Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 8496–8501, August 1996
Genetics

A three-hybrid system to detect RNA–protein interactions *in vivo*

DHRUBA J. SENGUPTA*†, BEILIN ZHANG‡, BRIAN KRAEMER‡, PASCALE POCHART*, STANLEY FIELDS*†, AND MARVIN WICKENS‡§

*Department of Molecular Genetics and Microbiology, State University of New York, Stony Brook, NY 11794; †Departments of Genetics and Medicine, Markey Molecular Medicine Center, University of Washington, Box 357360, Seattle, WA 98195; and ‡Department of Biochemistry, 420 Henry Mall, University of Wisconsin, Madison, WI 53706

Communicated by Larry Gold, NeXstar Pharmaceuticals, Boulder, CO, March 25, 1996 (received for review February 5, 1996)

ABSTRACT **RNA–protein interactions are pivotal in fundamental cellular processes such as translation, mRNA processing, early development, and infection by RNA viruses. However, in spite of the central importance of these interactions, few approaches are available to analyze them rapidly *in vivo*. We describe a yeast genetic method to detect and analyze RNA–protein interactions in which the binding of a bifunctional RNA to each of two hybrid proteins activates transcription of a reporter gene *in vivo*. We demonstrate that this three-hybrid system enables the rapid, phenotypic detection of specific RNA–protein interactions. As examples, we use the binding of the iron regulatory protein 1 (IRP1) to the iron response element (IRE), and of HIV trans-activator protein (Tat) to the HIV trans-activation response element (TAR) RNA sequence. The three-hybrid assay we describe relies only on the physical properties of the RNA and protein, and not on their natural biological activities; as a result, it may have broad application in the identification of RNA-binding proteins and RNAs, as well as in the detailed analysis of their interactions.**

A broad range of critical and unsolved biological problems converge on the specific binding of a protein to its RNA target. For example, the mechanisms and regulation of mRNA processing and translation rely on RNA–protein interactions to assemble the catalytic machineries involved and to interact with the RNA substrate. Similarly, replication of chromosome ends hinges on the assembly and activity of telomerase, an RNA–protein complex. Key decisions during early development rely on specific RNA–protein interactions to regulate the activity, stability and cellular localization of maternal mRNAs. RNA viruses, such as HIV and picornaviruses, exploit RNA–protein interactions to regulate infectivity and replication. Indeed, the interactions of viral trans-activator proteins, such as HIV Tat, with their RNA targets have been intensively investigated as targets for therapeutics.

Several systems have been devised to detect a range of RNA–protein interactions. For example, RNA–protein interactions can be assayed *in vivo* by placing an RNA binding site in an mRNA such that, when bound to a cognate protein, translation is repressed (1, 38). Methods based on phage display (2) and the antitermination properties of N protein in bacteria (39) facilitate analysis of interactions between RNA and proteins or peptides and the identification of specificity determinants. *In vitro* selection procedures can be used to identify RNAs that bind with high affinity to a protein of interest and to reveal those features of the RNA that are critical (3, 4).

In this report, we describe a genetic assay in which specific RNA–protein interactions can be detected rapidly in yeast, in a fashion that is independent of the biological role of the RNA or protein. Our approach is based on the yeast two-hybrid system (5, 6) which detects protein–protein interactions. The three-hybrid system presented here allows simple phenotypic properties of yeast, such as the ability to grow or to metabolize a chromogenic compound, to be used to detect and analyze an RNA–protein interaction.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Plasmid Constructions and Nomenclature. Plasmid nomenclature is as follows. Each plasmid is named by the protein or RNA it encodes, with designations that reflect the order (N to C terminal, or 5′ to 3′) of the components involved. Thus plasmids encoding hybrid protein 1 (the hybrid protein containing a DNA-binding domain linked to RNA-binding domain 1; see Fig. 2*A*) are named as follows. The LexA–MS2 hybrid, in which the LexA DNA binding domain is N terminal and the MS2 coat protein C terminal, is encoded by pLexA–MS2; the Gal4 DNA binding domain–MS2 coat protein hybrid is encoded by pGal4–MS2; LexA alone is encoded by pBTM116 (7). Plasmids encoding hybrid RNAs are based on the multi-copy vector pIIIEx426RPR (8), and so each is designated with an pIII prefix. Thus MS2–IRE RNA, in which the MS2 sites are 5′ of the iron response element (IRE), is encoded by pIII/MS2–IRE; IRE–MS2 RNA is encoded by pIII/IRE–MS2; IRE–TAR RNA by pIII/IRE–TAR; MS2 RNA by pIII/MS2; TAR–MS2 RNA by pIII/TAR–MS2 (TAR, trans-activation response element). In plasmids expressing hybrid protein 2 (the hybrid protein containing an activation domain linked to RNA-binding domain 2; see Fig. 2*A*), the activation domain (AD) is invariably N terminal. The AD–IRP1 hybrid is encoded by pAD–IRP1; the AD protein is encoded by pACTII (ref. 9; S. Elledge, personal communication); AD–Rev protein by pAD–Rev; IRP1 by pIRP1 (IRP1, iron regulatory protein 1).

Hybrid Protein 1 Family. The MS2 coat protein gene was amplified by PCR from the plasmid pKCO, an overexpression plasmid for coat protein similar to pTCT5 (10) using the primers 5′-CAGGTGGATCCATATGGCTTCTAACTT-TACT-3′ and 5′-TGCTAGGATCCTTAGTAGATGCCG-GAGTT-3′. The product was digested with *Bam*HI and ligated into the vector pBTM116 (7) to generate plasmid pLexA–MS2. The MS2 coat protein gene was released from pLexA–MS2 by *Bam*HI digestion, and inserted into the *Bam*HI site of pGBT9 (11) to generate pGal4–MS2.

