Rules of RNA specificity of hnRNP A1 revealed by global and quantitative analysis of its affinity distribution

Niyati Jain[a,1], Hsuan-Chun Lin[b,1], Christopher E. Morgan[a], Michael E. Harris[b,2], and Blanton S. Tolbert[a,2]

[a]Department of Chemistry, Case Western Reserve University, Cleveland, OH 44106; and [b]Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, OH 44106

Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) is a multipurpose RNA-binding protein (RBP) involved in normal and pathological RNA metabolism. Transcriptome-wide mapping and in vitro evolution identify consensus hnRNP A1 binding motifs; however, such data do not reveal how surrounding RNA sequence and structural context modulate affinity. We determined the affinity of hnRNP A1 for all possible sequence variants ($n = 16,384$) of the HIV exon splicing silencer 3 (ESS3) 7-nt apical loop. Analysis of the affinity distribution identifies the optimal motif 5′-YAG-3′ and shows how its copy number, position in the loop, and loop structure modulate affinity. For a subset of ESS3 variants, we show that specificity is determined by association rate constants and that variants lacking the minimal sequence motif bind competitively with consensus RNA. Thus, the results reveal general rules of specificity of hnRNP A1 and provide a quantitative framework for understanding how it discriminates between alternative competing RNA ligands in vivo.

hnRNP A1 | protein–RNA specificity | RNA structure | thermodynamics | binding kinetics

Gene expression is regulated by an ensemble of protein–RNA complexes that assemble and disassemble throughout the lifetime of a transcript (1–5). Knowledge of the determinants of RNA-binding protein (RBP) specificity is therefore essential to understanding the relative affinity for sites of association within the transcriptome. A characteristic feature of many RBPs is their ability to elicit biological function by binding at sites in RNAs that vary significantly in sequence and local structure. This broad specificity of RBPs challenges the simplistic description of RNA binding as either specific or nonspecific. Global profiling methods provide consensus sequence motifs that represent preferential binding sites by RBPs in the transcriptome (6–11). Although powerful, such approaches are usually nonquantitative. Moreover, they do not readily provide information on how surrounding sequence identity or positioning of preferred sequences within higher order structure alters affinity. Methods such as high-throughput sequencing kinetics (HTS-KIN), RNA Bind-n-Seq, and RNA MaP have recently been developed that allow the affinities and reaction kinetics of thousands of RNAs to be measured simultaneously (6, 12–15). These methods can potentially provide global information on the surrounding context of a given sequence motif; however, they have yet to see wide application for structure/function studies of RNA specificity.

Heterogeneous nuclear ribonucleoproteins (hnRNPs) represent a diverse family of RBPs that are implicated at most stages of posttranscriptional gene regulation (16, 17). The prototypical member of this family, hnRNP A1, regulates alternative splicing, nuclear export, translation, and other RNA processing events (17). HnRNP A1 has a modular domain organization consisting of tandem RNA recognition motifs (RRMs), collectively referred to as unwinding protein 1 (UP1), and an intrinsically disordered RGG-rich C terminus. UP1 is the primary RNA binding domain, whereas the C terminus mediates homologous and heterogeneous protein–protein interactions. The individual RRMs of UP1 have a

high degree of sequence homology and adopt nearly identical 3D structures (18–20). The hnRNPs, including A1, are thought to bind to pre-mRNAs combinatorially as dictated by their intrinsic specificity for RNA sequence and structure (11, 21). Thus, a comprehensive and quantitative description of hnRNP A1 specificity is required to understand its function in gene expression.

Early studies demonstrated that hnRNP A1 binds selectively to single-stranded RNA and the human telomeric sequence $(TTAGGG)_n$ (22). Selection amplification experiment(s) (SELEX) identified a high-affinity motif 5′-UAGGGA/U-3′ that now serves as a benchmark for sequence-specific hnRNP A1 binding (23). Transcriptome-wide studies of functional hnRNP A1 binding sites using motif finder algorithms further indicate specificity for a 5′-UAG-3′ motif (11, 24). In previous work, we determined a small angle X-ray scattering (SAXS) model of UP1 in complex with the HIV ESS3 stem loop ($SL3^{ESS3}$), which was guided by a 1.92-Å crystal structure of UP1 bound to a 5′-AGU-3′ oligomer (25). The structure indicates that RRM1 and the inter-RRM linker fold to form a nucleobase pocket that specifically recognizes 5′-AG-3′, which is part of the 7-nt $SL3^{ESS3}$ apical loop (5′-GAUUAGU-3′) (Fig. 1). Moreover, single- or double-cytosine substitutions within the apical loop reduce UP1 binding affinity (upward of 20-fold) by slowing down the rate of complex formation (26).

Significance

Human heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) is a key protein that functions in RNA metabolism both under normal and diseased cellular conditions. Despite playing a central role in eukaryotic biology, the mechanisms by which hnRNP A1 discriminates between cognate and noncognate RNA targets remain poorly understood. Here, we combined high-throughput sequencing analysis of equilibrium binding (HTS-EQ) experiments with independent biophysical measurements to reveal the complete hnRNP A1 specificity landscape. The data show that RNA sequence, motif copy number, spacing, and secondary structure determine specificity by modulating rates of productive hnRNP A1-RNA encounters. Thus, our work provides significant insights into the combinatorial factors that determine how hnRNP A1 identifies functional binding sites in vivo.
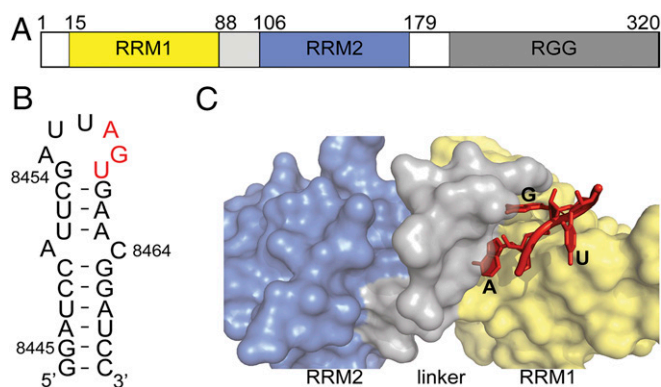
Fig. 1. UP1 domain of hnRNP A1 binds specifically to 5′-AG-3′ through its nucleobase pocket. (*A*) Domain organization of full-length hnRNP A1. (*B*) Secondary structure of the HIV ESS3 (NL4-3 strain) stem loop wherein apical loop residues involved in UP1 binding are colored red. (*C*) Surface representation of UP1 bound to 5′-AGU-3′ (red). UP1 color coding is the same as in *A*.

