

ARTICLE

Electronegative clusters modulate folding status and RNA binding of unstructured RNA-binding proteins

Steve Zaharias¹ | Talia Fargason¹ | Rory Greer² | Yuhua Song²  | Jun Zhang¹ 

¹Department of Chemistry, College of Arts and Sciences, The University of Alabama at Birmingham, Birmingham, Alabama, USA

²Department of Biomedical Engineering, The University of Alabama at Birmingham, Birmingham, Alabama, USA

Correspondence

Jun Zhang, Department of Chemistry, College of Arts and Sciences, The University of Alabama at Birmingham, Birmingham, AL 35294, USA.
Email: zhanguab@uab.edu

Funding information

United States National Science Foundation, Grant/Award Number: 2024964

Review Editor: Aitziber L. Cortajarena

Abstract

Electronegative clusters (ENCs) made up of acidic residues and/or phosphorylation sites are the most abundant repetitive sequences in RNA-binding proteins. Previous studies have indicated that ENCs inhibit RNA binding for structured RNA-binding domains (RBDs). However, this is not the case for the unstructured RBD in histone pre-mRNA stem-loop binding protein (SLBP). The SLBP RBD contains 70 amino acids and is followed by a phosphorylatable ENC. ENC phosphorylation increases RNA-binding affinity of SLBP to the sub-picomolar range. In this study, we use NMR and molecular dynamics simulations to elucidate the mechanism for this tight binding. Our NMR data demonstrate that the ENC transiently folds apo SLBP into an RNA-bound resembling state. We find that in the RNA-bound state, the phosphorylated ENC interacts with the loop region opposite to the RNA-binding site. This allosteric interaction stabilizes the complex and therefore enhances RNA binding. To evaluate the generality of our findings, we graft an ENC onto endoribonuclease homolog 1's first double-stranded RNA-binding motif (DRBM1), an unstructured RBD that shares no homology with SLBP. We find that the engineered ENC increases the folded species of DRBM1 and inhibits RNA binding. On the contrary, introducing basic residues to DRBM1 makes the domain more unfolded, enhances RNA binding, and mitigates the inhibitory effect of the engineered ENC. In summary, our study suggests that ENCs promote folding of unstructured RNA-binding domains, and their effects on RNA binding depend on the electropositive charges on the RBD surface.

KEYWORDS

acidic patch, allostery, electronegative clusters, intrinsically disordered proteins, poly-D/E, RNA-binding proteins, stem-loop binding protein

1 | INTRODUCTION

Intrinsically disordered protein regions lack persistent structure and account for more than a third of the

eukaryotic proteome (Pentony & Jones, 2010; Uversky et al., 2000; Uversky et al., 2005). These regions can serve as extended and dynamic binding interfaces and act as regulatory switches in many biological pathways,

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

including transcription, translation, signal transduction, and protein assembly (Bah et al., 2015; Csizmok & Forman-Kay, 2018; Kriwacki et al., 1996; Muller-Spath et al., 2010; Shoemaker et al., 2000). Disruption of any of these processes leads to a variety of diseases and cancers. About 25% of disease-associated missense mutations can be mapped onto disordered protein regions, and approximately 70% of proteins associated with human cancer contain relatively long unstructured regions (Iakoucheva et al., 2002; Oates et al., 2013; Vacic & Iakoucheva, 2012).

RNA-binding proteins are enriched with disordered protein regions that contain repetitive sequences (Wang et al., 2016). We have recently discovered that the most abundant repetitive regions in RNA-binding proteins are electronegative clusters (ENCs) made up of consecutive acidic residues and/or phosphorylation sites. Around one third of RNA-binding proteins contain ENCs of four amino acids or longer (Zaharias et al., 2021). The majority of ENCs are composed of acidic residues (poly-D/E patches). However, a portion of ENCs have embedded phosphorylation sites, which renders ENCs tunable. ENCs often flank RNA-binding domains (RBD). ENCs can stabilize the structure of neighboring RBDs and inhibit RNA binding through intramolecular interactions (Zaharias et al., 2021). These stabilizing and inhibitory effects have been observed for several proteins with structured RNA-binding domains, such as Nop15 and Hfq (Santiago-Frangos et al., 2017; Zaharias et al., 2021). When the ENC is just long enough to interact with the electropositive residues not responsible for specific RNA binding, it minimally affects specific RNA binding, but significantly suppresses nonspecific binding. Further elongating the ENC will significantly inhibit specific RNA binding (Zaharias et al., 2021).

However, the inhibitory effect of ENCs cannot be extended to *Drosophila* histone pre-mRNA stem-loop binding protein (SLBP). SLBP specifically binds to the histone mRNA stem loop, and the complex is maintained until the end of the histone mRNA lifecycle (Marzluff & Duronio, 2002). This tight binding prevents pre-mature degradation of histone mRNA by the exonuclease, 3' hExo, which binds to the other side of the stem-loop RNA (Tan et al., 2013). SLBP is unstructured throughout the whole protein, consisting of an N-terminal domain, a unique RBD, and a C-terminal ENC with four serine residues that can be phosphorylated (sequence: SNSDSDSD). The RBD and the C-terminal ENC constitute the minimal histone processing domain (Figure 1a) (Marzluff & Duronio, 2002). Phosphorylation of the SLBP ENC is required for efficient processing of histone pre-mRNA (Dominski et al., 2002). Instead of inhibiting RNA binding, ENC phosphorylation increases the RNA-binding

affinity of SLBP to the sub-picomolar range, which is the tightest protein RNA binding to the best of our knowledge (Zhang et al., 2014). Nonphosphorylated SLBP and the phosphomimetic (in which serine residues in the ENC are mutated to glutamic acid) have an identical RNA-bound structure, and the ENC regions cannot be observed by X-ray crystallography (Zhang et al., 2014). RNA binding enhancement partially stems from the unstructured apo state. Using fluorescence resonance energy transfer, our previous studies have shown that the phosphomimetic and phosphorylated SLBP are more compact and more stable than the nonphosphorylated protein (Davis et al., 2018; Zhang et al., 2014). However, it is still unclear how SLBP with the phosphorylated ENC achieves this extraordinarily tight binding at the atomic level. In addition to the elucidation of the biological function of SLBP, knowledge about this tight binding can be useful for the design of RNA-binding proteins. Considering the prevalence of ENCs, it is also of general interest to understand why ENCs exert opposite effects on RNA binding for structured and unstructured RBDs.

Using NMR, fluorescence polarization binding assays, and molecular dynamics (MD) simulations, this study answers these critical questions at the atomic level and elucidates regulatory roles of ENCs in RNA binding and protein folding. We find that the ENC transiently folds apo SLBP into a conformation that resembles the RNA-bound state. The RNA-bound resembling species is stabilized by ENC phosphorylation. Our NMR chemical shift perturbation and relaxation results show that in the RNA-bound state, the phosphorylated ENC interacts with the SLBP loop region that is distal to the RNA-binding site. Consistently, our MD simulations reveal a dynamic, but persistent network of salt-bridge and hydrogen bond interactions between the phosphorylated ENC and the distal loop. This interacting network is absent in nonphosphorylated SLBP and the mutant without the ENC. More interestingly, we found that it is a general phenomenon that ENCs can induce protein folding of unstructured RBDs. An engineered ENC can also prompt folding of the first double stranded-RNA binding motif (DRBM1) of endoribonuclease homolog 1 (DCL1), an unstructured RBD that shares no homology with SLBP. Introducing basic residues to DRBM1 eliminates the folded species, but significantly enhances RNA binding. Moreover, increasing the amount of positive charges on the surface of DRBM1 can attenuate the inhibitory effect of the engineered ENC. These results suggest that the effects of ENCs on RNA binding depends on surface charges of RBDs. Our work demonstrates the interplay between protein folding, RNA binding, and ENC regulation and suggests the potential application of ENCs in protein design.

