ACS OMEGA

Open Access

This article is licensed under CC-BY-NC-ND 4.0

http://pubs.acs.org/journal/acsodf

Article

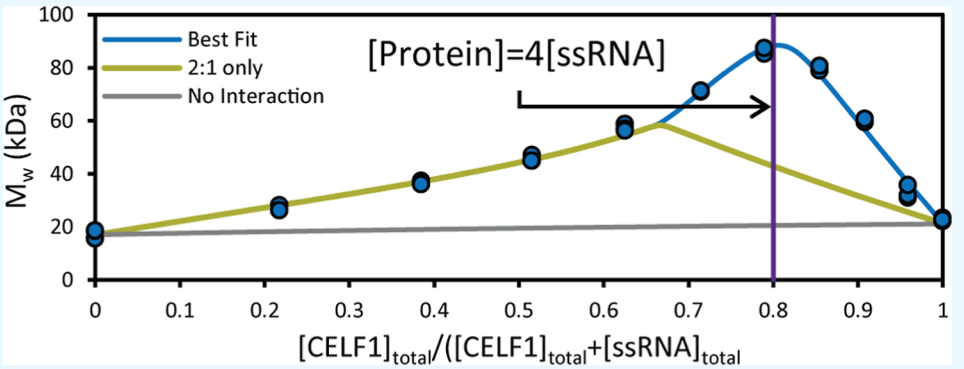
Multivalent Protein–Nucleic Acid Interactions Probed by Composition-Gradient Multiangle Light Scattering

Josephine C. Ferreon,* Natee Kongchan, Phoebe S. Tsoi, Kyoung-Jae Choi, Sophia Kenrick,* Joel Neilson,* and Allan Chris M. Ferreon*

Cite This: *ACS Omega* 2024, 9, 41003–41010

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Many RNA-binding proteins, such as TDP-43 or CELF1, interact multivalently with nucleic acid repetitive elements. The molecular stoichiometry of protein to nucleic acid is often difficult to assess, particularly by standard electrophoretic mobility shift assays (EMSAs). Here, we investigate the use of composition-gradient multiangle light scattering (CG-MALS) for quantifying binding affinity and stoichiometry for two RNA-binding proteins with their nucleic acid partners of varied sequence and length: TDP43's N-terminal RNA recognition motifs with both TG/GU-repeat ssDNA and ssRNA, respectively, and CELF1's two N-terminal RNA recognition motifs with (TG/UGUU/GU) repeats and an experimentally defined cognate GU-rich element (GRE). Our CG-MALS data derived from each of these interactions is consistent with expected ranges of binding affinity and stoichiometry for proteins binding to shorter nucleic acid repeats. Furthermore, we conclude that CG-MALS can be an excellent method for obtaining quantitative estimates even for high (>2) protein–nucleic acid stoichiometric ratios.

INTRODUCTION

Multivalent interactions occur universally in biological systems. Many homotypic and heterotypic protein–protein interactions and protein–nucleic acid interactions utilize multivalency to facilitate cooperativity and achieve cellular signaling (e.g., refs 1–7). For example, many RNA-binding proteins (RBPs) such as TDP-43 assemble in multiples upon UG-rich elements of various length, and alterations in the stoichiometric ratios between the protein and its RNA target may have significant consequences regarding normal cellular function.[4,8] However, current methods still present challenges in providing robust quantitative estimates of the molecular stoichiometry. Electrophoretic mobility shift assays (EMSAs), a widely used technique in protein–nucleic acid studies, require optimization of gel conditions to generate distinct bands. Even with optimized conditions, the number of bands is not formally indicative of the correct stoichiometry. Other biophysical techniques, such as fluorescence anisotropy, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), or biolayer interferometry (BLI), are indirect methods (i.e., based on changes in fluorescence, enthalpy heat, or "response units" with complex formation). They are generally best suited for quantifying interactions with a 1:1 stoichiometry and can suffer from artifacts for more complex interactions. For example, ITC may not readily distinguish between 1:1 and 2:2 stoichiometric ratios of molecular complexes. In contrast, multiangle static light scattering (MALS) measurements provide a label-free method of directly accessing the molecule's molar mass, thus directly measuring the essence of complex formation.[9,10]

Composition-gradient multiangle light scattering (CG-MALS) leverages the fact that changes in molar mass can be directly related to the reversible formation and dissociation of

Received: July 9, 2024
Revised: September 13, 2024
Accepted: September 18, 2024
Published: September 21, 2024

