Regulatory DNA binding peptides as novel tools for plant functional genomics

Lindhout, B.I.

Chapter 2

Effects of differently designed artificial polydactyl zinc finger transcription factors on the regulation of genomic target sites in yeast

Partially published as:

Leon W. Neuteboom, Beatrice I. Lindhout, Ingrid L. Saman, Paul JJ. Hooykaas and Bert J. van der Zaal, Biochem. Biophys. Res. Commun **339**, 263-270

ABSTRACT

We developed a vector system for easy assembly of zinc finger modules to be used in artificial transcription factors. Next to proteins that possess short canonical TGEKP linkers between all constituting zinc fingers we introduced 12 amino acid linkers between two three-finger units and between three two-finger units. Fusions of these zinc fingers and the VP16 activation domain were tested for their ability to regulate a repressed genomic locus containing contiguous or noncontiguous zinc finger binding sites in yeast. Remarkably, while the presence of longer linkers always badly affected the *in vivo* performance of transcription factors, this was not necessarily so for their *in vitro* DNA binding characteristics. In contrast to other studies, which were mostly confined to *in vitro* tests, we did not obtain evidence that superior artificial zinc finger transcription factors can be designed using longer linkers between individual fingers. Instead, canonical linkers within a highly regular backbone in combination with a contiguous 18 base pair DNA target site provided a sound base for polydactyl zinc finger design for regulation of genomic loci.

INTRODUCTION

Artificial transcription factors consisting of a polydactyl zinc finger (PZF) domain for DNA binding, linked to a protein domain which either activates or represses genes nearby the PZF binding site, have recently gained enormous attention as novel tools in molecular biology (see for recent reviews: 1-3). In addition, the possibility to induce site-specific double strand breaks in complex genomes with designed zinc finger nucleases has led to exciting prospects of effective methods for gene therapy (4, 5). In theory, it should now be possible to regulate or modify at will any gene from any organism for which sufficient genomic sequence information has been obtained. PZF domains are most successfully constructed using Cys_2-His_2 zinc finger (ZF) domains. Apart from being small, just 30 amino acid residues, Cys_2-His_2 fingers are highly modular and can be linked to increase the specificity of DNA binding.

Just a few amino acid residues within the alpha helix of a Cys_2-His_2 ZF domain are crucial for the recognition of a particular triplet of a DNA sequence. Via elegant experimental protocols optimized ZF sequences starting with 5' G or A have been elucidated (6-8), although in some cases the 'triplet code' is not yet fully specific (9) or there is some target site overlap when the ZF module

requires the presence of a G or T nucleotide in the adjacent finger's subsite (1). Selection of ZF sequences for specific recognition of triplets starting with 5' C or T proved to be more complicated (1). Only recently a promising ZF recognition code for most of the 5'-CNN triplets has been elucidated (10). Although a complete nondegenerate recognition code table for overlapping 4 bp sequences, which should account for the target site overlap exhibited by Cys_2-His_2 type zinc fingers, has been proposed (11), the practical value of this code remains uncertain. In any case, current knowledge of ZF modules and their target sites provides tremendous possibilities for successful construction of effective artificial transcription factors based on PZF modules.

There is no fixed method to combine different Cys_2-His_2 ZF modules into a PZF domain. The most common strategy involves PCR-mediated grafting of modules into a regular scaffold as exemplified by the natural zinc finger proteins Zif268 and Sp1, with a canonical 'TGEKP' five amino acid linker sequence between the different ZF modules (12). In those cases in which TGEKP linkers were used to connect all constituting ZFs (13, 14) it was observed that assembly of six-finger (6F) domains by connecting two three-finger (3F) domains via a TGEKP linker led to a relatively modest increase in affinity (50 to 70-fold) for their 18 bp target site compared to the constituting 3F domains for their corresponding 9 bp target sites. Since on theoretical grounds a much higher gain in affinity should be possible, it was hypothesized that the helical periodicity of ZFs connected via canonical linkers is just too short to match the helical periodicity of B-form DNA. Binding of more extended PZF moieties would therefore increase DNA unwinding, resulting in a structural tension that counteracts the stability of the complex. For this reason, the use of longer linkers at one or more positions within 6F proteins was expected to lead to a more stable protein-DNA complex and a dramatic increase in binding affinity. The first indications to support this view have been published by Kim and Pabo (15), who reported that connecting 3F proteins via a twelve amino acid linker resulted in a 6,000-fold tighter interaction with its target site than measured for its 3F components. Although this figure is frequently quoted elsewhere, the 6,000-fold improvement should be evaluated within its original context. The extremely low dissociation constants (K_d) reported for the 6F protein could not be measured via gel mobility shift assays, but had to be calculated as the ratio of observed k_{off}/k_{on} and thus leaned heavily on the extremely low k_{off} value. Interestingly, as reported in the same study, a 6F module with twelve (LRQKDGGGSERP) or nine (LRQKDGERP) amino acids between fingers 3 and 4 was able to span one or two extra base pairs present between the two 9 bp sequences of the half sites. With

a twelve amino acid linker, the affinity for 18, 18+1 and 18+2 bp target sites was very similar.

According to Kim and Pabo (15), longer linkers between two 3F peptides (2x 3F design) increase the binding strength to a target site, but at the expense of specificity. However, in an alternative configuration, after insertion of the amino acids G or GGS within the canonical linkers between fingers 2 and 3 and between fingers 4 and 5 (resulting in 3x two-finger (2F) peptides), the presence of longer linkers was reported to increase binding specificity (16). Although a 2x 3F peptide connected via a LRQKDGERP linker sequence bound with similar affinity to an 18 bp contiguous target site as 3x 2F variants, the 3x 2F peptides with two six amino acid linkers bound related target sites much weaker than the 2x 3F version and had a strong preference for noninterrupted 18 bp sites. Increasing the linker length to eight amino acids residues allowed recognition of noncontiguous target sites with one or two base pairs inserted between positions 6 and 7 and/or 12 and 13 of the original 18 bp sequence, thus in between 2F subsites. Two base pair insertions were not spanned efficiently. Since the LRQKDGERP linker used by Kim and Pabo (15) has been criticized because of the absence of a helix-capping motif for the preceding zinc finger (16), it is unfortunate that both Kim and Pabo (15) and Moore *et al*. (16) did not include a straight 6F peptide with only canonical linkers between individual fingers in their study.

Use of even longer linkers in a 2x 3F design, like a 20 amino acid flexible linker and in particular structured linkers, allowed *in vitro* binding to an 18 bp target site with 0-10 bp of nonbound DNA in between the 9 bp half sites (17). Although it can be argued that this loss in binding preference hampers the use of such designs for applications where only a single interaction is desired, the possibility to span intervening nucleotides can also be a valuable trait, especially when highly specific ZF moieties for certain DNA triplets are lacking. Nomura and Sugiura (18) demonstrated that 3F peptides with extra G and S residues between the E and K residues of a TGEKP sequence, started to exhibit preferences for extended binding sites. In a recent paper, comparing the effects of a polyarginine linker GRRRRRRRRQ with a polyglycine (G10) linker in a 2x 3F design, it was found that the bulky arginine linker led to a better interaction with an extended binding site than with an 18 bp contiguous site while the G10 linker allowed high affinity interaction with both (19). It remains to be demonstrated that the use of alternative linkers, not complying with helix-capping rules, truly contributes to more optimal PZF designs.

Apart from methodological differences leading to rather divergent biochemical characteristics, a more fundamental aspect leading to discrepancies

in PZF literature is probably the use of different frameworks to construct PZF domains. Although Beerli *et al.* (14) have shown that different backbones can lead to different affinities of the PZF proteins for their binding sites, with the Sp1C framework producing the highest affinity, there is apparently no consensus in the field and even mixtures of different backbones within a single PZF protein are frequently encountered in literature. In this work, we developed a cloning vehicle for rapid and flexible production of PZF proteins. To avoid any unnecessary bias and target site overlap, PZF proteins were in principle constructed using optimized 5'-GNN recognition helices (6) within a strict Sp1C framework and cognate binding sites comply with $(GNN)_6G$. The system also allows a choice of flexible linkers of different length at various positions in the multifinger protein. The sequence of each linker, including the longer linkers, complies with sequences known to lead to stabilizing α-helix C-capping interactions upon DNA binding. Furthermore, although PZF modules are intended for manipulating expression of genomic loci, systematic analyses of different types of PZF modules for their ability to be used in this manner are lacking. Therefore we introduced VP16-PZF fusions created with our novel cloning vehicle into yeast (*Saccharomyces cerevisiae*) strains that carry a stably integrated, repressed reporter gene with various ZF target sites. Analysis of reporter gene expression proved to be a robust test to determine which types of PZF modules were able to bind effectively to chromatin-embedded DNA. The *in vivo* gene expression data were combined with *in vitro* analyses for binding characteristics of the different proteins. To our knowledge, our study describes for the first time an elaborate and consistent series of data regarding the efficacy of different types of PZF domains for gene regulatory processes on a genomic locus. Unexpectedly, some of our data stand in sharp contrast with previous reports concerning the introduction of longer linkers in which conclusions were essentially based on *in vitro* data only. Although our *in vitro* data support other findings that an extra base pair present in binding sites can be spanned using longer linkers, we did not observe *in vivo* binding to interrupted target sites leading to reporter gene activity. The most optimal PZF modules possessed canonical linkers between all fingers and were highly specific for their corresponding uninterrupted target sites. 6F proteins also performed much better than 4F and 5F proteins for *in vivo* gene activation, a feature not observed previously. We did not find any evidence that 6F proteins with short, canonical linkers impose strain on the DNA helix that counteracts DNA binding affinity. The implications of these findings for the generation of artificial transcription factors are evaluated.

