

## METHOD

# Identification of RNAs that bind to a specific protein using the yeast three-hybrid system

DHRUBA J. SENGUPTA,<sup>1</sup> MARVIN WICKENS,<sup>2</sup> and STANLEY FIELDS<sup>1,3</sup>

<sup>1</sup>Departments of Genetics and Medicine, University of Washington, Seattle, Washington 98195, USA

<sup>2</sup>Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, USA

<sup>3</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

## ABSTRACT

We have adapted the yeast three-hybrid system to identify RNA ligands for an RNA-binding protein. In this assay system, a protein–RNA interaction is detected by the reconstitution of a transcriptional activator using two hybrid proteins and a hybrid RNA. The RNA molecule is tethered to the promoter of a reporter gene by binding to a hybrid protein consisting of the bacteriophage MS2 coat protein fused to the DNA-binding protein LexA; the RNA-binding domain to be analyzed is fused to the transcriptional activation domain of the yeast Gal4 protein; and the bifunctional RNA consists of binding sites for the coat protein and for the other RNA-binding domain. We built an RNA library such that short fragments of genomic DNA from yeast were transcribed in yeast together with binding sites for the coat protein. We screened this hybrid RNA library for RNAs that bound to the yeast Snp1 protein, a homolog of the human U1-70K protein. The screen yielded as the strongest positive the fragment of U1 RNA that contains loop I, which is known to bind to Snp1 in U1 snRNP. We also identified four other RNA ligands that produced weaker three-hybrid signals, suggesting lower affinities for Snp1 as compared to U1 RNA. In addition, this search also yielded a set of RNA sequences that can activate transcription on their own when bound to a promoter through a protein interaction.

**Keywords:** hybrid protein; reporter gene assay; Snp1 protein; transcriptional activator; U1 RNA

## INTRODUCTION

RNA–protein interactions play crucial roles in many biological processes, including transcription, splicing, and translation. These interactions are necessary for assembly of the machinery involved or for biological regulation. Elucidation of the constituents of an RNA–protein complex requires that protein partners can be identified for an RNA sequence of interest, and conversely, that RNA partners can be identified for a protein of interest.

The yeast three-hybrid system (SenGupta et al., 1996) detects the interaction between an RNA and a protein. In this system, a transcription factor is assembled by the use of an RNA bridge that brings a fusion protein containing a DNA-binding domain together with a fusion protein containing an activation domain. The assembly of this ternary complex depends on each fusion protein containing an RNA-binding domain that binds to a site in the RNA molecule. As fixed components, the

DNA-binding domain hybrid is the LexA protein fused to the bacteriophage MS2 coat protein, and the hybrid RNA contains two copies of the coat protein binding site. The three-hybrid system has been used to identify proteins that bind to RNA sequences such as the 3' end of histone mRNA (Wang et al., 1996; Martin et al., 1997) and an element in the 3' untranslated region of the *Caenorhabditis elegans fem3* mRNA (Zhang et al., 1997), and to detect and analyze known RNA–protein interactions (e.g., Bacharach & Goff, 1998).

We have now adapted this system to identify RNA ligands of an RNA-binding protein. We generated an RNA expression library of genomic sequences from the yeast *Saccharomyces cerevisiae* fused to coat protein binding sites, and screened it for RNAs that can bind the yeast Snp1 protein, a homolog of the human U1-70K protein. This search identified an RNA fragment containing the loop I sequence of *SNR19* RNA (U1 RNA), whose binding to Snp1 has been previously demonstrated by in vitro assays (Kao & Siliciano, 1992). Additionally, the search identified other Snp1-binding RNAs that yielded a weak signal in this assay, as well as RNA sequences that alone can activate transcription when bound to the promoter of a reporter gene.

Reprint requests to: Stanley Fields, Department of Genetics, University of Washington, Box 357360, Seattle, Washington 98195, USA; e-mail: fields@u.washington.edu.

