Biophysical Journal
Article

Biophysical Society

# Electrostatic Interactions between CSTF2 and pre-mRNA Drive Cleavage and Polyadenylation

Elahe Masoumzadeh,[1] Petar N. Grozdanov,[2] Anushka Jetly,[1] Clinton C. MacDonald,[2] and Michael P. Latham[1,*]
[1]Texas Tech University, Lubbock, Texas and [2]Texas Tech University Health Sciences Center, Lubbock, Texas

ABSTRACT Nascent pre-mRNA 3′-end cleavage and polyadenylation (C/P) involves numerous proteins that recognize multiple RNA elements. Human CSTF2 binds to a downstream U- or G/U-rich sequence through its RNA recognition motif (RRM) regulating C/P. We previously reported the only known disease-related CSTF2 RRM mutant ($CSTF2^{D50A}$) and showed that it changed the on-rate of RNA binding, leading to alternative polyadenylation in brains of mice carrying the same mutation. In this study, we further investigated the role of electrostatic interactions in the thermodynamics and kinetics of RNA binding for the CSTF2 RRM and the downstream consequences for regulation of C/P. By combining mutagenesis with NMR spectroscopy and biophysical assays, we confirmed that electrostatic attraction is the dominant factor in RRM binding to a naturally occurring U-rich RNA sequence. Moreover, we demonstrate that RNA binding is accompanied by an enthalpy-entropy compensation mechanism that is supported by changes in pico-to-nanosecond timescale RRM protein dynamics. We suggest that the dynamic binding of the RRM to U-rich RNA supports the diversity of sequences it encounters in the nucleus. Lastly, in vivo C/P assays demonstrate a competition between fast, high affinity RNA binding and efficient, correct C/P. These results highlight the importance of the surface charge of the RRM in RNA binding and the balance between nascent mRNA binding and C/P in vivo.

SIGNIFICANCE Alternative processing of nascently transcribed RNA is critical for the diversity of proteins produced by the genome. CSTF2 is part of the multiprotein complex that recognizes downstream G/U- or U-rich sequences necessary for mRNA cleavage and polyadenylation (C/P). CSTF2 binds to the mRNA through its RNA recognition motif (RRM). Here, we demonstrate that the RRM uses electrostatic forces and internal protein dynamics to bind to a natural U-rich sequence and that faster binding kinetics negatively regulate C/P in the cell. Together, our data suggest a mechanism for RNA binding by the RRM of CSTF2 whereby protein dynamics enable the domain to recognize a variety of G/U- or U-rich sequences and the kinetics of binding play a role in proper C/P.

## INTRODUCTION

In eukaryotes, the posttranscriptional cleavage and polyadenylation (C/P) of pre-mRNA is an essential step in mRNA maturation, altering mRNA stability, export, and translation efficiency. This process is also central to increasing the diversity of mRNAs, as ~70% of nascent pre-mRNAs have different C/P sites that lead to alternative isoforms of translated proteins (1,2). Within the nascent pre-mRNA, three *cis*-acting RNA elements form the core polyadenylation signal and have to be recognized for C/P to occur: 1) a conserved AAUAAA sequence located 10 to 30 nucleotides upstream of the cleavage site; 2) a less highly conserved G/U-rich element downstream of the cleavage site; and 3) the cleavage site (usually a CA dinucleotide), where the poly(A) tail of the mRNA is ultimately added (3–5).

The cleavage stimulation factor (CstF) is one of the major protein complexes that participates in the C/P reaction. CstF binds to the downstream G/U- and U-rich RNA sequences and participates in the regulation of C/P by modulating the binding of the C/P specificity factor to cleave the pre-mRNA (6–8). The CstF complex is a dimer of a heterotrimer composed of three proteins: CstF-77, CstF-64, and CstF-50. Of these, CstF-64 (gene symbol: *CSTF2)* is responsible for the RNA binding function of the CstF complex, although CstF-77 and CstF-50 also modulate CstF-64 RNA binding (9,10). CSTF2 contains five domains, the most studied of which is the RNA recognition motif (RRM) located at the N-terminus of the protein. Based on the x-ray crystal

*Submitted September 13, 2021, and accepted for publication January 7, 2022.*

*Correspondence: michael.latham@ttu.edu

Editor: Bernd Reif.

https://doi.org/10.1016/j.bpj.2022.01.005



structure of the yeast homolog Rna15p bound to one nucleotide of a five-nucleotide RNA (11), NMR binding data (12), and sequence homology, the CSTF2 RRM domain (hereafter referred to as RRM) interacts with the downstream U- or G/U-rich pre-mRNA sequence through three binding sites (Fig. 1 *A*) (8). Site I includes residues Y25 and R85 (*Homo sapiens* numbering), whose side chains interact with uracil or guanidine nucleobases through base-backbone hydrogen bonding interactions. Site II, which contains residues F19 and N91, and site III, which is formed by residues Y59 and F61, bind RNA via less specific nucleobase-aromatic side-chain stacking interactions (11). Even with these models, the exact mechanism (i.e., complete structure, role of RNA sequence, or effects of RNA/protein dynamics) by which the RRM recognizes U- and G/U-rich RNA sequences is yet to be fully understood.

We recently characterized the first known disease-associated mutant of the CSTF2 RRM domain, $CSTF2^{D50A}$, in which an aspartate residue (D) at position 50 of the RRM is mutated to an alanine (A) (Fig. 1 *A*) (13). As *CSTF2* is on the X chromosome, males carrying the single copy of this mutation are hemizygous and exhibit intellectual disabilities including delayed speech development and low verbal and nonverbal communication skills. We showed that substitution of the negatively charged aspartate with a neutral alanine in a non-RNA-binding surface loop of the RRM resulted in higher affinity for RNA, driven by a faster on-rate for the interaction. This D50A mutation led to reduced polyadenylation efficiency in vivo and a preference for processing at proximal (i.e., closer to the transcription start sites) C/P sites in the brains of mice (13). Considering the strong negative electrostatic field associated with the RNA phosphate backbone, charge-charge interactions are likely to be important for proper RNA-protein binding and ultimately for function in many RRM domain-containing complexes (14–16); yet, there have been relatively few studies that have investigated the effect of these interactions on RRM binding properties. Several reports on other RRM family members (e.g., U1A) have indicated that electrostatic forces contribute to the energy needed for the RRM-RNA interaction (17–21). However, in each of these studies, only the electrostatic interactions involving RRM residues within the RNA-binding pocket were considered, whereas the importance of the total surface charge of the protein has been overlooked.

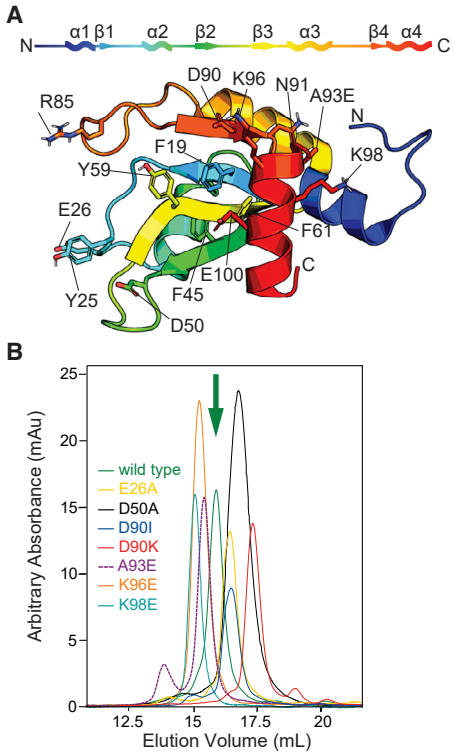


FIGURE 1 Structure of the CSTF2 RRM. (*A*) Solution-state structure of the RRM from CSTF2 (PDB: 6Q2I) (13). Binding site residues and the positions of mutants considered in this study are shown as sticks. (*B*) Size exclusion chromatogram illustrating how the surface charge mutants alter the elution volume when a no-salt buffer is used. The green arrow highlights the Ve for wild-type.

Here, we investigated the role of electrostatic interactions in the thermodynamics and kinetics of RNA binding and the regulation of C/P by the RRM domain of CSTF2. By combining mutagenesis with NMR spectroscopy and biophysical assays, we confirm that electrostatic attraction is the dominant factor in the RRM binding to a U-rich RNA sequence from the SV40 virus late transcription unit (SVL) polyadenylation site. Moreover, we demonstrate that RNA binding is accompanied by an enthalpy-entropy compensation mechanism that is supported by changes in fast timescale RRM protein dynamics. Finally, in vivo C/P assays show a negative correlation with changes in the overall charge of the RRM and a positive correlation with thermodynamics of RNA binding, highlighting the balance between pre-mRNA binding and C/P. The results presented here provide a deeper insight into the RNA binding mechanism of RRM-containing proteins and the importance of the global surface charge, which illuminates a possible role for RRM dynamics in U- and G/U-rich RNA sequence recognition.

