Programmable RNA-binding protein composed of repeats of a single modular unit

Katarzyna P. Adamala[a,1], Daniel A. Martin-Alarcon[b,1], and Edward S. Boyden[a,b,c,2]

[a]Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

The ability to monitor and perturb RNAs in living cells would benefit greatly from a modular protein architecture that targets unmodified RNA sequences in a programmable way. We report that the RNA-binding protein PumHD (Pumilio homology domain), which has been widely used in native and modified form for targeting RNA, can be engineered to yield a set of four canonical protein modules, each of which targets one RNA base. These modules (which we call Pumby, for Pumilio-based assembly) can be concatenated in chains of varying composition and length, to bind desired target RNAs. The specificity of such Pumby–RNA interactions was high, with undetectable binding of a Pumby chain to RNA sequences that bear three or more mismatches from the target sequence. We validate that the Pumby architecture can perform RNA-directed protein assembly and enhancement of translation of RNAs. We further demonstrate a new use of such RNA-binding proteins, measurement of RNA translation in living cells. Pumby may prove useful for many applications in the measurement, manipulation, and biotechnological utilization of unmodified RNAs in intact cells and systems.

RNA-binding protein | Pumilio | gene expression monitoring | protein engineering | translation initiation

Many scientific questions and bioengineering goals relate to the monitoring and control of RNA functions in living cells. A powerful strategy is to modify a target RNA by inserting an exogenous sequence such as MS2 or PP7, so that the corresponding RNA-binding protein can deliver a reporter or RNA modification enzyme to an RNA of interest (1–3). Ideally one could target unmodified RNA, both for simplicity and to preserve as much native RNA structure and function as possible (4, 5). It has been proposed that proteins such as the *Caenorhabditis elegans* Puf (6), the human PumHD (Pumilio homology domain) (7), or members of the pentatricopeptide family (8) could serve such a purpose. Each of these proteins is made of many similar units, each of which targets one RNA base. The most extensively studied protein architecture, in the context of single-stranded RNA targeting in mammalian cells, is the human PumHD (9–12). PumHD is a protein with 10 units, of which 8 units bind to the bases of an eight-nucleobase target RNA sequence (Fig. 1*A*), called the Nanos response element (NRE), in the reverse orientation 3′-AUAUAUGU-5′ (Fig. 1*B*) (13–19). X-ray structures of the PumHD–NRE complex indicate that three key amino acids interact with each RNA nucleobase (14, 20).

A number of pioneering studies have shown that modifications of the wild-type PumHD can indeed bind to many sequences other than the NRE (summarized in *SI Appendix*, Table S18), strongly pointing toward the modularity of PumHD (we here use the shorthand "Pum" to denote any protein homologous to, or derived from, PumHD). We set out to determine whether, given the rich set of previous findings related to Pum proteins, we could devise a set of four canonical protein modules, each of which targets one RNA base with high specificity and could be concatenated in chains of varying composition and length so as to bind desired target RNAs. A similar protein architecture, the transcription activator-like (TAL) effector, has been rendered in this single-module

form and has proven to be useful for targeting DNA with various proteins, because of its modularity (21, 22). There are four canonical TALE protein modules, each of which targets one DNA base with high specificity. If analogous Pum modules could be developed, they could be easily designed and used: Simply concatenate a chain of modules according to the sequence of a natural target RNA, and then the protein (perhaps equipped with various reporters and effectors) could be targeted to a desired RNA.

Results

A number of studies have mutated different units of PumHD to target different bases, testing various mutations in various cell-free or cellular contexts. Eleven of these studies used mammalian cells to explore 19 out of the 24 possible mutant units (i.e., three different bases at eight different sites; *SI Appendix*, Table S18). Because no single study tested PumHD variants binding to all four nucleotides at each unit's position, in the same condition, we first assessed whether all 24 PumHD single-unit mutants could target their respective 8-nt sequences. We used an assay commonly used in Pumilio evaluation and also useful in cell biology—RNA-based GFP complementation in mammalian cells (Fig. 2 *A* and *B*). This assay is sometimes seen as qualitative, because it does not indicate actual binding affinities, but it has proven useful in the study of RNA-binding proteins such as Pumilio because it allows for such interactions to be measured at a functional level in living cells (23–26). In particular, split fluorescent protein reconstitution was used to test on-target binding of three different Pum variants to NRE variants, and also previously used to visualize binding of PumHD variants to the mRNAs for human β-actin and NADH dehydrogenase subunit 6 (23–26). Based upon earlier literature,
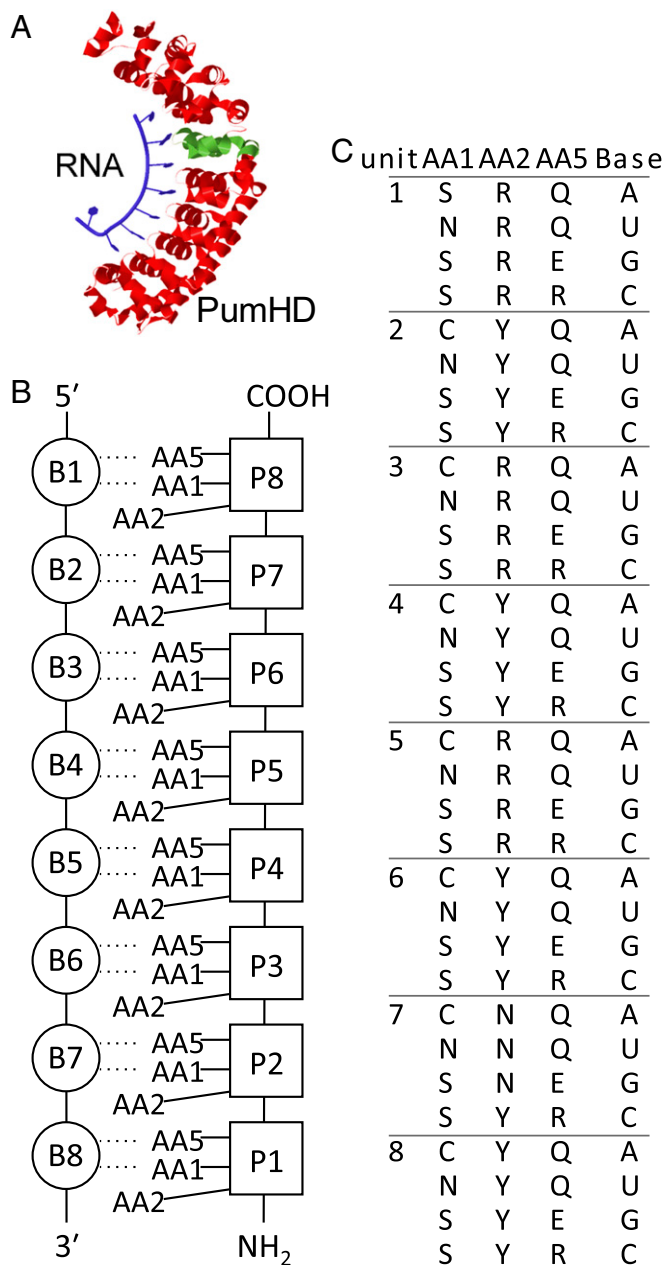
Fig. 1. A proposed amino acid code for universal PumHD binding. (A) Crystal structure of the wild-type human PumHD (red) with its cognate RNA (blue). One protein unit is highlighted in green; data are from PDB ID code 1M8X (14). (B) Schematic representation of RNA bases (labeled B1 to B8) and their respective PumHD protein units (labeled P8 to P1). Note the binding direction: The carboxyl terminus of the Pum protein binds to the 5′ end of the target RNA. Three amino acids (labeled AA1, AA2, and AA5) are key for recognizing the target nucleobases. (C) A proposed consensus sequence of key amino acids that allow PumHD to bind RNA sequences that have any base at any position in the eight-base target sequence.

unit	AA1	AA2	AA5	Base
1	S	R	Q	A
	N	R	Q	U
	S	R	E	G
	S	R	R	C
2	C	Y	Q	A
	N	Y	Q	U
	S	Y	E	G
	S	Y	R	C
3	C	R	Q	A
	N	R	Q	U
	S	R	E	G
	S	R	R	C
4	C	Y	Q	A
	N	Y	Q	U
	S	Y	E	G
	S	Y	R	C
5	C	R	Q	A
	N	R	Q	U
	S	R	E	G
	S	R	R	C
6	C	Y	Q	A
	N	Y	Q	U
	S	Y	E	G
	S	Y	R	C
7	C	N	Q	A
	N	N	Q	U
	S	N	E	G
	S	Y	R	C
8	C	Y	Q	A
	N	Y	Q	U
	S	Y	E	G
	S	Y	R	C

we hypothesized a consensus sequence for how to modify each unit of PumHD so that its base preference could be tuned to any of the four RNA bases (Fig. 1C). We adapted—from the TAL effector field—a Golden Gate assembly method to rapidly create PumHD variants (SI Appendix, Fig. S1 and Supplementary Results).

