

Molecular Cell, Volume 74

Supplemental Information

A Quantitative and Predictive Model

for RNA Binding by Human Pumilio Proteins

Inga Jarmoskaite, Sarah K. Denny, Pavanapuresan P. Vaidyanathan, Winston R. Becker, Johan O.L. Andreasson, Curtis J. Layton, Kalli Kappel, Varun Shivashankar, Raashi Sreenivasan, Rhiju Das, William J. Greenleaf, and Daniel Herschlag

Supplemental Figures and Tables

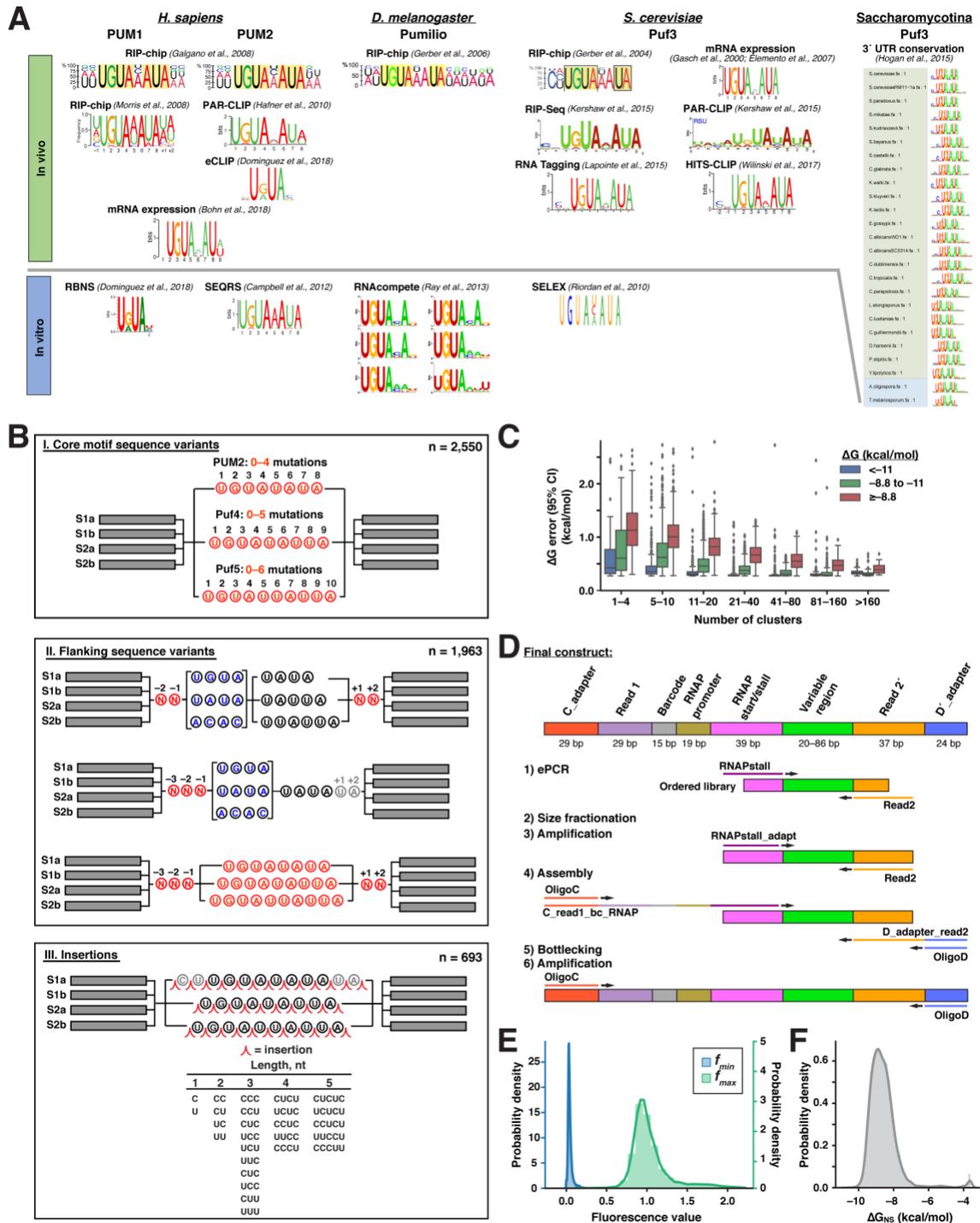


Figure S1. RNA Array Library Design, Preparation and Experiments, Related to Figure 1 and STAR Methods

(A) Representative sequence motifs previously identified for human PUM1, PUM2 and orthologous PUF proteins (Bohn et al., 2018; Campbell et al., 2012; Dominguez et al., 2018;

Elemento et al., 2007; Galgano et al., 2008; Gasch et al., 2000; Gerber et al., 2004; Gerber et al., 2006; Hafner et al., 2010; Hogan et al., 2015; Kershaw et al., 2015; Lapointe et al., 2015; Morris et al., 2008; Ray et al., 2013; Riordan et al., 2011; Wilinski et al., 2017). Copyright information for motifs that were reproduced from original figures: Kershaw et al., 2015, Hogan et al., 2015—CC BY 4.0 (<http://creativecommons.org/licenses/by/4.0/>); Gerber et al., 2006—Copyright (2006) National Academy of Sciences. Human PUM1 RBNS motif was obtained from the ENCODE database (Consortium, 2012; Dominguez et al., 2018); *D. melanogaster* Pumilio motifs were obtained from the CISBP-database (Ray et al., 2013).

(B) RNA libraries used in this study. Three broad categories of PUM2 binding site variants were designed: (I) mutations in the core binding site; (II) flanking sequence variants; (III) insertions. Each variant was represented in 2–4 scaffolds (S1a–S2b; Figure 1B). Within each category, the design was based on the previously established consensus motif of PUM2 and its orthologs; substantial additional sequence variation was achieved by applying the same types of perturbations to the consensus motifs of related PUF proteins, *S. cerevisiae* Puf4 and Puf5 (Gerber et al., 2004; Hogan et al., 2008; Hogan et al., 2015; Riordan et al., 2011; Wilinski et al., 2015). To prevent a preponderance of weak binders, in the mutation library (I) no more than two mutations were allowed in the highly conserved UGUA core of the PUF motifs. Numbers of variants in each category that were used for model fitting (Figure 3E,F) are indicated. Total numbers of designed variants and numbers of variants passing each criterion for inclusion in the final, high-quality data set are provided in Table S5.

(C) Affinity and number of clusters per variant affect measurement certainty. Box plots show measurement uncertainties from PUM2 replicate 1, with the box indicating the interquartile range, the whiskers indicating the minimum and maximum values and the points indicating outliers. Only variants with ≥ 5 clusters and ΔG values below -8.8 kcal/mol were included in subsequent analyses (see STAR Methods).