Hybrid RNA Family. We have created four plasmids for general use, pMS2-1, pMS2-2, pIII/MS2-1, and pIII/MS2-2 (diagrammed in Fig. 1). They provide two alternative routes for preparing identical plasmids encoding identical hybrid RNAs. Using pIII/MS2-1 and pIII/MS2-2, the plasmid encoding the hybrid RNA of interest can be generated in a single

Abbreviations: TAR, trans-activation response element; IRE, iron response element; AD, activation domain; IRP1, iron regulatory protein 1; β-gal, β-galactosidase.

§To whom reprint requests should be addressed. e-mail: wickens@biochem.wisc.edu.

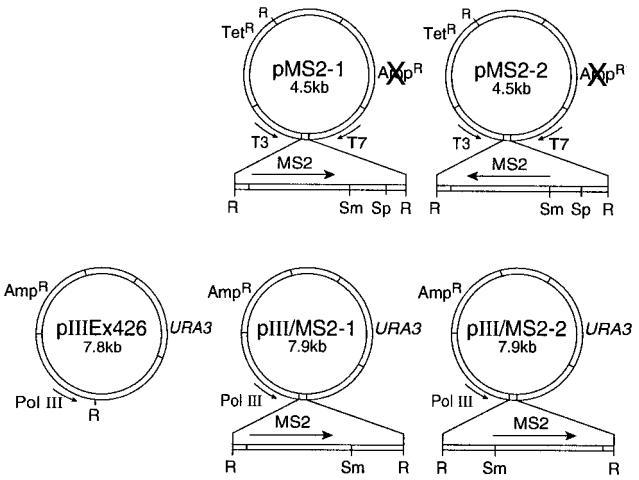

FIG. 1. Schematic diagram of plasmids used to generate hybrid RNAs. Arrows labeled MS2, two tandem MS2 coat protein binding sites, as described in the text and depicted in Fig. 2 (polarity of arrow is 5′ to 3′ in the sense-strand); T7 and T3, bacteriophage T7 and T3 RNA polymerase promoters, respectively (arrows indicate direction of transcription); pol III, RNA polymerase III promoter, including RNase P RNA sequences (arrows indicate direction of transcription); Tet^R, gene encoding tetracycline resistance in *Escherichia coli;* Amp^R, gene encoding ampicillin resistance in *E. coli* (eliminated by a small deletion in pMS2-1 and pMS2-2); *URA3,* the yeast *URA3* gene. pIIIEx426 designates pIIIEx426RPR, constructed by Good and Engelke (8). Restriction sites are abbreviated as follows. R, *Eco*RI; Sm, *Sma*I; Sp, *Spe*I. In pMS2-1 and pMS2-2, the *Sma*I and *Spe*I sites are unique, and can be used to insert a sequence of interest; in pIII/MS2-1 and pIII/MS2-2, only the *Sma*I site is unique. Origins of replication are not depicted. pIII/MS2-1, pIII/MS2-2, and pIIIEx426RPR are multicopy yeast plasmids. pMS2-1 and pMS2-2 are bacterial plasmids and do not propagate in yeast.

step; using pMS2-1 and pMS2-2, two steps are required to create essentially the same plasmid. However, the use of pMS2-1 and pMS2-2 may simplify screening for the desired clones.

pIII/MS2-1 and pIII/MS2-2 are yeast/*E. coli* shuttle vectors in which an RNA polymerase III promoter directs transcription of an RNA containing two tandem MS2 sites. A unique *Sma*I restriction site can be used to insert a sequence to be tested in the three-hybrid assay. This is the only unique insertion site for such sequences. In pIII/MS2-1, the *Sma*I site is 3′ of the tandem MS2 coat protein binding sites; in pIII/MS2-2, the *Sma*I site is 5′ of the MS2 sites. Both plasmids are multi-copy in yeast and carry a yeast *URA3* marker. In *E. coli*, the plasmids confer ampicillin resistance.

pMS2-1, pMS2-2, and PIIIEx426RPR (8) provide an alternative means of generating hybrid RNA plasmids. This cloning route requires two steps to reach the same end as is achieved in a single step using pIII/MS2-1 and pIII/MS2-2, but may simplify screening and cloning manipulations. pMS2-1 and pMS2-2 confer tetracycline resistance in *E. coli* and do not carry an RNA polymerase III promoter or any other yeast sequences. They carry two tandem MS2 coat protein binding sites flanked by bacteriophage T7 and T3 RNA polymerase promoters. Unique *Sma*I and *Spe*I sites can be used to insert a sequence of interest either 5′ (pMS2-2) or 3′ of MS2 sequences (pMS2-1). The sequence of any inserted DNA can be determined by using T7 and T3 promoter-specific primers. The region of these plasmids encoding MS2 coat protein binding sites (and any inserted additional sequences) can be excised as an *Eco*RI fragment. Thus, after inserting sequences of interest into pMS2-1 or pMS2-2, an *Eco*RI fragment can be excised and inserted into the unique *Eco*RI site of the vector, pIIIEx426RPR (ref. 8; Fig. 1) to generate a plasmid that can be used to produce the desired hybrid RNA in yeast. [pIIIEx426RPR, constructed by Good and Engelke (8), carries an RNA polymerase III promoter immediately adjacent to a unique *Eco*RI insertion site.] Such plasmids will carry the yeast *URA3* gene and confer ampicillin resistance in *E. coli*, and be identical to those generated using the pIII/MS2-1 and pIII/MS2-2 vectors.

