Promoter-targeted Phage Display Selections with Preassembled Synthetic Zinc Finger Libraries for Endogenous Gene Regulation

Caren V. Lund, Pilar Blancafort, Mikhail Popkov and Carlos F. Barbas III*

The Skaggs Institute for Chemical Biology and Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA

Regulation of endogenous gene expression has been achieved using synthetic zinc finger proteins fused to activation or repression domains, zinc finger transcription factors (TFZFs). Two key aspects of selective gene regulation using TFZFs are the accessibility of a zinc finger protein to its target DNA sequence and the interaction of the fused activation or repression domain with endogenous proteins. Previous work has shown that predicting a biologically active binding site at which a TFZ can control gene expression is not always straightforward. Here, we used a library of preassembled three-finger zinc finger proteins (ZFPs) displayed on filamentous phage, and selected for ZFPs that bound along a 1.4 kb promoter fragment of the human ErbB-2 gene. Following affinity selection by phage display, 13 ZFPs were isolated and sequenced. Transcription factors were prepared by fusion of the zinc finger proteins with a VP64 activation domain or a KRAB repression domain and the transcriptional control imposed by these TFZFs was evaluated using luciferase reporter assays. Endogenous gene regulation activity was studied following retroviral delivery into A431 cells. Additional ZFP characterization included DNase I footprinting to evaluate the integrity of each predicted protein:DNA interaction. The most promising TFZFs able to both up-regulate and down-regulate ErbB-2 expression were extended to six-finger proteins. The increased affinity and refined specificity demonstrated by the six-finger proteins provided reliable transcriptional control. As a result of studies with the six-finger proteins, the specific region of the promoter most accessible to transcriptional control by VP64-ZFP and KRAB-ZFP fusion proteins was elucidated and confirmed by DNase I footprinting, flow cytometric analysis and immunofluorescence. The ZFP phage display library strategy disclosed here, coupled with the growing availability of genome sequencing information, provides a route to identifying gene-regulating TFZFs without the prerequisite of well-defined promoter elements.

q 2004 Elsevier Ltd. All rights reserved.

Keywords: phage display selection; zinc finger; gene regulation; ErbB-2; immunofluorescence

*Corresponding author

Introduction

Transcription factors (TFs), mediator proteins, and chromatin remodeling proteins associate with the promoter regions of genes.1,2 The presence or absence of these proteins determines the timing

Supplementary data associated with this article can be found at doi: 10.1016/j.jmb.2004.04.057

Abbreviations used: TF, transcription factor; ZFP, zinc finger protein; TFZF, zinc finger transcription factor; UTR, untranslated region; ELISA, enzyme-linked immunosorbent assay; FACS, fluorescence-activated cell sorting; KRAB, Kruppel-associated box; HA, hema-glutinin; DAPI, 4',6-diamidino-2-phenylindole dihydro-chloride; P/PMR, polypurine/polypyrimidine mirror repeat; GFP, green fluorescent protein; VP64, tetrameric repeat of herpes simplex VP16's minimal activation domain.

E-mail address of the corresponding author: carlos@scripps.edu

0022-2836/S - see front matter © 2004 Elsevier Ltd. All rights reserved.
and duration of gene transcription. One of the most common classes of TFs contains the DNA-binding zinc finger transcription factors.\(^5,4\) In order to impart artificial control on gene transcription, synthetic zinc finger proteins (ZFPs) have been engineered to bind specific DNA sequences. A subset of these engineered proteins (TF\(_Z\)Fs) has been shown to control endogenous gene expression selectively when fused to activation or repression domains.\(^5-17\) The prediction of successful target sites for TF\(_Z\)Fs in endogenous promoters has been based on ZFP affinity, position within the 5′ untranslated region (UTR), and their location within DNase I-hypersensitive regions.\(^6-8,10\) Application of these criteria, however, has not resulted consistently in specific gene regulation. With the roles of endogenous TFs, mediator proteins, and chromatin remodeling proteins largely uncharacterized for any given promoter, a more empirical method of sampling binding sites across a promoter is required to achieve robust, imposed gene regulation.

Synthetic ZFPs have been derived from the framework of naturally occurring zinc finger TFs, including Sp1 and the murine Zif268, which have three zinc finger domains.\(^5\) The C\(_{2}\)H\(_{2}\) zinc finger domain consists of an α helix packed against two antiparallel β sheets and stabilized via coordination of a zinc ion. The side-chains of amino acid residues at the terminus of the β-sheet and continuing into the α helix, constitute the DNA reading-head of the domain that is responsible for sequence-specific interaction with the DNA. A single zinc finger domain typically recognizes 3 bp of DNA, primarily through major groove interactions. Synthetic ZFPs typically consist of a repeat of three to six zinc finger domains, allowing the resulting polydactyl proteins to recognize DNA sequences of 9–18 bp. Transcription factors have been constructed by fusion of an activation or repression domain to the ZFP, creating TF\(_Z\)Fs. Naturally occurring activators and repressors, or variants thereof, are commonly used with ZFPs, particularly the VP16 activation domain and the Krüppel associated box (KRAB).\(^5,18-20\)

The target of imposed gene regulation is not simply DNA, but rather the three-dimensional landscape of chromatin structure constructed from DNA:protein interactions and protein:protein interactions along a promoter. The linear sequence of DNA provides little indication of a promoter’s dynamic topology. Eukaryotic gene promoter sequences often stretch over many thousands of bases, and include both positive and negative regulatory elements. Genome-wide characterization of promoters employing deletion constructs, linker scanning, computational TF binding site analyses, and TF isolation, is ongoing.\(^21-25\) In addition, experiments have begun to address the chromatin state of promoters, the location of nucleosomes, and the nature of histone modifications.\(^26,27\) In spite of this information, the understanding of the precise mechanism of transcriptional control is far from complete for any promoter.