## RESULTS

### Optimization of hybrid RNA parameters

To modify the three-hybrid assay to allow searches for RNAs, we initially characterized features of the hybrid RNA that could guide construction of an optimal RNA library. Vectors for RNA expression contain the yeast *URA3* gene and produce fusions with the coat protein binding sites oriented either 5' (pIII/MS2-1) or 3' (pIII/MS2-2) of the other RNA element (SenGupta et al., 1996). Inserts appeared to be more stably maintained in *Escherichia coli* in the pIII/MS2-2 vector (data not shown), which was thus used for further experiments. To determine the most favorable size for genomic inserts into this vector, we took advantage of the plasmid pIII/IRE-MS2, which expresses the Iron Response Element (IRE) fused 5' to the coat protein binding sites (SenGupta et al., 1996). Transformation of this plasmid into the three-hybrid reporter strain L40coat along with a *LEU2*-containing plasmid encoding a Gal4 activation domain hybrid with the Iron Response Protein-1 (IRP1) yields a high level of  $\beta$ -galactosidase reporter gene activity. To assess whether a hybrid RNA containing a sequence of interest can function in the three-hybrid context, we inserted the sequence between the two defined binding sites, and assayed for  $\beta$ -galactosidase due to the IRE/IRP1 interaction. The effect of RNA insert length on the three-hybrid system was determined by ligating yeast genomic fragments of average size 400, 200, or 100 bp into the unique *Sma*I site in the pIII/IRE-MS2 plasmid and transforming the resulting plasmids into the IRP1-containing reporter strain. Of approximately 200–250 transformants per insert size class assayed for  $\beta$ -galactosidase, nearly all of those with 400-bp inserts showed no activity, 80% of those with 200-bp inserts were inactive, and 50% of those with 100-bp inserts were inactive. Sequence analysis of six independent clones with 200-bp inserts that were negative for  $\beta$ -galactosidase activity indicated that none was especially AT-rich or contained runs of more than four T nucleotides, indicating that polymerase III transcription termination was not likely causing loss of the three-hybrid signal. This experiment suggested that even modest lengths of additional RNA sequence might either destabilize the hybrid RNAs or disrupt the IRE/IRP1 interaction by alternative secondary structures, and that a small insert library might therefore be preferable.

We next reconstructed in the three-hybrid assay the RNA–protein combination that would serve as our test case for the library search: the interaction of the yeast Snp1 protein with the loop I sequence of U1 RNA. Snp1 belongs to the large and well-studied family of RNA Recognition Motif (RRM domain) proteins, of which several have been characterized in yeast proteins and others identified in yeast from the complete genome sequence. The binding of this protein to the 5' 45 nt of

U1 RNA (encoded by the *SNR19* gene) has been demonstrated by gel mobility shift assays (Kao & Siliciano, 1992), as has the binding of the homologous mammalian U1-70K protein to the same sequence in the human U1 RNA (Surowy et al., 1989). The entire Snp1 protein was used to construct an activation domain hybrid, and sequences encoding the 5' 45 nt or 5' 175 nt of U1 RNA were inserted into the RNA vector pIII/MS2-2. While both of these hybrid RNAs yielded a signal in the assay, the shorter U1 sequence led to approximately sixfold more  $\beta$ -galactosidase activity as compared to the longer sequence (50 U vs. 8 U). Taken together with the other experiments on the IRE-containing hybrid RNA, these results suggested that minimal binding sites without additional flanking RNA sequences might be optimal in this assay.

### Library search for Snp1-binding RNAs

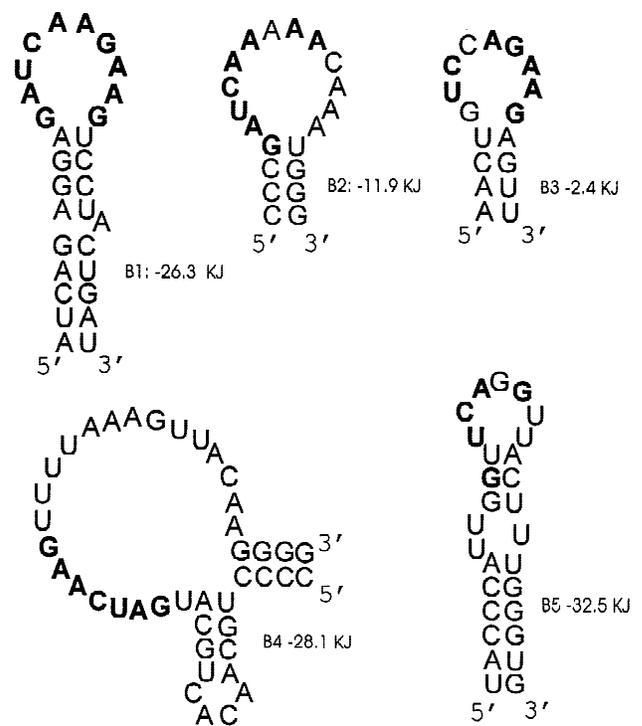
Yeast genomic DNA was partially digested with four restriction enzymes with four base-recognition sites, and fragments of 50–150 bp were gel-purified and ligated into pIII/MS2-2 to generate a library of  $>1.5 \times 10^6$  *E. coli* transformants. With an average insert size of 100 bp and a yeast genome size of  $\sim 1.3 \times 10^7$  bp, this library represents close to a 12-fold coverage of any sequence, or sixfold coverage of a sequence in the correct orientation for its normal transcription. The library was transformed into the L40coat strain containing the Snp1 activation domain hybrid plasmid. Transformants were plated on synthetic media lacking uracil, leucine, and tryptophan to select for the presence of the RNA plasmid, activation domain plasmid, and integrated MS2 coat protein fusion gene, respectively. The media also lacked histidine and contained 0.5 mM 3-aminotriazole to select for expression of the *HIS3* reporter gene. 3-aminotriazole inhibits the enzymatic function of the His3 protein and thus eliminates the background of colonies due to a low basal level of *HIS3* expression. From more than 1.5 million yeast transformants,  $\sim 250$  His<sup>+</sup> colonies formed after 4–5 days. Of these, 83 also expressed  $\beta$ -galactosidase activity.