Plasmids encoding hybrid RNAs used in this report were constructed as follows. A *Bam*HI–*Hin*dIII fragment from pVB536 (12), containing two tandem MS2 coat protein binding sites, was "filled-in" and cloned into the "blunted" ends of *Pst*I-cleaved pTET2. pTET2, a derivative of pBluescript IIKS(+), carries a Tet^R gene from YIp5 and has had its Amp^R inactivated. In pTET2, the *Xba*I site in the polylinker of pBluescript IIKS(+) has been changed to an *Eco*RI site, so that the DNA fragment encoding two MS2 sites and the desired other RNA sequence can be liberated as an *Eco*RI fragment and cloned into the pRPR vectors. The MS2 sites from pMS2-2 have also been cloned into pIIIEx426RPR (8) to generate pIII/MS2-2. RNA sequences of interest can be cloned directly into a unique *Sma*I site adjacent to MS2 sites to generate desired hybrid RNAs of the general form, RNAX–MS2. To fuse the IRE with MS2 sites, an *Ava*I fragment containing the IRE from rat ferritin light chain (derived from p16Bgl; ref. 13) was cloned into the *Sma*I site of pMS2-1 or pMS2-2, generating pMS2–IRE and pIRE–MS2. The plasmids were then cut with *Eco*RI and the appropriate fragments were cloned into the *Eco*RI site of pIIIEx426RPR (8). The resulting plasmids were designated pIII/MS2–IRE and pIII/IRE–MS2. In pIII/IRE–MS2, the RNase P promoter directs the synthesis of a predicted 320-nt transcript containing, from 5′ to 3′, RNase P leader sequence (84 nt), linker (28 nt), IRE (51 nt), linker (32 nt), MS2 binding sites (60 nt), linker (24 nt), and RNase P 3′ terminal sequence (41 nt). The HIV-1 TAR element was prepared by annealing the oligonucleotides 5′-CCCGGGTCTCTCTGGTTAGACCAGATCT-GAGCCTGGGAGCTC-3′ and 5′-ATCGGGTTCCCTAGT-TAGCCAGAGAGCTCCCAGGCTCAGATCT-3′ and filling in the overhanging 5′ ends. After suitable manipulations (details available upon request), this DNA segment was cloned upstream of two MS2 sites, generating pIII/TAR–MS2. In pIII/TAR–MS2, the RNase P promoter drives synthesis of a predicted 316-nt transcript containing, from 5′ to 3′, 84 nt of the leader sequence, 14 nt of linker sequence, 58 nt of TAR, 32 nt of linker region, 60 nt of the MS2 recognition sequence, 27 nt of linker region, and 41 nt of the 3′ terminus of RNase P RNA. In RAT–MS2 RNA, the TAR segment is present in the opposite orientation (i.e., antisense TAR).

Hybrid Protein 2 Family. The IRP1 gene was inserted into pACTII as follows. Plasmid pFRP32 (14), which carries the rabbit IRP1 gene, was cleaved with *Xba*I, the termini filled-in, and the fragment encoding IRP1 released from the vector by *Bsp*HI digestion. This IRP1 fragment was then inserted between the *Nco*I and *Sma*I sites of pACTII to generate pAD–IRP1. To create pIRP1, the IRP1 gene was released from plasmid pFRP32 (14) by cleavage with *Xba*I and *Bsp*HI, filled-in, and ligated to the filled-in *Nhe*I site of an *ADH1* promoter/terminator cassette (15) inserted into YEplac181 (16). pAD–Rev was created as follows. The HIV-1 Rev protein gene was amplified by PCR from the plasmid pDM121 (17) using the primers 5′-AGGCCCCGGGTATGGCAGGAA-GAAGCGGAGACAGC-3′ and 5′-AGTCCCCGGGCA-GACGGGCACACACTACTTGAAGC-3′. The PCR product was cleaved with *Sma*I and ligated into the vector, pACTII (ref. 9; S. Elledge, personal communication). To generate pAD–Tat, a portion of the HIV Tat protein (corresponding to the first exon) was amplified by PCR from the plasmid pBC12/CMV/t2 (18) using the primers 5′-GTCGGGATCCTAATG-GAGCCAGTAGATCCT-3′ and 5′-GTGACGGATCCT-TACTGCTTTGATAGAGAAAC-3′. The PCR product was

digested with *Bam*HI and ligated to the vector, pACT (19), which had been digested with *Bgl*II .

Yeast Strain Construction. A strain carrying an integrated copy of the LexA–MS2 coat protein fusion was created as follows. The *Sph*I fragment of pLexA–MS2 carrying the LexA–coat protein gene was used to replace the corresponding fragment of pGBT9.C, a Gal4 DNA-binding domain plasmid constructed in the vector Ycplac22 (16). The resultant plasmid was cut with *Spe*I and *Nhe*I and religated, to delete yeast replication elements, yielding the plasmid pLexA–coat. This plasmid was integrated into L40-*ura3* (*MAT***a**, *ura3–52*, *leu2–3112*, *his3*Δ*200, trp1*Δ*1, ade2, LYS2*::(lexAop)-*HIS3*, *ura3*::(lexAop)-*LacZ*; generous gift of T. Triolo and R. Sternglanz) to generate the yeast strain, L40-coat.

Yeast Methods. Yeast triple transformants were assayed for β-galactosidase (β-gal) activity either by restreaking onto plates containing 5-bromo-4-chloro-3-indolyl β-D-galactoside (colony color assay), or by direct measurement of enzyme activity using chlorophenol red β-D-galactopyranoside (20) as a substrate. Enzymatic assays represent the average of three to six independent transformants and were repeated in two to six independent experiments. Units of β-gal activity are defined as in ref. 21.

