Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 5525–5530, May 1997
Biochemistry

Design of polydactyl zinc-finger proteins for unique addressing within complex genomes

(molecule design/zinc-finger proteins/genome addressing/transcriptional regulation/gene therapy)

QIANG LIU, DAVID J. SEGAL, JAYANT B. GHIARA, AND CARLOS F. BARBAS III*

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

Communicated by Richard A. Lerner, The Scripps Research Institute, La Jolla, CA, March 28, 1997 (received for review March 24, 1997)

ABSTRACT **Zinc-finger proteins of the Cys_2-His_2 type represent a class of malleable DNA-binding proteins that may be selected to bind diverse sequences. Typically, zinc-finger proteins containing three zinc-finger domains, like the murine transcription factor Zif268 and the human transcription factor Sp1, bind nine contiguous base pairs. To create a class of proteins that would be generally applicable to target unique sites within complex genomes, we have utilized structure-based modeling to design a polypeptide linker that fuses two three-finger proteins. Two six-fingered proteins were created and demonstrated to bind 18 contiguous bp of DNA in a sequence-specific fashion. Expression of these proteins as fusions to activation or repression domains allows transcription to be specifically up- or down-modulated within human cells. Polydactyl zinc-finger proteins should be broadly applicable as genome-specific transcriptional switches in gene therapy strategies and the development of novel transgenic plants and animals.**

From the simplest of organisms to the most complex, transcriptional regulation is achieved primarily by proteins that bind nucleic acids. The advent of genomic sequencing and the availability of the complete sequences of several genomes provide new opportunities to study biology and to develop therapeutic strategies through specific modulation of the transcription of target genes. Of the DNA-binding motifs that have been manipulated by design or selection, the TFIIIA-related Cys_2-His_2 zinc-finger proteins have demonstrated the greatest potential for manipulation into general and specific transcription factors (1, 2). Each Cys_2-His_2 zinc-finger domain consists of approximately 30 amino acids and typically binds 3 base pairs of double-stranded DNA sequence (3, 4).

Specific delivery of a DNA-binding protein to a single site within a genome as complex as that found in humans, 3.5 billion bp, requires an address of at least 16 bp. Statistically assuming random base distribution, a unique 16-bp sequence will occur only once in 4^{16} or 4.3 billion nucleotides, roughly the same size of a human genome (3.5×10^9 bp). An 18-bp address would be specific within 68 billion base pairs of sequence. The 18-bp address could be specified by a protein containing six zinc fingers if the periodicity of the protein domains could match that of the DNA over this extended sequence. An address of this length would be sufficient to uniquely specify any locus within all known genomes. Although natural proteins containing long polydactyl arrays of zinc-finger domains have been inferred from sequence, no zinc-finger proteins have been demonstrated to bind such a long, contiguous DNA sequence. Structural studies of the five-finger human glioblastoma (GLI) protein–DNA complex (5) and biochemical studies of the nine-finger protein TFIIIA (6, 7) have demonstrated that DNA binding in these polydactyl proteins is dominated by the interactions of a select few fingers. Sequence-specific binding of more than three contiguous zinc-finger domains within polydactyl proteins has yet to be observed.

Manipulations of the zinc-finger motif have been guided by the x-ray crystal structures of the zinc-finger protein Zif268 (3, 8), the human GLI oncogene (5), and the *Drosophila* transcription factor Tramtrack (4) in complex with DNA. Previous work has demonstrated that novel zinc-finger proteins can be designed or selected from phage display libraries to bind diverse sequences with high affinity and specificity (9–16). However, selected proteins in each of these studies have had only three zinc fingers that bind a 9-bp site. While these reports provide hope for generating zinc fingers that could bind any given sequence, the practical application of this technology to target specific genes within the complex human genome requires the specific recognition of a much longer DNA sequence, at least 16 bp. To address this shortcoming we have designed, expressed, and characterized proteins with six contiguous zinc-finger domains that bind a contiguous 18-bp site.

MATERIALS AND METHODS

Design of Linker Sequence to Connect the Three-Finger Proteins. Coordinates for the Zif268–DNA complex were obtained from the Brookhaven Protein Data Bank. Model building was done with INSIGHTII (Biosym Technologies, San Diego). A continuous 20-bp double-stranded DNA molecule with a six-finger binding site (18 bp) was built from the coordinates for the DNA strands in the Zif268 complex (3). Two molecules of the three-finger protein were reintroduced onto each 9-bp half-site by overlapping the Zif268–DNA complex onto the modeled DNA. It was apparent that the linker length required to connect the F3 α-helix to the first β-strand of F4 was compatible with the length of the natural linker peptides, TGQKP and TGEKP. Hence, a peptide linker, TGEKP, was constructed between F3 and F4 after trimming off the extra residues from the C and N termini of the F3 and F4, respectively. The linker was built so as to maintain the positioning and hydrogen bond characteristics observed in the two natural linker regions of Zif268.

