SUPPLEMENTARY MATERIALS for:

A Cellular High-Throughput Screening Approach for Therapeutic *trans*-Cleaving Ribozymes and RNAi against Arbitrary mRNA Disease Targets

by

Edwin H. Yau[3,2,¶], *Mark C. Butler*[2], *and Jack M. Sullivan*[1-7,*]

[1]Research Service, VA Western New York Healthcare System, Buffalo, NY 14215;
Departments of [2]Ophthalmology (Ira G. Ross Eye Institute), [3]Pharmacology/Toxicology, [4]Physiology/Biophysics, [5]Neuroscience Program, University at Buffalo- SUNY, Buffalo, NY 14209; [6]SUNY Eye Institute; [7]RNA Institute, University at Albany- SUNY

[¶] Current Affiliation: University of California San Diego, Dept. of Hematology and Oncology

[*] Corresponding author

Keywords: high throughput screening; gene therapy; hammerhead ribozyme; RNAi; siRNA; shRNA; DNAzyme; RNA Drug Discovery; hereditary retinal degenerations, autosomal dominant retinitis pigmentosa;

Short title: Cell-Based Screening for Therapeutic Ribozymes and RNAi

Correspondence to:

Jack M. Sullivan, M.D., Ph.D.
Associate Professor of Ophthalmology, Pharmacology/Toxicology, and Physiology/Biophysics,
Ira G. Ross Eye Institute, University at Buffalo- SUNY, Buffalo, NY 14209;
Staff Physician-Scientist,
Veterans Administration Western New York Healthcare System
Medical Research, Building 20, Rm 245
3495 Bailey Ave.
Buffalo, NY 14215

(716)-862-6533
(716)-862-6526 (FAX)
js354@buffalo.edu
jackmsullivanmdphd@yahoo.com (**preferred e-mail**)

SUPPLEMENTARY METHODS

Plasmid Constructions and Cloning. **Construction of pUC-VAL.** pgVaL-Ad (Abdelmaksoud et al., 2009) is a slightly modified version of the original pgVaL plasmid constructed in pGEM7Zf (see Fig. 1 of Lieber and Strauss 1995). In pgVaL the central domain of the adenoviral VAI RNA scaffold is replaced with a large stem structure followed by an expected loop to harbor a hhRz sequence. These engineered changes obviated the role of the central domain in blocking activation of the dsRNA-activated protein kinase (PKR). Such adenoviral VAI- hhRz chimeras are expressed in mammalian cells from a strong intragenic RNA Pol-III promoter (A and B boxes) that is not impacted by engineered changes in the central domain. pgVaL-Ad was optimized for hhRz cDNA cloning efficiency by inserting an adapter to separate the immediately proximate Sal I and Pst I restriction sites of pgVaL into which the hhRzs cDNAs were ligated (Abdelmaksoud et al., 2009). While the pgVaL construct has proven useful, there are several limiting features of this hhRz RNA scaffold and its plasmid design. First, RNA folding analysis suggests that the expected loop element (see Fig. 1 of Lieber and Strauss, 1995) at the end of the highly stable engineered stem domain was not an open loop but had substantial secondary structure (**Fig. 1** in main paper). Second, the pgVaL plasmid harbored a wild type adenoviral VAII construct, of uncertain function. Third, the engineered central domain of the pgVaL vector contains a long (21 bp), GC rich, and stable stem structure, which makes DNA sequencing challenging from either direction. Effort was therefore directed to the design and construction of pUC-VAL, a next generation VAI scaffold plasmid for hhRz expression in mammalian cells.

pUC-VAL hhRz scaffold was designed to improve upon properties of the existing pgVal-Ad vector (Abdelmaksoud et al., 2009), such as creating a well-defined and stable single-stranded loop (not present in pgVaL) to support the function of hhRz RNA sequences, for ease of cloning of hhRz

cDNA inserts, and for DNA sequencing of those inserts. The parent plasmid for pUC-VAL, pNEB193 (2713 bp) (New England Biolabs, NEB #N3051S), a pUC19 derivative with additional restriction sites in the multiple cloning region, was first modified by directionally ligating a phosphorylated Hind III/Pst I adapter into the multiple cloning site (pNEB193-T7, 2731 bp) to create a T7 promoter either to support *in vitro* transcription or to prime forward sequencing. The adapter with Hind III and Pst I overhangs was formed by annealing two oligonucleotides (5' AGCTTTAATACGACTCACTATA***G***GGAGATCTGCA 3'; 5' GATCTCCCTATAGTGAGTCGTATTAA 3'), prior to ligation into pNEB193 plasmid cut with the same restriction enzymes. After ligation the mix was treated with Sph I (a unique restriction site between Hind III and Pst I in pNEB193 but not in the T7 adapter) to negatively select for parental plasmid (without inserts). The adapter also contains a unique Bgl II site (underlined) approximating the T7 transcriptional start point (bolded and italicized). pUC-VAL was generated by first directionally cloning the isolate gene for VAI RNA as a Xba I-BssH II fragment (243 bp) cut from pAdvantage (4392 bp) (Promega Inc., #E1711) into the identical sites of pNEB193-T7 to form pNEB193-T7-VAI. pAdvantage is a plasmid that contains both VAI and VAII genes of human adenovirus 2 within a 1724 bp Sal I/Hind III fragment representing bp 9,831-11,555 of the viral genome. The Xba I-BssH II fragment contains 31 bp of adenoviral sequence upstream of the RNA pol III transcription start of the VAI RNA and 53 bp of adenoviral sequence downstream of the RNA pol III U_4 transcriptional terminator or 12 bp downstream of the RNA pol III U_6 transcriptional terminator. The VAII gene was excluded in this construction. The native central domain of VAI was then replaced with an engineered stem-loop element to create a domain for hhRz insertion within the VAI scaffold RNA. The intent was to generate a large unhindered and flexible single stranded loop, stabilized by a stem, in order to allow the hhRz RNA component of the chimera the best opportunity

for catalytic function. This domain in pUC-VAL was designed using RNA secondary structure analysis algorithm (RNAStructure, vers. 4.2-5.2, Mathews et al., 2004; Mathews, 2006).

