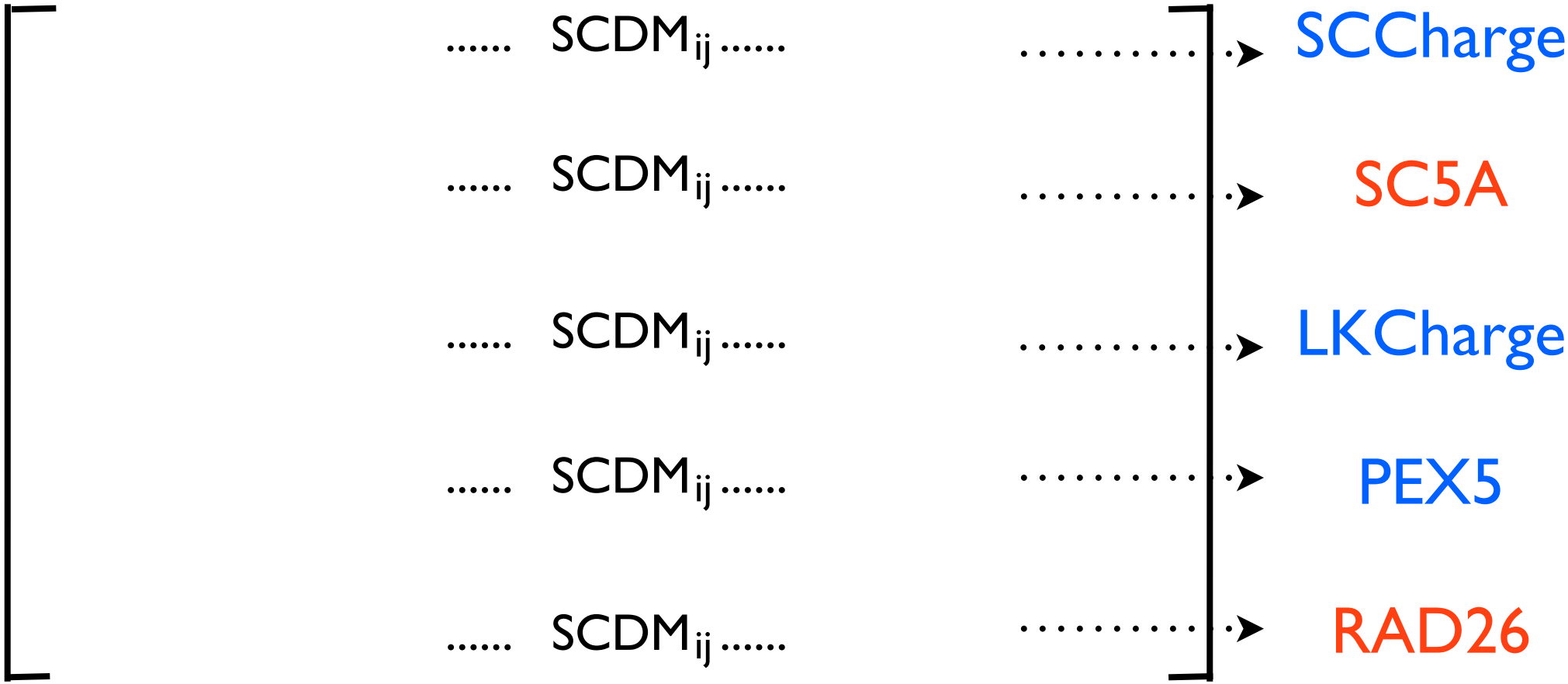
- Sequence alignment does not work
- Structure alignment does not work

What about using mathematical metric as IDP blueprint ?

Can we use SCDM to classify IDPS ?

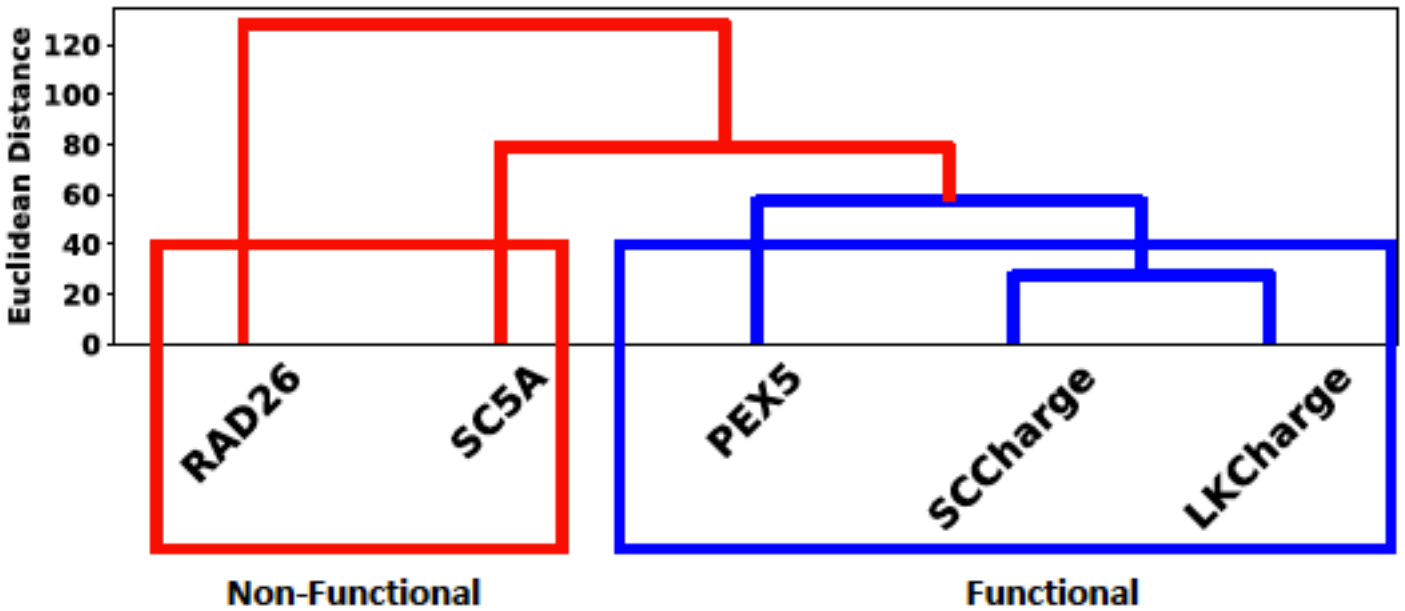
What about using mathematical metric as IDP blueprint ?

SCDM can detect functionally similar STE50



Huihui Ghosh (Biophysical Journal 2021)

SCDM can detect functionally similar STE50



SCCharge

SC5A

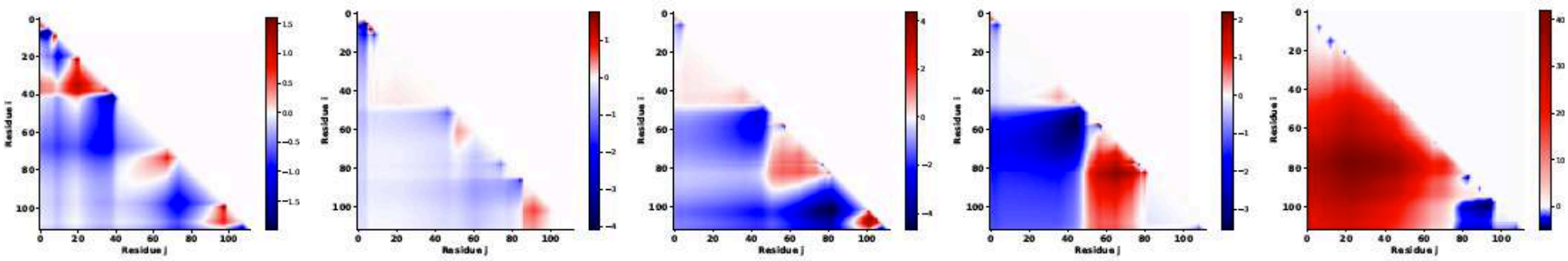
LKCharge

PEX5

RAD26

Huihui Ghosh Biophysical Journal 120, 1860 (2021)

