ACADEMIC PRESS

Methods 26 (2002) 123–141

METHODS

www.academicpress.com

Analyzing mRNA–protein complexes using a yeast three-hybrid system

David S. Bernstein, Natascha Buter, Craig Stumpf, and Marvin Wickens*

Department of Biochemistry, University of Wisconsin—Madison, Madison, WI 53706, USA

Accepted 22 January 2002

Abstract

RNA–protein interactions are essential for the proper execution and regulation of every step in the life of a eukaryotic mRNA. Here we describe a three-hybrid system in which RNA–protein interactions can be analyzed using simple phenotypic or enzymatic assays in *Saccharomyces cerevisiae*. The system can be used to detect or confirm an RNA–protein interaction, to analyze RNA–protein interactions genetically, and to discover new protein or RNA partners when only one is known. Multicomponent complexes containing more than one protein can be detected, identified, and analyzed. We describe the method and how to use it, and discuss applications that bear particularly on eukaryotic mRNAs. © 2002 Elsevier Science (USA). All rights reserved.

1. Introduction

The interactions of mRNAs and proteins are critical for a wide variety of biological processes, ranging from developmental decisions to the proliferation of certain viruses. For this reason, several methods have been developed to analyze RNA–protein interactions using molecular genetics [1–10]. These complement an array of biochemical approaches.

We focus here on one such genetic method, the yeast three-hybrid system [1]. In this method, an RNA–protein interaction in yeast results in transcription of a reporter gene. The interaction can be monitored by cell growth, colony color, or the levels of a specific enzyme. Because the RNA–protein interaction of interest is analyzed independent of its normal biological function, a wide variety of interactions are accessible. Published applications of the method to date include: discovery of proteins that bind to a known RNA sequence, confirmation of suspected interactions between an RNA and protein, mutational analysis of interacting RNAs and proteins, and discovery and analysis of multiprotein:RNA complexes. In addition, three reports suggest that it should be feasible to identify the previously unknown RNA partner of an RNA binding protein [11–13].

The three-hybrid system, like other genetic strategies, has the attractive feature that a clone encoding the protein of interest is obtained directly in the screen. The system offers the possibility of connecting RNAs and proteins on a broad and, perhaps, genomewide scale.

The different genetic methods to detect RNA–protein interactions have distinctive advantages. Several systems in eubacteria have been used to examine peptide–RNA interactions, and to select peptides with altered specificities (e.g., [14,15]). To our knowledge, only the three-hybrid system has been used to identify new, naturally occurring RNA binding proteins of biological significance.

In this article, we describe the system, summarize its published uses, and present protocols. We first present principles and background, then discuss the key elements of the system including vectors, strains, and hybrid RNAs. Finally, we consider specific applications: analyzing a known interaction, finding a protein, finding an RNA, and analyzing multiprotein complexes. The utility of the system has benefited from adaptations in many laboratories over the past few years. We relate some of those developments, and hope to help others mold the system to new ends.

2. Principles of the method

The general strategy of the three-hybrid system is diagrammed in Fig. 1A. DNA binding sites are placed

* Corresponding author. Fax: 608-262-9108.
E-mail address: wickens@biochem.wisc.edu (M. Wickens).

1046-2023/02/$ - see front matter
PII: S1046-2023(02)00015-4

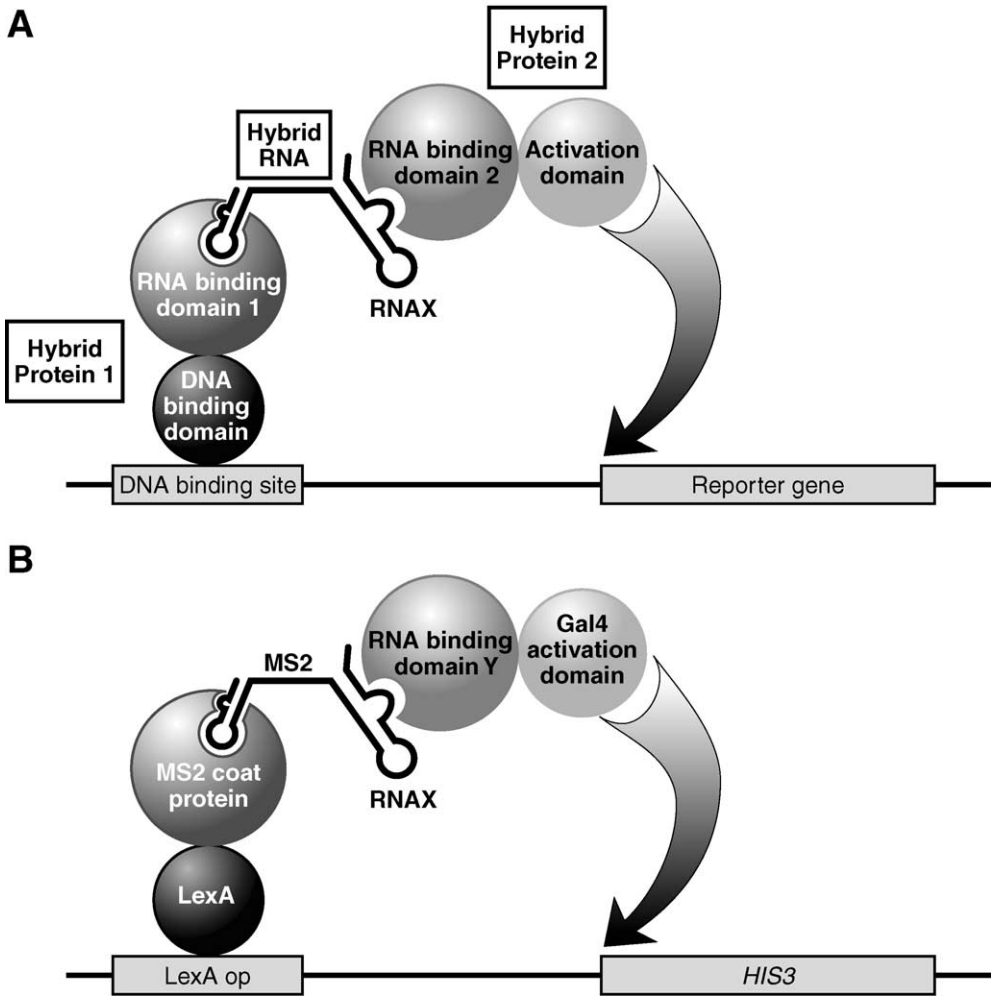

Fig. 1. Three-hybrid system to detect and analyze RNA–protein interactions. (A) General strategy of the system, (B) Specific protein and RNA components that typically have been used to date. Other RNA and protein components also can be used (see text). For simplicity, the following features are not indicated. Both *lacZ* and *HIS3* are present in strain L40coat and YBZ-1 under the control of lexA operators, and so can be used as reporters. Multiple LexA operators are present (four in the *HIS3* promoters and eight in the *lacZ* promoter), and LexA protein binds as a dimer. The hybrid RNA contains two MS2 binding sites, and MS2 coat protein binds as a dimer to a single site. Adapted, with permission from SenGupta et al. [1].

upstream of a reporter gene in the yeast chromosome. A first hybrid protein consists of a DNA binding domain linked to an RNA binding domain. The RNA binding domain interacts with its RNA binding site in a bifunctional (''hybrid'') RNA molecule. The other part of the RNA molecule interacts with a second hybrid protein consisting of another RNA binding domain linked to a transcription activation domain. When this tripartite complex forms at a promoter, even transiently, the reporter gene is turned on. Reporter expression can be detected by phenotype or simple biochemical assays.

The specific molecules most commonly used for three-hybrid analysis are depicted in Fig. 1B. The DNA binding site consists of a 17-nucleotide recognition site for the *Escherichia coli* LexA protein, and is present in multiple copies upstream of both the *HIS3* and *LacZ* genes. Hybrid protein 1 consists of LexA fused to bacteriophage MS2 coat protein, a small polypeptide that binds as a dimer to a short stem–loop sequence. The hybrid RNA (depicted in more detail in Fig. 3) consists of two MS2 coat protein binding sites linked to the RNA sequence of interest, X. Hybrid protein 2 consists of the transcription activation domain of the yeast Gal4 transcription factor linked to an RNA binding protein, Y.

Although the components depicted in Fig. 1B are most commonly used, others can be substituted. For example, the MS2 components can be replaced by either the histone mRNA 3′ stem–loop and the protein to which it binds (SLBP) (B. Zhang and M.W., unpublished); the NRE in *hunchback* mRNA's 3′ untranslated region (UTR) and its partner, Pumilio [16]; or hY5 and its partner, Ro60 [17]. Similarly, the Gal4 DNA binding domain can be used in place of LexA (B. Zhang and

M.W., unpublished). We are not aware of a systematic comparison of alternative components.

The three-hybrid approach has many of the same strengths and limitations of the two-hybrid system to detect protein–protein interactions. By introducing libraries of RNA or protein, cognate partners can be identified. As in two-hybrid screens, the challenge then becomes identifying those that are biologically relevant. For this purpose, mutations in the known RNA or protein component are very useful. Among the limitations of the system are certain technical constraints on producing hybrid RNAs and the low signal-to-noise ratio in screening a library when the genuine RNA–protein interaction is weak.

3. Initial considerations

The three-hybrid system has been used to analyze and identify specific RNA-binding proteins for a number of RNA targets. Table 1 summarizes directed tests of RNA and protein partners, and Table 2 summarizes successful screens of cDNA libraries. Several general properties of the system merit a brief discussion at the outset; some are considered in greater detail later.

3.1. Strength of interactions

The affinities of RNA–protein interactions detected by the three-hybrid system have K_d values from subnanomolar to micromolar (Table 1). Although levels of expression of the reporter gene generally correlate with measured in vitro affinities, the precise relationship between them has not yet been reported. Reporter expression levels presumably vary with the abundance of the two partners, the influence of endogenous yeast proteins and RNAs, and the reporter gene used. (See Ref. [18] for a discussion of related points in the two-hybrid system.)

3.2. The hybrid RNA

The RNase P RNA promoter, which is used to produce the hybrid RNA, has at least three desirable properties for the system: it is efficient, producing many copies of the hybrid RNA (up to 1300 copies per cell ([18]; B. Kraemer and M.W., unpublished)); a large portion of the natural RNase P RNA molecule can be replaced with foreign sequence [18]; and the use of this promoter and the RNase P RNA leader and trailer may cause the RNA to be nuclear. The high level of RNA likely enhances the signal because multicopy RNA plasmids yield higher reporter outputs than do centromeric plasmids (B. Zhang, B. Kraemer, and M.W., unpublished). On the other hand, the use of this RNA polymerase III promoter limits the RNAs that can be analyzed, both in length and in specific sequence (see Hybrid RNAs: Limitations): ideally, RNAs should be 150 or fewer nucleotides long, and should lack oligouridylate stretches of 4 or more nucleotides. However, these limitations are not universal: interactions have been detected using RNAs up to 1600 nt in length [19]. An alternative system exploiting an RNA polymerase II promoter has been described and should circumvent the sequence limitation [2].