The pUC-VAL plasmid construct (pNEB193-VAI-hhRz) was designed in four sequential steps using a series of adapters. Oligodeoxynucleotides were synthesized by Sigma GenoSys or Integrated DNA Technologies, annealed, and phosphorylated using T4 Polynucleotide Kinase (NEB) prior to ligation into linearized vectors. First, an annealed cohesive adapter (5' CTAGACCGGT 3' with 5' GATCACCGGT 3') was ligated between the Xba I and Bgl II sites of pNEB193-T7-VA1 to delete the Sal I and Pst I sites (to form pNEB193-T7-VA1-NoSalPst). Second, there are two Aat II sites in pNEB193-T7-VA1-NoSalPst, one within VAI and the second in an intergenic region of the plasmid. The second site was removed by complete digestion (linearization) with unique EcoO109 I in the intergenic region, then fill-in with T4 DNA polymerase and dNTPs, followed by pulse digestion with Zra I (blunt cutting isoschizomer of Aat II) and religation (to form pNEB193-T7-VA1-NoSalPst-1AatII). Third, an hhRz cloning adapter was cloned into the unique BstE II of pNEB193-T7-VA1-NoSalPst-1AatII. Sequences annealed to form a cohesive overhang adapter are:
5' GTTACCCCCGTGTAATCAACCACATACAATAAGTCGA**C**TTAA***GATATC***TGCAG 3' and
5' GTAACCTGCAGATATCTTAAGTCGACTTATTGTATGTGGTTGATTACACGGGG 3'. BstE II, Sal I and Pst I recognition sites are underlined. Unique Afl II motifs (bold) and EcoRV (bold italic) are included between the Sal I and Pst I sites to select for cloned hhRz or other PTGS cDNAs (see below). Correct orientation of adapter insertion was determined by sequencing of random plasmid DNA minipreps. Fourth, the last adapter was directionally cloned between the Pst I and Aat II sites to complete a 51 nt hhRz harbor and supporting stem. Annealed sequences forming this cohesive overhang adapter are:
5' GTACTACATATACATCCAAACACGGGGGTACCGACGT 3' and

5’ CGGTACCCCCGTGTTTGGATGTATATGTAGTACTGCA 3’.

The size of the pUC-VAL plasmid is 2935 bp, and contains a ColEI bacterial replication and a β-lactamase resistance gene. Unfortunately, the LacZ expression construct of pNEB193 cannot be used for blue/white screening for hhRz cDNA insertion because of constraints that occurred during pUC-VAL construction and the preexisting size of the VAI insert. Schematic of the predicted secondary structures of native VAI RNA, the pg-VaL-Ad RNA, and the pUC-VAL RNAs are shown (**Fig. 1** in main paper).

Efficient Ligation of hhRz cDNAs into pUC-VAL.

Ribozyme cDNAs (<50 base pairs (**bp**)) are generated by annealing and then phosphorylating two oligodeoxynucleotides to form cohesive adapters with Sal I and Pst I overhangs. These are directionally ligated between the cleaved Sal I/Pst I sites in pUC-VAL. Schematics of the standard hhRz RNA and cDNA constructs are shown (**Fig. 2** in main paper). pUC-VAL was digested to completion overnight in NEB 3 buffer at 37ºC with 20 units of Sal I and 20 units of Pst I. Enzymes were heat inactivated at 65ºC for 20 minutes and the preparations cooled on ice. Vector concentrations for ligation were calculated using the Ligate 1.3 program developed in this lab (Misasi and Sullivan, unpublished). This program uses the equations of Dugaciwicz et al. (1975) to calculate the optimal vector and insert concentrations for cohesive end or blunt ligations when the desired construct outcome is linear or circular products. For circular ligation of plasmids the size of pUC-VAL or pg-VaL the concentration of vector for optimum ligation (j/i = 5 or 10, number of moles of free ends of vector/total number of free ends in mix) is 2.16 nM or 1.08 nM, respectively. We used 50-100 fold molar excess of the hhRz (or other small) cDNA insert to drive the ligation reaction forward. T4 DNA ligase (New England Biolabs) was used at 10 units and the reaction conducted at room temperature (20-22ºC) overnight in T4 Ligase buffer, prior to heat inactivation (65ºC for 10

minutes). After ligation and before competent bacterial transformation the ligation mix was subjected to a restriction enzyme digest with EcoRV to negatively select for parental plasmids. Ligase reaction buffer was supplemented to be commensurate with 1X NEB Buffer 3 and split into two equal volumes. One of the reaction tubes received 10 units of Eco RV enzyme and the other received an equivalent amount of buffer. Digestion occurred for 1 hour at 37°C prior to heat inactivation. The two reactions (5 μl) were independent transformed into chemically competent *E. coli* K12 derivative bacteria (GC10 cells, Gene Choice, Frederick MD, > 10^9 transformants/μg; INVαF' cells, InVitrogen, > 10^8 transformants/μg). Transformation mixes were plated at equivalent volume onto LB-Ampicillin (100 μg/ml) plates and cultured overnight at 37°C. Colonies were picked the next day and grown in liquid LB broth with ampicillin (50 μg/ml) (6 mls) for plasmid DNA minipreps (Promega Wizard® Minipreps). Plasmid isolates were digested with Sal I and Pst I to confirm hhRz inserts by gel electrophoresis, or digested with Rsr II (cuts uniquely in our 6 bp extended Stem II hhRz) (linear) or digested with Rsr II and Xmn I to give predicted band sizes on agarose gels. The EcoRV restriction site between the Sal I and Pst I sites in VAI-hhRz vectors allowed for efficient negative selection of parental plasmid during cloning of hhRz cDNA sequences (in different samplings between 86% and 93% of clones selected contained the correct hhRz insert sequence).