We used a reference PumHD variant [used in a prior fluorescent protein reconstitution study (23), and that binds 3′-AUA-GAUGU-5′] to assess the efficacy of our hypothesized consensus sequence (Fig. 2F). Throughout our experiments we used two

PumHD proteins (a variable Pum, denoted Pum1, and a wild-type, denoted Pum2), each fused to one part of a split GFP, which bind two adjacent sequences before the stop codon of a transcript that codes for mRuby (Fig. 2B). Pum1 binding was assessed with on- as well as off-target RNA sequences (for off-target cases, purines were swapped with pyrimidines at all eight positions; see SI Appendix, Table S2 for all sequences used). We found that on-target RNA sequences supported effective Pum1 binding and green fluorescence, whereas off-target RNA sequences did not (individual examples, Fig. 2 C–E; population data, Fig. 2F; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and target sequence; $n = 3$ biological replicates each; for full statistics for Fig. 2, refer to SI Appendix, Table S1). All 24 PumHD variants had binding preferences for on-target vs. off-target sequences that were indistinguishable from the wild type ($P > 0.05$, Dunnett's post hoc test comparing target sequence vs. wild-type, for the ANOVA above). Thus, as expected given the prior literature, PumHD variants can indeed support any unit targeting any base.

To explore the robustness of PumHD binding in varying contexts, we tested binding of the wild-type PumHD with varying bases upstream and downstream of the NRE (Fig. 2G) and found successful binding, albeit with statistically significant differences in GFP reconstitution from one set of bases to another (statistics in SI Appendix, Table S1). Given that any protein–RNA interaction will be susceptible to environmental changes or RNA secondary structure arising from the specific sequences involved, this result suggests that PumHD variants should be vetted on a per-case basis. However, PumHD variants were generally capable of binding their target regardless of the bases immediately upstream and downstream of the core eight bases.

To assess the specificity of PumHD mutants for target sequence even more quantitatively, we assessed binding of two different PumHD variants to RNA targets that were on-target vs. those off-target at one, two, or three specific bases (see SI Appendix, Table S2 for all of the sequences used), using the GFP reconstitution method described above. Although some Pum-mediated GFP reconstitution was observed for RNA targets off by two bases, RNA targets off by three bases did not support GFP reconstitution any more than did completely different (i.e., off by eight bases) RNA sequences (Fig. 2H; $P = 0.9999$ for comparison of three vs. eight mismatches; Dunnett's post hoc test for the factor of mismatch number, after a two-way ANOVA with factors of mismatch number and Pum identity; $n = 3$ biological replicates; see SI Appendix, Table S1 for full statistics).

We then set out to make a set of four canonical protein modules, each of which targets one RNA base with high specificity (Fig. 3A). As in Fig. 2, we tested binding for both on-target and off-target Pum pairs in live mammalian cells, using GFP reconstitution. For simplicity we kept AA2 (the "stacking" amino acid) the same for all four modules in our design. Because most of the PumHD units of Fig. 1C had either Y or R for AA2, we decided not to use unit 7, which mostly used N. Then, we examined which units had been most thoroughly mutated by the most groups (SI Appendix, Table S18) and thus had been the most vetted in a variety of contexts and chose units 3 and 6 of PumHD as candidates for a Pumby module starting material. We screened variants of units 3 and 6 using the GFP reconstitution assay (see SI Appendix, Table S21 for all of the Pumby candidates that we tested). Using unit 3 and stacking amino acid R, assemblies that we tested seemed to hamper cell survival (SI Appendix, Fig. S6A). Using unit 3 and stacking amino acid Y, the tested assemblies did not hamper cell survival, but no Pum-mediated GFP reconstitution was observed (SI Appendix, Fig. S6B). Using unit 6 and stacking amino acid R, we found that the tested assemblies expressed well, but very weak Pum-mediated GFP reconstitution was observed for all tested sequences (SI Appendix, Fig. S6C). Finally, we tested unit 6 with stacking amino acid Y and found normal cell health and also
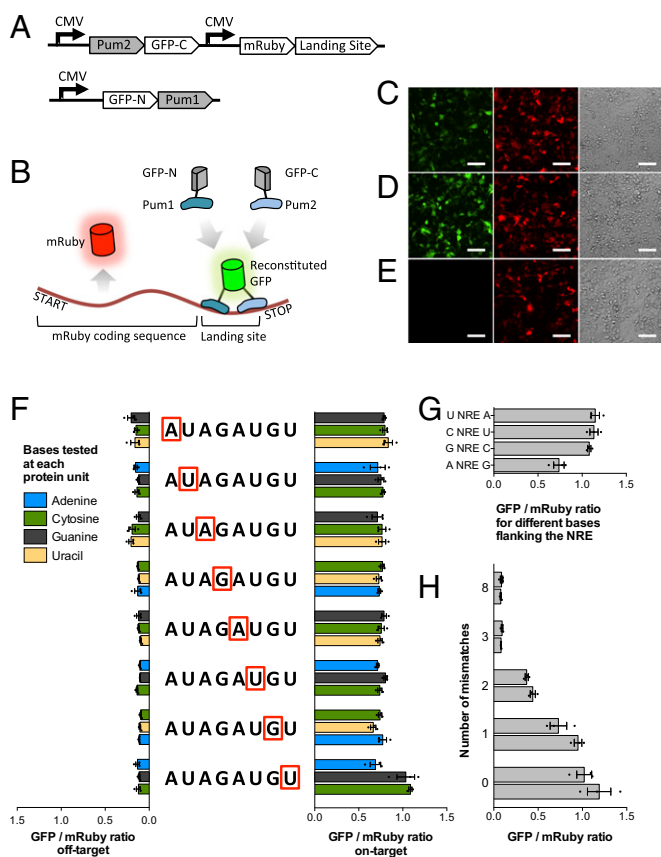
Fig. 2. Evaluation of PumHD mutant binding, with every unit mutated to target each of the four RNA bases. (A) Schematic of the plasmids used in the binding assay for validating the PumHD consensus sequence of Fig. 1C. (B) Schematic of the binding event that results from using the plasmids in A. A PumHD variant (denoted Pum1) and the wild-type PumHD (denoted Pum2) are each fused to one part of split GFP (for a full list of the sequences used in this figure, see *SI Appendix,* Table S2; for full statistics and *n* values of replicates, see *SI Appendix,* Table S1). Pum1 and Pum2 each target one 8-mer sequence within the landing site inserted before the stop codon of mRuby. The mRuby landing site transcript serves as a scaffold for GFP reconstitution upon PumHD binding, and the mRuby protein provides a control for overall cell density and transfection efficiency. (C–E) Representative fluorescent microscopy images of HEK293FT cells expressing the system of A, showing the green (GFP), red (mRuby), and bright-field channels for the same cells. (C) The transfected construct is PumHD with module 7 mutated to bind U (abbreviated 7-U), with on-target RNA present. (D) The transfected construct is PumHD 4-C, with on-target RNA. (E) The transfected construct is PumHD 4-C, with off-target RNA present. (All scale bars, 100 μm.) (F) Binding of on-target vs. off-target PumHD variants. We varied the target sequence of Pum1, changing each unit in turn to target each of the four bases of RNA, according to the key amino acid consensus sequence in Fig. 1C. The starting target sequence for Pum1 was 3′-AUAGAUGU-5′, which we mutated unit by unit to test the targeting of three other bases, at each position. Each cluster of three horizontal bars in this panel corresponds to the test results for the unit framed in red; the colors of the bars (blue, green, black, and yellow) indicate the specific base targeted according to the color key at left. The readout for this assay is fluorescence from reconstituted split GFP, normalized to mRuby expression. Bars to the right show the GFP/mRuby ratio for on-target Pum1 (i.e., in which the protein sequence exactly matches the RNA target in the landing site), and bars on the left show the ratio for off-target Pum1 (i.e., in which there are eight out of eight possible mismatches between Pum1 and its RNA target). (G) GFP/mRuby ratios for wild-type PumHD tested against the wild-type target sequence (called the NRE) flanked by different adjacent nucleotides. The bar at bottom, A NRE G, is for the pair of flanking bases used in the rest of Fig. 2. (H) Tolerance of PumHD to protein–RNA mismatches. We tested two Pum1 sequences against RNA targets with zero, one, two, three, and eight mismatches. Values throughout this figure are mean ± SEM.

GFP reconstitution (Fig. 3C; compare with Fig. 3D, which shows another example of the failed unit 3/stacking amino acid Y candidate highlighted in *SI Appendix,* Fig. S6B), which resulted in the hypothesized Pumby (Pumilio-based assembly) module set of Fig. 3B.

