Nucleic Acids Research, 1994, Vol. 22, No. 20 4031–4038

Strong transcriptional activators isolated from viral DNA by the 'activator trap', a novel selection system in mammalian cells

Matthias Gstaiger and Walter Schaffner*
Institute of Molecular Biology II, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received August 18, 1994; Accepted September 7, 1994

ABSTRACT

Transcription factors often contain activation domains that interact with the basic transcription machinery. We have developed a functional screening strategy in mammalian cells to selectively isolate activation domains from a library of random DNA inserts. For this, sonicated DNA fragments are cloned next to the DNA binding domain of GAL4 factor in a plasmid that also contains the SV40 origin of replication. Pools of fusion protein clones are transfected into CV-1-5GT monkey cells containing an SV40 T antigen gene under the control of a promoter with GAL4 binding sites. Plasmids that express functional transactivating fusion proteins activate the T antigen gene, thus promoting selective amplification of the plasmid in the mammalian host cell line. Using this method, we were able to select strong enhancer-type activation domains from the immediate early regions of two herpesviruses, namely pseudorabies virus and bovine herpesvirus 1. In both cases, the activation domains selected were homologues of the ICP4 regulatory protein of herpes simplex virus. The activation domain from pseudorabies virus is four times stronger than the activation domain of herpes simplex virus protein VP16 (Vmw65), making it the strongest activation domain characterized so far. This activator trap method should be useful for precisely localizing activation domain(s) in known factors, or to identify mammalian transcriptional adaptors that do not bind DNA and which may escape conventional detection methods.

INTRODUCTION

In mammalian cells, the DNA template for transcription by RNA polymerase II is not recognized as naked DNA; rather, the enzyme complex must be guided to the site of initiation by the help of DNA-bound transcription factors (1, 2; for review see 3). Such transcription factors often contain two major functional domains: a DNA-binding domain and a non overlapping activating domain for interaction with other factors of the transcriptional apparatus. Brent and Ptashne (1985) succeeded in producing chimeric GAL4 transcription factors by 'domain swapping'. This approach is now widely used to characterize activation domains of transcription factors.

Using GAL4-fusion proteins, we found that activation domains of mammalian transcription factors can be functionally classified into at least two types (4). 'Enhancer' type activation domains, such as the acidic domain of the herpes simplex virus VP16 (= Vmw65) protein, can activate transcription from remote and proximal positions. By contrast, glutamine-rich domains such as those of Oct-1 and Sp1, only stimulate transcription from a 'promoter' position proximal to the initiation site, and in response to a remote enhancer. Usually, activation domains have been identified by deletion analysis of known or putative transcription factors. A more demanding task is the selection of activation domains from a complex mixture of unrelated DNAs. Similar to a protocol of G. Schatz to detect fortuitous mitochondrial import sequences from a library of random *E.coli* DNA fragments (5, 6), M. Ptashne and colleagues demonstrated that about 1% of *E.coli* fragments conferred an activation potential to the DNA binding domain of GAL4 in yeast (7). While a positively charged amphipathic helix is the essential feature for mitochondrial import, the situation with activation domains seems more complex (8, 9). At first, a negatively charged amphipathic helix was postulated to be a common feature. Indeed, there is a preponderance of acidic residues in many activation domains, and some have the potential to form an alpha helix, although at low pH (10). While GAL4 and GCN4 activation domains are most likely folded in a beta sheet (11), evidence suggests that some activation domains may lack any ordered structure. In spite of striking examples of mammalian activation domains that work in yeast and vice versa, not all activation domains are universally active in eukaryotes. For example, the glutamine-rich domains of mammalian Oct and Sp1 factors fail to activate transcription in yeast (12), and the acidic domain of the metallothionein activating factor MTF-1 works in HeLa cells and frog oocytes but not in early frog embryos (13). Post-translational modifications, or the availability of specific mammalian 'adaptor' proteins that interact with an activation domain, may explain this selectivity.

*To whom correspondence should be addressed

We have developed a selection system to facilitate cloning of transcriptional activation domains in mammalian cells. This protocol for the isolation of activation domains is reminiscent to our previous 'enhancer trap' method, where we selectively replicated SV40 DNA which had incorporated an enhancer from a large collection of randomly sonicated DNA fragments (14, 15), and is also based on the 'trans-dependent replication' (TDR) assay of Rusconi *et al.* (1990). In the activator trap approach, activation domains are isolated from a GAL4 fusion library by means of selective plasmid replication, mediated in transfected monkey cells by a GAL4-inducible T antigen gene. When compared to the selection protocols for *cis*-acting DNA sequences (enhancers), selection of protein domains is subject to further constraints and thus technically more demanding.

Using our activator trap assay, we isolated several activation domains from random fragments of herpes viral DNA. In addition to the already well characterized activation domain of the herpes simplex virus tegument protein VP16 (Vmw65) (16), we isolated activation domains from the infected cell protein 4 (ICP4) homologues of pseudorabies virus (PRV) and bovine herpes virus 1 (BHV-1). All of the selected domains were subsequently tested in a conventional transcription assay and found to activate transcription both from proximal promoter as well as remote enhancer positions.

MATERIALS AND METHODS

Construction of plasmids

The target gene vector p5Gal4TAG for making the stably transformed CV-1 cell line CV-1-5GT is based on pH4-GGN-TAG (17). The GCN4 binding sites of pH4-GGN-Tag were replaced by a PstI/BamHI fragment derived from pG5E4 (42) containing five Gal4 binding sites. Expression vectors are all based on the plasmid pSCTEV and contain the DNA binding domain of GAL4 (4). Expression vector pSCTEV-GAL4(1−93)RV was obtained by introducing an oligonucleotide (A) with an EcoRV site (underlined) into the PstI and XbaI site pSCTEV-GAL4(1−93) and used for blunt end cloning of sonication products to the 5′ end of the GAL4 DNA binding domain.