***SEAP-RHO* Fusion RNA Construct**. The pSEAP2-Control vector (Clontech #631717, GenBank Accession No: U89938) (5.1 kB) expresses SEAP under the control of the SV40 early promoter. In order to evaluate predicted accessible regions of *RHO* mRNA (or any target mRNA) using the SEAP reporter protein, an Xba I-Fse I adapter was inserted into the 3'UTR of pSEAP2-Control plasmid at the intrinsic Xba I (T↓CTAGA) and Fse I (GGCCGG↓CC) sites to allow for efficient cloning of *RHO* mRNA secondary structure elements (adapter sequence: 5'-

TCTAGAGTC**GATATC**GGGGCGGCCGGCC-3'). The Xba I-Fse I adapter inserts a unique EcoRV restriction site (bolded) for negative selection of parental plasmid (without inserts) during in-frame cloning of *RHO* RNA elements (full Xba I and Fse I restriction sites are underlined). The resultant vector was named pSEAP-Ad. The primary sequence of the human *RHO* mRNA around site 250 (Leu57 codon) is predicted by secondary RNA structural analysis to fold into a secondary structure containing a large (33 nt), accessible single-stranded loop that is stabilized by a supporting stem. This accessible region has supported successful hhRz designs that promote significant and substantial cellular target protein suppression (Abdelmaksoud et al., 2009). The p*SEAP*-STOP-L57*RHO* vector was generated by cloning a partial cDNA that encodes the entire predicted *RHO* 250 region stem-loop secondary structural element into an Xba I/Fse I adapter in the 3'-UTR of pSEAP2-Control vector (adapter sequence: 5'-CTAGATGATAAGTGCTGGGCTTCCCCATCAACTTCCTCACGCTCTACGTCACCGCCAGCACCCCGACGCGGCCGG-3'). With this vector the local region of *RHO* mRNA is encoded as a fusion RNA within the *SEAP* 3'UT region. There is sufficient sequence in this adapter to promote folding of the secondary structural RNA element. This places the inserted *RHO* L57 loop region in the 3'UTR of SEAP mRNA immediately downstream of the SEAP protein STOP codon, allowing for translation of SEAP protein unperturbed from the inserted *RHO* RNA element. Any arbitrary sequence that contains one or more predicted RNA secondary structural elements could be ligated between these Xba I/Fse I sites. A vector schematic is shown below (see Fig. 6A). Cleavage by a hhRz within the fused RNA target element is expected to reduce the stability of the *SEAP* mRNA and lead to suppressed SEAP protein expression and secretion.

Bicistronic *RHO*-IRES-*SEAP* Fusion RNA vector

The bicistronic p*RHO*-IRES-*SEAP* vector was created using the internal ribosome entry site (**IRES**) of the encephalomyocarditis virus present in pIRES2-EGFP vector (Clontech, #6029-1; 5.3 kB). Full-length human *RHO* cDNA (containing the full 5'UT, the coding region, and the full 3'UT up to but not including the first dominant polyadenylation signal) was PCR amplified (primer sequences FOR: 5'-AGTATGGTACCAGATCTAAGAGTCATCCAGCTGGAG-3'; REV: 5'-ATCGTCGACCTACTGTGTGCCCCATTC-3') from a pCDNA3.1 (InVitrogen) plasmid construct containing the full-length human *RHO* mRNA coding sequence and ligated into the Bgl II/Sal I sites (underlined in primers) in the multiple cloning region of pIRES2-EGFP upstream of the IRES element. A unique Eco RI restriction site was added downstream of the IRES element, and the full *SEAP* cDNA from pSEAP2-control vector was ligated downstream as an Eco RI/Mfe I fragment to replace the EGFP cDNA. The SV40 polyA signal element from the original pIRES2-EGFP vector was used without any modification. The vector is designed to transcribe a single bicistronic mRNA that is translated into both *RHO* and *SEAP* proteins, by cap-dependent and cap-independent ribosome translation, respectively. Vector schematic is shown (**see Fig. 7A** in main paper). In this construct a CMV promoter expresses the bicistronic *RHO*-IRES-*SEAP* mRNA in human or mammalian cells. The *RHO* fragment can be replaced with an arbitrary target cDNA to achieve alternative bicistronic expression constructs for other target mRNAs in RNA Drug Discovery.

Construction of shRNA Expression Plasmids

Chemically synthesized and annealed shRNA cDNA constructs for RNAi were directionally ligated into the Bgl II-Xho I sites of the pSUPER-puro vector from Oligoengine (Brummelkamp et al., 2002). shRNA constructs were designed to target *SEAP* mRNA (*SEAP*i965 target sequence: 5'-AGACAUGAAAUACGAGAUC-3') or *RHO* mRNA (*RHO*i725 target sequence: 5'-

UGUUCGUGGUCCACUUCAC-3') or a scrambled control (target sequence: 5'-UGUUCGUGGUCCACUUCAC-3'). A shRNA was also constructed and expressed from pSUPER-puro that recapitulates the Rhoi2 agent previously reported by Cashman et al. (2005).