(D) Preparation of DNA array library. See STAR Methods and Table S7 for detailed sequence information and description of individual steps.

(E, F) Distribution of initial fit parameters for variant-independent parameters f_{\min} , f_{\max} , and ΔG_{NS} . (E) Histogram of the initial per-variant values for f_{\min} (blue) and f_{\max} (green) for variants that reached near-saturation of binding ($>95\%$ bound at highest concentration based on initial per-variant K_D) and that met other criteria for high-confidence binders during initial fitting, as defined in STAR Methods. (F) Histogram of the initial per-variant values for ΔG_{NS} ($= RT \ln K_{D,\text{NS}}$; STAR Methods). Values for PUM2 replicate 1 are shown.

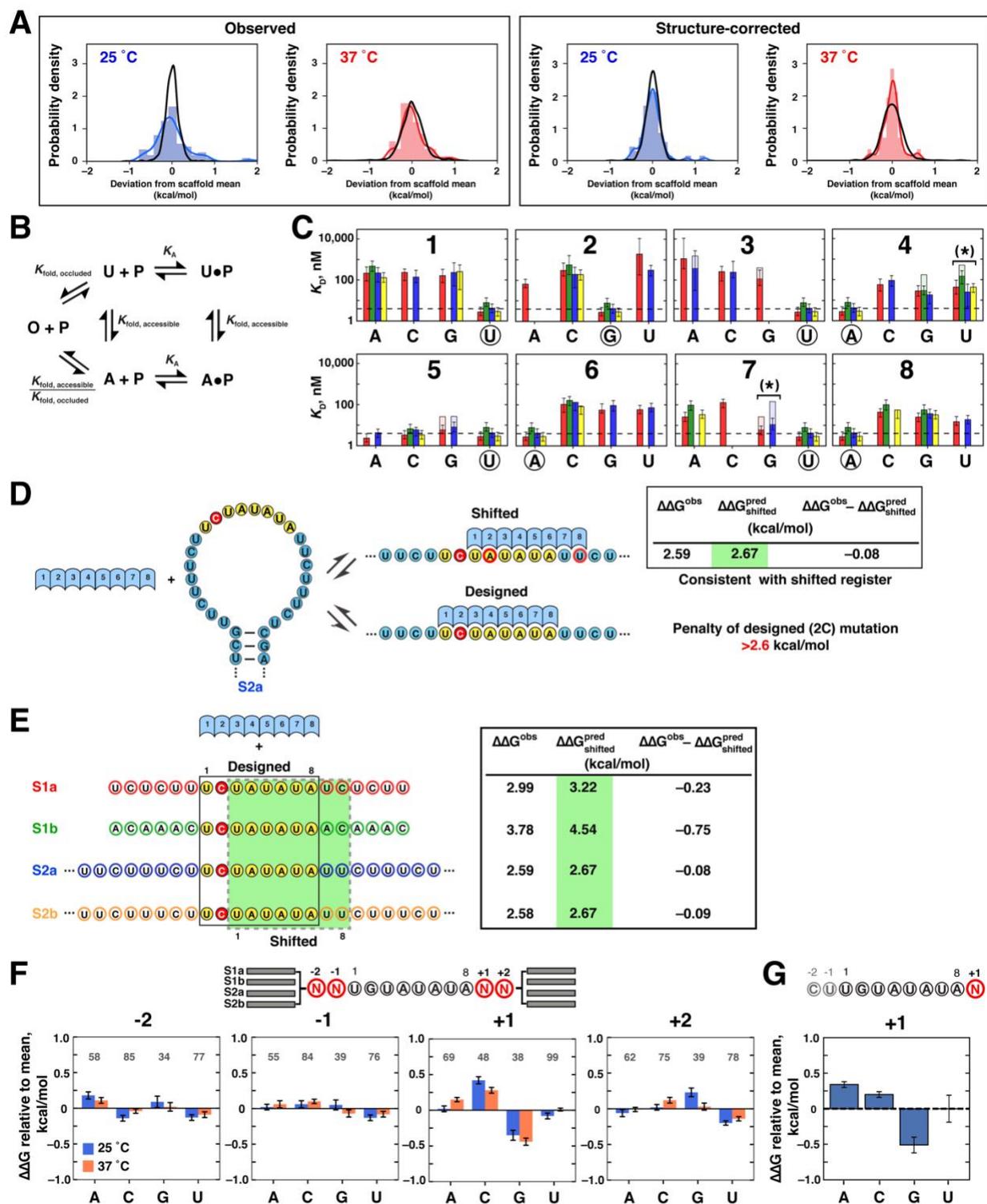


Figure S2. Analysis of Single-Mutant Variant Binding to PUM2, Related to Figure 2

(A) Assessing significance of scaffold differences. For each binding site variant (single mutants and UGUAUAUA consensus), we determined the deviations of individual scaffold affinities from

the scaffold mean for that variant. The distribution of these deviations was then compared to the distribution expected from experimental error (black line). The panels show distributions at 25°C (blue) and 37°C (red), before (left) and after (right) accounting for structure effects, as predicted by RNAfold ((Lorenz et al., 2011); see STAR Methods). A summary of the standard deviations of the displayed distributions is shown in Figure 2B.

(B) Model for accounting for RNA structure effects on PUM2 binding. U: unfolded RNA; P: protein; A: accessible RNA (i.e., structured RNA with the PUM2 binding site accessible); O: occluded RNA (i.e., PUM2 binding site occluded by base-pairing). A and O represent ensembles of the respective states; see Figure 2C). K_A : PUM2 association constant for the unstructured binding site (assumed to be the same for binding to U and A, but no protein binding occurs to O); $K_{\text{fold,accessible}}$: equilibrium constant for RNA folding from the unfolded state to the ensemble of structures in which the PUM2 binding site is accessible (determined by constraining the binding site to the single-stranded state in RNAfold; $K_{\text{fold,accessible}} = e^{-\Delta G_{\text{accessible}}/RT}$); $K_{\text{fold,occluded}}$: equilibrium constant for RNA folding from the unfolded state to the ensemble of structures in which the PUM2 binding site is occluded by structure.

(C) PUM2 affinities for UGUAUAUA single mutants at 37°C. Solid bars indicate mutant affinities after accounting for structure effects, and transparent regions indicate predicted structure effects. Asterisks indicate mutants for which affinities measured in different scaffolds were significantly different (10% FDR) prior to accounting for structure effects (differences between total bar heights). No significant differences remained after accounting for structure effects using RNAfold. Error bars indicate 95% CIs of the measured affinity values.