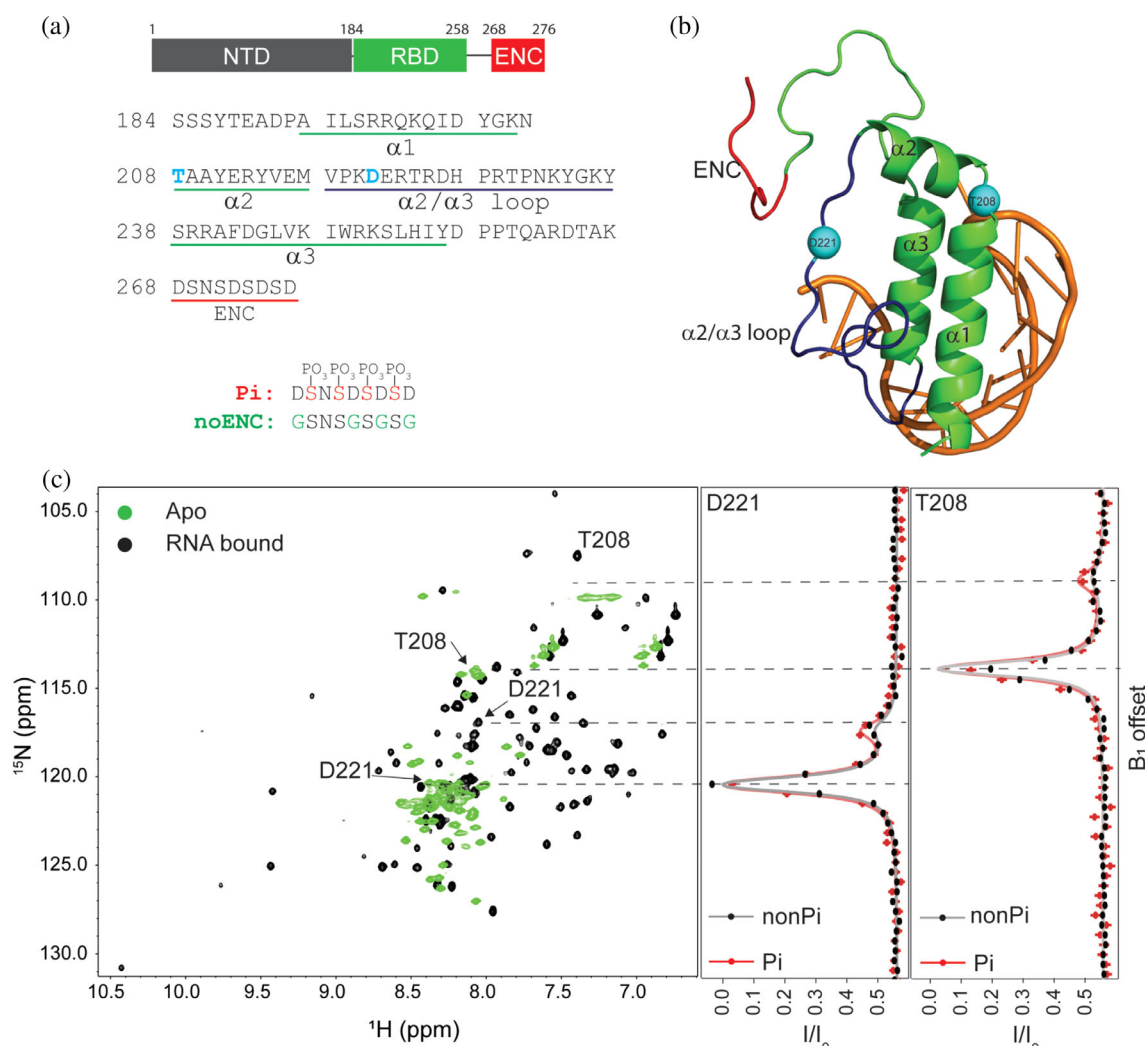


FIGURE 1 Phosphorylation of the stem-loop binding protein (SLBP) electronegative cluster (ENC) increases the stability and population of the conformer that resembles the RNA-bound state. (a) Domain architecture of SLBP, which consists of an N-terminal domain (NTD), an RNA-binding domain (RBD), and an ENC. The primary sequence encompassing RBD and ENC (residues 184–276) used in this study is shown. The ENC region sequences in phosphorylated and noENC constructs were shown below. The noENC construct is a control in which acidic ENC residues are replaced by Gly, and ser residues are nonphosphorylated. In this study, the NTD was not included in SLBP constructs, as it is not involved in RNA binding. (b) The structure model is prepared based on crystal structure of SLBP: RNA complex (PDB ID: 4TUX). T208 and D221 are shown in cyan, and RNA is in orange. The ENC and the $\alpha 2/\alpha 3$ loop are shown in red and dark blue, respectively. (c) Example chemical exchange saturation transfer (CEST) profiles of T208 and D221 in apo SLBP. CEST profiles for other residues are shown in Figures S1, S2, and S3. The B_1 offset has been converted into ppm and aligned with the HSQC on the left. Therefore, its Hz unit is not shown. The B_1 strength is 15 Hz.

2 | RESULTS

2.1 | apo SLBP transiently samples the conformation that resembles the RNA-bound state

Our previous studies have shown that the RBD of apo SLBP is unstructured, and phosphorylation of the C-terminal ENC increases molecular compactness and thermodynamic stability (Davis et al., 2018; Zhang et al., 2014). However, no significant difference in

secondary structure is found between the phosphorylated and nonphosphorylated SLBP based on circular dichroism (Davis et al., 2018). To investigate the role of the ENC, we prepared three SLBP constructs (Figure 1a): wild-type SLBP with the nonphosphorylated ENC (nonPi), SLBP with the phosphorylated ENC (Pi), and the construct with unphosphorylated Ser and all acidic tail residues mutated to Gly (noENC). Like many unstructured proteins, apo SLBP exists as an ensemble of numerous inter-exchanging conformations. In the ensemble, the conformation competent for ligand binding (excited

state) is sparsely populated. It is challenging to detect the sparsely-populated conformers using traditional NMR methods that detect only the highly populated ground state. To provide atomic-level insight into the apo RBD, we performed chemical exchange saturation transfer (CEST) experiments on these SLBP constructs in the apo state (Figure 1) (Vallurupalli et al., 2012). A CEST experiment employs a weak saturation field (B_1) applied along the ^{15}N dimension. Although the position of B_1 can be conveniently converted into chemical shift, it is named as B_1 offset in Hz, to distinguish from chemical shift (in ppm). When B_1 is applied at the chemical shift of the ground state (unfolded state in our case) of a resonance, it saturates the resonance and yields peak intensity attenuation or a major “dip”. When B_1 is applied at the chemical shift of the excited state (folded state in our case) that exchanges with the ground state, it also attenuates the resonance intensity of the ground state, yielding a minor “dip” (i.e., a CEST signal). In summary, CEST indirectly probes the sparsely-populated state by measuring how the NMR signal of the ground state changes along with the B_1 saturating field. A CEST profile conveys information about the exchanging process, such as population of the ground (P_A) and excited (P_B) states, exchange rates k_{ex} , rate constants for the ground-to-excited (k_{AB}) and excited-to-ground (k_{BA}) transitions, and the chemical shift for the excited state. It is noteworthy that chemical shift is sensitive to local environment and structure of an atom. Therefore, it is frequently used to probe the structural nature of the excited state.

We observed multiple residues demonstrating CEST signals for all three SLBP constructs in the apo state, suggesting that these constructs underwent conformational exchanges (Figures 1 and S1–S3). These extensive CEST signals indicate a large-scale conformational change of SLBP instead of only local structural adjustment. We assumed that they stemmed from the same chemical exchange process and performed a global fitting for CEST data. Representative CEST profiles for T208 and D221 are shown for apo nonPi and Pi constructs (Figure 1c). The minor dip positions for these residues are close to their chemical shifts in the RNA-bound state, suggesting the structure of the excited state resembles the RNA-bound state. It is noteworthy that the minor dip of T208 slightly differs from its chemical shift in the RNA-bound state. This kind of difference is also observed for some other residues showing CEST. SLBP is a small domain (~70 amino acids) that forms extensive interactions with the stem-loop RNA (Figure 1b). Most residues showing CEST undergo conformational change and are in close proximity to the stem-loop RNA. For example, T208 is located at the $\alpha 1/\alpha 2$ hinge adjacent to RNA (Figure 1b). Therefore, its chemical shift in the RNA-bound state is a

convolution of both conformational change and RNA proximity. However, the effect of RNA proximity is absent in the apo state in which the CEST data were collected. This explains the chemical shift difference between the CEST data and the RNA-bound HSQC. To probe the nature of the exchange, we further compared the chemical shift dispersion of the major and the minor state. Unstructured proteins have a narrow chemical shift distribution because residues are solvent exposed and consequently have a similar environment on average. However, structured proteins have a dispersed chemical shift distribution because residues have distinct local structure. We compared standard deviation of nonPi SLBP in the major and minor states. Overall, chemical shift standard deviation of the minor state is higher than that of the major state (6.8 vs. 4.9 ppm, Table S1). This analysis suggests that the minor state represented a more folded state compared with the major state. Overall, our CEST data indicate that apo SLBP constructs transiently sample a more folded conformation that may resemble the RNA-bound state.

We further compared the folding (k_{AB}) and unfolding (k_{BA}) rate constants and population of the excited state (P_B) for these three constructs (Table 1). The folding rate constants of these three constructs are similar. In contrast, the unfolding rate constant is reduced by more than threefold by phosphorylation. Correspondingly, the excited-state population increases from 0.3% to 1.3%. Our CEST results suggest that apo SLBP can sample the conformation that resembles the RNA-bound state, and phosphorylation of the ENC stabilizes and enriches the RNA-bound resembling species.

2.2 | The ENC of RNA-bound SLBP interacts with the loop region distal to the RNA-binding site

In the section above, we determined how phosphorylation of the ENC regulates the conformational exchange of apo SLBP. We continued to investigate the role of the

TABLE 1 CEST exchanging parameters of SLBP.