MATERIALS AND METHODS

Construction of vectors

All primers were purchased from Sigma Genosys (Haverhill, UK) and all enzymes were purchased from New England Biolabs (Ipswich, MA, USA) unless indicated otherwise. The polylinker of plasmid pBluescript II SK+ (Stratagene, La Jolla, CA, USA) was digested with SacI and KpnI and replaced by the sequence in Figure 1A. The resulting plasmid pSKN-SgrAI was used for the construction of ZF modules, exploiting different restriction sites producing 5'-CCGG-3' sticky ends. SfiI sites can be used for the excision of constructed ZF modules, while a variety of unique restriction sites in polylinkers I and II allows translational fusion(s) of additional regulatory effector domains at the N- and/or the C-terminus of the ZF moiety. A series of 16 pairs of oligonucleotides (84-mers) (Table 1), with the coding region optimized for recognition of a particular 5'-GNN-3' DNA target sequence, (6) was used to select the ZFs needed for the present study (outlined below). The herpes simplex virus VP16 transcriptional activation domain (20, 21) was obtained as a 210 bp KpnI/SalI fragment from the vector pRED16A (kind gift of Dr. Pieter Ouwerkerk, Leiden University, the Netherlands) and cloned into KpnI/XhoI-digested pSKN-SgrAI, upstream from the SfiI sites (Figure 1A). NotI sites were used to excise the fusion protein-coding sequences for recloning into the p426ADH-derived (22) yeast expression vector p426ADHl (Figure 1B). pSKN-SgrAI NotI fragments include the SV40 nuclear localization signal-coding sequence (Figure 1A) (23, 24), while the yeast vector p426ADHl adds an N-terminal FLAG-tag to the expressed proteins (Figure 1B).

Table 1. Oligonucleotides encoding ZF modules for GNN triplets. The sequences of the primary strands within the duplex regions are given. All oligonucleotides contain 5'-CCGG overhangs compatible with TGEKP linker construction (not shown). Codon usage has been varied, but rare codons were avoided. Each coding sequence corresponds with an optimal sequence to recognize a particular GNN target shown left (6).

triplet	oligonucleotide sequence from 5' to 3'
1: GGG	CGAAAAACCGTATGCTTGTCCAGAATGTGGTAAGTCCTTTTCTCGTAGCGATAAGCTCGTTCGCCACCAAAGGACTCATA
2: GGA	CGAGAAACCATACGCCTGTCCCGAGTGCGGTAAAAGCTTTAGTCAGCGTGCCCATCTTGAACGCCACCAGAGGACCCATA
3: GGT	CGAGAAGCCCTATGCGTGCCCTGAGTGTGGCAAGTCCTTTAGCACCTCGGGTCATCTGGTGCGTCACCAACGTACACATA
4: GGC	CGAGAAACCGTACGCGTGCCCTGAATGCGGTAAAAGCTTCTCTGACCCAGGTCACCTCGTCCGCCATCAGAGGACCCATA
5: GAG	CGAGAAGCCCTACGCTTGCCCGGAGTGTGGCAAGTCATTCTCCAGGAGCGACAATTTGGTTAGGCATCAACGCACACATA
6: GAA	CGAGAAACCGTACGCTTGCCCTGAATGCGGAAAAAGCTTCTCTCAGTCCAGCAACCTCGTGAGGCATCAGAGGACCCATA
7: GAT	CGAGAAGCCTTATGCCTGTCCTGAGTGCGGGAAGTCTTTTAGCACCTCGGGTAATCTCGTCCGTCACCAACGTACGCATA
8: GAC	CGAGAAGCCCTACGCATGCCCGGAATGCGGAAAATCCTTCTCCGATCCCGGTAACTTGGTTCGGCATCAGCGCACCCATA
9: GTG	CGAAAAGCCGTATGCGTGCCCCGAATGTGGGAAGTCCTTTTCTCGTAGCGACGAGCTTGTTCGCCACCAAAGGACCCATA
10: GTA	CGAGAAGCCTTACGCTTGCCCCGAGTGTGGCAAATCGTTCTCCCAGTCCTCCTCGTTGGTCCGTCATCAACGGACGCATA
11: GTT	CGAGAAACCCTACGCCTGCCCGGAATGCGGGAAGTCGTTTAGCACCTCCGGTAGCCTGGTGAGGCATCAGCGTACGCATA
12: GTC	CGAGAAACCGTACGCATGCCCTGAGTGTGGCAAGTCCTTCTCCGATCCGGGAGCACTCGTTCGGCATCAACGCACCCATA
13: GCG	CGAGAAACCGTATGCGTGTCCCGAATGCGGTAAAAGCTTTAGCCGTAGCGATGACCTTGTGCGTCACCAGAGGACCCATA
14: GCA	CGAGAAGCCCTACGCCTGCCCGGAGTGTGGAAAGTCGTTCTCCCAGAGCGGCGATTTGCGTCGGCATCAACGGACGCATA
15: GCT	CGAGAAACCGTATGCGTGCCCCGAATGCGGGAAATCCTTTTCGACCAGCGGTGAACTCGTGCGTCACCAGCGTACGCATA
16: GCC	CGAGAAGCCTTACGCTTGCCCGGAGTGTGGCAAGTCATTCAGCGACTGCCGTGATCTCGCAAGGCATCAACGCACCCATA

Generation of PZF proteins

Via stepwise PCR protocols using long overlapping primers, a sequence encoding E2C(Sp1) (Sp1C framework; 14) was generated (Figure 2A and 2C), available as a SfiI fragment for further cloning purposes.

Other PZF proteins were constructed using the pSKN-SgrAI vector developed in this study (Figure 2B and 2C). In principle, the PZF proteins generated the Sp1C framework (6, 14) with a constant choice of alanine preceding the first cysteine of each ZF. In this manner, the variation of amino acids was restricted to the sequence just before and in the recognition helix (positions -1 to 5), as these were chosen according to the GNN code published by Segal *et al*. (6), using the mutated versions for recognition of codons other than GGA, GTA, GAT, GTC, GCC, and GCA, as these were more specific. In a 25

µl reaction volume, 100-200 ng of SgrAI-digested pSKN-SgrAI (or this vector with one or more ZFs present) was ligated at 25˚C with a 1000-fold molar excess of annealed oligonucleotides in the presence of AgeI. Vector DNA, with oligonucleotides attached at both ends, was isolated from agarose gel, heated for 2' at 95˚C in STE, gradually cooled and annealed and transformed to the *Escherichia coli* strain DH5α. Depending on the order of ZF ligation and linker generation, two sets of three different types of 6F modules were constructed (Figure 2B). 6F modules PTF1 and E2C1, which possess canonical TGEKP linkers between all constituting ZFs, were constructed by sequential addition of oligo pairs corresponding with the matching ZFs into the (reoccurring) SgrAI site. For PTF1 the order of addition of the oligo pairs was 7, 12, 16, 4, 16, and 7 (Table 1). Considering the reading frame encoded by each oligo pair, the incoming oligos were fused at the 5' coding end of the growing ZF peptide-coding chain (the order of the ZF modules in the protein being anti-parallel to the order of appearance of the corresponding triplets in the binding site). For E2C1 this order was 1, 16, 2, 16, 14 and 9 (Table 1). 6F modules PTF2 and E2C2 were constructed by fusing two 3F-encoding domains (Figure 2B). DNA encoding the C-terminal half of the 6F protein was isolated as a SgrAI-BspEI fragment and cloned into a BspEI-digested plasmid encoding the N-terminal half of the 6F protein. When the incoming ZF moiety was present in the correct orientation, the N- and C-terminal 3F peptides were connected via a twelve amino acid linker (TGELGGGSGEKP) between fingers 3 and 4. 6F modules PTF3 and E2C3 were constructed by sequential addition of 2F modules (as SgrAI-BspEI fragments into BspEI-digested precursors) in such a manner that twelve amino acid linkers were formed between fingers 2 and 3 and between fingers 4 and 5, while standard five amino acid linkers were present between the other ZFs of the PZF protein encoded (Figure 2B). All constructs were checked by sequence analysis.