For direct selection, combinations of three plasmids were transformed into L40–*ura3*, in which the *HIS3* gene is under the control of LexA binding sites (as is the *lacZ* gene). Four independent transformants from each combination of plasmids were first streaked onto plates lacking tryptophan, leucine, and uracil; after two days the patches were replica-plated onto plates lacking histidine, tryptophan, leucine, and uracil containing 20 mM 3-aminotriazole.

RESULTS

Outline of the Method. The two-hybrid assay relies (*i*) on the ability of a protein such as the bacterial repressor LexA to tether another protein to DNA (22), and (*ii*) on a protein–protein interaction to bring a transcriptional activation domain hybrid into close proximity with the DNA-binding domain hybrid. We asked whether RNA–protein interactions could reconstitute transcription in an analogous fashion. This three-hybrid assay (diagrammed in schematic fashion in Fig. 2*A*) uses a known RNA-binding protein to tether an RNA containing the cognate binding site. As the known RNA-binding protein is fused to LexA, this hybrid protein should bring the RNA to which it is bound to a reporter gene regulated by LexA binding sites. In addition, the RNA is bifunctional (called here a "hybrid RNA") in that it also contains a binding site for a second RNA-binding protein. This second RNA-binding protein is present as a fusion to a transcriptional activation domain. Thus, the interaction of the RNA with both RNA-binding domains should result in the activation domain being present at the promoter of the LexA-regulated gene, and so should lead to transcriptional activation of the reporter. For the first RNA-binding protein, we used the coat protein of bacteriophage MS2. The MS2 coat protein, like the nearly identical protein from bacteriophage R17, recognizes a 21-nt RNA stem–loop in its genome with high affinity (25). The coat protein was joined to the bacterial DNA-binding protein, LexA, and expressed from a vector that also carries the yeast selectable gene *TRP1* (7). The LexA–coat protein hybrid would be held constant; hybrid protein 2 and the portion of the hybrid RNA with which it interacts would vary between experiments.

The details of the method are described using, as an example, the interaction of IRP1 with its RNA target, the IRE. The IRE is a stem–loop structure found in the untranslated regions of mRNAs encoding certain proteins involved in iron utilization, and binds specifically to IRP1 (26, 27). To analyze the IRE/IRP1 interaction in the three-hybrid system, we generated two additional plasmids. One is designed to produce

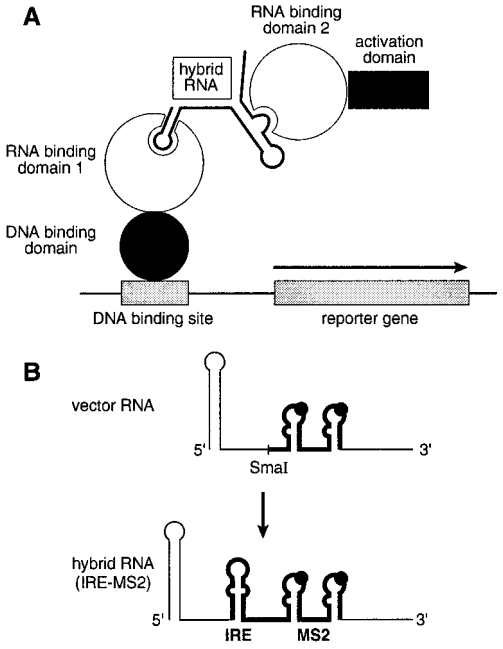

FIG. 2. Experimental approach. (*A*) Overview of the three-hybrid strategy to detect RNA–protein interactions. A hybrid protein containing a DNA-binding domain (e.g., LexA) with RNA-binding domain 1 (e.g., MS2 coat protein) localizes to the promoter of an appropriate reporter gene. A second hybrid protein containing a transcriptional activation domain (e.g., from Gal4) with RNA-binding domain 2 (e.g., IRP1) will activate transcription of the reporter gene when in close proximity to the gene's upstream regulatory sequences. A hybrid RNA containing sites recognized by the two RNA-binding proteins links the two hybrid proteins to one another, and the tripartite complex results in detectable expression of the reporter gene. For simplicity, the following details of the typical experimental design are not depicted in *A*. Multiple LexA binding sites are present in the promoter. Both MS2 coat protein and LexA bind to their target sites as dimers. Thus multiple copies of the hybrid RNA, each one of which contains two potential binding sites for MS2, may be bound simultaneously to the promoter. (*B*) Schematic structure of the hybrid RNAs. The hybrid RNAs are expressed from an RNase P promoter, using RNA polymerase III. They retain the 5′ stem–loop structure(s) and 3′ end of RNase P RNA, but are missing its internal portion. To facilitate cooperative binding to MS2 coat protein, the coat protein binding site is present in two copies (12, 23), each of which contains a point mutation (indicated by a black dot) that enhances this RNA–protein interaction (24). In the example shown, the stem–loop of the IRE has been inserted into the hybrid RNA vector, pIII/MS2-2, generating a plasmid encoding IRE–MS2 hybrid RNA. Thin line represents RNase P RNA sequences; bold lines represent IRE or MS2 binding site sequences; black dot represent point mutation in the loop of the MS2 binding site.

a fusion of IRP1 to the Gal4 Activation Domain, and was derived from the vector, pACTII (ref. 9; S. Elledge, personal communication) carrying the *LEU2* gene. The other plasmid is designed to encode the hybrid RNA, comprising two copies of the MS2 coat protein binding site and a single IRE. Two coat protein binding sites were used because binding to adjacent sites is cooperative (12, 23). Furthermore, a variant site, containing a single base change, was used because it substantially enhances binding to coat protein *in vitro* (24). The hybrid RNA was expressed from the vector pIIIEx426RPR (8), which uses the RNA polymerase III promoter and terminator from the *Saccharomyces cerevisiae* RNase P RNA gene (*RPR1*) to generate high levels of small RNAs in yeast. Such RNAs presumably do not enter pre-mRNA processing pathways. The plasmid encoding the hybrid RNA is multi-copy and carries the selectable gene, *URA3*. The RNA it encodes, diagrammed in Fig. 2*B*, contains a predicted 5′-terminal stem–loop of RNase

P RNA, followed by a single IRE, two MS2 sites, and the 3′ end of RNase P RNA.