A. 5′-GCTGATATCGT-3′
3′-ACGTCGACTATAGCAGATC-5′

Reporter constructs used for S1 nuclease assays bearing GAL4 binding sites are based on OVEC-1 (43) and described previously (44).

For activator trap experiments with the pseudorabies virus immediate early protein we used the expression vector pSCTEV-PRVIE containing a BamHI fragment (4815 bp) encoding the entire PRV IE protein (25). Plasmid pJuC (a kind gift of Martin Schwyzer) was used for cloning of the activation domain from the immediate early region of BHV-1 and contains the BHV-1 Jura HindIII C fragment (0.733 to 0.852 map units) described previously (27).

Activator trap assay

Twenty μg of plasmid DNA carrying the DNA segment to be analyzed for transcriptional activation domains was sonicated using the BRANSON sonifier system. Before blunt end cloning into pSCTEV-Gal4(1−93)RV, sonicated fragments of approximately 500 bp were selected by electrophoresis in a 1% agarose gel and repaired by treatment with T4 polymerase, Klenow polymerase and T4 kinase. The plasmid library was made by ligating the sonicated DNA fragments into the EcoRV site of pSCTEV-GAL4(1−93)RV and transformed into *E.coli* strain DH1 by electroporation using the BioRad genepulser system according to the instructions of the manufacturer. Pools of 500−1000 different clones were grown in LB broth for plasmid preparation and 0.1 to 1.0 μg of the plasmid pools transfected into CV-1-5GT cells using the DEAE dextran method (45). We found that the presence of GAL4(1−93) reduces the amplification efficiency of an active GAL4 fusion protein, presumably due to competition for the GAL4 binding sites. Also, to avoid the problem of replication 'parasites' (plasmids that are replicated but do not contain an activation domain), not more than 1000 different GAL4 fusion clones should be transfected per 100 mm dish and the amount of transfected pool DNA should be as low as possible (we started out with 0.3−1.0 μg, but in later experiments we routinely used 100 ng per 100 mm dish). In addition, we would like to point out that this highly sensitive assay is prone to contamination by previously used transactivator clones unless purity precautions of reagents similar as to those for PCR reactions are taken. Forty eight hours after transfection, low molecular weight DNA was isolated and digested with DpnI (17). Competent DH1 *E.coli* were transformed with aliquots of the DpnI digested plasmid DNA by electroporation and plated on ampicillin plates. Plasmid DNA of individual ampicillin resistant clones was isolated and 100 ng used for transfection into CV-1-5GT cells for measuring the transactivation potential of selected clones in a replication assay (17). For this purpose the plasmid DNA was harvested 48h after transfection, DpnI digested, transformed into *E.coli* strain DH1 and the activation potential was simply determined by the number of colonies transformed by plasmids replicated in CV-1-5GT cells. Alternatively, for measuring transactivation capacity of selected GAL4 fusion clones we cotransfected isolated clones together with a reporter plasmid bearing five GAL4 sites in a promoter position upstream from a luciferase gene into CV-1 cells and measured the induced luciferase activity. GAL4 fusion clones containing transcriptional activation domains were sequenced by the Sanger dideoxynucleotide technique (46).

Calculation of library size

In the previous 'enhancer trap' experiments where we selected active enhancers from random fragments of DNA, there was no constraint on reading frame and orientation (14, 15, 47, 48), which in protein fusions, as described here, reduces the chance for functional integration by a factor of 6. We calculated that from a 10 kb segment, which contains one activation domain of 50−150 aa length, a library of $25 \times 10 \times 6 = 1500$ clones should be tested to get in average one functional insert, assuming sonication fragments of 400 bp, and a window tolerance of ± 40 bp. If about 1000 clones are simultaneously transfected into one tissue culture plate of 100 mm diameter, screening of a 200 kb herpesvirus genome requires more than 30 000 clones (30 plates). Indeed, from 20 000 clones of the sonicated genome of human cytomegalovirus (CMV) we have recently isolated two different activation domains (M. Gstaiger, T. Stamminger, B. Fleckenstein and W. Schaffner, unpublished data), however we do not know the total number of such domains in CMV. It is generally assumed that 5% of all eukaryotic proteins are transcription factors (49), thus screening of some 50−100 kb

of cDNA may yield one activation domain. In the specific case of PRV, we selected one positive clone in duplicate from 5000, starting out with 10 kb of sonicated plasmid. From the 18.9 kb plasmid containing the 16.2 kb of the early region of the bovine herpesvirus 1, we have isolated the very same region of DNA (with slightly different integration junctions) from three