Construction of yeast reporter strains and strains for protein production

A NotI/KpnI fragment of plasmid pMELαUASc-10bp-UASc (kind gift of Dr. Karsten Melcher, University of Frankfurt, Germany)(25, 26), harbouring the *MEL1* gene driven by the *MEL1* core promoter containing one Mig1p and two Gal4p sites separated by 10 bp, was used for the cloning of either one or three ZF binding sites in the AatII site (Figure 3), which is located 79 bp upstream from the nearest TATA box (cloning details available upon request). All sequences were confirmed and subcloning into the vector pSKNN (sequence available upon request) provided a second NotI site directly downstream from the KpnI site, allowing further cloning into the NotI-digested pINT1 vector (27). Recombinant pINT1 plasmids harbouring the *MEL1* gene in the same orientation as the *APT1*

gene were linearized with SacI and NcoI and the DNA was used to transform yeast strain 21RΔmel1 (25) by a PEG/LiAc transformation method (28). Yeast colonies containing the *MEL1* gene at the *PDC6* locus were selected on YAPD medium containing 100 µg/ml G418. Insertions were checked by PCR after isolation of chromosomal DNA with the YeaStar genomic DNA kit (Zymo Research, Orange, CA, USA).

Expression plasmid p426ADHl (Figure 1B), harbouring the zinc finger-containing modules (or control constructs) derived from pSKN-SgrAI, were subsequently transferred into the reporter strains and colonies containing the 2µ vector were selected. Proteins to be used for gel shift assyas were isolated from yeast cells grown in liquid medium using CelLyticTM Y Cell Lysis Reagent (Sigma), including 2 mM DTT, according to the manufacturer's instructions. The levels of the correct FLAG-tagged proteins were verified by probing protein blots of total protein extracts with ANTI-FLAG M2-Alkaline Phosphatase Conjugate (Sigma).

Galactosidase assays

Yeast reporter strains producing proteins of interest were grown in liquid cultures (minimal medium, 1% casaminoacids, 3 g/l leucine, 2 g/l ade supplemented with either 2% glucose or 2% galactose) to an $OD_{600\ nm}$ of 0.7-1.0 and diluted in fresh medium to an OD_{600nm} of 0.55. Samples were centrifuged for 30" at 10,000x g and the supernatant was used for determination of α-galactosidase activity using *p*-nitrophenyl α-D-Galactopyanoside (Sigma) as substrate. A typical reaction contained 16 µl supernatant and 48 µl assay buffer (2 volumes 0.5 M NaOAc pH 4.5 and 1 volume 3.01 mg/ml *p*-nitrophenyl α-D-Galactopyanoside). Reactions were performed in microtiter plates at 30°C. Samples were drawn at regular intervals; further reaction was terminated by addition of 136 µl 1 M Na_2CO_3 and *p*-nitrophenol production was measured in a Bio-Rad model 2550 EIA Reader at 414 nm (Hercules, CA, USA). All samples were measured in triplicate with three independently grown cultures for each yeast target strain/p426ADHl-derived expression plasmid combination.

Gel shift assays and determination of dissociation constants

Samples of crude yeast cell extracts containing proteins of interest were assayed for DNA-binding activity using [α^{32}P]dCTP end-labelled (Klenow fragment, Amersham Biosciences, Little Chalfont, UK) synthetic oligonucleotide duplexes (Figure 3) essentially as described by Moore *et al*. (16) except that 20 mM HEPES (pH 7.2) was used as a buffer component instead of Bis-Tris propane (pH 7.0).

Each reaction contained 1 µg nonspecific competitor DNA (poly (dI-dC)) (Sigma). Binding reactions were performed in a 10 µl volume for 60' at room temperature after which the samples were loaded under voltage on a native 5% acrylamide gel in 0.5x Tris-borate buffer. Electrophoresis was performed at room temperature at 125 V per cm for 1.5 hr. Gels were dried on DE81 paper and exposed to X-ray film or to a phosphor screen (Molecular Dynamics, Sunnyvale, CA, USA). Radioactive signals were quantified using ImageQuant software. Apparent dissociation constants (K_d) for binding reactions were determined by Scatchard analysis essentially as described by Clemens *et al*. (29), with a constant amount of labelled binding site (15 pM) and increasing amounts of unlabelled binding sites in each reaction. After initial trials, conditions were -when possible- chosen in such a manner that at the lowest probe concentration used (about 10-fold lower than the estimated K_d) at least 50-60% was bound in complex with the PZF protein. For interactions with a K_d higher than 10 nM an accurate value could not be obtained since the PZF protein concentration in the yeast extracts were then too low to allow unbiased measurements.

RESULTS

Synthesis of PZF proteins and DNA binding sites

We have developed the vector pSKN-SgrAI (Figure 1A) for the rapid and flexible construction of different PZF proteins from a series of ZF oligomers with 5'-CCGG overhangs (Table 1). A procedure for building 6F proteins by connecting two PCR-generated DNA fragments with 5'-CCGG overhangs, which were generated after digestion with AgeI and XmaI, has been described (30). The PZF construction method in the present study further exploited 5'-CCGG overhangs by using additional flanking sequences in combination with the 5'-NNCCGGNN-recognizing enzymes SgrAI, AgeI, BspEI, and XmaI. Although more cloning steps are required, the repetitive addition of ZF-encoding oligomers into SgrAI-digested vector molecules proved to be simple and efficient and therefore the advantages of the system are obvious. A single series of sixteen primer duplexes suffices for the construction of fingers for all possible target sites based on GNN triplets and any other ZF-encoding oligos can be introduced as well. Moreover, the system easily allows the introduction of longer flexible linkers in case this would prove to be favourable. Furthermore, construction intermediates can be saved as valuable controls or starting points for the swift development of other PZF modules.

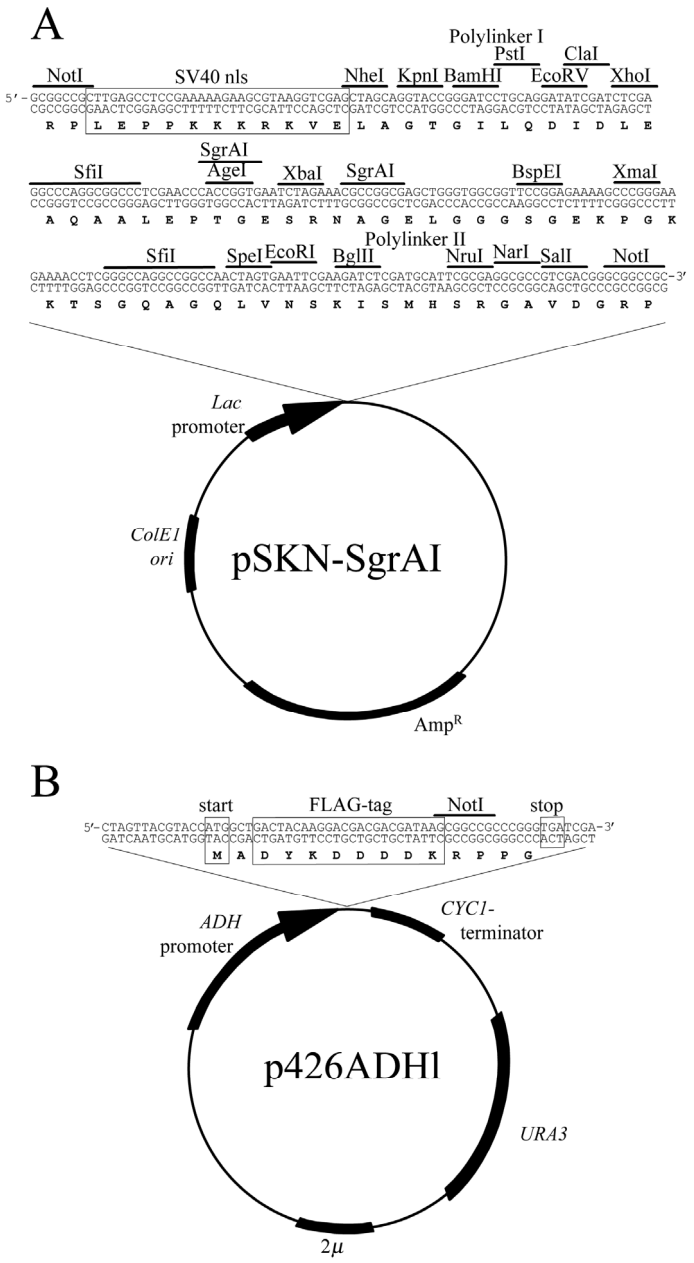

Figure 1: Plasmids used in this study; (**A**) pSKN-SgrAI for the construction of PZF-transcription factors. The inserted DNA sequence and the derived amino acid sequence are shown, coding from top to bottom; (**B**) p426ADHl for the expression of PZF-transcription factors in yeast. The novel polylinker shown was inserted in the SpeI and XhoI sites of p426ADH (22).