Here, we have studied the application of phage display to provide ZFP-based TFs that bind throughout a given promoter sequence with the aim of applying these proteins to endogenous gene regulation. This strategy resulted in the identification of a site at which a three-finger or a six-finger TF\(_Z\)F could provide regulation of ErbB-2 expression in human epidermoid carcinoma A431 cells. The site identified would not have been predicted from the linear sequence of the promoter, known TF binding sites, or chromatin characterization, suggesting that this approach can be used to synthesize gene regulators without prior characterization of a target gene’s promoter.

**Results**

**Phage display optimization and selection against an ErbB-2 promoter fragment**

Previous studies of zinc finger proteins using phage display have focused on the selection of zinc finger variants of altered specificity. These studies have utilized short pieces of DNA as the target for selection, typically oligonucleotides that present the 9 bp binding site for a three-finger protein.\(^9,28-30\) Our goal was to select zinc fingers that bound optimally within a 1.4 kb promoter fragment. The phage display library used in this study contained 9177 three-finger ZFPs, each capable of recognizing 9 bp of DNA. The construction of the library was based on a combinatorial assembly of a subset of defined zinc-finger DNA sequences recognizing GNN, ANN and TNN triplets.\(^9,14,30,31\) There were only a few TNN domains available at the time of the library’s construction, so the library is biased towards recognition of RNN triplets, where R = G or A. Theoretically, the double-stranded human genome contains 750 million (RNN)\(_3\) sites.\(^14\) In addition, many promoters contain GNN-rich regions of DNA, which makes the library particularly applicable to the isolation of promoter binding proteins.\(^32\)

As the length of the DNA target was far in excess of any used previously, we first evaluated the ability of the previously established phage display protocol to recover ZFPs. A six-finger protein, E2C, with a dissociation constant of 0.75 nM for an 18 bp site in the ErbB-2 promoter (see Figure 2), was used as a positive control.\(^5\) Using phage expressing the six-finger protein E2C on its surface, test selections were done to compare the recovery of E2C phage by a biotinylated 1.4 kb promoter or a biotinylated 18 bp hairpin oligonucleotide. Both selections had the E2C phage diluted into control phage bearing a different selective marker at a ratio of 1:10,000. The 1.4 kb promoter DNA yielded 90% fewer phage compared to the number of phage recovered using the 18 bp oligonucleotide.
Panning conditions were subsequently optimized, and phage selection using E2C phage was improved to 85% relative to the phage recovered with the 18 bp oligonucleotide.

The library of preassembled three-finger proteins was then selected over four rounds of panning using the 1.4 kb ErbB-2 promoter fragment. The first and second rounds of panning contained 125 nM promoter DNA per binding reaction. The third and fourth rounds contained 62 nM and 31 nM promoter DNA per binding reaction respectively, to increase the stringency of selection and to favor recovery of higher-affinity ZFPs. The effectiveness of each round of panning to select ZFPs specific for the target was initially evaluated using a phage enzyme-linked immunosorbent assay (ELISA) (Figure 1B). As a functional evaluation of the ZFPs recovered, DNA from each round of panning was subcloned into a retroviral vector and expressed as fusions with the activation domain, VP64, which is a tetrameric repeat of the minimal activation domain of herpes simplex virus VP16 activator protein. The ability of each pool of proteins to activate ErbB-2 expression in A431 cells was evaluated using fluorescence-activated cell sorting (FACS) analysis (Figure 1C). The unselected library did show some activation of ErbB-2 expression. However, proteins selected in subsequent rounds of phage display, particularly round 4, were more efficient activators of endogenous gene expression.

Characterization of selected zinc finger proteins

ZFP clones from round 4 of panning were selected for characterization. Five clones were

Figure 1. Phage display strategy with in vitro and in vivo characterization of the rounds. A, Using a 5′-biotinylated primer, the promoter of interest was PCR-amplified and conjugated to streptavidin-coated magnetic beads. Incubation of the phage displaying a preassembled synthetic zinc finger protein library with target-coated beads provided recovery of zinc finger proteins with binding sites in the promoter. Multiple rounds of selection were conducted with decreasing amounts of promoter target to select for a narrowed pool of proteins with high binding affinities. B, Phage ELISA analysis was used to determine the relative binding affinities of the unselected library (US) phage compared to phage selected from four subsequent rounds of panning. Phage were amplified from each round of panning and incubated with biotinylated ErbB-2 promotor fragment immobilized on ELISA plates. Bound phage were detected with an anti-M13/horseradish peroxidase antibody conjugate. Fluorescence of activated ABTS (2,2′-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) was detected at 405 nm and quantified by a microplate reader. The maximal signal of round 2 was normalized to 1. C, Each round of panning was subcloned into a retroviral vector for expression as fusion proteins with VP64 in A431 cells. The ability of the zinc finger fusion proteins from each round to activate endogenous ErbB-2 expression in A431 cells was monitored by FACS analysis. The tallest, dotted peak is A431 cells infected with stuffer DNA fused to VP64. The unselected library peak is the second tallest peak and is in bold. The third tallest peak is round 1 represented with a thin line. The peaks for rounds 2, 3, and 4 overlap, and are shown with a single bold line that is shifted furthest to the right.
initially sequenced and labeled 1A, 2A, 3A, 8A, and 9A. Subsequently, 25 more clones were screened for binding to the biotinylated 1.4 kb ErbB-2 promoter using an ELISA assay (data not shown). The clones with the best binding signals were designated 2, 4, 5, 6, 8, 11, 13, and 16, and were sequenced. On the basis of previous studies, the DNA sequence that each ZFP is predicted to bind is based on the amino acid sequence of each domain’s reading head. Table 1 lists the predicted DNA-binding sequence for each ZFP and the corresponding sites within the ErbB-2 promoter. Six of the ZFP proteins sequenced encoded pairs of identical sequence: 3A and 8, 4 and 11, and 5 and 13. Most of the proteins were predicted to bind with eight out of nine matches to sites in the ErbB-2 promoter, while 8A, 9A, 6, and 8 have perfect 9 bp matching sites in the promoter fragment.

