A novel genetic system to detect protein–protein interactions

Stanley Fields & Ok-kyu Song

Department of Microbiology, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

PROTEIN–protein interactions between two proteins have generally been studied using biochemical techniques such as crosslinking, co-immunoprecipitation and co-fractionation by chromatography. We have generated a novel genetic system to study these interactions by taking advantage of the properties of the GAL4 protein of the yeast *Saccharomyces cerevisiae*. This protein is a transcriptional activator required for the expression of genes encoding enzymes of galactose utilization[1]. It consists of two separable and functionally essential domains: an N-terminal domain which binds to specific DNA sequences (UAS_G); and a C-terminal domain containing acidic regions, which is necessary to activate transcription[2,3]. We have generated a system of two hybrid proteins containing parts of GAL4: the GAL4 DNA-binding domain fused to a protein 'X' and a GAL4 activating region fused to a protein 'Y'. If X and Y can form a protein–protein complex and reconstitute proximity of the GAL4 domains, transcription of a gene regulated by UAS_G occurs. We have tested this system using two yeast proteins that are known to interact—SNF1 and SNF4. High transcriptional activity is obtained only when both hybrids are present in a cell. This system may be applicable as a general method to identify proteins that interact with a known protein by the use of a simple galactose selection.

The basic strategy of these experiments is shown in Fig. 1. The native GAL4 protein, containing both domains, is a potent activator of transcription when yeast are grown on galactose media. The N-terminal domain binds to DNA in a sequence-specific manner but fails to activate transcription[2]. The C-terminal domain contains the activating regions but cannot activate transcription because it fails to localize to UAS_G (see, for example, ref. 4). Creation of two hybrid proteins, each containing one of the interacting proteins X and Y, allows the activating region of GAL4 to be brought to its normal site of action. Our test case for two interacting proteins uses the yeast SNF1 protein, a serine-threonine-specific protein kinase[5] and the SNF4 protein, which is physically associated with SNF1 and is required for its maximal activity (J. Celenza & M. Carlson, personal communication).

The constructs used here are shown in Fig. 2: GAL4(1-881) is the entire GAL4 protein of 881 amino acids[6]; GAL4(1-147) is the N-terminal 147 amino acids of GAL4, which binds to UAS_G; GAL4(1-147)-SNF1 is a hybrid protein containing the N-terminal domain of GAL4 fused in-frame to the entire coding sequence (633 amino acids)[5] of the SNF1 protein. The GAL4 N-terminal domain contains a nuclear-localization sequence[7], so we expect that at least some fraction of the GAL4(1-147)-SNF1 would be localized to the nucleus. SNF4 is the entire SNF4 protein (322 amino acids; J. Celenza, F. Eng and M. Carlson, personal communication); SNF4-GAL4(768-881) is a hybrid protein containing all but the last amino acid of SNF4 fused in-frame to the GAL4 activating region II. This activating region is sufficient when fused to the GAL4 DNA-binding domain to induce substantial transcriptional activity[3].

We introduced plasmids containing these different derivatives into strain GGY1::171[8], which is deleted for both GAL4 and GAL80 (a negative regulator of GAL4) and which contains a *GAL1-lacZ* fusion gene integrated at the *URA3* locus. Thus β-galactosidase activity is a measure of GAL4 function derived from the plasmid-borne GAL4 constructs. The strain also contains mutations of the *HIS3* and *LEU2* genes, which are the markers for plasmids containing the DNA-binding and activation domains, respectively. Transformants were grown in media which can induce transcription from UAS_G (2% galactose, 2% ethanol, 2% glycerol) and which did not contain leucine or histidine, as appropriate to maintain the plasmids.

Our results are shown in Table 1. There was no activity in the absence of any GAL4 plasmid (line 1); 4,000 units of β-galactosidase were produced when the native GAL4 protein was provided (line 2). As expected, the DNA-binding domain alone did not activate transcription (line 3), nor did the GAL4(1-147)-SNF1 fusion protein (line 4), indicating that the SNF1 domain does not function as an activating region. The intact SNF4 protein (line 5) and SNF4-GAL4(768-881) fusion protein also failed to activate transcription. When both GAL4(1-147)-SNF1 and SNF4-GAL4(768-881) were introduced however, 180 units of β-galactosidase were produced (line 7), an induction

TABLE 1 Transcriptional activation produced by hybrid GAL4 proteins

Plasmid	β-Galactosidase activity
1. None	<1
2. GAL4(1-881)	4,000
3. GAL4(1-147)	<1
4. GAL4(1-147)-SNF1	<1
5. SNF4	<1
6. SNF4-GAL4(768-881)	<1
7. GAL4(1-147)-SNF1; SNF4-GAL4(768-881)	180
8. GAL4(1-147)-SNF1; SNF4	7
9. GAL4(1-147); SNF4-GAL4(768-881)	<1

Plasmids were introduced[12] into yeast strain GGY1::171 (ref. 8) deleted for GAL4 and containing a *GAL1-lacZ* fusion gene in the chromosome. Transformants were grown in 2% galactose, 2% ethanol and 2% glycerol and assayed in triplicate for β-galactosidase activity, as described[13].