The vector system was used for the construction of two different sets of 6F proteins (Figure 2). The protein set PTF can be regarded as a random choice out of all possible 6F proteins that could have been made. Although not relevant for the present study, the unique 18 bp PTF recognition sequence is present in the first exon of the *PINOID* gene (AF232236, At2g34650) of the model plant species *Arabidopsis thaliana*. The E2C proteins are based on E2C(Sp1) as studied by Beerli *et al.* (14). E2C(Sp1), which differs from E2C1 in fingers 1 and 4 (Figure 2C), was included in our study as a reference PZF module. Data from the E2C series served predominantly to corroborate the 'random' PTF data. For the PTF set the individual ZFs matched with GNN triplets occurring in the 18 bp sequence 5'-GAT GTC GCC GGC GCC GAT-3', for the E2C set the GNN triplets matched with 5'-GGG GCC GGA GCC GCA GTG-3'. PTF1 and E2C1 represent 6F modules with canonical TGEKP linkers connecting all individual ZF moieties, while PTF2 and E2C2 contain longer linkers between fingers 3 and 4 and PTF3 and E2C3 between fingers 2 and 3 as well as fingers 4 and 5. Apart from these 6F final versions, precursors of these proteins were also used for experimentation.

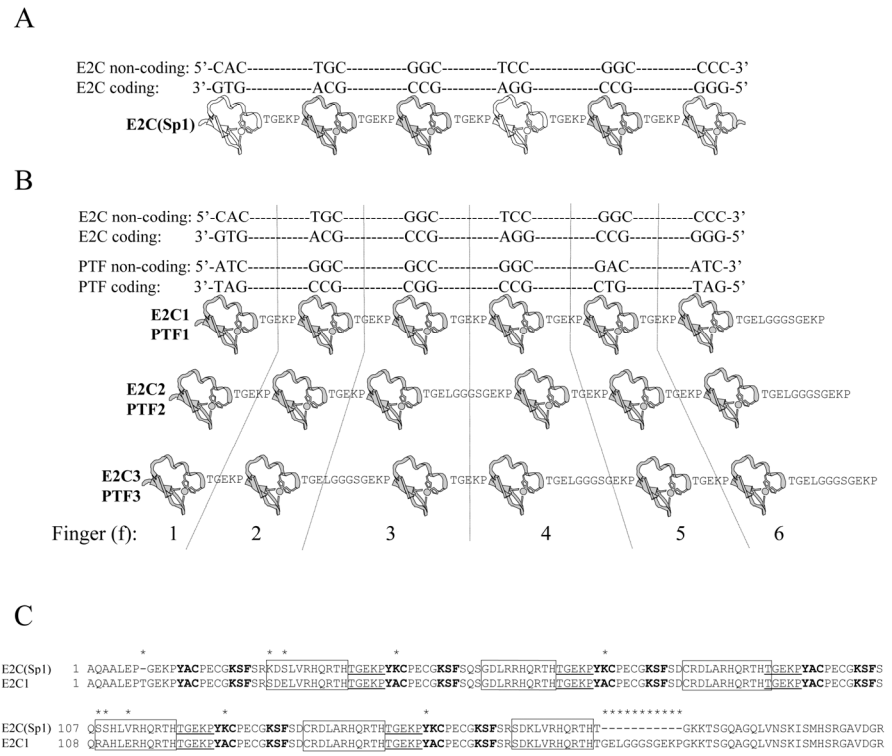

Figure 2: Overview of the 6F proteins used in this study; (**A**) The E2C(Sp1) 6F construct (sequence according to Beerli *et al.* (14)) with its cognate binding site; (**B**) 6F constructs developed with pSKN-SgrAI. The PTF and E2C contiguous binding sites are shown above the constituting fingers to indicate the design of the corresponding individual fingers f1 through f6; (**C**) Alignment of E2C(Sp1) and E2C1. Linkers are underlined, β-sheets are shown in boldface and α-helices are boxed. Differences caused by applying the optimized GNN code (6) and pSKN-SgrAI for construction of E2C1 are highlighted by asterisks.

Binding sites were designated 1P and 1E for single contiguous binding sites corresponding to PTF and E2C, respectively and 3P and 3E for three direct repeats of the same binding site (Figure 3). We accounted for any possible target site overlap problem that is inherent to some of the GNN triplet-binding ZFs, which require a G or T to be present at the 5' position in the neighbouring finger's subsite (1). Apart from a G residue downstream from the target site sequence, extra internal nucleotides were either T or TC (see Figure 3 and reference 16). As an equivalent strategy, G or GT nucleotides can be introduced (15). Both strategies agree with optimal binding affinity in case of target site overlap and do not substantially change the G/C content of the recognition

sequence. T or TC insertions were placed after every 6 bases and after 9 bases in the 18 bp PTF binding site, hence giving rise to the single target sites 1P9T9, 1P9TC9, 1P6T6T6, and 1P6TC6TC6 and the triple target sites 3P9T9, 3P9TC9, 3P6T6T6 and 3P6TC6TC6 as indicated in Figure 3. Correspondingly, interrupted target sites for E2C comprised 1E6T6T6, 1E9T9, 1E9TC9, 3E6T6T6 and 3E9T9 (Figure 3).

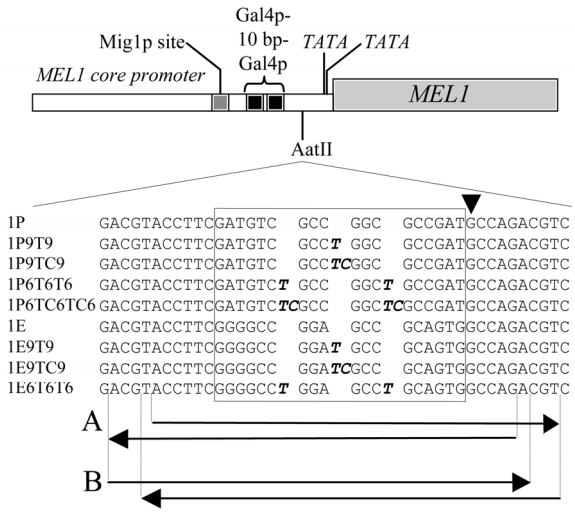

Figure 3: Organisation of the chromosomal reporter for *in vivo* assays. The *MEL1* gene (gray) is driven by the *MEL1* core promoter in which one Mig1P and two Gal4p sites are present. The two Gal4p sites are spaced by 10 bp, which abolishes background expression in the absence of galactose (26). The AatII site (5'-GACGTˇC-3') was used for the cloning of ZF binding sites. Different insertions of binding sites site are indicated in the multiple sequence alignment underneath. They were cloned using primer pairs indicated by arrows in (A) directly below the corresponding sequence. The arrow to the left represents the opposite strand of the sequence that is directly above it. Note that the AatII site is restored at the 3' coding end, but is absent at the 5' coding end allowing the selection of only sense-orientated constructs by restriction enzyme analysis. The triplets for which ZFs were designed are boxed and extra insertions within this 18 bp sequence are shown in italics. An extra G (indicated with a black triangle above the sequences) was presented following the last GNN triplet to avoid any bias in case G residue is required at this position. Constructs with three ZF target sites were obtained by ligation of three sets of primer pairs indicated in (A). Primer pairs used for end-labelling are shown in (B).