We validated the hypothesized Pumby module set of Fig. 3 using GFP reconstitution as in Fig. 2 (for a full list of the sequences used in Fig. 4, see *SI Appendix,* Table S4). We found that, for Pumby-based chains that were eight units long (abbreviated Pumby8 below), on-target Pum pairs resulted in significantly higher GFP reconstitution compared with off-target pairs (Fig. 4A; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and specific target sequence; $n = 3$ biological replicates each; see *SI Appendix,* Table S3 for full statistics for Fig. 4), as it had for the PumHD variants (Fig. 2F). We also explored the effect of varying flanking bases around the Pumby target sequence (as for PumHD variants in Fig. 2G) and again found successful binding, albeit with, as expected, quantitative differences in GFP reconstitution magnitude (Fig. 4B). We used purified PumHD variants as well as Pumby8 chains to measure K_d for on- vs. off-target pairs, obtaining K_d's in the nanomolar range for both Pumby8 and PumHD variants (*SI Appendix,* Fig. S8, Table S16, and *Supplementary Results*); off-target pairs had no detectable binding. We performed off-by-one, -two, and -three mismatch assessment for Pumby8, as we did for PumHD earlier, and found that some split GFP reconstitution was observed for one or two mismatched units, implying some degree of Pumby8 mismatch tolerance, but three mismatches did not support GFP reconstitution any more than did completely different (i.e., off by eight bases) RNA sequences (Fig. 4C; $P = 0.9999$ for comparison of three vs. eight mismatches; Dunnett's post hoc test across values of mismatch number, after the previous ANOVA; $n = 3$ biological replicates). We investigated the stability of Pumby8 proteins compared with PumHD proteins that bind the same RNA target sequence. We used a thermal assay, the measuring of fluorescence of SYPRO Orange as it is bound by unfolding protein. The resulting melting curves show that all Pum variants have a melting temperature (T_m) between 50–60 °C, Pumby8 and PumHD alike (*SI Appendix,* Fig. S7; for a full list of the sequences used in *SI Appendix,* Fig. S7, see *SI Appendix,* Table S15).

Having demonstrated the performance of Pumby chains eight units long (Pumby8 for short), we next explored Pumby chains that could bind to shorter or longer RNA sequences—ranging in length from 6 to 18 units long (denoted Pumby6 to Pumby18). We found that, for Pumby-based chains of variable length, on-target pairs resulted in significantly higher GFP reconstitution compared with off-target pairs (Fig. 4D; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and specific target sequence; $n = 3$ biological replicates; full statistics in *SI Appendix,* Table S3). The Pumby chains ranging from length 6 to length 18 were similar to Pumby8 in terms of their GFP reconstitution effects (Fig. 4D; statistics in *SI Appendix,* Table S3). Thus, Pumby modules can indeed support the generation of RNA-binding proteins that are specific and that are longer in length than wild-type PumHD, which have efficacy comparable to the 8-mer Pumby (Fig. 4A). We also explored sequences shorter than Pumby8, synthesizing and testing Pumby chains that were six units long, and found on-target pairs to yield significantly higher GFP reconstitution than off-target pairs (Fig. 4E; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and specific target sequence; $n = 3$ biological replicates), with no difference between any of the Pumby6's tested and the 4-U variant—that is, the equivalent of the truncated wild type, which was assessed in Fig. 4D ($P > 0.05$, Dunnett's post hoc test across specific target sequence for the ANOVA above).

We next developed a novel use of programmable RNA-binding proteins: the monitoring of translation in live cells. Our initial experiments showed how Pum proteins can recruit split GFP to produce green fluorescence in the presence of a target RNA (as in
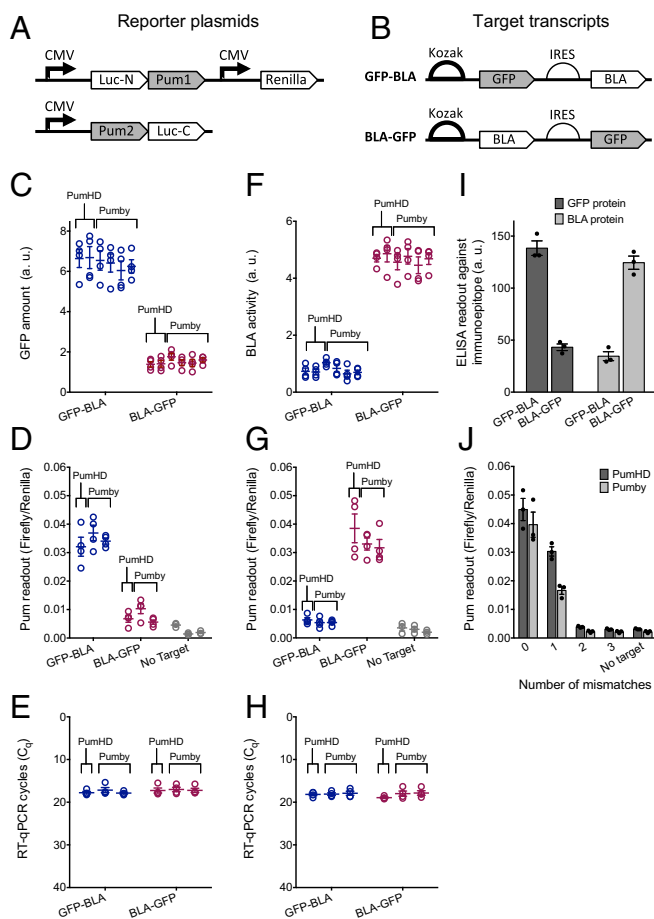
Fig. 3. Pumby: a proposed modular protein architecture for RNA binding. (A) Schematic representation of a protein architecture where concatenated chains of stereotyped Pumilio modules can bind target RNAs of variable length and sequence. (B) A proposed universal set of four modules, each of which can bind one RNA base when situated in any location in the chain of A. We call them Pumby modules. (C and D) Representative fluorescent microscopy images of HEK293FT cells expressing the system of Fig. 2A, showing the green (GFP), red (mRuby), and bright-field channels for the same cells. (C) The transfected construct is an eight-module Pumby chain (Pumby8 7-U, with on-target RNA; see *SI Appendix,* Table S4 for a full list of sequences used in this figure). (D) The transfected construct is an on-target but failed alternative version of Pumby, in which we concatenated unit 3 of PumHD and used stacking amino acid Y. (All scale bars, 100 μm.)

Fig. 2B). We only observed this useful result, however, when the target site was located within an open reading frame (ORF); putting a stop codon upstream of the target site resulted in no detectable GFP reconstitution (*SI Appendix,* Fig. S11). We hypothesized that, in the former case, ribosomal translation repeatedly displaces Pum-reconstituted GFP and allows for new split GFP halves to be bound to the newly freed sites, and reconstituted. Higher translation, thus, would produce a greater amount

of GFP reconstitution. We found similar results in a preliminary investigation using the endogenous gene ATF4 (*SI Appendix,* Fig. S2 and see *SI Appendix, Supplementary Results*) and thus designed a variation of this experiment that could hone in on the translation process itself. We used split firefly luciferase fused to split inteins (27–29) (Fig. 5A), which relies on protein splicing to produce a functional luciferase protein after the two halves are brought together by Pum binding to mRNA. To assess translation level independently from mRNA expression level, we devised Pum targets (8 nt in length) on the genes for GFP and β-lactamase (BLA; see *SI Appendix,* Table S6 for a full list of the protein and RNA sequences used in Fig. 5 and *SI Appendix, Supplementary Results* for how these target sequences were chosen). Expression of these genes was controlled by a Kozak sequence and an internal ribosome entry site (IRES), with the genes in either one order (GFP-BLA, Fig. 5B) or the reverse (BLA-GFP, Fig. 5B). The amount of protein expressed by the cells was roughly five times higher when the corresponding gene was immediately downstream of the Kozak sequence,

Fig. 4. Evaluation of Pumby targeting, with chains containing different numbers of Pumby modules, and different RNA targets. (A) As in Fig. 2F, but for 8-mer Pumby chains (also called Pumby8) targeting a reference RNA sequence as well as variants with every base changed to each of the other three bases of RNA (see *SI Appendix,* Table S3 for the full statistics associated with this figure and *SI Appendix,* Table S4 for the full list of sequences). (B) As in Fig. 2G, but for Pumby8 flanked by various combinations of bases upstream and downstream of the NRE (the RNA target of wild-type PumHD). (C) As in Fig. 2H, but now investigating the tolerance of Pumby8 to protein–RNA mismatches, with zero, one, two, three, and eight mismatches. (D) GFP reconstitution for on-target and off-target Pumby-RNA pairs, for Pumby chains of varying length. The 18-mer chain was tested against the sequence UUCGGCGGAAUGAUGGUU; the 6-mer assembly was tested against AUGGUU (i.e., the last six bases of the 18-mer). All other assemblies were tested against intermediate truncations of the 18-mer target sequence. (E) As in A, but now for 6-mer Pumby chains. Values throughout this figure are mean ± SEM.