For the 83 colonies active for both reporter genes, it was necessary to determine whether this activity was dependent on the presence of the Snp1 plasmid. We grew these colonies nonselectively in rich media, such that some of the cells lost the activation domain plasmid. This culture was then plated on media deficient in uracil but not leucine to identify those colonies that maintained the RNA, but not the activation domain plasmid. Colonies with this pattern were retested for  $\beta$ -galactosidase activity. Of the 83 transformants, only 13 lost  $\beta$ -galactosidase activity upon loss of the Snp1 plasmid. The other 70 expressed RNAs that were alone capable of activating reporter gene transcription when tethered to the promoter through the LexA-MS2 coat protein fusion.

The 13 Snp1-dependent RNA plasmids were reintroduced into the yeast reporter strain along with either the Snp1 plasmid or the pACTII vector. Nine of these RNAs led to at least twofold more  $\beta$ -galactosidase activity when present with the Snp1 hybrid than with the activation domain alone (Table 1). Sequence analysis of these nine RNAs indicated that the RNA yielding the greatest reporter gene activity (B1) contained 68 bases of U1 RNA, including the loop I. The Snp1 binding site in U1 consists of a 10-base loop (bold and underlined in Table 1) and a 9-bp stem with a single bulged nucleotide (Fig. 1, B1). Point mutations within the loop of the yeast U1 RNA affect splicing activity in vivo (Liao et al., 1990), whereas similar mutations in the loop of the human U1 RNA inhibit U1-70K protein binding in vitro (Surowy et al., 1989). We searched for sequences similar to the U1 loop nucleotides in the other four Snp1-dependent RNAs, and identified the best matches (Table 1), which included one RNA with an 8/10 match (B2) and two with 7/10 matches (B3 containing two such matches, and B4). The RNAdraw programs (Mazura & Wennborg, 1996) were then used to determine if a stable stem could be drawn that would display the best match as a loop (at 30 °C), and the corresponding stabilities for these structures (Fig. 1, B2–B5). For three of these RNAs (B2, B3, and B4), such a potential structure can be predicted. For the fourth (B5), a structure is shown in which a conserved G residue is base paired, but the short three-base stem containing this residue more likely would be single stranded when the protein binds.

### RNA activators

For the 70 RNA sequences that activated transcription independently of the presence of the Snp1 activation domain hybrid, we sequenced the 10 that displayed the greatest reporter gene activity. These plasmids defined four sets of sequences (Table 2). Retransformation of these RNA plasmids into a strain that did not



**FIGURE 1.** Predicted structures for Snp1 binders. Secondary structure predictions are shown for loop I of U1 RNA (B1) and U1-like sequences within RNAs (B2, B3, B4, and B5) that exhibit a three-hybrid signal dependent on the presence of the Snp1 hybrid. Nucleotide residues within the loop I region of U1 are shown in bold, as are the similar nucleotides within loops of B2 through B5.