(D) Assessment of alternative binding registers. Example of preferential binding in an alternative register accompanying a destabilizing mutation in the designed binding site. Affinity effects are shown as $\Delta\Delta G$ values, relative to the average affinity for the UGUAUAUA consensus across scaffolds. The predicted $\Delta\Delta G$ for the alternative register corresponds to the sum of the $\Delta\Delta G$ values for the two nonconsensus residues, 2A and 8U (red outlines), assuming energetic independence. Given that the observed $\Delta\Delta G$ ($\Delta\Delta G^{\text{obs}}$) matches the prediction for the alternative register, the measurement only establishes a lower limit for the penalty of the intended, 2C, mutation.

(E) Observed and predicted $\Delta\Delta G$ values for the best alternative register ($\Delta\Delta G_{\text{shifted}}^{\text{pred}}$) in each of the four scaffolds for the 2C variant (other variants with predicted alternative registers are listed in Table S1). The close agreement of alternative register predictions with observed affinities in the S1a, S2a, and S2b scaffolds suggests predominant binding in the shifted register, whereas the 0.75 kcal/mol difference in the S1b scaffold suggests preferential binding in the intended register, such that the observed value of 3.78 kcal/mol reflects the real effect of the 2C mutation.

(F) Flanking sequence effects on PUM2 binding. *Top*: Library design. Two bases upstream and downstream from the UGUAUAUA consensus were randomized ($N = A/C/G/U$; $n = 256$) and the resulting sequences were embedded in 2–4 scaffolds each. *Bottom*: Effects of each base at the indicated positions relative to the average affinity. The effects were determined by calculating the average affinity of variants containing the indicated base (regardless of the identities of other flanking bases) and subtracting the mean of all four bases. Error bars indicate 95% CI of the mean. Only variants with predicted structural effects < -0.5 kcal/mol were included, and the numbers of variants containing each base are indicated above the bars. Blue and orange bars indicate the sequence effects at 25°C and 37°C, respectively. The observed sequence effects at positions -2, -1 and +2 are small and are further dampened at 37°C, suggesting that residual structural effects are responsible for the small observed effect. The strongest effect was observed

at position +1, and the observed preference for G was maintained at 37°C, indicating a weak, base-specific interaction.

(G) Gel-shift measurements of sequence effects at position +1. *Top*: Oligonucleotide design; N = A/C/G/U. *Bottom*: Effects of each base relative to the mean affinity for all four variants. Averages and standard errors from two replicate experiments are shown.

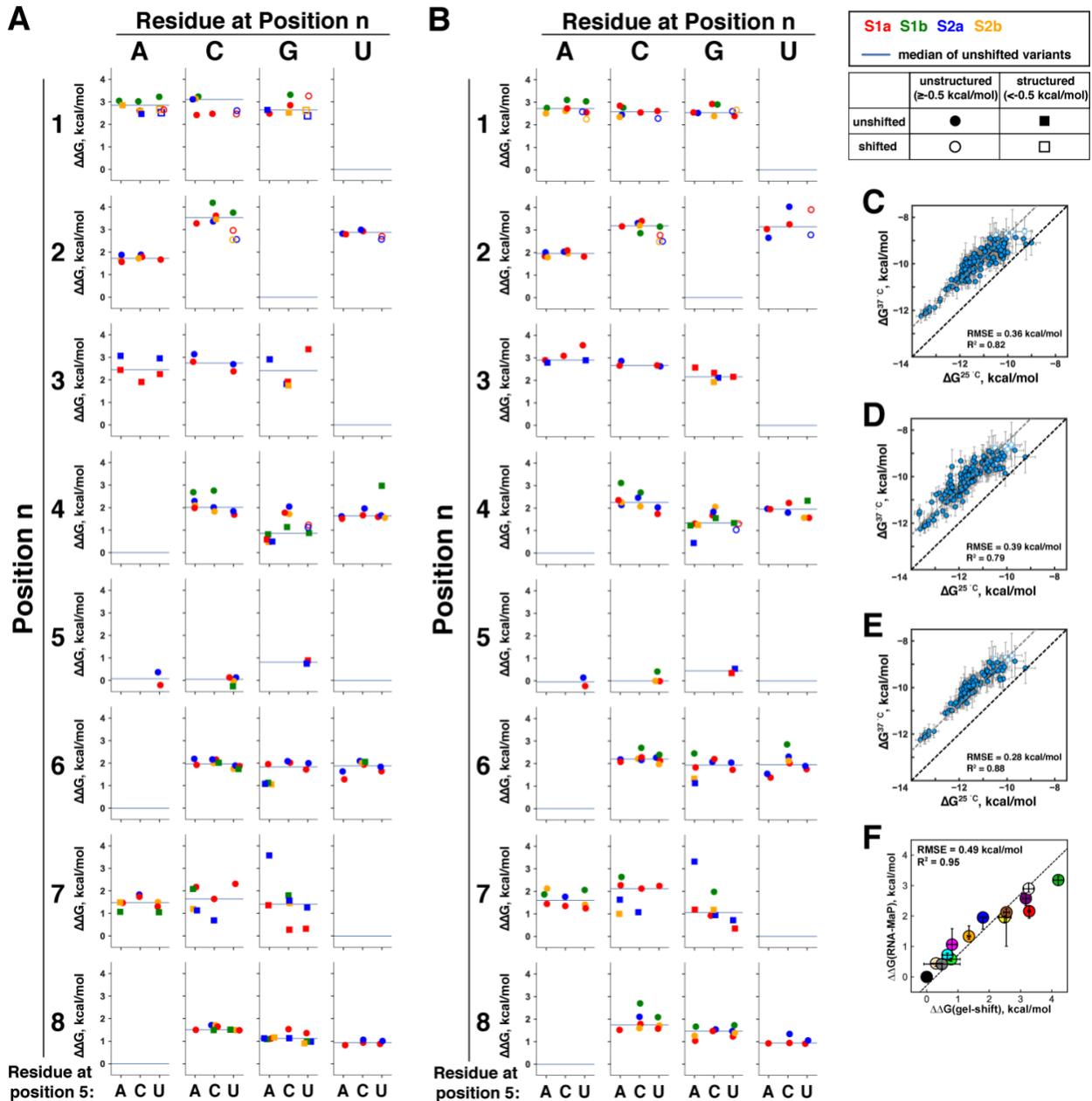


Figure S3. Determination of Single Mutant Penalties, Related to Figure 2

(A, B) Single mutation effects on PUM2 binding across backgrounds of different position 5 identities ($5 = A/C/U$) at 25°C (A) and 37°C (B). Symbols indicate the $\Delta\Delta G$ values for each mutation relative to the weighted mean of UGUA[A/C/U]AUA affinities across scaffolds (color-coded). As shown in Figure 2D and previously (Lu and Hall, 2011), residues A, C and U are bound with identical affinities at position 5, and other positions show high consistency of mutational penalties measured in 5A/C/U backgrounds, consistent with energetic additivity with respect to position 5. Thus, mutational penalties from all three backgrounds were used for determination of single mutant penalties, with median $\Delta\Delta G$ values indicated by horizontal lines. Circles indicate unstructured variants, while squares denote variants with predicted structure effects greater than 0.5 kcal/mol, and the displayed values correspond to structure-corrected affinities. Open symbols

indicate variants that have predicted alternative registers within 1 kcal/mol from the observed $\Delta\Delta G$ value; these variants were not included in determining the median. None of the variants in the 5A/C backgrounds were predicted to have alternative binding registers.

