Chemistry & Biology, Vol. 10, 533–540, June, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S1074-5521(03)00109-1

In Vivo Evolution of an RNA-Based Transcriptional Activator

Allen R. Buskirk,[1,2] Polina D. Kehayova,[1,2] Angela Landrigan,[1] and David R. Liu[1,*]
[1]Department of Chemistry and Chemical Biology
Harvard University
12 Oxford Street
Cambridge, Massachusetts 01238

Summary

From random RNA libraries expressed in yeast, we evolved RNA-based transcriptional activators that are comparable in potency to the strongest natural protein activation domains. The evolved RNAs activated transcription up to 53-fold higher than a three-hybrid positive control using the Gal4 activation domain and only 2-fold lower than the highly active VP16 activation domain. Using a combination of directed evolution and site-directed mutagenesis, we dissected the functional elements of the evolved transcriptional activators. A surprisingly large fraction of RNAs from our library are capable of activating transcription, suggesting that nucleic acids may be well suited for binding transcriptional machinery elements normally recruited by proteins. In addition, our work demonstrates an RNA evolution-based approach to perturbing natural cellular function that may serve as a general tool for studying selectable or screenable biological processes in living cells.

Introduction

In addition to its role as a transient carrier of genetic information within a cell, RNA is now known to play a functional role in several biological processes including tRNA processing, intron splicing, and peptide-bond formation during translation [1, 2]. The recent discovery of a class of small RNAs that block translation by base pairing to the 3′-untranslated region of mRNAs reveals that natural RNAs can also regulate gene expression [3]. O'Malley and coworkers recently discovered an RNA that plays a structural role in a protein-RNA complex that coactivates genes regulated by steroid hormone receptors [4, 5]. An RNA that functions as a transcriptional activation domain, however, has not yet been discovered in nature.

The repertoire of natural functional roles played by RNA suggests that a directed evolution approach might enable the discovery of artificial RNA sequences that perturb cellular functions. These intracellularly expressed RNAs may serve as useful probes of complex biological systems and as tools for identifying targets involved in cellular processes of interest. Three recent reports describing random peptide libraries coupled with phenotypic selection [6–8] have shown that peptide aptamers ("peptamers") within natural protein scaffolds can be used in a forward genetics manner to probe the function and mechanism of biological pathways. Although a small number of studies involving the evolution of functional RNAs from random sequence libraries in vivo have been reported [9–11], random RNA libraries have not to our knowledge been evolved in vivo to study natural cellular function.

We envision RNA as offering potential advantages over peptamers in experiments of this type. While the chemical functionality of RNA may be less diverse than that of peptides, a larger fraction of a random RNA pool may form stable secondary structures (through base pairing) compared with the fraction of similarly sized random peptides that can form well-folded motifs [12]. This ability may give random RNAs greater structural variation than is available to random peptides inserted into an exposed loop constrained by a stable protein scaffold. In addition, basic structure-function relationships within RNA aptamers can often be revealed using site-directed mutagenesis and covariance analysis coupled with secondary structure prediction, while analogous experiments on peptide sequences can be much more difficult. Finally, researchers have established general methods for rationally engineering RNA that enable its function to be modulated using antisense oligonucleotides or using ligand binding aptamers. These efforts have successfully generated sequence-regulated or allosteric functional RNAs [13], while analogous efforts to engineer conditionally active peptide aptamers have not been reported.

The complexity of eukaryotic transcriptional activation makes this process an ideal candidate for validating our approach to perturbing cellular function with evolved RNAs. We report here the evolution of RNA-based activation domains and their characterization using site-directed mutagenesis and secondary structure prediction. The most potent evolved RNAs activate transcription to a degree comparable to that of the strongest known natural protein activation domains. Our findings demonstrate the use of RNA evolution in vivo to perturb complex biological pathways and provide a basis for engineering sequence-specific or ligand-modulated RNA-based transcription factors.

Results

Transcriptional Activation Selection

Eukaryotic transcription factors typically consist of two modular protein domains: a DNA binding domain and an activation (or repression) domain. The recruitment model of Ptashne and Gann [14] suggests that the primary function of the activation domain is to make specific interactions with the RNA polymerase II holoenzyme that localize the proteins responsible for transcriptional initiation to a given promoter. This model does not require that recruitment occur through protein-protein interactions, and we hypothesized that RNA-based transcriptional activators could be evolved in vivo if an RNA library was localized to the promoter of a selectable

*Correspondence: drliu@fas.harvard.edu
[2]These authors contributed equally to this work.

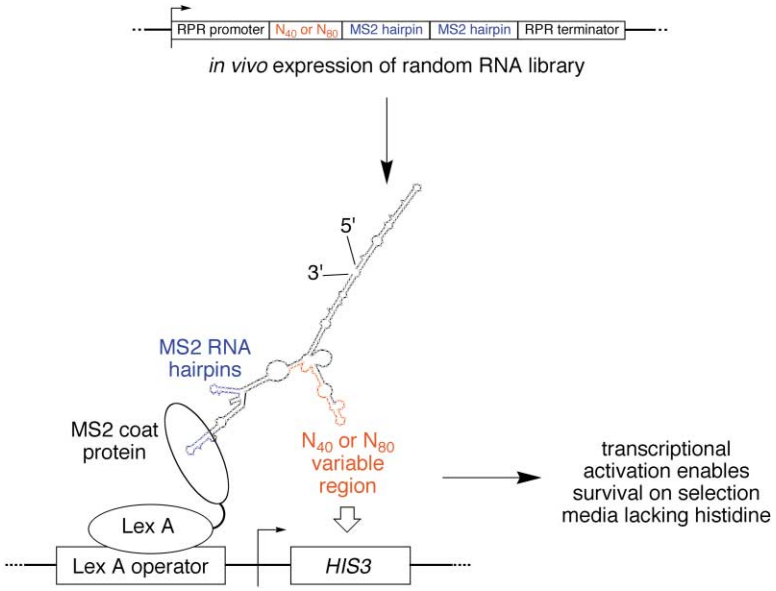

Figure 1. Library and Selection Design

RNA libraries containing a 5′ leader sequence, a random N_{40} or N_{80} region, two MS2 hairpins, and a terminator were expressed in yeast and localized to the promoter region of a *HIS3* gene by binding to a MS2 RNA binding protein fused to the DNA binding protein LexA [15]. RNA library members that activate transcription of *HIS3* allow survival on selection media lacking histidine.