SCDM can detect functionally similar STE50

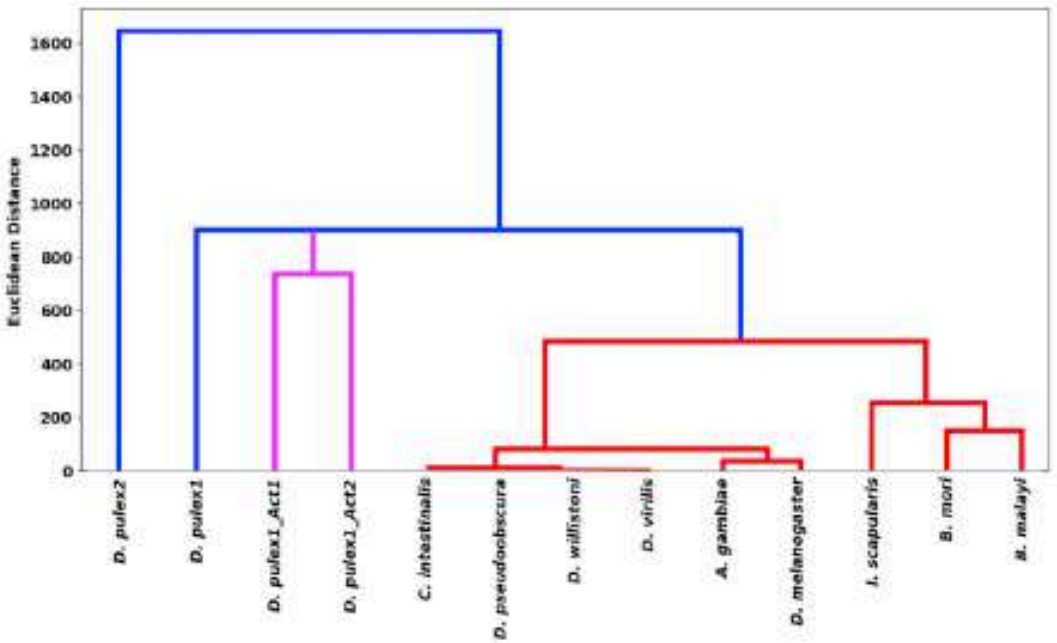


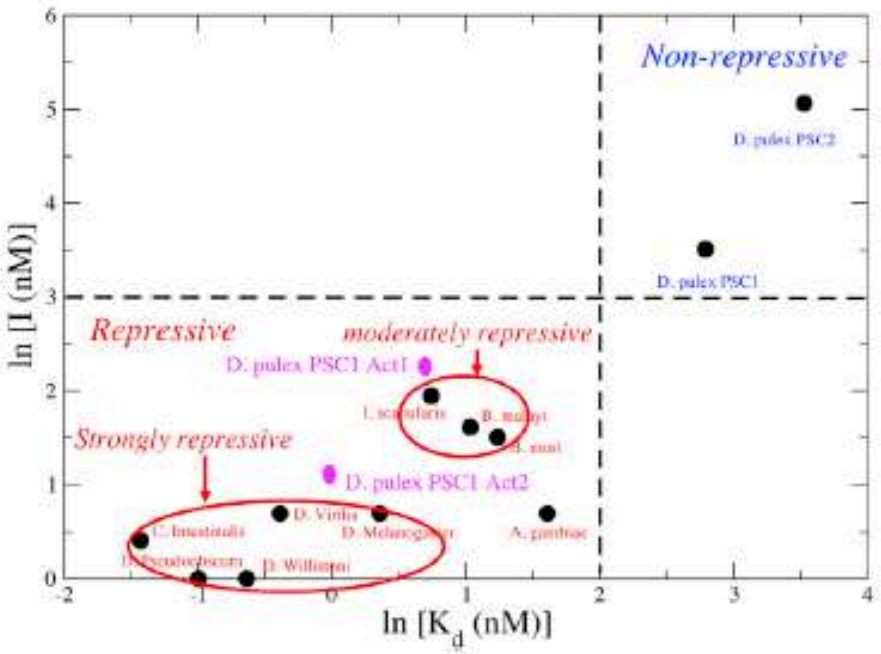
LKCharge PEX5 SCCharge SC5A RAD26

Functional non-functional

Huihui Ghosh (Biophysical Journal 2021)

SCDM can detect functionally similar PSC-CTR





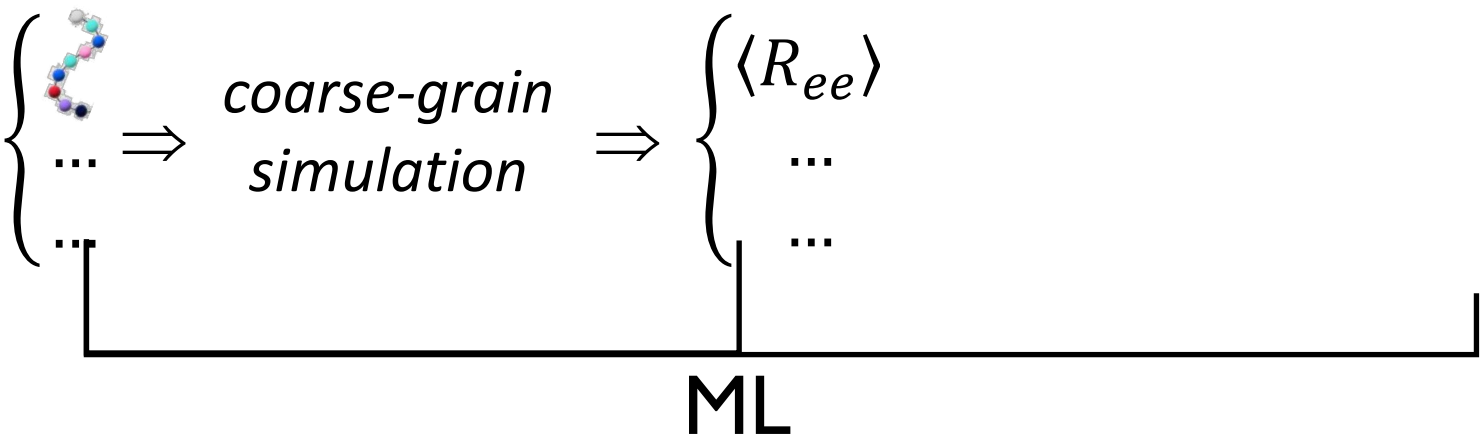
Huihui Ghosh (Biophysical Journal 2021)

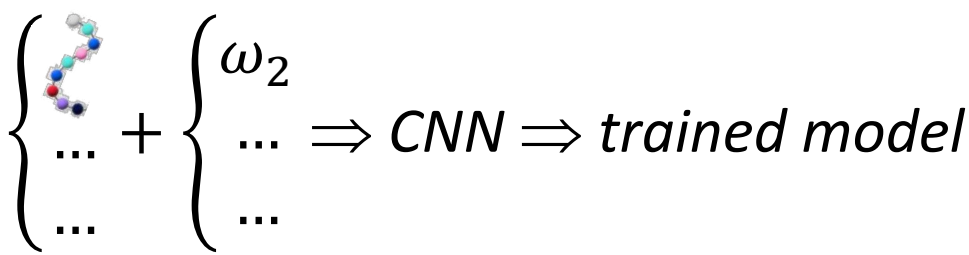
How to model sequence dependent non-electrostatics ?

We can use ML to train non-electrostatics H

$$\beta F(\langle R_{ee}\rangle) = -S(\langle R_{ee}\rangle) + \omega_2 f(\langle R_{ee}\rangle) + Q(c_s, [\,], \langle R_{ee}\rangle)$$

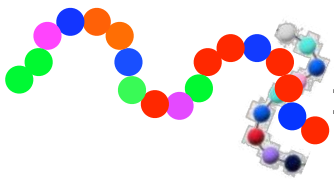
Zheng, Mittal, Lindorff-Larsen





$\Rightarrow$ *trained model* $\Rightarrow \omega_2 \Rightarrow$ ***H*** $\Rightarrow$ observables ($\langle R_{ij}\rangle$, ...)

Physics based Machine Learning (PML) trains H



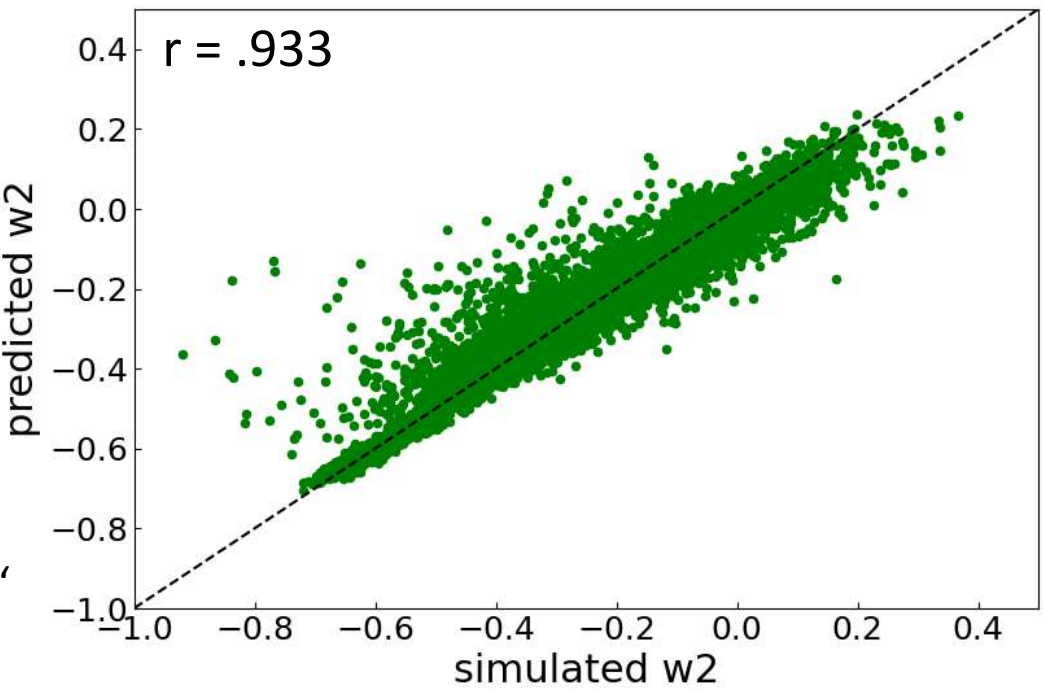
$\Rightarrow$ *trained model* $\Rightarrow \omega_2 \Rightarrow$ ***H***