In principle, multiple hybrid RNA molecules may be tethered simultaneously to a single promoter. This is possible because multiple LexA binding sites are present in the promoter, and because both LexA and MS2 proteins bind to their nucleic acid targets as dimers.

The IRE/IRP1 Interaction: Each Hybrid Component Is Essential. We introduced combinations of the three plasmids described above, as well as appropriate control plasmids, into the yeast strain, L40–*ura3,* a uracil auxotroph derived from the reporter strain L40 (28). L40–*ura3* contains a *lacZ* gene whose expression is regulated by LexA binding sites in the 5′ flanking sequence. The strain was transformed and selected for tryptophan, leucine, and uracil prototrophy. Transformants were assayed for *lacZ* expression by a colony color assay and by liquid assay using chlorophenol red β-D-galactopyranoside as substrate (21). As shown in Fig. 3, transformants carrying the LexA–MS2 coat protein and activation domain–IRP1 hybrids along with the hybrid RNA showed readily detectable β-gal activity in both assays (rows 1 and 2). Activity was detected with the IRE and MS2 RNA sites in either orientation with respect to one another (row 1 versus row 2). In the absence of any one of the hybrid components (rows 3–10), transformants displayed little or no activity, indicating that the hybrid RNA must be capable of binding simultaneously to both hybrid proteins, and that the resultant RNA–protein complex can trigger transcription. These controls (rows 3–10) also show the following. (*i*) Transcription is regulated by the LexA binding

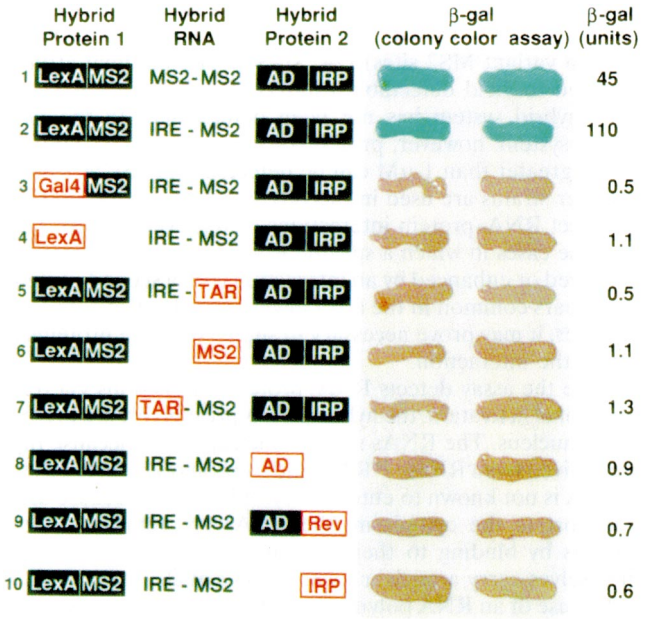

FIG. 3. β-gal activity in yeast transformants. Three plasmids, encoding the proteins and RNAs indicated at the left of the figure, were introduced into the yeast strain L40–*ura3* by transformation, plating on media lacking tryptophan, leucine, and uracil. Yeast triple transformants were assayed for β-gal activity either by the colony color assay or by direct measurement of enzyme activity. Duplicates of the colony color assay are shown. Units of β-gal activity are defined as in ref. 21. The two components present in each hybrid factor are depicted in the proper polarity, either N to C terminal, or 5′ to 3′. Proteins are indicated with black boxes, each box corresponding to a domain. Red boxes (rows 3–10) indicate components that differ from those present in the complete system (rows 1 and 2). MS2 in black ball, MS2 coat protein; IRE, iron response element; MS2, two MS2 RNA binding sites; TAR, stem–loop at 5′ end of HIV mRNAs (Tat binding site); IRP1, iron response protein [also called iron response element binding protein (IRE-BP) and iron response factor (IRF)]; AD, the activation domain of the Gal4 transcription factor of *S. cerevisiae;* Rev, an HIV transactivator.

sites, since a hybrid of the Gal4 DNA-binding domain and the MS2 coat protein was inactive (row 3). Similarly, *lacZ* was not activated by the combination of plasmids shown in row 2 when the DNA-binding site upstream of *lacZ* was that of Gal4 instead of LexA (not shown). (*ii*) Specific recognition of the MS2 coat protein binding sites is required, as elimination of either the MS2 coat protein (row 4) or its RNA binding sites (row 5) severely reduced transcriptional activation. (*iii*) Specific recognition of the IRE by IRP1 is required, as an RNA containing only the MS2 sites was inactive (row 6), as was a hybrid RNA containing MS2 sites fused to another RNA sequence, HIV TAR (row 7). IRP1 is essential, since either the activation domain alone (row 8) or the activation domain fused to HIV Rev, a different RNA binding domain (row 9), failed to activate. Similarly, the activation domain itself is required, since IRP1 alone was inactive (row 10). (*iv*) Since introduction of the LexA–coat protein hybrid, in the absence of either one of the other two required components, did not lead to significant β-galactosidase activity (rows 5–10), the coat protein does not itself possess a transcriptional activation domain. (*v*) Since the coat protein and IRP1 hybrids present with control RNAs did not lead to transcription (rows 5–7), coat protein and IRP1, as expected, do not bind to each other to result in transcription through a protein–protein interaction. Similarly, the 125 bases of RNase P RNA that are present in all of the hybrid RNAs, corresponding to their 5′ and 3′ termini, do not mediate interactions with the hybrid proteins. In sum, these data establish that RNA–protein interactions can be detected using this three-hybrid system, and that each interaction diagrammed in Fig. 2*A* is required.