CEST	noENC	nonPi	Pi
k_{AB} (s^{-1})	0.28 ± 0.01	0.46 ± 0.01	0.34 ± 0.01
k_{BA} (s^{-1})	80 ± 15	89 ± 6	24 ± 5
k_{ex} (s^{-1})	90 ± 15	99 ± 6	26 ± 5
P_A	$99.70 \pm 0.03\%$	$99.54 \pm 0.02\%$	$98.7 \pm 0.2\%$
P_B	$0.30 \pm 0.03\%$	$0.46 \pm 0.02\%$	$1.3 \pm 0.2\%$

Abbreviations: CEST, chemical exchange saturation transfer; SLBP, stem-loop binding protein.

ENC in the RNA-bound state. We compared RNA-bound HSQC spectra of noENC and Pi constructs, and found that major chemical shift perturbations (CSPs) were located at the ENC (residues 266–276) and the $\alpha 2/\alpha 3$ loop region (mainly residues 220–223, and 233) that is opposite to the RNA-binding site (Figures 2a, 1b, and S4a for location of the loop). As the loop region and the ENC have opposite charges, it is likely these regions interact with each other. We further measured ^{15}N longitudinal (R_1) and transverse relaxation rates (R_2), and heteronuclear NOE (^{15}N - ^1H hetNOE) (Figures 2b, S4). ^{15}N R_2 and ^{15}N - ^1H hetNOE are sensitive to molecular tumbling and local atomic motions. Structured protein regions have larger R_2 values relative to floppy regions, such as termini and loops. The average ^{15}N - ^1H hetNOE for structured protein regions is around 0.8. Floppy regions have much smaller (even negative) ^{15}N - ^1H hetNOE values.

Consistent with the CSP analysis, the $\alpha 2/\alpha 3$ loop and ENC regions have higher NOE values in the Pi construct than the noENC construct, which suggests that these regions become more rigid due to phosphorylation of the ENC (Figure 2b). It is noteworthy that the NOE values of these regions are lower than 0.8, the average value for structured protein regions. Therefore, these regions still

lack stable structure, as also supported by their low R_2 values (Figure S4b).

Our NMR results suggest that the intramolecular interaction between the ENC and the loop region are dynamic. To obtain an atomic-level depiction of the dynamic interaction, we performed 300 ns MD simulations for RNA-bound SLBP constructs (Figures 3 and S5). During these simulations, protein:RNA complexes were maintained for all SLBP constructs (Movies S1–S3). These systems reached equilibrium after 200 ns of simulation (Figure S5), and the last 100 ns of MD simulation trajectories were used for further analysis. Compared with the nonPi and noENC constructs, the phosphorylated ENC forms the most stable interactions with the rest of the complex (Movies S1–S3, SLBP_noD.mpg, SLBP_Pi.mpg, SLBP_WT.mpg). For each construct, we analyzed the percentage of time salt bridges and H-bonds were formed. We found that the phosphorylated ENC forms extensive and persistent salt bridges and H-bonds with basic residues in the loop region during the simulations (Table S2). In contrast, these interactions are mostly absent in the other two constructs (Table S2). Representative conformations for these three constructs are shown in Figure 3a. We also measured average pairwise distances

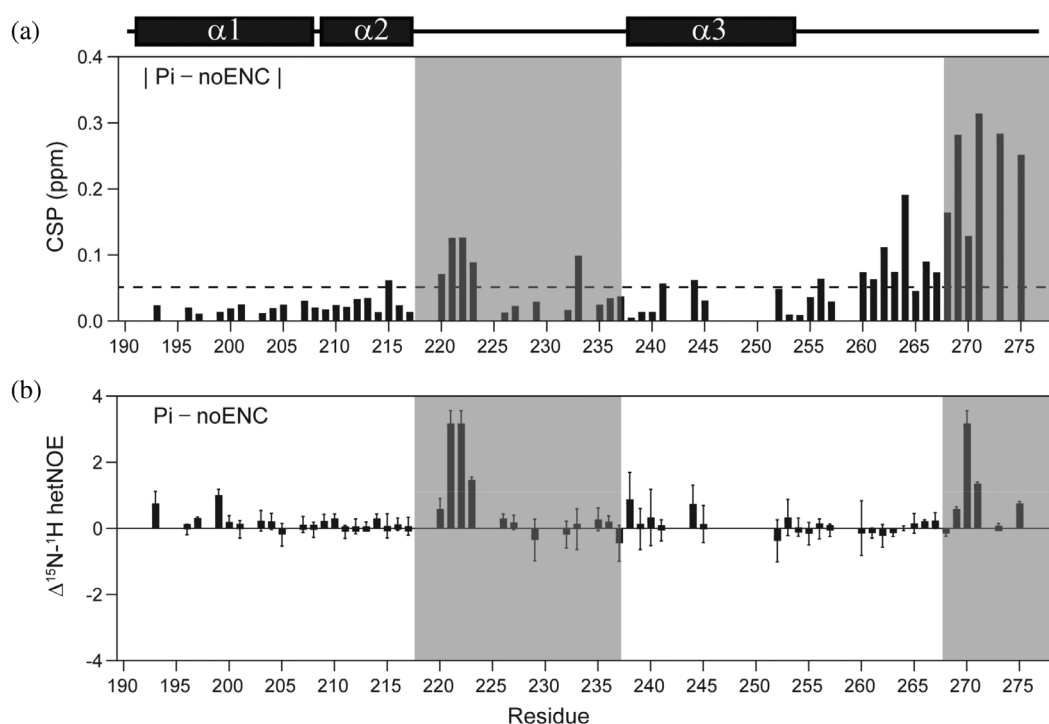


FIGURE 2 NMR analysis reveals the potential electronegative cluster (ENC) interacting sites on the stem-loop binding protein (SLBP): RNA complex. (a) Chemical shift perturbation (CSP) by ENC phosphorylation. CSP values were calculated as $|\delta^1\text{H}| + 0.1 \cdot |\delta^{15}\text{N}|$, where $\delta^1\text{H}$ and $\delta^{15}\text{N}$ are proton and nitrogen chemical shift differences, obtained by comparing the RNA-bound HSQC spectra of phosphorylated SLBP and noENC SLBP. The dash line denotes 0.05 ppm. (b) ^{15}N - ^1H hetNOE difference. The difference was calculated by subtracting the phosphorylated SLBP hetNOE from the noENC SLBP hetNOE. The error was estimated by the ratio of the spectral noise level to resonance intensities as detailed in the experimental procedure section. The gray shadows indicate the $\alpha 2/\alpha 3$ loop and ENC regions.