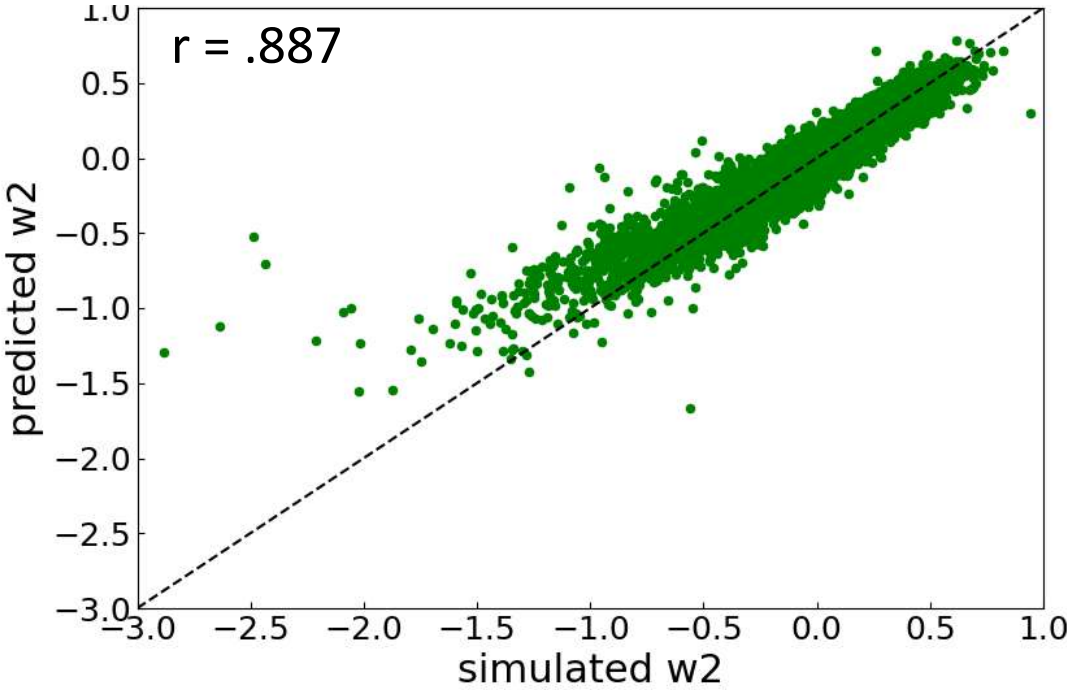
ML

$$\beta F(\langle R_{ee}\rangle) = -S(\langle R_{ee}\rangle) + \omega_2 f(\langle R_{ee}\rangle) + Q(c_s, [\,], \langle R_{ee}\rangle)$$

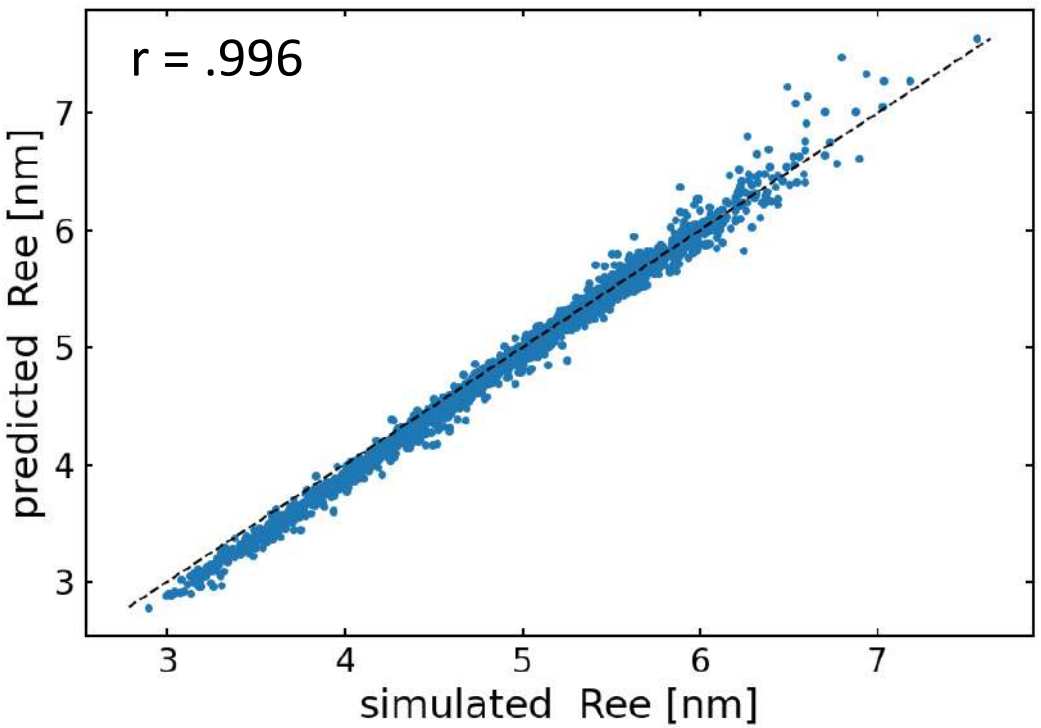
Physics

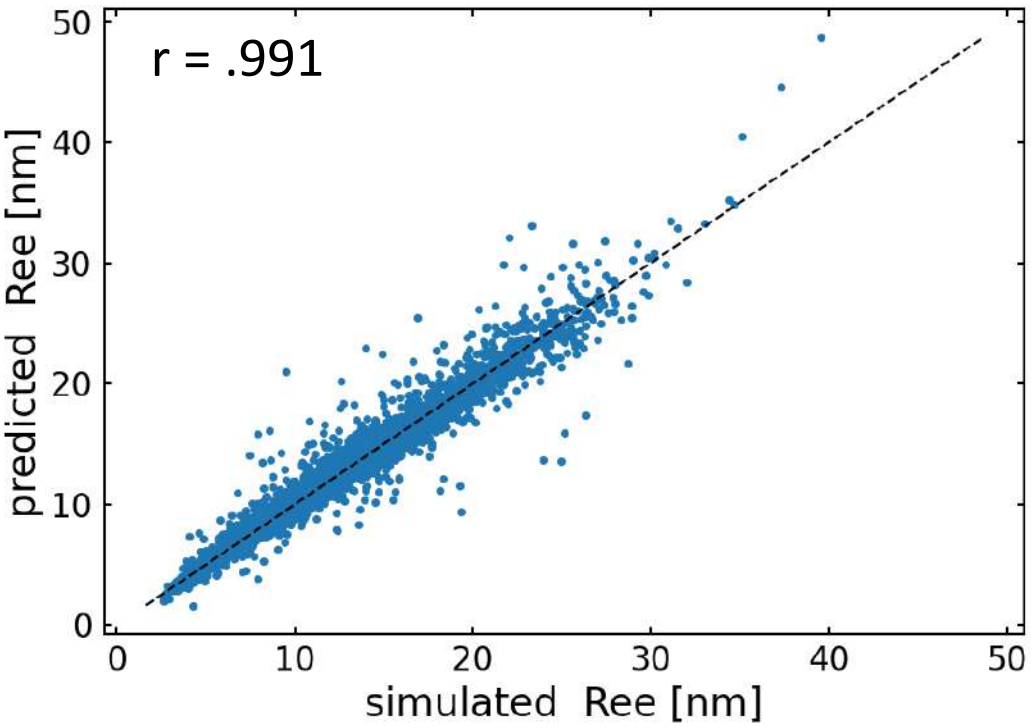
Simulated H can be mapped to an analytical H



