Biophysical *Journal*
Biophysical Letter

Biophysical Society

Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated

Yi-Hsuan Lin[1,3] and Hue Sun Chan[1,2,*]
[1]Department of Biochemistry and [2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; and [3]Molecular Medicine, Hospital for Sick Children, Toronto, Ontario, Canada

ABSTRACT Liquid-liquid phase separation of intrinsically disordered proteins (IDPs) is a major undergirding factor in the regulated formation of membraneless organelles in the cell. The phase behavior of an IDP is sensitive to its amino acid sequence. Here we apply a recent random-phase-approximation polymer theory to investigate how the tendency for multiple chains of a protein to phase-separate, as characterized by the critical temperature T^*_{cr}, is related to the protein's single-chain average radius of gyration $\langle R_g \rangle$. For a set of sequences containing different permutations of an equal number of positively and negatively charged residues, we found a striking correlation $T^*_{cr} \sim \langle R_g \rangle^{-\gamma}$ with γ as large as ~6.0, indicating that electrostatic effects have similarly significant impact on promoting single-chain conformational compactness and phase separation. Moreover, $T^*_{cr} \propto -$SCD, where SCD is a recently proposed "sequence charge decoration" parameter determined solely by sequence information. Ramifications of our findings for deciphering the sequence dependence of IDP phase separation are discussed.

The biological functions and disease-causing malfunctions of proteins are underpinned by their structures, dynamics, and myriad intra- and intermolecular interactions. Many critical cellular functions are carried out by intrinsically disordered proteins or protein regions (collectively abbreviated as "IDPs" here) with sequences that are less hydrophobic than those of globular proteins, but are enriched in charged, polar, and aromatic residues (1–6). At least 75% of IDPs are polyampholytes (7,8) in that they contain both positively and negatively charged residues (9,10). Accordingly, electrostatic effects are important in determining individual IDPs' conformational dimensions (8,11,12) and binding (13,14). Charge-charge interactions are often significant in the recently discovered phenomenon of functional IDP liquid-liquid phase separation as well (15–22). IDP phase separation appears to be the physical basis of membraneless organelles, performing many vital tasks. Recent examples include subcompartmentalization within the nucleolus (22) and synaptic plasticity (21). Malfunction of phase separation processes can lead to disease-causing amyloidogenesis (18) and neurological disorders (21). Speculatively, membraneless liquid-liquid phase separation of biomolecules might even have played a role in the origins of life (23).

Electrostatic effects encoded by a sequence of charges depend not only on the total positive and negative charges or net charge (24,25) but also the charge pattern (8). For IDPs, this was demonstrated by Das and Pappu (8), who conducted explicit-chain, implicit-solvent conformational sampling of 30 different sequences, each composed of 25 lysine (K) and 25 glutamic acid (E) residues (termed "KE sequences" hereafter). They found that the average radius of gyration, $\langle R_g \rangle$, is strongly sequence dependent, and is correlated with a charge pattern parameter κ that quantifies local deviations from global charge asymmetry (8). A subsequent analytical treatment of the KE sequences by Sawle and Ghosh (26) rationalized the trend through another charge pattern parameter called "sequence charge decoration" (SCD) that also correlates well with $\langle R_g \rangle$. For IDP phase separation, a recent sequence-dependent random-phase-approximation (RPA) approach we put forth (27,28) accounted for the experimental difference in phase-separation tendency between the wild-type and a charge-scrambled mutant of the 236-residue N-terminal fragment of DEAD-box RNA helicase Ddx4 (16).