(C–E) Scatterplots of RNA single mutant affinities for PUM2 at 25°C and 37°C. (C) Observed ΔG values for single mutants in position 5A/C/U backgrounds ($n = 161$). Open symbols indicate variants with affinities outside the reliable affinity threshold (-8.8 kcal/mol; not included in RMSE and R^2 determination). The grey dashed line indicates the mean offset between values measured at 25°C and 37°C, corresponding to a (20 ± 10) -fold decrease in affinity at 37°C. Horizontal error bars indicate weighted errors of replicate measurements at 25°C; vertical error bars indicate 95% CIs of the measured affinities at 37°C. (D) ΔG values after accounting for predicted structure (RNAfold). The poorer correlation than observed in part C is likely due to limited accuracy of structure predictions (Becker et al., 2019a). (E) Correlation of observed ΔG values at 25°C and 37°C for variants with no predicted structure ($\Delta G_{\text{fold}} > -0.5$ kcal/mol; $n = 107$). The correlation approaches that of replicate measurements (Figure 1E; RMSE = 0.26 kcal/mol; $R^2 = 0.96$).

(F) Comparison between single mutant affinities determined by gel-shift measurements and by RNA-MaP (similar to Figure 2F, but the RNA-MaP values correspond to 37°C rather than 25°C measurements of PUM2 binding). 1C: purple, 2A: yellow, 2C: green, 3A: white, 3G: red, 4G: orange, 4U: blue, 5G: wheat, 7C: brown, 7G: magenta, 9A: lime, 9C: cyan, 9U: grey. The RNA-MaP values are medians and 95% CIs of individual mutation penalties in 5A/C/U backgrounds, determined after accounting for secondary structure and shifting. The gel-shift values are averages and 95% CIs from 2–4 measurements. The better agreement of gel-shift values with 37°C than with 25°C RNA-MaP values (Figure 2F) is consistent with small structural effects on RNA-MaP-derived values at 25°C.

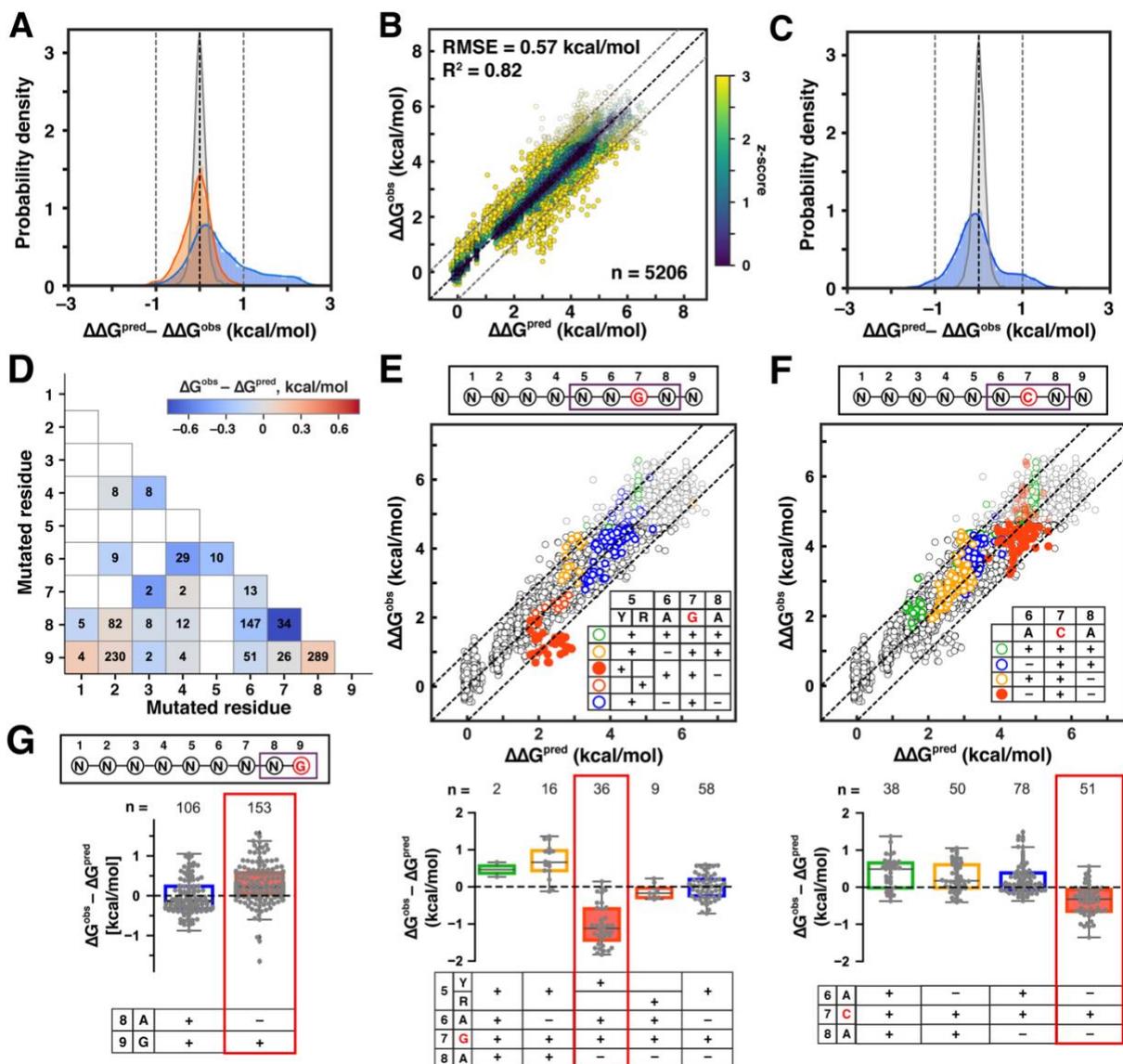


Figure S4. Developing a Predictive Model for PUM2 Specificity, Related to Figure 3

(A) Distribution of differences between model predictions and observed affinities through PUM2-binding model development. Blue: additive consecutive model (Figure 3A); orange: final model including flipping and coupling terms (Figure 3F); grey: the distribution of standard errors of measured $\Delta\Delta G$ values. Coupling did not noticeably affect the distribution, so the additive nonconsecutive model without coupling (Figure 3E) is not displayed.