End-Point RT-PCR

For detection of cellular VAI RNA expression, cytoplasmic RNA was extracted from transfected cell cultures (24-well transient transfections with 2µg of pNEB-VAI-hhRz-1 plasmid) at 24, 48, and 72 hours with Nonidet P-40 lysis buffer (0.5% v/v Nonidet P-40, 50mM Tris HCl, 140mM NaCl, 1.5mM $MgCl_2$) followed by pelleting of nuclei at 300 x g for 2 minutes at 4°C and stabilization in SideStep lysis buffer (Stratagene). SideStep lysis buffer lyses cells to release nucleic acids and stabilizes released nucleic acids by inactivating RNases. This allows for RT-PCR analysis of VAI RNA without the need to purify RNA. Because VAI RNA is small (~150nt), purification of VAI using silica columns often resulted in substantial loss of VAI RNA. DNase treatment was performed by neutralizing the lysis buffer with the addition of an equal volume of neutralization buffer, addition of $MgCl_2$ containing DNase digestion buffer (final $MgCl_2$ concentration 2.5 mM), and treatment with 2U TURBO DNase (Ambion) for 30 minutes at 30°C. DNase was inactivated by the addition of 4x volume of SideStep lysis buffer. cDNA synthesis was performed with 1µL of DNase treated, stabilized cytoplasmic RNA lysate with the AffinityScript RT system (Stratagene) using gene-specific primers for VAI RNA and Glyceraldehyde 3-phosphate dehydrogenase (**GAPDH**) RNA (VAI REV primer: 5'-TACCCCCGTGTTTGGATGTA-3'; GAPDH REV primer: 5'-GGTGCTAAGCAGTTGGTGGT-3') for 45 minutes at 45°C. PCR reaction was carried out with 1µL of cDNA reaction mixture or control (–RT) reaction mixture and *PfuUltra* hotstart PCR master mix (Stratagene) in a total volume of 50µL. Amplification was performed for 20 cycles (30 seconds of denaturing at 95°C, 30 seconds of annealing at 55°C, and 1 minute extension at 72°C). Primers for

VAI: FOR 5'-GTCCGCCGTGATCCATGC-3', REV 5'-TACCCCCGTGTTTGGATGTA-3'. Primers for GAPDH: FOR 5'-CGCTGAGTACGTCGTGGAG-3', REV 5'-GGTGCTAAGCAGTTGGTGGT-3'. The expected amplified products were 170 bp for VAI and 453 bp for GAPDH. Amplification results were analyzed by 2% agarose gel electrophoresis and visualized under UV illumination after ethidium bromide staining.

SUPPLEMENTAL RESULTS

SEAP Assay Development for High Throughput Screening for Lead Agents

HEK293S cell clones were engineered to stably express and secrete *SEAP* (HEK293S-*SEAP* cells). A HTS fluorescent system for assaying *SEAP* enzyme activity was developed. Serial dilutions of purified human PLAP (*SEAP* is an engineered, secreted form of PLAP) were assayed using the substrate 4-MUP to determine the effective linear dynamic range of measurement of the assay under Michaelis-Menten conditions ([substrate, S = 4-MUP] >> [enzyme, E = PLAP or *SEAP*]) ([S] = 50 μM, [E] ≤ 1.2 μM). 4-MUP is ideal as a substrate because it is stable at the high pH (9.8) optimum of *SEAP* enzyme, and because it has minimal background fluorescence (less than 2% of maximum measured fluorescent signal of the dephosphorylated product). *SEAP* efficiently converts (dephosphorylates) 4-MUP to 4-methylumbelliferone, which is strongly fluorescent. Using 50 μM 4-MUP, over a range of PLAP concentrations of 10^{-12} to 10^{-6} M, the *SEAP* assay output has an expected Boltzmann type curve, and can detect enzyme concentrations as low as 1 nM in a 96-well plate format (**Supp. Fig. 1A**). Within this curve, the dynamic range of measurement of 4-MUP conversion was linear over a 3-log PLAP enzyme concentration range from 1.2 nM to 1.2 μM (R^2 = 0.99; F=251.3, p = 0.004; slope = 0.435 ± 0.007; intercept= 1.40 ± 0.02) (**Supp. Fig. 1B**). Initial tests with PLAP operating on an alternative fluorescent phosphatase substrate (DDAO phosphate, Molecular Probes, D-6487, Life Technologies, Grand Island, NY) showed less sensitivity and a much

compressed dynamic range and was not further evaluated (data not shown). A chosen stable HEK293S-*SEAP* cell line secretes levels of *SEAP* at the high end of the linear range of the assay (**Supp. Fig. 1C**) (compare with Fig. 5B) and at a constant rate over 72 hours (R^2 = 0.96, F=66.4, p = 0.015; slope = 0.204 ± 0.004; intercept= 2.39 ± 0.006). The constant protein secretion rate in the HEK293S-*SEAP* cell line indicates a steady-state within the system of cytoplasmic steady-state *SEAP* mRNA concentration (affected by transcription and degradation rates), translation rates, and processing/trafficking streams. In stable expressing cell lines a substantial fraction of potential *SEAP* enzyme dynamic range is lost (e.g. ~2.6 log activity levels at time t = 0 in the selected cell line). We attempt to estimate the percent of dynamic range that is lost due to pre-synthesized *SEAP*. We find that the minimum detectable response (~1.8 log = 63 fluorescence units, Fig. 5A) and a maximum over practical assay time in the chosen stable cell line (~3.1 log = 1258 units, Fig. 5C) yields a total dynamic range of 1195 units. The (minimum) fluorescence seen at t=0 in the chosen cell line (~2.6 log = 398 fluorescence units) is approximately 33% of the total practical measure, which means that on the order of a third of the dynamic range is lost in the assay. Thus, a substantial fraction of *SEAP* mRNA already in the cellular processing stream toward translation cannot be suppressed during transient transfection due to practical delays for a PTGS expression plasmid to reach the nucleus, achieve active transcription, and the subsequent processing and trafficking of the PTGS RNA to the cytoplasm, where it can interact with preformed target mRNA to exert binding/cleavage. The practical dynamic range of *SEAP* assay from extracellular fluid of stable *SEAP* secreting cell lines occurs over about a 40-fold PLAP concentration range at the high end of the measured full linear dynamic range. This range proved sufficient for highly reliable identification of lead PTGS agents, but again with limited dynamic range of measurement. Enhancements would require placing the target expression constructs in the stable cell lines (containing *SEAP* cDNA) under experimentally inducible control (e.g. doxycycline induction system, Tet^{ON}) so that the transcription of the target to

form a steady state target mRNA population could occur only after experimental introduction of modulator following the transfection of PTGS agents (no preformed target mRNA). A simpler way to attempt to partially reopen the dynamic range is to remove preformed *SEAP* containing culture medium (with replacement) immediately after transfection prior to any expression of transfected PTGS RNA, which we used in the current study.