genetic marker. This approach requires three components: a yeast strain containing a selectable reporter gene, a method for tethering RNA to this reporter gene, and a vector that expresses a random library of stable RNAs. We used the yeast three-hybrid strain YBZ-1 developed by Wickens and coworkers [15] to provide the first two components. The expression of both a *HIS3* gene and a lacZ gene in YBZ-1 is driven by promoters that contain upstream LexA binding sites. The strain also expresses a LexA-MS2 fusion protein that binds both to these operator sites and also to the 19 bp MS2 RNA hairpin with extremely high affinity ($K_d = 2 \times 10^{-10}$ M [15]). RNAs containing an MS2 hairpin are therefore localized to the promoter of the *HIS3* and lacZ genes (Figure 1). Because expression of *HIS3* or LacZ can be selected or quantitated, respectively, Wickens, Fields, and coworkers used this system to discover cellular RNA targets of an RNA binding protein that was fused to a known protein transcriptional activator [16]. Encouragingly, sequences not requiring an RNA binding protein were also noted [16], suggesting that certain genome-encoded RNAs might be able to activate transcription without a protein transcriptional activator.

Expression and Selection of RNA Libraries

Stable expression of a random RNA library in vivo is a major challenge because unstructured RNAs can be rapidly degraded in the cell. To maximize the stability of our RNA libraries, we designed the variable region to lie within a larger RNA having known stable secondary structures at its 5′ and 3′ termini. In the pIII-MS2 vector constructed by Wickens and coworkers [17], RNA library members are transcribed by RNA polymerase III from the RNase P RNA (RPR) promoter [18] and are not modified or translated [15]. We inserted a random 40 base region (N_{40}) or 80 base region (N_{80}) into the transcribed region followed by two MS2 hairpins. The transcript ends with the RPR terminator to enhance stability of the 3′ end of the RNA library (Figure 1).

Prepared pIII-MS2 backbone DNA was ligated with a synthetic DNA cassette encoding the N_{40} or N_{80} library, amplified in *E. coli* (initial diversity of 1.1×10^7 *E. coli* transformants for each library), and transformed into YBZ-1 yielding 10^4–10^5 transformants. The yeast libraries were plated on media lacking histidine to select for *HIS3* transcriptional activation and expression. Initial survivors were each screened by plating on fresh media lacking histidine. Red colonies (which presumably lost the pIII-MS2 plasmid containing *ADE2*) and clones that failed to grow again were discarded. For the N_{40} library, clones passing the initial selection and screening were observed at a surprisingly high frequency of 0.2%. In contrast, the N_{80} library yielded a lower frequency of positives (0.01%). These results suggest that a significant fraction of our random RNA libraries are able to activate transcription when localized to a promoter.

Characterization of Initial Selected RNAs

We characterized 70 total survivors from both libraries by retransformation into fresh YBZ-1 cells and quantitation of β-galactosidase expression levels from cell extracts. As a positive control, we used the known three-hybrid interaction between MS2 hairpin-IRE RNA and the IRP-Gal4 fusion protein, which leads to recruitment of the strongly activating Gal4 domain to the LexA operator and activation of the reporter genes [19]. Results from β-galactosidase expression assays are shown in Figure 2 for the 11 strongest selected members of the N_{40} library. All eleven activate LacZ expression at least as strongly as the Gal4 positive control. One clone (N40-26) activated LacZ expression more than ten times as strongly as the Gal4 positive control (Figure 2). Because the library was selected on the basis of *HIS3* transcriptional activation yet characterized by activation of LacZ expression, these results indicate that survivors express general, rather than gene-specific, RNA-based transcriptional activators. From the N_{80} library, three clones of the 46 assayed demonstrated LacZ expression levels comparable to that of the positive control and were not further characterized.

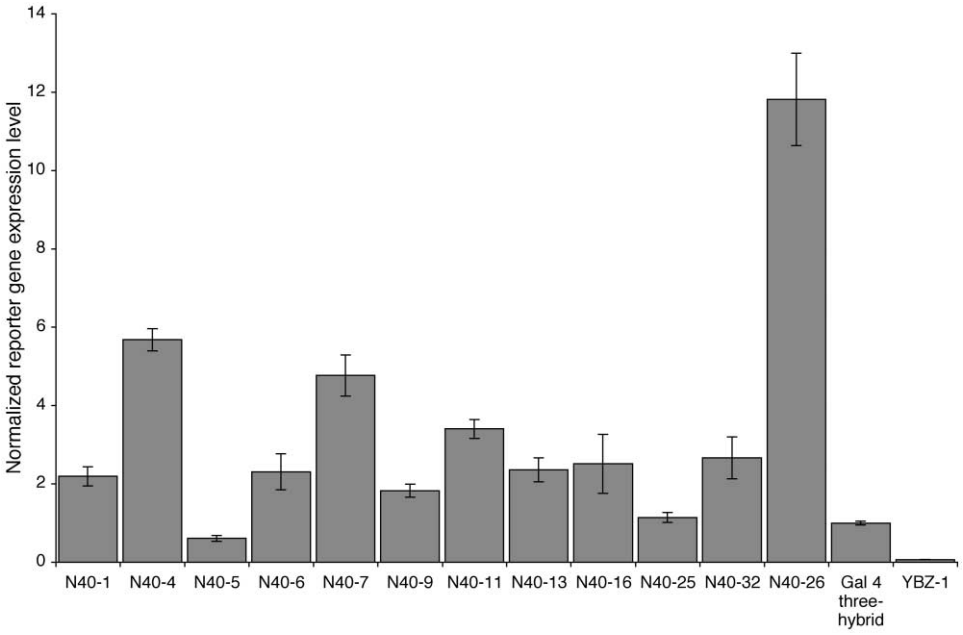

Figure 2. Transcriptional Activation Abilities of Original Selected RNAs

Quantitative β-galactosidase assays [32] of cell lysates from the 11 most potent N_{40} activators were performed at least three times each from independently grown clones. The average activity per clone is shown normalized relative to the Gal4 three-hybrid positive control (=1.0). Error bars reflect standard deviations.