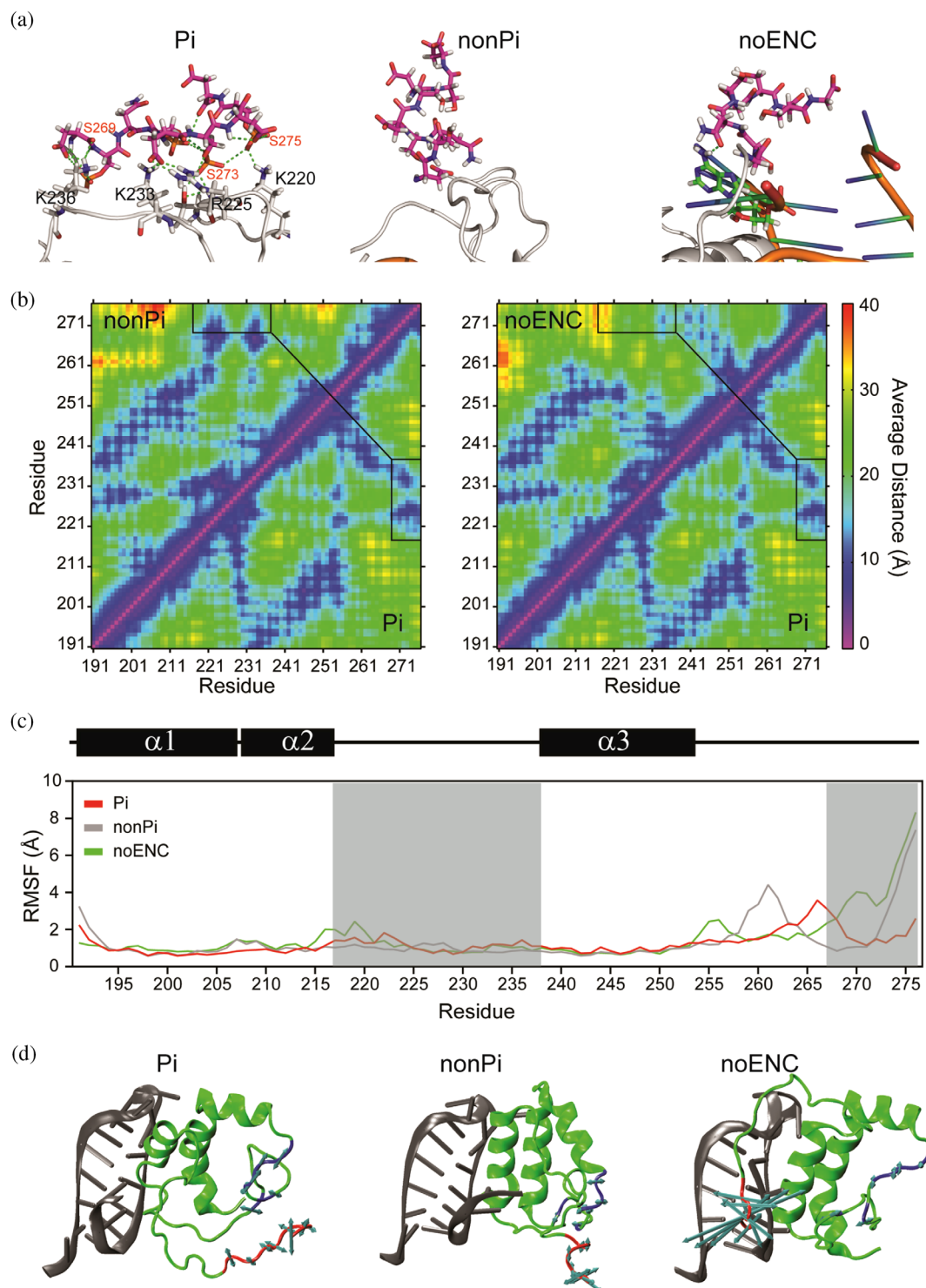


FIGURE 3 Stem-loop binding protein (SLBP) electronegative cluster (ENC) phosphorylation enhances its interaction with the loop opposite to the RNA-binding site. (a) Representative MD simulation snapshots of RNA-bound SLBP with different C-terminal tails. These representative models were selected from the dominant cluster of each construct. The C-terminal tails are shown in magenta sticks, and their interacting residues from the rest of the complexes are shown in gray sticks. Green dotted lines denote the salt-bridges or H-bonds. (b) Comparison of pair-wise residue distances for SLBP with different C-terminal tails. The black boxes highlight average distances between the $\alpha 2/\alpha 3$ loop region (218–237) and the C-terminal tail (268–276). The triangle regions below the diagonals is for phosphorylated SLBP, and the regions above the diagonals are for nonphosphorylated and noENC SLBP, respectively. (c) Comparison of root mean squared fluctuations (RMSFs) of SLBP residues between phosphorylated, nonphosphorylated, and noENC SLBP: RNA constructs. The gray shadows indicate the $\alpha 2/\alpha 3$ loop and ENC regions. (d) Principal component analysis of MD trajectories. The ENC or C-terminal tail was shown in red, and the $\alpha 2/\alpha 3$ loop is shown in blue. The aquamarine arrows indicate the direction and magnitude of principal motions.

for each residue (Figure 3b). These three constructs have similar pairwise distance patterns, consistent with their overall structural similarity. However, the phosphorylated ENC has a shorter distance to the loop region than the other two constructs (highlighted by the black boxes in Figure 3b). The conformational fluctuation analysis reveals that phosphorylation significantly reduces flexibility of the C-terminal tail (Figure 3c). To filter out local fluctuations and to highlight global motions, we performed principal component analysis (PCA) (David & Jacobs, 2014; Haider et al., 2008). Our PCA results reveal that the motion magnitude of the C-terminal tail is smaller for the Pi construct, relative to the noENC construct (Figure 3d). Moreover, apparent correlated motions are observed for the Pi construct in such a way that the phosphorylated ENC moves along with the $\alpha 2/\alpha 3$ loop to maintain the interactions (Movies S1–S3, PCA_SLBP_noD.mpg, PCA_SLBP_Pi.mpg, PCA_SLBP_WT.mpg). This kind of correlation is not observed for nonPi or noENC constructs. These MD simulation findings explain the dramatic increase in ^{15}N - ^1H hetNOE for phosphorylated SLBP. In summary, our NMR experimental results and MD simulations agree with each other and suggest that phosphorylation enhances the intramolecular interaction between the ENC and the $\alpha 2/\alpha 3$ loop in the RNA-bound state.

2.3 | An engineered ENC can modulate folding status and RNA binding of unstructured RNA-binding domains

Our results on SLBP suggest that an ENC can modulate protein folding and RNA binding of the neighboring domain. To test the generality of these findings, we aim to test the effects of a grafted ENC on an orthogonal unstructured RBD. To identify such a system, we calculated net charges and hydrophobicity of 5811 RBDs and generated a hydropathy plot (Figure 4a) (Consortium, 2020). The hydropathy plot can be used to predict folding status of a protein using the unfolding-folding boundary proposed by Uversky (Uversky, 2002; Uversky et al., 2000). In the hydropathy plot, unstructured domains cluster in the top-left corner, where repulsion of net charges outperforms hydrophobic interactions and causes unfolding.

From the hydropathy plot, we selected the first dsRNA binding motif (DRBM1) of endoribonuclease homolog 1 (DCL1), which has no homology to SLBP. DCL1 is a plant protein involved in processing pre-miRNA and is required for the biogenesis of miRNA in *Arabidopsis thaliana* (Wei et al., 2021). DRBM1 is a suitable model in that it has no native ENC and its RNA-bound structure is available (Figure 4b) (Wei et al., 2021). In addition, a previous study has used circular dichroism

to show that DRBM1 is unstructured in its apo state, which is consistent with our hydropathy plot analysis (Figure 4a) (Suarez et al., 2015). DRBM1 folds upon binding to its target RNA. The folded species can be detected in apo DRBM1 even without an ENC (Figure 4c). This is consistent with the fact that DRBM1 is located close to the unfolding/folding boundary in the hydropathy plot (Figure 4a). CEST analysis was conducted on apo DRBM1 constructs. Several DRBM1 residues display CEST signals (Figures 4c, S6, and S7), and an example CEST profile of G1772 is shown in Figure 4c. By aligning the CEST data with HSQC in the apo and RNA-bound states, we found that the minor dip corresponds to the RNA-bound state (Figure 4c). An engineered ENC significantly elevates the minor state population from 0.8% to 20% (Table 2). The folded species can even be observed in the apo spectrum of DRBM1 with engineered ENC at a lower contour level (Figure S8). Similar to the CEST data of SLBP, the grafted ENC only mildly increases k_{AB} , but dramatically reduces k_{BA} (>15 fold).

To investigate the interplay between net charges, RNA binding, protein folding, and ENC regulation, we created DRBM1 constructs with 3 and 9 arginine residues introduced at sites not involved in RNA binding (named as 3R, 9R, respectively) (Figure 4b). As expected, introducing 9 Arg residues completely eliminates CEST signals (Figure S9), suggesting the folded species in DRBM1 is reduced to an undetectable amount. This trend agrees with the hydropathy plot (Figure 4a) in which the 9R construct is located deep into the unfolded protein zone. These results suggest that it is generally applicable that ENCs can promote folding of unstructured RBDs.

Compared with SLBP ($\text{pI} = 10.0$), DRBM1 is less basic ($\text{pI} = 8.9$) and binds to RNA with a much lower affinity ($K_{\text{D}} = 8900 \pm 900$ nM) (Figure 5, Table 3). We found that 3R and 9R constructs bind to RNA with much higher affinities (Table 3). We also investigated the inhibitory effect of the ENC on various DRBM1 constructs. The inhibitory effect is indexed by K_{D} ratios of constructs with ENCs (+ENC) to constructs without ENCs (-ENC). We found that the inhibitory effect of the engineered ENC decreases from 43 to 1.9 fold along with the increase of electropositive residues (Table 3). This is consistent with the fact that increasing electropositive surface creates more interacting sites (Figure 4b) for the ENC and therefore mitigates its competition for RNA binding (Figure 6).

3 | DISCUSSION

Depending on their functions, RBDs bind to RNA with diverse affinities. Forming a stable complex with the histone mRNA is particularly important for SLBP, which prevents mRNA from being degraded by the 3'