These advances suggest that a deeper understanding of the fundamental relationship between single- and multiple-chain IDP properties is in order. It would be helpful, for instance, if experiments on single-chain properties can infer the conditions under which a protein sequence would undergo multiple-chain phase separation. We embark on this endeavor by first focusing on electrostatics, while leaving aromatic and other π-interactions—which can figure

Submitted March 8, 2017, and accepted for publication April 18, 2017.
*Correspondence: chan@arrhenius.med.toronto.edu
Editor: Rohit Pappu.
http://dx.doi.org/10.1016/j.bpj.2017.04.021

prominently in IDP behavior (16,27,29)—to future effort. To reach this initial goal, we apply RPA to the 30 KE sequences of length $N = 50$ to ascertain their phase-separation properties under salt-free conditions. Adopting our previous notation and making the same simplifying assumption that amino acid residues and water molecules are of equal size in the theory (27,28), the free energy $F_{\rm RPA}$ of the multiple-chain system of a given polyampholytic sequence with charge pattern $\{\sigma_i\} = \{\sigma_1, \sigma_2,...,\sigma_N\}$, where $\sigma_i = \pm 1$ is the sign of electronic charge of the ith residue, is given by (see Eqs. 13 and 40 of (28)):

$$\frac{F_{\rm RPA}a^3}{Vk_{\rm B}T} = \frac{\phi_m}{N}\ln\phi_m + (1-\phi_m)\ln(1-\phi_m) + \int_0^\infty \frac{dkk^2}{4\pi^2}\{\ln[1+\mathcal{G}(k)] - \mathcal{G}(k)\}, \quad (1)$$

where $a = 3.8$ Å is the Cα-Cα distance, V is the system volume, $k_{\rm B}$ is the Boltzmann constant, T is the absolute temperature, $\phi_m = \rho_m a^3$ is the volume ratio of amino acid residues wherein ρ_m/N is protein density, and $\mathcal{G}(k)$ is given by:

$$\mathcal{G}(k) = \frac{4\pi\phi_m}{k^2(1+k^2)T^*N}\sum_{i,j=1}^{N}\sigma_i\sigma_j\exp\left(-\frac{k^2}{6}|i-j|\right). \quad (2)$$

Here $T^* \equiv a/l_B$ is the reduced temperature. The Bjerrum length is $l_B = e^2/(4\pi\epsilon_0\epsilon_r k_{\rm B}T)$, where e is the elementary charge, ε_0 is the vacuum permittivity, and ϵ_r is the relative permittivity (27,28); $\epsilon_r \approx 80$ for water, but can be significantly lower for water-IDP solutions (28). Here ϵ_r is treated largely as an unspecified constant because our main concern is the relative $T^*_{\rm cr}$ values of different sequences.

We determined the phase diagrams of the 30 KE sequences from the free energy expression Eq. 1 using standard procedures (28). For each sequence, the highest temperature on the coexistence curve is the critical temperature $T^*_{\rm cr}$, which is the highest T^* at which phase separation can occur (Fig. 1 *a*). The critical temperatures of the KE sequences are highly diverse, ranging from $T^*_{\rm cr} = 0.089$ (sv1) to 8.570 (sv30). The variation of critical volume fraction $\phi_{\rm cr} \equiv \phi_{\rm m}(T^*_{\rm cr})$ from 0.0123 (sv30, sv24) to 0.0398 (sv1) is narrower. The KE sequences were originally labeled as sv1, sv2, …, sv30 in ascending values for Das and Pappu's charge pattern parameter κ, from the strictly alternating sequence sv1 with $\kappa = 0.0009$ (minimum segregation of opposite charges) to the diblock sequence sv30 with $\kappa = 1.0$ (maximum charge segregation) (8). Our RPA-predicted $T^*_{\rm cr}$ values follow largely, though not exactly, the same order: sv1 and sv30 have the lowest and highest $T^*_{\rm cr}$ values, respectively; however, e.g., sv24 rather than sv27 has the fourth largest $T^*_{\rm cr}$ and sv5, not sv2, has the second lowest $T^*_{\rm cr}$. If $\epsilon_r = 80$ is assumed, RPA predicts that 21 KE sequences can, but 9 KE sequences cannot, phase-separate at $T \gtrsim 300$ K (Fig. 1, *b* and *c*).

