Sequence dependent regulation of IDP conformation and function

Kingshuk Ghosh

University of Denver Department of Physics and Astronomy

IDPs have different regulators

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

Sequence matters in IDP function -8.0 **A** IDP function

Ddx4N phase separates but Ddx4CS does not Ddx4N phase separates but Ddx4CS does not \overline{a}

Nott et al Mol Cell (2015) entropy changes of the transition, which report on the microscopic interactions between molecules. $\frac{1}{2}$ The interaction parameters varied in a predictable way with interaction $\frac{1}{2}$ was found to decrease as Linking sequence to conformation and function

Hamiltonian (coarse grain) based theory

$H =$ *I* 1 **I** 2 *I* 3 *I* 4 $+$ ()() + () () + *R*¹³ R_{12} R_{13} R_{23} R_{12} R_{23} R_{23} $\frac{e^2}{R_{12}} - \frac{e^2}{R_{13}} - \frac{e^2}{R_{23}}$ — — chain connectivity electrostatics two body Three body repulsive H based polymer theory can be useful

Theory can compute ensemble average end-to-end distance

H based polymer theory can describe charge correlation

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

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

sequence specificity

Theory captures all-atom simulation of toy sequences

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

o[.]
Ia 15
Dic **Sawle and Ghosh (JCP 2015)**

 $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ and $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ and $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ Ghosh, Huihui, Phillips, Haider Annual Reviews of Biophysics (2022)

Modeling the Disordered Proteins (IDPs)

Experiment confirms predicted mutational hot spots

HYPK-WT (22 positive and 31 negative charges)

HYPK-4K v1

 $R_{ee, v1}$ < R $_{ee, v2}$ Prediction

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

> 0 shrink < 0 expand

Experiment confirms the predicted trend

$$
SCD_{lowsalt} = -27
$$

$$
SCD_{lowsalt} = -1.3
$$

Ionic strength coupled to patterning modulates conformation

More patterning metrics arise describing intra-chain sizes

More patterning metrics arise describing intra-chain sizes

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

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

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

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 ?

Huihui Ghosh (Biophysical Journal 2021)

SCCharge SCDM can detect functionally similar STE50

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 ?

Physics based Machine Learning (PML) trains H Physics hasad N … la chine …
… *Corping (DMI)* trains H **Physi**

Simulated H can be mapped to an analytical H Simulated *H* can be mapped to analytical *H*.

 2.5

Physics based ML allows prediction of nontrained observables

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 terms in *U^p* are either regularized subsequently or inconsequen-**Multi-chain H correctly captures phase se** *l* ³ ln *^Z*′ *p* Ω \mathbf{u} understood as an effective interaction interaction interaction in \mathbf{u} ctly captures pnase separation propens MED1, which drives phase separation of the transcriptional apparatus. Sequence dependent results of provides the phase separation in the phase separation of phase separatio

$$
\mathbf{H} = \frac{3}{2l^2} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2 + \left(\frac{1}{2} \sum_{\alpha,\beta=1}^{n_p} \sum_{\tau,\mu=1}^{N} \left[\frac{\sigma_{\tau} \sigma_{\mu} e^{-\kappa |\mathbf{R}_{\alpha,\tau} - \mathbf{R}_{\beta,\mu}|}}{|\mathbf{R}_{\alpha,\tau} - \mathbf{R}_{\beta,\mu}|} + v_2 \delta^3 (\mathbf{R}_{\alpha,\tau} - \mathbf{R}_{\beta,\mu}) \right]
$$

 $N \stackrel{\text{in } \gamma}{\longrightarrow}$ $\frac{1}{\sqrt{2}}$ *N* ϕ_s ln $\phi_s + \phi_c$ ln $\phi_c + \phi_w$ ln $\phi_w + \int \frac{d\theta}{4\pi}$ a fixed a fixed of the second term in the second term in the second term in the second term in the second term i $\left[\begin{array}{cc} 0 & \epsilon, \end{array} \right]$ vanishes. As intervalsed above, here we use a renormalized above, here we use a renormalized above, Δ^2 $\left|1+\phi\left(\frac{s\kappa}{\mu}+g_k\right)+\frac{r}{\mu}\left(\xi_k g_k-\zeta_k^2\right)\right|$ $\begin{bmatrix} 1 & \sqrt{\nu_k} & \$ $\int d^2k^2 \int (\xi, \xi) d^2k^2$ $t_1 + \phi_w \ln \phi_w + \int \frac{\omega \kappa}{4a^2} \ln \left[1 + \phi \left(\frac{S\kappa}{2} + g_k \right) + \frac{\phi}{2} \left(\xi_k g_k - \zeta_k^2 \right) \right]$ $\int 4\pi^2$ $\left\lfloor \sqrt{\nu_k} \right\rfloor / \nu_k$ Entropy of mixing **Electrostatic correlation** $\beta f = \frac{\phi}{\lambda}$ *N* $\ln \phi + \phi_s \ln \phi_s + \phi_c \ln \phi_c + \phi_w \ln \phi_w +$ $\int \frac{dkk^2}{4\pi^2} \ln \bigg[$ $1+\phi$ $\int \xi_k$ ν_k $+ g_k$ $+ \frac{\phi^2}{4}$ ν_k $(\xi_k g_k - \zeta_k^2)$ $\binom{2}{1}$ $\overline{}$