Adamala et al.

Fig. 5. Pumby-mediated monitoring of RNA translation in live cells. (*A*) Schematic of reporter plasmids used to measure translation. The plasmids encode for two Pum proteins (designed to bind to various sequences within the target RNAs shown in *B*), each fused to half of split firefly luciferase. One plasmid also encodes for a control gene, Renilla luciferase, which helps quantify transfection efficiency and cell density. (*B*) Schematics of two different target mRNAs used to systematically test the Pum vectors shown in *A*. Only one of the two target mRNAs is used in each experiment. The mRNAs contain sequences encoding for GFP and BLA behind strong (Kozak sequence) or weak (IRES) translation start positions. They are labeled GFP-BLA and BLA-GFP for the (GFP strong, BLA weak) and (BLA strong, GFP weak) conditions, respectively. Three Pums were targeted to each of the two ORFs, aiming for stretches of RNA with low secondary structure (see *SI Appendix*, Table S6 for a full list of the sequences used in this figure, *SI Appendix*, *Supplementary Results* for more on how these sequences were chosen, and *SI Appendix*, Table S5 for full statistics). (*C*) GFP levels (arbitrary units) measured for cells transfected with either GFP-BLA or BLA-GFP (with the choice of target transcript marked on the *x* axis), as well as both reporter plasmids. $n = 4$ biological replicates. (*D*) Firefly luciferase reconstitution (normalized to Renilla luciferase levels) mediated by Pum reassembly on RNA scaffolds, for three Pum binding sites in the GFP sequence, for cells transfected with either GFP-BLA or BLA-GFP (or no target) as well as both reporter plasmids from *A*; $n = 4$ biological replicates for the GFP-BLA and BLA-GFP cases; $n = 3$ biological replicates for the case of no target. (*E*) RT-qPCR measurement of the GFP transcript for the experiments of *D*, where C_q is the quantification cycle (50) ($n = 4$ biological replicates). (*F*) As in *C*, but for BLA activity (from the same set of biological replicates). (*G*) As in *D*, but for Pum binding sites in the BLA sequence. (*H*) As in *E*, but for the experiments of *G*. (*I*) Amount of GFP or BLA protein for cells transfected with one of the two target transcripts from *B*, as measured by ELISA against a small immunopeptide (6xHis) fused to either BLA or GFP. (*J*) Sensitivity of translation measurement to mismatches between Pum proteins and their target RNA. We tested two proteins in the role of Pum1 (a Pumby8 against a target in GFP and a PumHD against a target in BLA), each paired with the target transcript that would create high expression for their target gene (GFP-BLA and BLA-GFP, respectively), and varied the target RNA to have zero, one,

compared with when it was immediately downstream of the IRES; this was observed for both GFP (Fig. 5*C*; $P < 0.0001$ for factor of GFP location; two-way ANOVA with factors of GFP location and Pum type; see *SI Appendix*, Table S5 for the full statistics for Fig. 5, as well as n values for replicates) and for BLA (Fig. 5*F*; $P < 0.0001$ for factor of BLA location). The amount of translation did not depend on whether a Pumby8 or a PumHD was targeted to the mRNA sequence (Fig. 5*C*, $P = 0.6517$ for factor of Pum type; two-way ANOVA with factors of Pum type and GFP location; Fig. 5*F*, $P = 0.7198$ for factor of Pum type in the analogous BLA case).

We sought to independently verify the results of these translation measurements in a way that did not depend on the reporter nature of the two proteins (GFP and BLA) that we used in our demonstration but that could potentially apply to any protein. Thus, we fused GFP and BLA to 6xHis, an immunoepitope, and measured expression levels with ELISA, a standard way of gauging protein levels (Fig. 5*I*). As in the reporter-based readout (Fig. 5 *C* and *F*), we saw that the gene behind the Kozak sequence consistently yielded higher levels of protein production than the one behind the IRES (Fig. 5*I*; $P = 0.00024$ for variations in GFP protein level caused by position in the target transcript; $P = 0.0003$ for BLA protein level; multiple t tests using the Holm–Sidak method; see *SI Appendix*, Table S5 for full statistics). Thus, we were able to validate through both direct reporter detection and ELISA immunoepitope quantitation the modulation of translation by gene position in our constructs.

Having validated our assay, we next assessed the hypothesis that Pum-mediated luciferase reconstitution could also measure protein translation. We observed greater Pum-mediated luciferase reconstitution when Pums targeted the coding sequence behind the Kozak sequence than behind the IRES (for landing sites within GFP, Fig. 5*D*; $P < 0.0001$ for factor of GFP location; two-way ANOVA with factors of Pum type and GFP location; for landing sites within BLA, Fig. 5*G*; $P < 0.0001$ for factor of BLA location). Pumby8 and PumHD showed indistinguishable behavior in this experiment (Fig. 5*D*; $P = 0.5261$ for factor of Pum type; two-way ANOVA with factors of Pum type and GFP location; Fig. 5*G*; $P = 0.0854$ for factor of Pum type in the analogous BLA case). We verified that in these experiments mRNA levels were unaffected by the order of the ORFs within, using reverse-transcription quantitative PCR (RT-qPCR) to quantitate the amount of target transcript mRNA (Fig. 5*E*, $P = 0.2589$ for factor of GFP location; two-way ANOVA with factors of Pum type and GFP location; Fig. 5*H*, $P = 0.5634$ for factor of BLA location). The RT-qPCR mRNA counts for GFP were indistinguishable when Pumby8 vs. PumHD were used (Fig. 5*E*; $P = 0.6236$ for factor of Pum type; two-way ANOVA with factors of Pum type and gene order; Fig. 5*H*; $P = 0.1092$ for factor of Pum type). Thus, the Pum-based reconstitution assays, and the more conventional protein measurement assays above, represent mRNA translation and not mRNA transcript copy number change.

Next, we tested the tolerance of our translation monitoring assay, assessing mismatches between the Pum protein and its target RNA sequence. We mutated two particular Pum sequences (one PumHD and one Pumby8) to contain zero, one, two, or three mismatches (Fig. 5*J*). We observed some luciferase reconstitution above baseline when one unit was mismatched, but in this case even two mismatches were sufficient to effectively eliminate luciferase reconstitution (Fig. 5*J*; $P > 0.99$ for comparison of two or three mismatches vs. No Target; Dunnett's post hoc test across values of mismatch number, after two-way ANOVA with factors of Pum protein and mismatch number).

two, or three mismatches; we also included a case in which the target transcript was absent. Circles (*C*–*H*) and dots (*I* and *J*) represent individual data points; error bars show mean ± SEM.
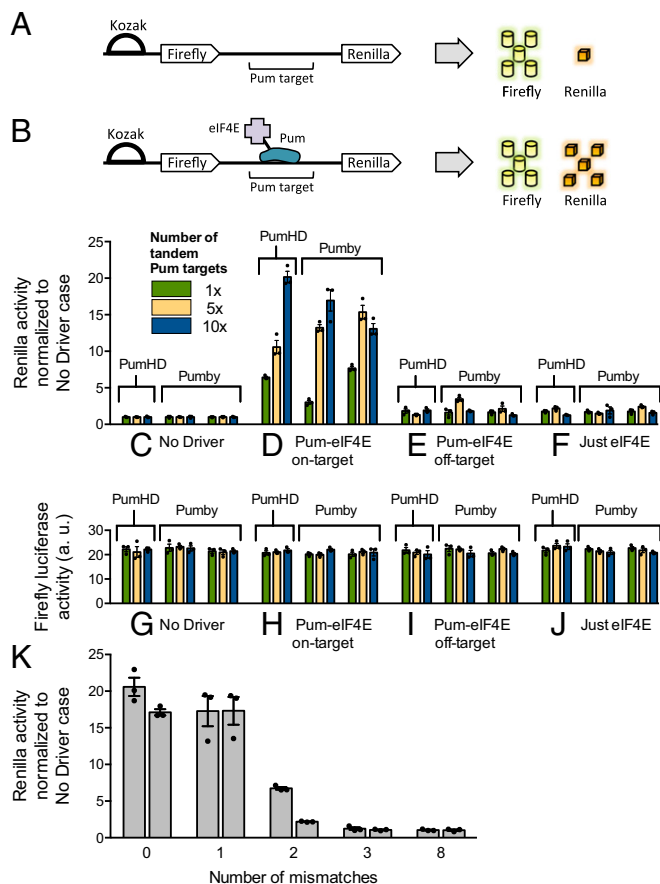
Fig. 6. Gene translation targeted to specific sequences by modular RNA-binding proteins. (A) Schematic of a reporter transcript containing genes for firefly and Renilla luciferases, with a Kozak sequence immediately upstream of firefly luciferase but not of Renilla luciferase; under these conditions, the Renilla ORF yields much lower levels of translation (30). (B) Schematic of how expressing translation initiation factor eIF4E fused to a Pum protein (from a separate driver plasmid) could in principle be used to drive translation of a downstream ORF, causing in this case the production of more Renilla luciferase. (C–F) Renilla luciferase activity as a measure of Pum-eIF4E–mediated translation initiation facilitation, using reporter transcripts bearing three different Pum target sites, in tandem repeats of 1, 5, or 10 copies in a row, in conjunction with various different driver plasmids. The data in C–F were normalized to their respective means in C (for a full list of the target binding sequences used in this figure, see SI Appendix, Table S8; for full statistics, see SI Appendix, Table S7). Specifically: C, Renilla levels when only the reporter plasmid of A is used, with no driver plasmid. (D) Renilla levels when the reporter plasmid of A is used with an on-target driver plasmid, as in B. (E) Renilla levels when the reporter plasmid is used with an off-target driver plasmid. (F) Renilla levels when the reporter plasmid is used with a driver plasmid where eIF4E is present but not fused to Pum. (G–J) Firefly luciferase activity, from the first ORF of the bicistronic luciferase vectors. (K) Sensitivity of translation initiation to mismatches between Pum-eIF4E and the RNA target. We tested two Pum proteins against targets with zero, one, two, and three mismatches. $n = 3$ biological replicates; values throughout are mean $\pm$ SEM.