Plasmid Construction. The six-finger protein, C7-C7, was constructed by linking two C7 proteins with the TGEKP linker peptide. Two C7 DNA fragments were created by PCR with two different sets of primers using pET3a-C7 as template (15), so the 5′ C7 was flanked by *Xho*I and *Cfr*101 sites at the 5′ and 3′ ends, respectively, and the 3′ C7 was flanked by *Cfr*101 and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/945525-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

Abbreviations: Zif268, murine transcription factor Zif268; MBP, maltose-binding protein; SV40, simian virus 40.

*To whom reprint requests should be addressed at: Department of Molecular Biology, BCC-515, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. e-mail: carlos@scripps.edu.

*Spe*I sites. The primer pairs for the generation of the 5′ C7 are 5′-GAGGAGGAGGAGGGATCCATGCTCGAG-CTCCCCTATGCTTGCCCTG-3′ and 5′-GAGGAGGAG-ACCGGTATGGATTTTGGTATGCCTCTTGCG-3′; for the 3′ C7 they are 5′-GAGGAGGAGACCGGTGAGA-AGCCCTATGCTTGCCCTGTCGAGTCCTGCGA-TCGCCGC-3′ and 5′-GAGGAGGAGACTAGTTCTAGA-GTCCTTCTGTC-3′. Then these two C7 DNA fragments were ligated into a pGEX-2T (Pharmacia) vector that had been modified with *Xho*I and *Spe*I sites introduced between the pGEX-2T cloning sites *Bam*HI and *Eco*RI. The *Cfr*101 enzyme site between the two C7 fragments encodes amino acids TG, part of the TGEKP linker peptide. The fidelity of the C7-C7 sequence was determined by DNA sequencing. The C7-C7 DNA fragment was then cut out from the pGEX-2T construct with *Xho*I/*Spe*I and cloned into a modified pMal-c2 (New England Biolabs) bacterial expression vector for the expression of C7-C7 maltose fusion proteins. For transfection experiments, the C7-C7 DNA fragment was removed via *Bam*HI/*Eco*RI excision and ligated into the corresponding sites of pcDNA3, a eukaryotic expression vector (Invitrogen). Like the generation of C7-C7 protein, the Sp1C-C7 protein was created by linking the PCR products of Sp1C (17) and C7, which were flanked with *Xho*I/*Cfr*101 and *Cfr*101/*Spe*I, respectively. Then the Sp1C-C7 fragments were ligated into the pcDNA3 eukaryotic expression vector or into the pMal-c2 bacterial expression vector. The DNA sequence of the Sp1C-C7 protein was confirmed by DNA sequencing.

For reporter gene assays of activation, the reporter genes were constructed by inserting six forward tandem repeats of the individual binding sites into the *Nhe*I site at the upstream of the simian virus 40 (SV40) promoter of pGL3-promoter (Promega). In the reporter gene assays for repression, six forward tandem copies of the C7-C7 binding sites were placed upstream of the SV40 promoter at the *Nhe*I site of pGL3-control (Promega).

Expression and Purification of Zinc-Finger Proteins. Proteins were overexpressed as fusions with the maltose binding protein using the Maltose fusion and purification system (New England Biolabs). The maltose fusion proteins were purified by using amylose resin-filled affinity column according to the manufacturer's instructions. Fusion proteins were determined to be greater than 90% homogeneous as demonstrated by Coomassie blue-stained SDS/PAGE gels. Protein concentrations were determined by amino acid analysis.

Gel Mobility-Shift Assays. To produce probes used in the gel mobility-shift assay, double-stranded oligonucleotides containing TCGA overhangs at the 5′ end of each strand were labeled with α-[^{32}P]dATP. The sequences of the primary strands within the duplex regions were 5′-GATGTATG-TAGCGTGGGCGGCGTGGGCGTAAGTAATGC-3′ (C7-C7 site), 5′-GATGTATGTAGCGTGGGCGGGGGC-GGGGTAAGTAATGC-3′ (SP1C-C7 site), 5′-GATGTATG-TAGCGGCGGCGGCGGCGGCGTAAGTAATGC-3′ [$(GCG)_6$ site], 5′-GATGTATGTAGCGTGGGCGTAAG-TAATGC-3′ (C7 site), and 5′-GATGTATGTAGGGG-CGGGGTAAGTAATGC-3′ (Sp1C site). For each binding