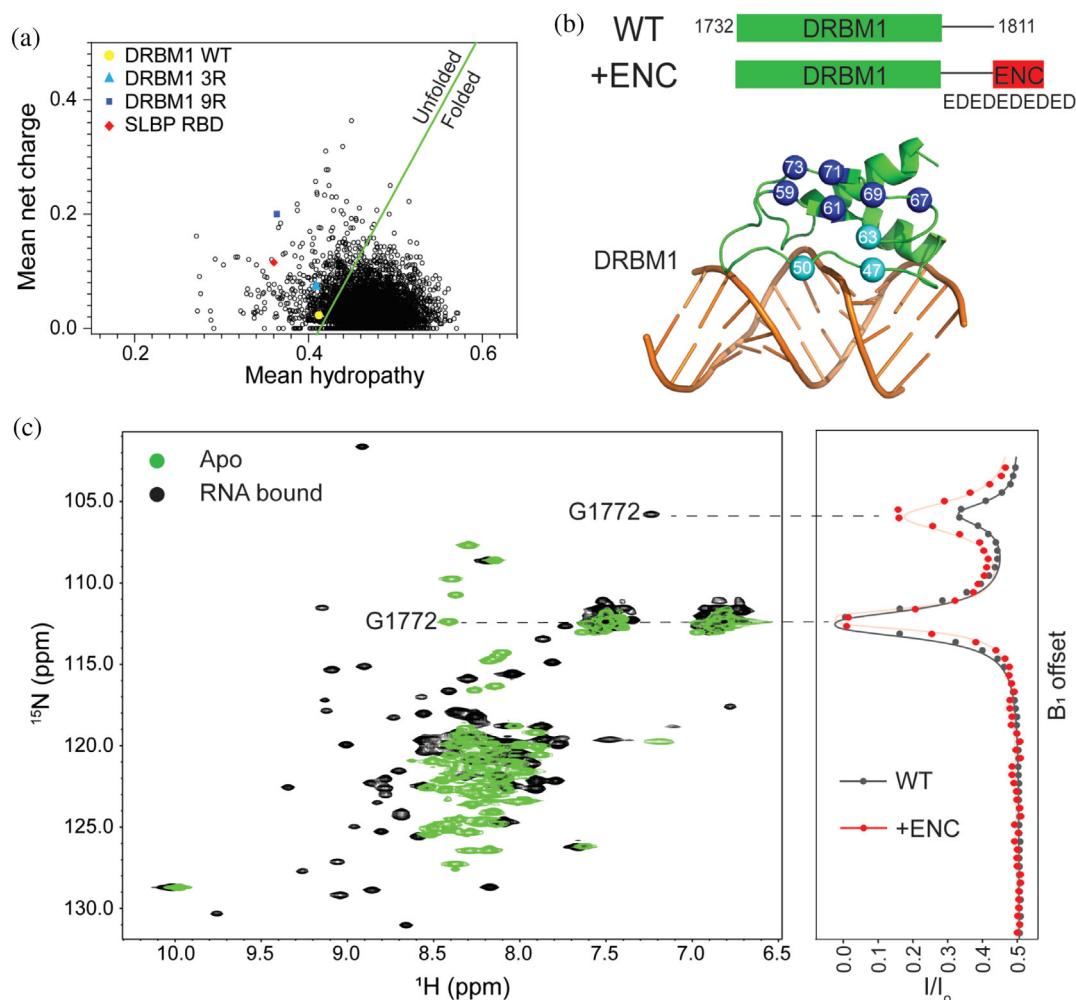


FIGURE 4 Electronegative cluster (ENC)-induced protein folding is observed in the unstructured DRBM1 of DCL1. (a) Hydropathy plot of 5811 RNA-binding domains. The green line is the boundary between folded and unfolded RNA-binding domains. The equation for the boundary line is $y = 2.785x - 1.151$ (Uversky et al., 2000). (b) Domain architecture for WT DRBM1 and the construct with the engineered ENC (+ENC). Cryo-EM structure of DRBM1: RNA complex (PDB ID: 7ELD), in which DRBM1 and RNA are shown as green and orange cartoons, respectively. Cyan spheres denote the DRBM1 residues mutated to Arg in the 3R construct at site 1747, 1750, and 1763. Dark blue spheres denote additional sites mutated to Arg in the 9R construct, 1759, 1761, 1767, 1769, 1771, and 1773. The molecular graphic was prepared using PyMOL. (c) G1772 CEST profiles of apo DRBM1. CEST profiles for other residues are shown in Figures S6 and S7. The B_1 offset has been converted into ppm and aligned with the HSQC shown on the left. Therefore, its unit in Hz is not shown. The B_1 strength is 15 Hz for ENC and 43 Hz for WT.

TABLE 2 CEST exchanging parameters of DRBM1.

CEST	WT	ENC
k_{AB} (s^{-1})	1.91 ± 0.07	3.09 ± 0.07
k_{BA} (s^{-1})	210 ± 20	14 ± 2
k_{ex} (s^{-1})	230 ± 20	15 ± 2
P_A	$99.17 \pm 0.02\%$	$80 \pm 2\%$
P_B	$0.8 \pm 0.02\%$	$20 \pm 2\%$

Abbreviations: CEST, chemical exchange saturation transfer; ENC, electronegative cluster.

exoribonuclease 1 (Tan et al., 2013; Zhang et al., 2014). Therefore, SLBP needs to continuously stay with the

stem-loop RNA until the end of the histone mRNA life-cycle. This functional requirement explains the necessity of the sub-picomolar affinity for SLBP, which is among the tightest protein RNA interaction to the best of our knowledge. It is interesting that this tight binding is achieved through a cooperation of an unstructured RBD and an intrinsically disordered ENC. Our work shows that the ENC increases RNA binding of SLBP through two ways: folding the apo SLBP RBD into a conformation similar to the RNA-bound state, and stabilizing the complex by interacting with the loop distal to the RNA-binding site (Zhang et al., 2014).

Introducing electropositive residues is an efficient approach to increase RNA-binding affinity. This strategy

is used in SLBP, in which basic residues R197, K200, R239, R240, and K247 directly interact with the stem-loop RNA. This strategy was also employed in our study to manipulate binding affinities of engineered DRBM1 (Table 3). However, extensive basic residues unfold a RBD when the charge repulsion outperforms the hydrophobic interaction. According to our hydropathy plot analysis, about 8% of RBDs are unstructured, like SLBP and DRBM1. SLBP RBD has extensive basic residues ($pI = 10.0$), but a small hydrophobic core consisting of four residues, P231, F242, V246, and W249. Although DRBM1 is also unstructured, it has a lower net charge ($pI = 8.9$) and a larger hydrophobic core (V1735, L1738,

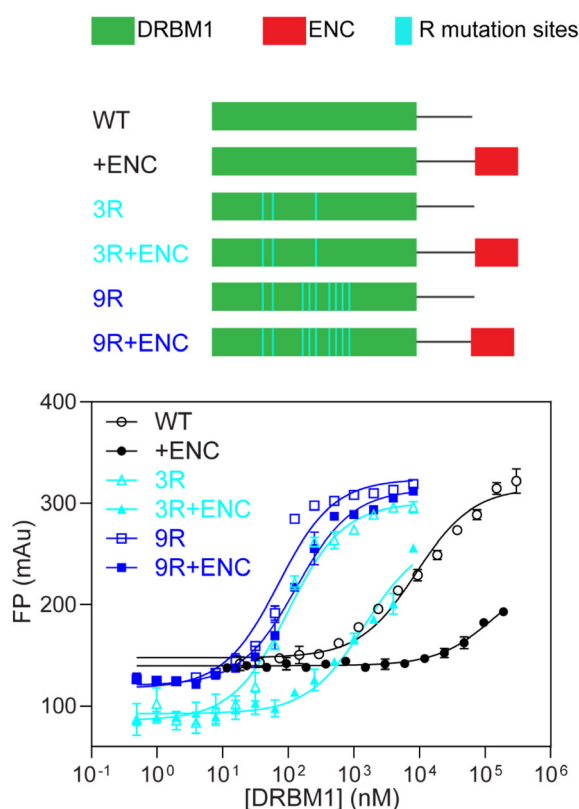


FIGURE 5 Introducing basic residues enhances RNA-binding affinity of DRBM1 and reduces the inhibitory effect of the engineered electronegative cluster (ENC). The mutation sites for the 3R and 9R constructs are indicated by cyan bars. The error was estimated from three individual measurements. Linker residues 1806–1811 were deleted in the 9R constructs to alleviate degradation.

TABLE 3 RNA-binding affinities of DRBM1 constructs.

K_D (nM)	–ENC	+ENC	K_D ratio of +ENC/–ENC
WT	8900 ± 900	$380,000 \pm 63,000$	43
3R	90 ± 11	1400 ± 228	16
9R ^a	66 ± 8	130 ± 10	1.9

^aLinker residues 1806–1811 were deleted in 9R constructs to prevent degradation.

V1762, V1764, and A1787). These differences in net charges and hydrophobicity between SLBP and DRBM1 correlate well with our CEST results that DRBM1 has a higher population of folded species than SLBP RBD. For both proteins, we show that ENC's can significantly increase the population of folded species, which can be used to manipulate folding status in protein engineering. The stabilizing effect is achieved by reducing unfolding rate instead of accelerating protein folding. This finding agrees with a previous study on structured proteins (Tzul et al., 2017). It is found that thioredoxins of the same family have similar folding rates but different unfolding rates (Tzul et al., 2017). This finding also suggests that flanking regions should be taken into consideration in predicting protein folding status, especially for proteins near the folded/unfolded boundary in the hydropathy plot.

Our previous study on the structured RBD of Nop15 has shown that the ENC's inhibitory effect depends on its length (Zaharias et al., 2021). The native ENC of Nop15 inhibits nonspecific RNA binding, while minimally affecting specific RNA binding. However, an ENC with a doubled size significantly inhibits specific RNA binding. This suggests that the inhibitory effect of ENC's depends on the relative size of ENC's and the electropositive surface on the neighboring RBD. For most structured RBDs, their electropositive surface is confined to RNA-binding sites. In this situation, ENC's and RNA tend to compete for RNA-binding sites. In contrast, SLBP has an extensive electropositive surface consisting of 16 basic residues. Only 5 of them interact with the stem-loop RNA, and the rest are scattered on the opposite side of the RNA-binding site. Moreover, the basic residues in the loop region appear to be arranged strategically, as they form dynamic, but persistent interactions with the phosphorylated ENC throughout the RNA bound MD trajectory. This ensures that the ENC is engaged in interacting with the loop region. With this configuration, the phosphorylated ENC of SLBP prefers the loop opposite, and is unable to compete for the RNA-binding site that interacts with RNA with extensive salt bridges, stacking interactions, and H-bonds (Zhang et al., 2014).