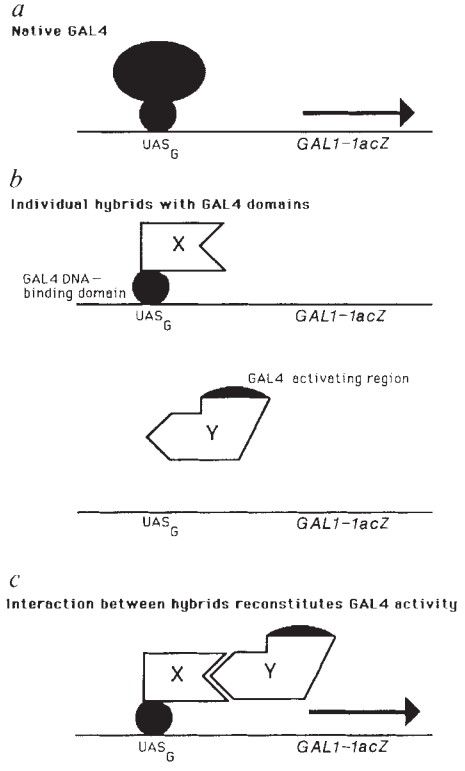

FIG. 1 Model of transcriptional activation by reconstitution of GAL4 activity. *a*, The native GAL4 protein contains both DNA-binding and activating regions and induces *GAL1-lacZ* transcription. *b*, Hybrids containing either the DNA-binding domain (upper) or activating region (lower) are incapable of inducing transcription. *c*, A protein–protein interaction between proteins X and Y brings the GAL4 domains into close proximity and results in transcriptional activity.

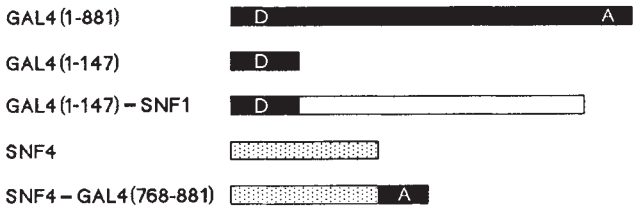

FIG. 2 Constructions used in our experiments. GAL4(1-881) is carried on plasmid pCL1, GAL4(1-147) on pMA424 (ref. 14), GAL4(1-147)-SNF1 on pEE5, SNF4 on pFF1 and SNF4-GAL4(768-881) on pNI12. D and A represent the DNA-binding and activating regions, respectively, of GAL4 protein.
METHODS. We constructed pCL1 by inserting the *Bam*HI fragment of pLKC15 (refs 3, 7) which contains the P_{ADH1}-*GAL4* gene into a YCp50 (described in ref. 15) derivative in which the *LEU2* gene replaces the *URA3* gene (from E. Stone). pEE5 is derived from pMA424 and pCC107 (from J. Celenza), which contains the *SNF1* gene in pUC18 (ref. 16) as follows. A *Pst*I site 12 base pairs upstream of the initiator methionine of *SNF1* was converted to a *Bam*HI site and then the *Bam*HI fragment containing the *SNF1* gene was inserted into the *Bam*HI site of pMA424; pFF1 was constructed by inserting the *Hin*dIII fragment of pFE27-2 (from J. Celenza, F. Eng and M. Carlson), which contains the *SNF4* gene, into the multi-copy yeast plasmid YEp13 (ref. 17). pNI12 was constructed from a pUC18 clone containing the *GAL4* gene and *ADH1* terminator (derived from pLKC15) from the *Nar*I site at residue 767 of *GAL4* to the 3′ *Bam*HI site. The *Nar*I site was converted to a *Kpn*I site, and the resulting *Kpn*I fragment ligated to pSL321 (from J. Celenza) at a *Kpn*I site 6 base pairs downstream of the penultimate amino acid of the *SNF4* gene. Finally, a *Hin*dIII fragment containing the *SNF4* promoter and gene fused to the GAL4(768-881) region was inserted into YEp13.

of several hundredfold over the background level of the single hybrids. Although the activity produced by the two hybrids is only 4.5% of that produced by the native GAL4 protein expressed from the strong *ADH1* promoter, it is sufficient to produce a dark blue colony on plates containing the indicator X-gal. Thus yeast cells containing the interacting hybrid proteins can be detected easily against a background of cells which contain only a single hybrid and are white.

This induction with the two hybrids does not result from the SNF4 protein binding to the GAL4(1-147)-SNF1 hybrid and converting it to a transcriptional activator; introduction of a plasmid encoding the intact SNF4 protein into cells carrying this hybrid was not sufficient for such high activity (line 8), although a low level of activity was observed. Furthermore, the induction is not due to the GAL4 activating region interacting with the GAL4 DNA-binding region without the need for the SNF1-SNF4 interaction (line 9). The transcription produced by the combination of the GAL4(1-147)-SNF1 and SNF4-GAL4(768-881) hybrids is strictly dependent on inducing media, as this and all other combinations of hybrid proteins produced <1 unit of β-galactosidase activity in glucose media.