Footprinting analysis of the zinc finger proteins along the promoter

DNase I footprinting analysis was used to confirm the binding of the selected ZFPs at their predicted sites within the promoter DNA. ZFP1A and 2A could not be purified in sufficient quantity for these studies. The other zinc finger proteins were footprinted on the fragment of DNA that contained the predicted sites (only data for ZFP16 are shown in Figure 3). Each protein was titrated over a 100-fold concentration range to confirm binding to the predicted sites. The proteins bound their predicted sites within the promoter DNA with affinities that ranged from 1.0 nM to 88 nM (Table 1). Seven out of nine clones showed additional, lower-affinity binding sites. None of the footprints observed for the predicted sites extended outside the expected 9 bp region by more than 3 bp, indicating the maltose-binding protein fused to each ZFP for protein purification did not interfere with the binding site analyses. For each footprint, the predicted binding site of each ZFP is marked with a lowercase letter. Proximal to site –285 is the poly-purine/poly-pyrimidine repeat region (P/PMR) of the ErbB-2 promoter. The P/PMR region is a GNN-rich region of DNA that is located within a DNase I-hypersensitive region of the ErbB-2 promoter.

Transfection and retroviral infection of zinc finger transcription factors

In order to study the potential of each of the selected ZFPs for gene regulation, the genes encoding them were transferred into mammalian expression vectors and expressed as fusions with the VP64 activation domain. The ability of the resulting transcription factors to activate a luciferase reporter gene driven by the 1.4 kb ErbB-2 promoter was studied in A431 cells (Table 2). TFs 1A and 11 showed the highest level of activation,
with 52.6-fold and 47.4-fold increases in luciferase expression over expression of luciferase in the absence of a transiently expressed transcription factor. TFZFs\textsubscript{2}, \textsubscript{6}, and \textsubscript{8} displayed less than twofold activation, while the level of activation derived from TFZFs\textsubscript{2A}, \textsubscript{8A}, \textsubscript{9A}, \textsubscript{13}, and \textsubscript{16} ranged from 3.2-fold to 15.2-fold. Initially, the various affinities derived from footprinting analysis (Table 1) were compared with activation of a luciferase reporter construct (Table 2). The TFZFs with affinities stronger than 52 nM generated at least fivefold activation of luciferase expression. Considering TFZFs\textsubscript{11} and \textsubscript{13}, whose affinities were determined to be similar for the same region of DNA, their activation of luciferase expression was 47-fold versus 5.8-fold, respectively. Therefore, there was no strict correlation between stronger affinity and greater activation. This may be due, in part, to the multiplicity of binding sites available to a three-finger protein in the 1.4 kb target region and elsewhere in the cell. However, the inability of transient transfection reporter assays to correlate activation with

Figure 3. DNase I footprinting analysis of ZFP16 and three six-finger proteins –642, –369 and –285. Storage phosphor autoradiograms of DNase I footprint titrations analyzed on 8 M urea/6% (w/v) polyacrylamide gels. For each gel: lane 1 (–) is the DNA fragment without digestion; lane 2 is the G + A ladder (L); lane 3, DNase I digestion of the labeled fragment without protein (D); lanes 4–10 contain 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM purified ZFP incubated with the DNA fragment. The location of the predicted binding sites are indicated to the right of each gel with lower-case letters. All reactions contained 15,000 cpm of \textsuperscript{32}P end-labeled PCR fragment. \(K_d\) values are calculated as described.\textsuperscript{54}
affinity or, more importantly, predict endogenous regulation, has been observed with other three-finger TFZFs. Therefore, the activity of the transcription factors on the endogenous gene was studied.

The ability of the TFZFs to activate or repress endogenous ErbB-2 expression was studied using retroviral delivery of the transcription factors into cells. The retroviral vector pMX, which contains a transcription factor expression cassette linked to an internal ribosome entry sequence (IRES) fused to the coding sequence of green fluorescent protein (GFP), was used for these studies. GFP expression from this construct is then correlated with transcription factor activity. Immunofluorescence studies were used to confirm the nuclear expression of zinc finger proteins and the cytoplasmic expression of GFP in retrovirally infected A431 cells.

Table 2 lists the values for activation and repression levels obtained with each TFZF. Transcriptional activation was studied using VP64, and repression was studied using the Kox-1 Kruppel-associated box (KRAB) domain. Six TFZF had greater than twofold activation of endogenous ErbB-2 expression in A431 cells. TFZF 1A and 16 enhanced expression 5.2-fold and 4.8-fold, respectively, relative to cells transduced with control virus. In contrast, only two of the ten TFZF were able to repress ErbB-2 expression greater than twofold: 2A with 2.2-fold repression, and 16 with 7.1-fold repression. Figure 4(a–d) presents the FACS profiles of the two best activating (1A and 16) and repressing (2A and 16) TFZFs. Only 16 was able to both up and down-regulate ErbB-2 expression. The six proteins that showed activation or repression of ErbB-2 expression had predicted binding sites within a GNN-rich region of DNA characterized previously as a P/PMR mirror repeat. The P/PMR region can form a triple helix and prevent binding of endogenous TFs adjacent to the neighboring TATAA box. The P/PMR region has been characterized to have an increased level of chromatin-relaxing proteins associated with it in cell-lines expressing high levels of ErbB-2, compared to cells with low-level ErbB-2 expression. Considering the substantial level of ErbB-2 expression in A431 cells, the P/PMR region would be expected to be an accessible region of DNA, supporting the hypothesis that the topology of a promoter varies from position to position, depending on endogenous factors.