3.3. Screens

In screening cDNA or RNA libraries for interacting partners, secondary screens of the initial "positives" are needed to winnow down candidates for further analysis (see Finding a Protein Partner for a Known RNA sequence, Step 5). One crucial secondary screen identifies positives requiring the "bait" to activate the reporter gene. Another screen identifies positives exhibiting the correct sequence specificity. Subtle mutations that perturb the biological function of the bait are ideal. Subtly mutant baits will bind most biologically irrelevant, "nonspecific" interactors, which can then be discarded. If subtle mutations are not available, cruder specificity tests (e.g., the antisense sequence) can be useful, almost surprisingly so (Table 2). Indeed, in the cDNA library screens carried out to date, virtually all positives that required the RNA to activate and possessed the correct sequence specificity proved to be the genuine interactor (Table 2). This may mean that few proteins bind nonspecifically with sufficient affinity to register in the system.

4. Key components: RNAs, vectors, and strains

4.1. Hybrid RNAs

4.1.1. General considerations

The RNA–protein interaction used to tether the hybrid RNA to the promoter must be specific and of high affinity. The RNA need not be structured. Several different interactions have been used; of these, by far the most common has been that between a short stem–loop RNA structure and bacteriophage MS2 coat protein.

Three features of the MS2 coat protein and its cognate binding sites make this pair attractive.

High affinity and selectivity. The stem–loop binding site used in the three-hybrid system binds to the coat protein with a k_d of 10^{-9}–10^{-10} [20]. The site contains a single nucleotide change that enhances binding [20].

Cooperativity. Binding of MS2 coat protein to adjacent sites is cooperative [21,22]. For this reason, two identical sites are used in the hybrid RNA.

The protein. MS2 coat protein is small and its RNA binding activity has been exploited for a variety of applications in the yeast *S. cerevisiae*.

Table 1
Test of previously known or suspected interactors

Protein component		RNA component		K_d	Refs.
Protein	Family	RNA	Length (nt)		
Cellular					
xlBrunoL3	Bruno	BREs	48	–	[44]
CUG-BP	Elav	long CUG repeats, UG repeats	[a]	–	[45]
CUG-BP + LYLQ	Elav	UG repeats	[a]	–	[45]
HBP (SLBP)	–	Histone RNA hairpin	28	–	[26]
IRP-1	Fe-S	IRE	51	10^{-10}–10^{-11}	[1]
Leu RS	–	bI4	1600	–	[19]
bI4 maturase	–	b14	1600	–	[19]
Mpt5	Puf	HO 3′UTR	134	–	[46]
PAB	RRM	Poly(A)	30	10^{-8}	[b]
Pop1	–	RPR1 P3	54	–	[47]
PTB	RRM	IRES	200	10^{-7}–10^{-8}	[b]
Pumilio	Puf (PUM-HD)	NRE	74	10^{-9}	[b], [16]
p50 (NF-kB)	–	ap50	57	10^{-9}	[23]
hnRNPA1	RRM	CD44 exonV5	200	–	[48]
hnRNPI, K	RRM, KH	hY1, 3	114	–	[49]
Rpp21, 29, 30, 38	–	H1 RNA	–	–	[50]
Smaug	–	nos TCE	90	–	[51]
yS14	–	yRps14B SL	59	5×10^{-7}	[29]
yS14	–	18S rRNA SL	70	3×10^{-6}	[29]
mTEP1	WD40	hvg1, 2, 4	99	–	[52]
TP1	WD40	mTR	393	–	[53]
TSR1IP	RRM	SL	146	–	[54]
U1-70K	RRM	U1 SL	45	–	[11]
Viral					
CaMV coat protein	Zn finger	CaMVpgRNA leader	87	–	[55]
HIV Gag	Retroviral CCHC	HIVy	139	–	[28]
RSV Gag	Retroviral CCHC	RSVMy	320	–	[27]
iso 4E	–	STNV	138	–	[c]
K-REV	Arg-rich	HERV-K RRE	433	–	[56,57]
Tap	–	MMPV CTE	138	–	[12]
Tat	Arg-rich	TAR	58	10^{-8}	[1]
Tat[d]	Arg-rich	TAR	55–126	–	[40,41]

[a] Number of nucleotides varied according to repeat number.
[b] Unpublished data from our laboratory.
[c] K.S. Browning, personal communication.
[d] Various combinations of Tat-1, Tat-2, and Tat-M with cyclin T were tested against TAR sequences from HIV-1, HIV-2, and SIVnd (see [40,41]).

RNA sequences to be tested can be inserted into the unique *Sma*I/*Xma*I and *Sph*I sites of any of the four RNA vectors depicted in Fig. 2. The hybrid RNA molecule that is transcribed in vivo from one of these plasmids consists of the sequence of interest, X, linked to two MS2 coat protein binding sites, and RNase P RNA 5′ leader and 3′ trailer sequences (Fig. 3).

The MS2 coat protein binding site is a highly structured stem–loop. The presence of this fold (as well as the structured 5′ region of RNase P RNA at the 5′ end of the hybrid molecule) should reduce the frequency of alternative structures involving the RNA sequence of interest. However, alternative structures can form that preclude interaction with both the MS2 coat protein and/or the protein of interest. This complication is illustrated in Fig. 4. RNA sequence X, which in its natural state is a short stem–loop (Fig. 4A), forms an alternative structure in which it base-pairs to the MS2 region. The resulting molecule (Fig. 4B) will not bind coat protein or its own protein partner. The formation of unproductive structures can be minimized by flanking the sequence with stable inverted repeats (a "GC clamp"), thereby restoring binding in the three-hybrid system Fig. 4C [23].

In our experience, most RNA inserts of suitable size produce RNAs of comparable and high abundance. RNA abundance may be important in determining the level of signal produced from the reporter gene: transferring the hybrid RNA gene from the normal high-copy vectors to low-copy vectors substantially reduces levels of *LacZ* expression.

The relative order of the RNA sequence of interest and the MS2 sites can affect signal strength. In the few cases that have been tested systematically, both orientations yield activation and are specific. However, in the IRE–IRP interaction, placing the IRE upstream of the

Table 2
Screens for cDNAs that interact with RNAs

Screen results				Screening information								Refs.
RNA	Length (nt)	Protein	Family	Trans-formants	[3-AT]	His$^+$	White	lacZ$^+$	RNA dependent	Positive interactors	Relevant in vivo	
mRNAs												
APP splice enhancer	72	CUG-BP	RRM	–	5 mM	–	–	–	–	2	–	[59]
		Siah-BP	RRM									
ash1 3′ UTR	126	Loc1p		2×10^6	5 mM	1500	66	66	11	4	2	[30]
ash1 E3	127	She3p		5×10^5	1 mM		49	23	9	1	1	[30]
fem-3 PME	74	FBF-1	Puf	5×10^6	5 mM	5000	100	60	12	3	3	[36]
		FBF-2										
Histone SL	35	SLBP-1	SLBP	2×10^6	5 mM	19	19	19	7	7	7	[63]
		SLBP-2	SLBP									
Histone SL	34	SLBP	SLBP	3×10^5	25 mM	4	4	4	1	1	1	[66]
nanos TCE A	29	Smaug		7.5×10^5	3 mM	77	–	35	9	1	1	[65]
prm1	37	MSY4	Y box	7.5×10^6	5 mM	113	–	10	5	1	1	[34]
tra-2 TGE	60	GLD-1	KH/STAR	6×10^5	5 mM	87	–		20	1	1	[62]
T. cruzi SL	111	XB1		2×10^6	2.5 mM	500	400	100	–	1	–	[60]
Viral RNAs												
HCV 3′X	98	L22	–	3×10^5	2 and 5 mM	30,000	6000	–	–	4	–	[61]
		L3	–									
		S3	–									
		mL3	–									
Influenza NP	50	GRSF-1	RRM	5×10^5	3 mM	120	–	120	10	2	2	[64]
Other												
hTR	158	L22		–	5 mM	–	–	–	–	2	–	[58]
		hStau	dsRBP	–								

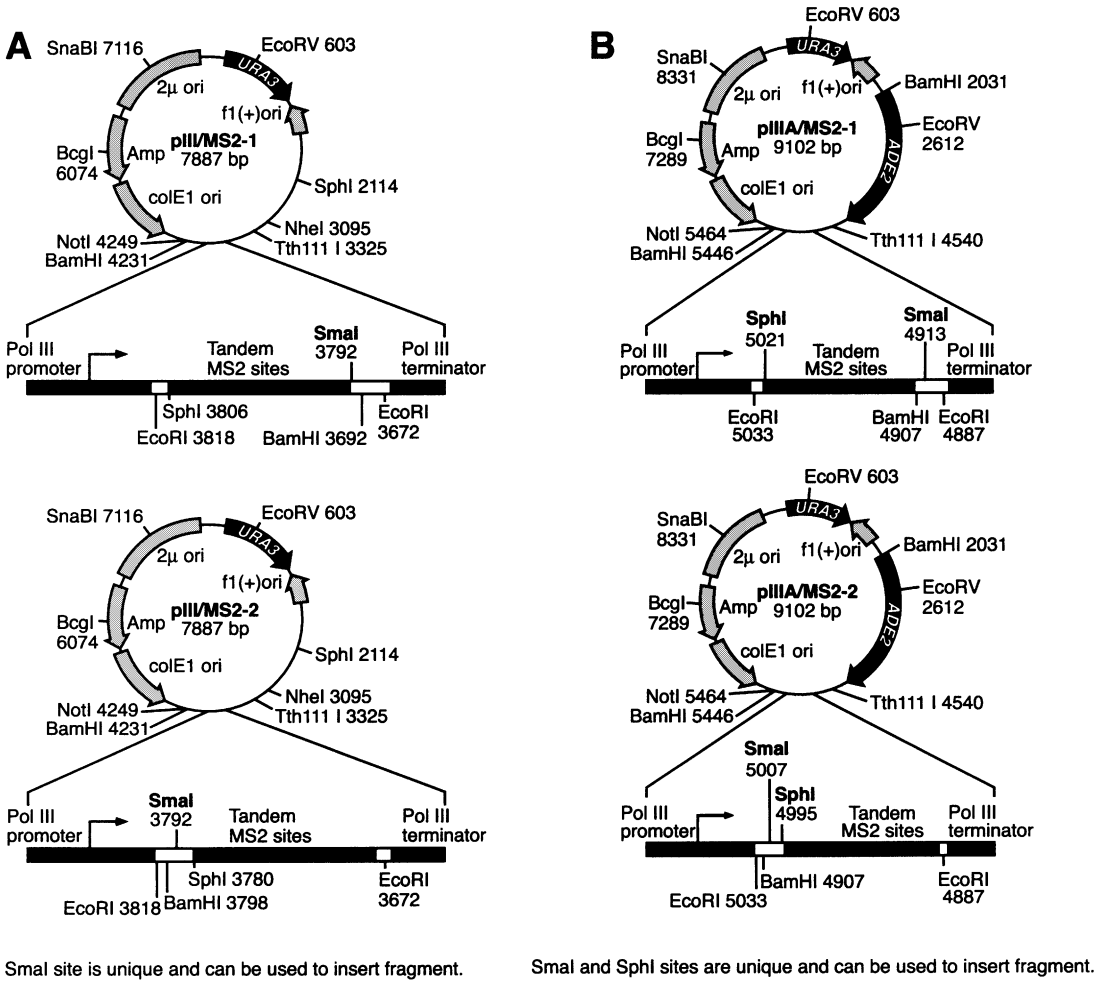

Fig. 2. Plasmid vectors used to express hybrid RNAs in the three-hybrid system. The restriction sites that are indicated either are unique or can be used to verify the presence of various elements on the plasmids. Sites suitable for insertion of sequences of interest are in boldface.