## MATERIALS AND METHODS

### Plasmids and mutagenesis

The RRM domain of *H. sapiens* CSTF2 (amino acids 1–107) was cloned into a pET22 bacterial expression vector with an amino terminal 6xHis-tag followed by a TEV protease cleavage site as previously described (9). RRM point mutations were made using a modified Quikchange (Stratagene) approach. Plasmids for the stem-loop assay for polyadenylation (SLAP) assay were made in a similar site-directed mutagenesis approach. All plasmids were verified by Sanger sequencing before use.

### Protein expression and purification

Expression and purification, via Ni-affinity resin, of wild-type and mutant RRM proteins were as previously described (9). For NMR experiments, transformed Rosetta (DE3) pLysS cells were grown in 2x minimal M9

media (22) using $^{15}NH_4Cl$ (1 g/L) and unlabeled D-glucose (3 g/L) as sole nitrogen and carbon sources, respectively. Protein expression was induced with 0.5 mM isopropyl-β-d-thiogalactopyranoside for 18 h at 25°C. Protein purification proceeded as previously described (9,13).

## Size exclusion chromatography

The size exclusion chromatography column (GE Healthcare Superedex 200 10/300 GI Chromatography column) was equilibrated in either 10 mM sodium phosphate (pH 6) and 0.05% w/v sodium azide or 10 mM sodium phosphate (pH 6), 200 mM sodium chloride, and 0.05% w/v sodium azide. 0.1–0.12 mg of each RRM mutant was loaded onto the column with a flow rate of 0.4 mL/min using a GE Healthcare AKTA Pure FPLC system. Resulting chromatograms were analyzed with the UNICORN software (version 7).

## Native PAGE

Native polyacrylamide gel electrophoresis (PAGE) analysis of RRM mutants was carried out using a precast mini-protein gradient page gel (Any Kd, BioRad). 4.5 μg of each mutant was loaded onto the gel, which was run at 200 V constant voltage for 8 h at room temperature. Protein bands were visualized by staining with Coomassie blue. Sample buffer was 62.5 mM Tris-HCl (pH 6.8), 25% glycerol, 20% bromophenol blue in water. PAGE running buffer was 2.5 mM Tris-HCl (pH 6.8), and 19.5 mM glycerol.

## Isothermal titration calorimetry

We used the simian virus 40 (SV40) late polyadenylation signal (SVL) as a model for a naturally occurring strong polyadenylation signal (8). The RRM proteins and SVL RNA (5′-AUUUUAUGUUUCAGGU-3′; Millipore-Sigma) were dialyzed overnight into 10 mM sodium phosphate, 0.05% w/v sodium azide (pH 6) buffer using a 2-kDa MWCO dialysis unit (ThermoFisher) and then degassed. Heats of binding were measured from the titration of RNA into the protein using a MicroCal iTC200 calorimeter (GE Healthcare) with a stirring rate of 1000 rpm at 27°C. For all titrations, isotherms were corrected by subtracting the heats of RNA dilution. The concentrations of proteins and RNA were calculated by absorbance spectroscopy with extinction coefficients of $\varepsilon_{280} = 5960\ M^{-1}\ cm^{-1}$ and $\varepsilon_{260} = 165.7\ M^{-1}\ cm^{-1}$, respectively. RRM concentrations ranged between ∼20 μM (D50A and D90K) to ∼40 μM (K98A/E), and the SVL concentrations ranged between ∼210 μM (D50A and D90K) to ∼500 μM (K98A/E). Binding isotherms were fit to "One Set of Sites" model in Origin 7. Reported values and errors are the average and standard deviation of at least two experiments.

## NMR experiments

NMR experiments were recorded in 10 mM phosphate buffer (pH 6.0), 1 mM TCEP, 0.05% w/v sodium azide, 0.1 mg/mL 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), and 10% $D_2O$ using an Agilent 600-MHz (14.1 T) DD2 NMR spectrometer equipped with a room temperature HCN $z$ axis gradient probe at 27°C. Data were processed with NMRPipe/NMRDraw (23) and analyzed with CCPN Analysis (24). Backbone amide $^1H$ and $^{15}N$ assignments for RRM mutants were verified with 3-dimensional $^{15}N$-edited NOESY-HSQC and TOSCY-HSQC experiments.

RNA titration experiments were performed by adding unlabeled SVL RNA to the $^{15}N$-labeled proteins and monitoring the change in amide chemical shifts in 2D $^{15}N$, $^1H$ HSQC spectra until complete saturation was reached. Amide chemical shift perturbations (CSPs) were calculated as $CSPs = \sqrt{(\delta^N_{WT} - \delta^N_{Mut})^2/25 + (\delta^H_{WT} - \delta^H_{Mut})^2/2}$, where $\delta^N$ and $\delta^H$ are the nitrogen and proton chemical shifts, respectively.

NMR-derived on- and off-rates for RNA binding were calculated from peaks in the titration spectra undergoing intermediate and fast exchange (∼10–20 residues) using a global two-site binding model in the program TITAN (25). The $K_D$ values in the TITAN fits were set to the average value obtained from ITC measurements and were not fitted during the calculation for the on- and off-rates. Reported errors in the kinetics for RNA binding were obtained from the covariance of the fit as reported by TITAN.

Backbone $^{15}N$ $R_1$ (longitudinal relaxation rate), $R_{1\rho}$ (rotating frame relaxation rate), and heteronuclear $\{^1H\}$-$^{15}N$ NOE relaxation data (26,27) for RRM wild-type and mutants in the absence and presence of SVL RNA were also acquired at 600 MHz and 27°C. $^{15}N$ $R_1$ and $R_{1\rho}$ relaxation rates were calculated from six parametrically varied timepoints ranging from 10–500 msec ($R_1$) or 2–60 msec ($R_{1\rho}$). $R_{1\rho}$ values were converted to transverse relaxation rates ($R_2$) using the relation $R_{1\rho} = R_1\ cos^2\theta + R_2\ sin^2\theta$, where $\theta = tan^{-1}(\nu_{SL}/\Omega)$ is the effective tilt angle of the rotating frame, $\nu_{SL}$ is the field strength of the applied spin lock (∼2000 Hz), and Ω is the offset of the peak from the $^{15}N$ carrier. Errors in the relaxation rates were calculated from the covariance matrix of the fit. $^{15}N$ $R_1$, $R_2$, and NOE data were used to calculate residue-specific backbone order parameters ($S^2_{NH}$) and effective correlation time for internal motions ($\tau_e$) and the overall global correlation time ($\tau_c$) from the model-free approach (28,29) using "model 2" and an axially symmetric diffusion model in Modelfree v4.2 (https://comdnmr.nysbc.org/comd-nmr-dissem/comd-nmr-software/software/modelfree). N-H bond lengths and $^{15}N$ chemical shift anisotropy were assumed to be 1.02 Å and −160 ppm, respectively. Errors in $S^2_{NH}$ were taken directly from Modelfree.

Backbone $S^2_{NH}$ values in the apo and SVL-bound states were then used to derive the change in conformational entropy upon SVL binding ($\Delta S_{conf}$) according to

$$\Delta S_{conf} = 0.88\ N_{res}\langle ln\left(1 - S^2_{NH,\ SVL}\right) - ln\left(1 - S^2_{NH,\ apo}\right)\rangle,$$

where $N_{res}$ is the total number of residues and "< >" denotes the average (30). Note, residues where at least one value of $S^2_{NH} < 0.8$ were used in the calculation, as this was previously shown to give a better correlation between the conformational entropy of the backbone amide group and $ln(1 - S^2_{NH})$ (30).

## SLAP assay

SLAP assay was performed as previously described (9,13). Briefly, 40,000 HeLa cells were transfected with a luciferase reporter gene containing two MS2 stem-loops, normalization control, and either wild-type MCP-CSTF2 or constructs containing the specific mutations in the RRM of human CSTF2 using LTX reagent (Thermo Fisher Scientific) in 24-well format. 36–48 h post transfection the cells were lysed as outlined in the Dual-Luciferase Reporter Assay System kit and processed accordingly to measure the luciferases abundance. The ratio of the relative expression of the luciferases was normalized to control cells that were not transfected with any MCP-CSTF2 construct.

# RESULTS

## Surface electrostatic attraction is the driving force for RNA-RRM interactions

Several point mutants within the human CSTF2 RRM were made focusing on surface-exposed, charged residues in the α4-helix (D90, K96, and K98) (Fig. 1 *A*). We also mutated a nonpolar residue in this region (A93) as well as F45 and E26, which is adjacent to binding site I residue Y25 (Fig. 1 *A*). The effect of each mutant on the theoretical pI

and electrostatic surface potential is given in Table S1 and Fig. S1, respectively. To characterize the change in total surface charge of these mutants, we used size exclusion chromatography (Fig. 1 *B*). Under no-salt conditions, electrostatic interactions between the column and the RRM should dominate the observed elution volume (Ve), which would then reflect the relative RRM surface charge (31,32). Conversely, when NaCl was added to the buffer, interactions between the RRM and the column solid phase were masked. Indeed, we observed a dramatic shift in the elution volumes of the wild-type RRM in the absence and presence of 200 mM NaCl (15.9 mL versus 17.0 mL, respectively; Figs. 1 *B* and S2 *A*). Generally, mutations that switched the charge of a residue by –2*e* (i.e., K98E and K96E, where a positive charge was replaced by a negative charge) had the smallest Ve (14.9 mL and 15.0 mL) in the absence of salt, whereas mutations that switched the charge by +2*e* (i.e., D90K, where a negative charge was replaced by a positive charge) had the largest Ve (17.1 mL; Table 1). The F45A mutant, which did not change the charge, eluted at the same Ve as wild-type RRM. When 200 mM NaCl was added to the buffer, all the mutants eluted at the same wild-type Ve of ~17.0 mL (Fig. S2 *A*).