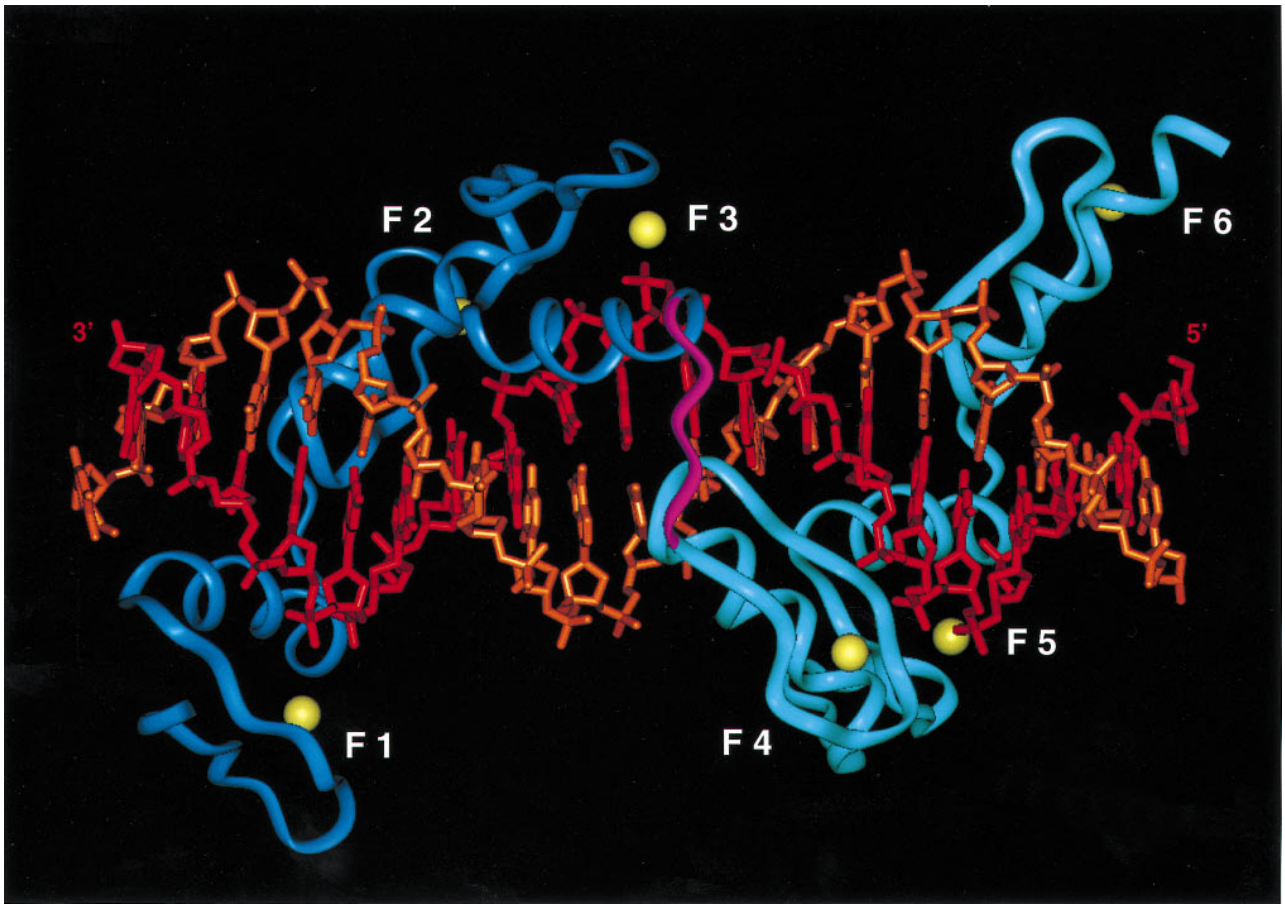

FIG. 1. Model of a six zinc-finger–DNA complex. The zinc fingers (numbered F1–F3 and F4–F6) are based on the structure of the three zinc-finger protein Zif268 (3). The six-finger protein was constructed by connecting F3 to F4 by building in the conserved linker sequence TGEKP, represented in magenta, with INSIGHTII. The guanine-rich strand, with which most of the base-specific contacts are made, is represented in red (3′-TGCGGGTGCGGCGGGTGCGA-5′). The zinc atoms are represented in yellow. The structure of the modeled linker is similar to that of the natural linker peptides between F1 and F2 (TGQKP) and between F2 and F3 (TGEKP), respectively (with an rmsd of 0.3 Å for backbone atoms). The modeled linker also retains the general position and hydrogen bond characteristics observed for the natural linkers.

reaction, 1.2 μg of poly(dI-dC) and 30 pM of labeled oligo were incubated with the C7-C7 maltose fusion protein (MBP-C7-C7) or Sp1C-C7 maltose fusion protein (MBP-Sp1C-C7) in 20 μl of 1× binding buffer (10 mM Tris·Cl, pH 7.5/100 mM KCl/1 mM $MgCl_2$/1 mM DTT/0.1 mM $ZnCl_2$/10% glycerol/0.02% Nonidet P-40/0.02% BSA) for 30 min at room temperature. The reaction mixtures were then run on a 5% nondenaturing polyacrylamide gel with 0.5× TBE buffer at room temperature. The radioactive signals were quantitated with a PhosphorImager (Molecular Dynamics) and recorded on x-ray films. The data were then fit using the KALEIDAGRAPH program (Synergy Software, Reading, PA) to give the equilibrium dissociation constants.