Another useful mRNA operation is translation initiation, previously demonstrated by fusing wild-type PumHD (or two of its mutants) to the translation activation factor eIF4E (30, 31). We assessed the performance of Pumby in this context by simultaneously measuring the expression of two ORFs from a single transcript (Fig. 6A). We created a transcript containing a Kozak sequence, a firefly luciferase ORF, and a Renilla luciferase ORF, in that order. The Kozak sequence allows translation of the more proximal firefly ORF, with only a weak spillover effect on the Renilla ORF. Between the ORFs were one of three

mRNA target sequences (for PumHD or Pumby binding), present in 1, 5, or 10 copies. We combined this target transcript with various Pum–eIF4E fusion proteins to drive translation (Fig. 6B; one Pum was a PumHD variant and two were Pumby8 chains; see SI Appendix, Table S8 for a full list of the sequences used in Fig. 6). We found that, compared with baseline Renilla expression with any of the nine target vectors on its own, expression with the correct on-target Pum–eIF4E driver increased Renilla luciferase translation by about an order of magnitude (Fig. 6 C and D; $P < 0.0001$ for post hoc comparison of these two conditions; Tukey's post hoc test after three-way ANOVA with factors of copy number, driver plasmid, and Pum type used throughout this paragraph; see SI Appendix, Table S7 for full statistics related to Fig. 6, as well as n values of replicates). More tandem repeats led to higher boosts in expression; for example, the 10× array produced several times higher expression than the 1× (Fig. 6D; $P = 0.0006$ for post hoc comparison of these two conditions). In contrast, expression was indistinguishable from baseline for off-target Pum proteins fused to eIF4E (Fig. 6E; $P = 0.9899$ for post hoc comparison of these two conditions), or for eIF4E administered alone (Fig. 6F; $P = 1$ for post hoc comparison of these two conditions). As a control, firefly luciferase activity did not vary with target copy number or Pum type (Fig. 6 G–J; $P = 0.7826$ and $P = 0.4676$ for each factor, respectively). Thus, Pum proteins make it possible to up-regulate translation of proteins without any need for modified translation initiation sites. We found that Pumby8 and PumHD had the same effect as each other throughout this experiment (Fig. 6 C–F, $P = 0.4656$ for factor of Pum type; Fig. 6 G–J, $P = 0.4676$ for factor of Pum type).

We tested the tolerance of our translation initiation assay to mismatches between the Pum protein and its target RNA sequence (Fig. 6K). We mutated two particular Pum sequences to contain zero, one, two, or three mismatches. We observed some translation above baseline for one or two mismatched units, but three mismatches were sufficient to effectively eliminate the Pum–eIF4E translation boost (Fig. 6K; $P = 0.9998$ for comparison of three vs. eight mismatches; Dunnett's post hoc test across values of mismatch number, after a two-way ANOVA with factors of Pum protein and mismatch number; $n = 3$ biological replicates each). Thus, Pums can mediate target-specific translation initiation, as well as PumHD. We also tested our Pum proteins in an assay for gene silencing (see SI Appendix, Fig. S4 and Supplementary Results), as well as further tests of Pum orthogonality (see SI Appendix, Fig. S3 and Supplementary Results) and found in all cases equivalent performance between Pumby8 and PumHD. Thus, through all these experiments we showed that PumHD and Pumby modules can enable a wide variety of protein-mediated mRNA measurements and perturbations, to be easily performed on unmodified mRNA sequences. We also discovered a new use of such RNA-binding proteins, the monitoring of translation level in living cells.

Discussion

We have discovered a modular protein architecture comprising four protein building blocks derived from the Pumilio protein that enable universal RNA targeting and engineered it for concatenation in chains ranging from 6 to 18 modules in length. Previous works had demonstrated, using proteins that bind to specific RNA sequences, the measurement of mRNA expression level (23, 24), imaging of mRNA dynamics (23–26, 32), and enhancement and suppression of mRNA translation (6, 30, 31, 33) with variants of natural RNA-binding proteins. We demonstrated that our Pumby architecture, which uses a single repeated module to support protein generation (analogous to the TALE design), enables performance equivalent to the original Pumilio protein. We also demonstrate a novel application of modular mRNA-binding proteins—the measurement of translation in live cells. This simple and modular technology may support, as the

ability to systematically map the static distribution of RNAs in situ becomes available (34, 35), the dynamic mapping and control of RNAs to assess their causal role in cellular processes such as those explored here. Pumby was able to support specific binding, with sequences differing by as few as two or three bases resulting in less, or even functionally zero, binding.

A significant part of this functionality in Pumby results from its modular architecture, which makes it possible to target sequences of varying length, not just eight bases long like with the wild-type Pumilio. Longer target sequences are less likely to be found at random in the transcriptome, which helps avoid off-target effects. Furthermore, some investigations require the recognition of a long target: Differentially spliced or highly repetitive transcripts, in particular, can only be uniquely identified through sequences longer than their constitutive parts. Pumby may allow for the creation of varying-length footprints for protection against nucleases or other RNA-binding proteins and may provide a malleable tool for tuning the energy balance of RNA secondary structure in living cells. Many engineering applications are also possible, such as assembling complex scaffolded protein-based reaction pathways in mammalian cells in an RNA-programmable fashion, as has been done before in bacteria (36).

RNA takes on complex secondary structures in live cells and is frequently bound by endogenous RNA-binding proteins; this behavior affects all technologies that rely on in vivo interactions with RNA. Pum proteins are no exception to this rule, and our use of several arbitrary target sequences should not be interpreted as evidence that any arbitrary Pum sequence will bind successfully, or that a Pum protein that worked in one cellular environment will work in all others. In our experiments, roughly three-fifths of the protein sequences we tested in a new RNA context behaved as expected (see *SI Appendix, Supplementary Results* for details on how this occurred in the context of our translation measurement experiments). Our several mismatch experiments, furthermore, showed that RNA sequences differing by one or two bases from a Pum's target sequence (but not three or more) can result in measurable binding effects using our assays. With these benchmarks in mind, researchers applying PumHD and Pumby to a new experiment should always validate new sequences in their final biological context.

Previous studies had probed whether PumHD variants could bind a wide diversity of NRE mutants. Here, in a single study, we tested PumHD variants binding to all four possible nucleotides at all positions under the same set of conditions. For many applications, especially if the number of bases targeted is not a key issue, or if a modular design is not required, this dataset may help with application of PumHD variants themselves to the mapping and control of RNA functions. Along these lines, other members of the Puf family have also been used to engineer selective binding between functional effector proteins and RNA targets. One of the most extensively studied is the *C. elegans* Fem-3 mRNA-binding factor 2 (FBF-2), which is an analog of PumHD (6, 37–40). Cooke et al. (41) linked wild-type FBF-2 to the translation activator GLD2 to trigger poly(A) signal addition and up-regulate translation in *Xenopus* oocytes. Conversely, they linked the FBF-2 domain to the translational repressor CAF1 to trigger poly(A) removal and subsequent translation down-regulation. Campbell et al. (6) also activated translation in human U2OS cells by fusing the yeast poly(A) binding protein to an FBF-2 protein mutant that targets a specific mRNA segment of the human cyclin B1. Such architectures, if tested with every unit mutated to bind every base, or if they yield single-module building blocks, may present the kinds of utility shown here for the Pumilio protein.