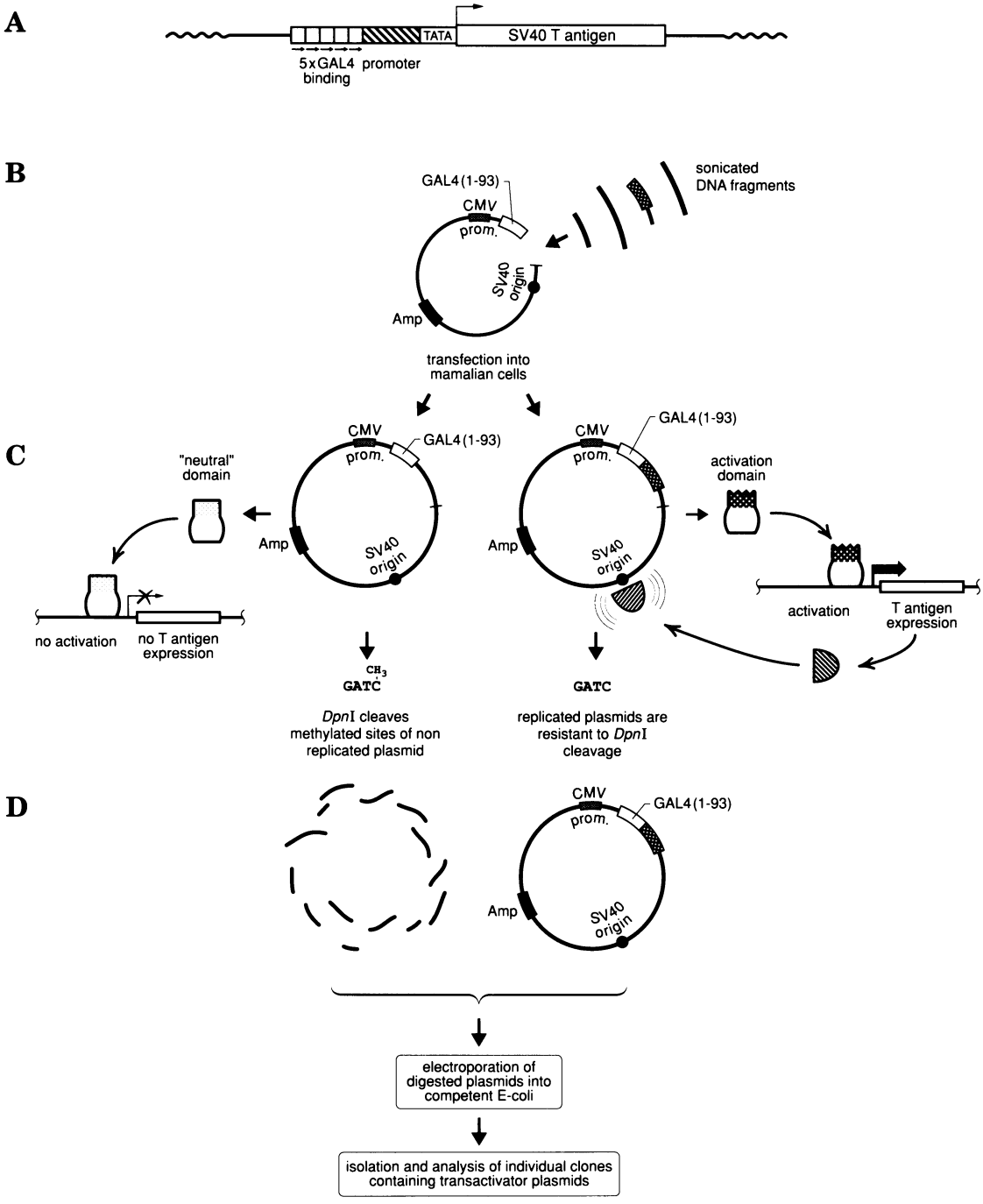

Figure 1. Selection of activation domains by the activator trap assay. **A)** Structure of the stably integrated target gene in CV-1-5GT cells used for activator selection. The vector used for transformation contains an inducible T antigen gene under the control of five GAL4 binding sites upstream from an MMTV promoter. **B)** Preparation of GAL4(1−93) fusion libraries. DNA to be analyzed for activation domains is fragmented by sonication and fragments of 300−600 bp are cloned into the EcoRV site of the expression vector pSCTEV-GAL4(1−93)RV. The GAL4 fusion library is transfected into the CV-1-5GT cells with the SV40 T antigen under the control of five GAL4 binding sites. **C)** Selective replication of plasmids encoding an activation domain. Expression of active GAL4 fusion proteins leads to the production of SV40 T antigen which induces replication of the SV40 origin-containing library plasmids. **D)** Isolation of replicated plasmids. Replicated plasmids become resistant to cleavage by DpnI enzyme and can be transformed into *E.coli*, which allows subsequent isolation of individual clones encoding activation domains. Finally, individual clones were tested again for transactivation funtion in the replication assay.

independent transfection experiments. No other positive clone was found among 12 000 tested. This is about two thirds of the theoretical expectation.

DNA transfection and RNA analysis

To obtain the stably transformed CV-1 cell line CV-1-5GT we transfected CV-1 cells with 15 μg p5GAL4Tag target gene and 1.5 μg of pSV2-Neo (50) and selected for neomycin resistant clones (51). This cell line CV-1-5GT with GAL4-factor dependent T antigen expression is available from our laboratory on request and has also been deposited with ATCC. For RNA mapping experiments, Hela cells were transfected by the calcium phosphate coprecipitation procedure (43) with 10 μg reporter plasmid, 1 μg activator plasmid, 1 μg of reference plasmid and 8 μg sonicated herring sperm DNA as carrier per 100 mm dish. The reference plasmid used in all transfection was OVEC-REF (43). After incubation of the cells for 36h, RNA isolation and S1 mapping were performed as described previously (43). Quantification of autoradiographs was performed with a phosphor imaging device (Molecular Dynamics) and the signals of the reporter genes were normalized to reference signals to correct for some variability in the transfection efficiency.

RESULTS

A transactivation dependent replication assay for the cloning of transcriptional activation domains

In order to set up a selection scheme for transactivation domains, we combined the enhancer trap approach with the transactivation dependent replication assay described previously (15, 17). An expression library was prepared, where the DNA to be analyzed was sonicated and fused to the DNA binding domain (1−93) of the yeast GAL4 activator (Figure 1B). These randomly generated fusion plasmids carried an SV40 origin of replication and were transfected in pools of approximately 300−1000 different clones per 100 mm plate into the monkey kidney cell line CV-1-5GT. This cell line, derived from CV-1, was stably transformed with the target vector p5GAL-Tag expressing the SV40 T antigen under the control of five GAL sites located at −226 to −131 bp from the TATA box (Figure 1A). If the DNA fragment fused to the GAL4 DNA binding domain encodes a transcriptional activation domain, SV40 T antigen gene expression is induced and, consequently, the fusion plasmids containing the SV40 origin are selectively replicated (Figure 1C). After isolation from the cells, plasmid DNA is digested with DpnI that selectively degrades nonreplicated plasmid DNA. Replicated plasmids encoding transcriptional actvation domains are resistant to DpnI digestion and therefore able to transform *E.coli* (Figure 1D). Individual clones were isolated from bacterial colonies, transfected again into CV-1-5GT cells and transcriptional activity was analyzed by measuring transactivator-induced plasmid replication under the control of T antigen in CV-1-5GT cells. The relative level of activation was scored by counting the number of *E.coli* colonies obtained. In subsequent experiments we simplified the transactivation test by cotransfecting the individual GAL4 fusion clones together with a GAL4-dependent luciferase gene and measuring luciferase activity.

