Promoter-specific Activation of Gene Expression Directed by Bacteriophage-selected Zinc Fingers

Y. Choo¹*, A. Castellanos², B. García-Hernández², I. Sánchez-García² and A. Klug¹

¹Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.
²Departamento de Proliferación-Diferenciación, Instituto de Microbiología Genética CSIC/ Universidad de Salamanca, Salamanca-37007 Spain

It has been shown that sequence-specific DNA-binding domains containing zinc fingers can be selected from libraries displayed on filamentous bacteriophage. The affinity and specificity of these peptides are well characterised in vitro, but few data are available to demonstrate specific DNA binding and discrimination between closely related DNA sequences in vivo. Transient transactivation assays were performed in mammalian cells, using expression plasmids which produce different amounts of a model transcription factor containing a phage-selected zinc finger DNA-binding domain, and reporter plasmids which carry systematic variations of the promoter sequence. When the intracellular concentration of the transcription factor was appropriate, activation of gene expression was absolutely dependent on a promoter having the same DNA sequence as that originally used to select the zinc finger domain by phage display. However, excessive intracellular concentrations of the transcription factor resulted in some less-specific DNA binding, leading to gene activation from similar promoters containing a maximum of two base changes. Thus, provided delivery is carefully controlled, highly specific control of gene expression in vivo can be achieved using artificial transcription factors containing phage-selected zinc finger DNA-binding domains.

Keywords: intermolecular recognition; DNA-binding protein; gene therapy; phage display; zinc finger

Introduction

The differentiation and development of an organism is effected by a programme of differential gene expression which is ordered in space and time; this is controlled mainly at the level of transcription, by protein transcription factors which bind to regulatory promoter DNA sequences found adjacent to specific genes (Mitchell & Tjian, 1989; Nikolov & Burley, 1997; Ptashne & Gann, 1990; Roeder, 1991). Artificially manipulating this programme would allow us to dissect developmental pathways, and to interfere with the aetiology of disease. Unfortunately, naturally occurring transcription factors are not normally suitable for applications where specific control over one gene is required, because the DNA sequences they bind are present in the promoters of many different genes (Kel et al., 1995). In addition, gene regulation by natural transcription factors often requires the concerted action of combinations of factors, which by themselves have little or no activity (Herschlag & Johnson, 1993; Milos & Zaret, 1992; Reith et al., 1994; Rivera et al., 1993).

One approach to the artificial control of gene expression is to use transcription factors containing potent activation or repression domains, and engineered with new DNA-binding specificities for rare binding sites associated only with the target gene, but probably distinct from the natural promoter. This strategy is feasible because transcription factors are structurally modular, with different domains responsible for DNA binding and transcriptional regulation (Mitchell & Tjian, 1989; Pabo & Sauer, 1992). These domains are often interchangeable such that chimaeric transcription factors with swapped characteristics can be easily engineered (Brent & Ptashne, 1985).

The manipulation of the DNA-binding specificity of transcription factors is less straightforward. Efforts in this field have focused on the zinc finger of the TFIIIA type: a small DNA-binding module found in eukaryotic transcription factors (Klug &
Rhodes, 1987; Miller et al., 1985). Zinc finger domains which bind to given DNA sequences in vitro can be selected by bacteriophage display (Choo & Klug, 1994a; Jamieson et al., 1994; Rebar & Pabo, 1994; Greisman & Pabo, 1997; for a review see Choo & Klug, 1995) and their specificity later assessed by screening for binding to 12 self-encoding DNA oligonucleotide libraries (Choo & Klug, 1994b). The DNA-binding characteristics of some phage-selected zinc fingers have been investigated in vitro, confirming that they function as well as their naturally occurring counterparts (Choo & Klug, 1994b; Jamieson et al., 1994; Rebar & Pabo, 1994). While the affinity of such phage-selected DNA-binding domains is quite high, usually falling in the nanomolar range for a three-finger peptide, the observed discrimination against closely related DNA sequences, by a two to tenfold increase in the apparent $K_d$ per point mutation, is rather modest.