Because $\langle R_g\rangle$ correlates positively with κ (8), this $T^*_{\rm cr}$ trend suggests that multiple-chain $T^*_{\rm cr}$ should correlate with single-chain $\langle R_g\rangle$. Indeed, a striking correlation (Fig. 1 *b*) satisfying the approximate power-law:

$$T^*_{\rm cr} \approx 9.8 \times 10^7 \langle R_g\rangle^{-5.83}, \quad (3)$$

with R_g in units of Å, is observed for the KE sequences. The variation of $T^*_{\rm cr}$ with $\langle R_g\rangle$ is very sharp: $T^*_{\rm cr}$ increases ~100 times whereas $\langle R_g\rangle$ decreases by $\lesssim$50%. Qualitatively, the positive ($T^*_{\rm cr}$)–$\langle R_g\rangle$ correlation may be understood by considering two extreme cases: the diblock, and the strictly alternating sequences (Fig. 2). For the diblock, attractive interactions are absent—cannot be satisfied—within most stretches of several (e.g., <6) residues. However, once a pair of opposite charges is in spatial proximity,

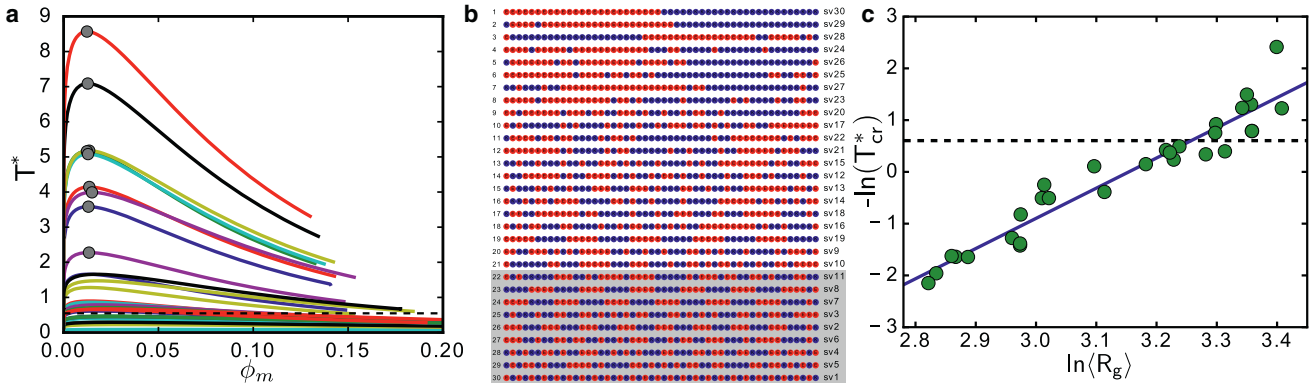

FIGURE 1 (*a*) Shown here are coexistence curves computed by RPA for KE sequences 1–30 in (*b*), listed in descending order of $T^*_{\rm cr}$ (except sequences 23 and 24, which have the same $T^*_{\rm cr}$), with K and E residues in red and blue, respectively; those with $T^*_{\rm cr} < 0.55$ (corresponding to $T < 300$ K when $\varepsilon_r = 80$) are shown on a gray background in (*b*). The "sv" sequence labels are those in Das and Pappu (8). Critical points ($T^* = T^*_{\rm cr}$) for several high-$T^*_{\rm cr}$ sequences are marked by circles in (*a*). (*c*) Shown here is a logarithmic correlation between RPA-predicted $T^*_{\rm cr}$ and $\langle R_g\rangle$ simulated in Das and Pappu (8) (*green circles*). The fitted line (*blue*) is $-\ln T^*_{\rm cr} = -18.4 + 5.83\ln\langle R_g\rangle$ with squared Pearson coefficient $r^2 = 0.92$. The dashed horizontal line represents $T^*_{\rm cr} = 0.55$.