The effect of incubation time on the phosphatase reaction was measured. After addition of the 4-MUP substrate (50 µM), serial fluorescent measurements were made every 30 minutes over twelve hours of incubation time with three different concentrations of PLAP enzyme [30 nM (2.5 units), 60 nM (5.0 units), and 120 nM (10.0 units)] (**Supp. Fig. 1D**). The assay was linear over at least 2 hours of incubation time for the various PLAP concentrations and then began to saturate in a concentration dependent manner. Control (no PLAP) showed no increase in fluorescence, which is expected given that the 4-MUP phosphate group is otherwise stable under the assay conditions. Using the same three concentrations of PLAP enzyme as above, serial fluorescent measurements were also made every 10 minutes over 2 hours of incubation time with 4-MUP substrate (50 µM). When measured every 10 minutes the assay is clearly linear over the first two hours at the three concentrations of PLAP used when conducting the assay in 96-well plates (**Supp. Fig. 1E**). Assay was linear over the measured time ($R^2 = 0.99$, $p<0.001$, $n=13$) for all concentrations of PLAP enzyme. A one hour incubation time was chosen for the standard assay because it is well within the linear range of the phosphatase reaction and there is no evidence for substrate depletion at this point over a range of enzyme concentrations that overlap with those achieved in stable cell lines. Given the linearity of the assay we expect that the levels of measured *SEAP* activity are directly proportional to the levels of secreted *SEAP* protein.

IRES as a Structured RNA Insulator Element

Bicistronic *Target*-IRES-*SEAP* mRNA represents a suitable model for development of PTGS agents against the "natural" target mRNA provided that the IRES and *SEAP* sequences do not influence the folding of the upstream target mRNA component. Using a bioinformatics RNA folding approach we investigated the extent to which the IRES element, intervening between the upstream full-length cDNA of the target and the downstream *SEAP* reporter cDNA, isolates the upstream (disease target) region of the bicistronic mRNA for independent folding. This issue is critical to this strategy because identification of lead PTGS candidates by an initial HTS type screen should occur under conditions that most closely simulate the true *in cellulo* native structure(s) and accessibility of the full length independent target mRNA. In tests of three bicistronic mRNAs engaging three different human retinal mRNA targets of relevance to human disease, the IRES element appears to strongly insulate the folding of the upstream target element from downstream IRES and SEAP components (**Supp. Fig. 2**). IRES elements are known to form stable secondary and tertiary structures that are required for recognition by the ribosome to initiate translation (Witwer et al., 2001). We hypothesized that the folding of the IRES structure occurs independent of the boundary RNA sequences of the upstream target and downstream *SEAP* components in the fusion (bicistronic) mRNA. If this hypothesis is correct, then there should be little impact on IRES RNA secondary structure folding whether these embracing elements are present or not. Similarly, if the IRES is a true RNA isolation element, then the folding of the upstream target will be little influenced whether the target is folded alone or as part of a larger bicistronic mRNA. We used MFold (vers. 3.2) to fold three complete bicistronic mRNAs involving known retinal target mRNAs: *RHO*-IRES-*SEAP*, *RPE65*-IRES-*SEAP*, and *RDH5*-IRES-*SEAP*. *RHO* is expressed in rod photoreceptors and *RPE65* and *RDH5* are both expressed in retinal pigment epithelium. The components of these bicistronic RNAs are shown (**Supp. Table 1**). We output the parameter P_{num}, which shows the frequency of base pairings of a each nt across the entire

span of the mRNA (Zuker and Jacobson, 1998; Zuker, 2003). Larger P_{num} values associated with a given nt indicate that is more likely to form base pairs with a larger number of other nts, or more likely to be accessible or in a single stranded region. Similar P_{num} maps across given regions of primary sequence suggest a similar range of local secondary structures. After folding the complete bicistronic mRNAs, we extracted the target, IRES, and SEAP components (regions) of each bicistronic mRNA fold for comparison to independently folded isolates of the same explicit component of the bicistronic RNA (**Supp. Fig. 2**). The P_{num} map of the upstream target fraction (*RHO*, *RPE65*, *RHD5*) of the bicistronic mRNAs is directly comparable to the P_{num} map of each target folded in isolation (without IRES and SEAP components) and appears little effected by the presence of the downstream components (IRES, SEAP) in the full length bicistronic mRNA for all three constructs. Similarly, the P_{num} map of IRES in isolation is directly comparable to the IRES fraction of a bicistronic mRNA when IRES is interposed between three different mRNAs and the *SEAP* element. The downstream SEAP region does show some variations for all three constructs. These outcomes identify IRES as an *RNA isolator element* that is used here to insulate folding of the upstream target component independent from other elements in the bicistronic fusion mRNA. Strong confirmatory evidence that the IRES folds stably *in cellulo* is identified by the fact that stably or transiently transfected cells abundantly express and secrete *SEAP* protein from the bicistronic mRNA expressed from the *RHO*-IRES-*SEAP* construct. This indicates that appropriate secondary and tertiary structures are formed in the IRES element of the bicistronic mRNA to allow ribosome recognition and initiation of cap-independent translation of the *SEAP* component of the bicistronic RNA.

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Fig. 1. *SEAP* Assay Development and Properties.

Supp. Fig. 1. *SEAP* Reporter Assay Dynamic Range. (**A**) Full range of measurement of the *SEAP* assay using the 4-MUP fluorescent substrate. Serial dilutions of PLAP enzyme were measured on the *SEAP* HTS platform using 96-well plates. (**B**) Linear dynamic range of measurement of the *SEAP* assay. Linear dynamic range of the 4-MUP conversion assay extends over 3-log orders of PLAP concentration (R^2 =0.99). (**C**) Engineered stable HEK293S-*SEAP* cell lines used in the present study express *SEAP* at a high level within the linear dynamic range of the assay, and produce *SEAP* at a constant rate (R^2 = 0.96). With the use of stable cell lines, even though *SEAP* expression levels were constant, there is a fraction (level of activity at time t = 0) of *SEAP* protein that cannot be knocked down by exogenously transfected VAI-hhRz agents due to the time course of hhRz or shRNA gene expression. (**D**) Effect of assay incubation time. Using various concentrations of PLAP enzyme, 4-MUP fluorescent substrate was added and serial measurements of fluorescence were made every 30 minutes for 12 hours using 96-well plates (n=3). The assay appeared linear over at least 3 hours of incubation time for the various PLAP concentrations and then began to saturate. (**E**) When measured every 10 minutes the assay is clearly linear over the first two hours at various concentrations of PLAP using 96-well plates (n=3). We chose to measure after one hour of incubation time with 4-MUP in all subsequent assays as the assay is clearly not substrate or enzyme limited over this time frame.