With the predicted binding sites and functional data collected, we were surprised to find that TFZF 9A and TFZF 16 did not share the same transcriptional control profiles. Both ZFP 9A and ZFP 16 are predicted to bind at site –369, with 9A being a perfect match. While TFZF 16 is able to both up-regulate and down-regulate ErbB-2 expression robustly, TFZF 9A up-regulates ErbB-2 expression only moderately. Studies are ongoing to determine if subtle differences in binding between these two proteins explains their differences in regulation.

Affinity of ZFP16 for its three potential binding sites

Since TFZF 16 was differentiated by its ability to both up-regulate and down-regulate, we sought to determine which of its three potential binding sites within a GNN-rich region of DNA characterized previously as a P/PMR mirror repeat.

The three finger (3Fn) ZFP proteins used in this work are listed with their predicted binding sites, the recognition sequences predicted for each ZFP (5'-3'), the number of times each binding motif was identified, the location of each predicted binding site in the promoter target, and the corresponding affinity of the ZFPs at the predicted sites. Sites that were observed in DNase I footprinting experiments and for which associated affinities are included.
sites within the ErbB-2 promoter sequences (Table 1) was responsible for the biological phenotype. The prediction of the biologically relevant site was first addressed by determining the affinity of ZFP16 for each of the three sites. Figure 3 shows the DNase I footprinting for each DNA:ZFP interaction site. Three independent binding reactions were used to calculate $K_d$ values. DNase I footprint analysis of ZFP16 bound on the fragment containing site −642 (Figure 3 panel a) shows one

Figure 4. FACS analysis and immunofluorescence of A431 cells transduced with zinc finger fusion proteins. a–d, FACS profiles are shown for cells transduced with the fusion proteins that displayed activation and repression of endogenous ErbB-2 expression. Black indicates the staining of A431 cells with IgG1 isotype control antibody, blue is the level of ErbB-2 expression in cells transduced with a stuffer fragment, and red is the level of ErbB-2 expression as a result of transduction with the fusion protein expression construct indicated in each panel. e–l, Maximal projections of 3D datasets show the total cellular distribution of TFZF (g and k). A431 cells were transduced with TFZF 16K (i–l) and control StufferK (e–h). The top panels show A431 nuclei in blue; the upper middle panels show GFP in green; the lower middle panels show TFZF in red; and the bottom panels show merge signals of TFZF (red), GFP (green), and nuclei (blue). TFZF localization in nuclei was observed in TFZF 16K-transduced cells (l). No TFZF was detected in control, StufferK-transduced A431 (h).
binding site that extends over the sequence 5'-GGGGAGGAA-3'. The affinity of ZFP16 for this site was determined to be 2.9 (±0.6) nM. ZFP16 was also predicted to bind a site at −369. Footprint analysis of ZFP16 on a 189 bp fragment containing the −369 binding site showed binding of ZFP16 at two regions. The predicted binding site on this fragment at −369 was 5'-TAGGAGGGA-3' and the associated affinity was determined to be 51.6 (±29) nM. ZFP16 also bound a second site with seven out of nine matches to the promoter sequence: 5'-GGGGAGGGG-3'. The affinity of ZFP16 for this site was not determined. The third predicted binding site for ZFP16 was site −285. The binding site extended over the predicted binding sequence, 5'-TTGGAGGGG-3', with an associated affinity of 5.9 (±1.8) nM. Additional binding sites were found between positions −243 and −218 that included three adjacent sites, each with seven out of nine matches with ZFP16's predicted DNA recognition:

5'-GAGAAGGAGGAGGTGGAGGAGGGCT-3'.

In vitro and in vivo characterization of three six-finger proteins for specific regulation of ErbB-2

In order to determine the biologically relevant site for TF216, three six-finger proteins were constructed: 6Fn-642, 6Fn-369 and 6Fn-285. Six-finger proteins recognize 18 bp of DNA and were designed to bind each of ZFP16's three predicted sites independently. The six-finger proteins were characterized by sequencing, by evaluation of binding to oligonucleotides displaying the target sequence, and by DNase I footprinting. Figure 3d shows the DNase I footprinting analysis of 6Fn-642 on a 180 bp fragment containing the predicted site. The site bound is 5'-CACATCCCTCCTTGA-3'. Figure 3e is the DNase I footprint of 6Fn-369 on a 189 bp fragment at the site 5'-GGCGCTAGGGAGGG-3'. Figure 3f is the DNase I footprint of 6Fn-285 on a 165 bp fragment at the site 5'-CGTCAACCTCCCCGGCTC-3'. The affinities of 6Fn-642, 6Fn-369 and 6Fn-285 for their respective sites were 4.8 (±1.1) nM, 4.8 (±1.3) nM, and 9.5 (±4.8) nM. It was interesting to find that two of the three six-finger proteins had lower affinities for the same target sites as their three-finger counterpart, ZFP16. Typically, six-finger proteins are associated with higher affinity.