In order for such a selection scheme to operate with the desired sensitivity, the T antigen gene in the host cell line must be silent in the absence of an activator but induced to high levels in the presence of the activator. For this purpose, several CV-1 cell lines transformed with the reporter construct p5GAL-Tag were tested for their inducibility of T antigen expression upon transfection of either the inactive clone pSCTGAL4(1−93) expressing the GAL4 DNA binding domain alone, as a negative control, or the transcriptional activator pSCTGAL4(1−93)-VP16(413−490) (henceforth referred to as GAL4-VP16). T antigen expression in this case was measured indirectly by the ability to replicate SV40 origin-containing plasmids. A cell clone (CV-1-5GT) with no basal T antigen expression, but high expression in the presence of GAL4-VP16, was chosen for further studies. We found that the level of T antigen expression upon transfection of several previously characterized GAL4-based activators correlated well with the relative activation potential of these activators, as measured by a conventional S1 nuclease transcript mapping (data not shown). In order to reduce the problem of having too many different fusion plasmids present in one cell, which would allow replication of 'parasite' plasmids (i.e. without a transcriptional activation domain of their own), we kept the amount of pool DNA transfected into each tissue culture dish as low as possible.

The herpes simplex virus VP16 activation domain can be selected in a transactivation dependent replication assay

The herpes simplex virus tegument protein VP16 is a strong transcriptional activator that activates the promoters of the immediate early genes IE175, IE110, and IE63 (reviewed in 18). Activation of these promoters requires interaction with the cellular DNA binding transcription factor Oct-1 and additional host cell factors (19−21). The VP16 protein contains a very strong acidic transcriptional activation domain which is located at the C-terminus (aa 413−490) of the protein (16, 22). Therefore we used this protein to test our activator trap system. We first made a GAL4 fusion library with inserts from the sonicated plasmid GAL4-VP16 (4). As a recipient for ligation to the C-terminal part of the GAL4 DNA binding domain, we constructed the vector pSCTEV-GAL4(1−93)RV. The ligation mix was transformed into *E.coli* by electroporation. We obtained 3000 recombinant clones that were split into 10 pools before preparation of plasmid DNA. Each plasmid pool was transfected into CV-1-5GT cells and analyzed for plasmid replication two days after transfection. From the pools that gave rise to the highest number of replicated plasmid, a total of eight individual colonies were picked and each clone was tested for activity in the replication assay (Table III). Among these, one clone showed high activity equal to that of the original plasmid GAL4-VP16 while the others were inactive (data not shown). As expected, sequence analysis showed that the isolated clone contained an in-frame insertion of a DNA segment encoding the VP16 activation domain. The entire 78 aa activator segment was fused to GAL4, creating at the junction a small duplication in the GAL4 sequence (75−93) (Table I).

A strong activation domain from the major immediate early protein of pseudorabies virus

During the immediate early phase of the infection cycle, herpesviruses express transcription factors that regulate a cascade of viral and cellular genes necessary for efficient replication of the virus. The major immediate early protein of pseudorabies virus (PRV IE) which belongs to the group of ICP4 related IE proteins of α-herpesviruses, is known to activate several viral and cellular promoters (23−25). In most cases, transcriptional

Table I. Amino acid sequence of activation domains isolated by the activator trap system

VP16-1 (413-490)	Lgtpaad**DLDMILKMDSLQDIKALL**gtpaaastqfpqiw
	APPTDVSLGDELHLDGEDVAMAHADALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG
PRVIEP-9 (1-120)	Lgtpaad*RKTPIEI*
	MADDLFDFIETEGNFSQLLAAAAAAAEEEGIASGPDGGSQGSRRRGSSGEDLLFGPGGLFSDDAAEAEAAVLAAAAGATRPPRPPSA
	QQQRHARRGSGEIVVLDDEDEEEDEPGSPAAGShrlesrs
BICP4-1 (1-78)	Lgtpaad*PTPSPPGHRTHPRPV*
	MSSPGRLELATLDLYDLIESADLGPTGGSEEDPALLDAARRAEARERRRAARAELAELWMVGGEDAESSEDDGAGDAdrlesrs
BICP4-11 (1-133)	Lgtpaad*PCVLTASRPTRPPTPSPPGHRTHPRPV*
	MSSPGRLELATLDLYDLIESADLGPTGGSEEDPALLDAARRAEARERRRAARAELAELWRMVGGEDAESSEDDGAGDAGATEGAEAE
	DAEIGEDAGAGDAGGAEDADVAECAEAEGAEGAEDADSAWAAAARALdrlesrs
BICP4-15 (1-92)	Lgtpaad*LLGGALLPPARSVSFAAARGPCRCPPPPPCVLTASRPTRPPTPSPPGHRTHPRPV*
	MSSPGRLELATLDLYDLIESADLGPTGGSEEDPALLDAARRAEARERRRAARAELAELWRMVGGEDAESSEDDGAGDAGATEGAEAE
	DAEIGdrlesrs

Numbers in brackets refer to the position of the activation domain within the viral protein. Amino acid sequences shown start with the leucine residue 93 of GAL4. Residues corresponding to viral amino acid sequences are boxed. Additional residues encoded by viral mRNA leader or vector linker sequences are shown in italics and lower case letters, respectively. The region corresponding to the duplicated GAL4 (75−93) portion in clone GAL4-VP16-1(413−490) is indicated with capital letters.