We further confirmed that the mutants affected the net surface charge of the RRM using native PAGE (Fig. S2 *B*). Mutants with a change in charge of –2*e* had the greatest mobility toward the cathode, followed by those with a change in charge of –1*e*, 0, and finally +1*e*. Mutants with a change in charge of +2*e* migrated the least (staying closer to the anode). These observations were consistent with the relative mobilities observed in the no-salt size exclusion chromatography (Fig. 1 *B*).

Having characterized the effects of these mutants on the net surface charge of the RRM, we next measured the apparent binding affinity ($K_d$) of the RRM for a 16-nucleotide polyadenylation region of the SV40 virus late transcription unit (SVL) RNA (5′-AUUUUAUGUUUCAGGU-3′) (8,13) via isothermal titration calorimetry (ITC) (Figs. 2 and S3). RRM mutants that became more positively charged displayed a greater binding affinity for SVL RNA (i.e., lower $K_d$), whereas RRM mutants that were more negative showed a weaker affinity for SVL RNA (Table 1). These data also showed that mutants with a greater positive charge had a more favorable enthalpy (Fig. 2 *B* and *C* and Table 1), likely reflecting stronger electrostatic interactions with the negatively charged ribose-phosphate backbone of the SVL RNA, compared with the mutants with a negative change in net charge. Correspondingly, a correlation was observed between ΔH and the Ve from the no-salt size exclusion chromatograms (Pearson's correlation coefficient, $R_P = -0.863$). As noted for the disease-associated D50A mutation (13), a clear enthalpy-entropy compensation trend was also observed among these charge mutants, whereby the large unfavorable entropy (i.e., positive value for –TΔS) was compensated by a favorable negative enthalpy (Fig. 2 *C*; $R_P = 0.999$). Because of the enthalpy-entropy compensation, a trend is also observed between the entropy and Ve (Fig. 2 *B*; $R_P = -0.888$). In total, these data are consistent with the hypothesis that differences in the thermodynamics of SVL RNA binding are due to changes in the net surface charge of the RRM (Fig. 2 *B*).

## Surface mutations do not alter the global structure of the RRM, but do affect interactions with RNA at RNA binding site residues

We used solution-state NMR spectroscopy to characterize the RRM mutants before and after adding SVL RNA. In all cases, the 2D $^{15}N,^{1}H$ heteronuclear single quantum correlation (HSQC) spectra showed CSPs (wild-type versus mutant) for several resonances across the CSTF2 RRM (Figs. 3 *A*, left, and S4, top). In general, the majority of the peaks did not shift, suggesting that the overall tertiary fold of the RRM does not change upon mutation. In the absence of RNA, D50A and E26A showed larger CSPs local to the mutation (arrows in Figs. 3 *A*, left, and S4, top),

**TABLE 1 Effect of Electrostatic RRM Mutants on RNA Binding**

| | Δe | Ve (mL) | $K_d$ (μM) | ΔH (kcal mol$^{-1}$) | ΔS (cal mol$^{-1}$ K$^{-1}$) | N | $k_{on}$ ($M^{-1}$ $s^{-1}$) | $k_{off}$ ($s^{-1}$) | $\Delta S_{conf}$ (cal mol$^{-1}$ K$^{-1}$) | Normalized Luciferase Signal |
|---|---|---|---|---|---|---|---|---|---|---|
| D90K | +2 | 17.13 | 0.608 ± 0.01 | −39.8 ± 1.0 | −104 ± 2.8 | 0.38 ± 0.01 | 5.43 ± 0.17 x $10^8$ | 329.9 ± 7.7 | −27.1 | 4.79 |
| D50A | +1 | 16.87 | 0.698 ± 0.06 | −32.4 ± 0.3 | −79.7 ± 0.9 | 0.62 ± 0.01 | 4.14 ± 0.36 x $10^8$ | 289.2 ± 1.8 | −30.0 | 5.12 |
| D90I | +1 | 16.57 | 1.26 ± 0.33 | −31.6 ± 1.6 | −78.2 ± 5.9 | 0.42 ± 0.02 | 2.54 ± 0.67 x $10^8$ | 320.6 ± 2.3 | −29.3 | 6.16 |
| E26A | +1 | 16.54 | 0.52 ± 0.01 | −26.4 ± 1.7 | −59.4 ± 5.7 | 0.66 ± 0.05 | 6.23 ± 0.10 x $10^8$ | 324.2 ± 1.9 | 12.6 | 6.13 |
| WT | 0 | 15.88 | 1.52 ± 0.17 | −24.6 ± 1.6 | −55.5 ± 5.4 | 0.70 ± 0.01 | 2.02 ± 0.22 x $10^8$ | 306.4 ± 1.4 | 10.8 | 6.59 |
| F45A | 0 | 15.88 | 1.01 ± 0.05 | −32.0 ± 1.7 | −79.3 ± 5.9 | 0.68 ± 0.05 | ND | ND | ND | ND |
| A93E | −1 | 15.39 | 3.88 ± 0.77 | −24.5 ± 1.8 | −56.6 ± 6.1 | 0.76 ± 0.09 | 7.44 ± 1.48 x $10^7$ | 288.7 ± 1.1 | 12.1 | 7.21 |
| K96E | −2 | 15.00 | 38.74 ± 4.2 | −16.1 ± 1.5 | −31.3 ± 4.7 | 0.90 ± 0.37 | 4.26 ± 0.46 x $10^6$ | 164.9 ± 0.7 | 13.0 | 1.56 |
| K98E | −2 | 14.94 | 8.70 ± 0.11 | −17.4 ± 1.2 | −34.8 ± 3.8 | 0.71 ± 0.01 | 1.26 ± 0.17 x $10^7$ | 109.3 ± 0.6 | ND | 6.76 |

Size exclusion elution volume, binding affinity, thermodynamics and kinetics of binding, conformational entropy, and normalized luciferase signal for RRM wild-type and mutants as obtained from size exclusion, isothermal titration calorimetry (ITC), NMR RNA titration, order parameters, and SLAP assay, respectively. Thermodynamics, kinetics, and dynamics information were determined at 27°C. The reported thermodynamic parameters are the average and standard deviation of at least two measurements, whereas the errors in binding rates are from the global fit of the data from TITAN.