To evaluate the biological effect of retroviral expression of each of the six-finger proteins fused to activation or repression domains, cell-surface expression of ErbB-2 was monitored by FACS analysis. Each of the constructs was infected into A431 cells in two independent experiments and the resulting up-regulation and down-regulation profiles of each are illustrated in Figure 5. 6Fn-285 showed no ability to up-regulate or down-regulate ErbB-2 expression in A431 cells, 6Fn-642 could slightly down-regulate ErbB-2 expression, and 6Fn-369 was able to up-regulate and down-regulate ErbB-2 expression approximately tenfold from basal expression levels. DNase I footprint analysis showed that 6Fn-369, with an affinity of 4.8 nM for site −369, does not bind at sites −285 or −642 with equal or better affinity (data not shown).

Immunofluorescence studies included A431 cells transduced to express the three-finger protein ZFP16 fused to KRAB (ZFP16K), the six-finger protein 6Fn-369 fused to KRAB (6F3K), and cells
Figure 6. TFZF and ErbB-2 visualization on A431. Maximal projections of 3D datasets show the total cellular distribution of ErbB-2 (d and f) and TFZF (g–i). A431 cells were transduced with the following fusion proteins: ZFP16K (a, d, g, and j), 6Fn-369K (b, e, h, and k), and control StufferK (c, f, i, and l). The top panels show nuclei in blue; the upper middle panels show ErbB-2 receptor in green; the lower middle panels show ZFP in red; and the bottom panels show merge signals of TFZF (red), ErbB-2 (green), and nuclei (blue). TFZF localization in nuclei was observed in ZFP16K-transduced and 6Fn-369K-transduced cells, and correlated with ErbB-2 down-regulation (j and k). No ZFP was detected in control Stuffer-transduced A431 (l).
transduced with a negative control DNA fragment fused to KRAB (StufferK) (Figure 6). Immunofluorescence studies confirmed the down-regulation of cell surface ErbB-2 expression shown by FACS analysis of ZFP16-KRAB and 6Fn369-KRAB (Figure 6d and e versus f). Immunofluorescence studies also showed a correlation of TFZF expression of ErbB-2 (Figure 6) and k versus l. The bottom three panels of Figure 6 have corresponding movies in Supplementary Material.

Discussion

Previous guidelines for the selection of TFZF for control of endogenous gene expression have included affinity, proximity to transcription start sites, and proximity to DNase I hypersensitive regions in cellular chromatin. None of these parameters have proven to be absolute rules for successful, specific gene regulation. With the roles of endogenous TFs, mediator proteins, and chromatin remodeling proteins largely uncharacterized for any given promoter, an empirical method of sampling different binding sites is needed. The availability of the phage display library used in this work, coupled with the growing availability of genome sequencing information, allowed us to identify gene-regulating TFZFs without the need for a well-characterized promoter, or a detailed zinc finger assembly strategy. Here, we describe isolation of ZFPs that bind along the ErbB-2 promoter, and characterize the ability of the selected zinc finger fusion proteins to up-regulate or down-regulate endogenous ErbB-2 expression in human epidermoid carcinoma A431 cells.

Phage display provided means for selecting ZFPs with high affinity for the binding sites provided by a 1.4 kb ErbB-2 promoter fragment. In previous work, protein:DNA affinity was considered a key requirement for achieving gene regulation with a synthetic transcription factor such as a TFZF. With more examples of endogenous gene regulation by TFZF available, the data indicate that the position of targeting is as important as affinity. For example, six-finger proteins pE2X, pE3Y and pE3Z were designed to regulate ErbB-2 and ErbB-3 expression. Only pE3Y both up-regulated and down-regulated ErbB-3 expression, and yet all three proteins had DNA-binding affinities of <10 nM. Similarly, only one of three TFZF targeted to the erythropoietin (EPO) promoter up-regulated EPO transcription, and only three of ten TFZF were able to up-regulate VEGF-A secretion, despite all these ZFPs having dissociation constants of <10 nM. The minimum affinity required for a ZFP to function as an effective TFZF amongst endogenous factors is not known. Theoretically, this value would vary depending on the accessibility of a binding site(s), the number of binding sites, and the position of the binding site(s) amongst promoter-associated proteins. The advantage of using the phage display method is that ZFPs can be isolated that bind various regions of the promoter and with a range of affinities. The resulting pool of proteins is screened to identify the combination of position and affinity that provides transcriptional control of the target gene. Here, the most successful TFZF had a dissociation constant of 52 nM for the biologically relevant target site, an affinity much lower than would be expected. However, the affinity of a ZFP for this site was improved to 4.8 nM with assembly of a six-finger protein. Therefore, even if the affinity that the phage display screen provides is not optimal, if function is demonstrated, an extended ZFP can typically improve the affinity.

The lack of a strict correlation between affinity and transcriptional regulation using TFZF has been acknowledged in the field. Investigation of the role of chromatin structure as it effects TFZF targeting is ongoing. One method of determining the accessible regions of chromatin-associated DNA, is DNase I-hypersensitive site mapping. The identification of a hypersensitive region was a valuable approach in the targeting of VEGF-A. However, the same approach in the design of TFZF selective for PPARγ2 regulation provided differing levels and extents of regulation of both of the PPARγ isoforms, γ1 and γ2. In the targeting of the EPO gene, there was no hypersensitive site associated with site −862, although additional characterization indicated proximity of the site to a positioned nucleosome. In this study, 6Fn-369 was able to up-regulate and down-regulate ErbB-2 expression, and did not map within the DNase I-hypersensitive region of the ErbB-2 promoter. Of the six TFZF that showed binding in the hypersensitive region by DNase I footprinting analysis, three showed activation of ErbB-2 expression, one showed repression of ErbB-2 expression, and one was unable to modulate the transcription of ErbB-2 in either direction. The inconsistency of targeting based on hypersensitive regions and the incomplete characterization of chromatin structure for many promoters, supports the use of an empirical method, without biases, to target TFZF as provided by phage display and a given linear DNA sequence.