Table II. Properties of isolated activation domains

GAL4 fusion clone	net charge	proximal activation[a,c]	remote activation[b,c]
GAL4 (1-93)		0.02	0.03
GAL4 (1-93) - VP16 (413-490)	-18	1.00	1.00
GAL4 (1-93) - PRVIEP-9	-15	3.98	4.31
GAL4 (1-93) - BICP4-1	-9	1.38	2.56
GAL4 (1-93) - BICP4-11	-24	1.37	2.18
GAL4 (1-93) - BICP4-15	-9	0.97	1.42

[a]Transcriptional activation on the reporter construct p5G-OVEC (Figure 2).
[b]Transcriptional activation on the reporter construct p2Sp1/β5G-OVEC (Figure 2).
[c]Levels of activation have been determined by phosphorimaging and are standardized to the activation level of GAL4-VP16(413−490) (Figure 3).

Table III. Summary of activation domain selection by the activator trap method

Plasmid DNA analyzed	total plasmid size (kb)	size of viral insert (kb)	number of tested library clones[a]	individual clones tested	number of positive clones
GAL4 (1-93) - VP16 (413-490)	5.20	0.36	3000	8	1
pSCT-PRIEV	9.26	4.80	5000	10	2
pJuc	18.90	16.20	12000	133	3

[a]total number of library plasmids transfected in pool experiments.

activation mediated by PRVIE may involve protein−protein interactions with DNA-bound transcription factors, rather than direct DNA binding. In an exhaustive deletion analysis, M. Green and colleagues defined an activation domain within the 1460 aa protein which encompasses the first 34 aa of the N-terminus of the protein (26).

We decided to extend these studies using our activator trap assay to determine the optimal configuration of the activator domain(s) from this large protein. After the transfection of pools from a library with fragments of pSCT-PRIEV into CV-1-5GT cells, we isolated ten individual clones and found two of them to be active when tested individually (Table III). They turned out to be of identical sequence, presumably representing duplicates from a single cloning event. The 382 bp insert corresponded to the N-terminus of the pseudorabies immediate early protein. These 120 aa included the above mentioned activating sequence. No other regions conferring transcriptional activity were found.

This result demonstrates that the activator trap system can specifically select for transcriptional activation domains from a large excess of functionally unrelated DNA, without significant background from fortuitous pseudo-activation domains.

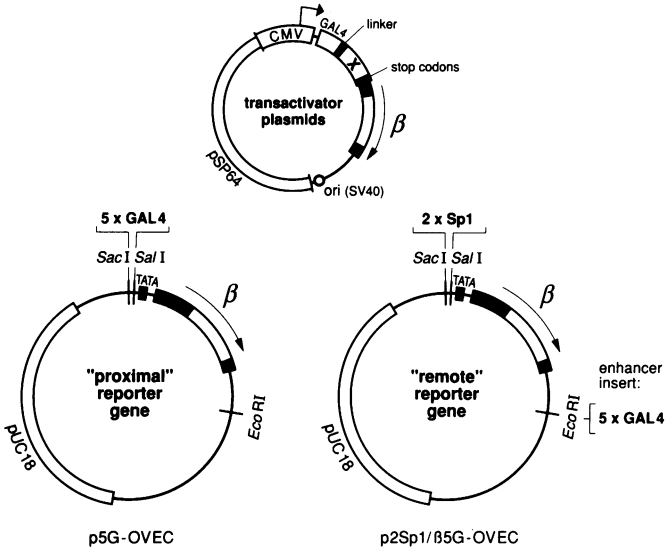

Figure 2. Schematic drawing of reporter constructs and the expression vector used in cotransfection experiments. All reporter constructs are derived from OVEC-1 (43). Reporter construct 5GOVEC for testing promoter type activation contains five GAL4 binding sites in a promoter position upstream from the TATA box, whereas enhancer-type activation was tested on the reporter 2SP1/β5G-OVEC that contains five GAL4 binding sites in an enhancer position downstream of the β globin reporter gene. All GAL4 fusion proteins were expressed from a plasmid based on pSCTEV-GAL4(1−93). Isolated activation domains fused to the GAL4 DNA binding domain are indicated with X (for details see Table I and Materials Methods).