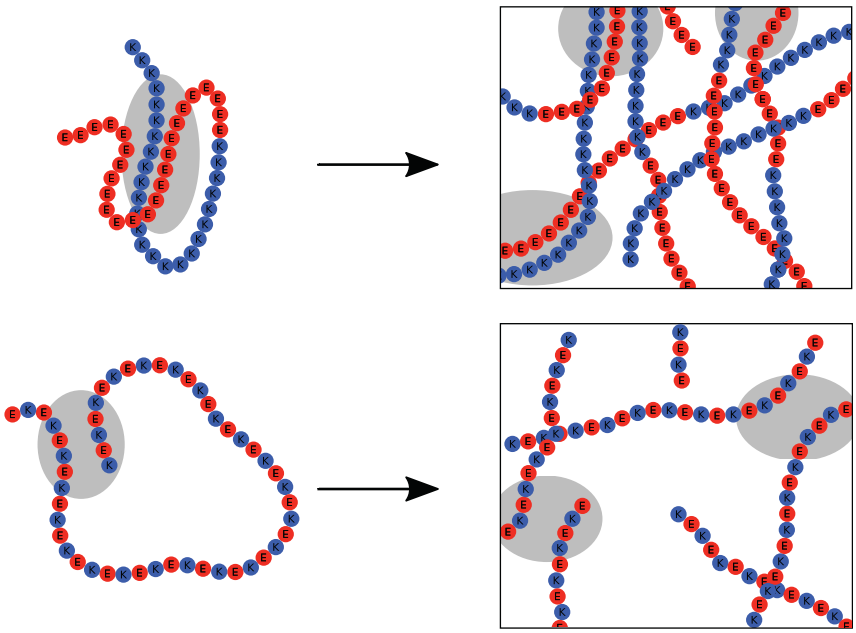

FIGURE 2 Schematics. Similar electrostatic effects are at play in single-chain compactness (*left*) and multiple-chain phase separation (*right*). (*Top*) Long stretches of like charges entail strong intra- and interchain attractions (*gray areas*). Favorable intrachain interactions are among residues that are nonlocal, i.e., more than a few residues apart, along the chain sequence. Most local interactions are repulsive because of the charge blocks. (*Bottom*) Attractions within and among polyampholytes that lack long charge blocks are weaker. Overall attractive interactions now require conformationally restrictive charge pairings and are weaker because of repulsion from neighboring like charges.

chain connectivity brings two oppositely charged blocks together, leading to a strong Coulomb attraction, thus a small $\langle R_g \rangle$ and a higher tendency to phase-separate (higher T^*_{cr}). In contrast, for the strictly alternating sequence, attractive Coulomb interactions that are already weakened relative to that of the diblock sequence require more conformational restriction, resulting in more open, large-$\langle R_g \rangle$ single-chain conformations and less tendency to phase-separate (lower T^*_{cr}).

It is instructive to compare the predictive power of κ and another charge pattern parameter SCD $\equiv \sum_{i<j}^{N} \sigma_i \sigma_j \sqrt{j-i}/N$ that has emerged from the analysis of Sawle and Ghosh (26). The two parameters are well correlated ($r^2 = 0.95$; see Fig. 7 of Sawle and Ghosh (26)), yet the variation of both T^*_{cr} and $\langle R_g \rangle$ of the KE sequences is significantly smoother with respect to SCD than κ (Fig. 3). For example, despite the large variation in κ for sv24, sv26, and sv28 (0.45, 0.61, and 0.77, respectively), their $\langle R_g \rangle = 17.6$, 17.5, and 17.9 Å (8), and their $T^*_{cr} = 5.16$, 5.08, and 5.18, are almost identical (Fig. 3 *b*). This similarity, however, is well reflected by their similar SCD $= -17.0$, -16.2, and -16.0. Indeed, a near-linear relationship ($r^2 = 0.997$):

$$T^*_{cr} \approx -0.314(\mathrm{SCD}), \tag{4}$$

is observed (Fig. 3 *b*). A likely origin of SCD's better performance is that it accounts for potential interactions between charges far apart along the sequence, whereas κ relies on averaging over five or six consecutive charges. Accordingly, SCD is less sensitive than κ to isolated charge reversals. The rather smooth SCD–$\langle R_g \rangle$ dependence is remarkable because the simulated $\langle R_g \rangle$ (8) bears no formal relationship with the variational theory from which SCD emerges (26). Future effort should be directed toward further assessment of these and other possible charge pattern parameters (30) as predictors for IDP conformational properties.