Retroviral expression of each selected ZFP as a fusion protein with an activation or repression domain was assessed by FACS analysis of ErbB-2. The most significant finding from the FACS data was that activation and repression of ErbB-2 expression were not usually accessible from the same DNA-binding domain. For example, TFZF targeting the erythropoietin (EPO) promoter was able to up-regulate and down-regulate ErbB-2 expression. Considering that both fusion proteins had the same DNA-binding domain, the up-regulation data suggests that the DNA-binding domain was able to bind the DNA of the promoter in the context of cellular chromatin. Therefore, the inability of TFZF to effect transcriptional control of ErbB-2 is hypothesized to be associated with the proteins recruited by KRAB or VP64 that were unable to form a functional unit or, if the functional unit was assembled, was
unable to modify the promoter’s transcriptional program.\textsuperscript{18,43} By sampling different sites along a promoter, sites amenable to VP64 and KRAB assembly and function can be elucidated to provide gene regulation. Site \(-369\) was a site from which activation and repression could be coordinated. In the context of the ErbB-2 promoter, site \(-369\) lies 20 bp from an upstream AP-2 binding site and 20 bp from a downstream p300 binding site. The AP-2 site acts in coordination with an upstream AP-2 site to activate ErbB-2 expression in breast cancer cells.\textsuperscript{44} AP-2 activation can be countered by an estrogen-suppressible enhancer in the first intron of the ErbB-2 gene.\textsuperscript{45} Binding of p300 in the ErbB-2 promoter provides a target for the adenovirus 5 E1A mediated repression of ErbB-2, which is in clinical trials for breast cancer treatment.\textsuperscript{16,46} Considering the characterization of endogenous factors for the ErbB-2 promoter, site \(-369\) is in proximity to regions accessible for the recruitment of activation and repression mediator proteins.

Once a successful three-finger TF\textsubscript{ZF} was identified, it was unclear which of ZFP\textsubscript{16}'s three potential binding sites was biologically relevant, or if binding at all three sites contributed to the transcriptional regulation observed. Six-finger fusion proteins, which have extended recognition of 18 bp of DNA, were designed to bind independently at each of TF\textsubscript{ZF}\textsubscript{16}'s three best match sites to evaluate transcriptional regulation derived from each site. Even before ZFPs were described, von-Hippel’s study of transcription factors and nucleotide sequence recognized that binding of extended sequences would increase the information content of a DNA-binding protein and reduce binding to other DNA sites.\textsuperscript{48} This theory is supported by the DNase I footprint analysis of the six-finger proteins assembled for this study. Figure 3d–f shows that only one 18 bp site was bound by each of the six-finger proteins on the fragments footprinted. This is in contrast to the upper panels of Figure 3 that show cross-reactive binding of the three-finger protein, ZFP\textsubscript{16}, outside of the region labeled as the best match site on each DNA fragment. Cross-reactivity of the three-finger ZFPs was most extensive in the P/PMR region that contains nine consecutive GNN triplets. Although the footprinting data were used to confirm the ability of the phage display selection method to isolate ZFPs that could be associated with sites in the promoter fragment, the footprinting analysis provided valuable insight into the type of recognition provided by three-finger proteins versus six-finger proteins. The subtlety of amino acid interactions within the ZFP protein structure as well as in proximity to DNA, continues to be an area of study for the zinc finger field.\textsuperscript{49} DNase I footprinting presents a ZFP with a biologically relevant diversity of binding sites not represented by specificity ELISAs, gel mobility-shift assays, CAST assays, or available microarray chips.\textsuperscript{38,50,51} While statistics predict a 9 bp perfect match site for a given ZFP should be once every

| Table 2. Luciferase reporter activation by TF\textsubscript{ZF}\textsubscript{8} and retroviral transduction mediated activation and repression of endogenous ErbB-2 expression by ZFPs |
|-------------------------------------------------|---------------|---------------|
| Zinc finger protein (ZFP) | Luciferase reporter assay | Retroviral transduction |
| | | Fold activation | Fold repression |
| 1A | 52.6 ± 4.4 | 5.2 ± 2.5 | – |
| 2A | 5.6 ± 3.1 | – | 2.2 ± 0.23 |
| 8A | 15.2 ± 10.2 | 2.9 ± 2.1 | – |
| 9A | 3.2 ± 0.7 | 3.2 ± 1.2 | – |
| 2 | 0.55 ± 0.2 | – | – |
| 3 | 1.85 ± 0.13 | – | – |
| 6 | 1.63 ± 0.21 | – | – |
| 11 | 474 ± 15.2 | 3.1 ± 2.0 | – |
| 13 | 5.8 ± 10.2 | – | – |
| 16 | 10.9 ± 2.1 | 4.8 ± 0.84 | 10.7 ± 3.6 |

A431 cells were cotransfected with each TF\textsubscript{ZF} listed and an ErbB-2 promoter-luciferase reporter construct.\textsuperscript{5} Luciferase activity in total cell extracts was measured 48 hours after transfection. Fold activation and standard deviation are derived from triplicate measurements. Fold activation and repression of the endogenous ErbB-2 promoter were evaluated in A431 cells using FACS analysis following retroviral expression of each TF\textsubscript{ZF} listed. Two independent transductions were used to determine the fold activation and repression from basal ErbB-2 expression levels and the corresponding standard deviation values.

2.6 \times 10^7 bp,\textsuperscript{50} this is an underestimate of the potential binding sites given the binding of eight out of 9 bp, or seven out of 9 bp match sites with affinities <100 nM as observed here. Taking into consideration this degeneracy, by increasing the number of contacts needed to provide a minimum threshold of binding affinity, the importance of using six-finger proteins for specific gene regulation is emphasized.