express the LexA-coat protein fusion resulted in no  $\beta$ -galactosidase activity, whereas cotransformation with a plasmid that expressed LexA-coat protein reconstituted activity (data not shown). Thus, these RNA activators require tethering to the promoter of the reporter gene to function. Although these RNAs might activate by different means—for example, by recruiting to the reporter gene promoter different RNA-binding proteins that possess an activation domain—we sought to determine if they possess any common sequence or struc-

**TABLE 1.** RNAs that bind to the Snp1 protein.

Transcribed sequence of the insert	Number of times isolated	$\beta$ Gal with Snp1 plasmid	$\beta$ Gal with pACT II plasmid	Fold activation	Position relative to open reading frame
B1 5' uaagauaucagagg <b>gaucaagaag</b> uccuacugaucaacaucgcuuccaauguagaaggacguua 3'	1	40	1.2	33	<i>SNR19</i> (U1 RNA coding sequence)
B2 5' <b>gaucaaaa</b> caaauggggagccuccgucugugacgguuua 3'	1	12	2.5	4.8	On the antisense strand of <i>SAC3</i> gene
B3 5' aauuugagugggagaucucuaauagugauaacu <b>uccagaagag</b> uuugcagaguccaugu <b>ucaagaa</b> uu 3'	4	6.2	1.6	3.9	Transcribed within <i>SNF5</i> mRNA
B4 5' cugcaacacugcau <b>gaucaagu</b> uuuuuaaguuaacaag 3'	1	3.8	1.7	2.2	May be transcribed as <i>YEL057C</i>
B5 5' uaaauuacccauu <b>ggucagg</b> uuacuuggguguaucagauua 3'	2	5.9	1.7	3.5	May be transcribed as <i>YOR206W</i>

TABLE 2. RNAs that transactivate.

	Transcribed sequence of the insert	Number of times isolated	$\beta$ -gal with RNA plasmid alone	Position relative to open reading frame
TA1	5' agagaugcuggccugaggcauu 3'	1	23	On the antisense strand of YLR152C
TA2	5' uaaugacuugcucaacaaaauagcgugaauuaucaacuaaguaaacaacggcucuu cucacaccugaaaaacaagu 3'	4	11	Adjacent to YBRWdelta17
TA3	5' uaaugguaaucaacugguucugcuggacggagccauucuaaaucuuu 3'	2	8.1	Adjacent to YOLCdelta3
TA4	5' aauuuuaucaauggaaaaguacgcgccgcccgcgugaacuaaagauuu 3'	3	9.4	On the antisense strand of YJL086C

tural motif. Use of the algorithms of Gorodkin et al. (1997) to align these sequences indicated several possible sets of folds based on both similar secondary structure and primary sequence conservation, and one of these sets is shown in Figure 2. Additional analysis involving mutagenesis of specific nucleotides will be necessary to establish if these structural features correlate with transcriptional activation.

## DISCUSSION

We describe an approach that should be generally applicable to the characterization of RNAs that are recognized by an RNA-binding protein. Using the three-hybrid system, we have screened a hybrid RNA library containing yeast genomic sequences and correctly identified as the strongest positive the sequence in U1 RNA that binds to the yeast Snp1 protein. We also found four other RNAs that bind to Snp1 and that produce lower reporter gene activity in the three-hybrid assay, suggesting a weaker affinity for Snp1. However, it is

unlikely that these other RNAs bind to Snp1 in vivo, and indeed, the RNA sequence that yields the greatest reporter gene activity of these four would not be expected to be present in a transcribed RNA. Structure predictions for these four RNAs indicate that at least three of them can be folded into a stem-loop that contains loop nucleotides resembling those critically important for U1 function.

The three-hybrid assay is rapid and simple to carry out, detects interactions that occur in the native conditions of the cell, and does not require purification of the RNA-binding protein. It should thus serve as a complementary strategy to approaches such as SELEX (Tuerk & Gold, 1990; Gold et al., 1995) that detect protein–RNA interactions in vitro. The SELEX procedure has the advantageous feature that it can sample much larger numbers of RNA sequences (up to  $10^{15}$ ) than can be expressed in the three-hybrid assay following transformation of yeast (typically up to  $10^7$ ). However, SELEX detects interactions under arbitrarily defined in vitro conditions, and generally requires pure RNA-binding protein to carry out the assay.

Genome sequencing projects have defined many new putative RNA-binding proteins, based on the presence of motifs such as the RRM domain, for which no other information is available. By contrast, other RNA-binding proteins have been implicated in defined processes via genetic or biochemical approaches, but their RNA ligands remain unidentified. In both of these cases, the three-hybrid application described here may be helpful in adding key functional data to these proteins. The fact that RNAs are identified that are probably not biologically relevant targets of the protein suggests that rapid secondary screens of function will be essential.