Hence, here we have questioned whether DNA binding by a phage-selected zinc finger domain is sufficiently discriminating to allow operations on specific DNA sequences in vivo. We used a CAT transactivation assay to investigate the ability of a model phage-selected DNA-binding domain to discriminate its target from a set of closely related promoter sequences. We found that this was possible with a high degree of specificity, but only when the intracellular concentration of zinc finger transcription factor was optimised. These results warrant the use of phage-selected DNA-binding domains to direct in vivo DNA manipulation, in particular the control of gene expression, but invoke the need for controlled delivery of the DNA-binding domain in order to achieve specific effects.

Results

**In vitro DNA specificity of phage-selected zinc fingers**

A phage display library of the three-finger DNA-binding domain of Zif268 has been described, in which fingers 1 and 3 have the wild-type sequence but finger 2 is randomised in eight positions which could function in DNA binding (Choo & Klug, 1994a). Affinity selections from this library were previously performed using the nine base-pair long Zif268 binding site, GCGTGGCCG, in which the underlined three-base subsite contacted by finger 2 was substituted by a different triplet. Many selections yielded new DNA-binding domains which bound to variants of the Zif268 binding site (Choo & Klug, 1994b).

One representative from these DNA-binding domains was arbitrarily chosen for further specificity experiments to be performed in vivo. This domain, denoted ZN1 (Figure 1), was originally selected against the sequence GCGTGAGCG from a phage display library based on the three-finger DNA-binding domain of the protein Zif268. The three zinc fingers (F1, F2 and F3) are aligned, regions of secondary structure are marked below the sequence, and key residue positions are numbered above the sequence relative to the first amino acid in the $\alpha$-helix (residue 1). The selected residues which fall in the randomised portion of finger 2 are shown in italics.

**Figure 1.** Amino acid sequence of the zinc finger DNA-binding domain ZN1, which was originally selected using the nucleotide sequence GCGTGAGCG from a phage display library based on the three-finger DNA-binding domain of the protein Zif268. The three zinc fingers (F1, F2 and F3) are aligned, regions of secondary structure are marked below the sequence, and key residue positions are numbered above the sequence relative to the first amino acid in the $\alpha$-helix (residue 1). The selected residues which fall in the randomised portion of finger 2 are shown in italics.

526 In vivo Action of Phage-selected Zinc Fingers

**Figure 2.** Determination of the apparent $K_d$ of ZN1 for the optimally bound DNA sequence GCGTGGCCG by full curve DNA-binding ELISA (○), contrasted to the signal obtained using the binding site of wild-type Zif268 (GCGTTGGCCG) in the same assay (■).
**In vivo DNA binding and activation of transcription**

The transient transactivation assay of Gorman et al. (1982) is commonly used to detect the interaction of a DNA-binding protein with specific DNA sequences in mammalian cells. Two plasmids are typically cotransfected into mammalian cells. One plasmid (the expression vector) produces a transcription factor containing a DNA-binding domain fused to an activation domain, while the second plasmid (the reporter vector) carries a specific DNA sequence of interest upstream of the chloramphenicol acetyl transferase (CAT) reporter gene and its minimal promoter (TATA box). Binding of the expressed transcription factor to the DNA sequence of interest causes activation of the CAT gene, which can be monitored by assaying for enzyme activity.

To investigate the level of synthetic transcription factor required for specific activation of gene expression, we used two expression vectors which produce different quantities of ZN1 fused to the acidic transcription activation domain from VP16 (Fields, 1993) and an 11 amino acid epitope tag derived from c-myc (Evan et al., 1985). The first was the pEF-BOS expression vector (Mizushima & Shigezaku, 1990) which drives high-level expression of the fusion protein from the strong promoter of human elongation factor 1a (EF1a). The second vector, pESP-SVTEXP (Shyam et al., 1986), drives expression from the relatively weaker SV40 promoter, resulting in lower intracellular levels of the fusion protein (Mizushima & Shigezaku, 1990). Western blotting of extracts containing equal amounts of protein taken from cells transfected with the two expression vectors reveals that there is roughly a 15-fold difference in the amount of transcription factor produced by the respective expression vectors (data not shown).