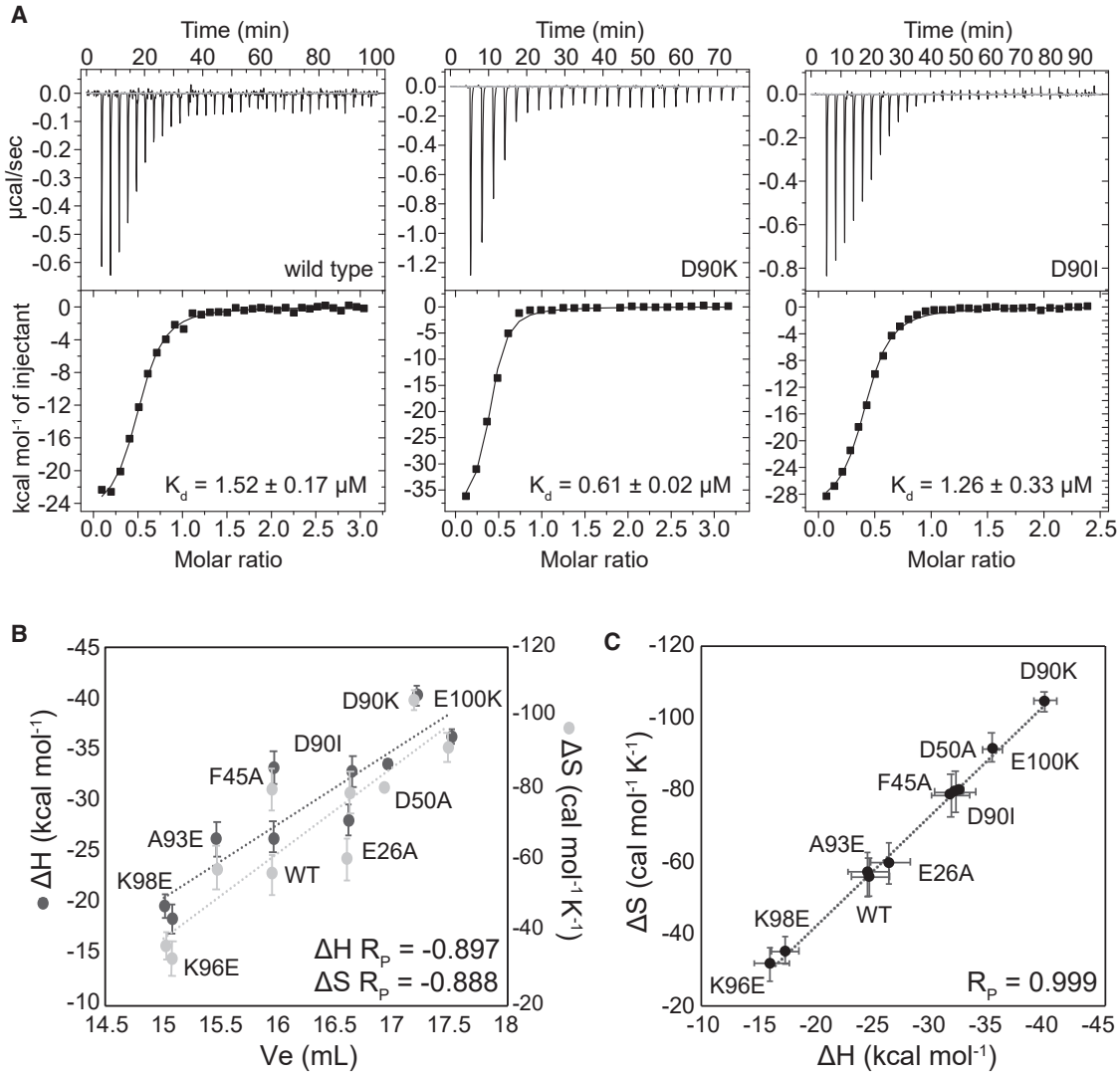


FIGURE 2 Electrostatic interactions drive the RNA-RRM interaction. (*A*) Representative ITC thermograms for wild-type (left), D90K (middle), and D90I (right) RRM binding to SVL RNA. (*B*) Plot of the extracted thermodynamic parameters ΔH (dark gray circles) and ΔS (light gray circles) versus the elution volume (Ve). The dotted line represents the line of best fit through the data, and the Pearson's correlation coefficient ($R_P$) is given for each data set in the lower right-hand corner. (*C*) Plot of ΔS versus ΔH highlighting the enthalpy-entropy compensation. The dotted line represents the line of best fit through the data, and $R_P$ is given in the lower right-hand corner.

whereas D90I and D90K showed the largest CSPs in regions far from D90 (boxed peaks in Figs. 3 *A*, left, and S4, top). In an effort to eliminate the local effect of CSPs from each mutant, we plotted the median CSP value for a given residue across all of the RRM mutants on the structure of the RRM domain (Fig. 3 *B*, black line, and *C*, left). As expected, the largest median CSPs were observed in the region around $\beta$4–$\alpha$4, which is where most of the mutants are found in the structure.

After adding saturating SVL RNA to the NMR samples, the mutant HSQC spectra again showed CSPs when compared with the SVL-bound wild-type RRM (Figs. 3 *A*, right, and S4, bottom). On average, the CSPs of RNA-bound RRM mutants were larger than those observed for the apo forms, with an average CSP of 0.09 for the SVL-bound constructs compared with 0.03 for the apo constructs. D50A and E26A again had the largest CSPs local to the site of their mutation (arrows in Figs. 3 *A*, right, and S4, bottom), whereas D90I and A93E had the largest CSPs spreading across the $\beta$-sheet (boxed peaks in Figs. 3 *A*, right, and S4, bottom). When plotted on the structure of the RRM (Fig. 3 *B*, red line, and *C*, right), the median SVL-bound CSPs followed a pattern similar to that observed for the apo form of the RRM, with noticeable CSPs across the $\beta$-sheet as well as in the $\alpha$1 and $\alpha$4 helices. We noted additional CSPs for Y25 and N91 residues (which are part of binding sites I and II, respectively) and the N-terminus. Larger CSPs were also observed for residues near the aromatic binding site residues F19 (site II) and F61 (site III). Thus, altering the net charge of the RRM via point mutation

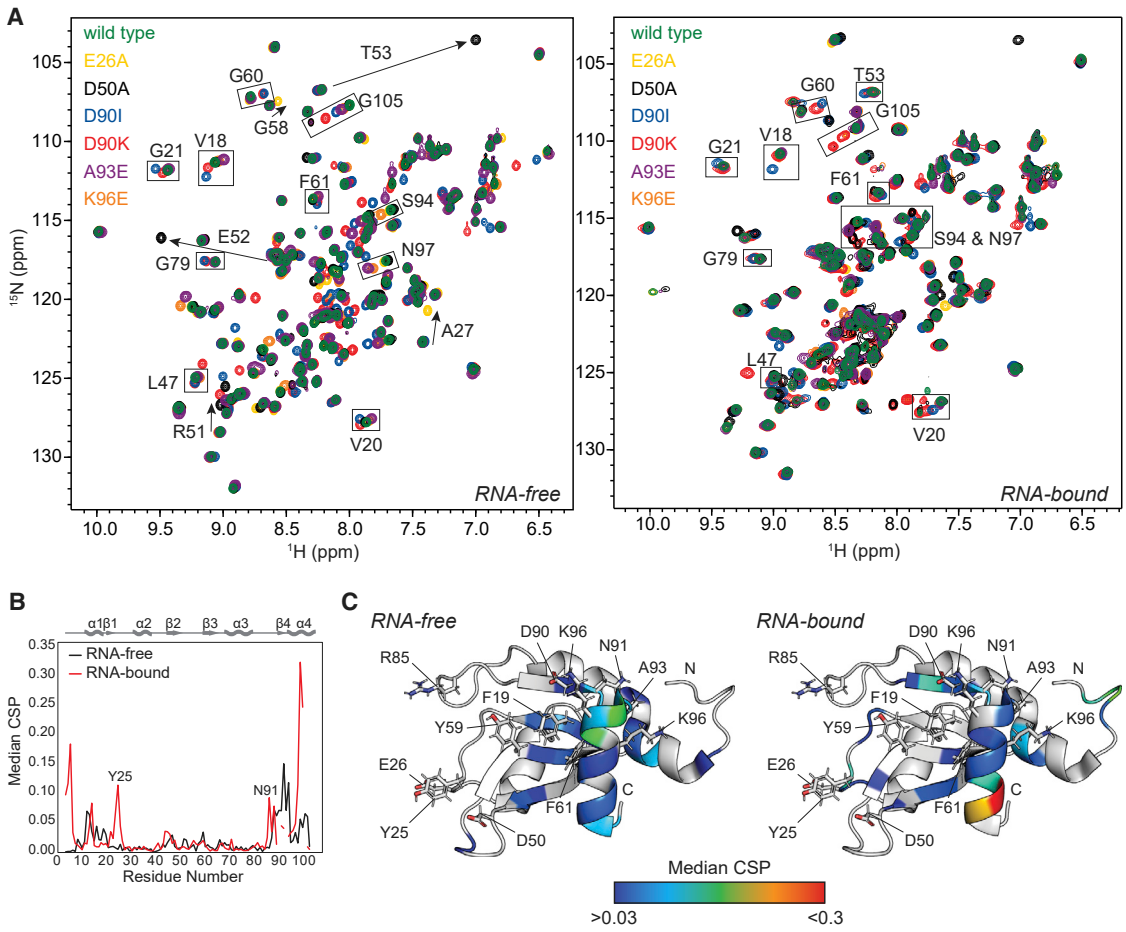


FIGURE 3 Surface charge mutants affect local structure and the C-terminal $\alpha$-helix of the RRM. (*A*) Overlays of 2D $^{15}$N,$^{1}$H HSQC spectra for wild-type and surface charge mutants in the RNA-free (left) and RNA-bound (right) states. In both, boxed peaks denote residues in some RRM mutants (e.g., D90I) far from the sites of mutation that experience CSPs, and arrows denote residues with large CSPs local to the region of the mutant (e.g., D50A). (*B*) Plot of the median chemical shift perturbation (CSP) versus residue number for the RNA-free (black) and RNA-bound (red) states. A diagram above the graph illustrates the RRM secondary structure. (*C*) Median CSPs highlighted on the solution-state structure of the RRM for the RNA-free (left) and RNA-bound (right) states. The color corresponds to the magnitude of the CSPs denoted by the bar below.

had structural effects on binding site residues in the SVL-bound form of the RRM domain.

## NMR titration data reveal that changes in RRM surface charge alter both the on- and off-rates of RNA binding

Using 2D line shape analysis of NMR titration experiments, we previously showed that the D50A mutation in the RRM increased the apparent on-rate for SVL binding without significantly affecting the apparent off-rate, resulting in a lower $K_d$ (13). For both the wild-type and D50A mutant, the apparent on-rates were faster than the rate of diffusion ($4.35 \times 10^8$ and $5.83 \times 10^8$ $M^{-1}sec^{-1}$, respectively, under these sample conditions), which we attributed to electrostatic attraction between the positively charged RRM and negatively charged RNA. To better understand the effects of electrostatics on RNA binding, we performed the same 2D line shape analysis on peaks from the SVL titration spectra in fast and intermediate exchange on the chemical shift timescale for a number of the RRM mutants (Figs. 4 *A* and S5). In most of the mutants, we observed a faster-than-diffusion on-rate, which would be the result of the attraction of the overall positive charge of the RRM and the negative charge of the RNA (Table 1). The K96E and K98E mutants, however, had ~47-fold and ~17-fold decreases in on-rate, respectively, compared with wild-type. As was also observed for the D50A mutant, the off-rate for the charge mutants showed relatively minor changes (ranging between 330 $s^{-1}$ to 290 $s^{-1}$ for D90K and A93E, respectively), again except for K96E and K98E, which had slower off-rates (164.9 $s^{-1}$ and 109.3 $s^{-1}$, respectively). Yet, the effect of this ~two- to three-fold change in off-rate for these two mutants may not be very substantial, considering the ~47-fold and ~17-fold changes in on-rates.