The discovery that a large fraction of the initial positives in our search were due to RNA activators suggests that this class of sequence is likely to prove a substantial component of other RNA libraries. These RNAs may fold into structures that directly recruit proteins of the transcriptional machinery to activate reporter gene expression. Alternatively, the RNAs may bind to RNA-binding proteins that are not normally involved in transcription but that possess fortuitous activation domains. A simple means to eliminate these RNA

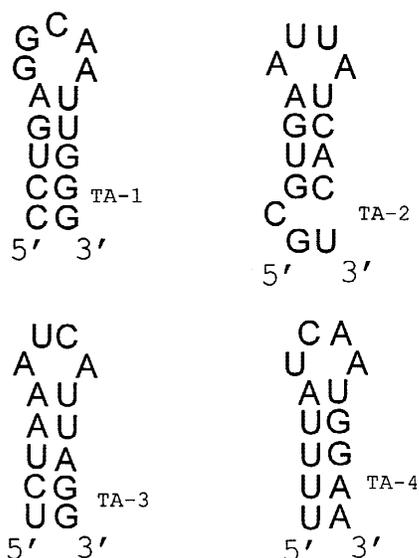


FIGURE 2. Predicted common structures within the RNA sequences that are transcriptional activators.

activators in the initial screening step is the inclusion of the yeast *ADE2* gene on the activation domain plasmid, because plasmid loss then leads to an *ade2* phenotype that is manifested by a red colony color. A comparable approach has been used to detect RNA-dependent positives in the screening of a cDNA library in an activation domain vector (Zhang et al., 1997).

Our demonstration that the three-hybrid assay can be used to identify RNA ligands for an RNA-binding protein from *S. cerevisiae* suggests that RNA libraries derived from other organisms may be fruitfully searched with this assay. The requirement for small inserts in the RNA expression plasmid and the relatively low transformation efficiency of yeast would favor the use of cDNA rather than genomic inserts for organisms with large genomes. Finally, RNA sequences that are capable of transcriptional activation when tethered to a promoter may find other uses in yeast genetic selections.

## MATERIALS AND METHODS

### Construction of small test libraries in pIII/IRE-MS2

pIII/IRE-MS2 expresses a hybrid RNA with binding sites for IRP1 and MS2 coat protein. A unique *Sma*I site between the sequences encoding the IRE and the coat protein binding sites was used to insert fragments of yeast genomic DNA. The genomic DNA was cut with *Alu*I or *Rsa*I, and fractionated on a 1.2% agarose gel. Fragments corresponding to 400, 200, and 100 bp were purified and ligated to the vector that had been cut with *Sma*I and treated with phosphatase.

### Construction of the yeast RNA library

Chromosomal DNA from *S. cerevisiae* strain S288C was partially digested in three digests: by *Mse*I (recognition site TTAA) and *Tsp*509I (AATT), by *Alu*I (AGCT) and *Rsa*I (GTAC), and by all four enzymes. Fragments ranging from 50 to 150 bp were purified from an agarose gel and the ends were made blunt with DNA polymerase I (Klenow fragment) where required. pIII/MS2-2 was digested with *Sma*I, treated with calf intestine phosphatase, and ligated to the genomic fragments, and the ligation mix was used to transform electrocompetent HB101 cells.

### Construction of U1 and Snp1 plasmids

The 5' 175 nt and 5' 45 nt of the *SNR19* (U1) gene were obtained by PCR using as a template pXL7 (Liao et al., 1990), incorporating *Sma*I sites into the PCR primers. The PCR fragments were digested with *Sma*I and ligated into *Sma*I-cut pIII/MS2-2 to construct pU1L1/MS2 and pU1L2/MS2, respectively. pSNP1/AD was constructed by ligating a PCR fragment containing the complete *SNP1* open reading frame into the *Bgl*II site of the activation domain vector pACT.