Supplemental Table 1. Properties of Dicistronic Retinal RNAs.
The size of the target, IRES, and SEAP components is indicated. Dicistronic mRNAs for *RHO*-IRES-*SEAP*, *RPE65*-IRES-*SEAP*, and *RDH5*-IRES-*SEAP* were generated *in silico* based upon the published sequences of the independent elements. GenBank accession sequences for the mRNAs of

human *RHO*, mouse *RPE65*, and mouse *RDH5* are NM_000532 (2767 nt), NM_134006 (1261 nt), and NM_029987 (1833 nt), respectively. For the human mRNA only the first 1532 nt were used that corresponds to the dominant polyadenylated mRNA. Full-length sequences of the upstream cDNAs from transcription start to just before the poly-adenylation signal site (e.g. AAUAAA) were fused with the full-length encephalomyocarditis virus IRES (609 nt), followed by the full length SEAP cDNA and SV40 polyA signal (from pSEAP2-control plasmid (Clontech)).

Supplemental Fig. 2 – RNA Secondary Structural Folding of Dicistronic mRNAs

MFold was run with 5% suboptimality with a window parameter of 3 for a maximum of 99 output structures and the P_{num} file was output. A P_{num}, map shows the frequency of base pairings of each single nt across the entire span of the mRNA. (**A**) *RHO*-IRES-*SEAP*. P_{num} maps of the full dicistronic mRNA (upper figure), and the location of the three components of the dicistronic RNA (*RHO*, IRES, *SEAP*) are indicated. In the middle row the individual elements of the folded dicistronic RNA are expanded (compare the abscissa in the middle row with the abscissa of the full folded dicistronic RNA in the upper aspect of the figure). In the bottom row are shown the corresponding components of the dicistronic RNA that were folded *independently* for comparison to the P_{num} maps of the expanded components *within* the dicistronic RNA. To understand the similarity of the folding within and independent of the dicistronic RNA compare the middle and bottom rows for the respective components of the dicistronic RNA. The *RHO* and IRES components of the *RHO*-IRES-*SEAP* dicistronic mRNA show very similar P_{num} features whether folded as part of the larger dicistronic mRNA or as an individual elements. The *SEAP* element, which is the last element in the dicistronic RNA, show more substantial differences. (**B**) *RPE65*-IRES-*SEAP*. The P_{num} maps of the full dicistronic mRNA and the individual components are shown (as a component of the dicistronic RNA or as an individually folded component). The *RPE65* and IRES components of the *RPE65*-IRES-*SEAP* dicistronic mRNA show similar P_{num} features whether folded as part of the larger

dicistronic mRNA or as an individual elements, while the terminal *SEAP* components show more variability. (**C**) *RDH5*-IRES-*SEAP*. The P_{num} maps of the full dicistronic mRNA and the individual components are shown (as a component of the dicistronic RNA or as an individually folded component). The *RDH5* and IRES components of the *RDH5*-IRES-*SEAP* dicistronic mRNA show similar P_{num} features whether folded as part of the larger dicistronic mRNA or as an individual elements, while the terminal *SEAP* components show more variability. For all targets the P_{num} maps of the upstream target regions of the folded dicistronic mRNAs were very similar to the P_{num} map when the full-length target was folded in isolation. Similarly, the IRES sequence strongly maintains its P_{num} map when folded as part of the dicistronic mRNAs. This strongly suggests that the IRES element insulates the upstream target in the dicistronic RNA.

Supplementary References

Abdelmaksoud H.E.B., Yau E.H., Zuker M., Sullivan J.M., 2009. Development of lead hammerhead ribozyme candidates against human rod opsin for retinal degeneration therapy. *Exp. Eye Res*. 88(5), 859-879.

Brummelkamp T.R., Bernards R., and Agami R., 2002. A system for stable expression of short interfering RNAs in mammalian cells. *Science* 296(5567), 550-553.

Cashman S.M., Binkley E.A., and Kumar-Singh R., 2005. Towards mutation-independent silencing of genes involved in retinal degeneration by RNA interference. *Gene Therapy* 12, 1223-1228.

Dugaiczyk A., Boyer H.W., and Goodman H.M., 1975. Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. *J. Mol. Biol.* 96, 171-184.

Lieber A., and Strauss M., 1995. Selection of efficient cleavage sites in target RNAs by using a ribozyme expression library. *Mol. Cell. Biol*. 15(1), 540-51.

Mathews D.H., Disney M.D., Schroeder S.J., Zuker M., and Turner D.H, 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* 101, 7287-7292.