The TAR/Tat Interaction and Specificity of the Assay. For the method to be of general utility, other RNA–protein interactions should also cause transcriptional activation, and the interactions should show sequence specificity. Toward this end, we tested a second RNA–protein interaction, the binding of the HIV-1 Tat protein to its RNA target, TAR. Binding of Tat to TAR, which comprises the first 59 nt of all HIV-1 transcripts, greatly enhances the viral infectivity and the production of viral RNA in infected cells (29). To test this interaction in the three-hybrid assay, we prepared a hybrid RNA containing MS2 sites and TAR, and a hybrid protein containing Tat (amino acid residues 1–72) linked to the activation domain. As shown in Table 1 (lines 1 to 4), the Tat/TAR interaction is readily detectable: β-gal production was stimulated when the coat protein and Tat hybrids were present along with a hybrid RNA possessing the cognate binding sites. Elimination of TAR (Table 1, line 2) or Tat (line 3) dramatically reduced stimulation, as did replacement of TAR with the anti-sense TAR sequence, RAT (line 4).

The sequence specificity of the Tat/TAR and IRP1/IRE interactions was demonstrated by examining matched and

Table 1. Tat/TAR interaction and specificity

Hybrid protein 1	Hybrid RNA	Hybrid protein 2	β-gal activity, units
1. LexA–MS2 coat	TAR–MS2	AD–Tat	30
2. LexA–MS2 coat	MS2	AD–Tat	0.4
3. LexA–MS2 coat	TAR–MS2	AD	0.5
4. LexA–MS2 coat	RAT–MS2	AD–Tat	0.4
5. LexA–MS2 coat	IRE–MS2	AD–Tat	0.4
6. LexA–MS2 coat	TAR–MS2	AD–IRP1	0.4
7. LexA–MS2 coat	IRE–MS2	AD–IRP1	110

Three plasmids, encoding the proteins and RNAs indicated, were introduced into the yeast strain L40–*ura3* by transformation, plating on media lacking tryptophan, leucine, and uracil. The β-gal activity of transformants was then determined using chlorophenol red β-D-galactopyranoside (17) as a substrate. Three to six transformants were assayed for each sample, and the experiment was repeated four times. A representative data set is shown.

mismatched combinations of RNA and protein. The activation domain–Tat hybrid led to transcription only in combination with TAR, not with an IRE (Table 1, line 1 versus line 5); similarly, the activation domain–IRP1 hybrid required an IRE, and did not activate with TAR (Table 1, line 6 versus line 7).

Direct Selection for an RNA–Protein Interaction and Construction of a General Yeast Reporter Strain. To test whether a selection could be used to detect an RNA–protein complex, we examined the Tat/TAR interaction using as a reporter the *HIS3* gene, which also is under the control of LexA binding sites in strain L40–*ura3* (Fig. 4). The growth of four independent transformants carrying each of five combinations of hybrid molecules was analyzed by plating on selective media containing 3-amino-1,2,4-triazole (3-AT). 3-AT, a competitive inhibitor of the enzyme encoded by *HIS3*, allows the growth of only those cells that express the HIS3 gene product at elevated levels. As shown in Fig. 4, only those cells possessing a functional Tat/TAR combination (row 1) grow significantly. Comparable results were obtained with the IRP1/IRE interaction (not shown), using the set of hybrid molecules depicted in Fig. 3. Because the strain used in Fig. 4 also carries the *lacZ* gene under the control of LexA binding sites, RNA–protein interactions detected by growth on selective media can be additionally assayed for β-gal activity.

To facilitate application of the three-hybrid strategy, we integrated the gene encoding the LexA–MS2 coat protein hybrid (which typically would not vary between experiments) into a chromosome in strain L40-*ura3*. Transformation of the new strain (L40–coat) with the appropriate matched plasmids for the IRP1/IRE interaction produced comparable β-gal activity as when the coat protein hybrid was expressed from a plasmid (Table 2). As a result, transformation of this strain with two plasmids is sufficient to detect or screen for an RNA–protein interaction.

DISCUSSION

The three-hybrid system described here provides a rapid and potentially versatile method to detect RNA–protein interactions *in vivo.* While this assay possesses many of the features of the two-hybrid system for the analysis of protein–protein interactions, hybrid RNAs differ from hybrid proteins in significant respects. The structures formed by these RNAs may often not be physiological, since the same RNA sequence placed into different contexts can adopt different conformations. However, our results suggest that local, relatively stable structures, such as those recognized by MS2 coat protein, IRP1 and Tat, form in unnatural contexts so as to be recognized by their cognate proteins, and can coexist in a single RNA molecule *in vivo*. The stem–loop structures of these RNA elements are relatively stable: ostensibly "unstructured" RNAs might be more prone to the formation of unproductive, alternative conformations, and could in principle limit the range of RNA–protein interactions that can be assayed. However, the formation of correct structures in only a fraction of the RNA molecules likely is sufficient to lead to transcriptional activation.