(B) Global fit to the additive consecutive model. Parameter constraints were identical to those used for fitting the additive nonconsecutive model (Figure 3E) and corresponded to the 95% confidence intervals from single mutant measurements (Figure 2E) or ± 0.4 kcal/mol, whichever was greater. Scatter plot of the predicted versus observed $\Delta\Delta G$ values is shown. The points are colored based on their z-scores for the deviation from predicted value (see Figure 3A). The black dashed line indicates perfect agreement with prediction, and dashed grey lines denote 1 kcal/mol deviation from predictions. While the additive consecutive model performed considerably better with globally fit parameters than with experimentally determined single mutant penalties (Figure

3A), there was substantial asymmetry and ten percent of variants had predicted affinities >1 kcal/mol from the observed value (versus 0.00% expected from measurement error); see part C.

(C) Blue: distribution of differences between values predicted by the additive consecutive model (using parameter values determined by global fit; part B) vs. observed values; grey: distribution of standard errors of measured affinities, as in part A.

(D) Double mutant coupling at 25°C. Colors indicate the difference of observed $\Delta\Delta G$ values ($\Delta\Delta G_{\text{obs}}$) from the sum of $\Delta\Delta G$ values for the individual mutations ($\Delta\Delta G_{\text{pred}} = \Delta\Delta G_1 + \Delta\Delta G_2$). Double mutants were defined as all sequences differing from the UGUA[A/C/U]AUAU reference sequence at two positions and predicted to be in the most favorable register with no flips (STAR Methods); only variants with predicted structure effects of <0.5 kcal/mol were included. Numbers of variants containing each combination of mutations are shown inside the respective boxes; only mutants represented by more than one variant in our library are shown. White boxes indicate absence of the respective mutants in the best register in our unstructured library: these absences are primarily due to the highly destabilizing effects of mutations in the 5' half of the motif, which in combination with a second mutation lead to binding outside our reliably measurable range and/or to shifting of the preferred binding site to an alternative binding register; low representation of 5G-containing mutants is due to predicted structure.

(E) Coupling between the 7G mutation and adjacent positions, assessed as the extent to which measured affinities for variants with the indicated residue combinations (inset) deviate from predicted values. The predictions are based on the additive nonconsecutive model (Figure 3E). All sequences containing the indicated combination of residues in the predicted best consecutive register were considered. Black residues in the inset table indicate consensus residues. The top panel shows the variants overlaid on the scatterplot of predicted and observed values for the entire library (same as Figure 3E). The box plots in the bottom panel show the distributions of differences from predicted values for each group of variants. The coupling term that was incorporated in the final model is highlighted by solid red circles (top) and red outline (bottom). To avoid overfitting, only most pronounced instances of apparent coupling were included in the final model.

(F) Coupling between the 7C mutation and flanking residues, as in *E*.

(G) Coupling between positions 8 and 9. As in *E* and *F*, coupling was assessed as the extent to which affinities for variants with the indicated residues deviated from predicted values. The panel illustrates the loss of the small stabilizing effect of 9G when position 8 is mutated (i.e., not A).

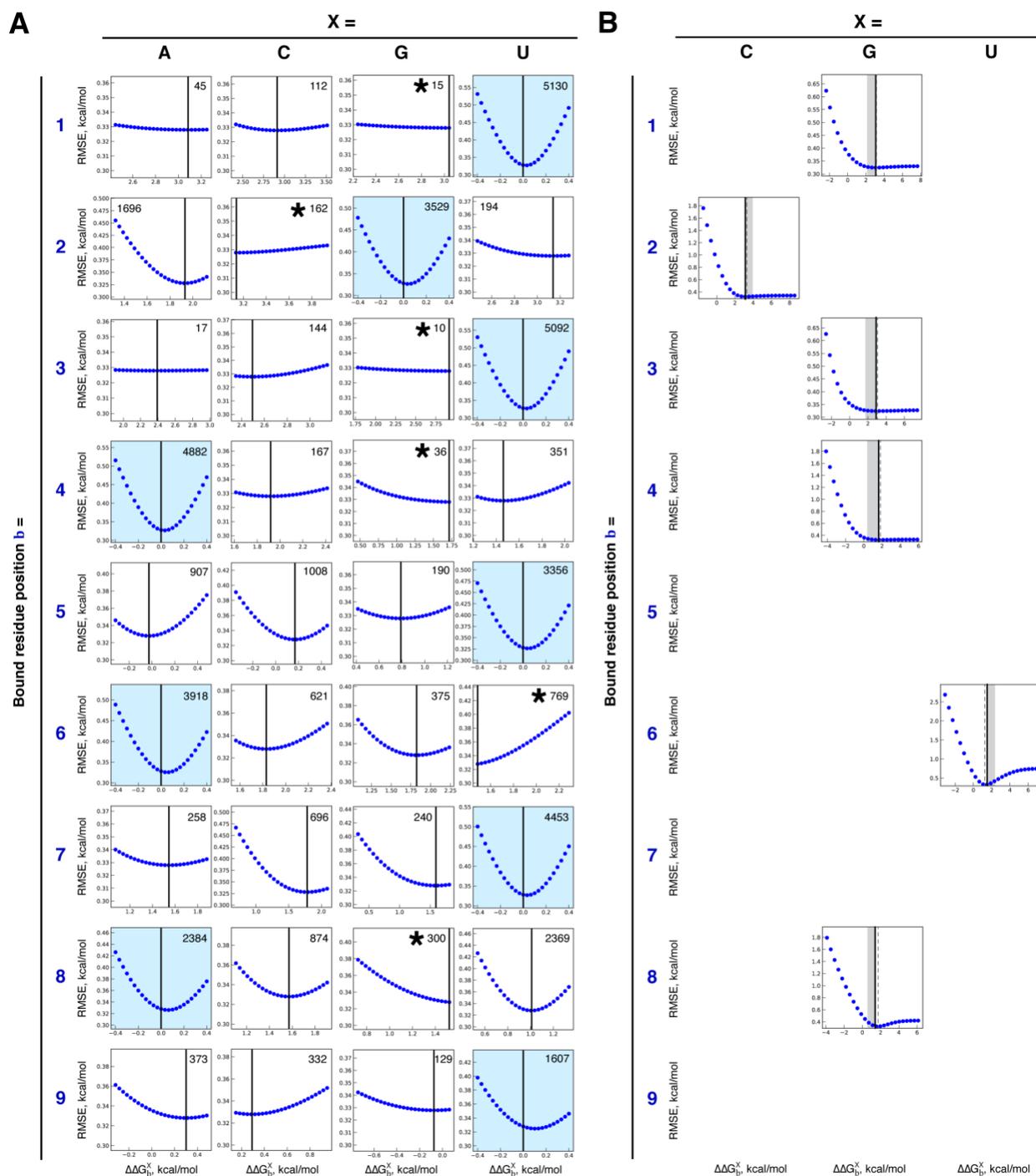


Figure S5. RMSE Dependence on Individual Fit Parameters in the Additive Nonconsecutive Model Including Coupling (Bound Residue Terms), Related to Figure 3