MS2 sites results in two- to threefold more transcription than does the alternative arrangement. Although RNA folding programs can be used to determine whether one arrangement is more likely to succeed, the accuracy of their predictions in vivo is problematic. In most cases, we have placed the RNA sequence of interest upstream of the MS2 sites; published reports have used both orientations.

4.1.2. Limitations

RNA sequences to be analyzed are restricted in two respects at present, though neither restriction is absolute. First, runs of four or more U's in succession can terminate transcription by RNA polymerase III. Second, typically, RNA inserts of lengths less than 150–200 nt yield higher signals: longer inserts commonly reduce the level of activation of the reporter (see below). In principle, both of these limitations might be overcome by using a different polymerase, such as any of the bacteriophage RNA polymerases. A system using RNA polymerase II has been described [2].

Four or more U's in succession can function as RNA polymerase III terminators, and so can prevent production of the desired hybrid RNA. The efficiency of termination at oligouridine tracts is context-dependent. Northern blotting can easily be performed to make certain that the hybrid RNAs are expressed at high levels. Termination in long runs of U's can be perturbed by appropriately placed single nucleotide substitutions.

The size of the RNA insert appears to be an important determinant of three-hybrid activity. In reconstruction experiments using known RNA–protein partners, RNA sequences that are less than 150 nt in length (e.g., TAR, IRE) typically yield substantial and specific reporter activation (Table 1). We have investigated the effects of additional RNA sequences flanking a known protein binding site. The addition of heterologous sequences to the IRE caused a reduction in the IRE–IRP1 three-hybrid signal. The effect on reporter gene expression was inversely proportional to the size of the insertion, with the addition of 150–200

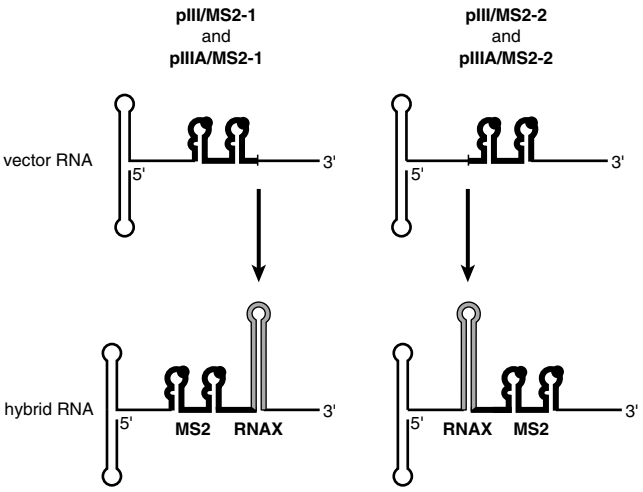

Fig. 3. Schematic diagram of the secondary structure of the hybrid RNA molecules. Thin lines: RNase P RNA 5′ leader and 3′ trailer sequences; bold lines: tandem MS2 recognition sites, including a point mutation that increases affinity (black dot); gray lines: inserted RNA sequence, depicted as a stem–loop.

nt commonly leading to almost complete abolition of the IRE–IRP1 three-hybrid signal [11]. Similarly, we have tested the effect of additional natural RNA sequence on the ability to detect the interaction of the yeast Snp1 protein, a homolog of the mammalian U1-70K protein, with the loop I of yeast U1 RNA. Insertion of even the RNA sequences that normally flank loop I, but are not part of the Snp1 binding site, decreased the three-hybrid signal due to Snp1-U1 interaction. Taken together, these experiments suggest a minimal binding site is preferable, and the optimal length of the RNA insert typically is less than 150–200 nt.

Nevertheless, substantial signals can be detected from longer RNAs. For example, Rho and Martinis used a 1600-nt RNA containing the yeast *cobI* intron to detect its interaction with two different proteins [19]. Similarly, we have detected the interaction of viral RNA of 500 nt with specific proteins (B. Kraemer and M.W., unpublished).

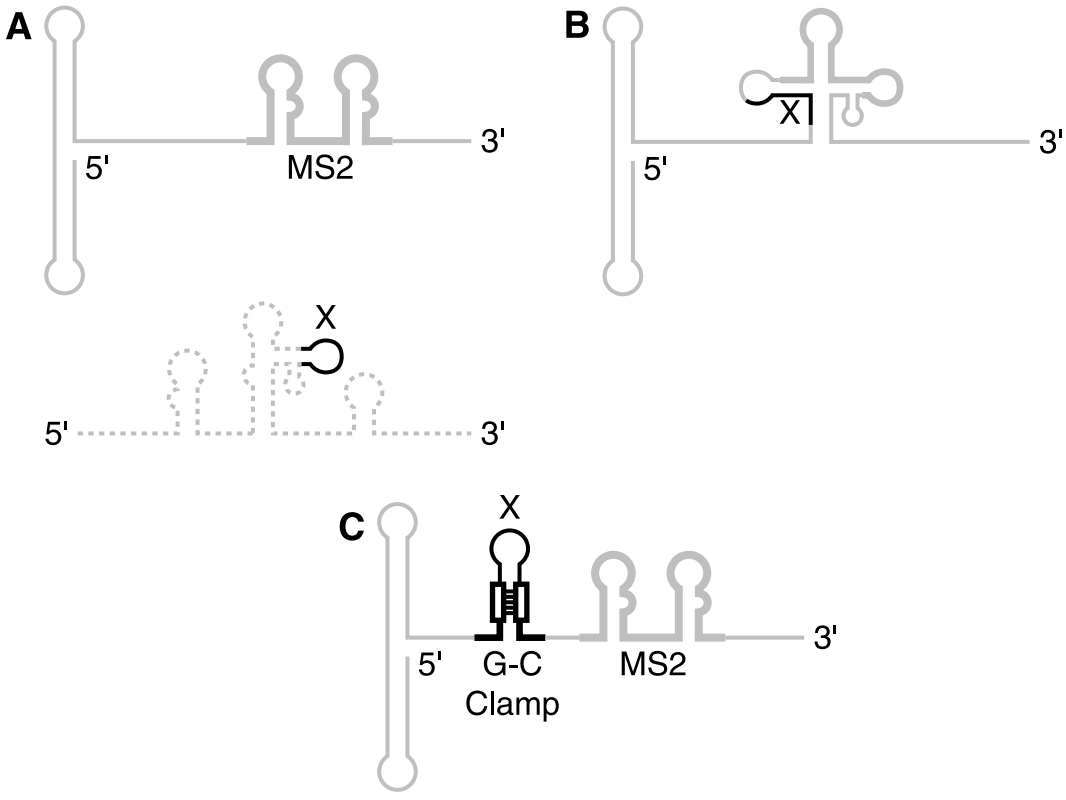

Fig. 4. "GC clamps" in a hybrid RNA molecule. A schematic diagram of the secondary structure of hybrid RNA molecules with and without the GC clamp. The RNA sequence that is being analyzed is a short stem–loop, X, indicated by a thin black line. Thin gray lines: RNase P RNA 5′ leader and 3′ trailer sequences; thick gray lines: tandem MS2 recognition sites; dashed gray lines: the natural RNA sequence in which RNA X resides; thick black line with boxes: "CG clamp." (A) Top: predicted structure of hybrid RNA vector sequence; bottom: RNA X in its natural context. (B) When inserted into the hybrid RNA (depicted in A), the new chimeric RNA (depicted in B) adopts a structure in which neither the stem–loop X nor the MS2 binding sites fold properly. (C) To restore the proper conformation, a GC clamp is inserted. See Ref. [25] for details.

4.2. Plasmid vectors

Several plasmids have been constructed to express hybrid RNA sequences, and are depicted in Fig. 2. Each of these is a multicopy plasmid containing origins and selectable markers for propagation in yeast and bacteria. They are described below. Additional information is available at www.biochem.wisc.edu/wickens, under ''Research/Three-Hybrid System.''

4.2.1. pIII/MS2-1 and pIII/MS2-2

pIII/MS2-1 and pIII/MS2-2 (Fig. 2A) are yeast shuttle vectors derived from pIIIEx426RPR, which was developed by Good and Engelke [18]. Sequences to be analyzed are inserted at the *Sma*I/*Xma*I site or *Sph*I site. pIII/MS2-1 and pIII/MS2-2 differ only in the relative position of these restriction sites and the MS2 binding sites. Both plasmids carry a *URA3* marker and produce hybrid RNAs from the yeast RNase P RNA (*RPR1*) promoter, an RNA polymerase III promoter. The *RPR1* promoter was chosen for two reasons. First, it is efficient, directing the synthesis of more than 1000 molecules per cell. Second, the transcripts produced from this promoter presumably do not enter the pre-mRNA processing pathway, and may not leave the nucleus. An alternative method using an RNA polymerase II promoter to generate the hybrid RNA has also been described [2].

4.2.2. pIIIA/MS2-1 and pIIIA/MS2-2

pIIIA/MS2-1 and pIIIA/MS2-2 (Fig. 2B) are similar to pIII/MS2-1 and pIII/MS2-2, but carry the yeast *ADE2* gene (in addition to *URA3*). The *ADE2* gene is exploited in screening for RNA binding proteins (see Finding a Protein Partner for a Known RNA Sequence, Step 2). The two plasmids differ from one another only in the relative position of the restriction sites and MS2 sites.

4.2.3. Commercially available vectors: pRH3′ and pRH5′

Modified vectors for producing hybrid RNAs are available from Invitrogen. These are derived from the vectors indicated above, but differ in the markers present on the plasmids and the cloning sites available for insertion of DNA. These vectors have been used successfully in several of the reports in Tables 1 and 2. Information on their details can be obtained at www.invitrogen.com. (The authors are not involved in their design, nor is this an endorsement.)

4.3. Yeast strains

The yeast reporter strain L40coat is derived from L40-ura3 (a gift of T. Triolo and R. Sternglanz, Stony Brook). The genotype of the strain is *MATa*, *ura3-52 leu2-3*, *112*, *his3-200*, *trp1-1*, *ade2*, *LYS2::(lexA op)-HIS3*, *LexA-MS2 coat (TRP1)*. The strain is auxotrophic for uracil, leucine, adenine, and histidine. Each of these markers is exploited in the three-hybrid system. Both the *HIS3* and *lacZ* genes have been placed under the control of *lexA* operators, and hence are reporters in the three-hybrid system. A gene encoding the LexA-MS2 coat protein fusion has been integrated into the chromosome.

Strain R40coat is identical to L40coat, but of opposite mating type.

Strain YBZ-1 is a derivative of L40coat (B. Zhang, S. Fields, and M.W., unpublished) of genotype *MATa*, *ura3-52*, *leu2-3*, *112*, *his3-200*, *trp1-1*, *ade2*, *LYS2::(LexAop)-HIS3*, *ura3::(lexA-op)-lacZ*, *LexA-MS2 MS2 coat (N55K)*. It contains a MS2 coat protein mutant (N55K) that decreases the K_d of the RNA–protein interaction from 3×10^{-9} to 2×10^{-10} M [24]. The coat protein mutants are also placed in tandem, creating an intramolecular coat protein dimer that enhances binding to the MS2 stem–loops. These modifications decrease the background of ''RNA-independent'' interactions and enable the detection of weaker interactions.

5. Identifying and analyzing known interactors

Assaying the interaction between a specific RNA and protein is straightforward. A yeast strain with an integrated reporter gene and LexA/MS2 coat protein is transformed with two plasmids: an activation domain plasmid carrying the protein of interest, and a hybrid RNA carrying the desired RNA sequence. If the protein and RNA interact, the reporter is expressed. *HIS3* and LacZ genes are the most common reporters used. Control experiments can be used to demonstrate that each component of the system is required for the interaction. Fig. 5 illustrates such a control using the IRE–IRP interaction, monitoring activation of *HIS3* by growth on selective media.