In vivo activity of VP16-PZF proteins

The ability of various VP16-PZF fusions to induce expression of a reporter gene on a chromosomal locus was tested in yeast. All single and triple, contiguous and noncontiguous target sites were cloned into the modified promoter of the *MEL1*

reporter gene (Figure 3). The different reporter constructs obtained in this way were all integrated into the *PDC6* locus in the same orientation. Growth of the 'target' strains in the presence of glucose results in active repression of the *MEL1* gene through the presence of a Mig1p site (25, 26, 31) in the *MEL1* core promoter (Figure 3). The actively repressed state of the genomic locus can be overcome by *ADH*-promoter-driven expression of the appropriate VP16-PZF fusion construct. Our assay provides an excellent model for the most common application of designed PZF-TFs. Moreover, the presence of two Gal4p sites (Figure 3) allows a direct comparison between the levels of VP16-PZF-mediated reporter gene expression in the presence of glucose and the 'maximum' level of MEL1 activity observed after galactose induction.

VP16-PTF1 and shorter precursors were tested for their ability to activate MEL1 expression in yeast strains possessing either one (Figure 4A) or three (Figure 4B) uninterrupted binding sites in the *MEL1* promoter. The *in vivo* test system clearly demonstrated that at least five ZFs are needed within a PTF module to activate expression of the reporter gene under repressed conditions. None of the VP16 constructs with four (VP16-PTFf3-6 and VP16-PTFf1-4) or three ZFs (VP16-PTFf1-3 and VP16-PTFf4-6) was able to induce detectable levels of α-galactosidase. As expected, neither PTF1 nor VP16 alone were able to trans-activate gene expression (Figure 4A and 4B); at the same time neither VP16-PTF1 nor any of the other *ADH* promoter-driven constructs mentioned in this paper were able to activate MEL1 expression in a control strain, lacking a ZF binding site in the AatII site of the modified *MEL1* promoter (data not shown).

The successful trans-activation of *MEL1* expression via the VP16-PTF1 protein in our test system allowed us to compare characteristics of the PTF1 module with those of differently designed 6F modules with longer linkers on both contiguous and interrupted binding sites. As shown in Figure 4C and 4D, the introduction of a longer linker between finger 3 and finger 4 (VP16-PTF2) dramatically reduced *MEL1* activation in the *in vivo* test system compared to VP16-PTF1, which possesses TGEKP canonical linkers between all fingers. VP16-PTF3, with 2 longer linkers dividing the 6F protein in a 3x 2F configuration, was completely inactive. Surprisingly, none of the VP16-PTF constructs was able to induce reporter gene expression in yeast strains with interruptions in the ZF binding site(s) (Figure 4C and 4D). Thus the PTF1 interaction with the contiguous 18 bp target site is the most effective interaction and is also very specific by not tolerating extra nucleotides enlarging the binding site.

Results obtained with 6F PTF modules were further extended by experiments performed with E2C(Sp1) and pSKN-SgrAI-constructed E2C versions of PZF domains. Strains carrying E2C binding sites or derivatives thereof showed

relatively high background levels (Figure 4E and 4F), indicating interference of other yeast factors. Nevertheless, VP16-6F proteins were able to further activate expression of the reporter gene. The E2C(Sp1) PZF module was somewhat less effective than the E2C1 module for *MEL1* activation on the 3E target sequence, but on a single target site it was much more effective (Figure 4E and 4F). However, reporter genes endowed with other target sites than the contiguous 18 bp sequence were also activated by VP16-E2C(Sp1). Although E2C(Sp1) might be prone to some nonspecific interactions compared to E2C1, due to the presence of two less specific ZFs in E2C(Sp1) (Figure 2C), the reason why more extended target sites seem to be recognized is not easily explained. The observed stimulation is specific for the cloned binding sites, because VP16-E2C(Sp1) did not stimulate expression in the yeast strain harbouring PTF target sites (data not shown). In any case, for the pSKN-SgrAI-generated E2C series as well as for the PTF series the interaction between a PZF domain with canonical linkers and a noninterrupted 18 bp target site proved to be the most efficient and specific for VP16-mediated trans-activation of a repressed genomic locus in yeast.

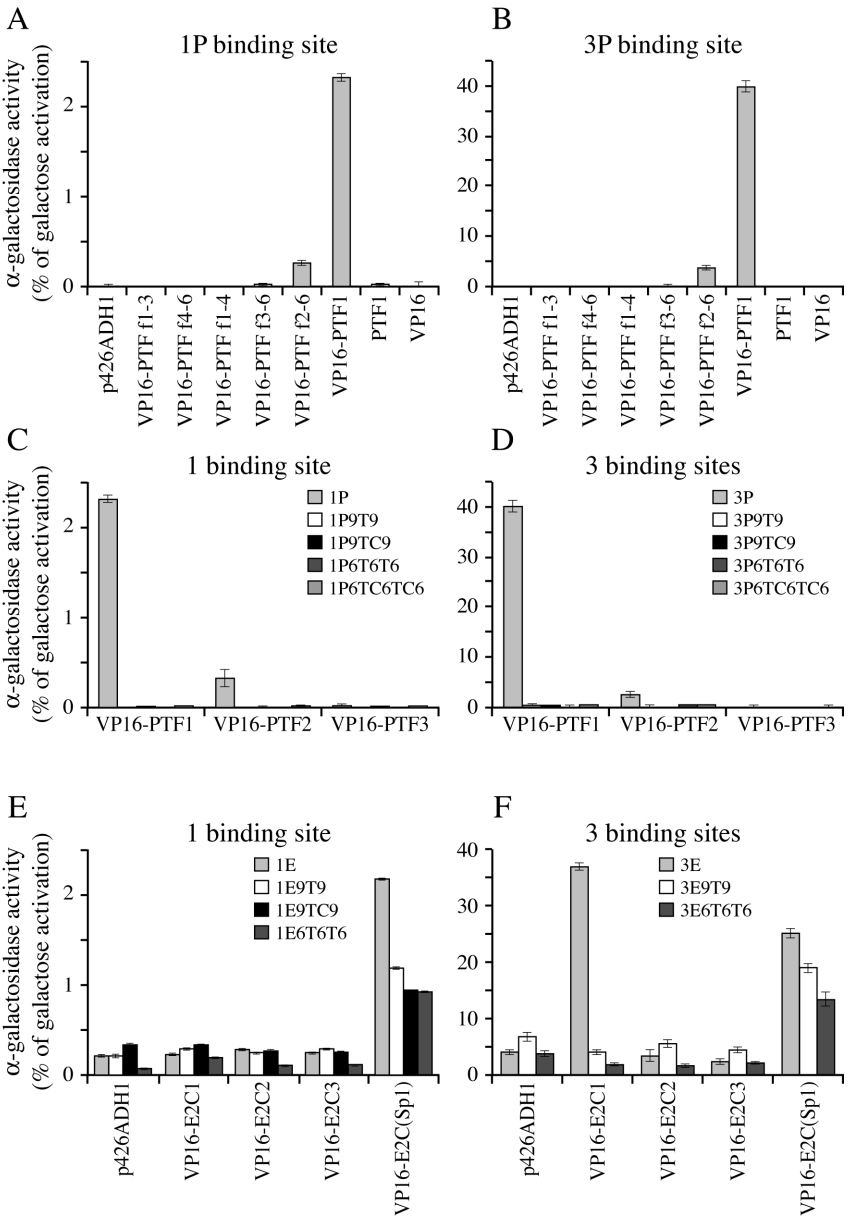
A
1P binding site
α-galactosidase activity
(% of galactose activation)
p426ADH1
VP16-PTF f1-3
VP16-PTF f4-6
VP16-PTF f1-4
VP16-PTF f3-6
VP16-PTF f2-6
VP16-PTF1
PTF1
VP16
B
3P binding site
C
1 binding site
1P
1P9T9
1P9TC9
1P6T6T6
1P6TC6TC6
VP16-PTF1
VP16-PTF2
VP16-PTF3
D
3 binding sites
3P
3P9T9
3P9TC9
3P6T6T6
3P6TC6TC6
E
1 binding site
1E
1E9T9
1E9TC9
1E6T6T6
VP16-E2C1
VP16-E2C2
VP16-E2C3
VP16-E2C(Sp1)
F
3 binding sites
3E
3E9T9
3E6T6T6

Figure 4: Effect of different PTF and E2C constructs and controls on MEL1 expression in yeast strains carrying one or three binding sites. MEL1 expression is shown as the percentage of expression measured in corresponding yeast target strains harbouring p426ADHI and grown in the presence of 2% galactose. For 3P9T9-containing strains the galactose-induced expression was found to be relatively low, but this did not affect results; (**A**) strains containing the 1P binding site with various PTF-derived constructs. VP16-PTFf4-6 possesses the last three fingers (f4 through f6) of the complete 6F PTF1 protein as indicated in Figure 2B. VP16-PTFf1-3 possesses the first three fingers. Similarly, the individual ZFs present in the four finger proteins (VP16-PTFf3-6 and VP16-PTFf1-4) and the five finger protein VP16-PTFf2-5 can be derived from Figure 2B; (**B**) strains containing the 3P binding site with various PTF-derived constructs; (**C**) strains with various single binding sites with PTF-derived constructs; (**D**) strains with various triple binding sites with PTF-derived constructs; (**E**) strains with various single binding sites with E2C-derived constructs; (**F**) strains with various triple binding sites with E2C-derived constructs.

