Biochemistry

 Volume 37, Number 4 January 27, 1998

Accelerated Publications

Structure-Based Design of a Dimeric Zinc Finger Protein†

Joel L. Pomerantz, Scot A. Wolfe, and Carl O. Pabo*

Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received October 6, 1997; Revised Manuscript Received November 10, 1997⊗

ABSTRACT: Designing DNA-binding proteins with novel sequence specificities may provide valuable tools for biological research and gene therapy. Computer modeling was used to design a dimeric zinc finger protein, ZFGD1, containing zinc fingers 1 and 2 from Zif268 and a portion of the dimerization domain of GAL4. ZFGD1 binds with high affinity and specificity to the predicted binding site, which contains two 6 base-pair symmetry-related zinc finger subsites separated by a 13 base-pair spacer. The DNA-binding specificity of this fusion protein is determined primarily by the zinc fingers and can be systematically altered through the substitution of the zinc fingers with variants selected by phage display. This zinc finger−GAL4 fusion may serve as a prototype for designed DNA-binding proteins that could exploit advantages of homo- and heterodimer formation, and the adaptability of the Cys_2His_2 zinc finger motif, to target virtually any site in the genome.

DNA-binding proteins play critical roles in the regulation of gene expression and in other key processes such as DNA replication and repair. Structural and biochemical studies have revealed many of the fundamental principles of protein−DNA recognition (*1*), and this information has provided a basis for the design and selection of novel DNA-binding proteins. Design methods that use known structural motifs and that exploit principles observed in natural biological regulatory systems can be used to create specific, high-affinity DNA-binding proteins for applications in biological research and gene therapy.

One strategy for generating proteins with desired specificities has involved the mutation or randomization of a DNA-binding module and the selection of sequence variants with the most favorable contact residues at the protein−DNA interface. Cys_2His_2 zinc finger proteins have been used for many of these studies (*2*−*6*) because the fold and DNA-docking arrangement of these fingers seems to make them especially versatile and adaptable (*7*−*11*). Using phage display techniques, novel three-finger proteins have been selected which bind to 9 base-pair target sites with dissociation constants of ∼100 pM (*6*). Design strategies based on the prediction of preferred contacts also have been applied to the construction of novel zinc finger proteins (*12*).

Structure-based methods for linking DNA-binding modules provide another powerful design strategy (*13*). In this approach, computer modeling is used to superimpose protein−DNA complexes in various registers, revealing arrangements that might allow heterologous modules to be fused with short

† Supported by NIH Grant GM31471. Graphics were prepared using equipment purchased with support from the PEW Charitable Trusts and the Howard Hughes Medical Institute.

* Author to whom correspondence should be addressed. 68-580, Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. Fax: (617) 253-8728.

⊗ Abstract published in *Advance ACS Abstracts,* December 15, 1997.

S0006-2960(97)02464-1 CCC: $15.00
Published on Web 01/13/1998

peptide linkers. The resulting chimeric DNA-binding proteins display sequence specificity for extended, chimeric DNA-binding sites. Using this strategy, zinc fingers have been fused to a homeodomain (*13*, *14*), to the TATA-binding protein (*15*), and to other zinc fingers (*16*). These chimeric proteins can regulate gene expression *in vivo*. The unique specificity of the zinc finger−homeodomain fusion protein already has been exploited in a model gene therapy system in mice (*17*).

Dimer formation, frequently employed by natural DNA-binding proteins to enhance the affinity and specificity of recognition, provides another attractive design strategy. The capacity for homo- and heterodimerization offers several potential advantages for DNA-binding proteins. The cooperativity that results from dimerization narrows the concentration range over which DNA binding occurs, resulting in a sharper transition between "on" and "off" regulatory states. Binding as a dimer also allows recognition of a longer DNA sequence (consisting of appropriately spaced half sites) and thus should enhance the specificity of target site recognition. In addition, heterodimer formation should allow recognition of asymmetric sites and open possibilities for designing networks of interacting proteins. Design strategies that employ dimerization also may provide a useful alternative to the covalent linkage of multiple DNA-binding domains. Large covalent assemblies might have higher absolute affinity for nonspecific DNA sites and might become kinetically trapped at inappropriate sites in the genome. Dimerization provides an alternative way of bringing multiple domains together as a functional recognition unit. Dimerization domains with modest protein−protein interaction energies may ensure that association could only occur cooperatively upon binding the specific DNA site.