The three-hybrid system has been applied to known protein–RNA interactions for a variety of purposes, using proteins from many different families and species (Table 1). Mutational analysis of the protein and RNA is relatively straightforward. For example, Edwards et al. (2001) used the three-hybrid assay to isolate a large collection of single amino acid substitutions that disturb binding of a protein (Pumilio) to its RNA target (a portion of the *hunchback* 3′UTR); these mutations cluster on a surface of the protein that is presumed to interact with RNA, while mutations that do not affect RNA binding lie elsewhere [25]. Mutants also can be selected with altered or relaxed binding specificities. For example, through reiterative rounds of mutagenesis of histone SLBP, multiple mutant proteins were isolated that showed relaxed RNA binding specificity [26]. In

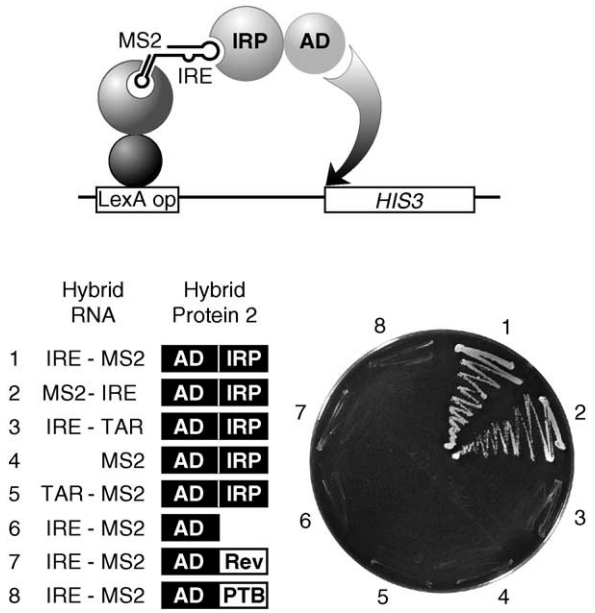

Fig. 5. IRE/IRP1 interaction monitored by activation of the HIS3. Plasmids encoding the indicated hybrid RNAs and activation domain fusions were transformed into strain L40coat. After selecting transformants for the presence of the plasmids, colonies were assayed for *HIS3* reporter activity. Only the two combinations that should lead to IRE–IRP1 interactions (lines 1 and 2) yield growth on media lacking histidine.

addition, the three-hybrid assay can reveal aspects of an RNA–protein interaction that are important in vivo, even when these are problematic to detect in vitro. For example, the NC portion of human immunodeficiency virus (HIV) GAG confers the protein's RNA binding specificity to HIVψ in the three-hybrid system, and mutants that perturb the interaction show parallel effects on packaging in infected cells [27]. In contrast, an in vitro gel shift assay did not discriminate NC and GAG binding [28].

The precise relationship between the affinity of an RNA–protein interaction (as measured in vitro) and the extent of reporter activation in the three-hybrid assay has not been analyzed extensively. While quantitative data are fragmentary, it is clear that variations in the strength of an interaction can be discriminated. In vitro binding and three-hybrid tests correlate well in the interaction of yS14 with its binding sites in rRNA and its pre-mRNA: the apparent K_d values (3 μM vs 0.5 μM, respectively) parallel the fivefold difference in β-galactosidase activity in three-hybrid assays [29]. Several RNA ligands that bind to HIV GAG in vitro with a range of affinities show a comparable range in the three-hybrid assay (B. Kraemer and M.W., unpublished). Judging by experience with two-hybrid assays, one might expect that extrapolating from one combination of protein and RNA to another will be difficult, since the levels of the components in the nucleus may vary, as may the interference by other cellular factors. Obviously, the absolute concentration of the components relative to the K_d also is critical.

"False-negatives" in a directed test can arise for many reasons, some trivial (e.g., the protein or RNA is not expressed or is imperfectly folded). However, there are a few instances in which a negative can be exploited: for example, it can be used to identify other proteins that enhance binding [30].

There are two definitions of "false" positives. "False" positives are sometimes defined as those interactions that register as positive but in fact do not reflect an RNA–protein interaction in the three-hybrid assay. This type of false positive appears to be rare. The second meaning of "false" positives is that the interactions occur in the three-hybrid system, but are not biologically relevant. There are three groups of false positives of this type.

1. *High-affinity, nonspecific interactions*. In principle, some proteins may interact with a restricted set of RNAs in the cell, but bind to many more RNAs in the three-hybrid system. Few such examples have been reported, but it seems likely they exist. On the other hand, the success of three-hybrid screens of cDNA libraries with specific regulatory elements as "bait" tends to suggest that few proteins have nonspecific affinities (at least for these sequences) in the nanomolar range. Obviously the mere fact that a protein binds to many RNA sequences with reasonable affinity does not preclude their being real targets in vivo.
2. *Bridges*. The protein that appears to interact with the RNA may in fact interact with a yeast cellular protein that binds to the RNA. This obviously is most germane when the activation domain hybrid is a yeast protein. For example, yeast She3p registers as binding to a portion of *ASH1* mRNA's 3′UTR in the three-hybrid assay, but does so because it interacts with cellular She2p which is bound to the hybrid RNA [30]. If *she2* is removed genetically, the apparent interaction disappears [30].
3. *RNA-independent interactions*. The protein of interest being introduced into yeast can bind directly to the LexA/MS2 coat protein fusion, independent of the hybrid RNA. This is an issue in performing screens for RNA-binding proteins using cDNA libraries, but is not commonly a problem in directed tests of any specific protein.

5.1. Assaying interactions: β-galactosidase activity and 3-AT resistance

The "strength" of an interaction is gauged by assaying the reporter gene's activity. Typically, this is done either by assaying β-galactosidase activity or by deter-

mining the level of resistance to 3-aminotriazole, which monitors *HIS3* activity. These two assays are described below.

5.1.1. β-Galactosidase

β-Galactosidase can be assayed by measuring the conversion of a lactose analog to a chromogenic or luminescent product. This assay can be performed using either colonies permeabilized on a filter or a cell lysate. The filter assay yields qualitative results, while the liquid assay is more quantitative.

5.1.1.1. Qualitative (filter) assays [31].

1. Streak onto the appropriate selective medium (SD −Leu −Ura) and grow overnight. Alternatively, place the nitrocellulose filter on the plate and streak directly onto the filter. This can yield cleaner results.
2. Immerse filter in liquid nitrogen for 5–20 s.
3. Allow filter to thaw on benchtop (approx 1 min).
4. Prepare Petri dish-sized circles of 3MM Whatman (Clifton, NJ) paper, place in a Petri dish, and saturate with Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH 7.0), supplemented with 300 μg/mL X-Gal. Remove excess buffer. Z buffer should be refrigerated. The X-Gal should be added fresh. The β-mercaptoethanol is optional; it slows the appearance of blue product, which can be helpful in comparing binding strengths.
5. Overlay filter onto Whatman paper and seal dish with Parafilm.
6. Incubate colony side up 30 min to overnight at 30 °C. Examine the filters regularly. A strong interaction (such as that between IRE and IRP) should turn blue within 30 min. With protracted incubation, weak interactions eventually yield a blue color. For this reason, it is important to examine the filters periodically to determine how long it takes for the color to develop.

5.1.1.2. Quantitative (liquid) assays. The specific activity of β-galactosidase can be determined in yeast cell lysates using any of a variety of substrates. Colorimetric assays using ONPG [32] or CPRG [33] are common; CPRG is more sensitive, but also more expensive. Alternatively, luminescent substrates provide high sensitivity yet are relatively inexpensive. Below we provide a protocol using a luminescent substrate, Galacton-Plus (Tropix, 47 Wiggins Avenue, Bedford, MA, Catalog No. BL300P). The assay requires an instrument to detect luminescence. The protocol below was designed for a Monolight 2010 Luminometer (Analytic Luminescent Laboratories, 11760 Sollento Valley Road, San Diego, CA). Certain details of the assay, such as sample volumes, will vary with the instrument used.

1. Inoculate 5-mL cultures of selective medium in triplicate for each interaction to be tested. Grow overnight.
2. Inoculate fresh selective medium to an OD of 0.1.
3. Grow to midlog phase (OD ∼0.8).
4. Pellet ∼1.0 OD unit worth of cells for each culture.
5. Resuspend pellet in 100 μL of lysis buffer (100 mM potassium phosphate, pH 7.8, 0.2% Triton X-100).
6. Lyse by freeze–thaw. This lysis requires three sequential cycles of freezing in liquid nitrogen for 10 s, followed by incubation at 37 °C for 90 s.
7. Vortex each tube briefly.
8. Pellet in microfuge at 12,000*g*.
9. Collect supernatant for Luminometer assays. If necessary, samples may be frozen at −70 °C.
10. Add 10 μL of lysate to luminometer tube. For strong interactions, it my be necessary to dilute the lysate to keep the assay in the linear range.
11. Dilute Galacton reagent 1:100 in reaction buffer (100 mM sodium phosphate, 1 mM magnesium chloride, pH 8.0). Add 100 μL to each luminometer tube.
12. Incubate at 25 °C for 60 min in the dark.
13. Measure luminescence as directed by Luminometer manufacturer. Use light emission accelerator II from Tropix.
14. Measure protein concentration in lysates by Bradford or equivalent assay.
15. Normalize light emission by protein concentration, yielding a specific activity.

5.1.2. 3-AT resistance

3-Aminotriazole (3-AT) is a competitive inhibitor of the *HIS3* gene product, His3p. Cells containing more His3p can survive at higher concentrations of 3-AT. Thus, the level of resistance to 3-AT monitors the strength of an RNA–protein interaction. An example of such an assay, titrating levels of 3-AT resistance, is shown in Fig. 6. Various combinations of plasmids expressing a protein (p1 or p2) and an RNA (r1, r2, etc) were introduced into a three-hybrid strain. Transformants were streaked onto plates lacking histidine, and containing either 0, 1, or 10 mM 3-AT, as indicated. (The leftmost plate is a control in which histidine is present; all cells grow, as expected.) p1 binds both r1 and r2, but binds r1 more tightly; as a result, cells containing the p1/r1 combination can grow on 10 mM 3-AT, while p1/r2 cells grow on 1 mM (but not 10 mM). Mutations that disrupt the binding site in the RNAs (r1* and r2*) do not bind p1, and so do not support growth even without 3-AT. A second protein, p2, does not bind to any of the RNAs, and so does not grow under selective conditions.

In our experience, weak interactions yield resistance to 1 mM 3-AT; strong interactions can yield resistance to 100 mM.