In vitro DNA-binding properties of PZF proteins

PZF modules for *in vitro* studies have been produced via a variety of methods, but mostly by using bacterial expression systems (e.g. 6, 8, 15, 19, 32) and also after *in vitro* transcription and translation (16, 17). To exclude that the rather unexpected results of the *in vivo* gene activation assay for the differently designed PZF domains were caused by the yeast system employed in our studies, several tests were performed. Immunodetection of the FLAG-tagged PZF proteins showed that the proteins of interest were indeed produced in yeast cells and at comparable levels, irrespective of the presence of longer linkers and a VP16 domain (Figure 5). Although not affecting binding parameters, but for reasons unknown to us, the presence of a fused VP16 domain consistently led to the appearance of supershifted bands when extracts of yeast cells were used for gel shift analyses (data not shown). The bound fraction within a single shifted complex was much easier to quantify and we therefore used extract from yeast strains producing 'free' PZF domains for further *in vitro* characterization. A typical gel shift pattern and derived Scatchard plot are given in Figure 6, as are all the calculated dissociation constants (Table 2). The yeast system was validated by the observation that the apparent K_d of 0.4 nM determined for the 6F protein E2C(Sp1) present in crude yeast extract with 1E (Table 2) was very close to the value of 0.5 nM reported by Beerli *et al.* (14), who used purified *E. coli*-produced protein. The K_d values for PTF1 and E2C1 with their 18 bp target sites (1.5 and 7.1 nM, respectively) also agree with 6F data reported in the same study. A drawback of the yeast system is that the PZF protein concentrations are relatively low, which hampers accurate determination of K_d values above 10 nM. This was the case for all 3F proteins (PTFf1-3, PTFf4-6, E2Cf1-3, E2Cf4-6) with all binding sites, as well as all 6F proteins with all binding sites not listed in Table 2, including the 6F proteins with two longer linkers (PTF3 and E2C3). In all those cases, we did not pursue further analysis of *in vitro* DNA binding.

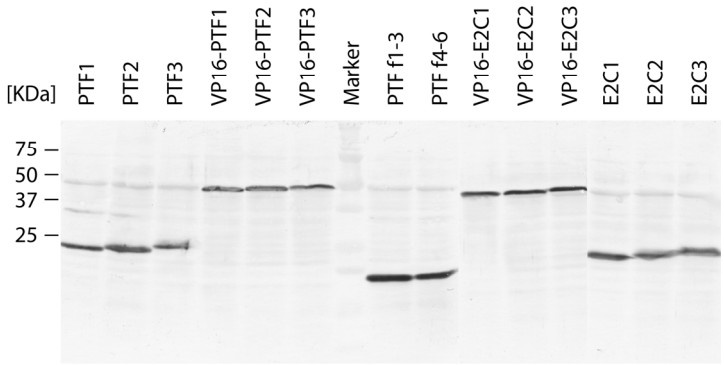

Figure 5: Western blot analysis of expression of a selection of PZF constructs in strains carrying three corresponding contiguous binding sites. All strains were grown in glucose-containing medium to an $O.D._{600nm}$ of 0.7-1.0 and diluted to an $O.D._{600nm}$ of 0.55. 100 µl of the dilution was spun down and the pellet was boiled in 15 µl 1x Laemmli buffer prior to loading.

The difference in K_d values of E2C(Sp1) and E2C1, 0.4 nM and 7.1 nM respectively, is most likely explained by the use of two different individual ZFs in the E2C1 version (Figure 2C). Although the combination of ZFs in E2C1 is more specific for the cognate $(GNN)_6$ target site when studied *in vivo*, as described above, the combination has reduced affinity compared to E2C(Sp1) in the *in vitro* binding assay. Remarkably, PTF1 bound about equally well to the 1P site as to the 1P9T9 site (Table 2) and still showed a relatively low K_d value (estimated just above 20 nM) for the interaction with the 1P9TC9 site. E2C(Sp1) and E2C1 have a clear preference for a contiguous binding site.

While 3x 2F designs of 6F proteins (PTF3 and E2C3) as produced in this study failed to interact with any binding site with a K_d below 10 nM and as such fell out of the scope of the *in vitro* analysis, the 2x 3F version of PTF (PTF2) exhibited several high affinity interactions. Considering the K_d values for the PTF series, introduction of one longer linker led to reduced affinity for the 1P site rather than to a higher affinity for the 1P9T9 site. For the 2x 3F E2C2 protein, an *in vitro* preference for a 9T9 binding site might exist, but could not be established under the experimental circumstances with both apparent K_d values estimated well above 10 nM.

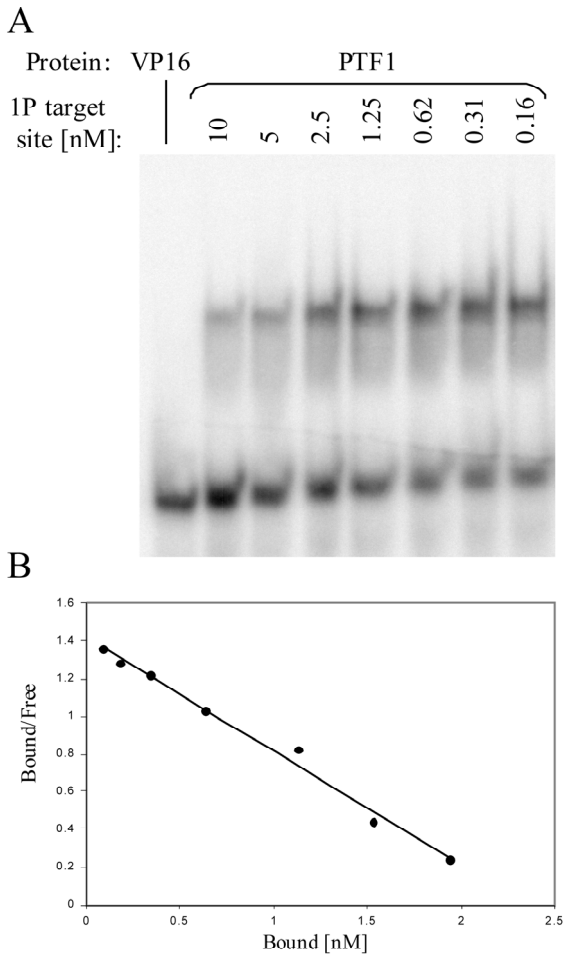

Figure 6: (**A**) Example of an electrophoretic mobility shift assay with crude extracts containing proteins VP16 and PTF1 and different amounts of unlabelled 1P target site (1P was used as a probe); (**B**) Scatchard plot corresponding to (A).

Table 2. Apparent *Kd* values in nM for E2C and PTF proteins and their target sites. Standard deviations are shown in parentheses.

Protein	Binding site	Apparent K_d [nM]
E2C(Sp1)	1E	0.4 (±0.1)
E2C1	1E	7.1 (±1.9)
PTF1	1P	1.5 (±0.7)
	1P9T9	1.9 (±0.2)
PTF2	1P	5.4 (±0.9)
	1P9T9	2.4 (±0.8)

DISCUSSION

Successful use of PZF-containing artificial transcription factors to modulate expression of eukaryotic genes requires that the PZF domains can efficiently interact with cognate recognition sites that are present within chromatin. The yeast system developed in this study offers the possibility to test transcription factors based on zinc finger technology for their potential to regulate a genomic locus and allows the production of different types of PZF proteins for *in vitro* analysis. As outlined in the introduction, existing differences in PZF structure - such as the sequence of linkers with or without a helix-capping motif as well as irregularities in the backbone of 6F designs- makes comparisons between our data and already available data rather complicated. Apart from that, a truly consistent series of 6F designs has never been investigated for *in vivo* activation on a genomic locus. The data presented in this study should be regarded as an attempt to fill in this gap, with a well established 6F peptide E2C(Sp1) included in the experiments as a reference.