© 2024 The Authors. Published by American Chemical Society

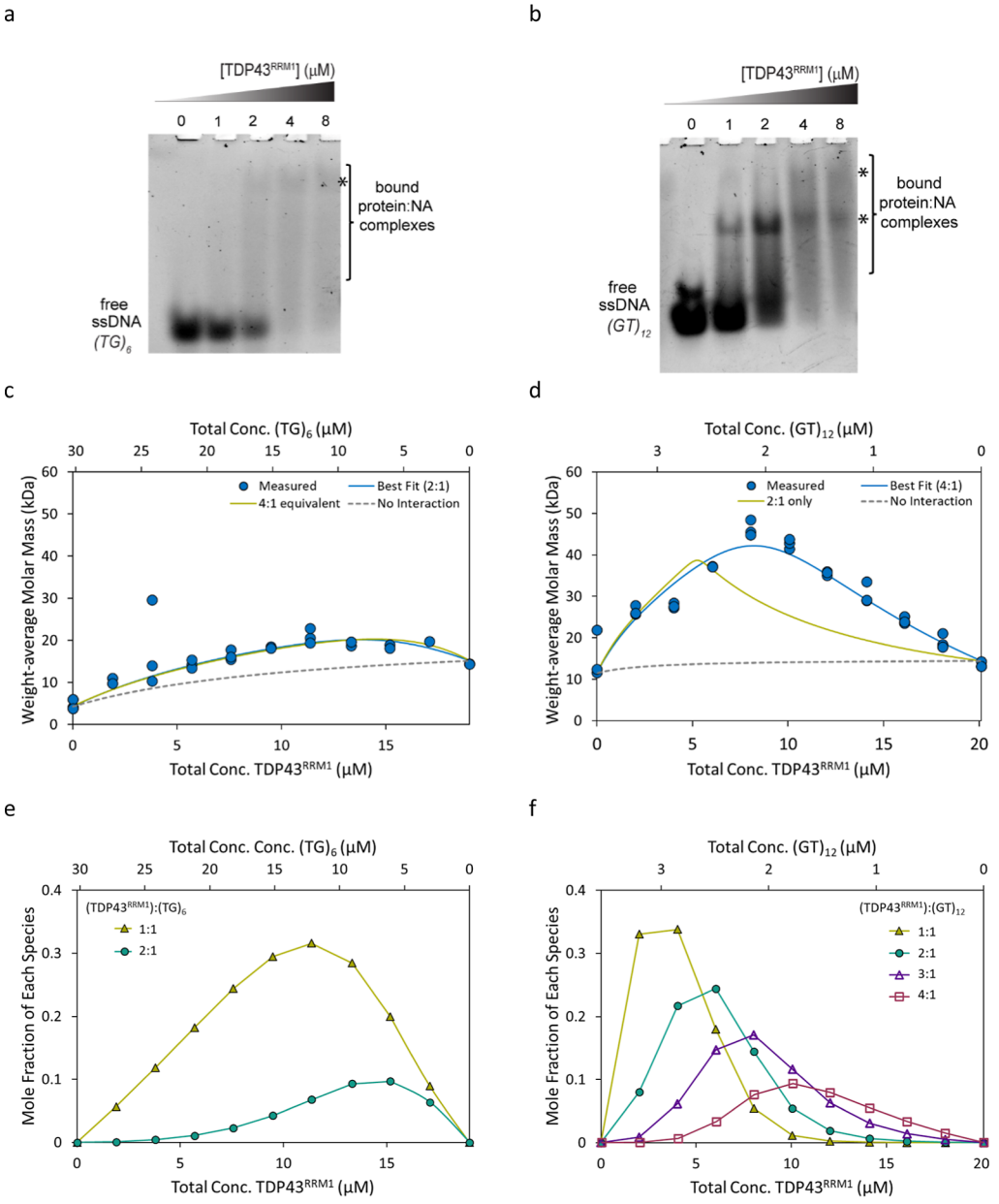


Figure 1. TDP43^{RRM1} interactions with $(TG)_6$ ssDNA and $(GT)_{12}$ ssDNA by EMSA (a,b) and CG-MALS (c–f). (a,b) EMSA of TDP43^{RRM1} and $(TG)_6$ ssDNA (a) and $(GT)_{12}$ ssDNA (b) showing disappearance of free ssDNA accompanied by band smearing (bracket) and less distinct bands representing complex formation (denoted by *). (c,d) Measured weight-average molar mass (M_w) from three independent CG-MALS experiments as a function of composition. The data was fit assuming 2 or 4 equiv sites on each ssDNA for the TDP43^{RRM1} protein. Measured molar mass data for all three experiments (blue circles) are shown along with the best fit (blue line), alternative nonfitting model (green line), and "no interaction" reference curve (dashed gray). The best fit for the $(TG)_6$ interaction includes two equivalent binding sites with K_d = 10.4 ± 3.2 μM; the best fit for the $(GT)_{12}$ interaction includes four equivalent binding sites with K_d = 1.34 ± 0.12 μM (average and standard deviation from three replicates; Table 1). (e,f) Concentration of each complex formed, as determined by the best fit of the CG-MALS data. The fraction of unbound protein and ssDNA has been left off for clarity.

specific complexes, each with its own equilibrium association constant.[10] In a typical experiment, each binding partner is prepared in solution at concentrations near the expected equilibrium constant of the interaction. Solutions of varying compositions are prepared automatically, injected into a multiangle light scattering (MALS) detector, and allowed to come to equilibrium inside the MALS flow cell for a user-specified amount of time. The weight-average molar mass (M_w) of the solution is a first-principles calculation based on the light scattering intensity and total concentration of the solution—either specified manually or measured using an inline concentration detector.[11] Association of one species with itself or with a binding partner results in a concentration-dependent increase in M_w; thus, multivalent complex formation is readily evident. The measured M_w as a function of composition is fitted to the appropriate expressions describing the equilibrium association of the analytes, which can include both self-interaction (e.g., dimerization) and heteroassociation. Like AUC or ITC, all interactions occur in solution without the need for immobilization or labeling. CG-MALS can be used to measure interactions with the equilibrium dissociation constant (K_d) from ∼100 pM[12] to

https://doi.org/10.1021/acsomega.4c06358
ACS Omega 2024, 9, 41003−41010

Table 1. Protein:Nucleic Acid (NA) Stoichiometry (n), Equilibrium Dissociation Constant (K_d), and Monomer Molecular Weight (MW) Determined by Fitting CG-MALS Data

interaction	nucleic acid	measured monomer MW (kDa)		$n_{protein}:n_{NA}$	K_d (μM)
		protein	NA		
TDP43^{RRM1}:$(TG)_6$	ssDNA	15.2 ± 0.5	4.51 ± 0.65	2:1[a]	10.4 ± 3.2
TDP43^{RRM1}:$(GT)_{12}$	ssDNA	14.3 ± 0.1	11.0 ± 0.4	4:1[a]	1.34 ± 0.12
CELF1^{RRM1-2}:$(UGUU)_4$	ssRNA	21.2 ± 0.3	4.81 ± 0.26	1:1[b]	0.02
				2:1[b]	4
CELF1^{RRM1-2}:$(GU)_{22}$	ssRNA	21.2 ± 0.4	16.8 ± 0.3	2:1[c]	0.0010 ± 0.0002
				4:1[d]	0.39 ± 0.03

All values are the average and standard deviation of three replicates, except for CELF1^{RRM1-2}:$(UGUU)_4$, which reports a single fit value across data from three experiments. [a]All binding sites are equivalent with indicated K_d. [b]Best fit of concatenated data is consistent with two binding sites—one strong and one weak. [c]Binding at first two sites is equivalent with indicated K_d. [d]Binding at the third and fourth sites is assumed equivalent with indicated K_d.

several mM[13,14] and has been used to quantify a wide variety of biomolecular interactions, including the self-association of insulin,[15,16] antibody−antigen interactions,[17] and multivalent interactions with complex stoichiometries.[12,18,19] The measured stoichiometry and affinity by CG-MALS agree well with other techniques, including SPR,[20] BLI,[17] AUC,[12,21,22] ITC,[23,24] and CryoEM.[17,19] Although often applied to measuring interactions under dilute solution conditions in the absence of molecular crowding or phase separations, as in this study, CG-MALS can also be applied to high-concentration measurements to understand what drives phase separation and other thermodynamic processes.[25,26] The automated CG-MALS measurements presented in this study can be applied to interactions among one or two binding partners and can accommodate any combination of self-association and heteroassociation.