## RNA library screen

The RNA expression library was transformed into L40coat containing pSNP1/AD by the method of Gietz et al. (1992) and transformants were plated onto synthetic complete (SC) plates lacking tryptophan, leucine, uracil, and histidine and containing 0.5 mM 3-aminotriazole. Colonies were picked after 4–5 days, patched on to fresh plates, and when grown, tested for  $\beta$ -galactosidase activity by Xgal filter assay.  $\beta$ -galactosidase-positive colonies were grown nonselectively overnight in liquid YEPD media, and plated on SC-tryptophan-uracil for single colonies. These colonies were replica-plated to SC-leucine to identify colonies that had lost the pSNP1/AD plasmid and maintained the RNA plasmid. Two such colonies were picked from each plate and patched onto SC-tryptophan-uracil plates and retested for  $\beta$ -galactosidase activity to determine which RNA plasmids were Snp1-dependent (white colonies) or Snp1-independent (blue colonies) for reporter gene activity. Snp1-dependent RNA plasmids were tested for  $\beta$ -galactosidase activity in the presence of the pACTII vector. The plasmid that resulted in the highest level of  $\beta$ -galactosidase activity (40 U) conferred resistance to 50 mM 3-aminotriazole, whereas the plasmid resulting in 3.8 U of  $\beta$ -galactosidase activity conferred resistance to 2.5 mM 3-aminotriazole.

### $\beta$ -galactosidase assay

For quantitative assay, the procedure described in Miller (1972) was used, except that the substrate was chlorophenol red- $\beta$ -D-galactopyranoside (Iwabuchi et al., 1993). Units were calculated as  $1,000 \times OD_{574}$  divided by the product of the volume (mL)  $\times$  time (min)  $\times$   $OD_{600}$ . For each transformant, at least two independent colonies were assayed and their values averaged.

## ACKNOWLEDGMENTS

We thank Gary Stormo for help with the RNA structure predictions, Alia Qureshi-Emili for technical assistance, and Michael Rosbash for the U1 plasmid. This work was supported by grants from the National Institutes of Health (GM56906) and the National Science Foundation (Science and Technology Center BIR9214821). S.F. is an Investigator of the Howard Hughes Medical Institute.

Received November 24, 1998; returned for revision January 6, 1999; revised manuscript received January 11, 1999

## REFERENCES

- Bacharach E, Goff SP. 1998. Binding of the human immunodeficiency virus type 1 Gag protein to the viral RNA encapsidation signal in the yeast three-hybrid system. *J Virol* 72:6944–6949.
- Gietz D, St Jean A, Woods RA, Schiestl RH. 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* 20:1425.
- Gold L, Polisky B, Uhlenbeck O, Yarus M. 1995. Diversity of oligonucleotide functions. *Annu Rev Biochem* 64:763–797.
- Gorodkin J, Heyer LJ, Stormo GD. 1997. Finding the most significant common sequence and structure motifs in a set of RNA sequences. *Nucleic Acids Res* 25:3724–3732.

- Iwabuchi K, Li B, Bartel P, Fields S. 1993. Use of the two-hybrid system to identify the domain of p53 involved in oligomerization. *Oncogene* 8:1693–1696.
- Kao HY, Siliciano PG. 1992. The yeast homolog of the U1 snRNP protein 70K is encoded by the SNP1 gene. *Nucleic Acids Res* 20:4009–4013.
- Liao X, Kretzner L, Seraphin B, Rosbash M. 1990. Universally conserved and yeast-specific U1 snRNA sequences are important but not essential for U1 snRNP function. *Genes & Dev* 4:1766–1774.
- Martin F, Schaller A, Eglite S, Schumperli D, Muller B. 1997. The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J* 16:769–778.
- Matzura O, Wennborg A. 1996. RNAdraw: An integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Comput Appl Biosci* 12:247–249.
- Miller JH. 1972. *Experiments in molecular genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- SenGupta DJ, Zhang B, Kraemer B, Pochart P, Fields S, Wickens M. 1996. A three-hybrid system to detect RNA–protein interactions in vivo. *Proc Natl Acad Sci USA* 93:8496–8501.
- Surowy CS, van Santen VL, Scheib-Wixted SM, Spritz RA. 1989. Direct, sequence-specific binding of the human U1-70K ribonucleoprotein antigen protein to loop I of U1 small nuclear RNA. *Mol Cell Biol* 9:4179–4186.
- Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.
- Wang ZF, Whitfield ML, Ingledue TC 3rd, Dominski Z, Marzluff WF. 1996. The protein that binds the 3' end of histone mRNA: A novel RNA-binding protein required for histone pre-mRNA processing. *Gene & Dev* 10:3028–3040.
- Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP. 1997. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* 390:477–484.