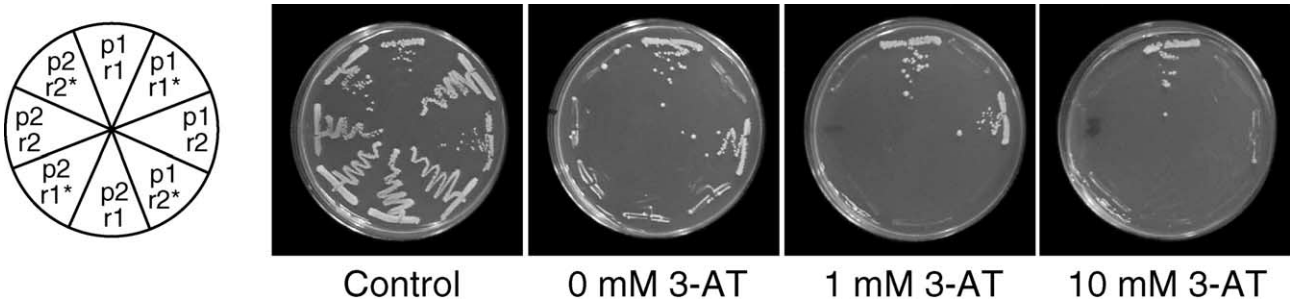

Fig. 6. RNA–protein interactions monitored by level of resistance to 3-AT. A sample growth assay on medium lacking histidine and containing 3-AT. The first plate is a control plate (−ura, −leu) demonstrating that all combinations are viable in the absence of selection for *HIS3*. The second, third, and fourth plates lack histidine and contain 0, 1, and 10 mM 3-AT, as indicated. Protein p1 interacts strongly with RNA r1 and moderately with RNA r2, but not with mutated RNAs, r1* and r2*. A related protein, p2, does not interact with either RNA.

5.1.2.1. 3-AT resistance assay.

1. Prepare a series of selective medium plates that lack histidine and contain increasing amounts of 3-AT (e.g., 0, 1, 5, 10, 15, 20 mM). Also prepare plates lacking leucine and uracil.
2. Transform the appropriate RNA and protein-encoding plasmids into a three-hybrid strain. Plate on SD −Leu −Ura medium. Incubate at 30 °C for 2 days.
3. Pick multiple average-sized colonies and streak for single colonies on −Leu −Ura and −His plates (while this is not ideal microbiological practice, it has worked reproducibly).
4. Incubate at 30 °C for 3–5 days. Growth should be assessed by the presence of individual colonies and not the smear of yeast from the initial streak. It is common to observe a few large colonies on high concentrations of 3-AT. These are probably the result of mutations in three-hybrid components or *HIS3*.

6. Finding protein partners of a known RNA sequence

The three-hybrid system can be used to identify a protein partner of a known RNA sequence. Typically, a three-hybrid reporter strain is created that expresses the RNA of interest as a hybrid molecule. A cDNA library is introduced into this strain by conventional transformation methods. When the RNA interacts with the protein produced from a cDNA, the *HIS3* gene becomes active and the yeast grow on media lacking histidine and/or containing 3-AT. Additional selections can be performed using the LacZ reporter and colorimetric assays. In principle, any selection that has been used in the two-hybrid system is applicable to the three-hybrid system as well. Several successful three-hybrid screens have been reported, and have yielded proteins of diverse families of RNA binding proteins (Table 2).

If the RNA–protein interaction is particularly strong, then the initial selection can demand high levels of expression of the reporter gene (e.g., high levels of 3-AT). A stringent selection eliminates weak (and potentially nonspecific) activators. However, a less stringent selection is often preferable because the "strength" of the interaction being sought is not known. In turn, this reduced stringency leads to a higher background, and greater need for subsequent screens. While high levels of 3-AT reduce background, they do so at the expense of the diversity of recovered cDNAs, as only the strongest interactors will be recovered. Screens have been performed with concentrations of 3-AT ranging from 0.5 mM (in the case of *fem-3* FBF [36]) to 25 mM (in the case of histone mRNA SLBP [63,66]).

The results of several screens published to date reveal that few positive clones satisfying selection criteria were isolated, but that the "correct" cDNA was among them (Table 2). In one remarkable example, 7.5 million transformants were screened, and after secondary selections based on RNA binding specificity, a single positive was identified that proved to be legitimate [34]. The number of positives obtained is determined by many variables, including the abundance of the cDNA in the library. Nonetheless, the published examples emphasize two practical points. First, at least for very sequence-specific, high-affinity RNA-binding proteins, one should arrange the stringency of the screen (e.g., the 3-AT concentration) to yield a reasonable number of initial positives. Second, subsequent screens that analyze the RNA sequence specificity of the protein are extremely useful.

Two categories of false positives are common and must be weeded out. Certain proteins bind nonspecifically to the RNA, while others activate the reporter independent of the RNA ("RNA-independent positives."). The second category can be very common in cDNA library screens, and it is important they be identified and discarded. Both categories of "false" positives can be identified by curing the RNA plasmid from the yeast, and retesting for activation in the resulting strain that lacks the hybrid RNA. Following this test, a new battery of RNA plasmids can be reintroduced to test for specificity.

6.1. Step 1: Introduce the RNA plasmid and cDNA library

The host strain is normally transformed with the RNA plasmid first. Cells containing the RNA plasmid are then transformed with a cDNA library fused with a transcription activation domain. [Hybrid RNAs usually are not toxic to the host cell. When they are toxic, one can cotransform both the RNA plasmid and the cDNA library (see below).] The transformation mix is plated out on media lacking both leucine (selecting for the cDNA plasmid) and histidine, as well as 3-aminotriazole (selecting for *HIS3* gene expression). Maintenance of the RNA plasmid (i.e., selection for *ADE2* or *URA3*) is not demanded. This permits cells that can activate *HIS3* without the RNA to lose the RNA plasmid, permitting the colony color screen below (see Step 2).

6.1.1. Protein libraries

Activation domain libraries prepared in *LEU2* vectors can be used with the hybrid RNA plasmids and yeast strains described above. Many such activation domain libraries have been prepared for use with the two-hybrid system. These may be obtained from individual laboratories and several commercial sources.

6.1.2. 3-Aminotriazole in the initial selection

We typically add 3-AT to select for stronger interactions. Some RNAs activate reporter genes weakly on their own, and many proteins appear to activate the reporter slightly, independent of the hybrid RNA; both situations yield "false positives." To eliminate weak activation by the RNA "bait," 3-AT should be titrated using a strain carrying the RNA plasmid and an "empty" activation domain plasmid prior to undertaking the initial transformation. Concentrations of 3-AT in the range 1–5 mM are a good starting point; they offer a reasonable balance between suppressing the background and permitting "real" positives to grow (see Table 2). In vitro data on the affinity of the RNA–protein interaction may be valuable for determining the concentration of the 3-AT that should be used; stronger interactions allow more 3-AT to be used.

6.1.3. Toxicity problems

Some RNAs are toxic to yeast and will make a titration prior to the screen impossible. In most cases, toxicity is reduced or eliminated by binding of a protein to the RNA. One simple explanation of the toxicity is titration of an essential yeast protein. In such cases, a cDNA library screen can be done by cotransforming the hybrid RNA plasmid and the cDNA library plasmids in at least a 10:1 molar ratio and plating on low levels of 3-AT. Hybrid proteins can also be toxic, but this too is sometimes mitigated by an RNA interaction. In some cases, yeast containing either the RNA or the protein grow poorly, but yeast containing both grow well.

6.2. Step 2: Eliminate RNA-independent false positives by colony color

Two classes of positives are obtained from the initial transformation. One class of transformants requires the hybrid RNA to activate *HIS3*; these are termed "RNA-dependent." A second class of positives activates *HIS3* with or without the hybrid RNA; these are termed "RNA-independent." The RNA-independent positives can carry proteins that may bind to the promoter regions of the reporter genes or proteins that interact directly with the LexA-MS2 coat protein fusion. The RNA-independent class of transformants can be very abundant, and can account for more than 95% of the total number of colonies.

To facilitate eliminating RNA-independent false positives, we exploit the *ADE2* gene on the pIIIA RNA plasmids. The host strain is an *ade2* mutant. When the level of adenine in the medium becomes low, cells attempt to synthesize adenine de novo and accumulate a red purine metabolite due to lack of the *ADE2*-encoded enzyme. This accumulation renders the cell pink or red in color. In contrast, cells carrying the wild-type *ADE2* gene are white.

In the initial transformation, selection is imposed only for activation of *HIS3* and the presence of activation domain plasmid, but not for maintenance of the RNA plasmid. For RNA-dependent positives, selection for *HIS3* indirectly selects for the RNA plasmid, which carries the *ADE2* gene; thus these transformants are white and must remain so if they continue to grow in the absence of exogenous histidine. On the other hand, RNA-independent positives do not require the RNA plasmid to activate the *HIS3* gene and so can lose that plasmid, which they do with a frequency of a few percent per generation. These false positives therefore yield pink colonies or white colonies with pink sectors.

The initial transformation plates are usually incubated at 30 °C for a week. This duration allows positives to accumulate *HIS3* gene product and grow, and also provides enough time for the color to develop. If the pink color is not strong after a week, incubation at 4 °C overnight can help. Pink or pink-sectored colonies are discarded, and the uniformly white colonies are picked for further analysis. We usually pick all the white colonies (typically a few large and many small, colonies) and patch onto media selecting again for both the cDNA plasmid and the RNA plasmid. Most of the small white colonies turn out to be RNA-independent and fail to grow. White colonies that are able to grow on the selective media are subject to further analysis.

The identification of RNA-dependent positives by colony color is not perfect; many, but not all,

RNA-independent positives can be identified and discarded. A majority of the white colonies may still be RNA-independent. It is important to rigorously eliminate the remaining RNA-independent activators from among the white colonies (Steps 4 and 5) before recovering plasmids in *E. coli*.

An additional and complementary means for eliminating false positives has recently been described [35]. In this protocol, *HIS3*+ colonies from the initial plating are patched onto media containing the drug 5-fluoroorotic acid (5-FOA). 5-FOA is toxic to yeast expressing the *URA3* gene. Cells requiring the hybrid RNA to activate the *HIS3* reporter will carry the hybrid RNA encoding plasmid at high copy number and so be 5-FOA sensitive. Cells not requiring the hybrid RNA to activate the *HIS3* reporter will carry the hybrid RNA plasmid at low copy number (or not at all) and so be relatively 5-FOA insensitive.

6.3. Step 3: Assay β-galactosidase activity

To corroborate that the *HIS3*+ colonies contain cDNAs that activate through the three-hybrid system, the level of expression of the *lacZ* gene is monitored. The assay is as described above (Testing Known Interactions). In strains L40coat and YBZ-1, the *lacZ* gene is integrated into the chromosome and placed under the control of LexA binding sites. In our experience, most *HIS3*+ transformants are also *LacZ*+; thus this step sometimes is delayed until after the specificity test (Step 5 below).

6.4. Step 4: Cure the RNA plasmid and test again

Most but not all of the RNA-independent false positives are eliminated by either the colony color or 5-FOA sensitivity assays in Step 2. To ensure that the positives are genuinely RNA-dependent, the RNA plasmid is removed by counterselection against *URA3*. Expression of the reporter gene is then monitored. Candidates that fail to activate the reporter genes are analyzed further.

6.4.1. URA3 counterselection

To select for cells that have lost the RNA plasmid, cells are plated on media containing 5-FOA. 5-FOA is converted by the *URA3* gene product to 5-fluorouracil, which is toxic. Cells lacking the *URA3* gene product can grow in the presence of 5-FOA if uracil is provided, while cells containing the *URA3* gene product cannot.

1. Replica-plate the positives from Step 2 to SD −Leu plates. Let the cells grow for a day, allowing cells to lose the plasmid.
2. Replica-plate onto SD −Leu + 0.1% 5-FOA plates. Incubate at 30 °C for a few days.
3. Cells that grow can be streaked on SD −Ura plate to confirm the loss of the RNA plasmid. A single pass through 5-FOA counterselection is usually sufficient.
4. Assay β-galactosidase activity.