The reporter vector used was pMCAT3, which contains the metallothionein promoter region upstream of the CAT gene (Luscher et al., 1989). Binding sites cloned into the reporter plasmid were of the generic sequence GCGXXXGCG, where the underlined nucleotides corresponding to the sub-site of the selected middle finger were varied. One set of reporter vectors carried only one copy of such a binding site, to test whether transcription can be controlled by a single molecule of the artificial transcription factor bound to the promoter. Another set of reporters which carried six consecutive binding sites (thus amplifying the signal from bound transcription factor) was used to detect transcription resulting from low levels of non-specific binding.

**Transcription activation from a single binding site**

Figure 3 (see also Table 1) shows the CAT assay result obtained in C3H10T1/2 cells, with the high-level expression vector (pEF-BOS) and a reporter plasmid bearing one copy of the transcription factor binding site GCGTGAGCG. Cotransformation of expression and reporter plasmids resulted in a strong induction of CAT activity relative to controls where the expression plasmid was cotransfected with pMCAT3, or where pMCAT3 was transfected alone. The result demonstrates that the phage-selected zinc fingers are capable of binding in vivo to the sequence used in their isolation by in vitro methods. Furthermore, when incorporated into engineered transcription factors this binding directs the downstream events which cause transcriptional activation of target genes. To test the specificity of transcription activation, the experiment was repeated using reporters with single copy binding sites containing the middle triplets TAA, GAT or TAT, which gave no significant increase over the background level of CAT expression (Table 1). Therefore, one copy of a
Table 1. The absolute values plotted in Figures 3, 4 and 5

<table>
<thead>
<tr>
<th>Expression and reporter plasmids</th>
<th>pBOS 1 site</th>
<th>pBOS 6 sites</th>
<th>pESP-SVTEXP 1 site</th>
<th>pESP-SVTEXP 6 sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>none†</td>
<td>0.029</td>
<td>0.071</td>
<td>0.036</td>
<td>0.029</td>
</tr>
<tr>
<td>none†</td>
<td>0.034</td>
<td>0.058</td>
<td>0.029</td>
<td>0.029</td>
</tr>
<tr>
<td>TGA</td>
<td>9.767</td>
<td>44.947*</td>
<td>10.537</td>
<td>5.037</td>
</tr>
<tr>
<td>CGA</td>
<td>ND</td>
<td>7.563</td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td>GGA</td>
<td>ND</td>
<td>7.842</td>
<td>0.031</td>
<td>0.031</td>
</tr>
<tr>
<td>AGA</td>
<td>ND</td>
<td>11.921</td>
<td>0.079</td>
<td>0.079</td>
</tr>
<tr>
<td>TCA</td>
<td>ND</td>
<td>10.623</td>
<td>0.025</td>
<td>0.025</td>
</tr>
<tr>
<td>TTA</td>
<td>ND</td>
<td>4.924</td>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>TAA</td>
<td>0.043</td>
<td>5.238</td>
<td>0.026</td>
<td>0.026</td>
</tr>
<tr>
<td>TGC</td>
<td>ND</td>
<td>5.041</td>
<td>0.035</td>
<td>0.035</td>
</tr>
<tr>
<td>TGT</td>
<td>ND</td>
<td>12.412</td>
<td>0.082</td>
<td>0.082</td>
</tr>
<tr>
<td>TGG</td>
<td>ND</td>
<td>5.217</td>
<td>0.037</td>
<td>0.037</td>
</tr>
<tr>
<td>GAA</td>
<td>ND</td>
<td>3.248</td>
<td>0.034</td>
<td>0.034</td>
</tr>
<tr>
<td>GCA</td>
<td>ND</td>
<td>3.456</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>GTC</td>
<td>ND</td>
<td>0.123</td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td>GTG</td>
<td>0.033</td>
<td>2.989</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>GTC</td>
<td>ND</td>
<td>0.024</td>
<td>0.037</td>
<td>0.037</td>
</tr>
<tr>
<td>GTT</td>
<td>ND</td>
<td>0.073</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>CAT</td>
<td>0.036</td>
<td>0.104</td>
<td>0.022</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Values correspond to the amount of [14C]chloramphenicol converted by a fixed amount of cell extract following induction of CAT activity using a given combination of expression and reporter plasmids. These vector combinations are given above the three right-hand columns, which list the amount of conversion in cpm ((x10^6), while the first column lists the middle triplets inserted into the generic binding site GCGXXXGCG present (in one or six copies) in the promoter of the reporter plasmid. Also given are values for two controls corresponding to the use of pMCAT3 without a binding site for the zinc fingers, transfected into cells alone (none†), or together with the appropriate expression vector for each experiment (none†). Converted chloramphenicol was measured while the CAT assay was in the linear range, except for that value marked with an asterisk, in which the enzyme activity is therefore likely to be somewhat underestimated.