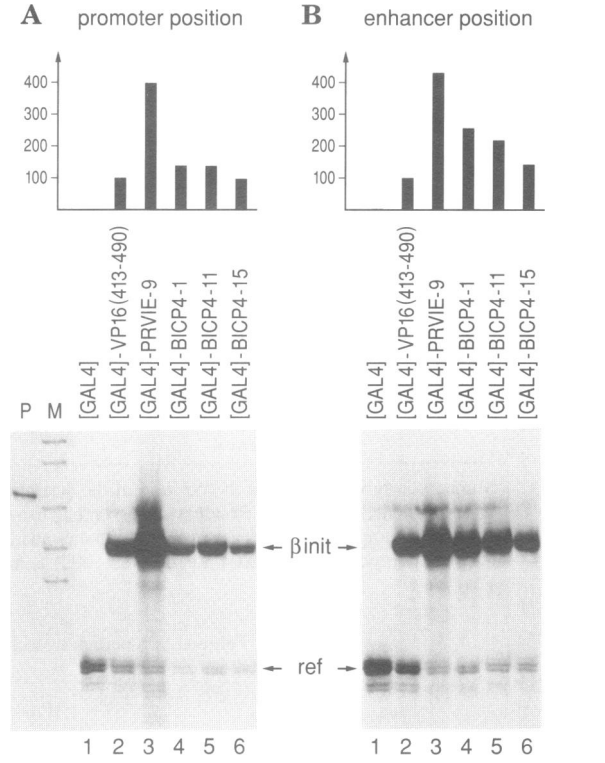

Figure 3. The selected activation domains are active from a remote enhancer position. Hela cells were transfected with 10 μg of β-globin reporter plasmids, 1 μg of expression vectors encoding the GAL4 fusion proteins and 1 μg of the reference plasmid OVEC-REF. β-globin RNA transcribed from the reporter plasmid was isolated from Hela cells, hybridized to a 93 nt S1 probe covering position −18 to +75 of a noncoding region of the β-globin reporter gene and mapped using S1 nuclease (43). Relative transcript levels were quantified by phosphorimaging and corrected for the reference signal. β init indicates the correctly intitiated transcripts of the reporter gene and ref indicates transcripts of the reference gene OVEC-REF. P, undigested S1 probe. M, HpaII digested pBR322 marker DNA. **A)** Transcriptional activity of isolated GAL4 fusion clones tested on the reporter plasmid 5GOVEC containing five GAL4 sites in a proximal promoter position. **B)** The same GAL4 fusion clones were tested for their ability to activate transcription from a downstream enhancer position (+1764) using the reporter plasmid 2Sp1/β-5GAL.

A novel transcriptional activation domain within viral protein BICP4 is specifically selected from BHV-1 DNA

We also searched for activation domains in a 16.2 kb immediate early region of the bovine herpes virus 1 genome (27). From the first round of selection we obtained 133 clones from a library of 12 000, three of which showed strong activity when tested individually (Table III). Strikingly, all three clones contained an insert derived from the same region within the 16.2 kb viral fragment, fused in the same reading frame to the GAL4 DNA binding domain. As depicted in Table 1, the activation domain of the three clones is derived from the N-terminus of BICP4, the bovine homologue of herpes simplex virus ICP4. This region is less conserved than other regions of herpesvirus ICP4 related proteins (28).

The activator trap system selects for enhancer-type activation domains

In the activator trap assay transactivation levels are determined only indirectly via DNA replication. We therefore decided to corroborate the transactivation capacity of the selected clones with quantitative RNA mapping using an S1 nuclease assay. The ability of the fusion proteins to activate a reporter gene containing five GAL4 DNA binding sites in a proximal promoter position was tested (Figure 2). As shown in Figure 3A, all of the selected fusion proteins activated transcription in this assay. The clone GAL4(1−93)-PRIEV-9 with the pseudorabies virus immediate early domain was almost four times as strong as GAL4-VP16. The GAL4 fusion proteins with the activation domain from the BICP4 protein were at least as active as the GAL4-VP16 when tested from a promoter position (summarized in Table II).

The viral activation domains selected by the activator trap all have a high content of negatively charged residues and contain a pattern of alternating acidic and hydrophobic residues that is reminiscent of a minimal synthetic activation domain deduced recently (29, 30, see also Discussion). This synthetic domain also works over large distances, i.e. it has the characteristics of an enhancer-type domain (4). In order to see whether the structural similarity of the isolated viral domains correlated with enhancer-type function, we tested the isolated clones on a reporter construct with five GAL4 DNA binding sites in a remote enhancer position downstream of the β-globin reporter gene (Figure 2). All of the fusion proteins were active from a downstream enhancer position (Figure 3B). Again the activation domains from the pseudorabies virus IE protein and BICP4 protein were more active than the strong activator GAL4-VP16, with the PRV IE domain being four times as active as the VP16 domain. The clone GAL4-BICP4-15 was somewhat less active than the other two BICP4 clones. This clone was found to contain a long stretch

of additional amino acids encoded by the normally untranslated 5′ leader sequence that may have impaired function (Table I).

DISCUSSION

Transcriptional activation domains are specifically selected in mammalian cells

We have developed a novel selection scheme for the isolation and characterization of transcriptional activation domains in mammalian cells. The feasibility of the assay was established by using sonicated DNA containing the activation domain from VP16. We used this strategy to select for very strong activation domains from the immediate early regions of PRV and BHV-1. The activation domain derived from PRV was selected from random fragments of a 10 kb plasmid encoding the entire major immediate early protein of PRV. This 127 aa domain encompasses the N terminus of the major 180 kDa immediate early protein, including 34 aa that were previously implicated in transcriptional activation by classical deletion analysis (26). From BHV-1, we have examined a 16.2 kb segment of the immediate early region, and repeatedly obtained overlapping protein-coding segments from the BICP4 N-terminus. The fact that the same region of the BICP4 protein was recovered in three independent clones underlines the specificity of our selection system.

Both the PRV IE protein and BICP4 are homologous to ICP4 of herpes simplex virus. ICP4 homologues serve a dual function as activators and represssors of herpesviral gene function (18). PRV IE, like other ICP4 members, can repress its own promoter in an autoregulatory manner (31−33), but activates the group of viral early promoters and also virtually any viral or cellular promoter when tested in transfection experiments (24, 25, 34). This raises the question of how PRV IE in infected cells manages to activate viral promoters without promiscuously activating a plethora of cellular promoters. We have previously proposed a model to account for this seeming paradox (25). As an activator, PRV IE may function in some cases without binding directly to DNA and thus would qualify as an adaptor. Among the activation domains we have selected, that of PRV is particularly strong. Repeatedly it was about four times as active as the one of VP16. Thus the PRV IE activation domain is the strongest activation domain we have encountered so far.