The *ADE2* marker on the RNA plasmid can be useful in monitoring the loss of the RNA plasmid. Cells that lose the plasmid will turn pink in a few days. If the number of the positives is small, then the use of 5-FOA, which is quite expensive, can be avoided. Cells can simply be grown in rich medium overnight, then spread onto SD-Leu plates. After a few days, some of the colonies become pink or show pink sectors. Uniformly pink colonies, which have lost the RNA plasmid, can be reassayed for β-galactosidase activity.

6.5. Step 5: Determine binding specificity using mutant and control RNAs

To test RNA binding specificity, reintroduce plasmids encoding various hybrid RNAs into the strains that have been cured of their original RNA plasmid. If the number of positives is small, the various RNA plasmids can be introduced by transformation. Otherwise, mating is used to introduce the plasmids. Strain R40coat can be used for this purpose. A sample protocol for the mating assay follows.

1. Grow lawns of separate R40coat transformants carrying a specific hybrid RNA plasmid (mutant vs wild type, for example) on SD −Ura plates.
2. Replica plate the grid of Ura-colonies from the 5-FOA plate to a YPD plate.
3. Replica plate the lawn from each R40coat strain to the same YPD plate.
4. Incubate the plates overnight at 30 °C to allow mating.
5. Replica plate to SD −Leu −Ura plate to select for diploids.
6. Assay β-galactosidase activity.

6.5.1. What RNAs should be used in specificity tests?

Ab initio, positives that survive Step 4 carry proteins that bind preferentially to the hybrid RNA relative to cellular RNAs. Although these positives recognize some features on the hybrid RNA, these need not be the ones recognized by the biologically relevant factor(s). Therefore, the ideal controls are subtle (e.g., point) mutations that affect the biological functions or interaction in a sequence-specific manner. When point mutations are not available, RNAs carrying the antisense sequence, small deletions, or unrelated elements of similar predicted structure can serve as cruder specificity controls.

6.5.2. What fraction of all RNA-dependent positives is "correct?"

The fraction of physiologically relevant positives is unpredictable at the outset, as it is a function of the

abundance of the protein (in a random library screen), strength of the interaction, and level of 3-AT used in the initial transformation, among other parameters. A synopsis of several published screens is shown in Table 2, and indicates the number of isolates obtained at each step of the screening protocol. For example, we identified the regulatory proteins, FBF-1 and FBF-2, by screening a *Caenorhabditis elegans* cDNA/activation domain library using a portion of the *fem-3* 3′UTR as bait [36] In this screen, 5 mM 3-AT was used in the initial transformation, selecting for reasonably strong interactions. In total, 5,000,000 transformants were analyzed, yielding approximately 5000 His+ Leu+ colonies. Of these, 100 were white. Sixty of these activated the *lacZ* gene. Twelve of the sixty proved to be genuinely RNA-dependent, and 3 of the 12 displayed the appropriate binding specificity. Each of these was FBF, the genuine regulator. Thus, RNA-dependent positives were 0.2% of the total number of colonies obtained in the initial transformation. Importantly, 25% of all RNA-dependent positives were FBF. Moreover, in this screen and the others reported to date, nearly all positives showing proper specificity were the genuine regulator.

6.5.3. What if no subtle RNA mutations are available?

Additional screens must be devised to identify the "correct" positives. Clearly, functional tests are ideal. The sequence of the cDNAs may be directly informative by comparison to known RNA binding proteins, or by comparison to the predicted molecular weight of the expected protein (based on, e.g., UV crosslinking.) Each case is idiosyncratic, and so no general discussion is offered here. However, secondary screens are critical.

6.6. Step 6: Identify the positive cDNAs

cDNA plasmids that display the predicted RNA binding specificity are recovered from the yeast cells and introduced into *E. coli* by transformation. The yeast cells can contain multiple cDNA plasmids, only one of which encodes the protein that binds to the RNA. Thus, plasmids should be isolated from multiple *E. coli* transformants, and reintroduced into yeast to ensure the correct plasmid has been obtained. The EZ yeast plasmid prep kit (Genotech, Inc., St. Louis, MO) and the QIAprep Spin Miniprep kit yeast protocol (QIagen, Valencia, CA) allow rapid isolation of plasmids from yeast. Alternatively, the following method works well in our hands [37].

1. Culture yeast colony overnight in liquid SD −Leu medium.
2. Pellet 1.5 mL of overnight; resuspend in 100 μL lysis solution (5% Triton X-100, 8% sucrose, 100 mM NaCl, 50 mM Tris–Cl, pH 8.0, 50 mM EDTA).
3. Add ca. 0.2 g acid-washed glass beads.
4. Vortex at high speed for 5 min Add 100 μL lysis solution. Mix.
5. Boil 3 min. Chill on ice briefly.
6. Spin at high speed in a microcentrifuge for 10 min at 4 °C.
7. Transfer 100 μL of the supernatant to a clean tube containing 50 μL 7.5 M ammonium acetate. Chill −20 °C for 1 h.
8. Spin at high speed in a microcentrifuge for 10 min at 4 °C.
9. Add 100 μL of supernatant to 200 μL of ice-cold ethanol to precipitate DNA.
10. Spin at high speed in a microcentrifuge for 10 min at 4 °C.
11. Wash pellet in 70% (v/v) ethanol.
12. Resuspend the pellet in 20 μL H_2O. Use 10 μL to transform *E. coli*.

For convenience, to determine how many plasmids are present in each yeast transformant, one can perform polymerase chain reactions (PCRs) using lysed yeast colonies.

6.7. Step 7: Functional tests or additional screens

Almost invariably, additional steps will be needed to identify those positives that are biologically meaningful. As stated earlier (see Step 5), those screens are idiosyncratic, depending on the interaction and organisms studied. In a screen with an RNA "bait," it is not surprising that one might identify irrelevant RNA-binding proteins, as well as the legitimate interactor.

7. Finding an RNA partner for a known RNA binding protein

The three-hybrid system can be used to identify a natural RNA ligand for a known RNA-binding protein by screening an RNA library with a protein-activation domain fusion as bait. Although only three experiments of this type have been reported to date, the objective is sufficiently important that we discuss both the published work and our own experience. These experiments have obvious parallels to conventional in vitro reiterative selection methods, including SELEX and "genomic SELEX" [38,39].

A prototypical screen identified RNAs that interact with the yeast Snp1 protein [11]. A Snp1p/activation domain protein was introduced into L40coat. A hybrid RNA library was next introduced by transformation. This library consisted of fragments of yeast genomic DNA of ca. 100 nucleotides transcribed along with MS2 coat protein-binding sites. The strongest positive from the screen was a fragment of U1 RNA that is known to

bind to Snp1p in U1 snRNP. This screen demonstrates that a three-hybrid screen can identify meaningful interactions from a large library of candidate RNAs.

A related screen exploited a randomized RNA sequence to find RNA ligands for a given RNA binding protein. RNA sequences that could interact with a Tap/activation-domain fusion were identified from an RNA library composed of the MPMV CTE randomized at seven positions. Six sequences were recovered from a possible 16,384. Of the six, one was wild type, one activated the reporter without the Tap/activation-domain fusion, one changed four of the seven residues, and three were single-nucleotide substitutions, demonstrating that individual nucleotides involved in an RNA–protein interaction can be identified. Compared with "standard" SELEX experiments, three-hybrid studies have the merit that they are performed in vivo; however, they are much more restricted in the number of RNAs that they can sample (ca. 10^8 vs 10^{15}), since each RNA requires a transformant.

A final screen used a different three-hybrid system to identify RNAs that bind to Dazl protein [25]. The system exploits the interaction of REV protein with RRE RNA to tether the hybrid RNA to the promoter. More importantly, the hybrid RNA is transcribed by RNA polymerase II rather than RNA polymerase III [2]. This permits analysis of long, U-rich RNAs; on the other hand, the RNA produced is probably rarer in abundance than the polymerase III-transcribed RNAs, and may enter mRNA maturation pathways. In the polymerase II system, Dazl protein was shown to recognize interrupted oligo(U) sequences, including ones that are likely biological targets [25].

7.1. RNA "autoactivation"

A three-hybrid search to identify RNAs follows much the same logic as that described above to identify proteins. One particular class of false positives that must be eliminated consists of RNAs that are able to activate expression of the reporter genes on their own, without interacting with the protein fused to the activation domain (i.e., "protein-independent" positives) (A. Ansari, personal communication; [11]). A step must be included to classify RNAs as protein-dependent, and thus worthy of additional analysis, or protein-independent, and thus of no further interest in this context. Additionally, the protein-dependent class of RNAs must be tested with the activation domain vector and other RNA binding protein fusions to identify those RNAs specific to the protein of interest.

7.2. Preparing a library

The strategy of using genomic DNA fragments to construct the hybrid RNA library can work with smaller genomes (as demonstrated for yeast in Ref. [11]); adaptations will probably be required to move to much larger genomes. Two hundred base pairs is a typical upper limit for hybrid RNA vector inserts (see Hybrid RNAs, Limitations). Less than a million transformants of a genomic library containing 200-bp inserts adequately represents the yeast genome. The number of transformants required to attain comparable coverage in other organisms is proportional to their genomes' size, and can become prohibitive. For example, approximately 300 million inserts would be required for comparable coverage of the human genome using a 200-bp insert size. In addition, increasing the number of nontarget sequences will elevate the background and so may complicate identifying the genuine target, and termination of polymerase III at oligouridylate tracts will discriminate against certain sequences in the library. For these reasons, it may be useful to construct RNA libraries that are preenriched for candidate interacting sequences, or to circumvent the length and sequence limitations of the RNA polymerase III vector (e.g., [2,13]).

7.2.1. Step 1: Introduce the activation domain plasmid and the hybrid RNA library

1. Transform a three-hybrid strain with the plasmid expressing the RNA binding protein as an activation domain hybrid, based on the pACTII vector.
2. Transform cells from a single colony carrying this plasmid with the RNA library, selecting on media lacking leucine, uracil, and histidine, and containing 0.5 mM 3-AT. Without any 3-AT, most of the transformants will grow on a plate lacking histidine.

7.2.1.1. Construction of an RNA library. The yeast RNA library used in our experiments was constructed as follows.

Chromosomal DNA from *Saccharomyces cerevisiae* was partially digested with the following four enzymes, listed with their recognition sequences: *Mse*I (TTAA), *Tsp*509I (AATT), *Alu*I (AGCT), and *Rsa*I (GTAC). The digests were pooled and fragments in the size range 50–150 nucleotides were purified from a preparative agarose gel. The ends of the digested DNA were filled in with the Klenow fragment of DNA polymerase I where required. The plasmid pIII/MS2-2 was digested with *Sma*I, treated with calf intestine phosphatase, and ligated to the blunt-ended genomic DNA fragments. The ligations were used to transform electrocompetent HB101 *E. coli*. DNA fragments cloned at the *Sma*I site of the pIII/MS2-2 are expressed such that the RNA sequence corresponding to the yeast genomic fragment is positioned 5′ to the MS2 coat protein binding sites within the hybrid RNA. More than 1.5 million *E. coli* transformants were obtained, pooled, and used to prepare plasmid DNA for the RNA library.

7.2.2. Step 2: Screen for activation of the second reporter gene

To ensure that activation of *HIS3* is not spurious, the level of expression of the *LacZ* gene is monitored.

1. Patch individual colonies that grew in the library transformation onto plates lacking leucine and uracil.
2. Carry out filter β-galactosidase assays as described above.