To begin exploring the potential advantages of dimerization in the design of novel DNA-binding proteins, we sought to construct a zinc finger protein that can form homodimers and heterodimers. Our approach involved the structure-based fusion of zinc fingers from Zif268 to the dimerization domain of GAL4. The GAL4 domain was chosen because structural information is available for this domain (*18*) and because it contains a coiled-coil motif, a simple, well-understood structure that can be further modified for design purposes. The GAL4 dimerization motif is also interesting because it docks to DNA and presumably would help to position and orient the fused zinc finger domains. Moreover, the dimerization motif does not appear to require specific sequences for binding.

In this paper, we report the design and characterization of the zinc finger−GAL4 fusion. We find that our designed protein binds the predicted DNA sequence as a dimer and that its sequence specificity is primarily determined by the zinc fingers. We also show that novel zinc fingers, selected by phage display, can readily be incorporated into this design. In principle, this strategy allows almost any sequence to be targeted through heterodimeric recognition of asymmetric binding sites.

MATERIALS AND METHODS

Construction and Expression of Fusion Proteins. A DNA fragment encoding residues 2 to 59 of Zif268 [numbering scheme of Pavletich and Pabo (*7*)], a glycine residue, and residues 41 to 100 of GAL4 (*18*) was generated by PCR and cloned into pGEX-2T (Pharmacia) to generate an in-frame fusion of ZFGD1 (zinc finger GAL4 dimerization construct 1) to glutathione-*S*-transferase (GST). The ZFGD2 construct was prepared by replacing the Zif268 coding sequences with those encoding fingers 2 and 3 of a peptide selected for binding to the nuclear hormone response element (*6*). The GST[1] fusion proteins were expressed and purified by standard methods (*13*) and then cleaved with thrombin, which was subsequently inactivated with PMSF. Proteins were stored in 50 mM Tris (pH 8.0), 100 mM KCl, and 10% glycerol.

Electrophoretic Mobility Shift Assays. DNA-binding reactions contained 20 mM Hepes (pH 7.9), 60 mM KCl, 0.75 mM dithiothreitol, 4% Ficoll-400, and BSA (300 μg/mL), with the appropriate protein(s) and binding site in a total volume of 10 μL. The binding studies shown in Figure 3 also contained 25 μg/mL sonicated salmon sperm DNA. Binding reactions were incubated at 30 °C for 40 min and then resolved in 4% nondenaturing polyacrylamide gels using Tris-glycine electrophoresis buffer (*13*). The amount of free DNA and of protein−DNA complexes were quantitated using a Molecular Dynamics Phosphorimager with ImageQuant 3.3 software. Duplex DNA sites used in these studies were derived by cloning the following sequences into the *Eco*RI and *Hin*dIII sites of pBSKII+ (Stratagene): 5′-GAATTCATGCCGCCCAGAGGACAGTCCTC-TGGGCGATGCAAGCTT-3′, 5′-GAATTCATGCCGCCCA-GAGGACAGTCCTCAAGGGTATGCAAGCTT-3′, 5′-GAATTCATGCACCCTTGAGGACAGTCCT-CAAGGGTATGCAAGCTT-3′, 5′-GAATTCCGATCGC-CCACGTATGCTAATGATGGGCGTACGAAGCTT-3′.

DNA fragments were excised with *Eco*RI and *Hin*dIII, and labeled by Klenow in the presence of [α-^{32}P]dATP.

Determination of Equilibrium Constants. Equilibrium dissociation constants (K_{eq}) were determined by linear regression using the Scatchard equation:

$$\frac{\theta}{[P]^2} = \frac{1}{K_{eq}} - \frac{\theta}{K_{eq}}$$

where θ represents the fraction of DNA bound and [P] represents the free protein concentration. In analyzing the data with this equation, we assumed that the free protein concentration would be essentially equivalent to the total protein concentration. This seemed reasonable since the total DNA concentration was always at least an order of magnitude lower than the protein concentration, and we also assumed that the free proteins would not dimerize significantly at the concentrations used in these experiments. The binding data fit well with these assumptions. Total active protein concentrations were determined by titrating DNA-binding sites, in the 1−50 μM concentration range, into reactions containing a constant amount of ZFGD1 or ZFGD2. The values listed in Table 1 are the averages of at least three determinations ± standard deviation.