The seemingly simple modular binding nature of PumHD masks a great wealth of complexity in the way that the diverse units of the protein contribute to overall protein binding. For example, it has been observed that stacking residues affect the specificity of base-binding differently at different units, that changes to the three key amino acids binding one base affect binding to neighboring bases as well as at the mutant site, and that C-terminal repeats are in general more specific than N-terminal repeats (6). PumHD variants from yeast and nematodes have been shown to bind nine-nucleobase RNA sequences even though they have only eight protein units (18). Human PumHD may bind the fifth RNA base in its target sequence using different in vivo binding modes depending on the base at that position (42). Pumby presents an array in which all units can be selected from the same set of four modules. Thus, Pumby may present a simplified context in which to insert Pumilio modules to study how specific amino acids contribute to the emergent properties of modular RNA binding, independent of position-specific effects. Such future insights into the architecture of Pumilio may not only provide basic science insights into this interesting class of proteins but also help with the design of next-generation RNA-binding tools.

Materials and Methods

Golden Gate Compatible Mammalian and Bacterial Expression Vectors. We prepared Golden Gate compatible mammalian expression vectors by eliminating BsaI sites from previously used vectors as follows. The human CMV major immediate-early gene enhancer/promoter expression vector, called pCI-CMV-GG, was made from the commercially available pCI vector (Promega) by removing BsaI sites from the CMV region (specifically from the β-globin/IgG chimeric intron located downstream of the enhancer/promoter) and from the ampicillin resistance gene. The BsaI site in the chimeric intron, and thus the introduced mutation, was outside of the two known intron splice sites (43). For lower expression levels we created a vector called pCI-GG-UB, in which we replaced the CMV promoter with the human poly-ubiquitin C (UBC) promoter and introduced a single point mutation to remove the BsaI site from the UBC promoter. The efficiency of the two newly mutated promoters was confirmed by comparing the expression of the firefly luciferase under the original promoters with that under the Golden Gate compatible mutated versions (data not shown). In both cases, the expression levels of luciferase from the original and mutated versions of the promoter were nearly identical. All key sequences are deposited at GenBank (KU900022–KU900031), and all key plasmids will be available from the nonprofit distributor Addgene.

Golden Gate Cloning of PumHD Variants. Our PumHD units were assembled by adapting the Golden Gate protocol from a prior TAL effector study (44). See *SI Appendix*, Fig. S1 for a general scheme of our cloning procedure. We first purchased—as synthetic oligonucleotides (IDT)—four base-specific variants of each of the eight RNA-binding units in PumHD, as well as nonsequence-specific units 0 and 9. The units were designed with BsmBI and BsaI restriction sites at the ends (see *SI Appendix*, Table S19).

To assemble the 10 units (eight RNA-binding units plus units 0 and 9) required for the PumHD architecture as used in Fig. 1, two intermediate pentamer assemblies were first prepared. The Golden Gate reaction (digestion with BsmBI at 37 °C and ligation with T7 ligase at 16 °C, repeated 25 times) created circular pentamers; for each PumHD assembly, one pentamer contained units 0, 1, 2, 3, and 4 and the second pentamer contained units 5, 6, 7, 8, and 9.

Any incorrect, noncircularized assemblies were digested with an ATP-dependent DNase that acts only on linear DNA (Plasmid-Safe ATP-Dependent DNase; Epicentre). The DNase digestion reaction mixture was then used as a PCR template to amplify the linear pentamers. The PCR, performed using Herculase polymerase (Herculase II Fusion DNA Polymerase; Agilent) yielded several unspecific products ("smudged bands"), as was previously described in the case of TAL assembly; this phenomenon has been attributed to polymerases "slipping" on repetitive templates, an occurrence that can be almost entirely avoided by preheating the PCR plus silicone oil to 98 °C and adding Herculase plus dNTPs to the hot mixture through the silicone oil. Pentamer products of the correct size were separated on a 2% (wt/vol) agarose gel and extracted from the gel. Two linear pentamers were assembled into the final construct by the second Golden Gate reaction, using BsaI (digestion with BsaI at 37 °C and ligation with T7 ligase at 16 °C, repeated 25 times) followed by a final digestion with Plasmid-Safe ATP-Dependent DNase. The digestion mixture was used to transform Z-Competent Stbl3 *Escherichia coli* (Zymo). Bacteria were always incubated at 30 °C, because slower growth is reported to prevent scrambling of the repetitive array plasmids. The plasmids were purified using standard Miniprep kits (Zymo). See *SI Appendix, Supplementary Results* and Fig. S1 for details on the design of the cloning procedure.

Golden Gate Cloning of Pumby. Proteins based on the Pumby module were assembled using the general Golden Gate scheme described above, with unit 6 of PumHD used on all positions in the assembly and Tyrosine as AA2 (the stacking amino acid). The full list of sequences used to prepare hexamers for Pumby construction is given in *SI Appendix*, Table S20. One major difference with PumHD is that the total length of Pumby chains may vary; consequently, the four base-specific variants of each Pumby unit were prepared with cloning overhangs to circularize into n-mer cloning intermediates of whatever length was needed. We used cloning intermediates with between three and six units to assemble final Pumby chains of up to 24 units. To create a 10-mer Pumby, for example, we prepared one hexamer and one tetramer to reach the total of 10 units in the final assembly. All bacterial amplification was done at 30 °C, as above. Because of difficulty in sequencing highly repetitive arrays, for each assembly three correct clones were selected, purified, and mixed (to minimize the chance of having undetected mutations because of lack of comprehensive sequencing coverage of the highly repetitive area). See *SI Appendix*, *Supplementary Results* and Fig. S1 for details on the design of the cloning procedure.

Transfections and Cell Culture. HEK293FT and HeLa cells were purchased from ATCC. All cells purchased from ATCC are tested for *Mycoplasma* contamination before shipping. All transfections of HEK293FT (used in all figures except *SI Appendix*, Fig. S4) and HeLa (used in *SI Appendix*, Fig. S4) cells were performed using Mirus X2 transfection reagent, according to the manufacturer's directions. Cells were grown in D10 medium (DMEM, supplemented with 10% vol/vol heat-inactivated FBS, 100 IU penicillin, 100 μg/mL streptomycin, and 1 mM sodium pyruvate). For imaging, cells were grown in Matrigel (Corning)-coated glass 24-well plates. For qPCR, luciferase, and BLA assays cells were grown in polystyrene six-well plates (Greiner Bio-One). In all experiments, cells used were no older than passage 18, typically passage 7–15. All batches of cells were assigned randomly to receive one set of transfected genes or pharmacological conditions vs. another. No blinding was used.

For transfection of cells in 24-well plates, we transfected 250 ng of plasmid with 250 ng of diluent DNA (pUC19 plasmid) to keep the total amount of DNA introduced at 500 ng per well of the 24-well plate. If multiple plasmids were cotransfected, they were always in equal proportion and the total amount of plasmid DNA was always 250 ng per well of the 24-well plate (plus 250 ng of pUC19, for 500 ng of total DNA). At 24 h posttransfection, we always exchanged the cell growth media with fresh D10 to remove any remaining transfection reagent.

PumHD and Pumby Binding in Live Mammalian Cells Measured via Pum-Mediated GFP Reconstitution Normalized to mRuby Red Fluorescence (the "Green Red Screen"). All images (Figs. 2 and 3 and *SI Appendix*, Figs. S6 and S11) were captured using cultured HEK293FT cells after a 60-h incubation posttransfection [48 h at 37 °C followed by 12 h at 30 °C, as has been done in previous split GFP experiments (23, 24)]. All images for samples presented in a given figure were taken with the same light source, filter cubes, and objective settings.