While several studies have indicated that artificial transcription factors with shorter PZF modules can influence gene expression *in vivo*, even on genomic loci (32-34), we did not find any transcriptional activation of the

repressed genomic yeast reporter locus via VP16-PZF modules with less than five ZF moieties. Apparently, our *in vivo* assay system puts up a high threshold for establishing contact with the binding site. The fact that three binding sites were in most cases a prerequisite for efficient activation of *MEL1* gene expression underscored this notion. Regarding the fact that most PZF modules to be constructed will be intended for *in vivo* applications, most notably on a genomic locus, a robust assay to test a logical series of PZF designs to function on a genomic locus was required. Most likely it is just because of the challenging *in vivo* requirements in our experimental setup that several practical conclusions regarding PZF design can now be drawn with certainty. First of all, efficient transcriptional stimulation of a repressed *MEL1* reporter locus can only be achieved when a PZF domain possesses an apparent K_d value for its cognate binding site that is below 10 nM under our experimental conditions. Remarkably, as a second point, it is absolutely clear that a low K_d value for a PZF-DNA interaction is no guarantee that this interaction will be useful *in vivo*. In addition to these conclusions it should be remarked that an apparent K_d value is a rather arbitrary unit as there is no strict consensus on a uniform method for establishing these values for protein-DNA interactions. K_d values can vary considerably when determined under different conditions (17) and a calculated value can therefore not direct PZF design for *in vivo* applications. However, the third and most important conclusion to be drawn from our experiments is that a very strict 6F design with only TGEKP canonical linkers between the individual fingers performs optimally *in vivo* with great specificity for contiguous 18 bp target sites. Therefore, it is very unlikely that a particular need for longer linkers exists in order to meet any progressing unwinding of the DNA caused by an increasing amount of zinc fingers. This conclusion is likely to be a universal lead for PZF design.

In their founding paper, Moore *et al.* (16) used longer linkers of six or eight amino acids in a 3x 2F design. In comparison with a 2x 3F design (with the nine amino acid LRQKDGERP linker in between fingers 3 and 4), PZFs with a 3x 2F design interacted more specifically with the contiguous 18 bp target sequence than with mutated recognition sites. While six amino acid linkers did not yet span 1 or 2 bp insertions between 2F subsites, eight amino acid linkers could bridge 1 bp insertions rather efficiently. In that respect, the design with eight amino acid linkers thus caused some loss of specificity but, on the other hand, created the possibility to span some extra nucleotides, which is interesting when highly specific ZF moieties for particular DNA triplets are lacking. Based on the findings of Moore *et al*. (16), we thus expected that our 3x 2F peptides with two twelve amino acid linkers (PTF3 and E2C3) would allow high affinity interactions also

with extended sites, including those with 2 bp insertions. Apart from that, supposing that a strict 6F design would be suboptimal due to underwinding of DNA in the PZF-DNA complex as postulated by others (6, 14), it seemed likely that further study of PZF domains with extended linkers might reveal a window for very high affinity interactions. However, by studying the effect of our VP16-PZF fusions on the expression of the genomic target locus in yeast, it became clear the 3x 2F design with a twelve amino acid linker was inferior to the 6F and 2x 3F designs as in none of the yeast reporter lines a significant enhancement of reporter gene activity was observed. So apparently, the 3x 2F design used in our study is unable to recognize any of the experimental target sites, including those with 1 or 2 bp insertions between 2F subsites.

The 2x 3F design, with the longer linker between fingers 3 and 4 (PTF2 and E2C2) showed a dramatic reduction in the trans-activating potential compared to the 6F modules with canonical linkers between all constituting ZFs. PTF2 could only induce gene expression via the contiguous binding site to about 10% of the value observed with PTF1 (Figure 4C and 4D), while E2C2 did not induce gene expression at all (Figure 4E and 4F). Remarkably, in our studies, the presence of extra base pairs within the binding site completely abolished trans-activation (Figure 4C-4F). These findings are apparently in contrast with previous investigations. Kim and Pabo (15) performed transient transfection assays to measure transcriptional repression due to binding of PZF proteins just downstream of the transcription start site of a reporter construct. A 2x 3F type PZF domain with 3Fs connected via a twelve amino acid linker interacted strongly with interrupted binding sites in which one (G) or two (GT) base pairs were inserted between the 9 bp half sites. Moreover, the same 2x 3F protein exhibited a virtually identical and extremely high affinity interaction *in vitro* for the noninterrupted and interrupted sites. These phenomena were not observed for the 2x 3F type PZF proteins used in our study. The reasons for these discrepancies are not immediately evident, but it should be noted that the 2x 3F proteins used by Kim and Pabo are different in at least two aspects: apart from being constructed with a Zif-type backbone, while we used the extremely regular Sp1C backbone, the twelve amino acid linker LRQKDGGGSERP used by Kim and Pabo might not allow optimal stabilization of the α-helix of the flanking ZF module after DNA binding (as is discussed below) and as such might introduce some lack of specificity for related binding sites. Support for this view has been provided in a study where it was shown in a plasmid-based reporter assay in yeast cells that introduction of a helix-stabilizing TGEKGGGGERP linker between constituting 2F and 3F domains of a partly Zif-based peptide contributed to a more specific interaction over the full length of a 15 bp binding site, but did not

increase binding affinity (35). However, as mentioned in the same paper, this effect might have been primarily caused by disruption of a particularly strong interaction between a 3F domain and a 9 bp sequence present in the target sites. Apart from that, interactions with related interrupted binding sites were not investigated.

For technical reasons, we could not produce an exact copy of a published longer linker sequence with our system. However, the TGELGGGSGEKP twelve amino acid linker sequence that can be introduced via our cloning vehicle complies with all known insights regarding the general properties of standard TGEKP-type linkers. Based on structure analysis, such linkers are predicted to contribute to high-affinity DNA binding and peptide stability due to the formation of a α_L C-cap at the end of the helical structure of the finger when it interacts with DNA. The K residue most likely stabilizes the protein-DNA complex through water-mediated contacts to the phosphate backbone, while the P residue marks the start of a new ZF domain (36-38). Recent NMR analysis and further structural comparisons strongly suggest that additional stabilization in the protein-DNA complex arises from hydrogen bonding between the backbone amide of the E residue with the side-chain O^{γ} of the T residue within the linker sequence (37). Considering these data, it is not at all surprising that mutations within the TGEKP linker sequence can result in severe reductions in the affinity of the protein for its cognate DNA binding site (29, 39). For these reasons the use of the longer linkers LRQKDGERP and LRQKDGGGSERP between 3F domains (16) has met criticism as these rather divergent sequences are unlikely to allow for the stabilizing α-helix capping (17). While Moore *et al.* (16) inserted extra amino acids between the G and E residues of the TGEKP linker, others (19, 35) decided to insert extra amino acids between the E and K residues, possibly in better agreement with the data mentioned above. In our system we also chose to insert the flexible amino acid sequences downstream from the TGE sequence of the linker. Although it is of course possible that superior longer linkers than the one used in our studies can be developed, it needs to be emphasized that the TGELGGGSGEKP linker per se does by no means have a detrimental effect on DNA binding. While the individual 3F peptides bound DNA with affinities that were too low to be determined, it is evident from the PTF data that linking two 3F moieties with just the canonical linker (PTF1) or with the TGELGGGSGEKP linker (PTF2) yielded 6F peptides with very similar affinities for 18 bp and 9T9 target sites (Table 2).

Altogether, based on our findings to induce expression of a repressed genomic locus in yeast, it is highly rewarding to construct 6F domains for *in vivo* regulation of endogenous genes and in particular when this domain follows a 6F

design with canonical TGEKP linkers at all positions. We did not find any evidence for the existence of a window for more effective interactions by adapting PZF modules and/or binding sites with, respectively, extra amino acid residues within connecting linker sequences and/or extra nucleotides within binding sites. When TGEKP linkers would indeed be too short to match the helical periodicity of a 6F protein with the helical periodicity of B-DNA (15, 16), the outcome of our experiments should have been very different. However, based on structural data the notion that the TGEKP linker sequence is too short for designing a 6F protein has recently been questioned (40). As already mentioned in the introduction, it should also be pointed out that the possible benefits of longer linker sequences in alternative PZF designs were not directly compared with a 6F version (15, 16) and were either only investigated *in vitro* (16) or in combination with a transient *in vivo* assay using plasmid-based reporter genes (15). Apart from that, our use of a strict Sp1C framework for construction of PZF domains out of individual ZF modules that consist of an identical number of amino acid residues only differing for amino acid substitution within the DNA recognition helix (6) avoids any bias caused by mixing ZFs with different designs. Moreover, the $(GNN)_n$ PZF modules used as well as their intended DNA target sites avoid problems with target site overlap that might cause suboptimal interactions (1, 17).