When we compared the calculated on- and off-rates with the change in net surface charge, we observed that positively charged mutants resulted in faster on-rates compared with

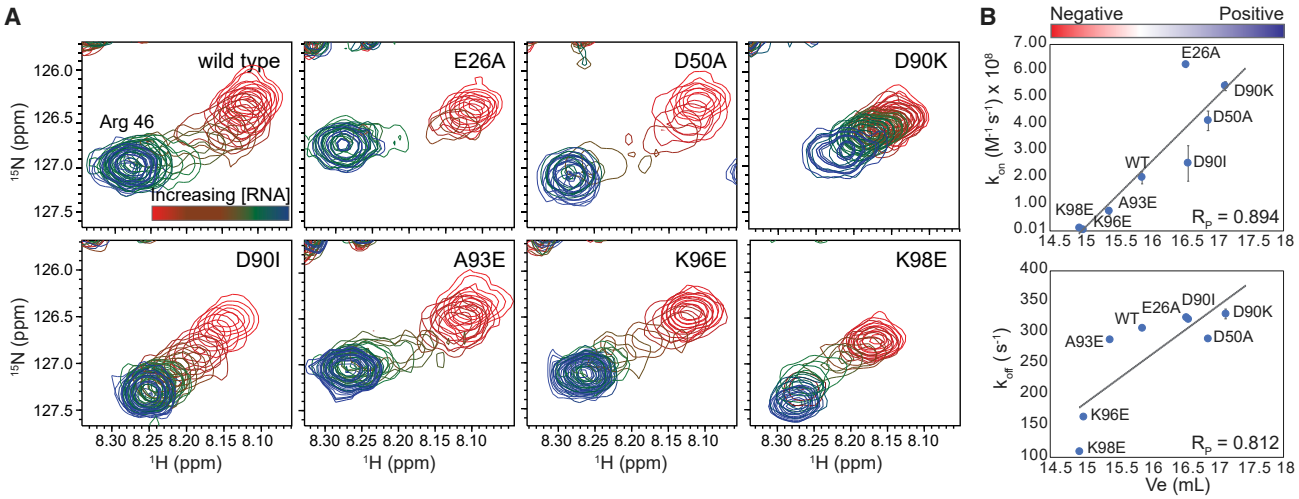


FIGURE 4 Surface charge mutants alter the kinetics of RNA binding. (*A*) Overlays for the SVL RNA titration of Arg46 for wild-type and mutant RRMs highlighting the intermediate exchange regime. RNA concentration for each titration increases from the blue to red shaded contours (from 0 to 2 molar equivalents). (*B*) Plots of $k_{on}$ (left) and $k_{off}$ (right) versus Ve. The dotted line represents the line of best fit through the data, and $R_P$ is given in the lower right-hand corner.

wild-type, whereas negatively charged mutants resulted in slower on-rates. For example, the on-rate for D90K was 5.43 x $10^8$ $M^{-1}s^{-1}$ and for K96E was 4.26 x $10^6$ $M^{-1}s^{-1}$ compared with 2.02 x $10^8$ $M^{-1}s^{-1}$ for wild-type. This trend resulted in a correlation between on-rate and Ve from the size exclusion chromatography ($R_P = 0.894$; Fig. 4 *B*). This correlation was expected since electrostatic attraction drives RNA-RRM association. We also observed a correlation between the off-rates and Ve ($R_P = 0.812$) where mutants with overall positive changes in surface charge had the fastest off-rates and mutants with overall negative changes had the slowest off-rates (Fig. 4 *B*). Thus, both rates were affected by surface electrostatic interaction between the RRM domain and RNA. Based on these data, we suggest that the on-rate drives the initial RNA binding event and the small differences in the off-rate could aid in the selectivity of RNA binding and/or account for the differences we observed for C/P efficiencies in vivo (see below).

## Changes in the backbone flexibility of the RRM are tied to the entropy of SVL RNA binding

Given the importance of electrostatics and enthalpy-entropy compensation in RRM binding to SVL RNA, we investigated the changes in fast timescale motions resulting from RNA binding to various RRM mutants. As previously demonstrated for the RRM D50A mutant, the enthalpy-entropy compensation led to an increase in the amplitude of fast timescale RRM backbone dynamics (13). Accordingly, $^{15}N$ $R_1$, $R_2$, and $\{^1H\}$-$^{15}N$ heteronuclear NOE relaxation data were recorded for several mutants at 14.1 T (600 MHz $^1H$ frequency) to study the nanosecond timescale dynamics that report on the backbone flexibility of RRM constructs. Similar values were observed across all the mutants for the $^{15}N$ relaxation data (Fig. S6), and heteronuclear NOE values were consistent with a lack of picosecond timescale motions throughout the folded portion of the RRM. Motional properties, including the global tumbling time ($\tau_c$), the backbone amide order parameter ($S^2_{NH}$), and time scale of internal motion ($\tau_e$), were extracted using the model-free approach (28,29). The $\tau_c$ for all of the RRM mutants in the RNA-free form was ~6.5 ns, which was similar to the wild-type $\tau_c$ (Table S2). Generally, there was similar backbone flexibility for all of the constructs, as gauged by $S^2_{NH}$, the amplitude of nanosecond timescale motions, with an average $S^2_{NH}$ of ~0.81 (Fig. S7 *A*, Tables S2, and S3). $^{15}N$ $R_1$, $R_2$, and heteronuclear NOE relaxation data for the K98E mutant did not fit well using the RRM structural model and axial diffusion model in the model-free approach; thus, we excluded it from this analysis.

After adding SVL RNA, the $^{15}N$ relaxation data exhibited greater dispersion across the mutants (Fig. S8). $\tau_c$ increased to ~8 ns for wild-type and all the mutant RRMs (Table S2), as expected for the larger RNA-bound complex. The average backbone flexibility of all the complexes slightly increased as denoted by a smaller average $S^2_{NH}$ of ~0.78 (Fig. S7 *B*, Tables S2, and S4). To understand the effect of SVL binding on the dynamics of the various RRM constructs in more detail, we examined the residue-specific deviation in $S^2_{NH}$ ($\delta S^2_{NH} = S^2_{NH,\,bound} - S^2_{NH,\,apo}$) between the RNA-bound and apo forms (Fig. 5 *A*). Although the backbone flexibilities of the various RRM constructs generally increased in the RNA-bound complex ($\delta S^2_{NH} < 0$), the $\beta$-sheet region largely became more rigid, which is illustrated by $\delta S^2_{NH}$ values greater than 0 in Fig. 5 *A*. Backbone flexibility of K96E is virtually unchanged in the $\beta$-sheet, whereas mutants that became more positively charged (e.g., D90K) showed the most rigidity in the $\beta$-sheet after

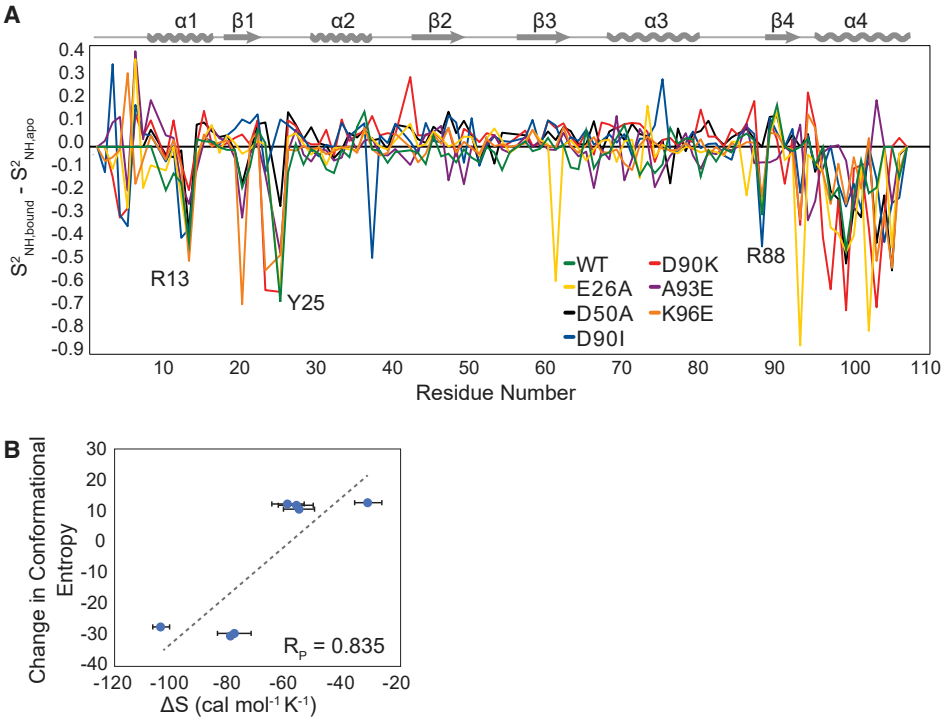


FIGURE 5 Changes in the amplitude of fast timescale motions are correlated with the entropy of binding. (*A*) Difference in the backbone amide order parameter ($S^2_{NH}$) for SVL-bound and apo RRM versus residue number. Differences for each mutant are colored according to legend. (*B*) Plots of the change in conformational entropy, calculated from the relaxation data in Tables S3 and S4, versus ΔS for SVL binding. The dotted line represents the line of best fit through the data, and $R_P$ is given in the lower right-hand corner.