DNaseI Footprinting Analysis. DNaseI footprinting was performed using the SureTrack Footprinting Kit (Pharmacia) according to the manufacturer's instructions. Two 220-bp DNA fragments containing single C7-C7 and Sp1C-C7 binding sites were synthesized by fusion PCRs, and then cloned into pcDNA3 vector. Two sets of primers: (*i*) *Eco*RIfootF, 5′-GAGGAGGAGGAATTCCGACATTTATAATGAACGTGAATTGC-3′, and C7-C73>5, 5′-TGCGCCCACGCCGCCCACGCGATGATTGGGAGCTTTTTTTGCACG-3′; and (*ii*) C7-C75>3, 5′-TCGCGTGGGCGGCGTGGGCGCAAAAAATTATTATCATGGATTCTAAAACGG-3′, and *Not*IfootB, 5′-GAGGAGGAGGCGGCCGCAGGTAGATGAGATGTGACGAACGTG-3′ were used with pGL3-promoter (Invitrogen) as template to generate the two overlapping subfragments of the C7-C7 footprinting probe. Then the two PCR products were used as template with *Eco*RIfootF and *Not*IfootB as primers to generate the 220-bp C7-C7 footprinting probe. The footprinting probe containing the Sp1C-C7 binding site was constructed the same way as the C7-C7 probe, except the oligos Sp1C-C73>5, 5′-TGCCCCGCCCCCGCCCACGCGATGATTGGGAGCTTTTTTTGCACG-3′, and Sp1C-C75>3, TCGCGTGGGCGGGGGGCGGGGCAAAAAATTATTATCATGGATTCTAAAACGG-3′ were used here to replace the C7-C73>5 and C7-C75>3 oligos. pcDNA3 vectors containing the binding sites for C7-C7 or Sp1C-C7 were then digested with *Eco*RI and *Not*I. The 220-bp fragments were gel-purified and end-labeled using Klenow polymerase and [^{32}P]dATP. Because there are no thymines in the *Not*I site, only the strand extended at the *Eco*RI site is radiolabeled. Approximately 2.3 × 10^4 cpm (0.1 pM) was then used in a 50-μl binding reaction containing 20 μg/ml of either BSA or purified binding protein (300 nM) in 1× binding buffer and 60 μg/ml poly(dI-dC) DNA. Optimal binding conditions were determined from gel shift assays. This reaction was incubated for 30 min at room temperature prior to the addition of 1 unit of DNaseI.

Luciferase Reporter Gene Assays. For the reporter gene assay experiments, 2.5 μg of the individual reporter DNA and 2.5 μg of the C7-C7-VP16 expression plasmids were transfected by calcium phosphate method (18) into HeLa cells that were passed the day before at 3 × 10^5 cells per well of the six-well culture plate. Eighteen hours later, the cells were washed and replenished with Dulbecco's modified Eagle's medium containing 10% newborn calf serum (GIBCO/BRL). Two days later, the cells were washed, lysed, and measured for luciferase activity using Wallac's 96-well LB96 luminometer with the luciferase assay system (Promega). The internal β-galactosidase activity control was measured by using a β-galactosidase reporter gene assay system (Tropix, Bedford, MA).

RESULTS AND DISCUSSION

Design of a Linker Peptide. To explore the possibility of connecting two three-finger protein molecules with a linker peptide, we undertook computer modeling studies based on the structure of the three zinc-finger Zif268–DNA complex (3). A six-finger–DNA complex, modeled by connecting finger 3 (F3) of Zif268 to finger 1 of a second Zif268 molecule [henceforth designated finger 4 (F4)], would help determine the length and sequence of a compatible linker peptide to be used in the construction of six-finger proteins. Study of the model (Fig. 1) suggested that it should be possible to produce a six-finger protein with a Thr-Gly-Glu-Lys-Pro (TGEKP) pentapeptide linker between F3 and F4 and that this polydactyl protein would most likely bind DNA containing the 18-nt site 5′-GCGTGGGCGGCGTGGGCG-3′. This pentapeptide constitutes the consensus peptide most commonly found linking zinc-finger domains in natural proteins. Prior to construction of the model, we had anticipated that the consensus peptide TGEKP would be insufficient to keep the periodicity of the