RNA Quantification for Translation Measurement Assays. RNA was quantified by RT-qPCR with a LightCycler480 (Roche), using a CellsDirect One-Step qRT-PCR Kit (Life Technologies). Hydrolysis probes were designed against the sequences of EGFP, BLA, and the N-terminal fragment of split luciferase using the Custom TaqMan Assay Design Tool (Life Technologies). Life Technologies did not disclose the sequence of the probes used in this work. HEK293FT cells were grown in 24-well plates, transfected at ∼70% confluence, and harvested after 24 h. For harvesting, cells were washed with DMEM (Corning), digested with 100 μL 0.05% trypsin-EDTA (Corning) for 5 min, diluted with 800 μL PBS, and transferred to 1.5-mL microtubes. Cells were centrifuged at 200 relative centrifugal force (rcf) for 5 min, resuspended in 1 mL PBS, and counted with a Scepter 2.0 Handheld Cell Counter (Millipore). A given cell number for each condition depending on availability (4,000 cells per condition for half of the biological replicates, 2,000 cells for the other half) was extracted, centrifuged at 200 rcf for 5 min, and resuspended in PBS. The cells were then treated according to the CellsDirect protocol. Briefly, cells from each condition were mixed with lysis buffer and frozen at −80 °C until further use, then lysed, digested with DNase I, and divided into RT-qPCR wells. The 20-μL reactions were carried out in 96-well plates (Roche). Each reaction included steps for reverse transcription (15 min at 50 °C) and 40 cycles of qPCR (30 s at 60 °C). Quantification cycle (C_q) calculations were carried out in the LightCycler480 software by the Fit Points Method (Roche). Statistical analysis of the C_q values was carried out in Microsoft Excel 2011, GraphPad Prism 6, and JMP Pro-11.

For experiments in Fig. 5, the data for GFP, BLA, and Pum-readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). HEK293FT cells for those experiments were harvested 72 h posttransfection.

For the gene silencing experiments of *SI Appendix*, Fig. S4, the Renilla luciferase, firefly luciferase, and RT-qPCR data for each sample were collected from the same biological replicates (HeLa cells grown and transfected at the same time, in adjacent wells of a microwell plate). Cells for those experiments were harvested 48 h posttransfection.

Orthogonality Tests. For the orthogonality tests of *SI Appendix*, Fig. S3, luciferase and APEX2 assays were performed on all technical replicates on the same day, with the same batch of reagents. APEX2 activity served as a transfection control; that is, we screened all our biological samples for peroxidase activity and used its presence as an indicator that the well had been successfully transfected with a target vector. We chose APEX2 for this purpose because it is a modified peroxidase that shows strong activity in the mammalian cytosol and to provide a verifiably translated gene in which to place the landing site. The landing site needed to be within the ORF of a translated gene, in order for a large amount of split firefly luciferase to be reconstituted (as described before for Fig. 2). We intended to exclude any samples that displayed zero peroxidase activity but in the end excluded none of our samples from the study for this reason. APEX2 activity was assayed with an Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit (Invitrogen). Each biological replicate consisted of the HEK cells from one 24-well plate well, transfected with three plasmids encoding the following: Pum fused to N-terminal split firefly luciferase, Pum fused to C-terminal split firefly luciferase, and APEX2 fused to the landing site. All replicates were transfected with the same Pum fused to C-terminal split firefly luciferase, so reconstitution was determined solely by the correspondence between the Pum fused to N-terminal split firefly luciferase and its binding site. Each tile in *SI Appendix*, Fig. S3 presents the average of three biological replicates.

Firefly and Renilla Luciferase Activity Assay. The activity of Renilla luciferase and firefly luciferase was measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's instructions. It is important to note that the measured luciferase activity, especially for the reconstituted split luciferase, differs significantly between experiments if the reconstituted luciferin reagent is allowed to go through more than one freeze–thaw cycle. This has been previously noted by others using a luciferase detection kit based on the same chemistry (29). For results described in this paper, each "batch" of experiments (samples directly compared with each other, that is, all biological replicates in a single figure panel) was analyzed using the same, freshly prepared batch of reagents.

For the translation quantification experiments of Fig. 5, the data for GFP, BLA, and Pum readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). The cell harvesting protocol for those experiments is described in *Materials and Methods*, *RNA Quantification for Translation Measurement Assays*.

For gene silencing experiments of *SI Appendix*, Fig. S4, the Renilla luciferase, firefly luciferase, and RT-qPCR data for each sample were collected from the same biological replicates (HeLa cells grown and transfected at the same time, in adjacent wells of a microwell plate). The cell harvesting protocol for those experiments is described in *Materials and Methods*, *RNA Quantification for Translation Measurement Assays*.

For the translation initiation experiments of Fig. 6, cells were harvested 36 h posttransfection by digestion with Glo Lysis Buffer (Promega), according to manufacturer's instructions.

BLA Activity Assay. The BLA activity assays were performed using GeneBLAzer In Vitro Detection Kit (Invitrogen) according to the manufacturer's instructions. For the translation measurement experiments of Fig. 5, the data for GFP, BLA, and Pum readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). The cell harvesting protocol for those experiments is described in *Materials and Methods*, *RNA Quantification for Translation Measurement Assays*.

Quantitative GFP Assay. The GFP activity was quantitated using GFP Quantitation Kit (BioVision) according to the manufacturer's instructions. For translation measurement experiments of Fig. 5, the data for GFP, BLA, and Pum readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). Thus,

the cell harvesting protocol for those experiments is described in *Materials and Methods*, *RNA Quantification for Translation Measurement Assays*.

His-Tag ELISA Expression Assay. A 6x poly-histidine tag (6xHis) was cloned at the N terminus of the GFP and BLA constructs used in the translation measurement experiments of Fig. 5. We measured expression of these proteins with a 6xHis-tag ELISA Kit (Abcam) according to the manufacturer's instructions.

Measurement of Native ATF4 Translation via Pum-Mediated Fluorophore Reconstitution. For the experiments described in *SI Appendix*, Fig. S2 *B and C*, HEK293FT cells were seeded and transfected with a pair of Pum GFP vectors and imaged as described above for the "green red screen." At 24 h posttransfection, 0.5 μM thapsigargin was added. Cells were imaged again after 12 h, as described above. Each experiment was performed in three biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). ATF4 protein expression was quantified using an ELISA Kit for Activating Transcription Factor 4 (Cloud-Clone Corp.). The cells were harvested at indicated time points and the ELISAs performed according to the manufacturer's instructions. Each experiment was performed in three biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate).

Protein Expression and Purification. A custom Golden Gate compatible bacterial expression vector was prepared, based on the pBadHisB (6xHis tag) vector backbone, removing BsaI site from the BLA coding sequence. Pum arrays were cloned into this vector as described above. His-tagged Pum variants were expressed in *E. coli* strain DH5α, grown in 100 mL RM media induced with 0.005% arabinose, at 18 °C, 200 rpm, for 18–24 h (until the colony reached OD_{600} of 0.7). Bacterial pellets were lysed with BugBuster Protein Extraction Reagent (5 mL per 1 g of wet bacteria paste; EMD Millipore) with lysozyme (0.50 mg/mL final concentration; Thermo Scientific). The proteins were purified using Talon Spin Columns (Clontech). The purified proteins were stored in aliquots in 25% (vol/vol) glycerol at −80 °C.

Binding of Pum Variants to RNA Measured by Fluorescence Anisotropy. We used fluorescence anisotropy to measure the kinetics of binding of purified Pum proteins to their cognate and noncognate RNA. Fluorescence anisotropy is widely used to investigate steady-state, dynamic equilibrium binding between protein and RNA (45–47).

The cognate and noncognate RNA targets for the purified Pum variant proteins were synthesized with 5′-labeled FAM, 6-carboxyfluorescein (IDT). The activity of the purified Pum variants was estimated with a saturation assay for each protein and its cognate RNA as described before (7). Fifty nanomolar cognate RNA was mixed with increasing concentration of the protein (measured by NanoDrop; Thermo Scientific) in the binding buffer (25 mM Tris·HCl, pH 7.5, 0.5 mM EDTA, 50 mM KCl, and 0.1 mg/mL BSA). The 100-μL samples were assayed, in duplicates, for fluorescence anisotropy using a Cary Eclipse fluorimeter (Varian) with Manual Polarizer Accessory (Varian). The cognate RNA is always the sequence exactly matching the whole Pum protein binding sequence, flanked as CCAGAAU*Pum_sequence*UUCG (for full list of sequences, see *SI Appendix*, Table S16) with flanking bases selected according to previously published studies (7, 23). Fluorescence anisotropy was calculated as a unitless ratio defined as $R = (I_= − I_⊥)/(I_= + 2I_⊥)$, where I is the emission intensity parallel ($I_=$) or perpendicular ($I_⊥$) to the direction of polarization of the excitation source. The stoichiometric point of each saturation plot was used to estimate the active protein fraction (See *SI Appendix*, Fig. S8 for example plots). The K_d of each protein to its cognate and noncognate RNA was subsequently measured, using the

protein concentration corrected to the active protein fraction, with constant concentration of RNA. The K_d was calculated from a nonlinear fit in IgorPro 6.22 of the anisotropy vs. protein concentration plot to the equation (48)

$$F([protein]) = \left(\left(\left(\left(([protein]*K_a + [RNA]*K_a + 1)\right.\right.\right.\right.$$
$$\left.\left.\left.- (([protein]*K_a + [RNA]*K_a + 1)^2 - 4*K_a^2*[RNA]*[protein])\right.\right.\right.$$
$$\left.\left.^{(.5)}\right)/(2*K_a)\right)*(F_b − F_0)/[protein]\left.\right) + F_0\right),$$

where [protein] is the concentration of the active fraction of the protein and [RNA] is the RNA concentration. Example anisotropy measurement plots are shown in *SI Appendix*, Fig. S8 and the K_d values for binding of PumHD variants and Pumby to cognate and noncognate RNA are shown in *SI Appendix*, Table S16.