RNA binding. Unlike the $\beta$-sheet, the C-terminal $\alpha$-helix of all the constructs, including wild-type, became strikingly more flexible upon RNA binding, with changes in $\delta S^2_{NH}$ ranging between $-0.9$ and $-0.7$ for some residues in the E26A and D90K (Figs. 5 *A* and S7). This observation was further supported by lower $\{^1H\}$-$^{15}N$ NOE values in this region (Figs. S6 and S8). Moreover, other residues, such as D12, R13, V20, Y25, and R88, which are mostly polar residues located on the RRM surface, also showed consistent increases in flexibility in all the constructs, as noted by decreases in $\delta S^2_{NH}$ of ~0.2–0.6 upon RNA binding. Together, these increases in flexibility were much larger than the increases in the rigidity of the $\beta$-sheet leading to the overall average decrease in $S^2_{NH}$.

To compare the NMR relaxation data with the thermodynamics of RNA binding determined from ITC, we used the backbone $S^2_{NH}$ values to calculate the change in conformational entropy upon substrate binding ($\Delta S_{conf}$). Using the calibration curve of backbone conformational entropy versus $S^2_{NH}$ derived from molecular dynamics simulations by Wand and co-workers (30), we determined that the changes in conformational entropy correlated with the ΔS for RNA binding (Fig. 5 *B* and Table 1; $R_P = 0.835$). For instance, the mutant with the most unfavorable entropy for binding (D90K with $\Delta S = -104$ kcal mol$^{-1}$) had the most unfavorable change in conformational entropy ($\Delta S_{conf} = -27.1$ kcal mol$^{-1}$), whereas the mutant with the least unfavorable entropy for binding (K96E with $\Delta S = -31.3$ kcal mol$^{-1}$) had a beneficial change in conformational entropy ($\Delta S_{conf} = 13.0$ kcal mol$^{-1}$). Heightened flexibility for residues D12, R13, V20, Y25, and R88 along with flexibility in the C-terminal $\alpha$-helix drove the changes in conformational entropy and compensated for the additional rigidification of the $\beta$-sheet. As noted above, K96E had nearly no change in backbone flexibility within the $\beta$-sheet, but because of the increased motions in the C-terminal $\alpha$-helix, K96E had the greatest change in conformational entropy. These data indicate that the RRM uses an increase in conformational entropy in the C-terminal $\alpha$-helix, possibly from local unfolding, to compensate partially for the loss of entropy, likely from restricting the conformational space of the single-stranded RNA ligand, to facilitate binding. Such a dynamic complex likely participates in the observed enthalpy-entropy compensation and may play a role in the selectivity of the CSTF2 RRM for generic U- and G/U-rich RNA sequences.

## RRM-RNA electrostatic interactions drive C/P in cells

To explore the effects of these biophysical characteristics of the CSTF2 RRM on C/P, we performed the SLAP assay (9,33). In SLAP, the specific interaction between an RRM fused with the MS2 bacteriophage coat-binding protein (MCP) and two copies of the cognate stem-loop RNA element promotes C/P in a luciferase reporter gene and subsequent luciferase signal (9,33). Although SLAP should be largely dependent on the fusion protein-stem loop RNA interaction and independent of RRM-RNA binding, we have previously shown that mutations to the RRM RNA

binding site residues and the disease-associated mutation D50A had an effect on the SLAP signal (9,13). Here, we also observed differences in SLAP upon mutation of surface electrostatic residues (Fig. S9 and Table 1). Surprisingly, the difference in SLAP luciferase signal was negatively correlated with Ve (except K96E, which had SLAP activity similar to the background). Generally, mutants with a negative change in surface charge and smaller Ve (e.g., K98E) had greater SLAP activity compared with wild-type, whereas mutants with a positive change in surface charge and larger Ve (e.g., D90I) had less SLAP activity (Fig. 6 *A*; $R_P = -0.88$, excluding K96E). Therefore, even in the presence of the specific MCP-stem loop RNA interaction, electrostatic attraction between the CSTF2 RRM and pre-mRNA alters C/P efficiency in cells.

In addition to the correlation with Ve, the SLAP activity for RRM mutants was also negatively correlated with the on-rate of RNA binding, meaning that the mutants with a faster on-rate (e.g., D90K and D50A) had less SLAP activity than wild-type, whereas mutants with slower on-rate (e.g., A93E and K98E) showed greater SLAP activity (Fig. 6 *B*; $R_P = -0.75$, excluding K96E). Although, we did not observe a good correlation between the SLAP activity and the RNA binding affinity ($R_P = 0.59$, excluding K96E), SLAP activity had a significant correlation with the enthalpy and entropy of RNA binding for RRM mutants (Figs. 6 *C* and *D*; $R_P = 0.85$ and 0.84, respectively, excluding K96E). This observation suggests that the mutant RRMs use the enthalpy-entropy compensation mechanism to overcome the effect of changes in surface charge during C/P in the cell. It is generally accepted that the ionic strength in vivo is ∼150 mM, whereas the ionic strength for our in vitro assays was 10 mM $NaPO_4$. Even with an order of magnitude difference in ionic strength between these two conditions, clear correlations are observed. Thus, any effects from electrostatic screening of the SVL RNA and/or RRM does not appear to be significant.

## DISCUSSION

The process of C/P of nascent pre-mRNA is complex, involving numerous proteins that must recognize multiple RNA elements. It has been suggested that the human CSTF2 RRM binds to the downstream U- or G/U-rich regulatory sequence of the pre-mRNA through a single-step process, but the details of specificity or the precise mechanism of binding are not clear (11). The importance of electrostatic interactions to protein structure, folding, ligand binding, and catalysis is highly realized (16,34–36). Here, we studied the role of electrostatic interactions in the

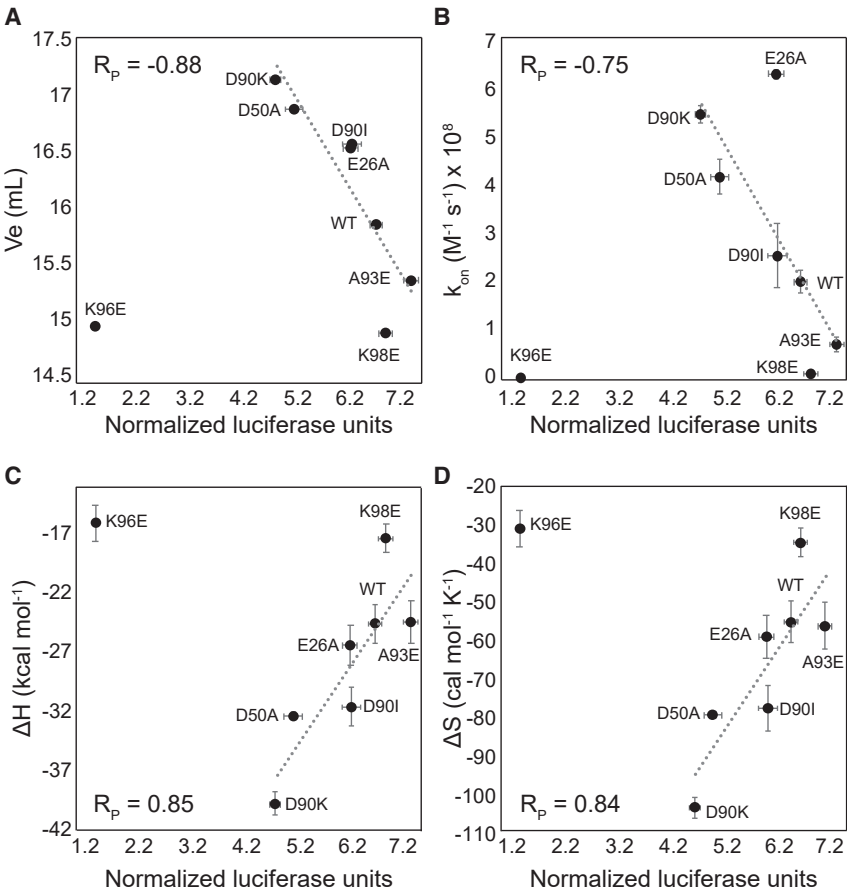


FIGURE 6 In vivo 3′-end cleavage and polyadenylation depend on the RRM surface charge. Correlations of Ve (*A*), $k_{on}$ (*B*), ΔH (*C*), and ΔS (*D*) versus SLAP activity (normalized luciferase units). In each graph, the dotted line represents the line of best fit through the data, and $R_P$ is given in a lower corner. Note that data from K96E, which gave a background SLAP signal, was excluded from the correlations.