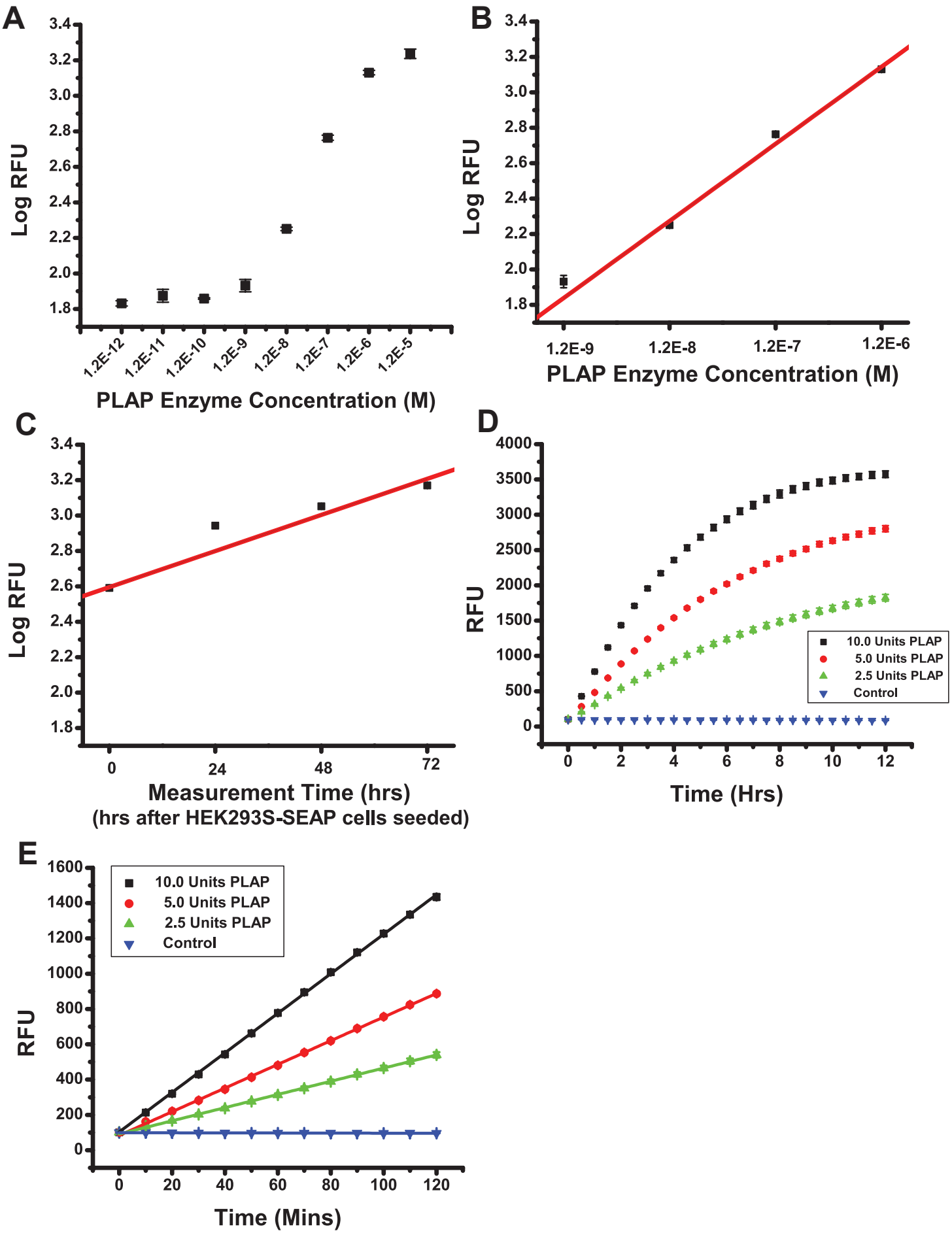
Mathews D.H. 2006. Revolutions in RNA secondary structure prediction. *J. Mol. Biol.* 359, 526-532.

Witwer C., Rauscher S., Hofacker I., and Stadler P., 2001. Conserved RNA secondary structures in picornaviridae genomes. *Nucleic Acids Res.* 29(24), 5079-5089.

Zuker M. and Jacobson A.B., 1998. Using reliability information to annotate RNA secondary structures. *RNA* 4(6), 669-679.

Zuker M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 31(13), 3406-3415.