In contrast to SLBP, DRBM1 only contains seven basic residues, and 5 of them are on the RNA-binding interface. Also DRBM1 completely depends on salt bridges to bind with RNA. Therefore, the engineered ENC

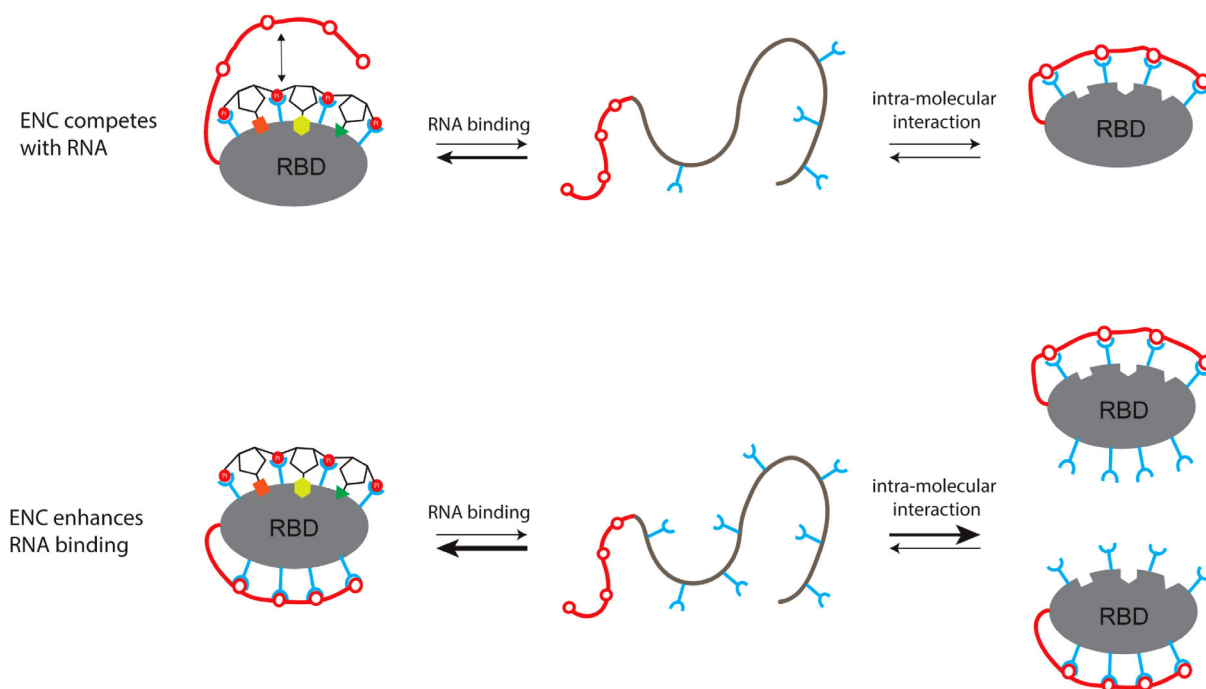


FIGURE 6 Effects of electronegative clusters (ENCs) on protein folding and RNA binding of unstructured RNA-binding domains. ENCs are shown in red, and electropositive residues are shown in cyan. Intra-molecular interactions between ENCs and electropositive sites promote protein folding. The impact of the intra-molecular interaction depends on distribution of electropositive surface. When electropositive surface is only limited to RNA-binding sites, ENCs inhibit RNA binding. The inhibitory effect can be reduced or reversed with increase of electropositive surface.

can efficiently inhibit RNA binding for DRBM1. By increasing electropositive surface on RBDs, the inhibitory effect of the ENC can be attenuated dramatically. This is due to the fact that introduced Arg residues form better binding sites for the engineered ENC than the RNA-binding site. Our work suggests that by changing the electropositive surface of RBDs, ENCs can play a continuum of roles from inhibiting to promoting RNA binding (Figure 6). This represents a promising start toward harnessing ENCs to modulate protein folding and RNA binding.

4 | MATERIALS AND METHODS

4.1 | Bioinformatic analysis

A hydropathy plot was generated using 5811 RBDs gathered from Uniprot (UniProt Consortium, T, 2018). We only selected proteins whose existence is confirmed and whose domain boundaries are annotated by PROSITE, Pfam, and SMART (El-Gebali et al., 2019; Letunic & Bork, 2018; Sigrist et al., 2013). Net charge was calculated by subtracting the number of aspartic and glutamic acid residues from the number of lysine and arginine residues and taking the absolute value divided by the total residue

number of the domain. Hydropathy was calculated using the Kyte and Doolittle scale (Kyte & Doolittle, 1982). This data was analyzed using an in-house Python script available upon request.

4.2 | Protein expression and purification

4.2.1 | SLBP constructs

SLBP (UniProt: Q9VAN6) constructs encompassing the RNA-binding domain (residues 184–267) and the ENC (residues 268–276) were from *Drosophila melanogaster* and cloned into pSMT3 (provided by Christopher Lima, Memorial Sloan Kettering Cancer Center, New York, NY). Mutagenesis PCR was used to create the noENC construct, where S268, S272, S274, and S276 were mutated to glycine. SLBP proteins were purified and phosphorylated as previously reported (Davis et al., 2018). Proteins were expressed at 22°C overnight in *Escherichia coli* strain BL21-CodonPlus (DE3) using 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), which was added when the OD₆₀₀ reached 0.6 AU (0.8 AU for isotope-labeled samples). Pelleted cells were re-suspended in 40 mL lysis buffer (25 mM Tris-HCl, pH 8.5, 1 M NaCl, 25 mM imidazole, 1 mM PMSF,

0.5 mg/mL lysozyme, 1 protease inhibitor tablet, and 0.2 mM tris(2-carboxyethyl)phosphine (TCEP)), and subject to three freeze–thaw cycles. The cells were lysed by sonication and centrifuged at 23,710 relative centrifugal force (RCF) at 4°C for 45 min. The supernatant was applied to 5 mL of Ni Sepharose™ excel resin (GE Healthcare) and washed with 200 mL of loading buffer (25 mM Tris–HCl, pH 8.5, 1 M NaCl, 25 mM imidazole, and 0.2 mM TCEP). The small Ubiquitin-like modifier (SUMO) tag was cleaved by adding 10 mL of loading buffer with 0.1 mg Ubiquitin-like-specific protease 1 (ULP1) overnight at 4°C. The sample was diluted using 20 mM Tris–HCl, pH 7.5, 4 M urea, and 1 mM TCEP and loaded onto a 5-mL HiTrap Heparin Column (GE Healthcare). The sample was eluted with a gradient from 0 to 2 M NaCl in 20 mM Tris–HCl, pH 7.5, 4 M urea, and 0.2 mM TCEP. The sample was then diluted using 20 mM Tris–HCl, pH 7.5, 4 M urea, and 1 mM TCEP before loading onto a 7.5-mL Source 15S column (Cytiva). The sample was eluted with a gradient from 0 to 2 M NaCl in 20 mM Tris–HCl, pH 7.5, 4 M urea, and 0.2 mM TCEP. The protein was then phosphorylated by subunit alpha of casein kinase II (CK2α) in a kinase: SLBP ratio of 1:10. Phosphorylation was performed over 24 hours in 50 mM Tris–HCl, pH 7.5, 10 mM MgCl₂, 0.2 mM TCEP, 250 mM Arg/Glu, and 1 mM ATP. Phosphorylation continued in an identical buffer with 200 mM NaCl at 25°C for another 24 hours. The sample was diluted using 20 mM Tris–HCl, pH 8.0, 0.2 mM TCEP and loaded onto a Mono Q 10/100 GL column (GE Healthcare). The sample was eluted with a gradient from 0 to 2 M NaCl in 20 mM Tris–HCl, pH 8.0, and 0.2 mM TCEP. The fractions that contained fully phosphorylated SLBP were pooled, and its phosphorylation status was confirmed using mass spectrometry on a Thermo Fisher Orbitrap Velos Pro. Mass spectrometry suggested that in total five phosphate groups were added onto SLBP. The sample was polished by a HiLoad 16/600 Superdex 75 pg column (Cytiva) in 200 mM Arg/Glu, 40 mM NaH₂PO₄, pH 6.0. Purities of proteins were confirmed using SDS-PAGE gel.