thermodynamics, kinetics, structure, and dynamics of RRM binding to RNA. To do so, we made point mutations to surface-exposed residues of RRM, altering the charge of single amino acids to change the overall surface charge of the protein (Fig. S1). Some of these residues (e.g., the aspartate at position 50) are conserved in many of the CSTF2 orthologs surveyed; other residues (e.g., E26 and F45) have conservative replacements in most metazoans (Fig. S10). Others are less highly conserved but tend to be similar in physical-chemical properties (e.g., D90, A93, K96, and K98). For the nine mutants examined here, a correlation between surface charge, as measured by interactions of the protein with the matrix of a size exclusion column in the absence of salt, and enthalpy was observed, indicating the importance of electrostatic attractions in RRM RNA binding (Fig. 2 *B*). Changing the surface charge altered the thermodynamics of RNA binding to result in enthalpy-entropy compensation. As a result of this compensation, we observed an unexpected correlation between surface charge and the entropy of binding (Fig. 2 *B* and *C*).

Generally, RRM surface charge mutants did not alter the overall structure of RRM. Furthermore, in each mutant, NMR CSPs were mostly local to the site of the mutation, both in the presence and absence of RNA (Fig. 2 *A*). However, the median CSPs for RNA-bound mutant RRMs revealed changes in the chemical environment for binding sites I and II residues Y25 and N91 as well as residues near F19 and F61 of binding sites II and III (Fig. 3 *B* and *C*). Therefore, changing the surface charge had a specific, targeted effect on RNA binding; this observation could help to further explain a mechanism for RNA binding to the RRM.

The RNA binding kinetics, extracted from NMR titrations, showed a correlation between the net surface charge and both the on- and off-rates. However, the effect of changing the charge was more significant for the on-rate (Fig. 4 *B* and *C*). We noted previously (13) that the effects of the D50A mutation were largely due to the faster on-rate of the mutant, and this result is upheld for the mutants in this study. The larger effect of surface charge on the on-rate suggests that the positive charge of the RRM surface is important for initiating the formation of the RNA-RRM complex (a long, nm-length-scale process); once RNA is bound, electrostatic interactions have a smaller contribution to RRM-RNA dissociation, since RNA association with the RRM is now driven by other interactions in binding pocket (e.g., other polar and/or aromatic ring stacking interactions, a small, Å-length-scale process). Thus, overall, off-rate varies very little with surface charge. Nonetheless, small differences in the off-rate could assist in the selectivity of RNA binding.

Protein conformational dynamics are important for the thermodynamics and kinetics of ligand binding. For example, conformational transitions that control access to the binding/active site may govern the observed kinetics of ligand binding and/or catalysis (37,38). Moreover, protein dynamics contribute to the entropy of binding through differences in the motions of the free and bound states (39–42). In reactions, the net change in entropy (i.e., $\Delta S$) results from changes in conformational distributions from the solvent, the loss of translational and rotational freedom of the ligand with respect to the protein, the internal conformations of the ligand, and the internal conformations of the protein (43). In protein-nucleic acid complexes, significant contributions to $\Delta S$ can arise from release of ions from around the nucleic acid (16,44). Here, we studied the internal conformations of the RRM, via fast timescale amide $^{15}N$ dynamics, upon mutation and observed a correlation between conformational entropy and net change in entropy (Fig. 5 *B*). It is important to note that the role of internal RNA conformational changes should be significant in the overall entropic penalty as well given that the RNA would likely sample a large structural landscape before RRM binding. Also, the release of water from the highly charged surface of the RRM, similar to what was recently described for the barnase-barnstar system (43,45), could contribute to the overall change in entropy upon RNA binding. Nevertheless, the protein conformational entropy term is of particular interest, as it appears to control some protein-protein binding reactions (39–42). Here, we observed an unusually strong enthalpy-entropy compensation across the RRM mutants. The significance of enthalpy-entropy compensation is under debate (46–48), but possible origins stem from changes in flexibility (i.e., entropy) that compensate exactly for changes in enthalpy upon binding (49,50) and/or from a bounded free energy change (46), though other possibilities have been proposed (see refs. in 40). Backbone flexibility data suggest that the RRM C-terminal $\alpha$-helix helps to compensate, in part, for the unfavorable RNA binding entropy (Figs. 5 and S8). Another possible mode for this compensation could be through local unfolding of the C-terminal $\alpha$-helix, which an analysis of C$\alpha$/C$\beta$ chemical shifts in the bound state or of residual dipolar coupling data could reveal. A possible example of local unfolding might be the K98E mutant, where the NMR relaxation data was not well fit to the wild-type RRM structure. Although it is unclear how general enthalpy-entropy compensation is among other RRM-containing proteins, other RRMs, like U1A, Sx1, PABN1, and nucleolin, also contain a C-terminal $\alpha$-helix that could participate in a compensation mechanism (19,51–53). The C-terminal $\alpha$-helix of the CSTF2 RRM leads into the structure of a hinge domain, which folds around an $\alpha$-helix from the CSTF3 monkey tail domain (54). It is interesting to consider how the effects of RNA binding to the RRM may translate to other members of the CSTF complex. Finally, the high compensation in the CSTF2 RRM might explain why there are a very limited number of known RRM mutations associated with a disease. Perhaps the dynamic structure can absorb the effect of mutations without leading to large differences in the $K_d$ for

RNA. Enthalpy-entropy compensation can also be interpreted as changes in the mechanism of binding (42,55,56); however, we hypothesized that the conformational entropy within the CSTF2 RRM might be the response to the binding specificity of RNA sequences. In fact, our observations compare favorably with those previously reported by Wand and co-workers for $Ca^{2+}$-calmodulin binding to different peptides (39–41). Differences in conformational entropy are thought to help $Ca^{2+}$-calmodulin bind to a variety (>200) of hydrophobic protein sequences.

Based on our biophysical and structural results, it was expected that increasing the positive surface potential of the RRM would lead to an increase in SLAP luciferase signal, because more CSTF2 would be bound to the RNA leading to more C/P. However, it might also be expected that no differences would be observed, since the MCP-RNA stem loop interactions should be the dominant interaction for SLAP. The analysis of SLAP results can also be complicated by the potential disruption of the many interactions that must take place within the CstF complex and the C/P machinery as a whole. Despite those arguments, our previous data demonstrated that altering the RRM changed the SLAP signal in a manner consistent with altered RRM characteristics (9,13). Here, we observed a negative correlation between charge and SLAP signal (Fig. 6) where the more negative mutants produced the greatest signal. We hypothesize that this negative correlation is the result of competition between RRM and MCP associating with the RNA. When the on-rate for the RRM increases due to an increased positive electrostatic potential, RRM drives the association with the nascent pre-mRNA, leading to a decrease in MCP-stem loop interactions and lower SLAP signal. On the other hand, when the on-rate decreases because the RRM is made to be more negatively charged, the MCP-stem loop interaction will dominate, generating a higher SLAP signal. We suggest that in the cell an RNA binding competition always exists for the CSTF2 RRM domain, since it must discriminate against various alternative RNA sequences during co-transcriptional C/P. In fact, mouse studies on the analogous *CSTF2*$^{D50A}$ mutation revealed nearly 1300 genes with altered 3′-ends and a ∼two-fold enrichment in shortened 3′-ends (13). For the wild-type protein, the kinetics for RNA binding is assisted by the dynamics of the RRM domain, which, within the context of the overall timing for C/P, leads to the polyadenylation site choice. However, when these kinetics are disrupted by mutation (i.e., D50A), the site choice is altered and a disease state occurs (13).

## CONCLUSION

In conclusion, we have investigated the role of electrostatic interactions in the binding of the RRM from human CSTF2 to a naturally occurring U-rich RNA sequence. In doing so, we demonstrated that the overall surface charge of the RRM drives the association of RRM and RNA and plays a role in stabilizing the bound state. At the same time, this panel of surface charge mutants exposed an enthalpy-entropy compensation mechanism that is promoted by flexibility within the RRM. We find that the enthalpy-entropy compensation observed upon changing the surface charge alters the C/P in vivo. We hypothesize that this mechanism assists the RRM in recognizing of a variety of U- and G/U-rich sequences in alternative C/P, thereby increasing the diversity of gene products.

## SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.01.005.

## AUTHOR CONTRIBUTIONS

E.M. and M.P.L. conceived research project. E.M, P.N.G., and A.J. performed experiments and analyzed data. C.C.M. and M.P.L. supervised the project. E.M. and M.P.L. wrote the manuscript, and all authors participated in discussion of the results and editing of the manuscript.