Phage display for the selection of TF\textsubscript{ZF}s has many advantages. First, the investigation is not limited to sites that fit the guidelines previously used for the selection of TF\textsubscript{ZF} binding sites. Second, the method provides TF\textsubscript{ZF}s more rapidly than by individual construction and testing. Finally, this empirical approach requires only the preassembled library of ZFPs and a linear sequence of DNA to allow selection. Our application of this approach has resulted in the successful regulation of the human ErbB-2 promoter. An advantage of using phage display is that each round of panning can be tested in various cell lines to determine which pool of TF\textsubscript{ZF}s is amenable to transcriptional regulation in a particular cell type. We suggest that this strategy holds potential for the rapid preparation of transcription factors for the characterization of genes and for therapeutic application via gene therapy.

Materials and Methods

Selection by phage display

Construction of zinc finger libraries has been
Phage ELISA

Phage ELISA was performed as described. Streptavidin-coated magnetic beads (Dynal) were optimized to bind the 1.4 kb target (0.05 μg target/ml beads). The beads were washed four times with zinc buffer (ZBA) (10 mM Tris, 90 mM KCl, 1 mM MgCl₂, 90 μM ZnCl₂), 5 mM DTT and once with ZBA, 5 mM DTT, 5% (w/v) non-fat dry milk. Target preparation was in ZBA, 5 mM DTT, 5% non-fat dry milk. Then 100 μl of ZBA and 2.5 μl of filtrate (10¹⁵ colony-forming units), 100 μl of ZBA, 5 mM DTT, 5% Blotto, 4 μg of sheared herring sperm DNA (Sigma), 294 μl of ZBA and 2.5 μl of filtrate were added to the beads and the samples were incubated for three hours at room temperature on a rotating wheel. After incubation, beads were washed once with ZBA, 5 mM DTT, 5% non-fat dry milk. Then 100 μl of filtered phage (10¹⁵ colony-forming units), 100 μl of ZBA, 5 mM DTT, 5% Blotto, 4 μg of sheared herring sperm DNA (Sigma), 294 μl of ZBA and 2.5 μl of filtrate were added to the beads and the samples were incubated for three hours at room temperature on a rotating wheel.

Luciferase assays

Luciferase assays were performed as described, except that A431 cells were used.

Antibodies

ErbB-2 expression was detected using TA-1 antibody (Calbiochem). Control staining was done using mouse IgG1-UNLB (SouthernBiotech). The secondary antibody for both was Cy5-labeled, affinity-purified donkey anti-mouse IgG (Jackson ImmunoResearch).

Retroviral gene targeting

For retroviral expression of the three-finger and six-finger proteins, the zinc finger-KRAB and zinc finger-VP64 coding regions were cloned into a modified pMX-ires-GFP using SfiI restriction sites (ires, internal ribosome-entry site; GFP, green fluorescent protein). As a control for the retroviral infection, a construct containing a stuffer fragment of DNA inserted at the Sfi sites was prepared. The stuffer sequence coded for a single-chain Fab modified with stop codons that prevent expression. The retroviral pMX-ires-GFP/zinc finger constructs were transiently transactivated into the packaging cell line 293 gag/pols by using Lipofectamine Plus (GIBCO/BRL). Three hours after transfection the medium was changed to 6 ml of fresh DMEM with 10% fetal calf serum (FCS) and penicillin/streptomycin/anti-mycotic antibiotics. Approximately 42 hours later, culture supernatants were used for infection of target cells in the presence of 8 μg/ml of Polybrene. Four infections were performed, the first after 42 hours of transfection, then 50, 66 and 74 hours following transfection. At 90 hours post-transfection, the medium was changed to 10 ml of fresh DMEM, 10% FCS, antibiotics. One week from the start of the transfection the cells were harvested for analysis.

Flow cytometric analysis

Cells were trypsinized and washed in FACS buffer (PBS, 1 mM EDTA, 25 mM Hepes (pH 7.0), 1% (w/v) BSA) prior to staining. Two wells for each sample were prepared with 10⁵ cells; one received 5 μg/ml of TA-1 antibody and the other 5 μg/ml of IgG1 control antibody in 100 μl of FACS buffer. After incubation on ice for one hour, cells were washed twice in FACS buffer. Bound antibodies were stained with Cy5-labeled donkey antimouse secondary antibody in 100 μl of a 1:100 (v/v) dilution in FACS buffer. Finally, the cells were washed twice in FACS buffer, resuspended in 500 μl of FACS buffer, and analyzed for their fluorescence with a Becton Dickinson FACSort.

Fold activation and repression were calculated based on gating the GFP-positive population (examples shown in Figure 4a–d). From the GFP-gated profiles, the geometric mean value for the center of the peak derived from the protein expression as a result of TFZF infection (in Figure 4a–d this peak is red) and the geometric mean value for the center of the peak derived from cells transduced with a stuffer DNA fragment fused to VP64 or KRAB (Figure 4a–d, the blue peak) were calculated using CellQuest software. To calculate fold activation, the TFZF derived peak value (red) was divided by the basal expression peak (blue). For fold repression, the basal expression peak (blue) was divided by the TFZF-derived peak (red).