7.2.3. Step 3: Eliminate protein-independent false positives

The identification of protein-independent false positives requires two successive steps. First, the activation-domain plasmid must be removed from the strain. Then the level of LacZ expression must again be determined. The colonies that are LacZ+ are cured of the activation domain plasmid. The plasmid is cured by growing a transformant overnight in YPD medium and then plating for single colonies on SD −ura to select for the RNA plasmid. These colonies are then replica-plated onto an SD −leu plate to determine which of the colonies lack the *LEU2* marker on the activation domain plasmid. Assay cells that are cured of the activation domain plasmid for β-galactosidase activity by filter assays, using the protocols above. Colonies that have lost activity upon loss of the activation domain plasmid contain possible RNA ligands of interest.

7.2.3.1. How common are protein-independent activators?. Protein-independent activators include hybrid RNAs that activate transcription when bound to a promoter. In our experiments to date, the frequency of such "activating RNAs" in a genomic library can sometimes be high. For example, 92% of all His+ LacZ+ positives obtained after selection in 0.5 mM 3-AT were protein-independent. However, at higher 3-AT concentrations only 10% were protein independent.

7.2.4. Step 4: Determine binding specificity

As in cDNA library screening, specificity tests are necessary in secondary screens to narrow down the number of candidates. Of course, the more subtle the mutation used, the better.

Transform cells from Step 3 that contain an RNA plasmid and appear to be protein-dependent for three-hybrid activity with control activation domain plasmids such as pACTII and an IRP1-activation domain fusion in pACTII. An RNA ligand that is specific for the RNA-binding protein used in the library screen should not produce a three-hybrid signal with these control plasmids.

7.2.5. Step 5: Sequence RNAs of interest

Determine the sequence of the RNA ligand and test the binding by alternative methods such as in vitro binding.

7.2.5.1. What fraction of protein-dependent activators is "correct?". In our experience with yeast Snp1 protein as "bait," we screened 2.5×10^6 transformants, and obtained 13 that were protein-dependent. Of these, the strongest by far was the appropriate segment of U1snRNA. Some of the other positives have weak sequence similarity to the relevant region of U1, and are now being analyzed further.

8. Multicomponent complexes

Multiple proteins often assemble on a single RNA to mediate the RNP's biological functions. Two proteins, P1 and P2, can interact with a single RNA in three permutations (Fig. 7). Each can be analyzed in the three-hybrid system.

8.1. Independent interactions

The two proteins interact with the RNA independently. Each protein binds to its own site, and is unaffected by the presence of the other protein. Interactions of the bI4 maturase and leucyl-tRNA synthetase with the *cob* bI4 intron are of this type, as determined in the three-hybrid system (Fig. 7A) [19].

8.2. Bridged interactions

One protein tethers the other to the RNA. A single RNA site is required for the apparent interaction of both proteins, but the binding of one protein requires the other. The interaction of She2p and She3p with the 3′UTR of *ASH1* mRNA is of this type, as shown in three-hybrid assays (Fig. 7B) [30].

8.3. Coupled interactions

The two proteins contact both the RNA and one another. In this instance, the balance among protein–protein and RNA–protein interactions determines the precise behavior in the three-hybrid system. The interaction of one protein may require prior association of an RNA–protein complex; alternatively, protein–protein interactions may be required to detect any association with the RNA. Complexes formed on *hunchback* [16], hY5 [17] and HIV-1 TAR RNAs [40,41] are of this type (Fig. 7C).

More complex RNPs, consisting of more than two proteins, also can be detected in the three-hybrid system, and are combinations of the simpler principles (Figs. 7A–C). A practical limitation in analyzing complexes containing many proteins (e.g., Fig. 7D) can be the number of plasmids and markers required.

The three-hybrid system can be permuted in many ways to analyze multiprotein complexes. One important

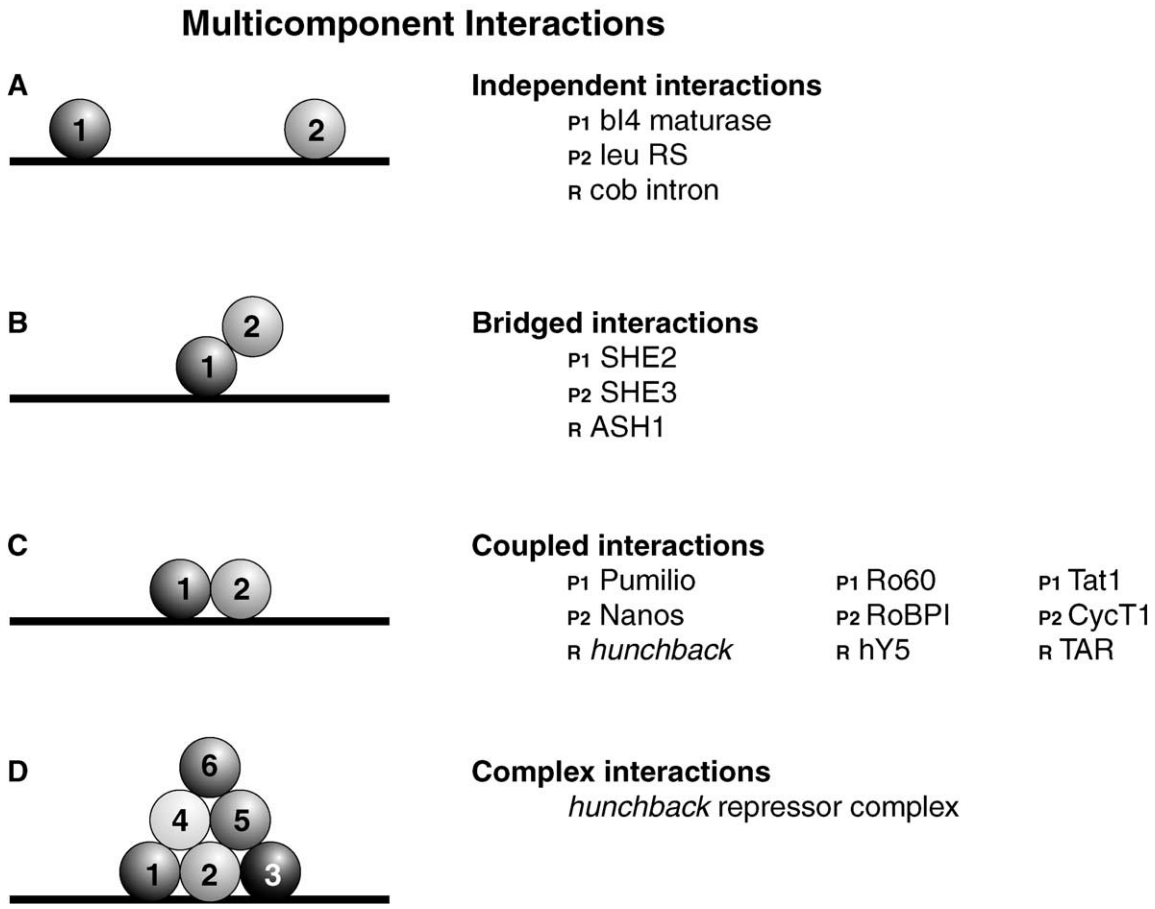

Fig. 7. Analyzing multicomponent complexes. A diagram of four permutations of complexes involving one RNA and two or more proteins. Each type of permutation has been analyzed in the three-hybrid system. Relevant citations are: (A) Ref. [19]; (B) Ref. [30]; (C) Refs. [16,17,40,41]; and (D) Ref. [42]. See text for details.

variable is the way in which the RNA is tethered. In one simple arrangement, the MS2 tether is exploited, and one or more protein components are supplied as nonhybrid molecules. In another arrangement, one component of the complex is linked to the DNA binding domain (BD), and takes the place of MS2 coat protein in the typical three-hybrid arrangement. This approach has been effective in directed tests [16,19,42] and in a screen [17,43]. In either arrangement, the results can reveal contributions of protein–protein and RNA–protein interactions. The following three examples from the literature illustrate this point.

- Tat was supplied on a multicopy plasmid as a nonhybrid protein to yeast expressing hybrid TAR-MS2 RNA and a CycT1/activation domain hybrid. In this fashion, the cyclin was found to interact with the TAR/Tat complex more avidly than with the RNA alone [40,41].
- In a three-hybrid screen, a She3p/activation domain fusion apparently interacted with yeast *ASH1* RNA [30]. Further analysis demonstrated that She3p did not bind directly to the RNA, but instead bound to endogenous yeast She2p, which bound the RNA [30].
- One protein (Nanos) interacts with an RNA (*hunchback*) when that RNA is tethered by a specific protein partner (Pumilio), but not when it is tethered by MS2 coat protein [16]. This result indicated a specific protein–protein interaction occurs in the context of the ternary complex with RNA.

Screens can be designed to identify components in complex RNPs, even when binding requires multiple proteins and the RNA. Three configurations for such screens are diagrammed in Fig. 8. Fig. 8A depicts the strategy used by Bouffard et al. [17] to find proteins that interact with Ro60 protein bound to hy5 RNA. The bait in this screen was a binding domain/pRo60 fusion bound to a nonhybrid hY5 RNA. Binding of the protein identified in the screen, RoBPI, required both hY5 RNA and Ro60 protein [17]. Fig. 8B depicts a screen searching for the fourth member of an RNP. In this case, the ''bait'' was a ternary complex containing two proteins

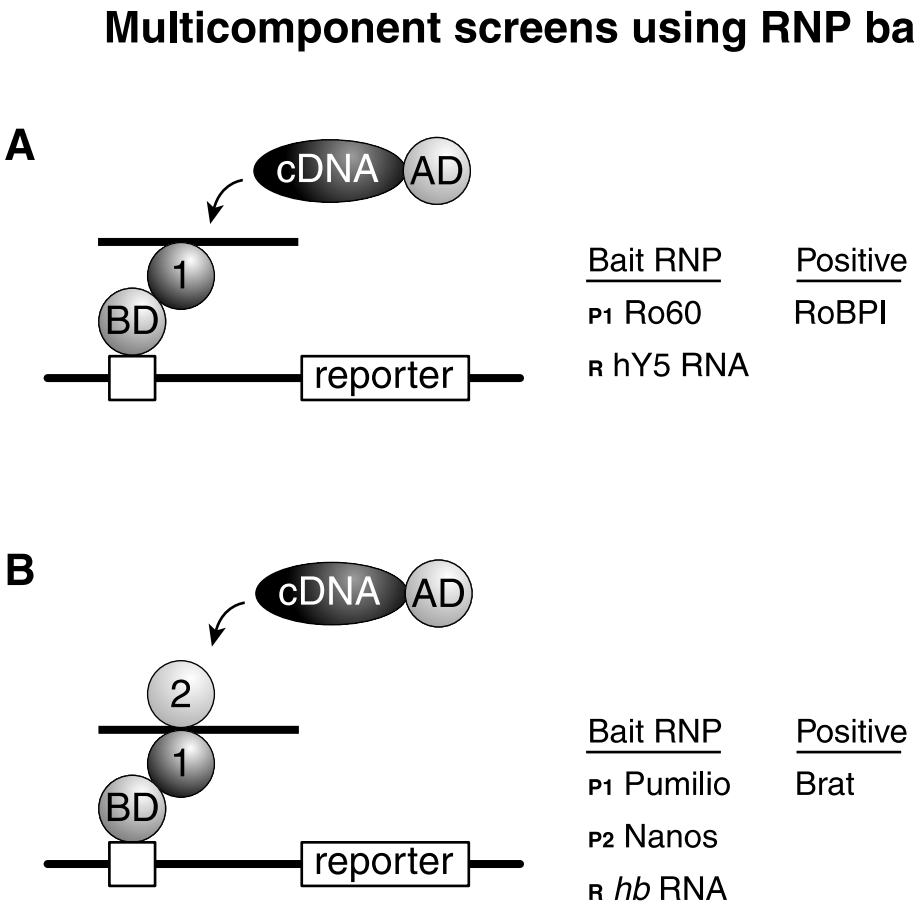

Fig. 8. Screens to identify proteins in multicomponent complexes: RNP "baits." Strategies used to look for multicomponent RNPs using the three-hybrid system. (A) A screen for proteins that interact with a protein–RNA complex. The DNA binding domain is fused to an RNA binding protein of choice (1) to tether the RNA. This was used by Bouffard et al. [17] to find proteins that interact with Ro60 protein bound to hY5 RNA. (B) A screen for a proteins that interact with a ternary protein–RNA complex. The DNA binding domain is fused to an RNA binding protein of choice (1) and a nonhybrid protein (2) is supplied to form the ternary complex with the RNA. This configuration was used by Sonoda and Wharton [42] to find proteins that interact with a ternary complex of Pumilio, Nanos, and *hunchback* RNA.