Stability of Pum Variants Measured by a Thermal Shift Assay. The T_m of purified PumHD and Pumby variants was measured using a thermal shift assay with SYPRO Orange (Invitrogen) dye according to the previously described protocol (51). Briefly, the 2.5 μM peptide samples were prepared in 100 mM Hepes (pH 7.4), 150 mM NaCl and 5× SYPRO Orange dye. Fluorescence vs. temperature was measured with a LightCycler480 (Roche) with a ramp rate of 1.2 °C/min. The melting temperature was obtained as a midpoint of the thermal unfolding curve by fitting the slope of the curve to the sigmoid equation in Igor Pro-6.37:

$$F = base + (max/(1 + exp((T_m − x)/(rate)))).$$

The reported T_m is an arithmetic average of four replicates; T_m obtained from all independent replicates was within 1 °C from the reported average value. See *SI Appendix*, Fig. S7 for melting plots and T_m results.

Statistics. The reasoning behind the sample sizes was not based upon a power analysis, because this study is primarily about exploring a new technology. As noted in ref. 49 and recommended by the NIH, "In experiments based on the success or failure of a desired goal, the number of [experiments] required is difficult to estimate...." We want to evaluate how a new technology works, and outcomes are not anticipatable, because the technology has not existed before, to our knowledge. As noted in ref. 49, "The number of experiments required is usually estimated by experience instead of by any formal statistical calculation, although the procedures will be terminated when the goal is achieved." In our case, we attempted to validate the tool by trying many different biological validations, in different contexts, as we have done in the past, to understand the biological impact of the tool in the context of different questions. Each experiment was repeated on a minimum of nine technical replicates; see *n* values given with each experiment.

1. Buxbaum AR, Haimovich G, Singer RH (2015) In the right place at the right time: Visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol* 16(2):95–109.
2. Lionnet T, et al. (2011) A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods* 8(2):165–170.
3. Tyagi S (2009) Imaging intracellular RNA distribution and dynamics in living cells. *Nat Methods* 6(5):331–338.
4. Re A, Joshi T, Kulberkyte E, Morris Q, Workman CT (2014) RNA-protein interactions: An overview. *Methods Mol Biol* 1097:491–521.
5. Chen Y, Varani G (2013) Engineering RNA-binding proteins for biology. *FEBS J* 280(16):3734–3754.
6. Campbell ZT, Valley CT, Wickens M (2014) A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* 21(8):732–738.
7. Abil Z, Denard CA, Zhao H (2014) Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *J Biol Eng* 8(1):7.
8. Coquille S, et al. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat Commun* 5:5729.
9. Filipovska A, Rackham O (2012) Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Mol Biosyst* 8(3):699–708.
10. Moore FL, et al. (2003) Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proc Natl Acad Sci USA* 100(2):538–543.
11. Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 8(6):479–490.
12. Wickens M, Bernstein DS, Kimble J, Parker R (2002) A PUF family portrait: 3′UTR regulation as a way of life. *Trends Genet* 18(3):150–157.
13. Spassov DS, Jurecic R (2002) Cloning and comparative sequence analysis of PUM1 and PUM2 genes, human members of the Pumilio family of RNA-binding proteins. *Gene* 299(1-2):195–204.
14. Wang X, Zamore PD, Hall TM (2001) Crystal structure of a Pumilio homology domain. *Mol Cell* 7(4):855–865.
15. Wang X, McLachlan J, Zamore PD, Hall TMT (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110(4):501–512.
16. Cheong C-G, Hall TMT (2006) Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci USA* 103(37):13635–13639.
17. Zamore PD, Williamson JR, Lehmann R (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA* 3(12):1421–1433.

18. Miller MT, Higgin JJ, Hall TM (2008) Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat Struct Mol Biol* 15(4): 397–402.
19. Qiu C, et al. (2012) Divergence of Pumilio/fem-3 mRNA binding factor (PUF) protein specificity through variations in an RNA-binding pocket. *J Biol Chem* 287(9): 6949–6957.
20. Chen Y, Varani G (2011) Finding the missing code of RNA recognition by PUF proteins. *Chem Biol* 18(7):821–823.
21. Miller JC, et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29(2):143–148.
22. Sander JD, et al. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* 29(8):697–698.
23. Ozawa T, Natori Y, Sato M, Umezawa Y (2007) Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* 4(5):413–419.
24. Yamada T, Yoshimura H, Inaguma A, Ozawa T (2011) Visualization of nonengineered single mRNAs in living cells using genetically encoded fluorescent probes. *Anal Chem* 83(14):5708–5714.
25. Yoshimura H, Inaguma A, Yamada T, Ozawa T (2012) Fluorescent probes for imaging endogenous β-actin mRNA in living cells using fluorescent protein-tagged pumilio. *ACS Chem Biol* 7(6):999–1005.
26. Tilsner J, et al. (2009) Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. *Plant J* 57(4):758–770.
27. Schwartz EC, Saez L, Young MW, Muir TW (2007) Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat Chem Biol* 3(1):50–54.
28. Chong S, et al. (1996) Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J Biol Chem* 271(36):22159–22168.
29. Selgrade DF, Lohmueller JJ, Lienert F, Silver PA (2013) Protein scaffold-activated protein trans-splicing in mammalian cells. *J Am Chem Soc* 135(20):7713–7719.
30. Cao J, et al. (2013) Light-inducible activation of target mRNA translation in mammalian cells. *Chem Commun (Camb)* 49(75):8338–8340.
31. Cao J, Arha M, Sudrik C, Schaffer DV, Kane RS (2014) Bidirectional regulation of mRNA translation in mammalian cells by using PUF domains. *Angew Chem Int Ed Engl* 53(19):4900–4904.
32. Tilsner J (2015) Pumilio-based RNA in vivo imaging. *Methods Mol Biol* 1217:295–328.
33. Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z (2012) Engineering RNA endonucleases with customized sequence specificities. *Nat Commun* 3:1147.
34. Chen F, Tillberg PW, Boyden ES (2015) Expansion microscopy. *Science* 347(6221): 543–548.
35. Lee JH, et al. (2014) Highly multiplexed subcellular RNA sequencing in situ. *Science* 343(6177):1360–1363.
36. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333(6041):470–474.
37. Campbell ZT, et al. (2012) Cooperativity in RNA-protein interactions: Global analysis of RNA binding specificity. *Cell Reports* 1(5):570–581.
38. Wang Y, Opperman L, Wickens M, Hall TMT (2009) Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc Natl Acad Sci USA* 106(48):20186–20191.
39. Opperman L, Hook B, DeFino M, Bernstein DS, Wickens M (2005) A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. *Nat Struct Mol Biol* 12(11):945–951.
40. Bernstein D, Hook B, Hajarnavis A, Opperman L, Wickens M (2005) Binding specificity and mRNA targets of a C. elegans PUF protein, FBF-1. *RNA* 11(4):447–458.
41. Cooke A, Prigge A, Opperman L, Wickens M (2011) Targeted translational regulation using the PUF protein family scaffold. *Proc Natl Acad Sci USA* 108(38):15870–15875.
42. Lu G, Hall TMT (2011) Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure* 19(3):361–367.
43. Matsumoto K, Wassarman KM, Wolffe AP (1998) Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J* 17(7):2107–2121.
44. Sanjana NE, et al. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 7(1):171–192.
45. Shi X, Herschlag D (2009) Fluorescence polarization anisotropy to measure RNA dynamics. *Methods Enzymol* 469:287–302.
46. Heyduk T, Ma Y, Tang H, Ebright RH (1996) Fluorescence anisotropy: Rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods Enzymol* 274: 492–503.
47. Dinman JD, ed (2013) *Biophysical Approaches to Translational Control of Gene Expression* (Springer, New York).
48. Qu X, Chaires JB (2000) Analysis of drug-DNA binding data. *Methods Enzymol* 321: 353–369.
49. Dell RB, Holleran S, Ramakrishnan R (2002) Sample size determination. *ILAR J* 43(4): 207–213.
50. Bustin SA, et al. (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55(4):611–622.
51. Biggar KK, Dawson NJ, Storey KB (2012) Real-time protein unfolding: A method for determining the kinetics of native protein denaturation using a quantitative real-time thermocycler. *Biotechniques* 53(4):231–238.