To evaluate the distribution of hnRNP A1 among the sea of alternative binding sites in the transcriptome requires a more complete understanding of specificity. Here, we comprehensively evaluate the determinants of specific UP1–SL3^{ESS3} interactions by measuring relative equilibrium association constants ($K_{A,rel}$) for all variants of a fully randomized 7-nt SL3^{ESS3} apical loop ($n = 16,384$). To achieve this goal, we used high-throughput sequencing analysis of equilibrium binding (HTS-EQ), a newly developed method that uses Illumina sequencing to quantify the protein or enzyme binding of thousands of RNAs simultaneously. Quantitative analyses of the resulting affinity distribution from HTS-EQ further establishes the consensus motif 5′-YAG-3′, but also comprehensively reveal how affinity for alternative RNAs is controlled by its position in the loop, spacing when two motifs are present, and the presence of unfavorable secondary structure. For a subset of SL3^{ESS3} variants studied, we further show that discrimination is primarily controlled by the association rate constant, and that variants lacking the 5′-YAG-3′ motif nonetheless compete with consensus RNA for binding. Thus, the results comprehensively reveal the rules of specificity of hnRNP A1, establish a mechanistic basis for discrimination, and provide a framework for understanding how it discriminates between alternative competing RNA targets in vivo.

Results

HTS-EQ Reveals the Global Affinity Distribution of UP1-SL3^{ESS3}. To evaluate the determinants of UP1-SL3^{ESS3} specificity comprehensively, we randomized the 7-nt SL3^{ESS3} apical loop and performed HTS-EQ (27) (Fig. 2*A*). HTS-EQ involves binding of a randomized RNA population to increasing concentrations of UP1 under equilibrium conditions, followed by separation of the free and bound RNAs by EMSA (HTS-EQ work flow is shown in Fig. S1). By analyzing the change in distribution of individual sequence variants quantified by Illumina sequencing as a function of RBP concentration allows measurement of the $K_{A,rel}$ values for all possible sequence variants calibrated to native SL3^{ESS3} [$K_{A,rel} = K_A(NNNNNNN)/K_A(GAUUAGU)$]. Variants that bind weaker than the native RNA have $K_{A,rel}$ values <1, whereas variants that bind stronger have $K_{A,rel}$ values >1.

Fig. 2*B* shows the resulting affinity distribution [i.e., distribution of the number of sequences among the range of observed $K_{A,rel}$ (ln$K_{A,rel}$) values]. The $K_{A,rel}$ values are best visualized as the natural log so that the magnitude reflects relative free energies between sequence variants. The histogram reveals native SL3^{ESS3} is optimized to bind UP1 because it lies toward the high-affinity region of the distribution; however, 722 variants bind with ln$K_{A,rel}$ values greater than the native sequence. Of the 722 variants, 72%

contain at least one 5′-AG-3′ motif and 98% of the top 50 variants each have a 5′-AG-3′. Calculation of an optimal sequence logo from the top 50 variants (Fig. 2*C*) shows an overall preference for U, A, and G, which is consistent with the base composition of the high-affinity 5′-UAGGGA/U-3′ sequence identified by SELEX (23); however, the SELEX sequence lies slightly to the low-affinity side of native SL3^{ESS3} (ln$K_{A,rel} = -0.24$) in the distribution, indicating the complexity of hnRNP A1 specificity. Nevertheless, HTS-EQ clearly reveals that the signature of a specific hnRNP A1 target minimally contains a 5′-AG-3′.

The relative binding affinities for different sequence variants are nonetheless clearly further tuned by the identity of neighboring sequences in the RNA. Such important features of specificity are not fully accounted for by models derived only from optimal sequence variants. To identify the local sequence and structure properties that determine specificity, we fit the entire distribution of $K_{A,rel}$ values to two analytical models of RNA sequence specificity. First, a position weight matrix (PWM) model was used that only considers the identity of a nucleobase at each position in the loop (28). The affinity distribution was also fit to a PWM model that includes interaction coefficients (IC values) between nucleobases that account for positive and negative effects of sequence variation between positions in the binding site (29, 30). The PWM model performs poorly at recapitulating the determinants of UP1-SL3^{ESS3} specificity, whereas inclusion of IC values greatly improves the correspondence between the model and experimental data ($R^2 > 0.7$) (compare Fig. 2*D* and Figs. S2 and S3). Therefore, binding specificity does not solely involve the selection of optimal AG nucleobases at individual positions in the loop but also must include additional contributions from the surrounding sequence context.
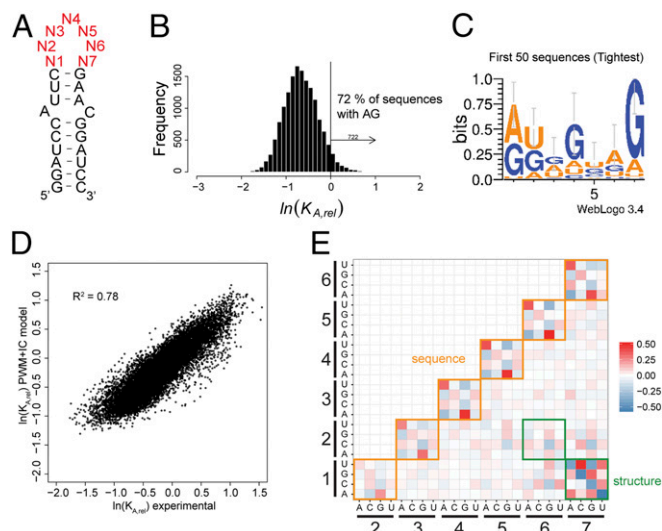