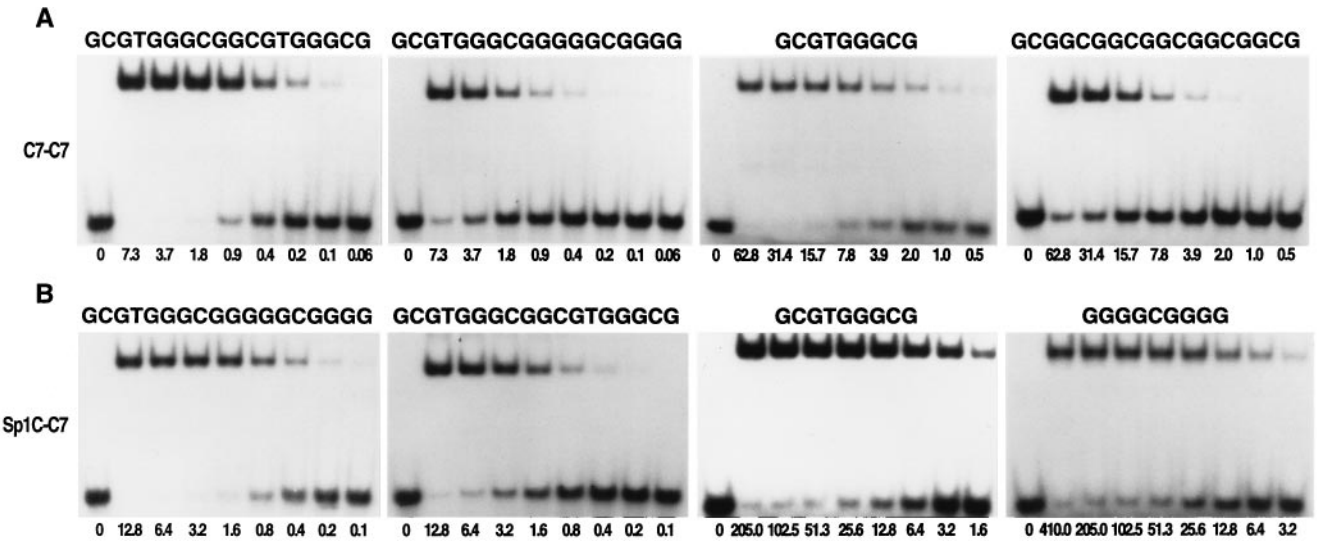

FIG. 2. Binding of the maltose-binding protein fusions (MBP)-C7-C7 and MBP-Sp1C-C7 with duplex DNA oligonucleotides containing various target sequences. (*A*) MBP-C7-C7 protein was used to shift the double-stranded DNA probes containing the target sequences listed on top of each panel [from left to right: C7-C7 site, Sp1C-C7 site, C7 site, and $(GCG)_6$ site]. The protein concentration is given in nanomoles (nM) beneath each lane with a 2-fold serial dilution from left to right in each panel. (*B*) MBP-SP1C-C7 protein was titrated into gel shift reactions with probes containing target sequences (from left to right: Sp1C-C7 site, C7-C7 site, C7 site, and Sp1C site) as listed on top of each panel. The protein concentration is labeled in nM beneath each lane, with a 2-fold serial dilution from left to right in each panel.

zinc-finger domains in concert with that of the DNA over this extended sequence because no natural zinc-finger proteins have been demonstrated to make DNA base-specific contacts with more than three contiguous zinc-finger domains, even though natural proteins containing more than three zinc-finger domains are quite common. Comparative studies of the constructed TGEKP linker with the natural linkers observed in the Zif268 structure indicated that this linker is as optimal a linker peptide as any novel linker sequence that could be designed. In binding this extended site the modeled six-fingered protein follows the major groove of DNA for approximately two turns of the helix. Such extended contiguous binding within the major groove of DNA has not been observed with any known DNA-binding protein.

Characterization of Affinity and Specificity of Two Six-Finger Proteins. To test our model we constructed two six-finger proteins. In the first protein designated C7-C7, two copies of C7, a phage display-selected Zif268 variant (15), were linked together via the TGEKP peptide. A second six-finger protein, Sp1C-C7, combines a designed variant of the three-finger Sp1 transcription factor, Sp1C (17), with the three-finger C7. The C7, Sp1C, C7-C7, and Sp1C-C7 proteins were overexpressed in *Esherichia coli* as fusions with maltose-binding protein (MBP) and purified. The affinities and specificities of these proteins were determined by electrophoretic mobility-shift assays (Fig. 2). The results of these studies are given in Table 1. The six-finger proteins C7-C7 and Sp1C-C7 bind their 18-bp target sequences, 5′-GCGTGGGCGGCGT-GGGCG-3′ and 5′-GCGTGGGCGGGGGCGGGG-3′, respectively, with 68- to 74-fold enhanced affinity relative to the three-finger C7 or Sp1C proteins. To examine the specificity of the C7-C7 protein we studied its binding to probes containing 4-bp differences in one half-site (Sp1C-C7 probe; 5′-GCGTGGGCGGGGGCGGGG-3′) and 2-bp differences in each of the finger 2 and 5 binding sites [$(GCG)_6$ probe; 5′-GCGGCGGCGGCGGCGGCG-3′). These studies revealed a preference for the designed target probe of 5-fold relative to the Sp1C-C7 probe and 37-fold preference over the $(GCG)_6$ probe. This, together with binding studies using a probe containing the 9-bp C7 half-site, 5′-GCGTGGGCG-3′, demonstrates that mutations spread across the binding site are more disruptive to binding than ones that occur at one end of the binding site. This behavior is expected of polydactyl proteins because mutations within a given finger binding site should affect the ability of both neighbor fingers to obtain their optimal mode of binding. Similar results were obtained for the Sp1C-C7 protein (Table 1). To further examine the binding of the C7-C7 and Sp1C-C7 proteins, DNaseI footprinting assays were performed (Fig. 3). These studies demonstrated that both