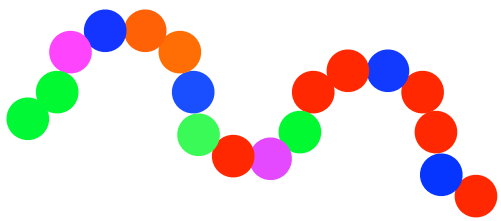


Physics based ML can predict sequence dependent size

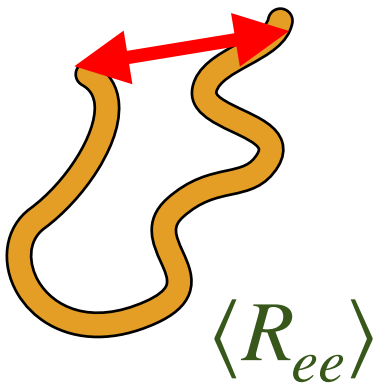




Physics based ML allows prediction of non-trained observables

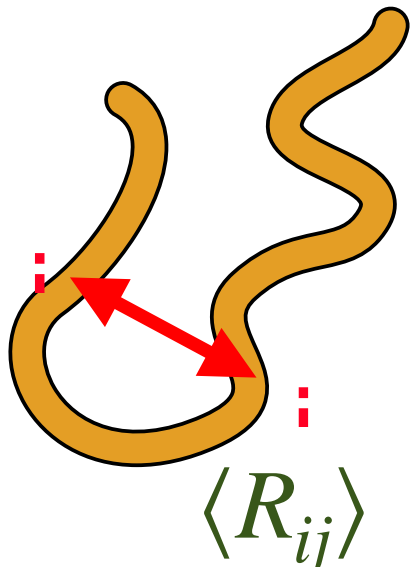


Trained model

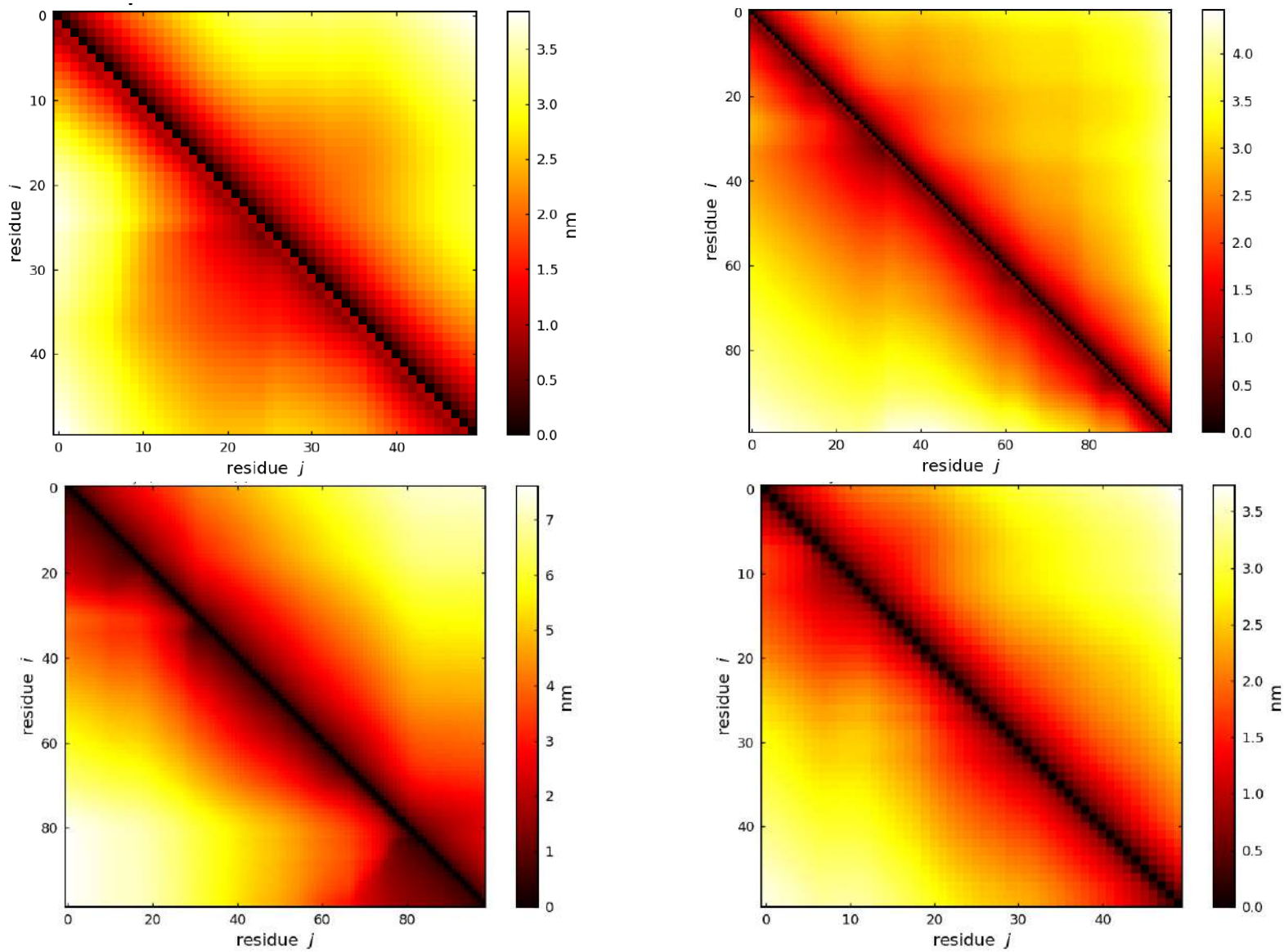


→ Hamiltonian →

non-trained observables

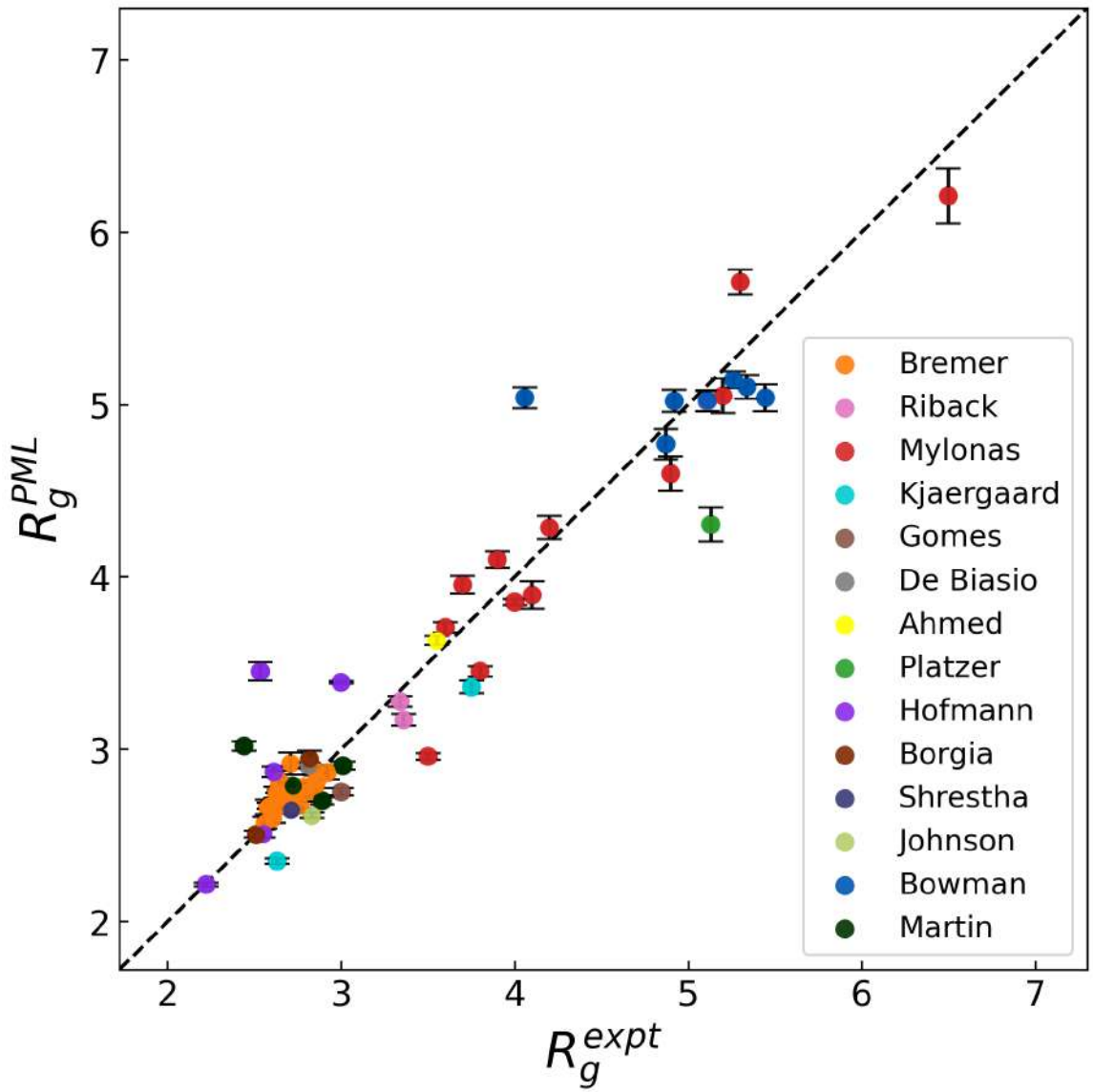


Physics based ML can predict distance map

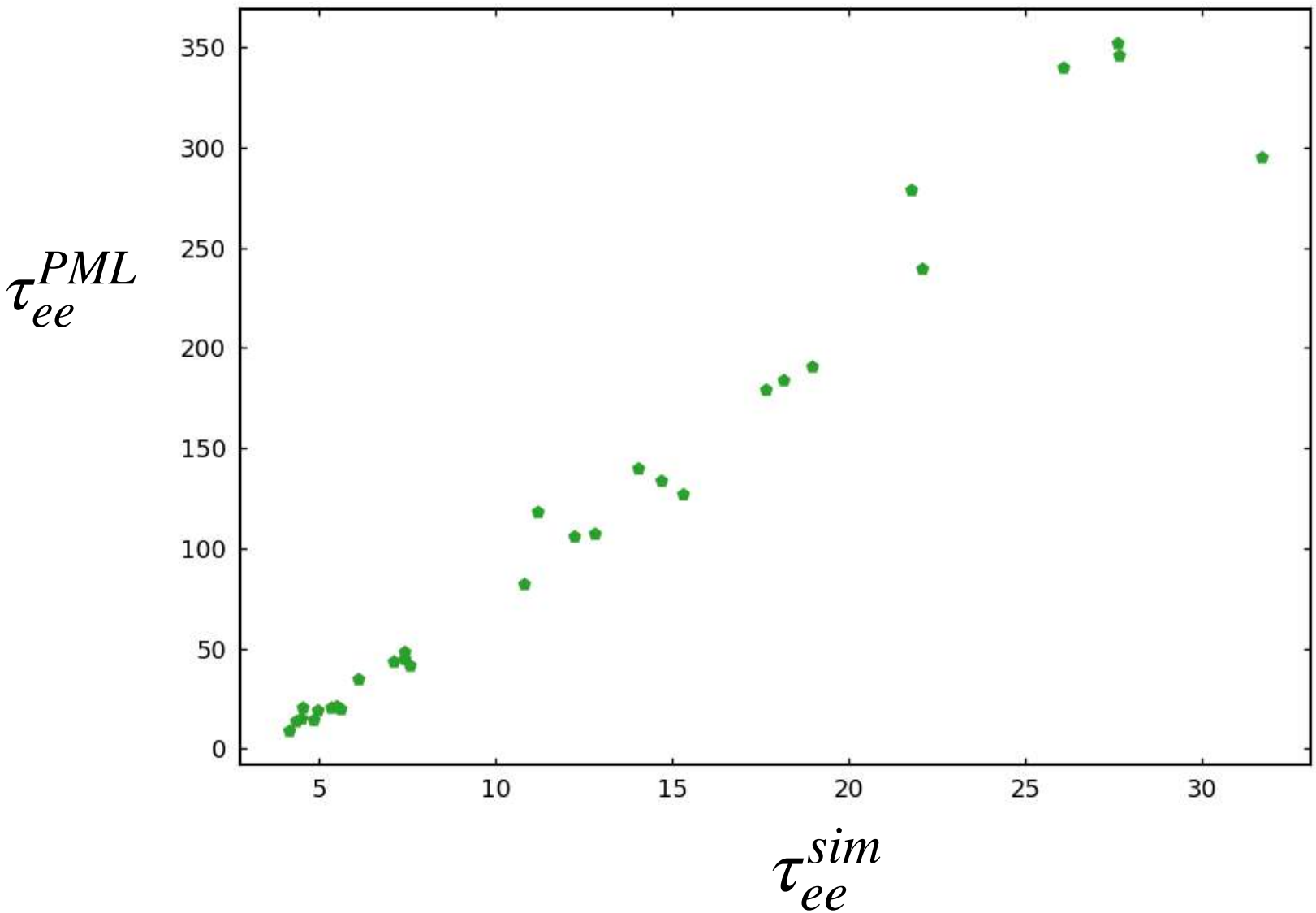


Houston et al JCTC (under revision)

Physics based ML also reproduces experimental data



Physics based ML can be used to predict simulated relaxation



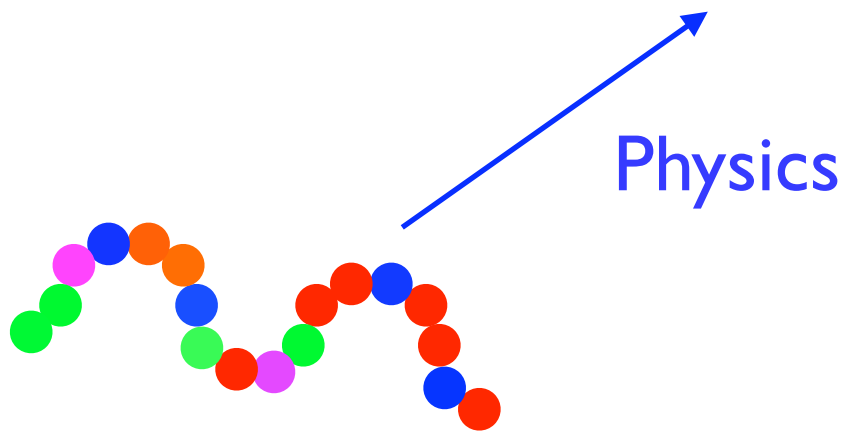
Multi-chain H correctly captures phase separation propensity