To explore the utilization of CG-MALS for the analysis of protein−nucleic acid multivalent interactions, we characterized interactions between two proteins and their nucleic acid partners: TDP43^{RRM1} with $(TG)_6$ and $(GT)_{12}$ ssDNA and CELF1 with GU-rich elements (GRE) made up of $(UGUU)_4$ and $(GU)_{22}$ repeats. For each interaction, CG-MALS data were collected in a three-part experiment, as shown in Figure S1: (1) measurement of the protein alone at 3−5 concentrations, (2) measurement of 8−11 mixtures of protein and oligonucleotide at varying concentrations and stoichiometric ratios to assess, and (3) measurement of the oligonucleotide at 3−5 concentrations. Three technical replicates were performed for each interaction. The change in M_w as a function of composition was then fit to an appropriate equilibrium association model to determine the single set of stoichiometries and corresponding equilibrium association constants (K_a) that fit the data.

We show that our CG-MALS data is consistent with some of the quantitative binding measurements and observations reported elsewhere.[8,27,28] Furthermore, even without detailed data analysis and model fitting, CG-MALS can provide minimum estimates of stoichiometry based on the maximum measured weight-average molar mass ($M_{w,max}$) and the molar composition of the protein and nucleic acid components where $M_{w,max}$ occurs. Complete fitting of CG-MALS data can provide not only quantitation of the complexes formed but also the equilibrium association constant for the formation of each complex and the individual affinity at each binding site (equilibrium dissociation constant, K_d).

RESULTS AND DISCUSSION

To evaluate CG-MALS as a method for characterizing multivalent interactions, we chose protein−nucleic acid pairs available to us. To avoid aggregation, we focused on truncated constructs: the first RRM domain of TDP43 (TDP43^{RRM1}) and the first two RRM domains of CELF1 (CELF1^{RRM1-2}). In all cases, at least two proteins are bound to each nucleic acid ligand. For two of the pairs, higher-order binding with four proteins per oligonucleotide was also present. The description and analysis for each of the protein−nucleic acid pair are detailed below.

TDP43^{RRM1} Binding $(TG)_6$ or $(GT)_{12}$ ssDNA. TDP43 is an RNA-binding protein, consisting of 2 RRM domains that recognize long GU repeats as well as single-strand DNA (ssDNA) TG repeats.[29,30] For this study, we focus on TDP43^{RRM1} and assess its interaction with multivalent nucleic acid repeats $(TG)_6$ and $(GT)_{12}$. Both DNA sequences have been previously reported as TDP43 cognate sequences. TG repeat elements were found near the human *CFTR* exon 9 gene[29] where TDP-43 can bind and regulate proper splicing activities to avoid exon skipping and ensure correct CFTR protein translation.[31] In other ssDNA studies, TDP43 was demonstrated to form higher-order assemblies (up to four TDP43 $^{RRM1-2}$) with 24 GT repeats.[8] We set out to test how many domains might bind $(TG)_6$ and $(GT)_{12}$ sequences and with what affinity, i.e., if two molecules of TDP43^{RRM1} could bind to $(TG)_6$, we expected four molecules to bind to $(GT)_{12}$ DNA. EMSA provides qualitative evidence that TDP43 binds to both $(TG)_6$ and $(GT)_{12}$. For $(TG)_6$, we observe the disappearance of the free ssDNA and the consequent smearing of the lanes with faint bands corresponding to bound states (Figure 1a). Similarly, for $(GT)_{12}$, we observed binding and the appearance of a major distinct band and a minor more shifted band (Figure 1b). The more pronounced bands for $(GT)_{12}$ suggest stronger binding and a tighter affinity. However, it is difficult to assess stoichiometry from these results. In contrast, the stoichiometry of the interactions is obvious when they are measured by CG-MALS (Figure 1c,d).

Triplicate experiments were performed to investigate each interaction, each one consisting of three gradients, as described in the Materials and Methods section. Each experiment was fit separately, and we present the average and standard deviations of all measured and fit parameters. The measured M_w for each species (Table 1) was in good agreement with the expected monomer molar mass (Table S1). Concentration gradient data for the protein and nucleic acid species alone (i.e., in the absence of the binding partner) revealed no measurable change

https://doi.org/10.1021/acsomega.4c06358
ACS Omega 2024, 9, 41003−41010

in M_w for concentrations up to 20 μM for TDP43^{RRM1}, 30 μM for $(TG)_6$, and 3.5 μM for $(GT)_{12}$, confirming the lack of reversible self-association under the conditions tested. Although the measured M_w was slightly greater than the expected value, it was invariant with the concentration, indicating the presence of a small fraction of irreversible aggregates. Taken together, these results indicate that any change in M_w upon mixing protein and oligonucleotide resulted from the specific, reversible heteroassociation of the two species, and only these complexes were included when fitting the CG-MALS data.

The heteroassociation gradients, which created 11 different combinations of protein and ssDNA, resulted in an increase in overall weight-average molar mass of the solution and indicated more than one protein bound to each ssDNA (Figure 1c,d). For $(TG)_6$, the maximum measured weight-average molar mass (M_w) occurs when TDP43^{RRM1} is mixed with $(TG)_6$ at a 2:1 molar ratio, with [TDP43^{RRM1}] ~ 14 μM and [$(GT)_6$] ~ 7 μM (Figure 1c), suggesting a 2:1 stoichiometry. The maximum M_w of 21.0 ± 1.8 kDa was consistent with two proteins bound to each oligonucleotide (Figure 1c), and this measured value was 40% higher than the M_w that would result in the absence of an interaction upon mixing these analytes at the same concentration (Figure 1c, gray dashed line). The best fit considers two binding sites with an equivalent affinity. Fitting three experiments independently resulted in an average equilibrium dissociation constant at each binding site of K_d = 10.4 ± 3.2 μM (Table 1). The concentrations of the complexes formed as a function of composition are shown in Figure 1e. Similar studies (TDP43^{RRM1} with sequence $(GT)_6$) by ITC[8] reported tighter affinities with K_d ~ 100 nM for protein binding to the first site and K_d ~ 10 nM to the second site, suggesting cooperativity between sites. The discrepancies could be attributed to the differences in sequence and experimental conditions. The ITC experiments were performed with lower buffer and salt concentrations (15 mM phosphate buffer, pH 6.8, containing 25 mM KCl) versus our buffer conditions (25 mM HEPES, 50 mM NaCl, pH 7.3). Also, since the CG-MALS experiments were conducted at concentrations from 3 to 30 μM $(TG)_6$ and most of the protein-oligonucleotide mixtures were made in conditions of excess $(TG)_6$ (Figure 1c,e), higher affinity interactions may have been masked, and the effect of cooperativity may have been negligible.