Fig. 2. Determination of the affinity distribution for UP1 binding to all possible variants of the 7-nt SL3^{ESS3} apical loop. (*A*) Secondary structure of SL3^{ESS3} showing the constant base-paired stem loop and the randomized 7-nt apical loop region (red). (*B*) Affinity distribution of $K_{A,rel}$ values for UP1 binding to all possible sequence variants in the randomized pool of SL3^{ESS3} variants. The vertical line corresponds to the value for the reference native SL3^{ESS3} (ln$K_{A,rel} = 0$). Seven hundred twenty-two variants bind with higher affinities than native SL3^{ESS3}, of which 72% contain at least one 5′-AG-3′ dinucleotide. (*C*) Probability sequence logo calculated for the 50 highest affinity SL3^{ESS3} variants. (*D*) Comparison of experimental $K_{A,rel}$ values and $K_{A,rel}$ values predicted using a PWM model that includes interaction coefficients (IC values) between nucleobases. (*E*) Comparison of the IC values derived from fitting the $K_{A,rel}$ affinity distribution to Eq. S7 as described in *SI Materials and Methods*. The magnitude of the IC values is indicated by differences in color: red, positive interactions; blue, negative interactions.

BIOCHEMISTRY

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Recasting the IC values as a heat map offers insights into sequence and structure that modulate binding affinity (Fig. 2E). For example, negative IC values are observed between N1–N7 and N2–N6 for sequence combinations that can form Watson–Crick pairs. This result indicates that base-pairing interactions that reduce the loop size are highly unfavorable. Also, the subsets of variants containing stable loop sequences UUCG, GNRA, and UNCG have distributions that are shifted to lower $K_{A,rel}$ values (Fig. S4). In contrast, the heat map shows positive couplings for 5′-UA-3′ and 5′-AG-3′ dinucleotides as well as, to a lesser extent, 5′-CA-3′ dinucleotides at adjacent positions in the loop (Fig. 2E).

The affinity distribution of variants containing a 5′-AG-3′ but lacking an unfavorable secondary structure is clearly shifted to higher $K_{A,rel}$ values compared with AG variants that contain secondary structure (Fig. 3A). Approximately 200 of 722 variants with $K_{A,rel}$ values >1 lack at least one AG in the loop. To identify the positive determinants at adjacent positions, high-affinity variants were aligned according to the AG dinucleotide in their sequence regardless of its position in the loop (Fig. 3B). A sequence probability logo calculated from this alignment further identifies C/U located 5′ to 5′-AG-3′ and reveals a G positioned 3′ as an additional positive determinant. A probability logo of high-affinity variants ($K_{A,rel} > 1$) lacking 5′-AG-3′ shows a polyG pattern (Fig. 3C). Nonetheless, the optimal specificity determinant is clearly a 5′-YAG-3′ motif that may occur at multiple positions in the loop. Comparison of the $K_{A,rel}$ values for this motif at each possible position in the loop reveals that UAG is optimal at N5–N7 (Fig. 3 E and F). This observation contrasts with the position of

UAG at N4–N6 in the native HIV ESS3 loop (Fig. 3E). Sequence variants with two 5′-AG-3′ dinucleotides, on average, bind with higher affinity than the native sequence; however, an optimal spacing of at least two nucleotides is required for the additive effect to be observed (Fig. 3D). Thus, analysis of the $K_{A,rel}$ affinity distribution reveals an unfavorable loop structure; establishes 5′-YAG-3′ as the minimal sequence specificity motif; and, most importantly, defines how local sequence context modulates the contribution of this motif to affinity.

Analysis of the Thermodynamic and Kinetic Basis for UP1-SL3^{ESS3} Specificity. To understand the thermodynamic contributions to UP1 specificity, we randomly selected SL3^{ESS3} variants from different regions of the affinity distribution and quantitatively analyzed their binding properties by calorimetric titrations (Fig. 4A and Fig. S5). The $K_{A,rel}$ values measured by HTS-EQ and calorimetry correlate for the selected SL3^{ESS3} variants (Fig. 4B and Fig. S5), indicating HTS-EQ provides an accurate depiction of the distribution of binding affinities. Although the $K_{A,rel}$ correlation between HTS-EQ and isothermal titration calorimetry (ITC) is good ($R^2 = 0.85$), a few outliers are observed (Fig. S5). Variations in sample quality for ITC and inherent noise in the high-throughout sequencing data that increases with lower affinity substrates are likely major contributors to the ~15% discrepancy.

Consistent with observations from HTS-EQ, UP1 binds SL3^{ESS3} variants that contain at least one 5′-AG-3′ with greater affinity than non-AG variants, and sequences with two 5′-AG-3′ motifs bind slightly stronger than sequences with just one, likely reflecting a higher probability for UP1 to interact productively with these variants. On average, SL3^{ESS3} variants with no 5′-AG-3′ bind approximately sevenfold weaker than native SL3^{ESS3}. This observation is consistent with our previous results using single-cytosine substitutions, wherein UP1 binds the loop sequence 5′-GAUU-CGU-3′ 10-fold weaker than 5′-GAUUAGU-3′ (26). The thermodynamic signatures reveal that the reduction in binding affinity for non-AG variants manifests primarily as a decrease in favorable association enthalpy ($\Delta\Delta H = 8$ kcal/mol).