$$H = \frac{3}{2l^2}\sum_{\alpha=1}^{n_p}\sum_{\tau=1}^{N-1}\left(\mathbf{R}_{\alpha,\tau+1}-\mathbf{R}_{\alpha,\tau}\right)^2 + \frac{1}{2}\sum_{\alpha,\beta=1}^{n_p}\sum_{\tau,\mu=1}^{N}\left[\frac{\sigma_\tau\sigma_\mu e^{-\kappa|\mathbf{R}_{\alpha,\tau}-\mathbf{R}_{\beta,\mu}|}}{|\mathbf{R}_{\alpha,\tau}-\mathbf{R}_{\beta,\mu}|} + \upsilon_2\delta^3\left(\mathbf{R}_{\alpha,\tau}-\mathbf{R}_{\beta,\mu}\right)\right]$$

Entropy of mixing

Electrostatic correlation

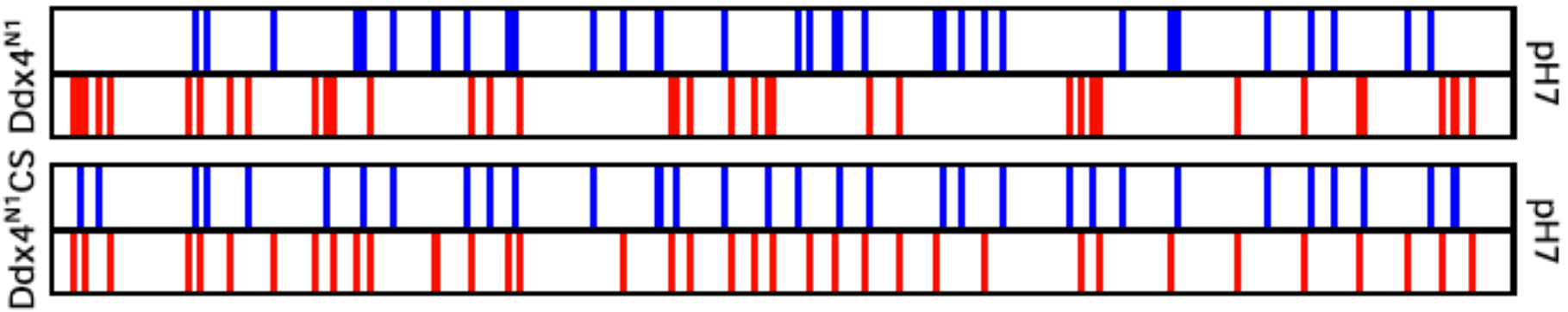
$$\beta f = \frac{\phi}{N}\ln\phi + \phi_s\ln\phi_s + \phi_c\ln\phi_c + \phi_w\ln\phi_w + \int\frac{dkk^2}{4\pi^2}\ln\left[1+\phi\left(\frac{\xi_k}{\nu_k}+g_k\right)+\frac{\phi^2}{\nu_k}\left(\xi_k g_k-\zeta_k^2\right)\right]$$