Our results extend those of a previous study[9] which indicate that the GAL80 protein can be converted into a transcriptional activator by the insertion of an acidic sequence and allowing the resultant protein to interact with a DNA-bound GAL4 derivative. Similarly, analysis of the herpes simplex virus protein VP16 indicates that it can activate transcription by binding to host proteins bound to DNA[10] and analysis of the protein product of c-*fos* indicates that it can activate transcription by binding to the DNA-bound c-*jun* protein product[11]. These examples, however, involve proteins naturally involved in regulating transcription, whereas the system we describe should be applicable to almost any two interacting proteins. The system requires that the interaction can occur within the yeast nucleus, that the GAL4-activating region is accessible to the transcription machinery, and that the GAL4(1-147)-protein X hybrid is itself not a potent activator.

We envisage that this system could be used as a method for the genetic selection of proteins that interact with a known protein, if the cloned gene for the known protein is available. Yeast containing this protein as a GAL4(1-147) hybrid would be transformed with a clone bank of genomic or complementary DNA sequences fused to the GAL4 activating region. The double transformants would be selected for ability to grow on galactose or screened for blue colour on indicator plates for those able to express the *GAL1-lacZ* fusion. Such an approach could be useful in identifying, for example, proteins that interact with oncogene-encoded products, or for mapping the specific interacting domains of proteins known to form a complex. Additionally, it could be of value in designing therapeutic peptides to interact with proteins of bacterial or viral origin. □

Received 3 May; accepted 12 June 1989.

1. Johnston, M. *Microbiol. Rev.* **51,** 458-476 (1987).
2. Keegan, L., Gill, G. & Ptashne, M. *Science* **231,** 699-704 (1986).
3. Ma, J. & Ptashne, M. *Cell* **48,** 847-853 (1987).
4. Brent, R. & Ptashne, M. *Cell* **43,** 729-736 (1985).
5. Celenza, J. L. & Carlson, M. *Science* **233,** 1175-1180 (1986).
6. Laughon, A. & Gesteland, R. *Molec. cell. Biol.* **4,** 260-267 (1984).
7. Silver, P. A., Keegan, L. P. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **81,** 5951-5955 (1984).
8. Gill, G. & Ptashne, M. *Cell* **51,** 121-126 (1987).
9. Ma, J. & Ptashne, M. *Cell* **55,** 443-446 (1988).
10. McKnight, J., Kristie, T. & Roizman, B. *Proc. natn. Acad. Sci. U.S.A.* **84,** 7061-7065 (1987).
11. Curran, T. & Franza, B. R. Jr *Cell* **55,** 395-397 (1988).
12. Hinnen, A., Hicks, J. B. & Fink, G. R. *Proc. natn. Acad. Sci. U.S.A.* **75,** 1929-1933 (1978).
13. Yocum, R., Hanley, S., West, R. Jr & Ptashne, M. *Molec. cell. Biol.* **4,** 1985-1998 (1984).
14. Ma, J. & Ptashne, M. *Cell* **51,** 113-119 (1987).
15. Kuo, C.-L. & Campbell, J. L. *Molec. cell. Biol.* **3,** 1730-1737 (1983).
16. Yanisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33,** 103-119 (1985).
17. Broach, J. R., Strathern, J. N. & Hicks, J. B. *Gene* **8,** 121-133 (1979).

ACKNOWLEDGEMENTS. We thank Mark Johnston, John Celenza, Jun Ma, Grace Gill, Roger Brent and Rog Yocum for strains and plasmids. Supported by the NSF, Procter and Gamble Company and the New York State Science and Technology Foundation.

CORRIGENDUM

An eIF-4A-like protein is a suppressor of an *Escherichia coli* mutant defective in 50S ribosomal subunit assembly

Kayoko Nishi, Francoise Morel-Deville, John W. B. Hershey, Terrance Leighton & Joachim Schnier

Nature **336,** 496-498 (1988).

IN this letter parts of the DNA sequence and thus the derived protein sequence are revised at the following positions: nucleotides from position 610 to 615 are -ggcacc- encoding the amino acids GT. The nucleotides g (713), c (850) and g (874) are deleted. The amino acid sequence corresponding to the new reading frame from nucleotides 712 to 873 is TRELAMQV-SDHARELAKHTHLDIATITGGVAYMNHAEVFSENQDIV-VATTGRLL. The sequence from position 1233 to 1246 is -gcgtgagcg-. One -g- is inserted between nucleotides 1277-1278 and between 1280-1281. The corresponding new amino acid sequence from nucleotides 1231 to 1281 of the original sequence is KRERVHELANWLREAG. Identical amino acids between eIF-4A and the SrmB protein increase to 108 out of 382 amino acids (28.3%).

Since the uncorrected protein sequence of the *srmB* gene was contributed by K. Nishi and J. Schnier to the scientific correspondence by P. Linder *et al.* (*Nature* **337,** 121-122; 1989), the new sequence also changes the number of identical amino acids among the eIF-4A-like proteins described in that publication. The amino acids which now are common to all 8 eIF-4A-like proteins increase to 48/408 (11.8%). The SrmB protein shows divergence from this set of proteins at only 4 positions. □