As predicted from the sequence specificity model derived from fitting the affinity distribution measured by HTS-EQ, some variants form new Watson–Crick base pairs within the apical loop that are inhibitory to binding (Fig. 2E). Indeed, ^{1}H NMR of the imino region of 5′-GGUGACC-3′ and 5′-UGUGGCA-3′ (proposed pairing interactions underlined here and throughout remaining text) show new imino signals with chemical shifts consistent with GC and AU base pairs (Fig. S6). The new base pairs reduce the loop from 7 to 3 nt. Calorimetric titrations of UP1 with 5′-GGUGACC-3′ and 5′-UGUGGCA-3′ show binding affinities are weaker relative to native SL3^{ESS3} as a consequence of a significant loss of binding enthalpy (~$\Delta\Delta H = 15$ kcal/mol).

We next investigated the kinetic contributions to specificity by performing biolayer interferometry (BLI) studies with selected SL3^{ESS3} variants (Fig. 4 D–F, Fig. S7, and Table S1). Kinetic analysis shows that all six variants dissociate from UP1 with comparable apparent first-order off-rate constants ($k_{off, app} = 4.5–6.2 \times 10^{-3}$ s^{-1}), whereas the apparent second-order on-rate constants for association ($k_{on, app} = 0.3–1.1 \times 10^5$ M$^{-1}\cdot$s^{-1}) show a modest but larger sequence dependence. SL3^{ESS3} variants with one 5′-AG-3′ bind with slightly slower $k_{on, app}$ values compared with 5′-GUAGGAG-3′; however, non-AG variants bind, on average, 2.5-fold slower, and 5′-UGUGGCA-3′ binds approximately fourfold slower. Decreases in $k_{on, app}$ values were previously determined as the predominant mechanism that contributes to weaker affinities for cytosine-substituted SL3^{ESS3} constructs, and similar observations have recently been made for other protein–RNA systems (13, 25, 27). Thus, for some SL3^{ESS3} variants, specificity is driven, in large part, by the frequency with which UP1 collides productively with a cognate RNA that minimally exposes a conformationally flexible 5′-AG-3′ motif. However, a more detailed kinetic
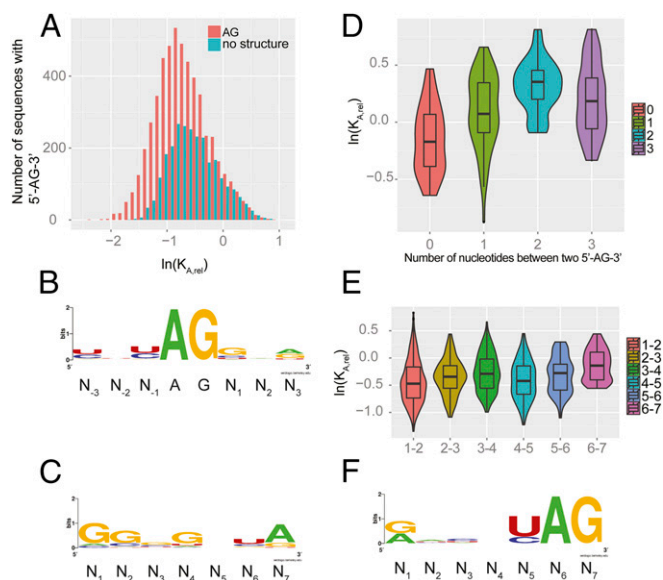

Fig. 3. Analysis of sequence and structure attributes that modulate binding affinity of hnRNP A1. (A) Histogram comparison of the $K_{A,rel}$ affinity distributions of sequence variants with a 5′-AG-3′ dinucleotide motif (AG, red) and the subset of these sequences filtered to remove sequence variants with predicted base-pairing interactions between N1N2 and N6N7 (no structure; steel blue). (B) Sequence probability logo for the highest affinity ($K_{A,rel} > 1$) variants aligned with respect to the position of the 5′-AG-3′ in their sequence. Because all of the variants included in the analysis contained an AG at the aligned positions, their binary probability is 2 by definition. (C) Sequence probability logo of the highest affinity SL3^{ESS3} loop variants that lack a 5′-AG-3′ dinucleotide in their sequence. (D) Violin plot of the distributions of $K_{A,rel}$ values for SL3^{ESS3} loop sequence variants with two 5′-AG-3′ motifs plotted according to the number of nucleotides between the two AG sequences. (E) Violin plot of the probability density of $K_{A,rel}$ values for SL3^{ESS3} loop sequence variants with only one 5′-AG-3′ motif plotted according to the position of the AG in the loop (1–2 = A1G2; 2–3 = A2G3, etc.). (F) Sequence probability logo for highest affinity SL3^{ESS3} loop sequence variants with A6G7.
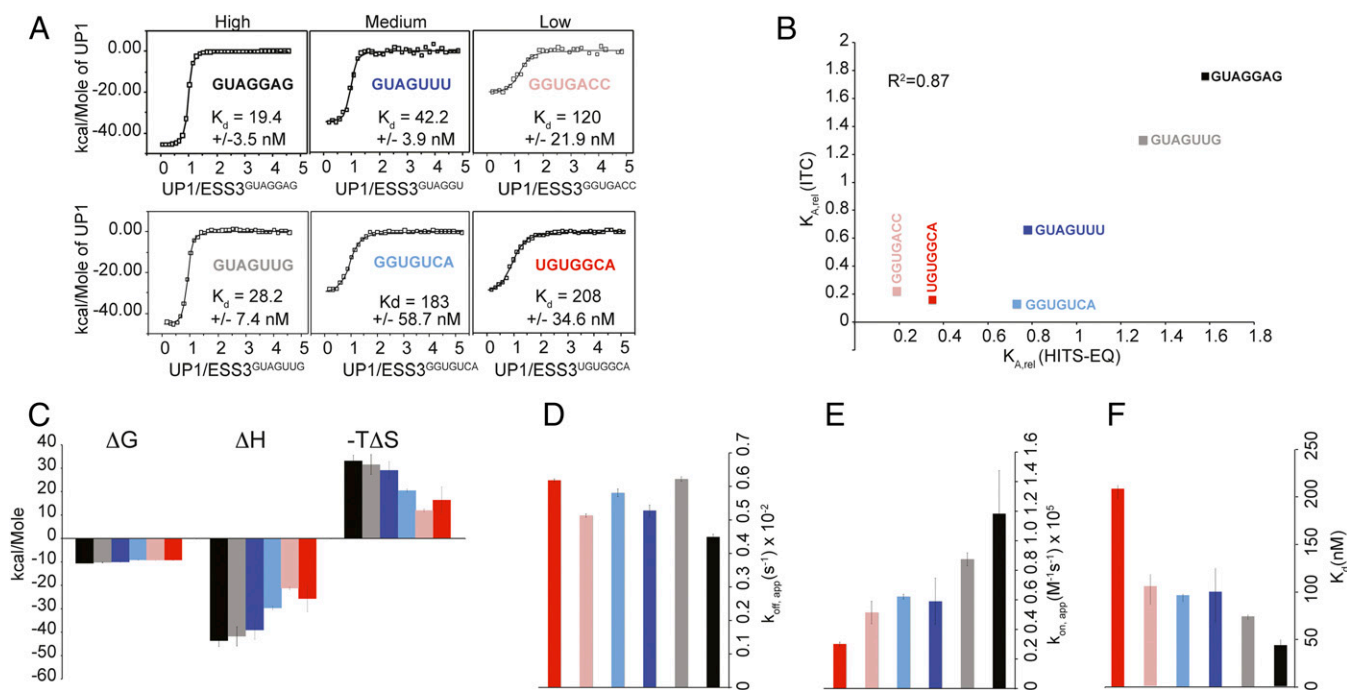
Jain et al.