Common features of activation domains

Activation domains of mammalian transcription factors can be subdivided into at least two functional classes (4). One, exemplified by the glutamine-rich domains of Oct and Sp1 factors, mediates transcriptional activation only from a promoter position, and in response to an enhancer. The other, exemplified by the widely used acidic domain from the viral activator VP16, has the ability to activate from remote enhancer as well as from proximal promoter positions. We have previously found that the domains from GAL4, ITF-2, TFE3 and MTF-1 belong to this latter enhancer-type activator class. All of the viral domains isolated by the activator trap can be classified as 'enhancer-type' domains because of their ability to activate over a long distance. From a detailed analysis of the VP16 activation domain it is evident that hydrophobic amino acids, notably phenylalanine and leucine, are at least as crucial for activation function as the acidic residues (29, 30, 35). In particular, we have found that multiple copies of a minimal core motif with alternating acidic and hydrophobic aminoacids (4×DDFDL) and even the simplified version (4×DDLDL) can act as strong enhancer domains when fused to the GAL4 DNA binding domain (30). Interestingly, the only regions within the PRV IE and the bovine BICP4 proteins with similarity to the minimal motif (4×DDFDL) were also the ones selected in our assay and shown to be transcriptionally active, while other BICP4 regions with patches of acidic residues were apparently not active (Table I). In addition, the activation domain from the IE62 protein from Varicella zoster virus, another ICP4 homologue, is also located at the N terminus and contains a pattern of acidic and hydrophobic residues reminiscent to the VP16 activation domain (36). This structural similarity may indicate that the activation domains from these proteins contact the same partners in the transcription apparatus. In agreement with this idea, VP16 interacts with the basal cellular transcription factor TFIIB *in vitro* (1, 37, 38). ICP4 from herpes simplex virus seems to even form a triple complex with TFIIB and TBP, and the ability to form a complex correlates with ICP4 activator function (39). Further studies will have to show whether the simple activation 'core' motif from VP16 and similar motifs in the activation domains of PRV IE protein and BICP4 are responsible for direct contacts to the basal transcription machinery. However, it should also be mentioned that, despite seeming similarities, there can also be qualitative differences between acidic domains. The 78 aa activation domain of VP16 can be divided into independently active N and C terminal subdomains with similar amino acid pattern (4). From these, only the C terminal part contacts Drosophila $TAF_{II}40$, a subunit of TFIID whereas each half-domain seems to be important for complex formation with TFIIB (38). In addition, only the C terminal 40 aa subdomain functionally interacts with the yeast ADA2 protein, whereas the N terminal half does not (S.J. Triezenberg, personal communication).

Why selection in mammalian cells?

Brent and and Ptashne (1985) originally demonstrated the validity of the domain concept for transcriptional activators by reconstructing a transcriptional activator in yeast through fusion of a prokaryotic lexA DNA binding domain to the activation domain of GAL4 (40). They also exploited the yeast system to select for fortuitous activation domains present in the genome of *E.coli*. About 1% of random *E.coli* DNA fragments rescued reporter gene expression in yeast, as measured by blue colony staining for β galactosidase activity. When subsequently analyzed, the common structural feature among these domains was a preponderance for negatively charged residues, irrespective of whether they actually encoded parts of genuine *E.coli* proteins or were in the wrong reading frame (7). By contrast, our selection system seems to be more specific. All of the positive clones that we have isolated with our mammalian system encode domains from a particular region within a long viral reading frame. Fragments such as prokaryotic vector sequences, that could not possibly stem from a transcriptional regulator protein, and out-of-frame segments of viral provenance, were not isolated.

We have not as yet ascertained the limits of the factor trap assay described here. Certainly, it is not constrained to viral activators: in a reconstruction experiment, the enhancer-type activation domains from the eukaryotic factors TFE3 and GAL4 were highly amplified in CV-1-5GT cells, as measured by transformation of *E.coli* with DpnI resistant plasmids (data not shown). However, promoter-type activation domains (4), or

domains not active in SV40 T antigen-expressing CV-1 cells, would not be isolated under the present conditions. To circumvent these problems, other target gene constructions, other cell types and/or other viral replicons may have to be used.

We think that the mammalian expression system has major advantages for the detection of transactivation domains. The absence of a mammalian cell-specific component could prevent cloning of some mammalian activation domains in yeast. Also some activation domains have to be properly phosphorylated in order to become active in a mammalian cell (reviewed in 41). Indeed, we have recently isolated an activation domain from human cytomegalovirus DNA, that functions in monkey CV-1 and HeLa cells but not in yeast (in collaboration with T. Stamminger and B. Fleckenstein, unpublished data). Selectivity for activation in mammalian cells was also observed with a domain derived from the mammalian basal transcription apparatus (M. Gstaiger and W. Schaffner, unpublished data). Recently we also demonstrated that promoter-type activation domains (4) are inactive in yeast (12). Whether this selectivity is based on mammalian cell type-specific posttranslational modification and/or dependence on mammalian adaptor proteins is not clear. Taken together these findings have convinced us that the best milieu for selection of mammalian activation domains is the mammalian cell.

The activator trap system should be particularly useful to identify and precisely delineate an activation domain within a known factor, as well as for the identification of mammalian transcription factors that do not bind DNA, such as coactivators/adaptors which in most cases can be expected to escape the conventional detection methods.

ACKNOWLEDGEMENTS

We are indebted to Dr Thomas Stamminger for valuable discussions and helpful suggestions. We thank Drs Christopher Hovens, Pamela Mitchell, Philippe Douville and Martin Schwyzer for critical reading of the manuscript, and Fritz Ochsenbein for excellent artwork. This work was supported by the Kanton of Zurich.