designed zinc finger transcription factor bound to an upstream promoter can regulate specific gene expression.

Effect of DNA-binding protein concentration on specificity in vivo

CAT assays using one copy of the zinc finger binding site suggest either that mutant sequences do not interact appreciably with ZN1, or that weak binding does not result in measurable transcriptional activation. To examine the DNA-binding specificity of ZN1 in greater detail, the transactivation assay was made more sensitive by using reporter plasmids which contain six copies of the same binding site upstream of the CAT gene. Figure 4 (see also Table 1) shows that using the high-level expression vector (pEF-BOS) and a reporter bearing six copies of the sequence GGCTGAGGCG, an even greater increase in the level of CAT expression was obtained. Sequence-specificity was investigated using nine other such reporter constructs which systematically encompass the possible single base-pair deviations from the optimal triplet TGA, and a small number of additional reporters in which the binding sites contained two base changes or completely different middle triplets. In these cases, significant activation of transcription can be detected from promoters with one, or even two, base deviations from the optimal 9 bp sequence. The mutations caused cumulative decreases in transcriptional activation in that promoters with only a single base change were capable of stronger activation than those with two base changes, while no CAT activation was observed from promoters with three base changes (Figure 4). The data shown in Table 1 and Figure 4 were derived from transfections of C3H10T1/2 cells, but similar data (not shown) were obtained in COS1 cells, meaning that the results are not particular to the use of any one cell line. Thus, in contrast to the activation studies performed using one copy of the zinc finger binding site as a promoter, multiply repeated promoters reveal degrees of non-specific occupancy of closely related binding sites in the presence of high concentrations of DNA-binding domain.

Non-specific binding of proteins to DNA can be reduced by limiting the concentration of the reagents to the Kd value for the specific reaction. Accordingly, the observed activation of gene expression from non-specific promoters might be eliminated by reducing the intracellular concentration of artificial transcription factor. The experiment of Figure 4 was therefore repeated using the vector pESP-SVTEXP to drive low-level expression of ZN1. Figure 5 shows that under these conditions, activation is reduced but occurs only from GGCTGAGGCG: any deviations from this sequence completely suppress CAT expression. Hence limiting the level of the artificial transcription factor in the cell reduces binding to non-specific DNA sequences, abolishing transcription activation from mutant promoters.

Discussion

Promoter-specific activation of gene expression has been described in cell culture, using a representative phage-selected DNA-binding domain which was converted to a transcription factor by fusion with the activation domain from VP16. Significant activation of gene expression was recorded by the CAT assay when one binding site for the artificial transcription factor was present upstream of the reporter gene, and under these circumstances no non-specific activation of gene expression was detected from a limited set of promoters. However, a degree of non-specific activation of gene expression was seen when the assay was made more sensitive by using promoters with multiple binding sites containing a series of systematic mutations. With six binding sites, substantial activation of transcription was noted from promoters which carried point mutations. All-or-none activation particular to the correct promoter could.
However be restored when the intracellular concentration of transcription factor was reduced to an appropriate level by the use of a less potent expression vector.