For $(GT)_{12}$, the measured M_w suggests higher-order binding and higher affinity compared to the interaction with $(TG)_6$. Even at ~10 times lower overall concentrations of oligonucleotides (i.e., concentrations up to 3.5 μM $(GT)_{12}$ compared to concentrations up to 30 μM $(GT)_6$), a significant increase in M_w is observed, reflecting both the increase in affinity and stoichiometry. The maximum measured M_w of 46.3 ± 1.9 occurs where [TDP43^{RRM1}] = 4[$(GT)_{12}$] = 8.3 μM (Figure 1d) and is significantly higher than the maximum possible M_w that could be achieved if each $(GT)_{12}$ bound only two TDP43^{RRM1} at these concentrations (39 kDa), as shown by the green alternative nonfitting model in Figure 1d. Thus, the best fit is unambiguously consistent with four equivalent binding sites each with K_d = 1.34 ± 0.12 μM (Figure 1d, blue line), confirming a higher affinity for TDP43^{RRM1} as compared to $(TG)_6$. The concentrations of each complex formed are shown in Figure 1f. Since the light scattering intensity measured by MALS is proportional to the weight-average molar mass, even a small amount of high-molecular-weight species can significantly impact the data, giving high confidence to the observation of the 4:1 species even at relative concentrations of 10% mol/mol or less. The RSD <10% on the K_d across three technical replicates further increases confidence in the fit.

CELF1^{RRM1-2} Binding $(UGUU)_4$ or $(GU)_{22}$ ssRNA. CELF1 is also an RNA-binding protein, consisting of 3 RRM domains with the N-terminal RRMs (RRM1 and RRM2) separated from the C-terminal RRM3 via a long disordered linker.[32,33] For the current studies, we focused on CELF1^{RRM1-2} and its interactions with its defined GU-rich target elements (GREs). Similar to the investigation of TDP43-binding ssDNA, we observed significant differences in both the stoichiometry and affinity for CELF1^{RRM1-2} depending on the length of the ssRNA binding partner. Similarly, we observed no significant self-association of either CELF1^{RRM1-2} or its binding partners.

Initially, we investigated CELF1^{RRM1-2} binding to $(UGUU)_4$. As expected, EMSA showed disappearance of free ssRNA and appearance of smeary band(s) that migrate higher upon increasing protein concentration but without clear insight into the stoichiometry. Triplicate CG-MALS experiments were performed to assess the interaction. In all three experiments, the maximum M_w (~30 kDa) was achieved when the concentration of protein was approximately twice the concentration of ssRNA (overall mole fraction of CELF1^{RRM1-2} = [CELF1]/([CELF1]+[$(UGUU)_4$]) ~ 0.67), consistent with a 2:1 stoichiometry (Figure 2b). The best fit, considering all three experiments together, suggested two nonequivalent binding sites. CELF1^{RRM1-2} binds strongly at the first site with K_d ~ 0.02 μM and more weakly to the second site with K_d ~ 4 μM (Table 1). The results presented consider a single global fit to all three experiments. Significant experiment-to-experiment variation was observed, as seen in Figure 2b, likely due to the presence of aggregates in CELF1^{RRM1-2}. Although CG-MALS confirmed an overall 2:1 stoichiometry for this interaction, the current data were insufficient to conclude whether the two binding sites are equivalent or nonequivalent, as shown by the comparison of the blue best fit line and the green alternative fit line in Figure 2b. However, a binding model considering two nonequivalent binding sites is consistent with literature studies, suggesting that the secondary site could accommodate only 1 RRM domain, resulting in a weaker affinity,[28] and thus is presented as the best fit.

We then investigated CELF1^{RRM1-2} interactions with the longer RNA $(GU)_{22}$. With EMSA, we observed binding, which resulted in an upward shifting band. Importantly, the interaction appears to occur at lower concentrations of CELF1, compared with $(UGUU)_4$, suggesting a higher affinity interaction (Figure 3a). With CG-MALS, we can fully quantify this interaction. The maximum measured M_w (86.1 ± 1.2 kDa) suggests that we can see the formation of at least a 4:1 stoichiometric CELF1^{RRM1-2}:$(GU)_{22}$ complex (Figure 3b). The best fit included the formation of three species with the following protein:ssRNA stoichiometry: 1:1, 2:1, and 4:1 (Table 1). The maximum measured M_w is nearly 50% higher than the maximum possible M_w if only 2:1 binding is considered (Figure 3b, green line), clearly indicating that higher-order assemblies are reversibly formed under these conditions.

The 1:1 and 2:1 complexes appear to form with equivalent affinity at each binding site (K_d = 1.0 ± 0.2 nM), which is over 1000× stronger than the affinity for CELF^{RRM1-2} for

https://doi.org/10.1021/acsomega.4c06358
ACS Omega 2024, 9, 41003−41010

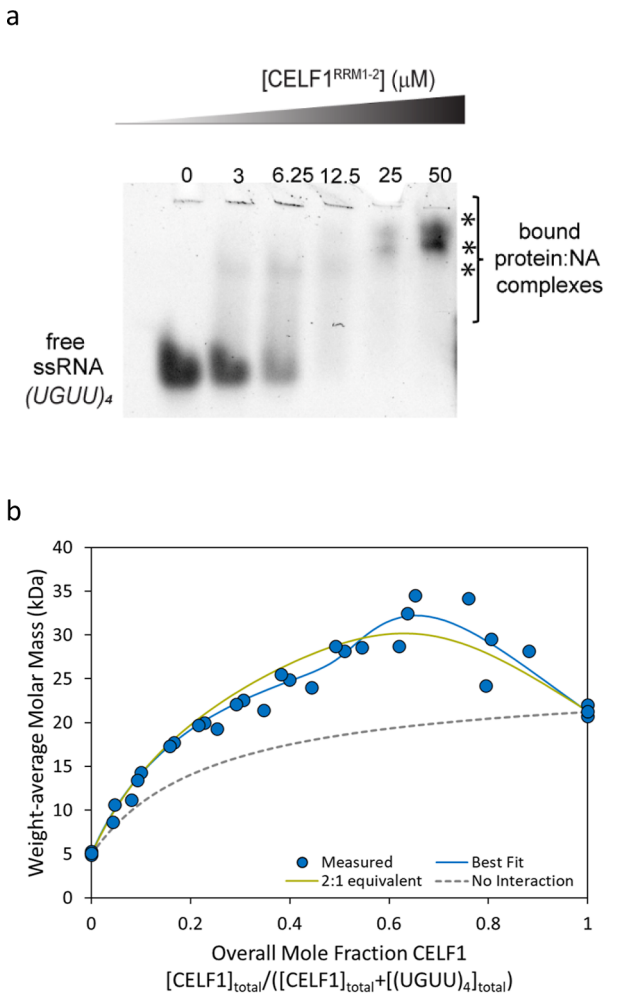


Figure 2. CELF1^{RRM1-2} and $(UGUU)_4$ ssRNA interactions by EMSA (a) and CG-MALS (b). (a) EMSA of CELF1^{RRM1-2} and $(UGUU)_4$ ssRNA showing disappearance of free ssRNA accompanied by shifted bands representing complex formation (denoted by *). (b) M_w measured by CG-MALS (blue circles) as a function of composition for three experiments, each with a maximum CELF^{RRM1-2} concentration of 7−8.5 μM and a maximum concentration of $(UGUU)_4$ of 9−20 μM. The best fit (blue line) considering two nonequivalent binding sites (Table 1) is shown by the blue line. An alternative binding model considering two equivalent binding sites, each with K_d = 3.6 μM, is indicated by the green line. The "no interaction" reference curve (gray dashed line) is shown for reference.