4.2.2 | DRBM1

The gene encoding DRBM1 (residues 1732–1811) of the DCL1 (UniProt: Q9SP32) was synthesized by GenScript and cloned into pSMT3. The 3R mutant has the following mutations, E1747R, E1750R, and E1763R. The 9R mutant has the following mutations T1759R, T1761R, D1767R, V1769R, V1771R, and V1773R, in addition to the 3R mutations. To alleviate protein degradation, linker

residues 1806–1811 were deleted for the 9R constructs. DRBM1 mutants were created by mutagenesis PCR. DRBM1 proteins were expressed at 22°C overnight in *E. coli* strain BL21-CodonPlus (DE3) using 0.5 mM IPTG, which was added when the OD₆₀₀ reached 0.6 AU (0.8 AU for ¹⁵N-labeled samples). Mutants were expressed in the same manner. Once pelleted, the cells were re-suspended in 40 mL lysis buffer (25 mM Tris–HCl, pH 8.5, 1 M NaCl, 25 mM imidazole, 1 mM PMSF, 0.5 mg/mL lysozyme, 1 protease inhibitor tablet, and 0.2 mM TCEP), and subjected to three freeze–thaw cycles. The cells were lysed by sonication and centrifuged at 23,710 RCF at 4°C for 45 min. The supernatant was applied to 5 mL of Ni Sepharose™ excel resin (GE Healthcare) and washed with 200 mL of loading buffer (25 mM Tris–HCl, pH 8.5, 1 M NaCl, 25 mM imidazole, and 0.2 mM TCEP). The SUMO tag was cleaved by adding 10 mL of loading buffer with 0.1 mg of ULP1 and the column was placed on a shaker overnight at 4°C. The sample was then diluted using 25 mM MES–NaOH, pH 6.0, and 0.2 mM TCEP and loaded onto a 7.5-mL Source 15S column (Cytiva). The sample was eluted with a gradient from 0 to 2 M NaCl in 20 mM MES–NaOH pH 7.5, and 0.2 mM TCEP. The sample was then diluted using 20 mM Tris–HCl, pH 8.0, 0.2 mM TCEP and loaded onto a Mono Q 10/100GL column (Cytiva). The sample was eluted with a gradient from 0 to 2 M NaCl in 20 mM Tris–HCl, pH 8.0, and 0.2 mM TCEP. The sample was finally loaded onto a HiLoad 16/600 Superdex 75 pg column (Cytiva) in 20 mM MES–NaOH, pH 6.5, 100 mM NaCl, and 0.2 mM TCEP. The purities of were then confirmed using SDS-PAGE gel.

4.3 | Fluorescence polarization assays

Fluorescence polarization assays were carried out using 10 nM 5' fluorescein-labeled RNA mixed with DRBM1 constructs at concentrations ranging from 300,000 to 0.488 nM by 2-fold serial dilutions in 20 mM HEPES–NaOH, pH 7.0, 0.02% Tween, 0.2 mM TCEP, and 50 mM NaCl. A 100 µL of samples were mixed in black flat-bottom 96-well plates (Costar) by shaking at 220 rpm for 5 min, followed by incubation at 37°C for 30 min, and incubation at 25°C for 20 min. The RNA sequence was the stem-loop form of pri-miR172 RNA (5'GGGCUGC UGUGGCAUUUCGAUGCUGC AUCGGC-3').

The fluorescence polarization data were gathered at 22°C using a BioTek synergy 2 plate reader with an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The binding affinities were determined using non-linear regression for one-site interaction using

GraphPad Prism 7. The fluorescence polarization anisotropy F_p was fitted using the quadratic equation below, where the fitting parameters $F_{\min}$, $F_{\max}$, and K_D are the fluorescence polarization anisotropy baseline, plateau, and dissociation constant, respectively. $[P_T]$ is the total protein concentration and $[L_T]$ is the total RNA concentration (10 nM). Errors of K_D were calculated based on three independent measurements.

$$F_p = F_{\min} + (F_{\max} - F_{\min})$$

$$\left\{ \frac{([P_T] + [L_T] + K_D) - \{([P_T] + [L_T] + K_D)^2 - 4[P_T][L_T]\}^{0.5}}{2[L_T]} \right\}$$

4.4 | NMR experiments

4.4.1 | Chemical exchange saturation transfer NMR

Stem-loop binding proteins were exchanged into 40 mM NaH_2PO_4 , pH 6.0, 200 mM Arg/Glu for NMR measurements. A 15, 25, and 45-Hz saturation pulse (B_1) was applied from 102 to 133 ppm at an interval of 0.5 ppm. Protein concentrations for phosphorylated SLBP, non-phosphorylated SLBP, and noENC were 170, 280, and 750 μM , respectively. DRBM1 constructs were exchanged into 20 mM MES-NaOH, pH 6.5, 50 mM NaCl, 1 mM TCEP, and 10% D_2O for NMR measurements. Protein concentrations for wild-type, ENC, and 9R constructs were 715, 1100, and 730 μM , respectively. A saturation pulse (B_1) with strength of 15, 43, or 45 Hz was applied from 103 ppm to 134 at an interval of 0.5 ppm. CEST experiments were collected at 298 K on a Bruker Avance III-HD 600 and 850 MHz spectrometer installed with a cryo-probe. CEST data were analyzed using ChemEx (Vallurupalli et al., 2012). ChemEx performs a numerical fitting by minimizing predefined Chi-squared (χ^2) value. We assumed a two-state model during the data fitting. The fitting procedure minimizes the difference between experimental and theoretical peak intensities by optimizing relaxation rates, chemical exchange rates, and populations of conformations. All values were unrestrained except for the peak positions, which were held constant and read from the HSQC. The fitting quality was tested by the Kolmogorov–Smirnov p -value that is calculated by ChemEx, and all our data fitting passed the p -value test (Vallurupalli et al., 2012). Backbone assignment of DRBM1 was obtained from BMRB (ID: 19104 and 19105). The residues not included in CEST analysis due to missing resonance, overlapping resonance, or lack of assignments can be found in Table S3.

4.4.2 | Chemical shift perturbation analysis

The phosphorylated SLBP and noENC proteins were exchanged into 10 mM MES-NaOH, pH 6.0, 100 mM NaCl, 10 μM EDTA, 1 unit RNase inhibitor, 0.22 mM TCEP, and 10% D_2O for NMR measurements. Samples were then mixed with stem-loop RNA (5'-AAAGGCC-CUUUUCAGGGCCA-3') with a ratio of 1:1.5. Protein concentrations for phosphorylated SLBP and the noENC constructs were 380 and 560 μM , respectively. HSQC experiments were collected at 298 K on a Bruker Avance III-HD 600 MHz spectrometer installed with a cryo-probe. The NMR data were processed using NMRPipe and analyzed using NMRViewJ (Delaglio et al., 1995; Johnson & Blevins, 1994). Triple resonance experiments HNCA, HNCACB, HN(CO)CA, CBCA(CO)NH, and HNCO were used to assign backbones of apo and RNA-bound SLBP (BMRB ID 25364 and 25365). The backbone assignment of the noENC construct was performed using the same set of experiments, and the assignment of the RNA-binding domain is almost identical to the wild-type protein, except mutated residues and their neighboring residues. The residues not assigned in our NMR analysis can be found in Table S3.

4.4.3 | R_1 , R_2 , and ^{15}N - ^1H hetNOE NMR analysis

The ^{15}N longitudinal and transverse relaxation rates (R_1 and R_2) together with ^{15}N - ^1H hetNOEs for backbone ^{15}N atoms were recorded for phosphorylated and noENC SLBP constructs bound to RNA with a ratio protein:RNA of 1–1.5. Protein concentrations for phosphorylated SLBP and the noENC constructs were 380 and 560 μM , respectively. Samples were exchanged into 10 mM MES-NaOH, pH 6.0, 100 mM NaCl, 10 μM EDTA, 1 unit RNase inhibitor, 0.22 mM TCEP, and 10% D_2O . For R_1 measurements, intensities at 20, 60, 100, 200, 400, 600, 800, and 1200 ms were measured. The inter-scan delay was set to 2 s. For R_2 measurements, intensities were gathered at 17, 34, 51, 68, 102, 136, 204, and 271 ms and a recycle delay of 2.0 s was used. These rates were fitted using a single exponential decay as below:

$$I(t) = I_0 * e^{-t * R_{1,2}}$$

Where $I(t)$ is the intensity at time point t , I_0 is the signal intensity at $t = 0$, and $R_{1,2}$ is the longitudinal and transverse relaxation rates, respectively. The error in R_1 and R_2 measurements was estimated from data fitting. ^{15}N - ^1H hetNOEs were obtained from the signal intensity ratio acquired with and without proton saturation using a 2.5 s

recycling delay. Experiments were collected at 298 K on a Bruker Avance III-HD 850 MHz spectrometer installed with a cryo-probe. The error of hetNOEs difference between the noENC and the Pi constructs was estimated

by $\sqrt{\left(\frac{\sigma}{I}\right)_{\text{noENC, NOE}}^2 + \left(\frac{\sigma}{I}\right)_{\text{noENC, ref}}^2 + \left(\frac{\sigma}{I}\right)_{\text{Pi, NOE}}^2 + \left(\frac{\sigma}{I}\right)_{\text{Pi, ref}}^2}$,

where σ is the spectral noise level, I is the resonance intensities, NOE and ref are spectra with the proton saturation pulse on and off, respectively. The NMR data were processed using NMRPipe and analyzed using NMRViewJ (Delaglio et al., 1995; Johnson & Blevins, 1994).