www.manaraa.com

Fig. 4. Independent biophysical measurements correlate with HTS-EQ and reveal insights into the binding mechanism. (*A*) Six representative isotherms for UP1 titrated into SL3^{ESS3} variants from the three regions of the affinity distribution: high, medium, and low. Titrations were performed at 298 K, 140 mM K^{+}, and pH 6.5. Processed isotherms were fit to a 1:1 binding model. Reported dissociation constants correspond to average values of two measurements. (*B*) Comparison of the relative binding affinities measured by HTS-EQ and ITC for the selected SL3^{ESS3} variants. The relative binding affinities for each variant are calibrated to native SL3^{ESS3} (GAUUAGU). The plot shows that the two methods provide comparable information regarding the relative binding affinities and that HTS-EQ provides a true depiction of specificity. Fig. S5 shows the comparison between HTS-EQ and ITC for all 15 SL3^{ESS3} variants ($R^2 = 0.85$). (*C*) Comparison of the thermodynamic signatures of UP1 titrations with SL3^{ESS3} variants. Thermodynamic values represent average ± SD for two replicates. (*D* and *E*) Summary of kinetic binding parameters for UP1 with select SL3^{ESS3} variants reveals the predominant driver of specificity is a change in association rates, whereas dissociation rates are mostly unperturbed. Kinetic values represent average ± SEM for two replicates. (*F*) Dissociation constants for UP1 with select SL3^{ESS3} variants as determined from kinetic binding parameters. Note that the trend in the equilibrium dissociation constants measured by BLI and ITC correlate but are, on average, 1.5-fold different.

study of SL3^{ESS3} variants is needed to understand fully if the differences in $k_{on, app}$ values result from possible intermediates in the binding pathway. Nevertheless, the trend in the equilibrium dissociation constants measured independently by calorimetric titrations (Fig. 4*A*) and BLI (Fig. 4*F*) correlate, indicating that the kinetic parameters indeed offer some insights into the mechanisms by which hnRNP A1 discriminates cognate from noncognate RNA targets.

UP1 Uses the Same Binding Surface to Interact with 5′-AG-3′ and Non–AG-Containing RNAs. We previously showed UP1 interacts with native SL3^{ESS3} through its RRM1 domain and inter-RRM linker to form a 1:1 complex (25). Although capable of binding RNA, RRM2 is not considered to be the preferred high-affinity binding surface. To test whether SL3^{ESS3} variants compete for the same surface on UP1, we carried out competitive calorimetric titrations with SL3^{ESS3} loop sequence variants 5′-GUAGGAG-3′ and 5′-GCACUUU-3′. For simplicity, we refer to these RNAs as (AG)SL3^{ESS3} and (AC)SL3^{ESS3}, respectively. Titration of (AC)SL3^{ESS3} with UP1 shows that it binds with a dissociation constant of 344 nM; however, (AC)SL3^{ESS3} does not bind to a preformed UP1–(AG)SL3^{ESS3} complex under these conditions (Fig. 5*A*). This observation is consistent with the ~18-fold stronger binding affinity of UP1 for (AG)SL3^{ESS3} compared with (AC)SL3^{ESS3} (Fig. 5 *A* and *B*). Conversely, (AG)SL3^{ESS3} binds with reduced apparent affinity for UP1 when titrated into a reaction containing preformed UP1–(AC)SL3^{ESS3} complexes (Fig. 5*B*). As expected for competitive binding, the apparent affinity of (AG)SL3^{ESS3} for UP1 decreases as the concentration of (AC)SL3^{ESS3} increases.