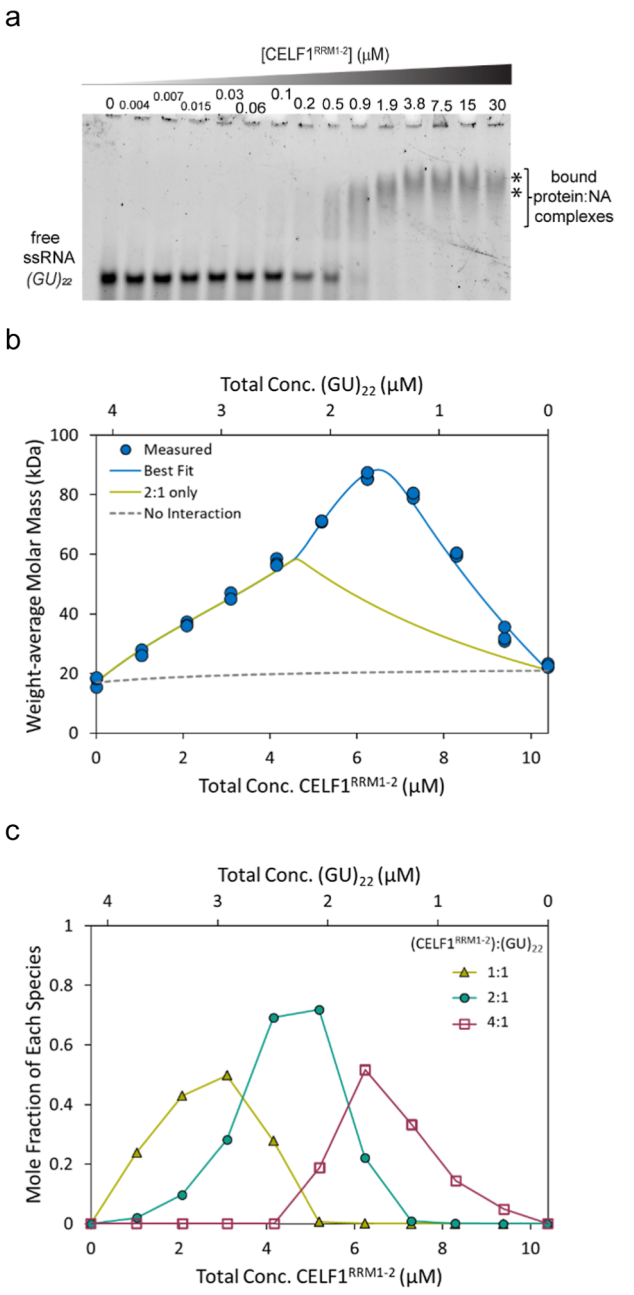


Figure 3. CELF1^{RRM1-2} and $(GU)_{22}$ ssRNA interactions by EMSA (a) and CG-MALS data and simulations (b,c). (a) EMSA of CELF1^{RRM1-2} and $(GU)_{22}$ ssRNA showing disappearance of free ssRNA accompanied by shifted bands representing complex formation (denoted by *). (b) Measured M_w vs total [CELF1^{RRM1-2}] and total $[(GU)_{22}]$ for triplicate CG-MALS experiments. The best fit (blue line) consists of complexes with 1:1, 2:1, and 4:1 stoichiometry. Including only 1:1 and 2:1 interaction (green line) is not sufficient to capture the measured molar mass. The "no interaction" reference curve (gray dashed line) is shown for reference. *c*, Calculated concentration of each complex formed across the CG-MALS heteroassociation gradient. The concentration of unbound monomer has been left off for clarity.

$(UGUU)_4$. This might reflect an increase in cooperative binding affinities as expected for multivalent interactions.[34] Under these conditions, a 4:1 complex also appears, and the third and fourth binding events were assumed to have equivalent affinity (K_d = 390 ± 30 nM). In general, we would also expect to observe the formation of the 3:1 complex; however, the model excluded the formation of this species, possibly due to the high concentration at which the measurements were performed (Figure 3c). It is possible that the 3:1 complex would also be evident at lower concentrations closer to K_d, where more binding sites would be unsaturated.

In conclusion, we show that CG-MALS is an excellent technique for obtaining estimates of binding affinities and molecular stoichiometry associated with multivalent interactions. Even when the binding interactions become significantly more complex and involve high stoichiometric ratios (>2), the $M_{w,max}$ can provide minimum estimates of molecular stoichiometric ratios. Better quantitative estimates can be obtained with CG-MALS data simulations and estimated K_d values from experimental studies for both simple stoichiometry (≤2) and complex interactions by combining data across

https://doi.org/10.1021/acsomega.4c06358
ACS Omega 2024, 9, 41003−41010

multiple concentration ranges. Thus, CG-MALS could be a useful technique in investigating the molecular stoichiometry involved in large multivalent protein−nucleic acid complexes. The multivalent interactions of these transcription factors and RNA-binding proteins with nucleic acid repeat sequences are required for their cellular functions. CELF1 is linked to myotonic dystrophy (DM) type 1 disease associated with the toxic effect of $(CUG)_n$ expansion.[33,35,36] TDP-43 aberrant aggregation with RNA has been linked to amyotrophic lateral sclerosis (ALS) disease.[4,37,38] There might be critical protein:RNA stoichiometry thresholds that may distinguish between physiological and pathological complexes and could benefit from the CG-MALS technique.

MATERIALS AND METHODS

Cloning, Protein Expression, and Purification. *$TDP43^{RRM1}$ (101−192).* The plasmid was prepared by site-directed mutagenesis of $TDP43^{RRM1-2}$ (101−277), containing both RRM domains, by changing the codon for amino acid V193 for the stop TAG codon. To construct the expression plasmid of $TDP\text{-}43^{RRM1-2}$ tagged with N-terminal His6, DNA fragments encoding TDP-43 (101−277) were amplified from the template plasmid (Addgene 27462) and then ligated into *Eco*RI/AvrII digested pET302NT-His (Invitrogen).