(A) Bound residue terms ('b' in Figure 3F; Table 1, 'Term I'). The sensitivity of the global fit to each model parameter was individually assessed by varying the $\Delta\Delta G$ value across the indicated range, with the values of other parameters kept constant. $\Delta\Delta G$ values were constrained to measured single mutant penalties (median ± 0.4 kcal/mol) or to the 95% confidence intervals of the measured median if they exceeded 0.4 kcal/mol (Figure 2E and Table S2). Vertical black lines

indicate the fit values. Asterisks indicate parameters for which the fit value was at the limit of the allowed range (see part B). The number of constructs containing each parameter in the best register (and any additional registers within 0.4 kcal/mol from the best one) is indicated. The slight offset from 0 observed for the consensus residues appears to be due to residual RNA structure (STAR Methods). As expected, the RMSE dependence was shallower for nonconsensus residues than for consensus residues due to the lower representation in the library, and the shallowness was most pronounced for the highly destabilizing 5' mutations, which were only weakly represented in the preferred binding registers.

(B) Six binding terms for which the best fit value was at the limit of the allowed range (asterisks in part A) were refit without constraints on their $\Delta\Delta G$ value. The gray area indicates the original constraints. The solid lines indicate the original fit value, and the dashed line indicates the fit value after removing the constraints. In all cases, removing the constraints resulted in fit values within 0.2 kcal/mol of the original fit value. The small deviations could originate from small RNA structural effects or additional weak couplings.

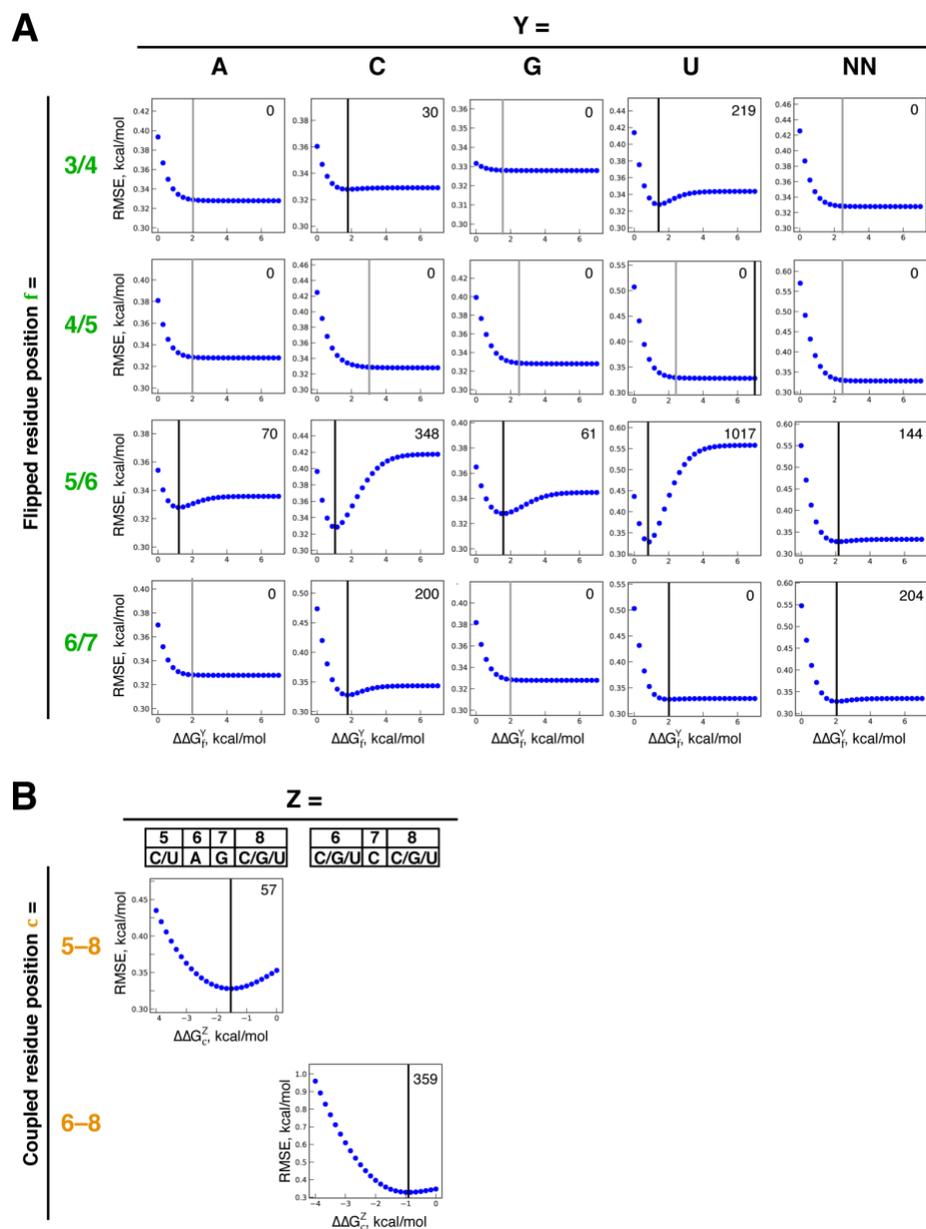


Figure S6. RMSE Dependence on Individual Fit Parameters in the Additive Nonconsecutive Model Including Coupling (Flipping and Coupling Terms), Related to Figure 3

(A) Flipping terms ('f' in Figure 3F; Table 1, 'Term II'). As in part A, the sensitivity of the global fit RMSE to each term was assessed by varying the $\Delta\Delta G$ value across the indicated range, while the values of other parameters were kept constant. Black lines indicate the fit values; gray lines indicate lower limits for terms whose upper limits were undefined because these parameters were never featured in the most stable binding register (largely as a result of energetic redundancy between flipped positions). 'NN' indicates 2-nt flips of any sequence.

(B) Coupling terms. RMSE sensitivity was assessed as in parts A–C. The 8/9 coupling term is not separately shown, as its value is determined by the position 9 binding terms (A; Table 1).