It is possible that lower affinity SL3^{ESS3} variants bind to UP1 in a nonspecific manner in which the protein associates weakly with

multiple different sites on the RNA. To test the binding mode directly, ^{1}H-^{13}C heteronuclear multiple quantum coherence (HMQC) titrations were performed using a (^{13}C)U selectively labeled (AC)SL3^{ESS3} construct. Fig. 5*C* shows that despite its lower affinity, UP1 binds (AC)SL3^{ESS3} in a site-specific manner because only a subset of the correlation peaks show significant chemical shift perturbations. In the presence of an equimolar amount of unlabeled (AG)SL3^{ESS3}, the ^{1}H-^{13}C correlation peaks of (AC)SL3^{ESS3} are identical to its unbound form (Fig. 5*D*). Fig. 5*E* shows that adding twofold excess (AG)SL3^{ESS3} to a preformed UP1–(AC)SL3^{ESS3} complex effectively displaces (AC)SL3^{ESS3}. Collectively, the calorimetric and NMR titrations reveal that SL3^{ESS3} variants compete for the same binding surface on UP1 and that UP1 preferentially associates with (AG)SL3^{ESS3} over (AC) SL3^{ESS3} when both are present at similar concentrations. The observation that different SL3^{ESS3} variants compete for the same surface on UP1 is significant because it verifies that the relative binding affinities determined from HTS-EQ reflect the same binding mode.

The UP1 HTS-EQ Affinity Distribution Allows Prediction of Potential microRNA Targets. The RNA specificity model derived from the UP1-SL3^{ESS3} HTS-EQ affinity distribution reveals sequence and structure attributes that should extend to other biologically relevant stem loop systems. To test the predicative capabilities, we applied the UP1-SL3^{ESS3} trained model to microRNAs (miRNAs) derived from miRBase. Fig. S8 shows the predicted affinity distribution for a subset ($n = 1,193$) of miRNAs that contain hairpin loops >3 nt long (*Materials and Methods*). The distribution indicates that hnRNP A1 has a range of affinities for miRNA targets that reflect both positive and negative contributions of RNA
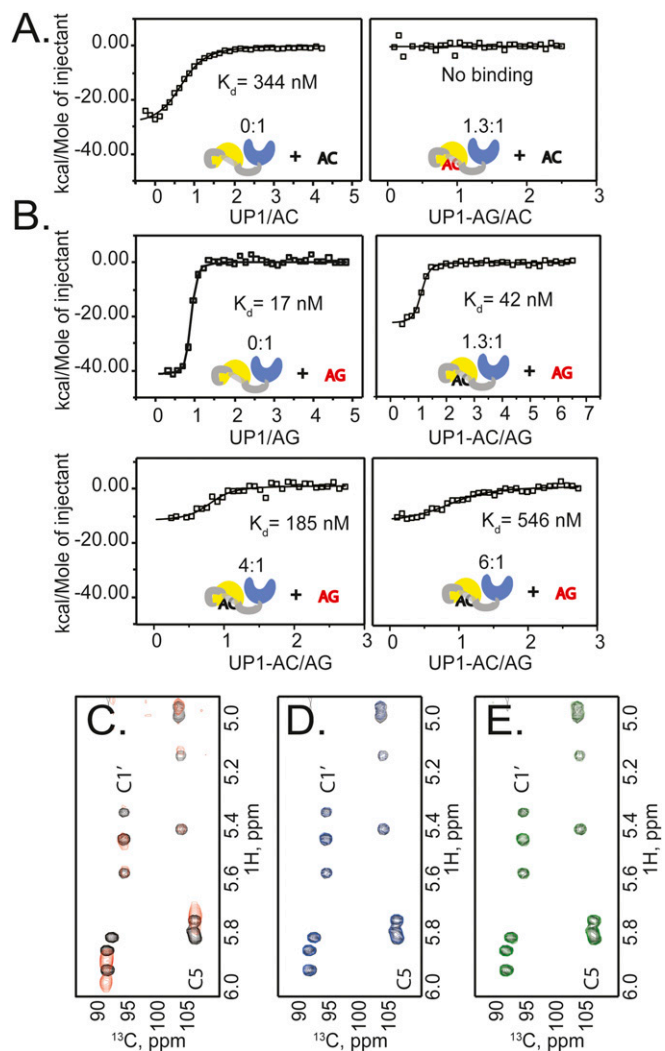
Fig. 5. SL3^ESS3 variants compete for the same UP1 binding surface. (*A*) Calorimetric titrations of (AC)SL3^ESS3 with free UP1 and UP1 preloaded with (AG)SL3^ESS3 at a 1.3:1 (RNA/protein) ratio show that (AC)SL3^ESS3 is unable to bind to a preformed UP1–(AG)SL3^ESS3 complex. (*B*) Calorimetric titrations of (AG)SL3^ESS3 into UP1 preloaded with increasing ratios of (AC)SL3^ESS3 show (AG)SL3^ESS3 can still bind UP1, although with decreasing apparent affinities. (*C*) ^{1}H-^{13}C HMQC titration of UP1 with (^{13}C)U-selectively labeled (AC)SL3^ESS3 indicates UP1 binds site specifically to the apical loop. (*D*) Inclusion of unlabeled (AG)SL3^ESS3 with (^{13}C)U-selectively labeled (AC)SL3^ESS3 at equimolar concentrations reveals UP1 no longer binds (AC)SL3^ESS3. (*E*) Addition of twofold excess of (AC)SL3^ESS3 into UP1 preloaded with (^{13}C)U-selectively labeled (AC)SL3^ESS3 shows (AG)SL3^ESS3 can effectively displace bound (AC)SL3^ESS3. Collectively, these results indicate that SL3^ESS3 variants compete for the same binding surface on UP1 and that UP1 preferentially binds RNA elements with a 5′-AG-3′ motif. Titrations were performed at 298 K, 140 mM K⁺, and pH 6.5.

sequence and structure. Indeed, the model recapitulates trends in relative affinities for 20 of 22 miRNAs for which experimental binding data to hnRNP A1 were determined using a chemiluminescent microtiter assay (31). Significantly, the model predicts the relative affinities of mir-18a (119 nM) and let-7a (12 nM), for both of which biogenesis is regulated by hnRNP A1 (32, 33). Consistent with attributes derived from the PWM + IC model, the hairpin loops of mir-18a and let-7a each contain at least one unpaired 5′-UAG-3′ motif. Thus, the RNA specificity model derived from the UP1-SL3^ESS3 HTS-EQ data can be used to predict other biologically relevant RNA targets of hnRNP A1. Application of the

model to miRNAs hints that hnRNP A1 might be a general accessory factor for a subset of primary miRNAs.