The $TDP43^{RRM1}$ plasmid was transformed into *Escherichia coli* BL21 star-competent cells and grown at 37 °C in Terrific Broth medium containing carbenicillin antibiotic. Cells were induced for expression with 1 mM IPTG when the OD_{600} reached between 0.8 and 1.0. Growth was allowed to continue overnight at 18 °C, and then cells were harvested by centrifugation at 4,000×g for 20 min. Cells were resuspended in lysis buffer (1.5 M KCl, 25 mM Tris, pH 7.5) containing 1 mM PMSF and the 1X Xpert protease inhibitor (GenDEPOT) and lysed by using a homogenizer (Avestin). The lysate was centrifuged at 4 °C for 1 h at 50,000×g. The supernatant was applied to nickel agarose beads (GoldBio) and washed extensively with lysis buffer containing 20 mM imidazole before elution with lysis buffer containing 200 mM imidazole. The elution was dialyzed to 150 mM NaCl, 25 mM Tris buffer pH 7.5, concentrated to a 2 mL volume, and purified using size exclusion chromatography (SEC) using a Superdex 75 column (Cytiva).

$CELF1^{RRM1-2}$. The His_{10}-$CELF1^{RRM1-2}$-expressing plasmid was transformed into *E. coli* BL21 star-competent cells and grown at 37 °C in LB medium. Cells were induced for expression with 0.4 mM IPTG when the OD_{600} reached between 0.8 and 1.0. Growth was allowed to continue overnight at 18 °C, and then cells were harvested by centrifugation at 4,000×g for 20 min. Cells were resuspended in lysis buffer (20 mM HEPES, 500 mM NaCl, 1 mM TCEP, 10 mM Imidazole, pH 7.5) supplemented with a protease inhibitor (GoldBio Problock). The cell suspension was lysed using a homogenizer (Avestin). The lysate was centrifuged at 4 °C for 30 min to 1 h at 20,000×g. The cleared lysate is filtered to 0.45 μm using a vacuum filter and applied to Ni-NTA beads equilibrated with the lysis buffer. The beads were washed with 40 column volumes of the same lysis buffer with 1 M NaCl. The protein was eluted with an imidazole step gradient (100, 250, and 500 mM imidazole in 50 mM HEPES, 500 mM NaCl, 1 mM TCEP, pH 7.5).

Protein and DNA Concentration Determination. Protein concentrations were calculated based on the UV absorbance extinction coefficient (ε, ml/(mg·cm)) at 280 nm: $TDP43^{RRM1}$ = 1.16 and $CELF1^{RRM1-2}$ = 0.49. Oligonucleotides $(TG)_6$, $(UGUU)_4$ were obtained from IDT (Integrated DNA Technologies, Inc., Coralville, IA) and $(GU)_{22}$, $(GT)_{12}$ from Sigma. Oligonucleotide concentrations were calculated based on UV absorbance extinction coefficient (ε, ml/(mg·cm)) at 280 nm: $(TG)_6$ = 18.21, $(GT)_{12}$ = 18.12, $(UGUU)_4$ = 14.03, and $(GU)_{22}$ = 14.86. The molecular weights of the protein and RNA are summarized in Table S1.

Electrophoretic Mobility Shift Assay (EMSA). The EMSA binding reactions were prepared from 2-fold dilutions of protein:nucleic acid mix and incubated at room temperature for 10 min prior to gel electrophoresis. For nonfluorescent conjugated nucleic acid, the gel was stained with SYBR Gold (Invitrogen) according to the manufacturer's instructions. All gels were imaged by using ChemiDoc with appropriate filters. The protein concentrations, nucleic acid concentrations, and gel running conditions are as follows: $TDP43^{RRM1}$ (1−8 μM); $(TG)_6$ ssDNA (2 μM) or $(TG)_{12}$ ssDNA (1 μM) in buffer (20 mM HEPES, 50 mM NaCl, 1 mM $MgCl_2$, 10% glycerol, 1 mM TCEP, pH 7.3); 10% tris-glycine (TG, Bio-Rad) gel, 90 min at 100 V, 4 °C, in 0.5× TAE buffer. $CELF1^{RRM1-2}$ (3−50 μM) and $(UGUU)_4$ ssRNA (5 μM) in buffer (10 mM HEPES, 12.5 mM Tris, 70 mM NaCl, 12.5 mM KCl, 1 mM $MgCl_2$, 10% glycerol, 0.5 mM TCEP, pH 7.4); 10% Tris-Glycine (TG, Bio-Rad) gel, 60 min at 80 V, 4 °C, in 0.5× TAE buffer. $CELF1^{RRM1-2}$ (4 nM−50 μM) and $(GU)_{22}$ ssRNA (200 nM) in buffer (20 mM HEPES, 50 mM NaCl, 1 mM $MgCl_2$, 10% glycerol, 1 mM TCEP, pH 7.3); 10% TG (TG, Bio-Rad) gel, 90 min at 80 V, 4 °C, in 0.5× TAE buffer.

Composition Gradient Multiangle Light Scattering (CG-MALS). Proteins were passed through size exclusion chromatography (Superdex 75) prior to CG-MALS experiments to ensure the removal of high-MW aggregates. Proteins and oligos were diluted in experimental buffer (25 mM HEPES, 50 mM NaCl, and 1 mM TCEP, pH 7.3). CG-MALS data were collected using a fully automated three-syringe pump Calypso II instrumentation (Wyatt Technology, LLC) linked to a DAWN (Wyatt) multiangle light scattering detector equipped with a 661 nm laser and UV/vis detector (Agilent Technologies) collecting absorbance at 280 nm. Each experiment follows a concentration gradient of one species (A) to assess self-association, followed by a "crossover" gradient with varying concentrations of A and species B to characterize the heterointeractions, and finally, a concentration gradient of species B to assess its self-association. An example of one such experiment is shown in Figure S1. Initial concentrations for the different protein:nucleic acid complexes are approximately as follows: $TDP43^{RRM1}$:$(TG)_6$ are ∼24 μM and ∼14 μM, respectively; $TDP43^{RRM1}$:$(GT)_{12}$ are ∼21 μM and ∼5 μM, respectively; $CELF1^{RRM1-2}$:$(UGUU)_4$ are ∼9 μM and ∼19 μM, respectively; and $CELF1^{RRM1-2}$:$(GU)_{22}$ are ∼11 μM and ∼4 μM, respectively. Final protein and oligo concentrations were adjusted based on the UV absorbance signal detected in the CALYPSO software (Wyatt). Three replicates were performed. All analysis, including model fitting and data simulations, was performed using CALYPSO software (Wyatt). Light scattering data were fit to the following equations.[9,10,20]