REFERENCES

1. Roberts, S. G., Ha, I., Maldonado, E., Reinberg, D., and Green, M. R. (1993) *Nature* **363**, 741–744.
2. Koleske, A. and Young, R. (1994) *Nature* **368**, 466–469.
3. Tjian, R. and Maniatis, T. (1994) *Cell* **77**, 5–8.
4. Seipel, K., Georgiev, O., and Schaffner, W. (1992) *EMBO J.* **11**, 4961–4968.
5. Baker, A. and Schatz, G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3117–3121.
6. Kaiser, C. A., Preuss, D., Grisafi, P. and Botstein, D (1987) *Science* **235**, 312–317.
7. Ma, J. and Ptashne, M. (1987) *Cell* **51**, 113–119.
8. Hahn, S. (1993) *Cell* **72**, 481–483.
9. Struhl, K. (1991) *Current Biology* **1**, 188–191.
10. Donaldson, L. and Capone, J. P. (1992) *J. Biol. Chem.* **267**, 1411–1414.
11. Van Hoy, M., Leuther, K. K., Kodadek, T., and Johnston, S. A. (1993) *Cell* **72**, 587–594.
12. Künzler, M., Braus, G., Georgiev, O., Seipel, K., and Schaffner, W. (1994) *EMBO* **13**, 641–645.
13. Xu, L., Rungger, D., Georgiev, O., Seipel, K., and Schaffner, W. (1994) *Biol. Chem. Hoppe-Seyler* **375**, 105–112.
14. Weber, F., de Villiers, J., and Schaffner, W. (1984) *Cell* **36**, 983–992.
15. Tognoni, A., Cattaneo, R., Serfling, E., and Schaffner, W. (1985) *Nucl. Acids Res.* **13**, 7457–7472.
16. Triezenberg, S. J., Kingsbury, R. C., and McKnight, S. L. (1988) *Genes Dev.* **2**, 718–729.
17. Rusconi, S., Severne, Y., Georgiev, O., Galli, I., and Wieland, S. (1990) *Gene* **89**, 211–221.
18. Hayward, S. (1993) *Semin Virol* **4**, 15–23.
19. Stern, S., Tanaka, M., and Herr, W. (1989) *Nature*po **341**, 624–630.
20. Thompson, C. and McKnight, S. (1992) *TIG* **8**, 232–235.
21. Wilson, A. C., LaMarco, K., Peterson, M.G. and Herr, W. (1993) *Cell* **74**, 115–125.
22. Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988) *Nature* **335**, 563–564.
23. Ben Porat, T., Veach, R. A., and Hampl, H. (1983) *Virology* **124**, 411–424.
24. Green, M. R., Treisman, R., and Maniatis, T. (1983) *Cell* **35**, 137–148.
25. Thali, M., Rusconi, S., and Schaffner, W. (1990) *J. Mol. Biol.* **215**, 301–311.
26. Martin, K. J., Lillie, J. W., and Green, M. R. (1990) *Genes Dev.* **4**, 2376–2382.
27. Wirth, U., Vogt, B., and Schwyzer, M. (1991) *J. Virol.* **65**, 195–205.
28. Schwyzer, M., Vlcek, C., Menekse, O., Fraefel, C., and Paces, V. (1993) *Virology* **197**, 349–357.
29. Regier, J. L., Shen, F., and Triezenberg, S. J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 883–887.
30. Seipel, K., Georgiev, O., and Schaffner, W. (1994) *Biol. Chem. Hoppe-Seyler* **375**, 963–970.
31. Muller, M. T. (1987) *J. Virol.* **61**, 858–865.
32. Disney, G. H., McKee, T. A., Preston, C. M., and Everett, R. D. (1990) *J. Gen. Virol.*, **71**, 2999–3003.
33. Vlcek, C., Kozmik, Z., Paces, V., Schirm, S., and Schwyzer, M. (1990) *Virology* **179**, 365–377.
34. Abmayr, S. M., Workman, J. L., and Roeder, R. G. (1988) *Genes Dev.* **2**, 542–553.
35. Cress, W. D. and Triezenberg, S. J. (1991) *Science* **251**, 87–90.
36. Perera, L. P., Mosca, J. D., Ruyechan, W. T., Hayward, G. S., Straus, S. E., and Hay, J. (1993) *J. Virol.* **67**, 4474–4483.
37. Lin, Y. S., Ha, I., Maldonado, E., Reinberg, D., and Green, M. R. (1991) *Nature* **353**, 569–571.
38. Goodrich, J., Hoey, T., Thut, C., Admon, A., and Tjian, R. (1993) *Cell* **75**, 519–530.
39. Smith, C. A., Bates, P., Rivera, G. R., Gu, B., and DeLuca, N. A. (1993) *J. Virol.* **67**, 4676–4687.
40. Brent, R. and Ptashne, M. (1985) *Cell* **43**, 729–736.
41. Hunter, T. and Karin, M.(1992) *Cell* **70**, 375–387.
42. Carey, M., Leatherwood, J., and Ptashne, M. (1990) *Science* **247**, 710–712.
43. Westin, G., Gerster, T., Muller, M. M., Schaffner, G., and Schaffner, W. (1987) *Nucl. Acids Res.* **15**, 6787–6798.
44. Seipel, K., Georgiev, O., and Schaffner, W. (1993) *Nucl. Acids Res.* **21**, 5609–5615.
45. Gerster, T., Matthias, P., Thali, M., Jiricny, J., and Schaffner, W. (1987) *EMBO J.* **6**, 1323–1330.
46. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
47. Serfling, E., Lubbe, A., Dorsch, H. K., and Schaffner, W. (1985) *EMBO J.* **4**, 3851–3859.
48. Weber, F. and Schaffner, W. (1985) *Nature* **315**, 75–77.
49. He, X. and Rosenfeld, M. (1991) *Neuron* **7**, 183–196.
50. Southern, P. J. and Berg, P. (1982) *J Mol Appl Genet* **1**, 327–341.
51. Maniatis, T., Fritsch, E., and Sambrook, J. (1989) Molecular Cloning: A Laboratory Manual. *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.*