Whether or not our efforts to refrain from introducing other variables than the different amino acids required for binding to a particular DNA triplet contribute to the performance of the PZF proteins generated is still uncertain due to lack of comparable data. However, the demonstration that the most regular PZF design has the most valuable traits for gene regulation *in vivo* strongly indicates that designing biologically active PZF domains in such a manner will be a successful and reproducible strategy.

ACKNOWLEDGEMENTS

This work was supported by grant 050-10-123 from the NWO Programma Genomics. We are grateful to Dr. Karsten Melcher (University of Frankfurt, Germany) for materials and Dr. Pieter Ouwerkerk and Dr. Paul van Heusden (University of Leiden, the Netherlands) for materials and valuable discussions. We thank Peter Hock (University of Leiden) for assistance in figure preparations and Dr. Marianne Rots (Therapeutic Gene Modulation, University of Groningen, the Netherlands) for critical reading of the manuscript.

REFERENCES

1. **Blancafort, P., Segal, D.J. and Barbas III, C.F.** (2004) Designing transcription factor architectures for drug discovery. *Mol. Pharmacol.*, **66**, 1361-1371.
2. **Jamieson, A.C., Miller, J.C. and Pabo, C.O.** (2003) Drug discovery with engineered zinc-finger proteins. *Nat. Rev. Drug. Discov.*, **2**, 361-368.
3. **Jantz, D., Amann, B.T., Gatto, G.J. and Berg, J.M.** (2004) The design of functional DNA-binding proteins based on zinc finger domains. *Chem. Rev.*, **104**, 789-799.
4. **Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C.** (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646-651.
5. **Porteus, M.H. and Carroll, D.** (2005) Gene targeting using zinc finger nucleases. *Nat. Biotechnol.*, **23**, 967-973.
6. **Segal, D.J., Dreier, B., Beerli, R.R. and Barbas III, C.F.** (1999) Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. USA*, **96**, 2758-2763.
7. **Dreier, B., Segal, D.J. and Barbas III, C.F.** (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J. Mol. Biol.*, **303**, 489-502.
8. **Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and Barbas III, C.F.** (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **276**, 29466-29478.
9. **Liu, Q., Xia, Z.Q. and Case, C.C.** (2002) Validated zinc finger protein designs for all 16 GNN triplet targets. *J. Biol. Chem.*, **277**, 3850-3856.
10. **Dreier, B., Fuller, R.P., Segal, D.J., Lund, C., Blancafort, P., Huber, A., Koksch, B. and Barbas III, C.F.** (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, 280, 35588-97
11. **Sera, T. and Uranga, C.** (2002) Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry*, **41**, 7074-7081.
12. **Jamieson, A.C., Miller, J.C. and Pabo, C.O.** (2003) Drug discovery with engineered zinc-finger proteins. *Nat. Rev. Drug Discov.*, **2**, 361-368.
13. **Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas III, C.F.** (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. USA*, **94**, 5525-5530.
14. **Beerli, R.R, Segal, D.J., Dreier, B. and Barbas III, C.F.** (1998) Toward controlling gene expression at will: specific regulation of the *erbB-2/HER-2* promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. USA*, **95**, 14628-14633.
15. **Kim, J.S. and Pabo, C.O.** (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA*, **95**, 2812-2817.

16. **Moore, M., Klug, A. and Choo, Y.** (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two finger units. *Proc. Natl. Acad. Sci. USA*, **98**, 1437-1441.

17. **Moore, M., Choo, Y. and Klug, A.** (2001) Design of polyzinc finger peptides with structured linkers. *Proc. Natl. Acad. Sci. USA*, **98**, 1432-1436.

18. **Nomura, W. and Sugiura, Y.** (2003) Effects of length and position of an extended linker on sequence-selective DNA recognition of zinc finger peptides. *Biochemistry*, **42**, 14805-14813.

19. **Imanishi, M., Yan, W., Morisaka, T. and Sugiura, Y.** (2005) An artificial six-zinc finger peptide with polyarginine linker: Selective binding to the discontinuous DNA sequences. *Biochem. Biophys. Res. Commun.*, **333**, 167-173.

20. **Sadowski, I., Ma, J., Triezenberg, S. and Ptashne, M.** (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, **335**, 563-564.

21. **Cousens, D.J., Greaves, R., Goding, C.R. and O'Hare, P.** (1989) The C-terminal 79 amino acids of the herpes simplex virus regulatory protein, Vmw65, efficiently activate transcription in yeast and mammalian cells in chimeric DNA-binding proteins. *EMBO J.*, **8**, 2337-2342.

22. **Mumberg, D., Müller, R. and Funk, M.** (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, **156**, 119-122.

23. **Dunn, J.J., Krippl, B., Bernstein, K.E., Westphal, H. and Studier, F.W**. (1988) Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene*, **68**, 259-266.

24. **Lassner, M.W., Jones, A., Daubert, S. and Comai, L.** (1991) Targeting of T7 RNA polymerase to tobacco nuclei mediated by an SV40 nuclear location signal. *Plant Mol. Biol.*, **17**, 229-234.

25. **Melcher, K., Sharma, B., Ding, W.V. and Nolden, M.** (2000) Zero background reporter plasmids. *Gene*, **247**, 53-61.

26. **Melcher, K. and Xu, H.E.** (2001) Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete *GAL* gene repression. *EMBO J.*, **20**, 841-851.

27. **Meijer, A.H., Ouwerkerk, P.B.F. and Hoge, J.H.C.** (1999) Vectors for transcription factor cloning and target site identification by means of genetic selection in yeast. *Yeast*, **14**, 1407-1416.

28. **Gietz, D., St Jean, A., Woods, R.A. and Schiestl, R.H.** (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucl. Acids Res.*, **20**, 1425.

29. **Clemens, K.R., Zhang, P., Liao, X., McBryant, S.J., Wright, P.E. and Gottesfeld, J.M.** (1994) Relative contribution of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *J. Mol. Biol.*, **244**, 23-35.

30. **Segal, D.J.** (2002) The use of zinc finger peptides to study the role of specific transcription factor binding sites in the chromatin environment. *Methods*, **26**, 76-83.

31. **Nehlin, J.O., Carlberg, M. and Ronne, H.** (1991) Control of yeast *GAL* genes by MIG1 repressor: a transcriptional cascade in the glucose response. *EMBO J.*, **10**, 3373-3377.

32. **Liu, P.Q., Rebar, E.J., Zhang, L., Liu, Q., Jamieson, A.C., Liang, Y., Qi, H., Li, P.X., Chen, B., Mendel, M.C., Zhong, X., Lee, Y.L., Eisenberg, S.P., Spratt, S.K., Case, C.C. and Wolffe, A.P.** (2001) Regulation of an endogenous locus using a panel of

designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J. Biol. Chem.*, **276**, 11323-11334.

33. **Zhang, L., Spratt, K., Liu, Q., Johnstone, B., Qi, H., Raschke, E.E., Jamieson, A.C., Rebar, E.J., Wolffe, A.P. and Case, C.C.** (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. *J. Biol. Chem.*, **275**, 33850-33860.

34. **Beerli, R.R., Dreier, B. and Barbas III, C.F.** (2000) Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl. Acad. Sci. USA*, **97**, 1495-1500.

35. **Bartsevich, J.J. and Juliano, R.L.** (2000) Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries. *Mol. Pharmacol.*, **58**, 1-10.

36. **Elrod-Erickson, M., Rould, M.A., Nekludova, L. and Pabo, C.O.** (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171-1180.

37. **Laity, J.H., Dyson, H.J. and Wright, P.E.** (2000) DNA-induced α-helix capping in conserved linker sequences is a determinant of binding affinity in Cys_2-His_2 zinc fingers. *J. Mol. Biol.*, **295**, 719-727.

38. **Wuttke, D.S., Foster, M.P., Case, D.A., Gottesfeld, J.M. and Wright, P.E.** (1997) Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA Sequence: determinants of affinity and sequence specificity. *J. Mol. Biol.*, **273**, 183-206.

39. **Choo, Y. and Klug, A.** (1993) A role in DNA binding for the linker sequences of the first three zinc fingers in TFIIIA. *Nucl. Acids Res.*, **21**, 3341-3346.

40. **Peisach, E. and Pabo, C.O.** (2003) Constraints for zinc finger linker design as inferred from X-ray crystal structure of tandem Zif268-DNA complexes. *J. Mol. Biol.*, **330**, 1-7.