The secondary structures of the 11 most active clones from the N_{40} library were predicted using the mfold method [20]. While the MS2 hairpins and 5′ leader sequence in the constant region maintained the same predicted secondary structure among the active clones, both the variable regions as well as the constant regions near their junction were predicted to adopt a wide variety of different structures (data not shown). These results suggest that transcriptional activator RNAs may operate through a variety of different mechanisms and possibly a variety of different targets, or that several different RNA structures can form interactions with the same target leading to transcriptional activation.

Evolution of More Potent Activators

To determine if these initial clones could be further evolved toward stronger transcriptional activation, we increased the stringency of the selection for *HIS3* expression by adding 3-aminotriazole (3-AT), a competitive inhibitor of His3 activity, to the growth media [17]. Freshly transformed YBZ-1 yeast expressing N40-26 can grow on selection media containing 1 mM 3-AT, while freshly transformed yeast that express RNAs with activities below that of the positive control fail to grow in the presence of 1 mM 3-AT. These results indicate that 3-AT can be used to increase the dynamic range of the selection and therefore can enable more potent transcriptional activator RNAs to be distinguished from less active sequences.

We generated a library of variants of our strongest initially selected RNA activator, N40-26, in which each of the 40 bases in the variable region was randomly mutated at a frequency of 20%. DNA sequencing of 14 library members before selection revealed an average of 9.0 mutations per clone, close to the anticipated value. High stringency selection of this library (1.4×10^5 yeast transformants from an original diversity of 5.5×10^7 *E. coli* transformants) in the presence of 1 mM 3-AT yielded 40 survivors containing 32 unique sequences. Each of the 32 evolved clones was characterized by retransformation and β-galactosidase assay (Figure 3). Fifteen of the clones possessed a transcriptional activation activity higher than that of the starting clone N40-26. Only one clone (m26-12) was much less active than the parental N40-26 RNA. The most active evolved clone, m26-29, activates transcription of the reporter gene more than 5-fold stronger than N40-26 and 53-fold stronger than the Gal4 activation domain positive control.

To compare the m26-29 RNA with one of the most potent [21] and well-characterized natural protein transcriptional activators known, we expressed a LexA-VP16 fusion protein from the ADH promoter on a single copy vector in L40-ura3 to mimic the expression levels of the LexA-MS2 fusion protein. Remarkably, the m26-29 RNA activated gene expression only 2-fold lower than VP16 fused directly to LexA (Figure 3), even though the RNA-based activator requires an additional interaction (between the MS2 hairpin and the MS2 binding protein domain) that may decrease the efficiency of transcriptional activation. Taken together, these results indicate that the mutagenesis and high stringency selection strategy applied to N40-26 resulted in the evolution of significantly improved RNAs that rival the effectiveness of the most potent known transcriptional activator proteins.

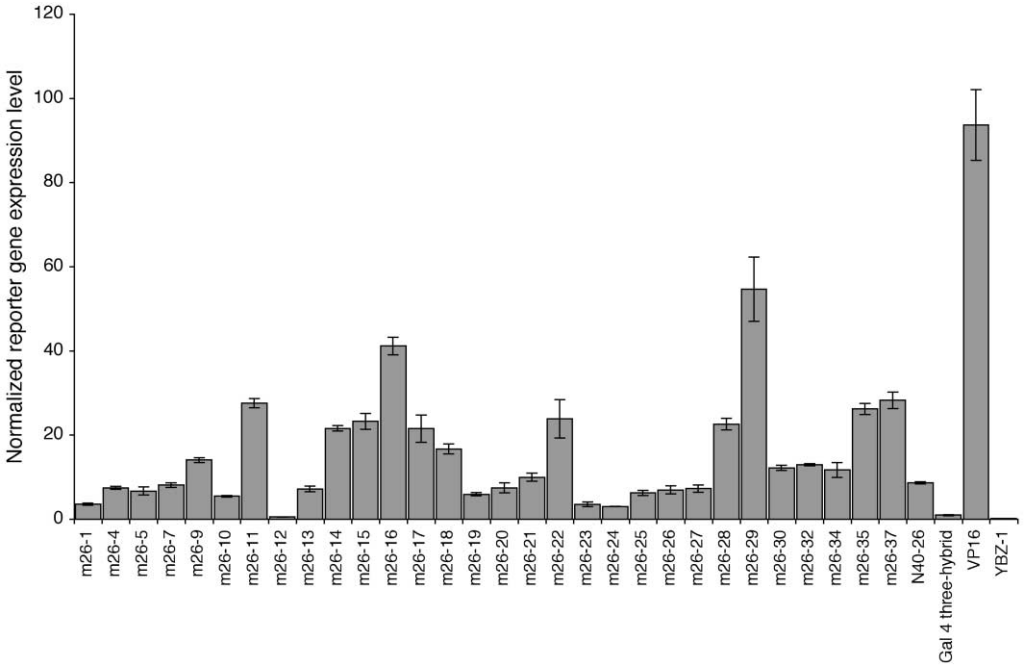

Figure 3. Transcriptional Activation Abilities of Evolved N40-26 Variants

Quantitative β-galactosidase assays [32] of cell lysates from 30 evolved N40-26 mutants were performed at least three times each from independently grown clones. The average activity per clone is shown normalized relative to the Gal4 three-hybrid positive control (=1.0). Error bars reflect standard deviations.