Determination of Dissociation Rates. Binding reactions were prepared with saturating amounts of protein and were

[1] Abbreviations: GST, glutathione-*S*-transferase; BSA, bovine serum albumin; PMSF, phenylmethanesulfonyl flouride.

Table 1: DNA-Binding Affinity and Specificity of ZFGD1

binding site	K_{eq} ($\times 10^{-19}$ M^2)[a]	half-maximal binding (nM)[b]	k_{off} ($\times 10^{-5}$ s^{-1})[a]	$t_{1/2}$ (min)
5′-CGCCCAGAGGACAGTCCTCTGGGCG-3′	7.8 ± 0.4	0.9	7.4 ± 0.1	157
5′-CGCCCA**CGTATGCTAATGA**TGGGCG-3′	10 ± 6	1.0	12.2 ± 0.8	95
5′-CGCCCAGAGGACAGTCCTCAAGGGT-3′	nd	14.6[c]	nd[d]	nd
5′-ACCCTTGAGGACAGTCCTCAAGGGT-3′	nd	22.0[c]	nd	nd

[a] The average of at least three determinations ± standard deviation. [b] Given in concentration of ZFGD1 monomer. [c] The concentration at which 50% of the free DNA was shifted. The DNA−protein complex was not stable to electrophoresis when one or both of the zinc finger subsites was altered. Since we do not know the stoichiometry of nonspecific binding or the number of nonspecific sites on the DNA probe, we cannot reliably estimate K_{eq} for binding to these probes. [d] nd: not determined.

equilibrated as described above for 30 min. Following the addition of a vast excess of cold binding site (relative to the total protein concentration), and, after continuing incubation at 30 °C, portions of the binding reaction were loaded onto a 4% polyacrylamide gel at various time points up to 4 h. The fraction of DNA bound was quantitated and the dissociation rate constant k_d was determined by linear regression using the equation $\ln(\theta/\theta_0) = -k_d t$. The values listed in Table 1 are the averages of at least three determinations ± standard deviation.

RESULTS

Structure-based design methods were used to explore how a dimerization domain might be fused to the Zif268 zinc fingers. The crystal structures of the Zif268−DNA (*7*) and GAL4−DNA (*18*) complexes were superimposed by aligning phosphates in various registers. One especially favorable alignment suggested that the GAL4 dimerization domain might be fused onto the C-terminal end of a set of one or more fingers (essentially replacing the GAL4 DNA-binding domain with a set of Cys_2His_2 zinc fingers). In this model, the Cα of glycine 59 from Zif268 was only 2.8 Å away from the Cα of serine 41 from GAL4 (Figure 1). These two residues also seemed to provide "natural" linkage sites because glycine 59 occurs at the C-terminus of finger 2 in Zif268 and because serine 41 is the first amino acid in the GAL4 linker (which connects the DNA-binding domain of GAL4 to the coiled-coil dimerization element). The chimeric protein in this model is predicted to bind to a 25 base-pair site with the sequence 5′-CGCCCAGAGGACAGTC-CTCTGGGCG-3′. The 6 base-pair zinc finger subsites (underlined) are present on each end of the extended site. The central 13 base pairs are derived from a portion of the GAL4-binding site that lies under the coiled-coil dimerization motif and associated polypeptide linker. This region of the GAL4 site does not appear to mediate sequence-specific protein−DNA contacts.

To test this model, we constructed the fusion protein ZFGD1, which contains fingers 1 and 2 of Zif268, a glycine residue, and residues 41−100 from GAL4 (Figure 1). Structural information is not available for residues 66−100 of GAL4, but these residues were included because they are known to form part of the GAL4 dimerization domain (*18*). ZFGD1 was tested for binding to the predicted site in electrophoretic mobility shift experiments. ZFGD1 bound to the predicted site with high affinity, with a K_{eq} of 7.8 × 10^{-19} M^2 corresponding to half-maximal binding at a monomer concentration of 0.9 nM (Figure 2A, lanes 1−5, Figure 2B, and Table 1). When tested with a probe containing substitutions in one zinc finger subsite, ZFGD1 did not form a complex stable to electrophoresis (Figure 2A lanes 6−10), but binding (at higher concentrations) could be monitored by the disappearance of the free DNA. We found that 50% of this altered DNA site was shifted at a ZFGD1 monomer concentration of 14.6 nM (data not shown, Table 1). ZFGD1 behaved similarly with a probe containing substitutions in both zinc finger subsites (Figure 2A, lanes 11−15) and shifted 50% of this altered DNA site at a ZFGD1 monomer concentration of 22 nM (data not shown, Table 1). This set of binding studies clearly validates the modeling and indicates that ZFGD1 binds as a dimer to the predicted site.