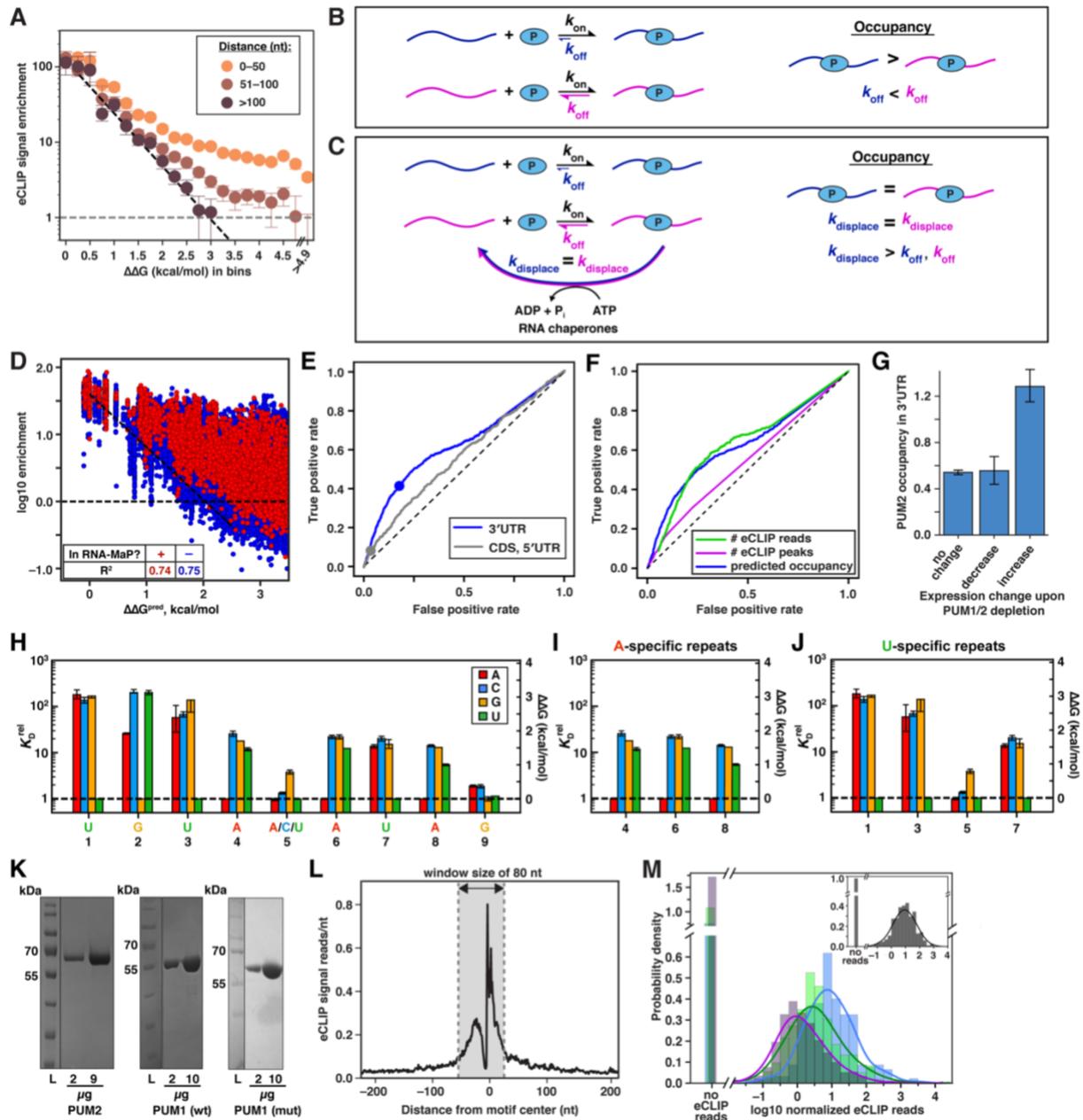


Figure S7. Assessment of PUM2/RNA Interactions in Vivo, Related to Figure 6

(A) Scatterplot comparing the median eCLIP enrichment across sites in bins of predicted relative affinity. Colors indicate sites with a nearby UGUA sequence within 50 nt, between 50–100 nt, or greater than 100 nt from the predicted site. The black dashed line indicates the expected change in the eCLIP signal for different predicted $\Delta\Delta G$ values, relative to the eCLIP signal in the lowest $\Delta\Delta G$ bin. Grey dotted line indicates expected background enrichment ($=1$). Error bars indicate the 95% confidence intervals on the median. Numbers of sites per bin are provided in Table S6.

(B,C) Schematic representation of models for RNA occupancy in vivo in which active RBP ('P') displacement in vivo (e.g., by ATP-dependent RNA chaperones) yields identical occupancies for targets with different affinities. (A) Equilibrium model: Two targets with different affinities are

shown, in blue and purple. The association rate constants (k_{on}) for RBP binding to the two RNAs are identical, and the affinity difference is reflected in the greater dissociation rate constant (k_{off}) for the second target (as is common and is observed for PUM2; data not shown). The occupancy (right) is greater for the first target due to its lower K_D value ($K_D = k_{off}/k_{on}$). (B) Kinetic model: If the rate constant of active displacement ($k_{displace}$) is greater than either of the intrinsic dissociation rate constants, the occupancies are equal and independent of the affinity for the RNA targets (assuming equal k_{on} values).

(D) Predicting RBNS enrichments with the thermodynamic model. RBNS enrichments of 11mer sequences from publicly available PUM1 (5 nM) and protein-free input datasets were used (Consortium, 2012; Dominguez et al., 2018), and data are shown for the 3.5 kcal/mol range within which a linear relationship between predicted affinities and RBNS enrichments was observed. Given the uneven distribution of 11mers across the predicted affinity range (Table S6), the R^2 values are based on data subsampled to have the same number of randomly selected points in each 0.5 kcal/mol $\Delta\Delta G^{pred}$ bin (see STAR Methods for details).

(E) Receiver operating characteristic (ROC) curves for classification of PUM1 and PUM2 targets based on their predicted relative occupancy in their 3'-UTR region (blue) or outside the 3' UTR (grey). True targets are defined as mRNAs that were significantly upregulated upon PUM1 and PUM2 knockdown in (Bohn et al., 2018) (see STAR Methods). Predicted targets are genes that have predicted relative PUM1 and PUM2 occupancy greater than or equal to a variable threshold. Dots correspond to the true positive and false positive rate using a threshold equal to 1 (i.e., occupancy equivalent to one consensus site).

(F) ROC curves for classification of PUM1 and PUM2 targets based on their predicted relative occupancy in the 3'-UTR region (blue) vs. PUM2 eCLIP signal across the 3' UTR (green: number of abundance-normalized eCLIP reads; magenta: number of eCLIP peaks). While eCLIP read counts and thermodynamics-based occupancies predict regulation to a similar degree, note that the number of eCLIP peaks in the 3' UTR (as defined in published eCLIP data (Consortium, 2012; Van Nostrand et al., 2016)) is not predictive of regulation.

(G) Mean predicted relative occupancies across 3' UTRs of mRNAs that are unaffected, downregulated, or upregulated upon PUM1 and PUM2 knockdown (Bohn et al., 2018). Upregulated mRNAs have significantly increased predicted PUM1 and PUM2 occupancy relative to unaffected genes ($p = 0.001$, Mann-Whitney U test). Error bars indicate 95% CIs of the mean.