 $\mathsf{Lin},\, \mathsf{Brady},\, \mathsf{Chan},\, \mathsf{Ghosh}\, \mathsf{JCP}\,(2020)$

Lin, Brady, Chan, Ghosh JCP (2020)

PML can predict LLPS propensity for non-charge mutations Flory-Huggins fit to exp. $\overline{}$ \mathcal{L}

Charges may not be fully ionized

Beyond monopole electrostatics is needed

 $\mathbf{F}^{(1)}$ condensation, resulting in $\mathbf{F}^{(2)}$ and conformation in IDPs. Sequence 1 (S1) has $\mathbf{F}^{(3)}$ Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

charge modulation, sequence and chain conformation are coupled

Fig. 1: September 2: September 2: S
1: September 1: September 1: September 1: September 2: September 2: September 2: September 2: September 2: S formed by condensation, resulting in charge and conformation in IDPs. Sequence $\mathcal{L}(\mathcal{L}|\mathcal{L})$ $\mathbf{F}^{(1)}$ condensation, resulting in $\mathbf{F}^{(2)}$ and conformation in IDPs. Sequence 1 (S1) has $\mathbf{F}^{(3)}$ Phillips, Muthukumar, Ghosh PNAS Nexus (2024)

Beyond monopole electrostatics is needed \mathbf{r} peyonic and salt in the state in production in the single with data, we use a single with data, we use a single single set of each of the independent of

- Combinatorial entropy of ionization **<u>A</u>**
Frank of ionization \bullet Complete \mathcal{L} is a centration in pico molar (in million in milli molar), with salt concentration (in milli molar), with \mathcal{L}
- Counterion translational entropy $F_2(\alpha_+, \alpha_-)$
	- Ion-pair formation equilibrium • Ion-pair formation equili
	- **•** Chain free energy coexist in two states in the charge-conformation landscape giving

ergy (related to equilibrium constant) of ion pair formation arising Material, eqs. S1-S6, for details of these three terms). *F*4 is the energy (related to equilibrium constant) of ion pair formation arising where, *f*⁺ = *N*+*/N* and *f*[−] = *N*−*/N* are sequence charge fractions, *p*˜ ≡ *p/b* is the (nondimensional) distance between a pair of ions, *^B* ≡ *`B/b*, *`B* is Bjerrum length (*e*²*/*4πϵ0ϵ*kBT*), and δ = ϵ*/*ϵ*l* is the dielectric mismatch between water's dielectric constant, ϵ = 80, and that of the local chain environment, ϵ*l*. Typical values of dielectric mismatch for IDPs are δ ∈ [1*.*3, 2*.*7], assuming ϵ*^l* ∈ β*F*⁵ = 3 ² (*^x* [−] ln (*x*)) ⁺ ω3*B* 2 3 2π*x* ✓ ◆³ + 2*`*˜ *BQ* ³ 2π*x* ✓ ◆¹*/*² ⁺^Ω ³ 2π*x* ✓ ◆³*/*² dependence of ω2 that would arise from temperature dependent solvation effects. Two additional contributions to Ω are, Ω^c−^d and Ω^d−d, resulting from directionally averaged charge–dipole and dipole–dipole interactions approximated as delta function po-**4 |** *PNAS Nexus*, 2024, Vol. 3, No. 9

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

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

if all entropy of ionization

\n
$$
F_1(\alpha_+, \alpha_-)
$$

$$
t \n topy\n $F_2(a_+, a_-)$
$$

on equilibrium
$$
\frac{\beta F_4}{N} = -[f_+(1-\alpha_+) + f_-(1-\alpha_-)] \frac{\tilde{\ell}_B}{\tilde{p}} \left(\delta + \frac{1}{2}\right)
$$

$$
f_{\angle}(\alpha_{+}, \alpha_{-})
$$
\nrelation equilibrium

\n
$$
\beta F_4 \quad \text{for } (\alpha_{-}, \beta_{-}) \leq (\alpha_{-}, \beta_{-}) \leq (\beta_{-}, \beta_{-})
$$

Theory describes conformation and charge state

Phillips, Muthukumar, Ghosh PNAS Nexus (2024) *^p*˜ = 0*.*55*,* = 1*.*3*,* !² = 1*.*275. Other model parameters are: ⇢ = 1 ⇥ ¹⁰⁶ mol*/*L (˜⇢ ⇡ ³*.*³ ⇥ ¹⁰⁸), `*^B* = 7*.*¹² ˚A

Sequence patterning regulates charge patterning from a large pool of sequences having same sequence patterning regulates conformation and charge state a subset of photophoral conditions between σ

of coexistence and School and School, included by the SCD, included by the SCD, included by the SCD, included b

Charge and conformation landscape may exhibit bistability

 \circ

 \circ

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

National Science Foundation WHERE DISCOVERIES BEGIN