The utility of this zinc finger−GAL4 fusion rests on the expectation that specificity will be determined primarily by the zinc fingers, which can be designed or selected to recognize desired target sites. Although the GAL4 linker and dimerization element contact DNA, they do not make any base-specific contacts in the crystal structure (*18*). Binding studies with ZFGD1 confirmed that the central 13 base-pair region of this site, where the GAL4 linker and dimerization elements are expected to contact the DNA (Figure 1), makes little contribution to sequence-specific recognition. In these experiments, ZFGD1 was titrated with a probe that contained the appropriate zinc-finger-binding sites but which had the central 13 base-pair region replaced by a sequence unrelated to the natural GAL4-binding site. ZFGD1 bound to this altered probe with an affinity similar to that observed for the original, modeled site (Figure 2A, lanes 16−20): the equilibrium dissociation constant (K_{eq}) was 10 × 10^{-19} M^2, which corresponds to half-maximal binding at a monomer concentration of 1.0 nM (Figure 2B, Table 1). Dissociation rates were also compared at the original and substituted sites and were found to be quite similar, 7.4 × 10^{-5} s^{-1} and 12.2 × 10^{-5} s^{-1}, respectively (Figure 2C, Table 1). Since such extensive sequence changes can be tolerated in the central 13 base-pair region, while changes in the flanking regions significantly reduce affinity, we conclude that the binding specificity of the protein is primarily determined by the zinc fingers.

To test the adaptability of this design, we also constructed ZFGD2, a variant containing novel zinc fingers obtained through phage display. The fingers in ZFGD2 had been selected in a previous study (*6*) for binding to the subsite 5′-AAGGGT-3′. Having variant zinc fingers fused to the same dimerization region also allowed us to test whether a heterodimer would bind specifically. A mixture of the ZFGD1 and ZFGD2 proteins was tested for the ability to recognize a hybrid site containing appropriately spaced 6