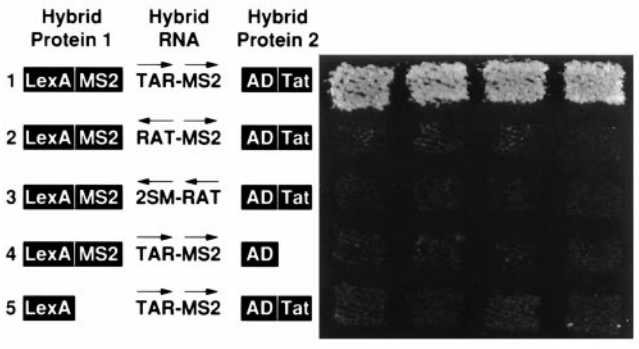

FIG. 4. Direct selection for RNA–protein interaction. Combinations of three plasmids were transformed into L40–*ura3*, in which the *HIS3* gene is under the control of LexA binding sites (as is the *lacZ* gene). Four independent transformants from each combination of plasmids were first streaked onto plates lacking tryptophan, leucine, and uracil; after 2 days the patches were replica-plated onto plates lacking histidine, tryptophan, leucine, and uracil and containing 20 mM 3-amino-1,2,4-triazole. The hybrid proteins and RNAs encoded by each of the three plasmids are indicated to the left. Arrows indicate the polarity of RNA sequences. RAT–MS2 RNA contains the antisense TAR sequence 5′ of MS2 sites; 2SM–RAT RNA contains antisense TAR 3′ of antisense MS2 sequences. Other RNAs and plasmids are described in Fig. 3 and Table 1.

Table 2. LexA–coat protein integrant facilitates the three-hybrid assay

Hybrid RNA	Hybrid protein 2	β-gal activity, units
1. MS2–IRE	AD–IRP1	110
2. MS2–IRE	AD–Tat	5.4
3. MS2–IRE	AD	3.4
4. None	None	<1

The indicated plasmids encoding a hybrid RNA and hybrid protein 2 were transformed into yeast strain L40–coat, which constitutively expresses the LexA–MS2 coat protein hybrid, by selection for *URA3* and *LEU2* expression. β-Gal activity was determined as described in Fig. 3. Plasmids are as described in Fig. 3 and Table 1.

The dissociation constants of the interactions we have detected, based on *in vitro* experiments under roughly physiological conditions, vary from ≈0.01–0.1 nM (IRE/IRP1) (26, 27, 29, 30) to 1–10 nM (Tat/TAR, and MS2 coat protein/tandem variant MS2 sites) (24, 31–34). The minimal affinity required to yield detectable transcriptional activation in the three-hybrid system has not been determined. In the two-hybrid system, however, protein–protein interactions with K_d values greater than 1 μM can be detected (35, 36); as similar reporter strains are used in the two assays, it may be possible to detect RNA–protein interactions of relatively low affinity. In those cases in which a specific RNA–protein interaction is stabilized or enhanced by an interaction with a second protein, as appears common in the assembly of complex RNA–protein particles, it may prove necessary to supply multiple proteins to detect the interaction.

Since the assay detects RNA–protein interactions via transcriptional activation, the hybrid RNAs presumably are present in the nucleus. The RNAs may be nuclear either because they are derived from RNase P RNA and its promoter (since RNase P RNA is not known to enter the cytoplasm), or because after transport to the cytoplasm, the RNAs are returned to the nucleus by binding to their cognate hybrid proteins, which themselves carry a nuclear localization signal.

The use of an RNA polymerase III promoter to produce the hybrid RNA may restrict the range of sequences that can be analyzed in the three-hybrid system, because RNA polymerase III often terminates at runs of four or more consecutive uridine residues (37). This complication may be circumvented through the use of alternative RNA polymerases, such as those from bacteriophages T7 and SP6. The effect of potential polymerase III terminators in an RNA sequence of interest can be assessed directly by inserting the entire RNA sequence between MS2 and IRE elements and then assaying the IRE/IRP1 interaction in the three-hybrid system.

Our results provide a potentially powerful means by which to examine RNA–protein interactions *in vivo*, toward a variety of ends. At a minimum, the method should be capable of defining domains, as well as single amino acid residues or nucleotides, that are necessary *in vivo* for RNA–protein interactions that have been previously characterized. Additionally, this method should be useful for identifying and cloning the

genes for RNA–binding proteins that recognize biologically important RNA sequences, such as those that control the processing, translation, location, and stability of specific mRNAs, and the packaging and infectivity of RNA viruses. Proteins that interact with such sequences may be identified by using the large variety of existing libraries of genomic and cDNA sequences in activation domain vectors. The specificity demonstrated in the examples described here suggests that such searches should be feasible. In such searches, recognition of an RNA element by its specific binding protein must be detectable against a background due to cDNAs encoding RNA binding domains with little sequence specificity. The three-hybrid assay may also be applied to identify RNA ligands by preparing a library of hybrid RNAs, each of which carries a different artificial or cellular sequence fused to coat protein binding sites. Such a library may enable the identification, for example, of specific mRNAs that bind to a defined protein. Other potential applications include a facile means to screen *in vivo* for inhibitors of a known RNA–protein interaction (e.g., that between Tat and the TAR element); a means to screen for proteins that facilitate a known weak RNA–protein interaction (as appears to be common in mRNA processing); and an *in vivo* method to identify or assay synthetic RNA oligonucleotides with selective affinity for defined proteins, analogous to *in vitro* approaches that exploit reiterative selections (3, 4). Finally, it may be possible to extend this method ito generate a four-hybrid system for the analysis of RNA–RNA interactions *in vivo*, using two fixed protein hybrids and two different hybrid RNAs.

Transcriptional activation in the three-hybrid system relies only on the physical, and not the biological, properties of the RNA. The RNA–protein interactions are assayed in an entirely foreign context, having nothing to do with the normal function of the RNA molecule. As a result, a wide variety of RNA–protein interactions should be amenable to analysis.