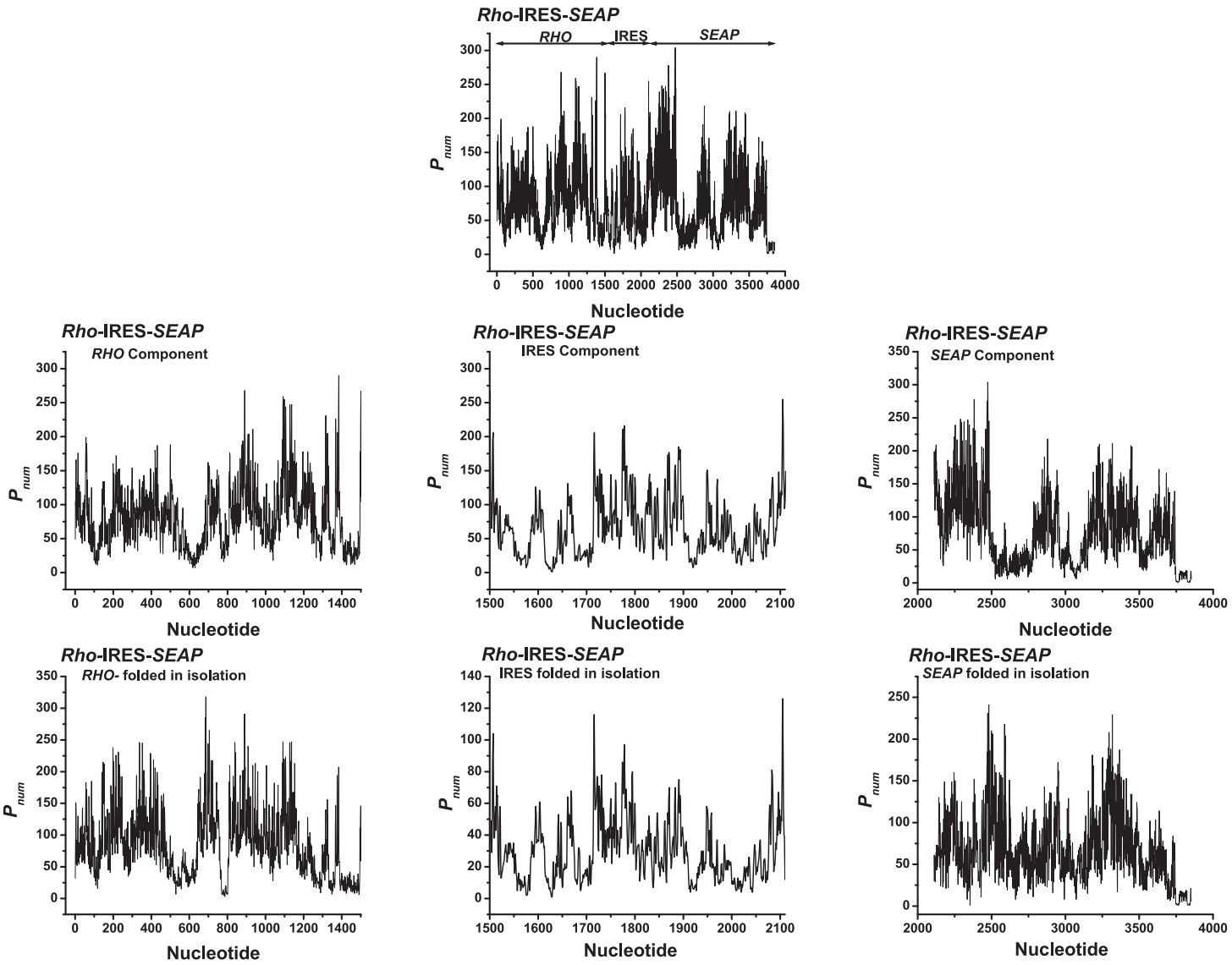
Yau et al. Supp. Fig 1



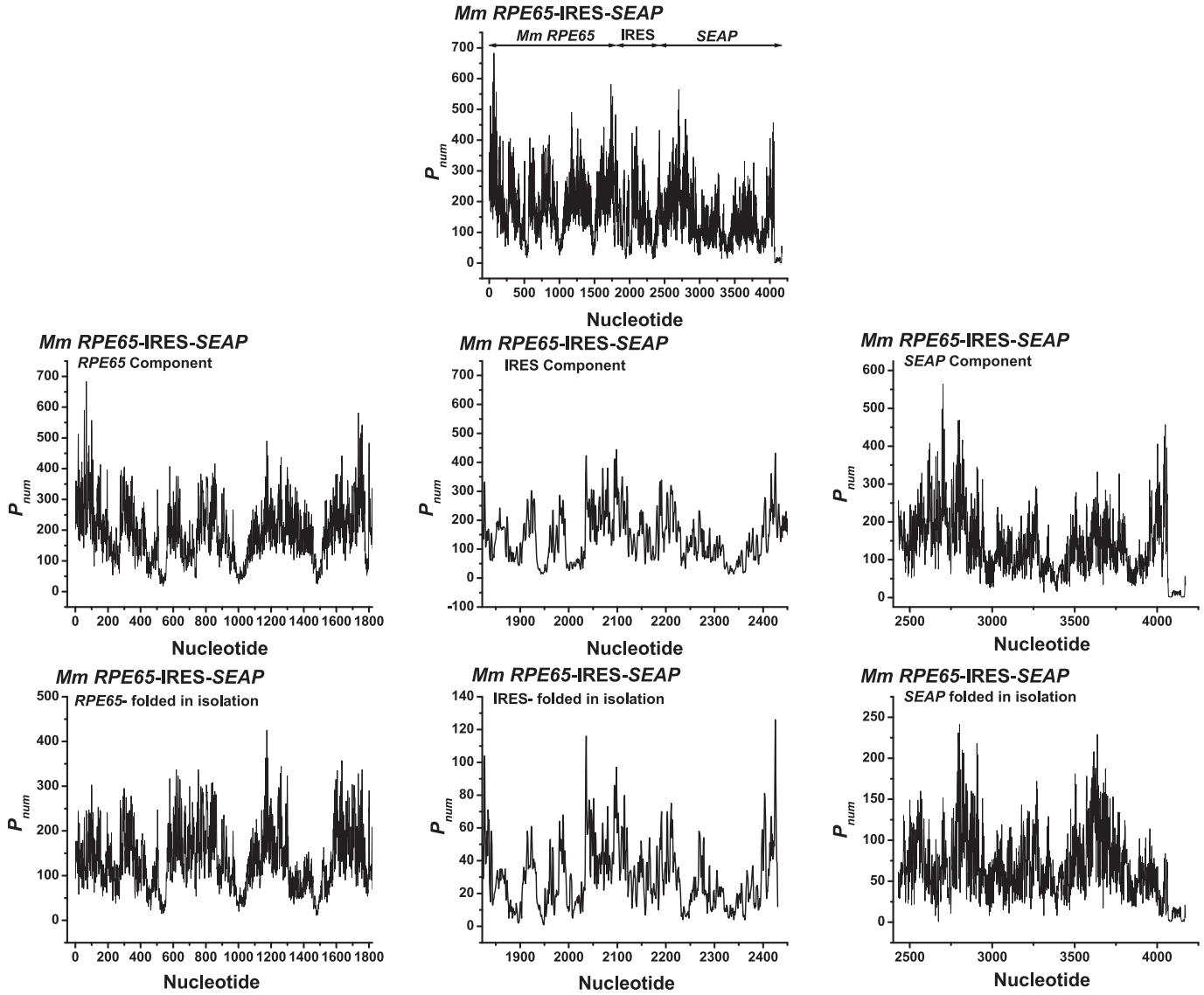
Supplementary Table 1. Properties of Different Model Dicistronic mRNAs *in Silico*

	Target Component	**IRES Component**	***SEAP* Component**	**Total Size**
Size	variable	609 nt	1742 nt	variable
Dicistronic mRNAs				
RHO*-IRES-*SEAP	1-1501	1502-2110	2111-3852	3852
RPE65*-IRES-*SEAP	1-1821	1822-2430	2431-4172	4172
RDH5*-IRES-*SEAP	1-1234	1235-1843	1844-3585	3585

Supplementary Fig. 2A



Supplementary Fig. 2B



Supplementary Fig. 2C