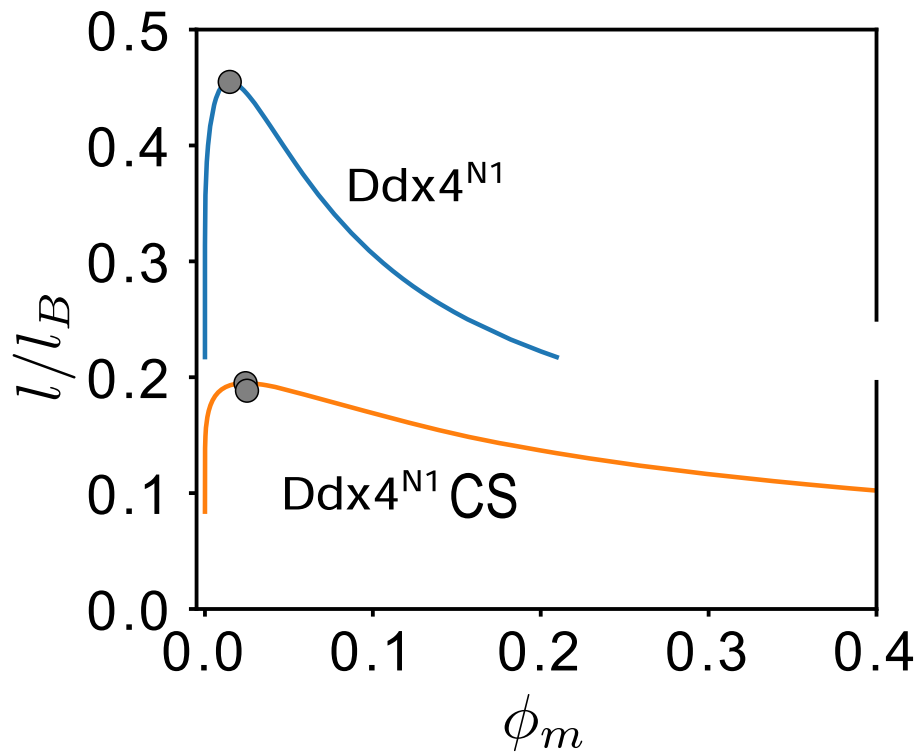


Sequence dependence

Lin, Brady, Chan, Ghosh JCP (2020)

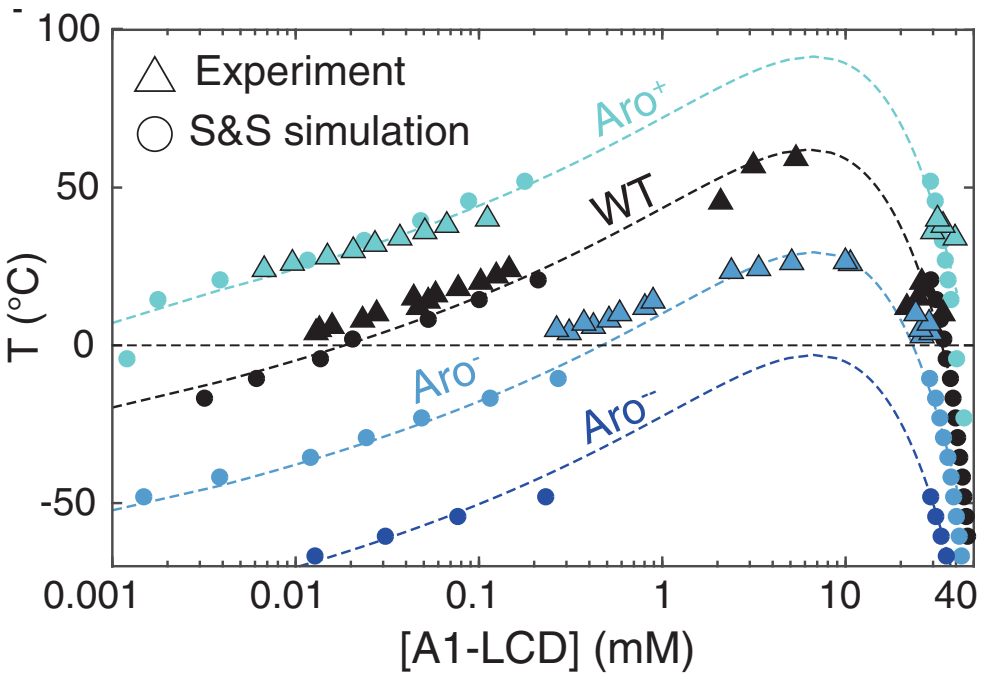
Multi-chain H correctly captures phase separation propensity

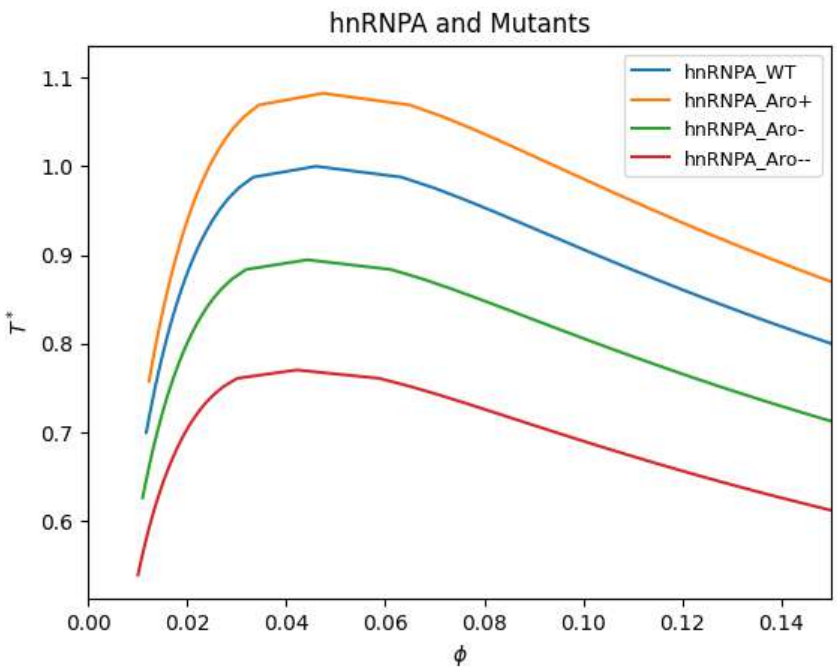




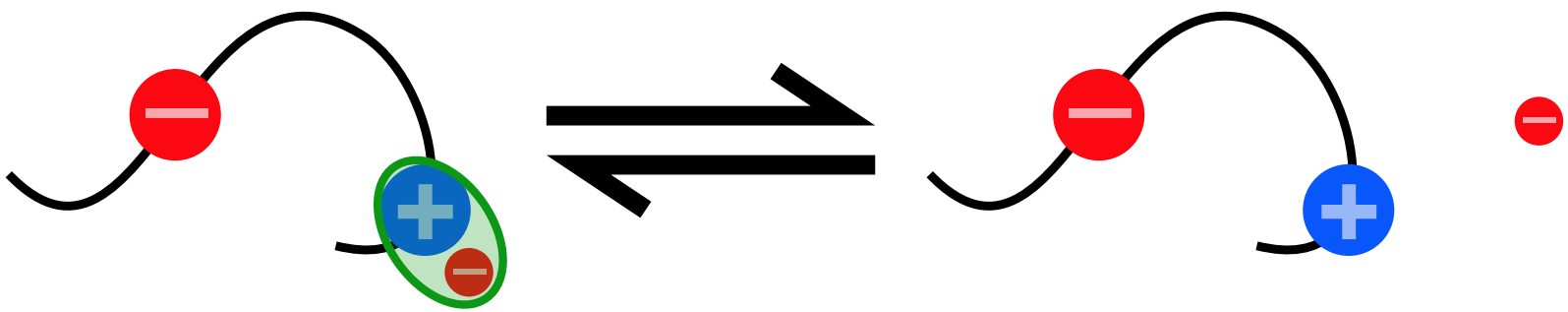
Lin, Brady, Chan, Ghosh JCP (2020)