Table 1. The equilibrium dissociation constants of zinc-finger proteins

Zinc-finger protein	Binding site	K_d, nM
C7-C7	C7-C7 GCGTGGGCGGCGTGGGCG	0.46
	Sp1C-C7 GCGTGGGCGGGGGCGGGG	2.4
	C7 GCGTGGGCG	6.1
	$(GCG)_6$ GCGGCGGCGGCGGCGGCG	17.3
Sp1C-C7	Sp1C-C7 GCGTGGGCGGGGGCGGGG	0.55
	C7-C7 GCGTGGGCGGCGTGGGCG	1.8
	C7 GCGTGGGCG	4.9
	Sp1C GGGGCGGGG	27.4
C7	C7 GCGTGGGCG	31.8
Sp1C	Sp1C GGGGCGGGG	40.8

The binding affinities of purified MBP-C7-C7, MBP-Sp1C-C7, MBP-C7, and MBP-Sp1C to the above-listed target sequences were measured by protein titration with gel mobility-shift assays, and are expressed as dissociation constants K_d, which were determined with the KALEIDAGRAPH program.

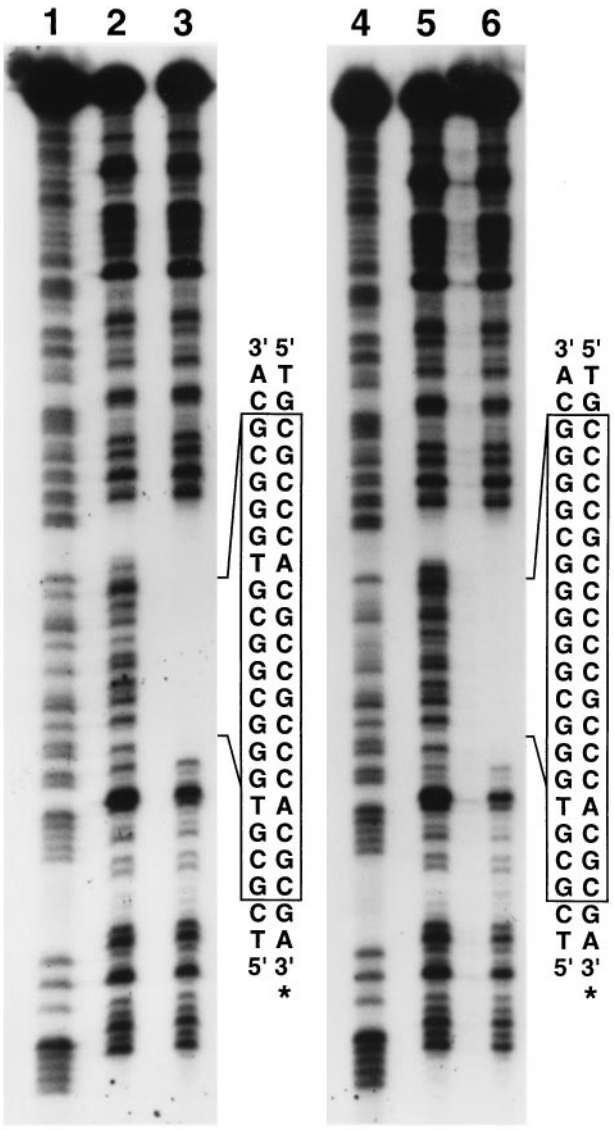

FIG. 3. DNaseI footprint of MBP-C7-C7 and MBP-Sp1C-C7. A 220-bp radiolabeled fragment containing the binding site for MBP-C7-C7 (lanes 1–3) or MBP-Sp1C-C7 (lanes 4–6) was incubated with either 20 μg/ml of BSA (lanes 2 and 4) or the cognate binding protein (300 nM, lanes 3 and 6) in 1× binding buffer for 30 min. DNaseI footprinting was then performed using the SureTrack Footprinting Kit (Pharmacia) according to the manufacturer's instructions. Boxed region indicates the binding site sequence. Asterisk indicates the 3′-labeled strand. Lanes 1 and 4: G + A ladders.

MBP fusions protected DNA-binding sites slightly longer than the 18-bp target site. This is most likely due to steric blockade by the MBP fusion at the N terminus of the protein and a decapeptide epitope tag at the C terminus of the protein.

Transcriptional Activation and Repression. To examine the specificity of the six-finger proteins in living cells, we constructed eukaryotic expression vectors that fuse the C7-C7 and Sp1C-C7 proteins to the nuclear localization signal from the SV40 large T antigen (Pro-Lys-Lys-Arg-Lys-Val) (19) and the transcriptional activation domain from the herpes simplex virus VP16 protein. These plasmids were cotransfected into the human HeLa cell line, with reporter plasmids expressing the firefly luciferase gene under control of the SV40 promoter (pGL3-promoter). The reporter plasmids were constructed with C7-C7, Sp1C-C7, C7, and $(GCG)_6$ binding sites placed upstream of the SV40 promoter. The results of these studies with the C7-C7 protein are given in (Fig. 4*A*). Both C7-C7 and