Characterization of Evolved Activators

To test whether the evolved RNAs require the MS2 protein-mediated localization to the LexA promoter, we introduced the plasmids expressing two representative active clones (m26-11 and m26-15) into the yeast strain L40-ura3 which lacks the LexA-MS2 fusion protein but is otherwise identical to YBZ-1 [15]. As expected, the resulting cells were unable to survive on media lacking histidine, indicating that localization of the evolved RNAs to the reporter gene is required for transcription activation.

An alignment of the sequences of 31 evolved N40-26 variants is shown in Figure 4. All evolved N40-26 variants were closely related, with the sole exception of the much less active m26-12 clone (data not shown). Surprisingly, the consensus sequence is the same as the N40-26, suggesting that N40-26 is already somewhat optimized in its ability to activate transcription despite the significant improvements in activity upon mutagenesis and reselection. The 31 active sequences contained an average of 4.5 mutations each, indicating that only about 50% of the introduced mutations allowed RNAs to survive the higher stringency selection. These mutations were clustered at positions 4–15, 19–22, 34, and 39–40 within the 40 base variable region (Figure 4).

Three subsequences (bases 16–18, 23–33, and 35–38) are highly conserved among the evolved N40-26 variants (Figure 4). Interestingly, these conserved subsequences correspond to three of the four regions of predicted secondary structure [20] (Figure 5). Bases 17–18 (CC) are predicted to participate in pairing with the G-rich end of the 5′ constant region; bases 24–30 may be involved in base pairing with the 5′ constant region; and bases 35–38 may pair with four bases in the 3′ terminator.

```
                  1    1    2    2    3    3    4
            1  5  0    5    0    5    0    5    0
N40-26      CGCGGAAGAATGCT-CCCCCAAGTGGATGCCTAAGCCTCTT
m26-10      CGCGGAAGAATGAT-CCCCCAAGTGGATGCCTAAGCCTCTT
m26-36      CGCGGAAGAATTCT-CCCCCAAGTGGATGCCTAAGCCTCTT
m26-31      CGCGGAAGAATGCT-CCCCCGAGTGGATGCCTAAGCCTCAT
m26-9       CGCGGAGGAATGCT-CCCCCGAGTGGATGCCTAATCCTCTT
m26-14      CGCGGAAGCATACT-CCCCCGAGTGGATGCCTAAGCCTCTT
m26-15      CGCGGACCAATGGT-CCCCCAAGTGGATGCCTAAGCCTCTT
m26-34      CGCGGAAGATTGCC-CCCCCAAGTGGATGCCTAAACCTCTA
m26-29      CGCGGAAGATTGTT-CCCCCAAGTGGATGCCTAAACCTCAT
m26-28      CGCGGAAAAATACT-CCCCCAAGTGGATGCCTAAACCTATT
m26-17      CGCGGAAGAATGAT-CCCCCAGGTGGATGCCTAAGCCTCTA
m26-30      CGCGGAACAATGCT-CCCCCAGCTGGATGCCTAAGCCTCTA
m26-20      CGGGGAAGAATGCT-GCCCAAAGTGGATGCCTAAGCCTCTT
m26-4       CGCGGAAGAATGCT-GCCCAACCTGGATGCCTAAACCTCTT
m26-32      CGCGGAAGAATGCT-GCCCAACCTGGATGCCTAAACCTCTT
m26-13      CTCGGAACAATGCC-CCCCCAAGTGGATGCCTAAACCTCAT
m26-35      CTCGGAAGAATTCT-CCCGATAGTGGATGCCTAAGCCTCTT
m26-16      CGCGGAAGAAGAGT-CCCCCGAGTGGATGCCTAAACCTCTT
m26-18      CGCGGAACAATGTT-GCCCCAAATGGATGCCTAAGCCTCTT
m26-25      CGCGGAAGAACACT-CCCCCACGTGGATGCCTAAGCCTCCTT
m26-21      CGCGGAAGACTGCA-GCCCCAGGTGGATGCCTAAACCTCGGT
m26-39      CCCGGAAGACTGCA-GCCCCAGGTGGATGCCTAAACCTCTT
m26-33      CGCGGAAGACTTCA-GCCCCAAGTGGATGCCTAAGCCTCTTA
m26-1       CGCGGGTGTAAGCC-CCCCCAGGTGGATGCCTAAGCCTCTT
m26-23      CGCGTAAACATGCG-GCCCCAAGTGGATGCCTAAGCCTCCT
m26-24      CGCGGAAAGATGAT-CACCCGAGTGGATGCCTAAGCCTCTT
m26-11      CGCGCGAGTATACT-CCCCCAAGCGGATGCCTAAGCCTCTT
m26-19      CGCAGAAGATACCT-CCCCCAAGTGGATGCCTAAGCCTCAT
m26-26      AACAGAAGAATACT-CCCCCAAGTGGATACCTAAGCATCTT
m26-37      CGCGGAAAAATTCTTCCCCCAAGTGGATGCCTAAGCCTCTT
m26-38      CGCGGAAAAATTCTTCCCCCAAGTGGATGCGTAAGCCTCTT
m26-7       CGCGGAAGAACGCT-CCCCGACGTGGATGCCTATTGTCCTT
consensus   CGCGGAAGAATGCT-CCCCCAAGTGGATGCCTAAGCCTCTT
```