The results show that it is possible to induce gene expression in vivo from specific promoters using phage-selected zinc fingers fused to transcriptional activation domains. The ability of the engineered transcription factor to activate transcription from closely related promoters correlated well with its in vitro DNA-binding specificity. On the contrary, the function in vivo of many naturally occurring transcription factors (including homeo-domains (Hayashi & Scott, 1990), basic region leucine zipper proteins (Lamb & McKnight, 1991) and the prototype Gal4 transcription factor (Vashee et al., 1993)) cannot be accounted for merely by their in vitro DNA-binding preferences, probably because they often require additional proteins to form the complexes which regulate gene expression. Thus while the actions of natural transcription factors are unpredictable a priori, those of artificial factors are a simple extension of their well-established in vitro DNA-binding characteristics. Hence it follows that transcription factors such as those we have described are preferable to natural factors for the artificial control of gene expression.

Even so, the specificity of transcriptional control clearly depends on the intracellular concentration of transcription factor. According to the law of mass action, high transcription factor concentration favours promoter binding, leading to stronger activation, but at the same time increases binding to closely related DNA sequences. Excessive intracellular concentrations of protein can therefore result in a loss of DNA-binding specificity and aberrant gene activation. Our results reflect this and highlight the necessity for fine tuning of the transcription factor concentration such that only specific complexes are formed in vivo. In applications of artificial gene regulation this can be achieved using appropriate expression vectors, as here, or by means of alternative delivery protocols which attain the right balance between maximal activation of the target gene and a minimum effect on other genes.

Differential regulation of very similar promoters by proteins containing phage-selected zinc fingers has been demonstrated above. In the CAT assay system used, where the only variable in each experiment is the 9 bp sequence of the zinc finger...
binding site in the reporter plasmids, these results indicate highly specific protein-DNA interactions \textit{in vivo}. This raises the possibility of using zinc fingers to discriminate between normal and point-mutated DNA, for example cellular proto-oncogenes and transforming mutants which differ by a single base-pair. A referee has made the valid point that a useful next step might be to ascertain whether the highly specific interactions we have noted using the transient activation assay can be reproduced in cells where the reporters are stably integrated in chromosomes. Although such data are not yet available, we have previously shown that a designed zinc finger protein can inhibit the expression of a stably integrated fusion oncogene expressed in a mouse cell line (Choo \textit{et al}. , 1994).

An issue related to the specificity of DNA binding \textit{in vivo} is the length of the target binding site. At present, zinc finger design has been based on three-finger domains which directly contact at most ten base-pairs (Elrod-Erickson \textit{et al}. , 1996; Isalan \textit{et al}. , 1997; Swirnoff & Milbrandt, 1995). While binding sites of this length would normally be sufficient to specify a unique locus within any given gene, a complex genome would be predicted to contain thousands of identical binding sites. Hence to target a unique DNA site in the genome would require a DNA-binding domain which specifies 13 to 16 base-pairs, i.e. containing four or five zinc fingers. However, while longer arrays of zinc fingers might bind to rarer binding sites, the discrimination against any mutations in those sites might be expected to be relatively weaker. As with the intracellular concentration of transcription factor, which must be limited to a “window” which allows for specific high-level transcriptional activation, a compromise will probably be required between the need to bind long stretches of DNA and to discriminate strongly between similar sequences, which in practice will depend on the application.

\textbf{Materials and methods}

\textit{In vitro} DNA-binding assays

Zinc finger phage selections and binding site signatures were performed as described (Choo & Klug, 1994 a,b). The method for the calculation of apparent $K_d$ for ZNI was exactly as described by Choo & Klug (1994b).

\textbf{Eukaryotic expression constructs}

Genes to be cloned into expression plasmids were generated by the polymerase chain reaction (PCR) using phage DNA as the template. PCR primers included appropriate initiation and termination codons and restriction sites for cloning. Expression constructs were generated by inserting in-frame fusions between the activation domain of herpes simplex virus VP16 (Fields, 1995) and zinc fingers into the NotI site of pEF-BOS.
(Mizushima & Shigezaku, 1990), or between the HindIII and XhoI sites of pESP-SVTEXP (Shyam et al., 1986).