Sp1C-C7 (data not shown) stimulated the activity of the promoter in a dose-dependent fashion. In the C7-C7 case, a >300-fold stimulation of expression above background was observed for plasmids containing the C7-C7 binding site, whereas a similar concentration of protein stimulated expression of plasmids containing the C7 and Sp1C-C7 only about 3-fold. The *in vivo* specificity of this protein, indicated by an approximately 100-fold activation of the reporter plasmid bearing the proper binding site over plasmid containing a variant of the binding site, exceeds that determined in the *in vitro* binding assays described in Table 1 by approximately 5- to 10-fold. This enhanced specificity may be due to interactions generated by the maltose-binding protein at the N terminus of the C7-C7 fusion protein, which was used in the *in vitro* binding assays. Difficulty in producing the purified protein in a fully folded natural state may also contribute to the reduced specificity in the *in vitro* assays.

A distinct advantage of protein-based strategies over antisense or small synthetic molecules is that they can modulate transcription up or down by binding to sites within promoter or coding regions or to sites removed a great distance from these regions. The previous experiments demonstrated specific activation or up-regulation of a target gene. However, since sequence-specific repression may be applicable in gene therapy strategies that target the transcription of viral genomes such as HIV-1, we have also studied the ability of the polydactyl C7-C7 protein to down-modulate transcription of reporter plasmids expressing the firefly luciferase gene under control of the SV40 promoter and enhancer (pGL3-control). In this study we have constructed a eukaryotic expression vector that fuses the C7-C7 protein to the nuclear localization signal from the SV40 large T antigen and to a 43-amino acid peptide representing the minimal human Krüppel-associated box-A (KRAB-A) repression domain (20–22). The KRAB-A domain is one of the strongest transcriptional repressors identified to date in higher organisms, and functions in a distance- and orientation-independent manner. The KRAB-A domain has been identified in approximately one-third of all natural Cys_2-His_2 zinc-finger proteins and is the only widely distributed transcriptional repressor motif known (23). Cotransfection of these plasmids into HeLa cells resulted in a 93% decrease in luciferase transcription compared with controls (Fig. 4*B*). Together with the activation experiments described above, these results demonstrate that designed six-finger proteins can specifically address their target sequences in living cells.

In conclusion, we have demonstrated that polydactyl arrays of as many as six zinc-finger domains can bind a contiguous 18-bp DNA sequence with high affinity and specificity and that these proteins can function in human cells to either activate or repress transcription. Previous studies have demonstrated the feasibility of evolving three-fingered proteins to bind novel sequences using the technique of phage display (10, 11, 13, 15, 16). Together with the data presented here, a general strategy for the generation of proteins with 18-bp recognition sites is the independent evolution of two three-fingered proteins using phage display followed by fusion of these proteins with the consensus Krüppel-type linker peptide TGEKP. This strategy should be of general utility in the construction of zinc-finger proteins with genome-specific addressing potential, though exceptions may arise. Transcriptional control at the DNA level presents distinct advantages over antisense (24), ribozymes (25), and dominant-negative strategies (26), because only a single site needs to be occupied as compared with targeting the many copies of mRNA or protein that could be produced from a single gene. Although there have been recent stunning advances in the design of DNA-binding pyrrole-imidazole polyamides (27), genetically codable transcriptional modifiers will always have distinct advantages in certain applications.

We believe that polydactyl zinc-finger proteins of the type described here will find broad application in future gene therapy strategies. With respect to strategies that target HIV-1, a polydactyl zinc-finger protein specific for a conserved region of the viral genome and tethered to a KRAB-A repressor domain should inhibit transcription of the virus and hence its replication. Further, a similar protein with specificity for the human chemokine receptor gene CCR5 (28) might be used in a prophylactic gene therapy strategy to produce resistance to infection by HIV-1. Application of these proteins to engender viral resistance or to otherwise alter the phenotype of transgenic plants and animals should likewise be broad. Although we have focused our discussion on DNA recognition, zinc-finger proteins of the Cys_2-His_2 type have also been demonstrated to specifically bind DNA–RNA hybrids (14) and should be malleable via selection to binding double-stranded RNA sequences, extending the scope and application of this recognition motif even further.