Figure 4. Alignment of Variable Region Sequences from Evolved N40-26 Variants

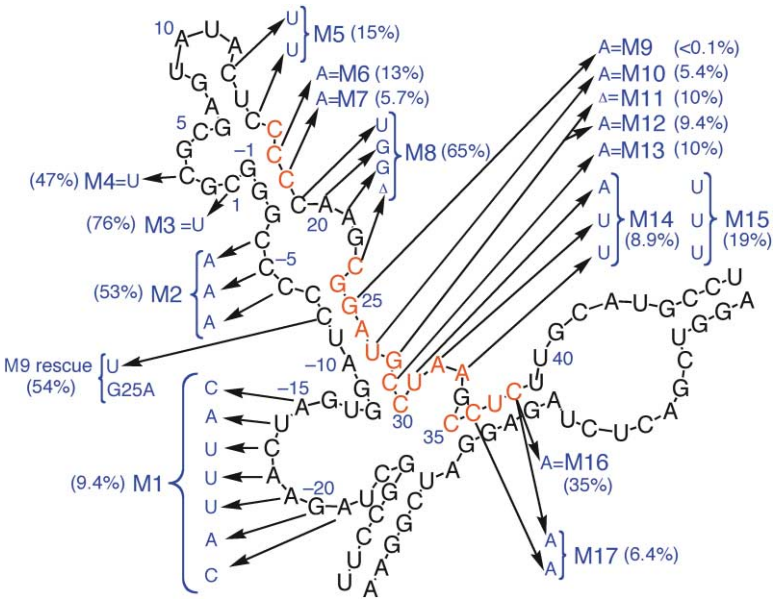

Figure 5. Predicted Secondary Structure and Strategy for Site-Directed Mutagenesis of m26-11

Each set of single- or multiple-base mutations is labeled M1 through M17 to correspond with the data listed in Table 1. Highly conserved bases among N26-40 variants are shown in red. Activities of each mutation set are listed in parentheses as a percentage relative to the activity of unmutated m26-11.

It is noteworthy that among the evolved sequences, m26-12 uniquely lacks several of the conserved secondary structures and is also by far the weakest activator among the evolved N40-26 variants. In addition, the lack of predicted base pairing between conserved variable region bases (as opposed to between the variable and constant regions) among all of the evolved N40-26 variants is surprising and may suggest geometric constraints imposed by the structure of the constant region that disfavor base pairing within the variable region. In summary, the conserved subsequences emerging from random mutagenesis and reselection together with their predicted secondary structures collectively suggest several candidate structural elements that could play roles in transcriptional activation.

Structure-Activity Analysis of an Evolved RNA

An attractive feature of RNA aptamers is the possibility of using secondary structure prediction together with site-directed mutagenesis to infer and test structure-function relationships. We systematically installed a series of 16 single or multibase site-directed mutations (Table 1 and Figure 5) in the variable and constant regions of one of the most active evolved N40-26 variants, m26-11, and measured the ability of the resulting mutants to activate β-galactosidase transcription.

We first perturbed nucleotides predicted to participate in base pairing within the three highly conserved regions described above. Variable region bases C17, C18, and C19 in m26-11 are predicted to pair with the GGG at the end of the 5′ constant region (Figure 5). Mutation of C17 to A (M6) or mutation of C18 to A (M7) reduces transcriptional activation by 8-fold and 17-fold, respectively (Table 1). The highly conserved GGAUGCC representing bases 24–30 is also predicted to pair with the 5′ constant sequence. The G25A mutant (M9) possesses no measurable transcriptional activation activity (<0.1%). Similarly, a variety of mutations among the other bases predicted to form secondary structures in this region (M10, M11, M12, and M13), reduce activity by 10- to 20-fold (Table 1). The role of base pairing involving bases 35–39 was probed by the single mutation of C38 to A (M16), which resulted in a 3-fold reduction in transcriptional activation, as well as by a larger perturbation changing C36 and C38 to A36 and A38 (M17), which caused a 16-fold loss in activity. These results further highlight the importance of these three regions in transcriptional activation.

To test aspects of the predicted structural model within the largest conserved region (bases 23–33), we generated a secondary mutation designed to restore the activity of the least active m26-11 mutant (M9, G25A). The structural model in Figure 5 predicts that base 25 of the variable region pairs with base −8 of the constant region. Replacing C(−8) with U, predicted to restore base pairing with the inactive G25A mutant, rescues transcriptional activation ability to 54% of the unmutated m26-11 (Table 1). The ability of a single compensating mutation at base −8 to restore the activity of an inactive point mutant provides strong support for the predicted secondary structure in this region. In addition, this result demonstrates that base pairing, but not base-specific contacts, at positions −8 and 25 are required for transcriptional activation.

Site-directed mutations outside of the three conserved regions predicted to participate in base pairing resulted in smaller losses in activity. Mutations M2, M3, M4, and M5 perturb constant and variable region bases upstream of the first conserved region and resulted in 1.3- to 7-fold decreases in activity. Indeed, bases 1–16 can even be replaced with an unrelated 26 base sequence without significant loss of transcriptional activation (data not shown). Mutation of the nonconserved bases 19–23 (M8) likewise resulted in less than 2-fold loss of activity. These findings suggest that mutations are more tolerated in regions predicted not to participate in base pairing. In support of this relationship between predicted base pairing and functional importance, mutating bases 31–33 (predicted to form an unpaired bulge

Table 1. Transcriptional Activation Abilities of Site-Directed Mutants of m26-11 Shown in Figure 5

Mutation	Genotype	% Activity Relative to m26-11
M1	A(−21)C G(−20)A A(−19)U A(−18)U C(−17)U U(−16)A A(−15)C	9.4% ± 1.3%
M2	C(−6)A C(−5)A C(−4)A	53% ± 2.7%
M3	C1U	76% ± 7.6%
M4	C3U	47% ± 4.7%
M5	C13U C15U	15% ± 2.4%
M6	C17A	13% ± 2.0%
M7	C18A	5.7% ± 1.2%
M8	C19U A20G A21G ΔC23	65% ± 11%
M9	G25A	<0.1%
M9 rescue	C(−8)U G25A	54% ± 5.0%
M10	U27A	5.4% ± 0.15%
M11	ΔG28	10% ± 1.4%
M12	G28A	9.4% ± 1.4%
M13	C29A	10% ± 2.0%
M14	U31A A32U A33U	8.9% ± 0.63%
M15	A32U A33U	19% ± 1.1%
M16	C38A	35% ± 2.5%
M17	C36A C38A	6.4% ± 0.5%

Quantitative β-galactosidase assays [32] of cell lysates were performed three to nine times each from independently grown clones, and average values are reported as the percentage of transcriptional activation relative to m26-11. Standard deviations are shown following each value.

between the two conserved putative stems) impaired activity by as little as 5-fold (M14 and M15), despite the highly conserved nature of these three nucleotides.