## ACKNOWLEDGMENTS

We would like to thank Marella Canny (TTU) and Dr. Shiva Moaven (TTU) for technical and intellectual contributions. This work was supported by NIH grant 1R35GM128906 (M.P.L.); the Presidents' Collaborative Research Initiative of Texas Tech University System (P.N.G., C.C.M, and M.P.L.); the Department of Cell Biology and Biochemistry, Texas Tech University Health Sciences Center (P.N.G. and C.C.M.); South Plains Foundation (P.N.G. and C.C.M.); and the Welch Summer Scholars Program (A.J.).

## REFERENCES

1. Glisovic, T., J. L. Bachorik, …, G. Dreyfuss. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 582:1977–1986.
2. Barabino, S. M. L., and W. Keller. 1999. Last but not least: regulated poly(A) tail formation. *Cell.* 99:9–11.
3. Zhao, J., L. Hyman, and C. Moore. 1999. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* 63:405–445.
4. Beaudoing, E., S. Freier, …, D. Gautheret. 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 10:1001–1010.
5. Martin, G., A. R. Gruber, …, M. Zavolan. 2012. Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. *Cell Rep.* 1:753–763.
6. Beyer, K., T. Dandekar, and W. Keller. 1997. RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *J. Biol. Chem.* 272:26769–26779.
7. MacDonald, C. C. 2019. Tissue-specific mechanisms of alternative polyadenylation: testis, brain, and beyond (2018 update). *Wiley Interdiscip. Rev. RNA.* 10:e1526.
8. MacDonald, C. C., J. Wilusz, and T. Shenk. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs

downstream of the cleavage site and influences cleavage site location. *Mol. Cell. Biol.* 14:6647–6654.

9. Grozdanov, P. N., E. Masoumzadeh, …, C. C. MacDonald. 2018. The structural basis of CstF-77 modulation of cleavage and polyadenylation through stimulation of CstF-64 activity. *Nucleic Acids Res.* 46:12022–12039.
10. Yang, W., P. L. Hsu, …, G. Varani. 2018. Reconstitution of the CstF complex unveils a regulatory rolefor CstF-50 in recognition of 3-end processing signals. *Nucleic Acids Res.* 46:493–503.
11. Pancevac, C., D. C. Goldstone, …, I. A. Taylor. 2010. Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors. *Nucleic Acids Res.* 38:3119–3132.
12. Pérez Cañadillas, J. M., and G. Varani. 2003. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.* 22:2821–2830.
13. Grozdanov, P. N., E. Masoumzadeh, …, C. C. MacDonald. 2020. A missense mutation in the CSTF2 gene that impairs the function of the RNA recognition motif and causes defects in 3′ end processing is associated with intellectual disability in humans. *Nucleic Acids Res.* 48:9804–9821.
14. Tao, J., and A. D. Frankel. 1993. Electrostatic interactions modulate the RNA-binding and transactivation specificities of the human immunodeficiency virus and simian immunodeficiency virus Tat proteins. *Proc. Natl. Acad. Sci. U S A.* 90:1571–1575.
15. Corley, M., M. C. Burns, and G. W. Yeo. 2020. How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell.* 78:9–29.
16. Yu, B., B. M. Pettitt, and J. Iwahara. 2020. Dynamics of ionic interactions at protein-nucleic acid interfaces. *Acc. Chem. Res.* 53:1802–1810.
17. Katsamba, P. S., D. G. Myszka, and I. A. Laird-Offringa. 2001. Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. *J. Biol. Chem.* 276:21476–21481.
18. Tang, Y., and L. Nilsson. 1999. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* 77:1284–1305.
19. Varani, L., S. I. Gunderson, …, G. Varani. 2000. The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat. Struct. Biol.* 7:329–335.
20. Nagai, K., C. Oubridge, …, P. R. Evans. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature.* 348:515–520.
21. Law, M. J., M. E. Linde, …, I. A. Laird-Offringa. 2006. The role of positively charged amino acids and electrostatic interactions in the complex of U1A protein and U1 hairpin II RNA. *Nucleic Acids Res.* 34:275–285.
22. Azatian, S. B., N. Kaur, and M. P. Latham. 2019. Increasing the buffering capacity of minimal media leads to higher protein yield. *J. Biomol. NMR.* 73:11–17.
23. Delaglio, F., S. Grzesiek, …, A. Bax. 1995. A multidimensional spectral processing system based on pipes. *J. Biomol. NMR.* 6:277–293.
24. Vranken, W. F., W. Boucher, …, E. D. Laue. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* 59:687–696.
25. Waudby, C. A., A. Ramos, …, J. Christodoulou. 2016. Two-dimensional NMR lineshape analysis. *Sci. Rep.* 6:1–8.
26. Barbato, G., M. Ikura, …, A. Bax. 1992. Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry.* 31:5269–5278.
27. Kay, L. E., L. K. Nicholson, …, D. A. Torchia. 1992. Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T1 and T2 values in proteins. *J. Magn. Reson.* 97:359–375.
28. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.
29. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559–4570.
30. Sharp, K. A., E. O'Brien, …, A. J. Wand. 2015. On the relationship between NMR-derived amide order parameters and protein backbone entropy changes. *Proteins.* 83:922–930.
31. Zhu, Y., M. Potschka, …, C. Cai. 2001. A method for the quantitation of charge by size exclusion chromatography demonstrated with components of ficoll 400. *Macromol. Chem. Phys.* 202:61–72.
32. Potschka, M. 1988. Size-exclusion chromatography of polyelectrolytes: experimental evidence for a general mechanism. *J. Chromatogr. A.* 441:239–260.
33. Hockert, J. A., H. J. Yeh, and C. C. MacDonald. 2010. The hinge domain of the cleavage stimulation factor protein CstF-64 is essential for CstF-77 interaction, nuclear localization, and polyadenylation. *J. Biol. Chem.* 285:695–704.
34. Zhou, H. X., and X. Pang. 2018. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* 118:1691–1741.
35. Warshel, A., P. K. Sharma, …, M. H. M. Olsson. 2006. Electrostatic basis for enzyme catalysis. *Chem. Rev.* 106:3210–3235.
36. Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.
37. Eisenmesser, E. Z., D. A. Bosco, …, D. Kern. 2002. Enzyme dynamics during catalysis. *Science.* 295:1520–1523.
38. Boehr, D. D., D. McElheny, …, P. E. Wright. 2006. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science.* 313:1638–1642.
39. Frederick, K. K., M. S. Marlow, …, A. J. Wand. 2007. Conformational entropy in molecular recognition by proteins. *Nature.* 448:325–329.
40. Lee, A. L., S. A. Kinnear, and A. J. Wand. 2000. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat. Struct. Biol.* 7:72–77.
41. Marlow, M. S., J. Dogan, …, A. J. Wand. 2010. The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* 6:352–358.
42. Tzeng, S.-R., and C. G. Kalodimos. 2012. Protein activity regulation by conformational entropy. *Nature.* 488:236–240.
43. Wand, A. J., and K. A. Sharp. 2018. Measuring entropy in molecular recognition by proteins. *Annu. Rev. Biophys.* 47:41–61.
44. Pletka, C. C., R. Nepravishta, and J. Iwahara. 2020. Detecting counterion dynamics in DNA–protein association. *Angew. Chem. Int. Ed. Engl.* 59:1465–1468.
45. Caro, J. A., K. W. Harpole, …, A. J. Wand. 2017. Entropy in molecular recognition by proteins. *Proc. Natl. Acad. Sci. U S A.* 114:6563–6568.
46. Sharp, K. 2001. Entropy–enthalpy compensation: fact or artifact? *Protein Sci.* 10:661–667.
47. Fox, J. M., M. Zhao, …, G. M. Whitesides. 2018. The molecular origin of enthalpy/entropy compensation in Biomolecular recognition. *Annu. Rev. Biophys.* 47:223–250.
48. Pan, A., T. Biswas, …, S. P. Moulik. 2015. Enthalpy-entropy compensation (EEC) effect: a revisit. *J. Phys. Chem. B.* 119:15876–15884.
49. Weber, G. 1995. Van't Hoff revisited: enthalpy of association of protein subunits. *J. Phys. Chem.* 99:1052–1059.
50. Dunitz, J. D. 1995. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* 2:709–712.
51. Handa, N., O. Nureki, …, S. Yokoyama. 1999. Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature.* 398:579–585.
52. Deo, R. C., J. B. Bonanno, …, S. K. Burley. 1999. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell.* 98:835–845.

53. Allain, F. H.-T., P. Bouvet, …, J. Feigon. 2000. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* 19:6870–6881.

54. Moreno-Morcillo, M., L. Minvielle-Sébastia, …, C. D. MacKereth. 2011. Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CF IA complex. *Structure*. 19:534–545.

55. Koch, C., A. Heine, and G. Klebe. 2011. Tracing the detail: how mutations affect binding modes and thermodynamic signatures of closely related aldose reductase inhibitors. *J. Mol. Biol.* 406:700–712.

56. King, N. M., M. Prabu-Jeyabalan, …, C. A. Schiffer. 2012. Extreme entropy-enthalpy compensation in a drug-resistant variant of HIV-1 protease. *ACS Chem. Biol.* 7:1536–1546.