**Reporter plasmids**

The reporter plasmids were constructed by inserting one or six copies of the target sequence into the unique BgIII site of the reporter vector pMCAT3 (Luscher et al., 1989). Reporter plasmids with one copy of the target sequence were as follows: p1MCAT3-ST1 (5'-GGCTGAAGCG-3'). p1MCAT3-2 (5'-GGCTGAAAGCG-3'), p1MCAT3-ST2 (5'-GGCTGAAGCG-3') and p1MCAT3-MST2 (5'-GGCTGAAGCG-3'). Reporter plasmids containing six tandem copies of the target sequence were: pMCAT3-ST1 (5'-GGCTGAAGCG-3'), pMCAT3-6 (5'-GGCTGAGACCG-3'), pMCAT3-7 (5'-GGCGTACGGC-3'), pMCAT3-8 (5'-GGCTGAAGCG-3'), pMCAT3-1 (5'-GGCTGACGGC-3'), pMCAT3-MST1 (5'-GGCTGAAAGCG-3'), pMCAT3-2 (5'-GGCTGAAAGCG-3'), pMCAT3-3 (5'-GGCTGACGGC-3') and pMCAT3-13 (5'-GGGCAGACC-3'). pMCAT3-14 (5'-GGCTGACGGC-3'), pMCAT3-9 (5'-GGCTGACGGC-3'), pMCAT3-10 (5'-GGCTGTCGCC-3'), pMCAT3-3 (5'-GGCTGACGGC-3'), pMCAT3-5 (5'-GGCTGACGGC-3'), pMCAT3-11 (5'-GGCTGACGGC-3'), pMCAT3-12 (5'-GGCTGACGGC-3') and pMCAT3-ST2 (5'-GGCTGACGGC-3').

**CAT transactivation assays**

C3H10T1/2 and COS1 cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 5% (v/v) foetal bovine serum (FBS), penicillin G (100 units/ml) and streptomycin sulphate (100 µg/ml). C3H10T1/2 and COS1 cells were transiently co-transfected with 10 ng of the reporter plasmid and 10 µg of the expression vector. The DNA used in transfections was prepared by the standard alkaline lysis method. Cells were trypsinized and plated at 2 × 10^5 cells per 100 mm diameter plate; on the following day, cells were supplied with fresh medium, and at three hours later calcium phosphate-DNA coprecipitates (1 ml per 100 mm dish) were added. Following 16 to 18 hours, the precipitates were removed, the cells were washed twice in PBS and given complete medium. After 48 hours, the cells were harvested for CAT and luciferase assays as described (de Wet et al., 1987; Gorman et al., 1982). The vector RSVL (de Wet et al., 1987), which contains the Rous sarcoma virus long terminal repeat linked to luciferase, was used as an internal control to normalize for differences in transfection efficiency.

All assays were performed as duplicate independent transfections. Equivalent amounts of protein (15 to 25 µg as determined with the Bio-Rad protein kit and bovine serum albumin as a standard) were used in each CAT conversion reaction, which proceeded for one hour such that values were within the linear range of the assay. Relative CAT activities were determined by quantifying the amounts of [14C]chloramphenicol present in spots corresponding to the two acetylated species cut out from the thin layer chromatographs. Plasmid pGSEC (which has five consensus 17-mer GAL4-binding sites upstream from the minimal promoter of the adenovirus Elb TATA box) and pM1VP16 vector (which encodes an in-frame fusion between the DNA-binding domain of GAL4 and the activation domain of herpes simplex virus VP16) were used as a positive control of the transient co-transfection assays (Sadowski et al., 1992).

**Acknowledgements**

We thank D. Martin-Zanca for advice, L. Fairall for suggesting improvements to the manuscript, J. Girdlestone for the pMCAT3 vector, I. Sadowski for plasmids pG5EC and pM1VP16, and V.N. Rao for plasmid pESP-SVTEXP. Research in I. S. G.’s laboratory is supported by Fundacion Internacional Jose Carreras (FIJC-94/INT), European Commission (BMHI-C-T96-0375) and Fundación Científica of the AECC.

**References**


Edited by T. Richmond

(Received 9 May 1997; received in revised form 25 July 1997; accepted 7 August 1997)

Note added in proof: Following the submission of this manuscript, polydactyl proteins comprising head-to-tail fusions of two three-finger DNA-binding domains were described by Liu et al. (Liu, Q., Segal, D. J., Ghiara, J. B. & Barbas, C. F., III (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5525–5530).