$$\frac{R}{K^*} = \sum_{i,j} (iM_A + jM_B)^2 [A_i B_j]$$

https://doi.org/10.1021/acsomega.4c06358
ACS Omega 2024, 9, 41003−41010

$$K_{ij} = \frac{[A_iB_j]}{[A]^i[B]^j}$$

$$[A]_{total} = \sum_{i,j} i[A_iB_j],\ [B]_{total} = \sum_{i,j} j[A_iB_j]$$

where R is the Rayleigh ratio; K_{ij} is the equilibrium association constant of the A_iB_j complex; A and B are constituent monomers; i and j represent the stoichiometric numbers of A and B in the A_iB_j complex; M_A and M_B are the molar masses of monomers A and B, respectively; and K^* is a constant defined below

$$K^* = \frac{4\pi^2\left(\frac{dn}{dc}\right)^2 {n_0}^2}{N_a {\lambda_0}^4}$$

where (dn/dc) is the refractive index increment of a sample−solvent combination; n_0 is the refractive index of the solvent; N_a is Avogadro's number; and λ_0 is the vacuum wavelength of incident light.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c06358.

Theoretical molecular weights (MW) of protein/nucleic acid components and CG-MALS example method with varying protein and nucleic acid concentrations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Josephine C. Ferreon − *Department of Biochemistry and Molecular Pharmacology, Baylor College of Medicine, Houston, Texas 77030, United States*; orcid.org/0000-0002-3175-5700; Email: josephine.ferreon@bcm.edu

Sophia Kenrick − *Wyatt Technology, LLC, Santa Barbara, California 93111, United States*; Email: sophia_kenrick@waters.com

Joel Neilson − *Department of Integrative Physiology, Baylor College of Medicine, Houston, Texas 77030, United States*; Email: neilson@bcm.edu

Allan Chris M. Ferreon − *Department of Biochemistry and Molecular Pharmacology, Baylor College of Medicine, Houston, Texas 77030, United States*; Email: allan.ferreon@bcm.edu

Authors

Natee Kongchan − *Department of Integrative Physiology, Baylor College of Medicine, Houston, Texas 77030, United States*

Phoebe S. Tsoi − *Department of Biochemistry and Molecular Pharmacology, Baylor College of Medicine, Houston, Texas 77030, United States*

Kyoung-Jae Choi − *Department of Biochemistry and Molecular Pharmacology, Baylor College of Medicine, Houston, Texas 77030, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c06358

Author Contributions

Conceptualization, J.C.F.; Methodology, A.C.M.F. and J.C.F.; Software and Formal Analysis, P.S.T., S.K., A.C.M.F., and J.C.F.; Investigation, K.-J.C., P.S.T., A.C.M.F., and J.C.F.; Visualization and Writing—Original Draft, Review, and Editing: A.C.M.F., J.R.N., S.K., and J.C.F.; Supervision, J.R.N., A.C.M.F., and J.C.F.; Funding Acquisition, J.R.N., A.C.M.F., and J.C.F.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the Recombinant Protein Production and Characterization Core (RPPCC) for access to the CG-MALS instrumentation (NIH S10 OD030276). This work was also supported by grants from the NIH R01 GM122763 and Welch Q-2097-20220331 (J.C.F.), the Adrienne Helis Malvin Medical Research Foundation and the Samuel Waxman Cancer Research Foundation (J.R.N), and NIH grant R01 NS105874 (A.C.M.F.).

■ REFERENCES

(1) Ruthenburg, A. J.; Li, H.; Patel, D. J.; David Allis, C. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (12), 983−994.

(2) Lee, C. W.; Ferreon, J. C.; Ferreon, A. C. M.; Arai, M. A.; Wright, P. E. Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (45), 19290−19295.

(3) Li, P.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483* (7389), 336−340.

(4) Hallegger, M.; Chakrabarti, A. M.; Lee, F. C. Y.; Lee, B. L.; Amalietti, A. G.; Odeh, H. M.; Copley, K. E.; Rubien, J. D.; Portz, B.; Kuret, K.; et al. TDP-43 condensation properties specify its RNA-binding and regulatory repertoire. *Cell* **2021**, *184* (18), 4680−4696.e22.

(5) Choi, K. J.; Quan, M. D.; Qi, C.; Lee, J. H.; Tsoi, P. S.; Zahabiyon, M.; Bajic, A.; Hu, L.; Prasad, B. V. V.; Liao, S. C. J.; et al. NANOG prion-like assembly mediates DNA bridging to facilitate chromatin reorganization and activation of pluripotency. *Nat. Cell Biol.* **2022**, *24* (5), 737−747.

(6) Tsoi, P. S.; Choi, K. J.; Leonard, P. G.; Sizovs, A.; Moosa, M. M.; MacKenzie, K. R.; Ferreon, J. C.; Ferreon, A. C. M. The N-Terminal Domain of ALS-Linked TDP-43 Assembles without Misfolding. *Angew. Chem., Int. Ed. Engl.* **2017**, *56* (41), 12590−12593.

(7) Sharma, R.; Choi, K. J.; Quan, M. D.; Sharma, S.; Sankaran, B.; Park, H.; LaGrone, A.; Kim, J. J.; MacKenzie, K. R.; Ferreon, A. C. M.; et al. Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization. *Nat. Commun.* **2021**, *12* (1), 5579.

(8) Rengifo-Gonzalez, J. C.; El Hage, K.; Clement, M. J.; Steiner, E.; Joshi, V.; Craveur, P.; Durand, D.; Pastre, D.; Bouhss, A. The cooperative binding of TDP-43 to GU-rich RNA repeats antagonizes TDP-43 aggregation. *Elife* **2021**, *10*, No. e67605.

(9) Arisaka, F.; Niimura, Y.; Minton, A. P. Comparison of composition-gradient sedimentation equilibrium and composition-gradient static light scattering as techniques for quantitative characterization of biomolecular interactions: A case study. *Anal. Biochem.* **2019**, *583*, 113339.

(10) Attri, A. K.; Minton, A. P. Composition gradient static light scattering: a new technique for rapid detection and quantitative characterization of reversible macromolecular hetero-associations in solution. *Anal. Biochem.* **2005**, *346* (1), 132−138.

(11) Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **1993**, *272*, 1−40.