Discussion

RNA sequence and structure contribute to specific protein recognition; therefore, identifying these intrinsic properties is necessary to interpret fully how protein–RNA networks control gene expression. Although they have proven transformative, most transcriptome-wide methods probe RBP specificity without obtaining information on the underlying structural context; they also do not provide quantitative information about relative affinities for alternative binding sites. Here, we applied HTS-EQ, a newly developed high-throughput method that allows measurement of the relative equilibrium constants of an RBP to thousands of RNAs simultaneously (27). We used HTS-EQ to evaluate interactions between the UP1 domain and a randomized pool of HIV SL3^ESS3 variants containing all possible loop sequences. Quantitative sequence specificity modeling of the resulting distribution of binding affinities allowed the intrinsic determinants of hnRNP A1 recognition to be identified. HnRNP A1 is generally characterized as nonspecific owing to its wide range of cellular functions; however, crosslinking immunoprecipitation (CLIP) and RNAcompete studies identified similar consensus patterns that center around a composite UAGG motif (11, 24, 34). Consistent with those results, the highest affinity SL3^ESS3 variant identified from HTS-EQ has a GUAGGAG sequence and 76% of the top 50 variants all contain UAG. Despite this correspondence, we were unable to determine a clear 7-nt consensus pattern from the HTS-EQ distribution because the minimal 5′-AG-3′ motif can occupy multiple and energetically similar registers around the apical loop. Indeed, quantitative modeling of the affinity distribution revealed favorable couplings between UA and AG dinucleotides at adjacent positions throughout the loop, which is consistent with 5′-AG-3′ being the most frequently observed dinucleotide identified in variants that bind tighter than native SL3^ESS3. Thus, we conclude that the specificity of hnRNP A1 centers on a conformationally exposed 5′-AG-3′ dinucleotide but that binding affinities are modulated by the surrounding sequence and structural context.

The preference for 5′-AG-3′ is understood by analysis of the UP1-AGU crystal structure, which shows both rA and rG engage in stereospecific contacts within the nucleobase pocket. Purines are preferred at both positions within the pocket due to favorable van der Waals and cation–π stacking interactions. Pyrimidines, due to their smaller ring size, are unlikely to satisfy these interactions with equivalent energetics. An adenine in the first position makes specific hydrogen bond interactions with functional groups contributed from Val90 and Arg88. At the second position, the guanine is primarily selected through hydrogen bonds with Gln12 and Lys15. Although some of these interactions can be satisfied with other dinucleotides, the combined effects of size and functional group specificity are unique to 5′-AG-3′. We previously hypothesized the hydrogen bond interaction between the N6 amino group of adenine and the α-carbonyl oxygen of Arg88 triggers a conformational change in UP1 that transmits across the inter-RRM linker to RRM2. Arg88 is involved in one of two conserved salt-bridge interactions that stabilize the relative RRM orientations and the nucleobase pocket. Mutation of Arg88 resulted in an ∼18-fold reduction in binding affinity for native SL3^ESS3, presumably by causing the nucleobase pocket to misfold (25). Thus, the ability of rA to induce a conformational change might increase the apparent thermodynamic contribution of rA at this position, because no other nucleobase can simultaneously fulfill the stacking interactions and hydrogen bond with Arg88.

It is confounding that 5′-AG-3′ can determine functional specificity, given its potential to be highly redundant in genomes. As revealed in the HTS-EQ distribution, contextual features surrounding 5′-AG-3′ modulate binding affinities of UP1 for SL3^ESS3 variants and individual biophysical measurements indicate the differences in affinities manifest as perturbations in association rate constants. Therefore, hnRNP A1 achieves specificity, in

part, from the frequency with which it collides productively with RNA targets that contain 5′-AG-3′ sequences embedded within optimized environments. Although the UP1 domain of hnRNP A1 retains binding affinity for non–AG-containing RNAs, the biological significance of those associations will also depend on physiological conditions within the cell. Transcript abundance, posttranscriptional modifications, and subcellular localization will each play a role in determining the extent to which hnRNP A1 occupies non-AG transcripts. Of note, hnRNP A1 has been shown to partition to RNA granules under different physiological conditions, including those physiological conditions associated with ALS pathogenesis (35). It will be of great interest to determine the sequence and structural composition of transcripts that contribute to this property.

A key function of hnRNP A1 is to regulate alternative splicing events, in which it associates with intronic or exonic splicing silencers to occlude components of the splicing apparatus. The results presented here suggest that hnRNP A1 might more globally affect splicing outcomes by binding directly to 3′-acceptor sites that contain consensus 5′-YAG-3′ motifs. Along those lines, hnRNP A1 was shown to proofread binding of U2AF to 3′-acceptor sites that contain AG but not CG dinucleotides (36). As revealed by quantitative modeling of the HTS-EQ distribution, CG dinucleotides contribute negatively toward UP1-SL3^{ESS3} recognition at most positions around the 7-nt apical loop (Fig. 2E). Thus, the ability of hnRNP A1 to regulate splicing likely derives from how rapidly it forms stable complexes with target transcripts that are also being competed for by other RBPs.

In summary, we applied HTS-EQ to determine the global affinity distribution of hnRNP A1 by evaluating binding of its UP1 domain to a randomized pool of HIV SL3^{ESS3} variants. The distribution reveals hnRNP A1 retains binding affinity for a wide range of RNA targets; however, specificity is decisively determined by the local context of 5′-YAG-3′ motifs. Therefore, the work presented here represents a quantitative and comprehensive evaluation of hnRNP A1 specificity, and it raises interesting questions with regard to the combinatorial factors that determine how hnRNP A1 identifies functional binding sites in vivo.

Materials and Methods

The overall work flow of the HTS-EQ method is summarized in Fig. S1. Details regarding sample preparation, data analysis, and data interpretation for HTS-EQ and the biophysical studies can be found in *SI Materials and Methods*.