We are grateful to Peter Garber and Liz Barlow for experimental assistance; to Eckard Wimmer, Judith Kimble, and members of the Fields and Wickens laboratories for comments on the manuscript; and to Bryan Cullen, Steve Elledge, Rick Eisenstein, David Engelke, Alan Frankel, Rolf Sternglanz, Olke Uhlenbeck, and William Walden for plasmids and yeast strains. We appreciate the excellent assistance of Laura VanderPloeg and Adam Steinberg in preparing figures. This work was supported by National Institutes of Health Grant GM31892 to M.W. and National Institutes of Health Grants GM49065 and CA28146 to S.F.

1. Stripecke, R., Oliveira, C. C., McCarthy, J. E. & Hentze, M. W. (1994) *Mol. Cell. Biol.* **14,** 5898–5909.
2. Laird-Offringa, I. A. & Belasco, J. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 11859–11863.
3. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. (1995) *Annu. Rev. Biochem.* **64,** 763–797.
4. Szostak, J. W. (1992) *Trends Biochem. Sci.* **17,** 84–93.
5. Fields, S. & Song, O. (1989) *Nature (London)* **340,** 245–246.
6. Chien, C.-T., Bartel, P. L., Sternglanz, R. & Fields, S. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9578–9582.
7. Bartel, P. L. & Fields, S. (1995) *Methods Enzymol.* **254,** 241–263.
8. Good, P. D. & Engelke, D. R. (1994) *Gene* **151,** 209–214.
9. Legrain, P., Dokhelar, M.-C. & Transy, C. (1994) *Nucleic Acids Res.* **22,** 3241–3242.
10. Gott, J. M., Willis, M. C., Koch., T. H. & Uhlenbeck, O. C. (1991) *Biochemistry* **30,** 6290–6295.
11. Bartel, P. L., Chien, C.-T., Sternglanz, R. & Fields, S. (1993) in *Cellular Interactions in Development: A Practical Approach,* ed. Hartley, D. A. (Oxford Univ. Press, Oxford), pp. 153–179.
12. Bardwell, V. J. & Wickens, M. (1990) *Nucleic Acids Res.* **18,** 6587–6594.
13. Eisenstein, R. S., Tuazon, P. T., Schalinske, K. L., Anderson, S. A. & Traugh, J. A. (1993) *J. Biol. Chem.* **268,** 27363–27370 .
14. Patino, M. M. & Walden, W. E. (1992) *J. Biol. Chem.* **267,** 19011–19016.
15. Ammerer, G. (1983) *Methods Enzymol.* **101,** 192–201.
16. Geitz, R. D. & Sugino, A.(1988) *Gene* **74,** 527–534.
17. Hope, T. J., Huang, X., McDonald, D. & Parslow, T. G. (1990) *Proc. Natl Acad. Sci. USA* **87,** 7787–7791.
18. Cullen, B. R. (1986) *Cell* **46,** 973–982.
19. Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A. E., Lee, W.-H. & Elledge, S. (1993) *Genes Dev.* **7,** 555–569.
20. Bartel, P. L., Roecklein, J. A., SenGupta D. J. & Fields, S. (1996) *Nat. Genet.* **12,** 72–77.
21. Iwabuchi, K., Li, B., Bartel, P. & Fields, S. (1993) *Oncogene* **8,** 1693–1696 .
22. Brent, R. & Ptashne, M. (1985) *Cell* **43,** 729–736 .
23. Witherell, G. W., Wu, H.-N. & Uhlenbeck, O. C. (1990) *Biochemistry* **29,** 11051–11057.
24. Lowery, P. T. & Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15,** 10483–10493.
25. Uhlenbeck, O. C., Carey, J., Romaniuk, P. J., Lowary, P. T. & Beckett, D. B. (1983) *J. Biomol. Struct. Dyn.* **1,** 539–552.
26. Klausner, R. D., Rouault, T. A. & Harford, J. (1993) *Cell* **72,** 19–28.
27. Melefors, O. & Hentze, M. W. (1993) *Bioessays* **15,** 85–90.
28. Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. (1993) *Cell* **74,** 205–214.
29. Haile, D. J., Hentze, M. W., Rouault, T., Harford, J. & Klausner, R. D. (1989) *Mol. Cell. Biol.* **9,** 5055–5061.
30. Barton, H. A., Eisenstein, R. S., Bomford, A. & Munro, H. N. (1990) *J. Biol. Chem.* **265,** 7000–7008.
31. Cullen, B. R. (1992) *Microbiol. Rev.* **56,** 375–394.
32. Gait, M. J. & Karn, J. (1994) *Trends Biochem. Sci.* **18,** 255–259 .
33. Slice, L. W., Lodner, E., Antelman, D., Holly, M., Wegrzynki, B., Wang, J., Toome, V., Hsu, M. C. & Nalin, C. M. (1992) *Biochemistry* **31,** 12062–12068.
34. Calnan, B. J., Biancalana, S., Hudson, D. & Frankel, A. D. (1991) *Genes Dev.* **5,** 201–210 .
35. Fields, S. & Yang. (1995) *Nucleic Acids Res.* **23,** 1152–1156.
36. Estojak, J., Brent, R. & Golemis, E. A. (1995) *Mol. Cell. Biol.* **15,** 5820–5825.
37. Geiduschek, E. P. & Tocchini-Valentini, G. P. (1988) *Annu. Rev. Biochem.* **57,** 873–914.
38. MacWilliams, M. P., Celander, D. W. & Gardner, J. F. (1993) *Nucleic Acids Res.* **21,** 5754–5760.
39. Harada, K., Martin, S. S. & Frankel, A. D. (1996) *Nature (London)* **380,** 175–179.