Discussion

We have described the in vivo selection of RNA sequences capable of activating transcription with potency comparable to the most active known protein transcriptional activation domains. Through a combination of further evolution, systematic site-directed mutagenesis, and secondary structure prediction, we elucidated structure-function relationships that identify regions of the evolved RNAs that play important functional roles. The potency of our evolved activators—up to 53-fold higher than a Gal4 three-hybrid positive control—is surprising given that the most active previously reported genomic RNA sequences with transcriptional activation properties [16] are 5-fold less potent than the same Gal4 three-hybrid positive control [19]. Indeed, independent work by Ptashne and coworkers [22] used a similar selection (without additional rounds of mutagenesis and reselection) on a smaller, 10 base random region to isolate transcriptional activating RNAs that are 10-fold less potent than intact Gal4 and have no sequence homology to the RNAs described here. The significantly higher potency of the 40 base variable region RNAs evolved in this work suggests that the secondary structural diversity available to longer random RNAs may be required to activate transcription with high potency. Collectively these findings demonstrate that RNA is capable of folding into stable structures that present a compatible surface for recruiting the transcriptional machinery.

While we believe recruitment to be the most likely mechanism of action of these RNAs, we cannot rigorously exclude the possibility of a more complex activation mechanism such as one in which the RNA acts as a decoy for transcriptional inhibitors. However, the requirement of MS2 protein-mediated localization for activity, together with preliminary results indicating that deletion of specific recruitable components of the transcriptional machinery significantly decreases the activity of our RNAs (P.D.K., A.R.B., and D.R.L., unpublished data), further supports simple recruitment as the mechanism of activation.

We found a surprisingly large fraction (~0.2%) of our initial random N_{40} library was able to activate transcription. Our work parallels previous studies by Ptashne and coworkers that report 0.1% to 1% of short random peptides fused to a Gal4 DNA binding domain are capable of activating transcription [23, 24], although the most active peptide fusion was reported to activate transcription 1.6-fold as potently as intact Gal4. Given the significant differences between the physical properties of RNA and proteins, our results collectively imply that there are many different but comparably effective solutions for recruitment of the eukaryotic transcription initiation complex. This likely reflects both many possible targets as well as multiple sites per target for productive binding that leads to transcriptional activation. The fact that nonnatural RNA-protein interactions can activate transcription lends further support to the recruitment model [14] by demonstrating that simple binding mechanisms distinct from those used in nature may be sufficient for mediating an important and ubiquitous biological function. RNA's lack of positive charges, ability to make hydrophobic interactions, and abundant negative charges—features found in protein transcriptional activators [25, 26]—apparently provide RNA with an effective chemical repertoire to interact with the transcriptional machinery.

Although the N_{40} library yielded a high frequency of transcriptional activators, the N_{80} library yielded significantly fewer. We initially hypothesized that a larger random region might offer a greater frequency of positives

because of its much higher frequency of containing a specific required secondary structure [27]; this reasoning may hold true when comparing 40-base-variable regions to the 10-base-variable regions described by Ptashne and coworkers [22] that yielded a much lower frequency ($\sim$1 in 10^6) of positives. Based on the 20-fold lower frequency and lower average activities of transcriptional activators in the N_{80} library compared with the N_{40} library, we additionally speculate that the smaller N_{40} library balanced secondary structures required for high activity with minimizing the presence of unstructured single-stranded regions prone to intracellular degradation, and that at longer lengths, RNA instability can become limiting.

Our studies identify three regions within the most active evolved RNAs as particularly crucial for the observed activity. Gratifyingly, the sequence conservation within these regions, their predicted secondary structures, and the results of site-directed mutagenesis experiments are all consistent with a model in which these three subsequences (bases 17–19, 24–30, and 35–39) play key roles in transcriptional activation, possibly by forming essential base paired structures. Surprisingly, these findings suggest that extensive base pairing between the variable and constant regions is required for activity. The flanking constant regions, when paired with the variable sequences, may therefore provide a sufficiently large and well-ordered scaffold to enable effective interactions with the as yet unidentified target.

The approach to perturbing a biological function of interest (in this case, transcriptional activation) using RNA evolution in vivo requires an efficient selection or high throughput screen but is attractive because it does not require knowledge of any targets involved in the biological process of interest. In addition, while the more common RNA evolution approach of in vitro selection using previously identified and purified biological targets may not yield optimal desired activities when expressed in vivo, the approach described here evolves RNAs on the basis of their activities in natural cellular contexts. The well-characterized nature of several of the RNAs evolved in this study provide a promising start for efforts to identify the cellular target mediating RNA-based transcriptional activation using genetic or affinity-based methods. In addition, the identification of crucial bases within the evolved RNAs may enable the engineering of regulated RNA-based transcriptional activators that require the presence or absence of specific ligands. For example, it may be possible to evolve an RNA linker region that transduces a small molecule binding event [13] into a conformational rearrangement in the critical stem region in order to either activate or repress transcription. In theory, this approach may also be used to study selectable or screenable functions unrelated to transcriptional activation.