In summary, we have quantified a close relationship between single-chain conformational compactness of polyampholytes and their phase-separation tendency. The above

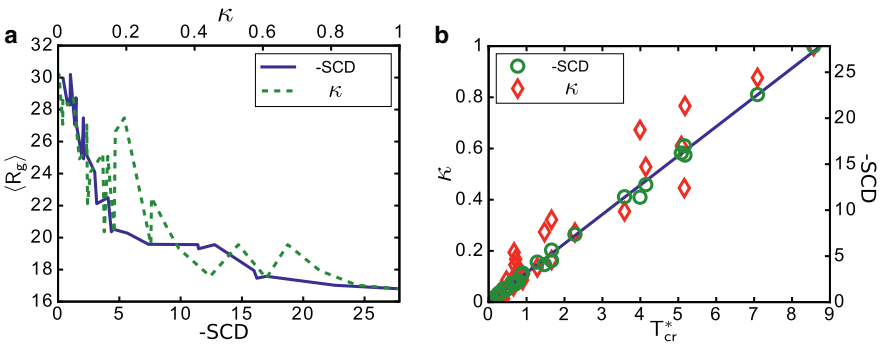

FIGURE 3 Charge-pattern parameters. (*a*) Given here is the single-chain $\langle R_g \rangle$ in Das and Pappu (8) versus the κ-parameter of Das and Pappu (8) (*top horizontal scale*) and the SCD parameter of Sawle and Ghosh (26) (*bottom scale* for $-$SCD). (*b*) Given here is the variation of RPA-predicted T^*_{cr} with κ (*left vertical scale*) and $-$SCD (*right vertical scale*).

RPA results were derived with a short-range cutoff for Coulomb interactions to account for residue sizes (28,31). If we had adopted an unphysical interaction scheme without such a cutoff, similar trends would still hold although the scaling relations Eqs. 3 and 4 would be modified, respectively, to $T^*_{cr} \sim (R_g)^{-3.57}$ and $T^*_{cr} \approx -0.490(\mathrm{SCD})$. Thus, in any event, basic physics dictates a rather sharp positive correlation between T^*_{cr} and $\langle R_g \rangle$. This connection should be further explored by both theory and simulation (30,32) to help decipher the sequence determinants of IDP phase separation.

AUTHOR CONTRIBUTIONS

Y.-H.L. and H.S.C. both designed research, performed research, analyzed data, and wrote the article.

ACKNOWLEDGMENTS

We thank Julie Forman-Kay and Robert Vernon for helpful discussions.

This work was supported by Canadian Cancer Society Research Institute grant No. 703477, Canadian Institutes of Health Research grant No. MOP-84281, and computational resources provided by SciNet of Compute Canada.