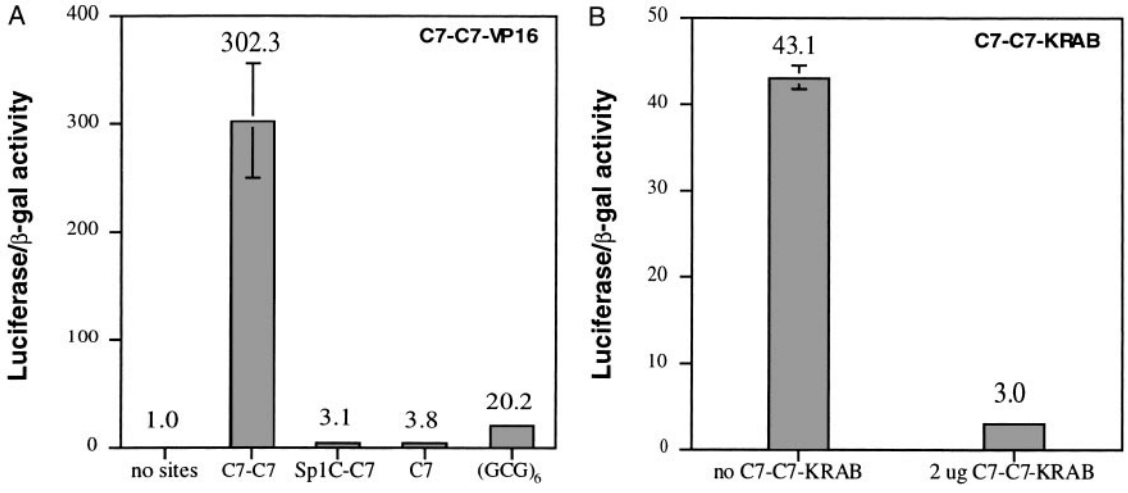

FIG. 4. Transcriptional regulation mediated by six-finger proteins in living cells. (*A*) HeLa cells were transiently transfected in triplicate with 2.5 μg of the indicated reporter plasmids and 2.5 μg C7-C7-VP16 expression plasmid. Luciferase activities were measured 48 hr later and normalized to the control β-galactosidase activity. The relative light units are given on top of each column (error bar = SD). (*B*) HeLa cells were transfected with 2.5 μg of the indicated reporter plasmids and either no C7-C7-KRAB expression or 1 μg of the C7-C7-KRAB expression plasmid by using LipofectAmine (GIBCO/BRL) as the transfection reagent. Luciferase activities were measured 48 hr later, and normalized to the control β-galactosidase activity. The relative light unit values were labeled on top of each column (error bar = SD).

We acknowledge Peter Wright and Joel Gottesfeld for insightful discussions. This study was supported by National Institutes of Health Grant GM53910 and The Skaggs Institute for Chemical Biology to C.F.B., and by funds provided by the Breast Cancer Fund of the State of California through the Breast Cancer Research Program of the University of California, Grant 2FB-0138 to Q.L. C.F.B. acknowledges an investigator award from The Cancer Research Institute.

1. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4,** 1609–1614.
2. Berg, J. M. & Shi, Y. (1996) *Science* **271,** 1081–1085.
3. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252,** 809–817.
4. Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366,** 483–487.
5. Pavletich, N. P. & Pabo, C. O. (1993) *Science* **261,** 1701–1707.
6. Clemens, K. R., Liao, X., Wolf, V., Wright, P. E. & Gottesfeld, J. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10822–10826.
7. Hayes, J. J. & Tullius, T. D. (1992) *J. Mol. Biol.* **227,** 407–417.
8. Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. (1996) *Structure (London)* **4,** 1171–1180.
9. Choo, Y., Sanchez-Garcia, I. & Klug, A. (1994) *Nature (London)* **372,** 642–645.
10. Choo, Y. & Klug, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11163–11167.
11. Jamieson, A. C., Kim, S. H. & Wells, J. A. (1994) *Biochemistry* **33,** 5689–5695.
12. Pomerantz, J. L., Sharp, P. A. & Pabo, C. O. (1995) *Science* **267,** 93–96.
13. Rebar, E. J. & Pabo, C. O. (1994) *Science* **263,** 671–673.
14. Shi, Y. & Berg, J. M. (1995) *Science* **268,** 282–284.
15. Wu, H., Yang, W. P. & Barbas, C. F. I. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 344–348.
16. Greisman, H. A. & Pabo, C. O. (1997) *Science* **275,** 657–661.
17. Shi, Y. & Berg, J. M. (1995) *Chem. Biol.* **2,** 83–89.
18. Brasier, A. R., Tate, J. E. & Habener, J. F. (1989) *BioTechniques* **7,** 1116–1122.
19. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. (1984) *Cell* **39,** 499–509.
20. Margolin, J. F., Friedman, J. R., Meyer, W. K.-H., Vissing, H., Thiesen, H.-J. & Rauscher, F. J. I. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 4509–4513.
21. Witzgall, R., O'Leary, E., Leaf, A., Onaldi, D. & Bonventre, J. V. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 4514–4518.
22. Pengue, G., Calabro, V., Bartoli, P. C., Pagliuca, A. & Lania, L. (1994) *Nucleic Acids Res.* **22,** 2908–2914.
23. Kim, S.-S., Chen, Y.-M., O'Leary, E., Witzgall, R., Vidal, M. & Bonventre, J. V. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 15299–15304.
24. Stein, C. A. & Cheng, Y.-C. (1993) *Science* **261,** 1004–1012.
25. Yu, M., Ojwang, J., Yamada, O., Hampel, A., Rapapport, J., Looney, D. & Wong-Staal, F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6340–6344.
26. Yen, P. M. & Chin, W. W. (1994) *Mol. Endocrinol.* **8,** 1450–1454.
27. Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996) *Nature (London)* **382,** 559–561.
28. Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R. & O'Brien, S. J. (1996) *Science* **273,** 1856–1862.