PML can predict LLPS propensity for non-charge mutations

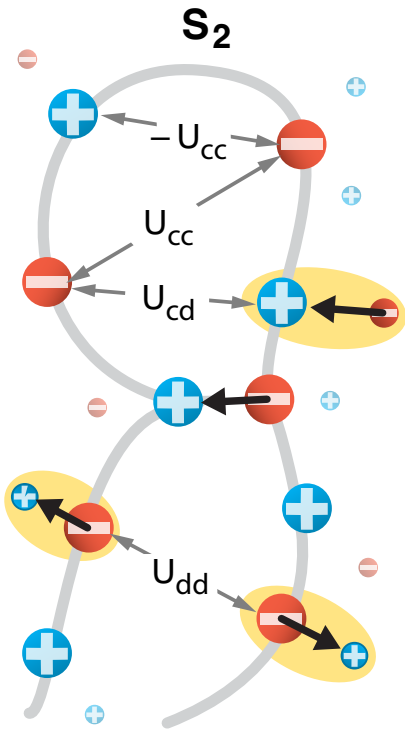




Charges may not be fully ionized



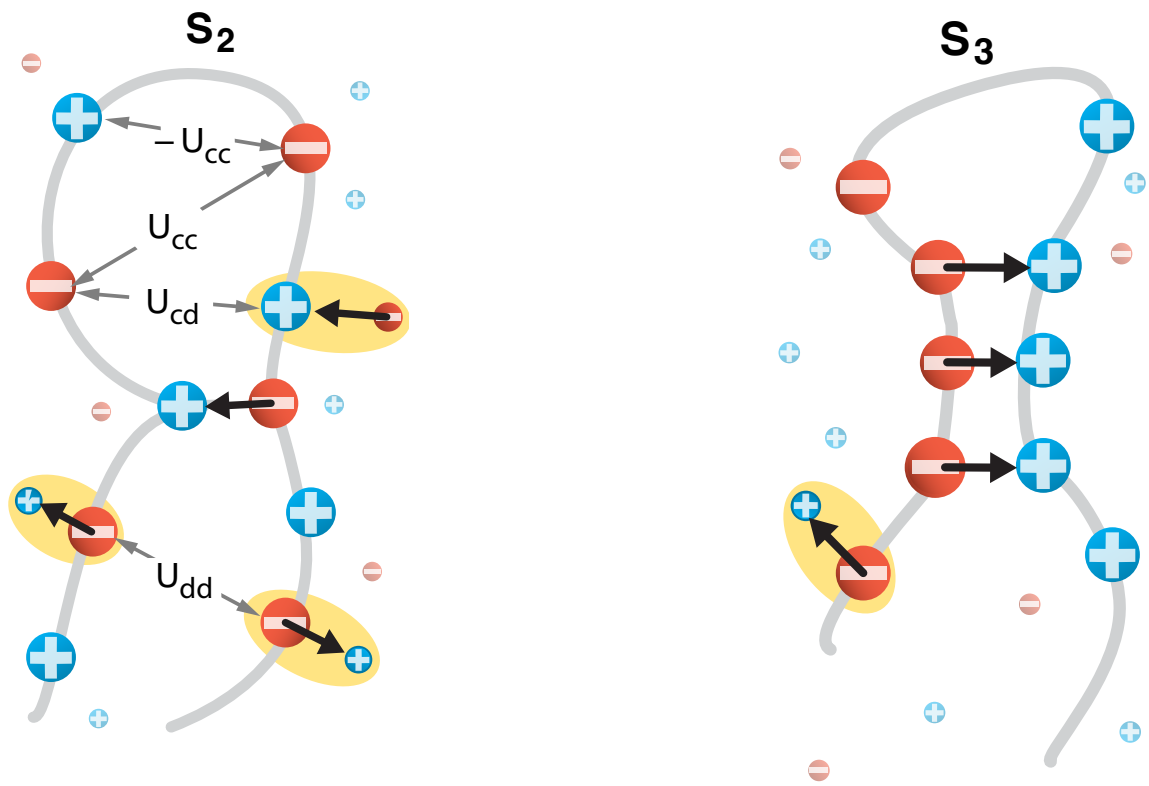
Beyond monopole electrostatics is needed



dipole formation

Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

charge modulation, sequence and chain conformation are coupled



Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

Beyond monopole electrostatics is needed

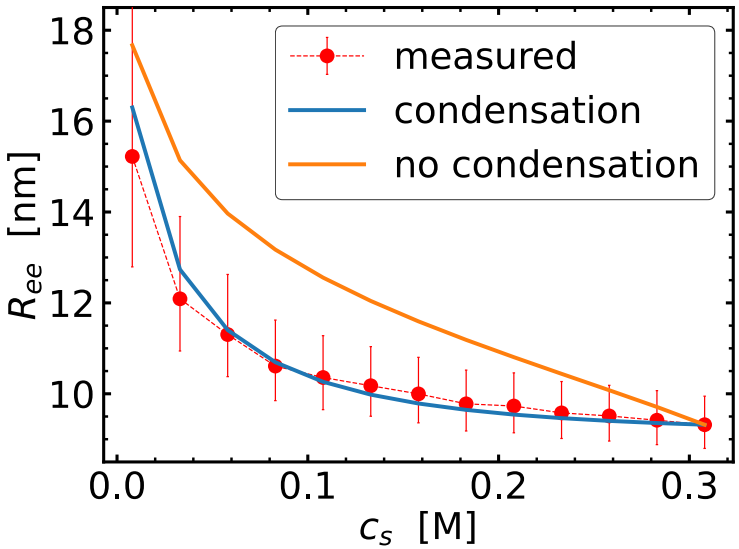
- Combinatorial entropy of ionization $F_1(\alpha_+, \alpha_-)$
- Counterion translational entropy $F_2(\alpha_+, \alpha_-)$
- Ion-pair formation equilibrium $\frac{\beta F_4}{N} = -[f_+(1-\alpha_+) + f_-(1-\alpha_-)]\frac{\tilde{\ell}_B}{\tilde{p}}\left(\delta + \frac{1}{2}\right)$
- Chain free energy

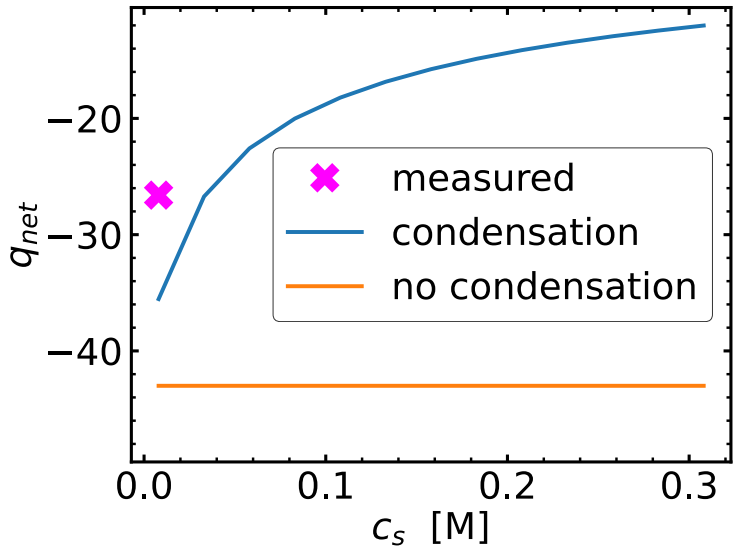
$$\beta F_5 = \frac{3}{2}(x - \ln(x)) + \frac{\omega_3 B}{2}\left(\frac{3}{2\pi x}\right)^3 + 2\tilde{\ell}_B Q\left(\frac{3}{2\pi x}\right)^{1/2} + \Omega\left(\frac{3}{2\pi x}\right)^{3/2}$$

$$\Omega_{c-d} = \omega_{cd}\frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1}(c_m d_n + c_n d_m)(m-n)^{-1/2}$$

$$\Omega_{d-d} = \omega_{dd}\frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1} d_m d_n (m-n)^{-1/2},$$

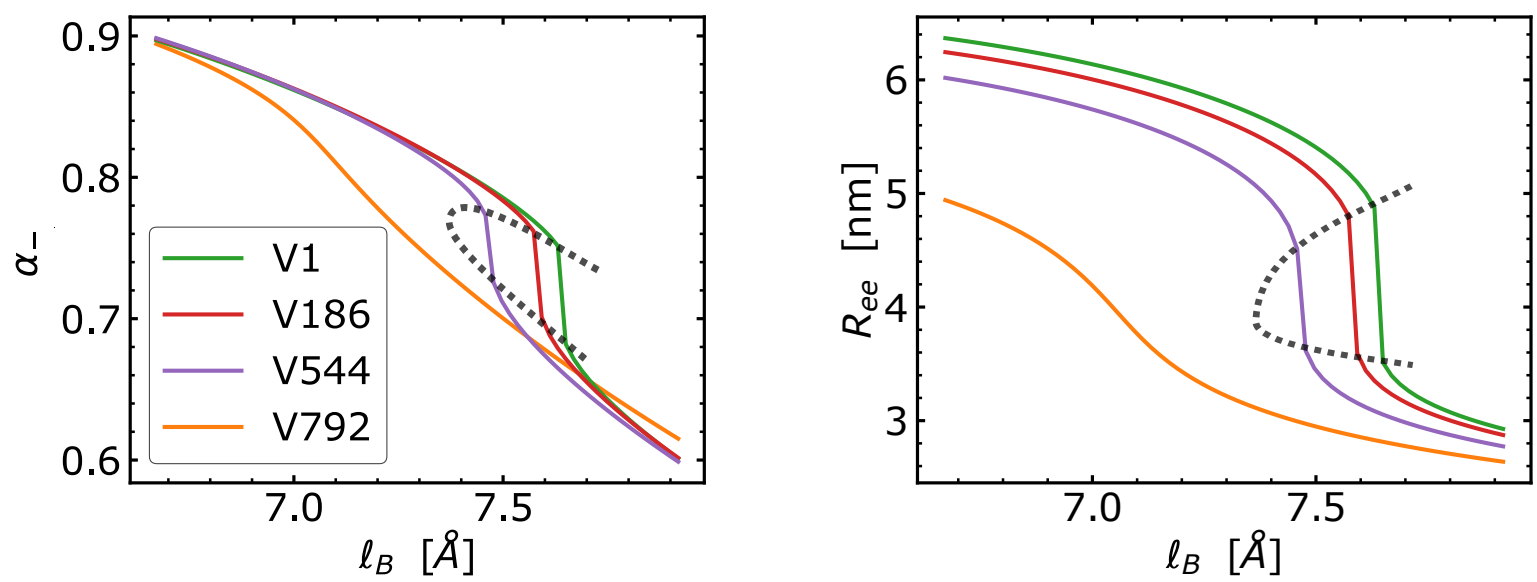
Theory describes conformation and charge state