REFERENCES

1. Uversky, V. N., J. R. Gillespie, and A. L. Fink. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins.* 41:415–427.
2. Tompa, P. 2012. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* 37:509–516.
3. Forman-Kay, J. D., and T. Mittag. 2013. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure.* 21:1492–1499.
4. van der Lee, R., M. Buljan, …, M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114:6589–6631.
5. Chen, T., J. Song, and H. S. Chan. 2015. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Curr. Opin. Struct. Biol.* 30:32–42.
6. Das, R. K., K. M. Ruff, and R. V. Pappu. 2015. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 32:102–112.
7. Sickmeier, M., J. A. Hamilton, …, A. K. Dunker. 2007. DisProt: the database of disordered proteins. *Nucleic Acids Res.* 35:D786–D793.
8. Das, R. K., and R. V. Pappu. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA.* 110:13392–13397.
9. Higgs, P. G., and J.-F. Joanny. 1991. Theory of polyampholyte solutions. *J. Chem. Phys.* 94:1543–1554.
10. Dobrynin, A. V., R. H. Colby, and M. Rubinstein. 2004. Polyampholytes. *J. Polym. Sci. Part B Polym. Phys.* 42:3513–3538.
11. Liu, B., D. Chia, …, C. C. Gradinaru. 2014. The effect of intrachain electrostatic repulsion on conformational disorder and dynamics of the Sic1 protein. *J. Phys. Chem. B.* 118:4088–4097.
12. Song, J., G. N. Gomes, …, H. S. Chan. 2015. An adequate account of excluded volume is necessary to infer compactness and asphericity of disordered proteins by Förster resonance energy transfer. *J. Phys. Chem. B.* 119:15191–15202.
13. Borg, M., T. Mittag, …, H. S. Chan. 2007. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc. Natl. Acad. Sci. USA.* 104:9650–9655.
14. Csizmok, V., S. Orlicky, …, J. D. Forman-Kay. 2017. An allosteric conduit facilitates dynamic multisite substrate recognition by the SCFCdc4 ubiquitin ligase. *Nat. Commun.* 8:13943.
15. Toretsky, J. A., and P. E. Wright. 2014. Assemblages: functional units formed by cellular phase separation. *J. Cell Biol.* 206:579–588.
16. Nott, T. J., E. Petsalaki, …, A. J. Baldwin. 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell.* 57:936–947.
17. Brangwynne, C. P., P. Tompa, and R. V. Pappu. 2015. Polymer physics of intracellular phase transitions. *Nat. Phys.* 11:899–904.
18. Molliex, A., J. Temirov, …, J. P. Taylor. 2015. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell.* 163:123–133.
19. Bergeron-Sandoval, L.-P., N. Safaee, and S. W. Michnick. 2016. Mechanisms and consequences of macromolecular phase separation. *Cell.* 165:1067–1079.
20. Pak, C. W., M. Kosno, …, M. K. Rosen. 2016. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell.* 63:72–85.
21. Zeng, M., Y. Shang, …, M. Zhang. 2016. Phase transition in postsynaptic densities underlies formation of synaptic complexes and synaptic plasticity. *Cell.* 166:1163–1175.e12.
22. Feric, M., N. Vaidya, …, C. P. Brangwynne. 2016. Coexisting liquid phases underlie nucleolar subcompartments. *Cell.* 165:1686–1697.
23. Keating, C. D. 2012. Aqueous phase separation as a possible route to compartmentalization of biological molecules. *Acc. Chem. Res.* 45:2114–2124.
24. Mao, A. H., S. L. Crick, …, R. V. Pappu. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:8183–8188.
25. Müller-Späth, S., A. Soranno, …, B. Schuler. 2010. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:14609–14614.
26. Sawle, L., and K. Ghosh. 2015. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* 143:085101.
27. Lin, Y.-H., J. D. Forman-Kay, and H. S. Chan. 2016. Sequence-specific polyampholyte phase separation in membraneless organelles. *Phys. Rev. Lett.* 117:178101.
28. Lin, Y.-H., J. Song, …, H. S. Chan. 2017. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *J. Mol. Liq.* 228:176–193.
29. Song, J., S. C. Ng, …, H. S. Chan. 2013. Polycation-π interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family. *PLOS Comput. Biol.* 9:e1003239.
30. Holehouse, A. S., R. K. Das, …, R. V. Pappu. 2017. CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys. J.* 112:16–21.
31. Ermoshkin, A. V., and M. Olvera de la Cruz. 2003. Polyelectrolytes in the presence of multivalent ions: gelation versus segregation. *Phys. Rev. Lett.* 90:125504.
32. Ruff, K. M., T. S. Harmon, and R. V. Pappu. 2015. CAMELOT: a machine learning approach for coarse-grained simulations of aggregation of block-copolymeric protein sequences. *J. Chem. Phys.* 143:243123.