(H) Relative affinities for bound residues 1–9 (corresponding to PUM2 repeats 'R1'–'R9') determined by global fitting to the final model (additive nonconsecutive with coupling; Table 1, 'Term I'; Figure 3F). Error bars indicate 95% confidence intervals of the median, determined by bootstrapping analysis (STAR Methods).

(I) Relative affinities of A-specific PUM2 repeats determined by global fitting.

(J) Relative affinities of PUM2 repeats with conserved U-specific residues.

(K) SDS-PAGE gels showing the purity of final purified PUM1 and PUM2 constructs. *L*: protein ladder.

(L) Average number of eCLIP read stops per nucleotide (nt) around 3,703 PUM2 consensus binding sites ($\Delta\Delta G_{pred} < 0.5$ kcal/mol) within annotated 3' UTR, CDS, or 5' UTR regions of expressed genes in K562 cells. The grey area indicates the interval relative to the binding site over which any eCLIP read stop was included in the calculation of eCLIP enrichment (80 nt window, offset by 15 nt relative to the center of the binding site).

(M) Distribution of eCLIP signal for representative affinity bins from Figure 6A: -0.12 – 0.12 kcal/mol (blue), 0.88 – 1.12 kcal/mol (green), 1.88 – 2.12 kcal/mol (magenta). The eCLIP signal is the eCLIP read counts divided by transcript abundance (see STAR Methods). *Inset*: Distribution of eCLIP signal for UGUUAUAU-containing sites ($n = 745$).

Table S1. Single mutant variants with alternative registers that have predicted affinities within 1 kcal/mol from the observed affinity, which would lower the apparent K_D by at least 20%. Related to Figure 2 and STAR Methods.

	Designed site	Shifted site	$\Delta\Delta G^{obs}$	$\Delta\Delta G^{pred}_{shifted}$	$\Delta\Delta G^{obs} - \Delta\Delta G^{pred}_{shifted}$
S1a	<u>A</u> GUAUAUA	UAUAUAUC	2.68	3.22	-0.54
S2a		UAUAUAUU	2.55	2.67	-0.12
S2b		UAUAUAUU	2.70	2.67	0.03
S1a	<u>C</u> GUAUAUA	UAUAUAUC	2.49	3.22	-0.73
S2a		UAUAUAUU	2.64	2.69	-0.03
S1a	<u>G</u> GUAUAUA	UAUAUAUC	3.29	3.22	0.07
S2a		UAUAUAUU	2.42	2.67	-0.25
S2b		UAUAUAUU	2.65	2.67	-0.02
S1a	<u>U</u> AUAUAUA	UUUAUAUA	1.70	2.64	-0.94
S1a	<u>U</u> CUAUAUA	UAUAUAUC	2.99	3.22	-0.23
S1b		UAUAUAAC	3.78	4.54	-0.75
S2a		UAUAUAUU	2.59	2.67	-0.08
S2b		UAUAUAUU	2.58	2.67	-0.09
S1a	<u>U</u> UAUAUA	UAUAUAUC	2.73	3.22	-0.49
S2a		UAUAUAUU	2.60	2.67	-0.07
S1a	UG <u>U</u> GUAUA	UGUAUAUC	1.26	1.53	-0.27
S2a		UGUAUAUU	1.16	0.97	0.19
S1a	UGUAUA <u>C</u> A	UAUACAUC	2.35	3.23	-0.88

Table S2. Experimentally determined single mutant penalties. Related to Figure 2.

Bound residue position b =	$\Delta\Delta G_b^X$ (kcal/mol)			
	X =			
	A	C	G	U
1	2.85*	3.11	2.64	0.00
2	1.72	3.53	0.00	2.87
3	2.45	2.75	2.41	0.00
4	0.00	2.01	0.87	1.64
5	0.08	0.05	0.81	0.00
6	0.00	1.96	1.84	1.89
7	1.48	1.64	1.41	0.00
8	0.00	1.51	1.12	0.94
9	0.08	0.52	-0.34	0.00

* Position 1–8 values correspond to single mutant medians from Figure S3A. Position 9 values are based on measurements of randomized flanking sequences around the UGUAUAUA consensus (Figure S2F)

Table S7. Oligonucleotide sequences. Related to STAR Methods.

	Name	Sequence (sense)
Components of the RNA array construct (see Figure S1D)	C_adapter	AATGATACGGCGACCACCGAGATCTACAC
	Read1	TCTTTCCCTACACGACGCTCTTCCGATCT
	Barcode*	NNNNNNNNNNNNNNNN
	RNAP promoter	TTTATGCTATAATTATTTTC
	RNAP start/stall sequence**	ATGTAGTAAGGAGGTTGTATGGAAGAC <u>G</u> TTTCCTGGATCC
	Read2'	AGATCGGAAGAGCGGTTTCAGCAGGAATGCCGAGACCG
	D'_adapter	ATCTCGTATGCCGTCTTCTGCTTG
Primers used in library preparation	RNAPstall	GTAAGGAGGTTGTATGGAAGACGTTTCCTGGATCC
	Read2	CGGCATTCTGCTGAACCGCTCTTCCGATCT
	RNAPstall_adapt	GTAAGGAGGTTGTATGGAAGACGTTTCCTGGAT
	C_read1_bc_RNAP	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTNNNNNNNNNNNNNNNNNTTTATGCTATAATTATTTTCATG TAGTAAGGAGGTTGTATGGAAGACGTTTCCTGGAT
	D_read2	CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCTGCTGAAC CGCTCTTCCGATCT
	OligoC	AATGATACGGCGACCACCGAGATCTACAC
	OligoD	CAAGCAGAAGACGGCATAACGAGAT
DNA oligonucleotides used in RNA array preparation	Fiducial_chip***	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTCTTGGGTCCACAGGACACTCGTTGCTTTCCAGATCG GAAGAGCGGTTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTC TTCTGCTTG
	Fiducial_flow	/5TYE563/GGAAAGCAACGAGTGTCTGTGGACCCAAG
	Fluorescent_stall'	GGATCCAGGAACGTCTTCCATACAACCTCCTTACTACAT/3AlexF647N/
	Biotin_D_Read2	/5BiosG/CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCTG CTGAACCGCTCTTCCGATCT
	Dark_stall'	GGATCCAGGAACGTCTTCCATACAACCTCCTTACTACAT
	Dark_read2	CGGCATTCTGCTGAACCGCTCTTCCGATCT

* Randomized 15mer region.

** Underlined 'C' indicates the RNA polymerase stall site.

*** Corresponds to 5' – [C_adapter] [Read1] CTT GGG TCC ACA GGA CAC TCG TTG CTT
TCC [Read2'] [D'_adapter].