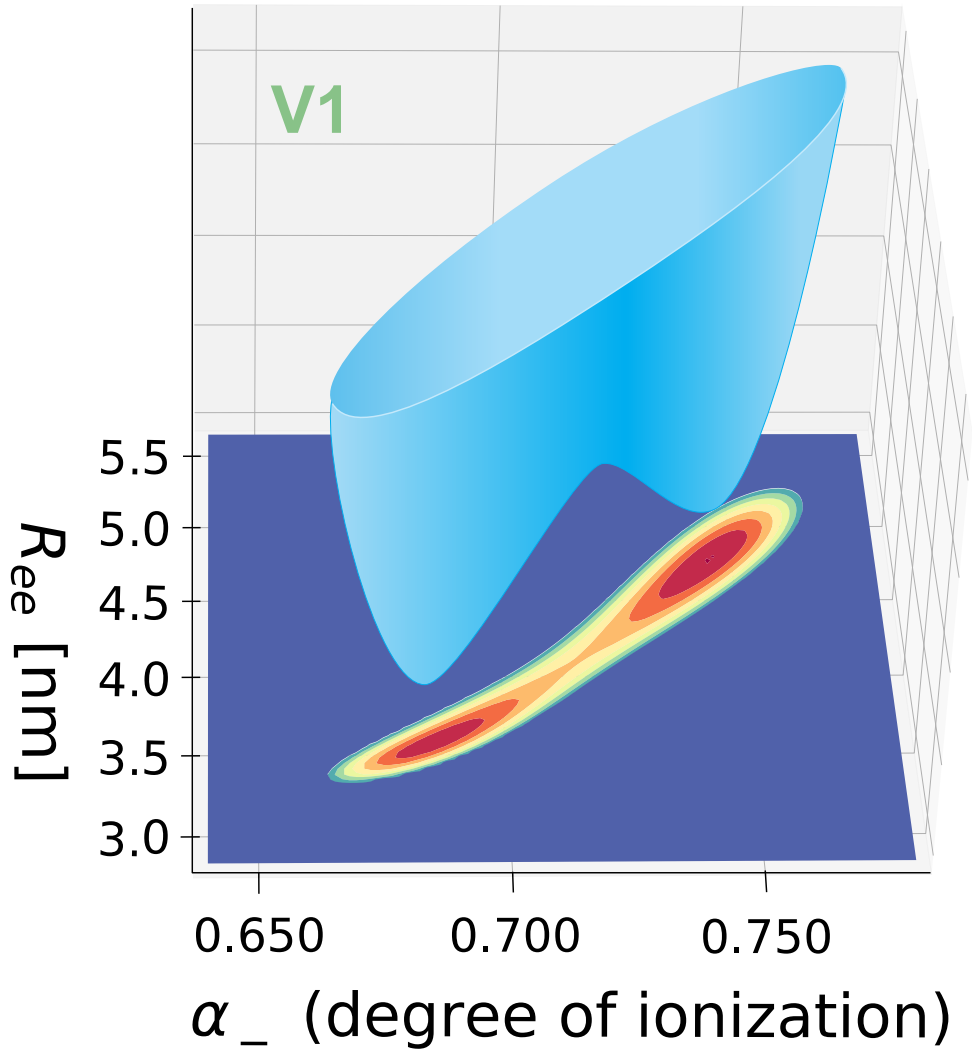
Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

Sequence patterning regulates conformation and charge state



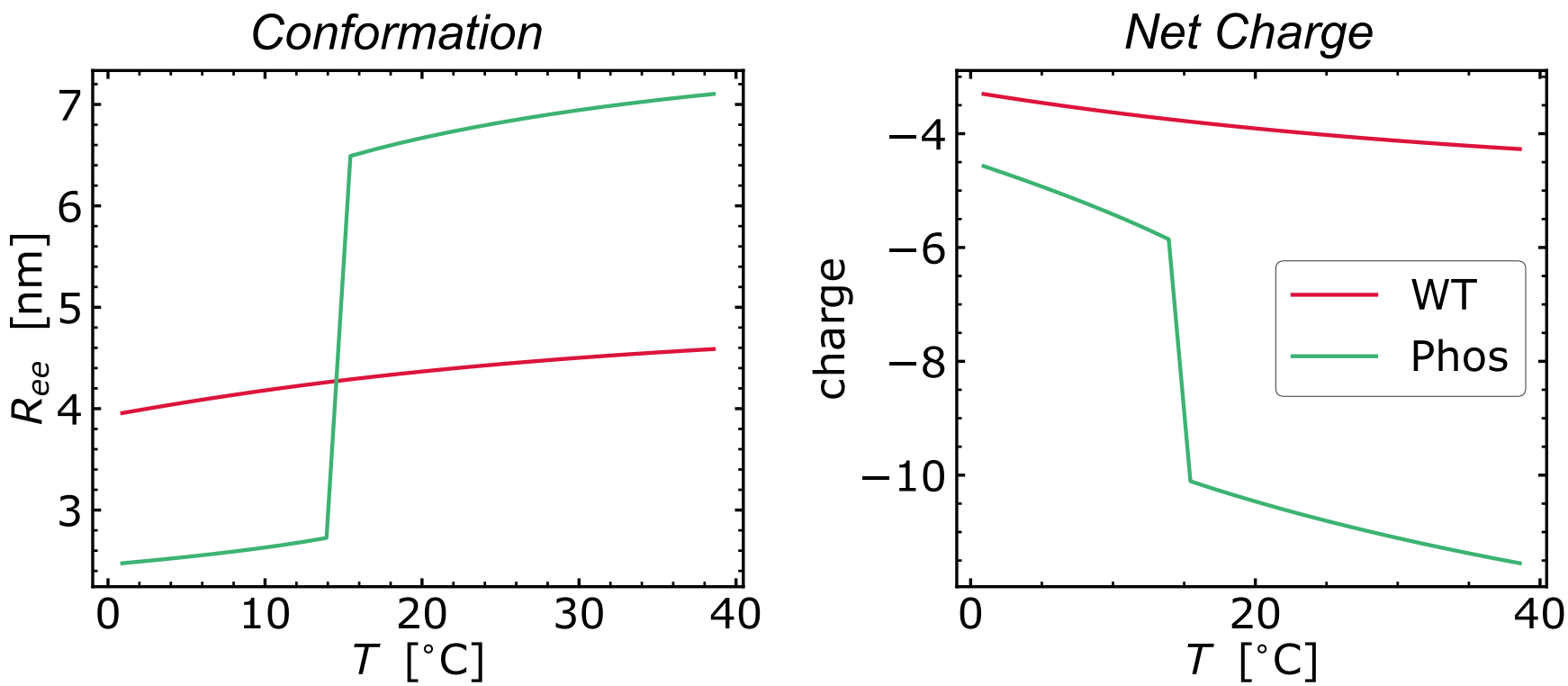
Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

Charge and conformation landscape may exhibit bistability



Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

Phosphorylation can harness charge and conformational fluctuation



Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

Conclusion

- Hamiltonian based analytical theory can be useful to dissect different regulators of chain conformation

- Sequence based theory can help understand function

Acknowledgment

Jonathan HuiHui

Michael Phillips

Austin Haider

Andrew Torres

Liliana Houston

Kari Gaalswyk

Taylor Firman , Luke Sawle

Martin Gruebele, Hue Sun Chan, Dan Fisher, Muthukumar

RESEARCH CORPORATION for SCIENCE ADVANCEMENT
A foundation dedicated to science since 1912.

National Science Foundation
WHERE DISCOVERIES BEGIN

NIH National Institute of General Medical Sciences