4.5 | Molecular dynamics simulations

The crystal structures of RNA-bound SLBP (PDB ID: 4TUW) was used as a starting structure to simulate the interaction between the phosphorylated tail and the SLBP: RNA complex. The phosphorylated tail and missing residues were built using Coot (Emsley & Cowtan, 2004). We then used distance restraints inferred from our CSP data to simulate the structure of RNA-bound SLBP using Xplor-NIH (Schwieters et al., 2003; Schwieters et al., 2006). The energy minimized model was used as the starting structure for MD simulations.

The Amber20 package with AMBER force fields ff14SB, phosaa14SB, and RNA.OL3 was used for MD simulations (Case et al., 2005). We analyzed RNA-bound SLBP in three states: phosphorylated, non-phosphorylated, and noENC. TIP3P water molecules were used as solvent with 0 mM NaCl. The charge of the protein: RNA complexes was neutralized with the relevant amount of Na^+ . The sizes of the simulated systems were approximately $20 \text{ nm} \times 20 \text{ nm} \times 20 \text{ nm}$. The MD simulation protocol (Yang & Song, 2016) was identical for all simulated systems including: (Uversky et al., 2000) steepest descent energy minimization of the solvent water with restrains on the protein and ions; (Uversky et al., 2005) 20 ps constant number-pressure-temperature (NPT) MD simulation at 50 K and 1 bar to equilibrate solvent water with restrains on the protein and ions; (Pentony & Jones, 2010) heating up the system to 300 K via a series of 10-ps constant number-volume-temperature (NVT) MD simulations at 50, 100, 150, 200, 250, and 300 K; (Kriwacki et al., 1996) 300-ns production NPT MD simulations at 300 K and 1 bar. In the production MD simulations, a 2-fs time step was used with SHAKE constraints on all bonds involving hydrogen. Long-range electrostatic interactions were calculated with the particle-mesh Ewald method. The cut-off for the Lennard-Jones potential was set at 1.0 nm. The root mean square deviations (RMSDs) of the protein complexes were monitored over time in each simulation to determine if the system has reached equilibrium.

The DBSCAN (density based spatial clustering of applications with noise) program from Amber20 was used in the clustering analysis of the equilibrated SLBP MD trajectories (Ester et al., 1996). DBSCAN considers points to be as part of a single cluster if there are at least some other points within a neighborhood range (ϵ). The DBSCAN algorithm then generates a structure for the population of each cluster known as a mediod. The mediod structure in the largest cluster was then used as the representative conformation. The root mean square fluctuations (RMSF) of the SLBP constructs were calculated for each residues and averaged over the course of the equilibrated MD trajectories. The distances between each residue in the SLBP constructs were measured throughout the course of the equilibrated MD trajectories. The average of these values were then calculated and used to produce the density matrices. Hydrogen bond formation was calculated between the ENC (residues 268–276) and the loop region (residues 218–237) of the SLBP constructs. A hydrogen bond was assigned to residues when the distance between the donor and acceptor were $<4 \text{ \AA}$ and the angle of the donor-hydrogen-acceptor was less than 30° . The occupancy of each hydrogen bond was calculated by the percentage of time that the hydrogen bond existed over the course of the equilibrated MD simulations. PCA was performed using the CPPTRAJ scripts of Amber20. PCA visualizations were presented using VMD 1.9.3 (Humphrey et al., 1996).

AUTHOR CONTRIBUTIONS

Steve Zaharias: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); visualization (equal); writing – review and editing (equal). **Talia Fargason:** Data curation (equal); formal analysis (equal); investigation (equal). **Rory Greer:** Data curation (supporting); investigation (supporting); methodology (equal); validation (equal). **Yuhua Song:** Data curation (equal); formal analysis (equal); investigation (equal); validation (equal); writing – review and editing (equal). **Jun Zhang:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (lead); investigation (equal); methodology (equal); project administration (lead); resources (lead); supervision (lead); visualization (equal); writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGMENTS

We want to thank Dr. Ronald Shin, the manager of UAB Central Alabama High-Field NMR facility, Dr. Charles Schwieters for his help with Xplor-NIH, and Dr. Guillaume Bouvignies for his assistance with ChemEx. This work was supported by the United States National Science Foundation (MCB 2024964 to Jun Zhang).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Yuhua Song  <https://orcid.org/0000-0002-2583-850X>

Jun Zhang  <https://orcid.org/0000-0002-5842-7424>

REFERENCES

- Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*. 2015;519:106–9.
- Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, et al. The amber biomolecular simulation programs. *J Comput Chem*. 2005;26:1668–88.
- Consortium TU. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2020;49:D480–9.
- Csizmok V, Forman-Kay JD. Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications. *Curr Opin Struct Biol*. 2018;48:58–67.
- David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol Biol*. 2014;1084:193–226.
- Davis KB, Zhang Z, Karpova EA, Zhang J. Application of tyrosine-tryptophan fluorescence resonance energy transfer in monitoring protein size changes. *Anal Biochem*. 2018;557:142–50.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995;6:277–93.
- Dominski Z, Yang XC, Raska CS, Santiago C, Borchers CH, Duronio RJ, et al. 3' end processing of *Drosophila melanogaster* histone pre-mRNAs: requirement for phosphorylated drosophila stem-loop binding protein and coevolution of the histone pre-mRNA processing system. *Mol Cell Biol*. 2002;22:6648–60.
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427–d432.
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004;60:2126–32.
- Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the second international conference on knowledge discovery and data mining*. Portland, OR: AAAI Press; 1996.
- Haider S, Parkinson GN, Neidle S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys J*. 2008;95:296–311.
- Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(33–38):27–38.
- Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*. 2002;323:573–84.
- Johnson BA, Blevins RA. NMR view: a computer program for the visualization and analysis of NMR data. *J Biomol NMR*. 1994;4:603–14.
- Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA*. 1996;93:11504–9.
- Kyte J, Doolittle RF. A simple method for displaying the hydrophobic character of a protein. *J Mol Biol*. 1982;157:105–32.
- Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res*. 2018;46:D493–d496.
- Marzluff WF, Duronio RJ. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr Opin Cell Biol*. 2002;14:692–9.
- Muller-Spath S, Soranno A, Hirschfeld V, Hofmann H, Ruegger S, Reymond L, et al. From the cover: charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci USA*. 2010;107:14609–14.
- Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, et al. D²P²: database of disordered protein predictions. *Nucleic Acids Res*. 2013;41:D508–16.
- Pentony MM, Jones DT. Modularity of intrinsic disorder in the human proteome. *Proteins*. 2010;78:212–21.
- Santiago-Frangos A, Jeliakov JR, Gray JJ, Woodson SA. Acidic C-terminal domains autoregulate the RNA chaperone Hfq. *Elife*. 2017;6.
- Schwieters C, Kuszewski J, Clore GM. Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc*. 2006;48:47–62.
- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson*. 2003;160:65–73.
- Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci*. 2000;97:8868–73.
- Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res*. 2013;41:D344–7.
- Suarez IP, Burdisso P, Benoit MP, Boisbouvier J, Rasia RM. Induced folding in RNA recognition by *Arabidopsis thaliana* DCL1. *Nucleic Acids Res*. 2015;43:6607–19.
- Tan D, Marzluff WF, Dominski Z, Tong L. Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex. *Science*. 2013;339:318–21.
- Tzul FO, Vasilchuk D, Makhatadze GI. Evidence for the principle of minimal frustration in the evolution of protein folding landscapes. *Proc Natl Acad Sci USA*. 2017;114:E1627–32.
- UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2018;46:2699.
- Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem*. 2002;269:2–12.
- Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins*. 2000;41:415–27.
- Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit*. 2005;18:343–84.
- Vacic V, Iakoucheva LM. Disease mutations in disordered regions—exception to the rule? *Mol Biosyst*. 2012;8:27–32.
- Vallurupalli P, Bouvignies G, Kay LE. Studying “invisible” excited protein states in slow exchange with a major state conformation. *J Am Chem Soc*. 2012;134:8148–61.
- Wang C, Uversky VN, Kurgan L. Disordered nucleome: abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, bacteria and archaea. *Proteomics*. 2016;16:1486–98.

- Wei X, Ke H, Wen A, Gao B, Shi J, Feng Y. Structural basis of micro-RNA processing by dicer-like 1. *Nat Plants*. 2021;7:1389–96.
- Yang H, Song Y. Structural insight for roles of DR5 death domain mutations on oligomerization of DR5 death domain–FADD complex in the death-inducing signaling complex formation: a computational study. *J Mol Model*. 2016;22:89.
- Zaharias S, Zhang Z, Davis K, Fargason T, Cashman D, Yu T, et al. Intrinsically disordered electronegative clusters improve stability and binding specificity of RNA-binding proteins. *J Biol Chem*. 2021;297:100945.
- Zhang J, Tan D, DeRose EF, Perera L, Dominski Z, Marzluff WF, et al. Molecular mechanisms for the regulation of histone mRNA stem-loop-binding protein by phosphorylation. *Proc Natl Acad Sci USA*. 2014;111:E2937–46.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zaharias S, Fargason T, Greer R, Song Y, Zhang J. Electronegative clusters modulate folding status and RNA binding of unstructured RNA-binding proteins. *Protein Science*. 2023;32(5):e4643. <https://doi.org/10.1002/pro.4643>