(12) Halling, D. B.; Kenrick, S. A.; Riggs, A. F.; Aldrich, R. W. Calcium-dependent stoichiometries of the KCa2.2 (SK) intracellular

https://doi.org/10.1021/acsomega.4c06358
ACS Omega 2024, 9, 41003−41010

domain/calmodulin complex in solution. *J. Gen. Physiol.* **2014**, *143* (2), 231−252.
(13) Esfandiary, R.; Hayes, D. B.; Parupudi, A.; Casas-Finet, J.; Bai, S.; Samra, H. S.; Shah, A. U.; Sathish, H. A. A systematic multitechnique approach for detection and characterization of reversible self-association during formulation development of therapeutic antibodies. *J. Pharm. Sci.* **2013**, *102* (9), 3089−3099.
(14) Esfandiary, R.; Parupudi, A.; Casas-Finet, J.; Gadre, D.; Sathish, H. Mechanism of reversible self-association of a monoclonal antibody: role of electrostatic and hydrophobic interactions. *J. Pharm. Sci.* **2015**, *104* (2), 577−586.
(15) Attri, A. K.; Fernandez, C.; Minton, A. P. Self-association of Zn-insulin at neutral pH: investigation by concentration gradient–static and dynamic light scattering. *Biophys. Chem.* **2010**, *148* (1−3), 23−27.
(16) Attri, A. K.; Fernandez, C.; Minton, A. P. pH-dependent self-association of zinc-free insulin characterized by concentration-gradient static light scattering. *Biophys. Chem.* **2010**, *148* (1−3), 28−33.
(17) Pallesen, J.; Murin, C. D.; de Val, N.; Cottrell, C. A.; Hastie, K. M.; Turner, H. L.; Fusco, M. L.; Flyak, A. I.; Zeitlin, L.; Crowe, J. E.; et al. Structures of Ebola virus GP and sGP in complex with therapeutic antibodies. *Nat. Microbiol.* **2016**, *1* (9), 16128.
(18) Mitchell, S. L.; Ismail, A. M.; Kenrick, S. A.; Camilli, A. The VieB auxiliary protein negatively regulates the VieSA signal transduction system in Vibrio cholerae. *BMC Microbiol.* **2015**, *15*, 59.
(19) Zhao, M.; Wu, S.; Zhou, Q.; Vivona, S.; Cipriano, D. J.; Cheng, Y.; Brunger, A. T. Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **2015**, *518* (7537), 61−67.
(20) Some, D. Light-scattering-based analysis of biomolecular interactions. *Biophys. Rev.* **2013**, *5* (2), 147−158.
(21) Monterroso, B.; Alfonso, C.; Zorrilla, S.; Rivas, G. Combined analytical ultracentrifugation, light scattering and fluorescence spectroscopy studies on the functional associations of the bacterial division FtsZ. protein. *Methods* **2013**, *59* (3), 349−362.
(22) del Castillo, U.; Alfonso, C.; Acebron, S. P.; Martos, A.; Moro, F.; Rivas, G.; Muga, A. A quantitative analysis of the effect of nucleotides and the M domain on the association equilibrium of ClpB. *Biochemistry* **2011**, *50* (12), 1991−2003.
(23) Nishino, T.; Rago, F.; Hori, T.; Tomii, K.; Cheeseman, I. M.; Fukagawa, T. CENP-T provides a structural platform for outer kinetochore assembly. *EMBO J.* **2013**, *32* (3), 424−436.
(24) Zhu, H. L.; Fernandez, C.; Fan, J. B.; Shewmaker, F.; Chen, J.; Minton, A. P.; Liang, Y. Quantitative characterization of heparin binding to Tau protein: implication for inducer-mediated Tau filament formation. *J. Biol. Chem.* **2010**, *285* (6), 3592−3599.
(25) Voges, M.; Herhut, M.; Held, C.; Brandenbusch, C. Light-scattering data of protein and polymer solutions: A new approach for model validation and parameter estimation. *Fluid Phase Equilib.* **2018**, *465*, 65−72.
(26) Rakel, N.; Bauer, K. C.; Galm, L.; Hubbuch, J. From osmotic second virial coefficient (B22) to phase behavior of a monoclonal antibody. *Biotechnol. Prog.* **2015**, *31* (2), 438−451.
(27) Moosa, M. M.; Tsoi, P. S.; Choi, K. J.; Ferreon, A. C. M.; Ferreon, J. C. Direct Single-Molecule Observation of Sequential DNA Bending Transitions by the Sox2 HMG Box. *Int. J. Mol. Sci.* **2018**, *19* (12), 3865.
(28) Edwards, J.; Malaurie, E.; Kondrashov, A.; Long, J.; de Moor, C. H.; Searle, M. S.; Emsley, J. Sequence determinants for the tandem recognition of UGU and CUG rich RNA elements by the two N–terminal RRMs of CELF1. *Nucleic Acids Res.* **2011**, *39* (19), 8638−8650.
(29) Buratti, E.; Baralle, F. E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **2001**, *276* (39), 36337−36343.
(30) Lukavsky, P. J.; Daujotyte, D.; Tollervey, J. R.; Ule, J.; Stuani, C.; Buratti, E.; Baralle, F. E.; Damberger, F. F.; Allain, F. H. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **2013**, *20* (12), 1443−1449.
(31) Buratti, E.; Dork, T.; Zuccato, E.; Pagani, F.; Romano, M.; Baralle, F. E. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* **2001**, *20* (7), 1774−1784.
(32) Ladd, A. N.; Charlet, N.; Cooper, T. A. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol. Cell. Biol.* **2001**, *21* (4), 1285−1296.
(33) Edwards, J. M.; Long, J.; de Moor, C. H.; Emsley, J.; Searle, M. S. Structural insights into the targeting of mRNA GU-rich elements by the three RRMs of CELF1. *Nucleic Acids Res.* **2013**, *41* (14), 7153−7166.
(34) Deng, Y.; Efremov, A. K.; Yan, J. Modulating binding affinity, specificity, and configurations by multivalent interactions. *Biophys. J.* **2022**, *121* (10), 1868−1880.
(35) Lee, J. E.; Cooper, T. A. Pathogenic mechanisms of myotonic dystrophy. *Biochem. Soc. Trans.* **2009**, *37* (6), 1281−1286.
(36) Miller, J. W.; Urbinati, C. R.; Teng-Umnuay, P.; Stenberg, M. G.; Byrne, B. J.; Thornton, C. A.; Swanson, M. S. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* **2000**, *19* (17), 4439−4448.
(37) Mackenzie, I. R.; Rademakers, R.; Neumann, M. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol.* **2010**, *9* (10), 995−1007.
(38) Ling, S. C.; Polymenidou, M.; Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **2013**, *79* (3), 416−438.

https://doi.org/10.1021/acsomega.4c06358
ACS Omega 2024, 9, 41003−41010