1. Singh G, Pratt G, Yeo GW, Moore MJ (2015) The clothes make the mRNA: Past and present trends in mRNP fashion. *Annu Rev Biochem* 84:325–354.
2. Iadevaia V, Gerber AP (2015) Combinatorial control of mRNA fates by RNA-binding proteins and non-coding RNAs. *Biomolecules* 5(4):2207–2222.
3. Van Assche E, Van Puyvelde S, Vanderleyden J, Steenackers HP (2015) RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol* 6:141.
4. Mitchell SF, Parker R (2014) Principles and properties of eukaryotic mRNPs. *Mol Cell* 54(4):547–558.
5. Rissland OS (2017) The organization and regulation of mRNA-protein complexes. *Wiley Interdiscip Rev RNA* 8(1):1–17.
6. Ascano M, Gerstberger S, Tuschl T (2013) Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Curr Opin Genet Dev* 23(1):20–28.
7. Jayaseelan S, Doyle F, Tenenbaum SA (2014) Profiling post-transcriptionally networked mRNA subsets using RIP-Chip and RIP-Seq. *Methods* 67(1):13–19.
8. Licatalosi DD, et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456(7221):464–469.
9. Cook KB, Hughes TR, Morris QD (2015) High-throughput characterization of protein-RNA interactions. *Brief Funct Genomics* 14(1):74–89.
10. Campbell ZT, Wickens M (2015) Probing RNA-protein networks: Biochemistry meets genomics. *Trends Biochem Sci* 40(3):157–164.
11. Huelga SC, et al. (2012) Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Reports* 1(2):167–178.
12. Jankowsky E, Harris ME (2015) Specificity and nonspecificity in RNA-protein interactions. *Nat Rev Mol Cell Biol* 16(9):533–544.
13. Buenrostro JD, et al. (2014) Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol* 32(6):562–568.
14. Lambert N, et al. (2014) RNA Bind-n-Seq: Quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* 54(5):887–900.
15. Ozer A, et al. (2015) Quantitative assessment of RNA-protein interactions with high-throughput sequencing-RNA affinity profiling. *Nat Protoc* 10(8):1212–1233.
16. Geuens T, Bouhy D, Timmerman V (2016) The hnRNP family: Insights into their role in health and disease. *Hum Genet* 135(8):851–867.
17. Jean-Philippe J, Paz S, Caputi M (2013) hnRNP A1: The Swiss army knife of gene expression. *Int J Mol Sci* 14(9):18999–19024.
18. Ding J, et al. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* 13(9):1102–1115.
19. Xu RM, Jokhan L, Cheng X, Mayeda A, Krainer AR (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* 5(4):559–570.
20. Shamoo Y, Krueger U, Rice LM, Williams KR, Steitz TA (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. *Nat Struct Biol* 4(3):215–222.
21. Dreyfuss G, Kim VN, Kataoka N (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 3(3):195–205.
22. Ishikawa F, Matunis MJ, Dreyfuss G, Cech TR (1993) Nuclear proteins that bind the pre-mRNA 3′ splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. *Mol Cell Biol* 13(7):4301–4310.
23. Burd CG, Dreyfuss G (1994) RNA binding specificity of hnRNP A1: Significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* 13(5):1197–1204.
24. Bruun GH, et al. (2016) Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biol* 14:54.
25. Morgan CE, et al. (2015) The first crystal structure of the UP1 domain of hnRNP A1 bound to RNA reveals a new look for an old RNA binding protein. *J Mol Biol* 427(20):3241–3257.
26. Rollins C, Levengood JD, Rife BD, Salemi M, Tolbert BS (2014) Thermodynamic and phylogenetic insights into hnRNP A1 recognition of the HIV-1 exon splicing silencer 3 element. *Biochemistry* 53(13):2172–2184.
27. Lin H-C, Niland CN, Zhao J, Jankowsky E, Harris ME (2016) Analysis of binding landscapes: C5 protein contributes to ribonuclease P specificity of pre-tRNA 5′ leaders in the ground state and transition state for binding. *Cell Chem Biol* 23(10):1271–1281.
28. Stormo GD (2015) DNA motif databases and their uses. *Curr Protoc Bioinformatics* 51:2.15.1–2.15.6.
29. Zhao Y, Ruan S, Pandey M, Stormo GD (2012) Improved models for transcription factor binding site identification using nonindependent interactions. *Genetics* 191(3):781–790.
30. Guenther UP, et al. (2013) Hidden specificity in an apparently nonspecific RNA-binding protein. *Nature* 502(7471):385–388.
31. Towbin H, et al. (2013) Systematic screens of proteins binding to synthetic microRNA precursors. *Nucleic Acids Res* 41(3):e47.
32. Michlewski G, Cáceres JF (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat Struct Mol Biol* 17(8):1011–1018.
33. Michlewski G, Guil S, Cáceres JF (2010) Stimulation of pri-miR-18a processing by hnRNP A1. *Adv Exp Med Biol* 700:28–35.
34. Ray D, et al. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499(7457):172–177.
35. Molliex A, et al. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163(1):123–133.
36. Tavanez JP, Madl T, Kooshapur H, Sattler M, Valcárcel J (2012) hnRNP A1 proofreads 3′ splice site recognition by U2AF. *Mol Cell* 45(3):314–329.
37. Jain N, Morgan CE, Rife BD, Salemi M, Tolbert BS (2016) Solution structure of the HIV-1 intron splicing silencer and its interactions with the UP1 domain of heterogeneous nuclear ribonucleoprotein (hnRNP) A1. *J Biol Chem* 291(5):2331–2344.
38. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 7(8):1534–1550.
39. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
40. Johnson BA, Blevins RA (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4(5):603–614.

BIOCHEMISTRY

BIOPHYSICS AND
COMPUTATIONAL BIOLOGY

www.manaraa.com