Significance

We describe an approach to studying biological function using random RNA libraries coupled with in vivo selections. Using this approach, we have evolved RNA transcriptional activators with potencies comparable to the most active natural protein-based activation domains such as VP16. The high frequency of finding active RNAs in our selection for transcriptional activators suggests that features of protein structure necessary for transcriptional activation can be mimicked effectively by nucleic acids. Additional rounds of diversification and selection, systematic site-directed mutagenesis, and secondary structure prediction together identified regions of the evolved RNA sequences that likely play important roles in transcriptional activation. Evolution of random RNA libraries in vivo may be a powerful tool for dissecting complex biological function.

Experimental Procedures

Yeast Strains and Media

Media consisted of yeast nitrogen base (Sigma, St. Louis, MO), 4% dextrose, and synthetic drop out supplements lacking histidine or histidine and uracil (Clontech, Palo Alto, CA). Yeast were cultured either in liquid medium or on agar plates at 30°C. *S. cerevisiae* strains YBZ-1 (*MAT*a, *ura3-52*, *leu2-3*, *112*, *his3-200*, *trp1-1*, *ade2*, *LYS2::(LexA op)-HIS3*, *ura3::(LexA op)-LacZ*, *LexA-MS2-MS2 coat (N55K)*) and L40-ura3 (*MAT*a, *ura3-52*, *leu2-3*, *112*, *his3-200*, *trp1-1*, *ade2*, *LYS2::(LexA op)-HIS3*, *ura3::(LexA op)-LacZ*) were a gift from Professor Marvin Wickens [15].

Construction of Plasmids and RNA Libraries

Plasmids encoding the RNA libraries were based on the yeast shuttle vector pIIIa-MS2 [15] (a gift from Professor Marvin Wickens). Library-encoding sequences were cloned directly into the plasmid using the unique SphI and XmaI sites. The plasmid carries a *URA3* marker as well as the *ADE2* gene that can be used to screen for false positives in the selection. Plasmid pIIIa/IRE-MS2 expresses a fusion of the *ir*on *r*esponse element (IRE) and the MS2 hairpin (5′IRE-MS2-3′) from the RPR promoter, and plasmid pAD-IRP expresses a fusion of the *ir*on *r*egulatory *p*rotein (IRP) and the Gal4 activation domain driven from the ADH promoter. Random single stranded N_{40} or N_{80} libraries were generated on an Applied Biosystems Expedite 8909 DNA Synthesizer or purchased from Sigma-Genosys (The Woodlands, TX), respectively. Blunt-ended double-stranded library inserts were synthesized by primer extension using the Klenow fragment of *E. coli* DNA Pol I from a constant primer binding site in the synthetic library oligonucleotides, digested with SphI and XmaI, and ligated into precut pIIIa/MS2 backbone to provide pIIIa/MS2-N40 and pIIIa/MS2-N80. Library-encoding plasmids were amplified by transformation into electrocompetent DH10B *E. coli* (Invitrogen, Carlsbad, CA) and isolated by plasmid purification. Constrained by the modest transformation efficiencies of yeast and our large variable region (40 bases), our libraries only cover a tiny fraction of possible sequence space even though the DNA encoding the library should contain >99% of sequences with >20% similarity to the N40-26 parent based on the analysis[28] of Knight and Yarus.

LexA-VP16 was expressed from the ADH promoter on p416ADH-LV, a single copy yeast shuttle vector, to mimic the expression of LexA-MS2 in YBZ-1. LexA (1-202) was amplified from the LexA-Cyc8 plasmid, a gift from Kevin Struhl [29], using the primers GGGGGGGGATCCCAGCCAGTCGCCGTTGCGAAT and GGGGGGG CTAGCATGAAAGCGTTAACGGCCAGG and digested with BamHI and NheI. VP16 (residues 413–489) was amplified from the C7-VP16 plasmid, a gift of Roger Beerli [30], with the primers CCGCCGGGAT CCGCTCCCCCGACCGATGTCAGC and CCGCCGCTCGAGTTAAC CGTACTCGTCAATTCCAAG (designated VC), and digested with BamHI and XhoI. These digested fragments were ligated into NheI- and XhoI-digested pET23a vector (Novagen, Madison, WI). The LexA-VP16 region was amplified from the resulting plasmid using the primer CCGCGGACTAGTATGAAAGCGTTAACGGCCAGGC and the primer VC above and subcloned into p416ADH ([31], purchased from the ATCC [Manassas, VA]) using SpeI and XhoI sites. All constructs were verified by DNA sequencing. Molecular biology enzymes were purchased from New England Biolabs (Beverly, MA).

Selection and Assay Protocol

For the selection experiments, the RNA expression plasmid was transformed into YBZ-1 using a standard lithium acetate procedure. Transformants were selected on media lacking histidine. Plasmid DNA was extracted via glass bead lysis and phenol extraction, ethanol precipitated, and then amplified in *E. coli*. Selection survivors were initially screened by restreaking on media lacking histidine and uracil prior to assaying. Selection at higher stringency was performed in an identical manner, with the addition of 1 mM 3-aminotriazole to the media.

Retransformed clones were assayed for β-galactosidase activity using a liquid *o*-nitrophenyl-β-galactopyranoside (ONPG) assay [32]. Activity was calculated as Miller units and normalized to the Gal4-based positive control as explained in the figures. Assay values represent the average of at least three independent cultures of each clone.

Secondary Structure Prediction

Secondary structures of selected RNA sequences were individually predicted with the mfold program [20] using the most recent optimized parameters (predicted for 37°C; these parameters are currently not available for 30°C). Although the most thermodynamically stable structures were used in our analysis, all structures within 5% of the minimal energy were considered. For m26-11, the most stable predicted structure (shown in Figure 5) is 5.6 kcal/mol more stable than the next lowest energy structure.