Construction and characterization of the six-finger proteins

For the construction of the six-finger proteins, two three-finger proteins binding each of the 9 bp half-sites of each of the three 18 bp target sequences were constructed by grafting the appropriate DNA recognition helices into the framework of the three-finger protein Sp1C. DNA fragments encoding the two three-finger proteins were assembled from six overlapping oligonucleotides as described. The oligonucleotides used were based on DNA recognition helices characterized from finger 2 variants of Zif268. Six-finger 6Fn-654 was designed to bind 5'-GTCAGGGGGAATGTTGGA-3' and was assembled from domains pGTG, pAAAG, pmGAG, pmGGG, pmGAT and GTG. Six-finger 6Fn-369 was designed to bind 5'-GGCGCTAGGAGGGACGAC-3' and was assembled from domains pmGAG, pmGCT, pAAG (t), pAGG (j), pmGAC and PMGAC. Six-finger 6Fn-275 was designed to bind 5'-CCAGTTGGAAGGGGCGAG-3' and was assembled from domains pGCA, pmGTT, GGA, pmGAG, pmGAC, pmGAG.

Purification and footprinting

Zinc finger-coding DNA was subcloned into a modified pMAL-c2 (New England Biolabs) bacterial expression vector and transformed into XL-1 Blue (Stratagene). Protein was purified using the Protein Fusion and Purification System (New England Biolabs) following the manufacturer’s protocol, except that ZBA/5 mM DTT was used as the column buffer. Protein purity was determined by staining 4%–12% Novex gels with Coomassie brilliant blue. The protein concentration was determined by Bradford assay with bovine serum albumin (BSA) standards. The radiolabeled DNA fragments were generated by PCR using the human ErbB-2 promoter cloned into pG3® with High Fidelity PCR Master (Roche) with 60 pmol of the following (5'-3')-labeled primers (designated with an F) and 3' primers (designated with a B):

342F (GCATTTGAGAAATTTAGATAAGCTTTTTG)
(GGATTACAGGCTAGGCGACCGATCCG) 450F (GGAGTCAAGAGCAGCCTCACGAGTGAGGAAGACC) (GGCTTCGTGCCAGGTGAGGAGAC) (CCAAAAATTTTGGACCTCCTTAGATGCT) (CCAGGCTTATTTTGAATATACAAAAATGG) (CCTGGACTATAAAGGTTAAGATTGGCGACC) 1110B (GCTTACACTTCTCCCTCTTCGCG) (CCTAGGGAATTTATCCCGGACTTCC) 1150B (CGAAGGCTGGATCGCAACTCTCC) 1090F (GCAGAGAGGGGAGAAGTAAGGGAAGGC) 1270B (GCTCCAAACTTCTCCTTACTTCC)

Buffer for PCR with all primers except 1270B and 1150B included 10% (v/v) DMSO. The PCR program was two minutes at 94 °C, 30 seconds at 94 °C, 30 s at 55 °C, 30 seconds at 72 °C, three cycles, seven minutes at 72 °C, then 4 °C. The footprinting assay was carried out in triplicate as described. The 5 × TKMC buffer (50 mM Tris–HCl (pH 7.0), 50 mM KCl, 50 mM MgCl2, 25 mM CaCl2,) was supplemented with 50 μM ZnCl2 and 5 mM DTT was added fresh to the binding reactions. Thymidines at 72 °C included 10% (v/v) DMSO. The PCR program was two minutes at 94 °C, seven minutes at 55 °C, and were subsequently digested with 15 μl of a 0.00012 unit/μl solution of DNase I, 1 mM DTT (Roche Diagnostics) Samples were electrophoresed through an 8 M urea/6% poly-acrylamide gel. The gels were exposed on phosphorimager plates, recorded by a PhosphorImager SI (Molecular Dynamics), and subsequently analyzed using Image-Quant (Molecular Dynamics) and Kaleidagaph software (Synergy, Reading, PA) to give Kd values as described.

Immunocytochemical analysis of TFZF localization

For analysis of TFZF expression in transduced A431 cells (TFZFZ FP16K and 6Fn3K), cells were seeded on poly(L-lysine)-coated Lab-Tek coverglasses and incubated for 30 minutes. Cells were washed with copious amounts of PBS, incubated with 4% (v/v) paraformaldehyde for 20 minutes, washed again, and incubated in a humidifying chamber at room temperature for one hour with a mixture of biotinylated rat anti-HA mAb (2 μg/ml, Roche Diagnostics, Indianapolis, IN) and rabbit anti-human ErbB-2 polyclonal antibody (5 μg/ml, Calbiochem, La Jolla, CA) in FACS buffer, 0.1% (w/v) saponin. The cells were then stained for one hour at room temperature with the mixture of Cy5-conjugated donkey anti-rabbit IgG polyclonal antibodies and Streptavidin/ rhodamine Red-X (both from Jackson Immunoresearch, West Grove, PA) diluted 1:100 (v/v) in FACS buffer, 0.1% (w/v) saponin. Finally, the cells were incubated with 4',6-diamino-2-phenylindole dihydrochloride (DAPI) solution for five minutes, washed with PBS, and covered with SlowFade Antifade reagent. Four-color (DAPI, GFP, rhodamine Red-X, and Cy5) three-dimensional datasets were collected with a DeltaVision system (Applied Precision, Issaquah, WA). Exposure times were 0.2–0.5 s (2-binning), and images were obtained with either 60× or 100× magnification oil objectives. Three-dimensional reconstruction was generated by capturing 150 μm serial sections along the z-axis. Images were deconvolved (based on the Agard-Sadat inverse matrix algorithm) and analyzed with softWorX Version 2.5.

Acknowledgements

The authors thank Joel Gottesfeld, Christian Melander and David Segal for their invaluable assistance. This work was supported by the US National Institutes of Health CA86258 and DK61803. C.V.L. is a Skaggs Predoctoral Fellow.

References


46. Chen, H., Yu, D., Chinnadurai, G., Karunagaran, D.


Edited by F. Schmid

(Received 14 January 2004; received in revised form 21 April 2004; accepted 22 April 2004)