(Nanos and Pumilio) and an RNA [42]. The screen yielded a third protein, Brat, from a library of activation domain fusions [42].

9. Prospects

The simple three-hybrid system (Fig. 1) has been used to analyze a wide variety of known or suspected interactions, and has been extended to analyze complexes containing two or more proteins. This likely presages its application to even more complex situations. For example, complexes containing multiple RNAs might readily be analyzed and even more "layers" of protein likely can be penetrated to yield new components.

Further enhancements of the method may facilitate applications in several areas. Factors that enhance or prevent an RNA–protein interaction should be accessible through simple adaptations of the system. These factors include proteins that participate in or modulate the complex in vivo, as well as synthetic compounds that can be used to biological or clinical advantage. Adaptations of the system may also be used to identify RNA targets of a known RNA-binding protein, complementing biochemical approaches.

With the proliferation of sequence databases and genomic information, a small bit of protein or RNA sequence based on a three-hybrid screen may be enough to suggest whether the interaction is biologically important. It may be possible to create RNA–protein linkage maps within families of RNA-binding proteins or RNA elements, or, more ambitiously, throughout a genome.

Acknowledgments

We are grateful to the Media Laboratory of the Biochemistry Department of the University of Wisconsin for help with figures, and to Carol Pfeffer with pre-

paring the manuscript. We appreciate the helpful comments and suggestions of members of the Wickens laboratory. The laboratory is supported by research grants from the National Institutes of Health.

References

[1] D.J. SenGupta, B. Zhang, B. Kraemer, P. Pochart, S. Fields, M. Wickens, Proc. Natl. Acad. Sci. USA 93 (1996) 8496–8501.
[2] U. Putz, P. Skehel, D. Kuhl, Nucleic Acids Res. 24 (1996) 4838–4840.
[3] B.R. Cullen, Methods Enzymol. 328 (2000) 322–329.
[4] S.G. Landt, R. Tan, A.D. Frankel, Methods Enzymol. 318 (2000) 350–363.
[5] H. Peled-Zehavi, C.A. Smith, K. Harada, A.D. Frankel, Methods Enzymol. 318 (2000) 297–308.
[6] C. Jain, J.G. Belasco, Methods Enzymol. 318 (2000) 309–332.
[7] D.W. Celander, K.A. Bennett, D.E. Fouts, E.A. Seitz, H.L. True, Methods Enzymol. 318 (2000) 332–350.
[8] H. Kollmus, H. Hauser, Methods Enzymol. 318 (2000) 363–374.
[9] E. Paraskeva, M.W. Hentze, Methods Enzymol. 318 (2000) 374–384.
[10] I.A. Laird-Offringa, Methods Mol. Biol. 118 (1999) 189–216.
[11] D. SenGupta, M. Wickens, S. Fields, RNA 5 (1999) 596–601.
[12] Y. Kang, H.P. Bogerd, J. Yang, B.R. Cullen, Virology 262 (1999) 200–209.
[13] J.P. Venables, M. Ruggiu, H.J. Cooke, Nucleic Acids Res. 29 (2001) 2479–2483.
[14] C. Jain, J.G. Belasco, Cell 87 (1996) 115.
[15] K. Harada, S.S. Martin, A.D. Frankel, Nature 380 (1996) 175–179.
[16] J. Sonoda, R.P. Wharton, Genes Dev. 13 (1999) 2704–2712.
[17] P. Bouffard, E. Barbar, F. Brière, G. Boire, RNA 6 (2000) 66–78.
[18] P.D. Good, D.R. Engelke, Gene 151 (1994) 209–214.
[19] S.B. Rho, S.A. Martinis, RNA 6 (2000) 1882–1894.
[20] P.T. Lowary, O.C. Uhlenbeck, Nucleic Acids Res. 15 (1987) 10483–10493.
[21] V.J. Bardwell, M. Wickens, Nucleic Acids Res. 18 (1990) 6587–6594.
[22] G.W. Witherell, J.M. Gott, O.C. Uhlenbeck, Prog. Nucl. Acid Res. Mol. Biol. 40 (1991) 185–220.
[23] L.A. Cassiday, L.J. Maher III, Biochemistry 40 (2001) 2433–2438.
[24] F. Lim, M. Spingola, D.S. Peabody, J. Biol. Chem. 269 (1994) 9006–9010.
[25] T.A. Edwards, S.E. Pyle, R.P. Wharton, A.K. Aggarwal, Cell 105 (2001) 281–289.
[26] F. Martin, F. Michel, D. Zenklusen, B. Müller, D. Schümperli, Nucleic Acids Res. 28 (2000) 1594–1603.
[27] E.-G. Lee, A. Yeo, B. Kraemer, M. Wickens, M.L. Linial, J. Virol. 73 (1999) 6282–6292.
[28] E. Bacharach, S.P. Goff, J. Virol. 72 (1998) 6944–6949.
[29] S.W. Fewell, J.L. Woolford, Mol. Cell. Biol. 19 (1999) 826–834.
[30] R.M. Long, W. Gu, E. Lorimer, R.H. Singer, P. Chartrand, EMBO J. 19 (2000) 6592–6601.
[31] L. Breeden, K. Nasmyth, Cold Spring Harbor Symp. Quant. Biol. 50 (1985) 643.
[32] J.H. Miller, Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1972.
[33] K. Iwabuchi, B. Li, P. Bartel, S. Fields, Oncogene 8 (1993) 1693–1696.
[34] H.G. Davies, F. Giorgini, M.A. Fajardo, R.E. Braun, Dev. Biol. 221 (2000) 87–100.
[35] Y.W. Park, S.L. Tan, M.G. Katze, BioTechniques 26 (1999) 1102–1106.
[36] B. Zhang, M. Gallegos, A. Puoti, E. Durkin, S. Fields, J. Kimble, M. Wickens, Nature 390 (1997) 477–484.
[37] K. Robzyk, Y. Kassir, Nucleic Acids Res. 20 (1992) 3790.
[38] C. Tuerk, L. Gold, Science 3 (1990) 505–510.
[39] T. Shtatland, S.C. Gill, B.E. Javornik, H.E. Johansson, B.S. Singer, O.C. Uhlenbeck, D.A. Zichi, L. Gold, Nucleic Acids Res. 28 (2000) E93.
[40] P.D. Bieniasz, T.A. Grdina, H.P. Bogerd, B.R. Cullen, EMBO J. 17 (1998) 7056–7065.
[41] P.D. Bieniasz, T.A. Grdina, H.P. Bogerd, B.R. Cullen, J. Virol. 73 (1999) 5777–5786.
[42] J. Sonoda, R.P. Wharton, Genes Dev. 15 (2001) 762–773.
[43] P. Bouffard, F. Brière, R.J. Wellinger, G. Boire, BioTechniques 27 (1999) 790–796.
[44] P.J. Good, Q. Chen, S.J. Warner, D.C. Herring, J. Biol. Chem. 275 (2000) 28583–28592.
[45] N. Takahashi, N. Sasagawa, K. Suzuki, S. Ishiura, Biochem. Biophys. Res. Commun. 277 (2000) 518–523.
[46] T. Tadauchi, K. Matsumoto, I. Herskowitz, K. Irie, EMBO J. 20 (2001) 552–561.
[47] W.A. Ziehler, J. Morris, F.H. Scott, C. Millikin, D.R. Engelke, RNA 7 (2001) 565–575.
[48] N. Matter, M. Marx, S. Weg-Remers, H. Ponta, P. Herrlich, H. König, J. Biol. Chem. 275 (2000) 35353–35360.
[49] G. Fabini, R. Raijmakers, S. Hayer, M.A. Fouraux, G.J.M. Pruijn, G. Steiner, J. Biol. Chem. 276 (2001) 20711–20718.
[50] T. Jiang, C. Guerrier-Takada, S. Altman, RNA 7 (2001) 937–941.
[51] S. Crucs, S. Chatterjee, E.R. Gavis, Mol. Cell 5 (2000) 457–467.
[52] V.A. Kickhoefer, A.G. Stephen, L. Harrington, M.O. Robinson, L.H. Rome, J. Biol. Chem. 274 (1999) 32712–32717.
[53] L. Harrington, T. McPhail, V. Mar, W. Zhou, R. Oulton, M.B. Bass, I. Arruda, M.O. Robinson, Science 275 (1997) 973–977.
[54] N. Ismaïli, D. Pérez-Morga, P. Walsh, M. Cadogan, A. Pays, P. Tebabi, E. Pays, Mol. Biochem. Parasitol. 106 (2000) 109–120.
[55] O. Guerra-Peraza, M. De Tapia, T. Hohn, M. Hemmings-Mieszczak, J. Virol. 74 (2000) 2067–2072.
[56] H.P. Bogerd, H.L. Wiegand, J. Yang, B.R. Cullen, J. Virol. 74 (2000) 9353–9361.
[57] J. Yang, H.P. Bogerd, S.-Y. Le, B.R. Cullen, RNA 6 (2000) 1551–1564.
[58] S.-Y. Le, R. Sternglanz, C.W. Greider, Mol. Biol. Cell 11 (2000) 999–1010.
[59] A. Poleev, A. Hartmann, S. Stamm, Eur. J. Biochem. 267 (2000) 4002–4010.
[60] P. Xu, L. Wen, G. Benegal, X. Wang, G.A. Buck, Mol. Biochem. Parasitol. 112 (2001) 39–49.
[61] J. Wood, R.M. Frederickson, S. Fields, A.H. Patel, J. Virol. 75 (2001) 1348–1358.
[62] E. Jan, C.K. Motzny, L.E. Graves, E.B. Goodwin, EMBO J. 18 (1999) 258–269.
[63] Z.F. Wang, M.L. Whitfield, T.C.I. Ingledue, Z. Dominski, W.F. Marzluff, Genes Dev. 10 (1996) 3028–3040.
[64] Y.W. Park, J. Wilusz, M.G. Katze, Proc. Natl. Acad. Sci. USA 96 (1999) 6694–6699.
[65] A. Dahanukar, J.A. Walker, R.P. Wharton, Mol. Cell 4 (1999) 209–218.
[66] F. Martin, A. Schaller, S. Eglite, D. Schumperli, B. Muller, EMBO J. 16 (1997) 769–778.