Acknowledgments

The authors are grateful to David Bernstein and Prof. Marvin Wickens for the three-hybrid system strain and plasmids as well as their valuable advice. The LexA-Cyc8 plasmid and C7-VP16 plasmids were gifts from Professor Kevin Struhl and Professor Roger Beerli, respectively. We thank Keith Fandrick for assistance in generating the libraries, and Michael Sacerdote for helpful discussions. This research was supported by the American Cancer Society (#RSG-02-066-01-MGO), the National Science Foundation (#MCB-0094128), and the National Institutes of Health (#1R01GM65400-01). P.D.K. is a Howard Hughes Medical Institute Predoctoral Fellow, and A.R.B. is supported in part by the National Institutes of Health Molecular, Cellular, and Chemical Biology Training Grant #5 T32 GM07598-25.

Received: March 6, 2003
Revised: April 28, 2003
Accepted: April 30, 2003
Published: June 20, 2003

References

1. Doudna, J.A., and Cech, T.R. (2002). The chemical repertoire of natural ribozymes. Nature *418*, 222–228.
2. Moore, P.B., and Steitz, T.A. (2002). The involvement of RNA in ribosome function. Nature *418*, 229–235.
3. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858–862.
4. Lanz, R.B., McKenna, N.J., Onate, S.A., Albrecht, U., Wong, J., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell *97*, 17–27.
5. Lanz, R.B., Razani, B., Goldberg, A.D., and O'Malley, B.W. (2002). Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). Proc. Natl. Acad. Sci. USA *99*, 16081–16086.
6. Norman, T.C., Smith, D.L., Sorger, P.K., Drees, B.L., O'Rourke, S.M., Hughes, T.R., Roberts, C.J., Friend, S.H., Fields, S., and Murray, A.W. (1999). Genetic selection of peptide inhibitors of biological pathways. Science *285*, 591–595.
7. Geyer, C.R., Colman-Lerner, A., and Brent, R. (1999). "Mutagenesis" by peptide aptamers identifies genetic network members and pathway connections. Proc. Natl. Acad. Sci. USA *96*, 8567–8572.
8. Blum, J.H., Dove, S.L., Hochschild, A., and Mekalanos, J.J. (2000). Isolation of peptide aptamers that inhibit intracellular processes. Proc. Natl. Acad. Sci. USA *97*, 2241–2246.
9. Ferber, M.J., and Maher, L.J., 3rd. (1998). Combinatorial selection of a small RNA that induces amplification of IncFII plasmids in Escherichia coli. J. Mol. Biol. *279*, 565–576.
10. Soukup, G.A., and Maher, J.J., 3rd. (1998). Selection and characterization of RNAs that relieve transcriptional interference in Escherichia coli. Nucleic Acids Res. *26*, 2715–2722.
11. Zimmerman, J.M., and Maher, L.J., 3rd. (2002). In vivo selection of spectinomycin-binding RNAs. Nucleic Acids Res. *30*, 5425–5435.
12. Keefe, A.D., and Szostak, J.W. (2001). Functional proteins from a random-sequence library. Nature *410*, 715–718.
13. Soukup, G.A., and Breaker, R.R. (2000). Allosteric nucleic acid catalysts. Curr. Opin. Struct. Biol. *10*, 318–325.
14. Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.
15. Bernstein, D.S., Buter, N., Stumpf, C., and Wickens, M. (2002). Analyzing mRNA-protein complexes using a yeast three-hybrid system. Methods *26*, 123–141.
16. Sengupta, D.J., Wickens, M., and Fields, S. (1999). Identification of RNAs that bind to a specific protein using the yeast three-hybrid system. RNA *5*, 596–601.
17. Zhang, B., Kraemer, B., SenGupta, D., Fields, S., and Wickens, M. (2000). Yeast three-hybrid system to detect and analyze RNA-protein interactions. Methods Enzymol. *318*, 399–419.
18. Good, P.D., and Engelke, D.R. (1994). Yeast expression vectors using RNA polymerase III promoters. Gene *151*, 209–214.
19. SenGupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S., and Wickens, M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. Proc. Natl. Acad. Sci. USA *93*, 8496–8501.
20. Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. *288*, 911–940.
21. Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4–VP16 is an unusually potent transcriptional activator. Nature *335*, 563–564.
22. Saha, S., Ansari, A., Jarell, K., and Ptashne, M. (2003). RNA sequences that work as transcriptional activating regions. Nucleic Acids Res. *31*, 1565–1570.
23. Lu, X., Ansari, A.Z., and Ptashne, M. (2000). An artificial transcriptional activating region with unusual properties. Proc. Natl. Acad. Sci. USA *97*, 1988–1992.
24. Ma, J., and Ptashne, M. (1987). A new class of yeast transcriptional activators. Cell *51*, 113–119.
25. Ponticelli, A.S., Pardee, T.S., and Struhl, K. (1995). The glutamine-rich activation domains of human Sp1 do not stimulate transcription in Saccharomyces cerevisiae. Mol. Cell. Biol. *15*, 983–988.
26. Struhl, K. (1995). Yeast transcriptional regulatory mechanisms. Annu. Rev. Genet. *29*, 651–674.
27. Wedel, A.B. (1996). Fishing the best pool for novel ribozymes. Trends Biotechnol. *14*, 459–465.
28. Knight, R., and Yarus, M. (2003). Analyzing partially randomized nucleic acid pools: straight dope on doping. Nucleic Acids Res. *31*, e30.
29. Tzamarias, D., and Struhl, K. (1994). Functional dissection of the yeast Cyc8-Tup1 transcriptional co-repressor complex. Nature *369*, 758–761.
30. Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F., 3rd. (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. USA *95*, 14628–14633.
31. Mumberg, D., Muller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119–122.
32. Pryciak, P.M., and Hartwell, L.H. (1996). AKR1 encodes a candidate effector of the G beta gamma complex in the Saccharomyces cerevisiae pheromone response pathway and contributes to control of both cell shape and signal transduction. Mol. Cell. Biol. *16*, 2614–2626.