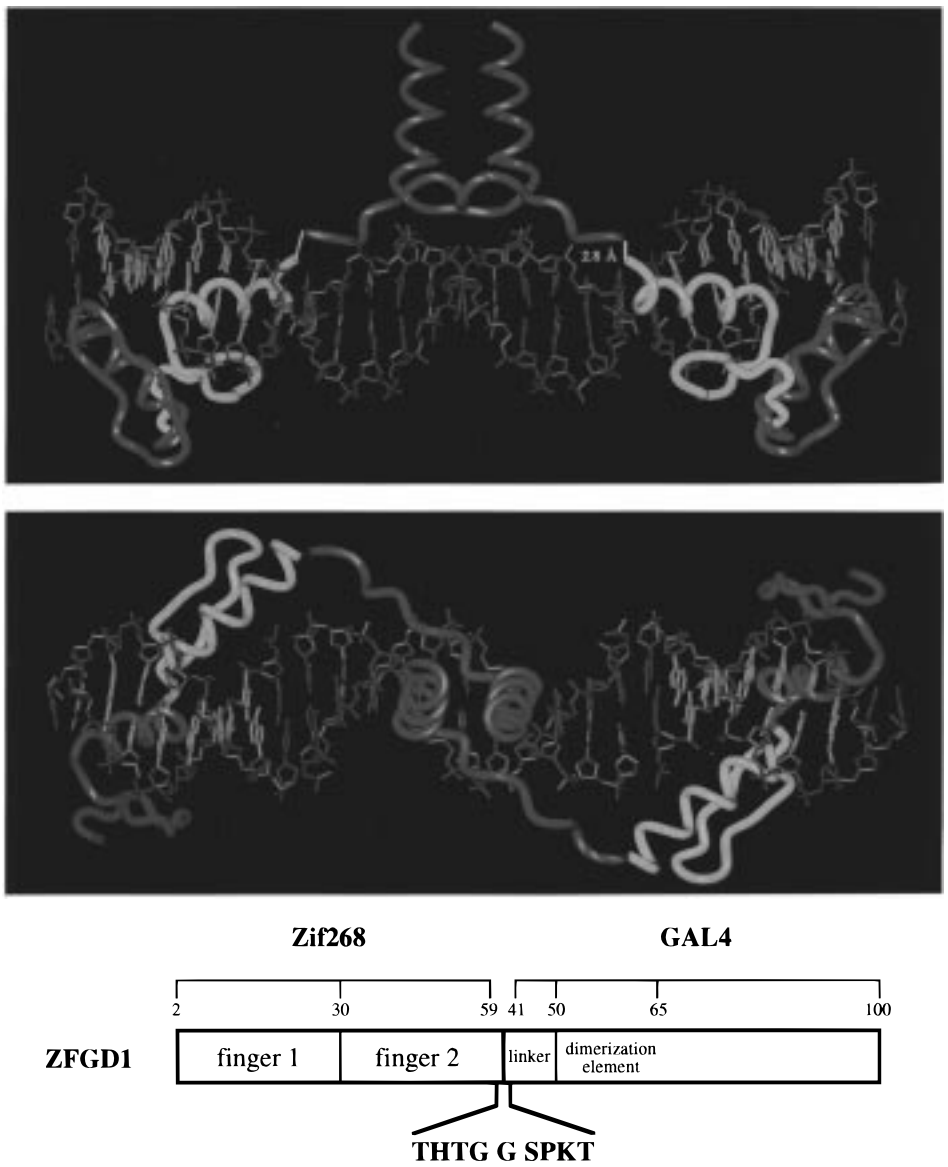

FIGURE 1: Model of a Zif268−GAL4 chimeric DNA binding/dimerization domain. (Top panel) Finger 1 of Zif268 is shown in red, finger 2 in yellow, and the GAL4 dimerization element in purple. The DNA is blue with the base pairs in the zinc finger subsites (TGGGCG) highlighted in cyan. The hybrid protein recognizes a sequence of the form 5′-CGCCCAN$_{13}$TGGGCG-3′. The C_α of glycine 59 from Zif268 (the C-terminus of finger 2) is 2.8 Å from the C_α of serine 41 from GAL4 (the first residue in the linker). (Middle panel) A view with the model rotated 90 degrees. This figure was generated with Insight II (Biosym Technologies, San Diego, California). (Bottom panel) Schematic indicating the domain structure and linker region of ZFGD1.

base-pair subsites for the corresponding zinc fingers. These experiments were performed in the presence of sonicated salmon sperm DNA to increase the stringency and to simulate binding in the presence of genomic DNA. ZFGD1 alone showed no binding to the hybrid heterodimeric site (Figure 3, lanes 5−8) at concentrations at which there was efficient binding to the ZFGD1 homodimer site (lanes 1−4). The addition of ZFGD2 allowed recognition of the heterodimeric site (lanes 9−12). At equivalent monomer concentrations, ZFGD2 alone effectively recognized its cognate homodimer site (lanes 17−20), but did not bind efficiently to the hybrid site (lanes 13−16). These data demonstrate that the zinc finger−GAL4 design (Figure 1) can accommodate substitution of the zinc fingers to change sequence specificity. As a result, heterodimers can be used to target asymmetric sites containing two appropriately spaced 6 base-pair subsites for the zinc fingers.

DISCUSSION

ZFGD1 may serve as a useful prototype for the continued development of zinc finger dimerization strategies. This current design fuses the remarkable versatility and adaptability of the Cys_2His_2 zinc fingers to a simple coiled-coil-based dimerization motif that allows for homo- or heterodimerization. ZFGD1 binds with high affinity and specificity to the site predicted from the initial modeling and design studies (Figure 1). The sequence specificity of the fusion protein is determined primarily by the zinc fingers, allowing substitution of fingers to target new sites and to generate heterodimers capable of binding asymmetric sequences.

It should be feasible to target the zinc finger−GAL4 fusion to a wide variety of sites by selecting or designing zinc fingers that will recognize the appropriate 6 base-pair subsites. Sequential selection methods for phage display (*6*)

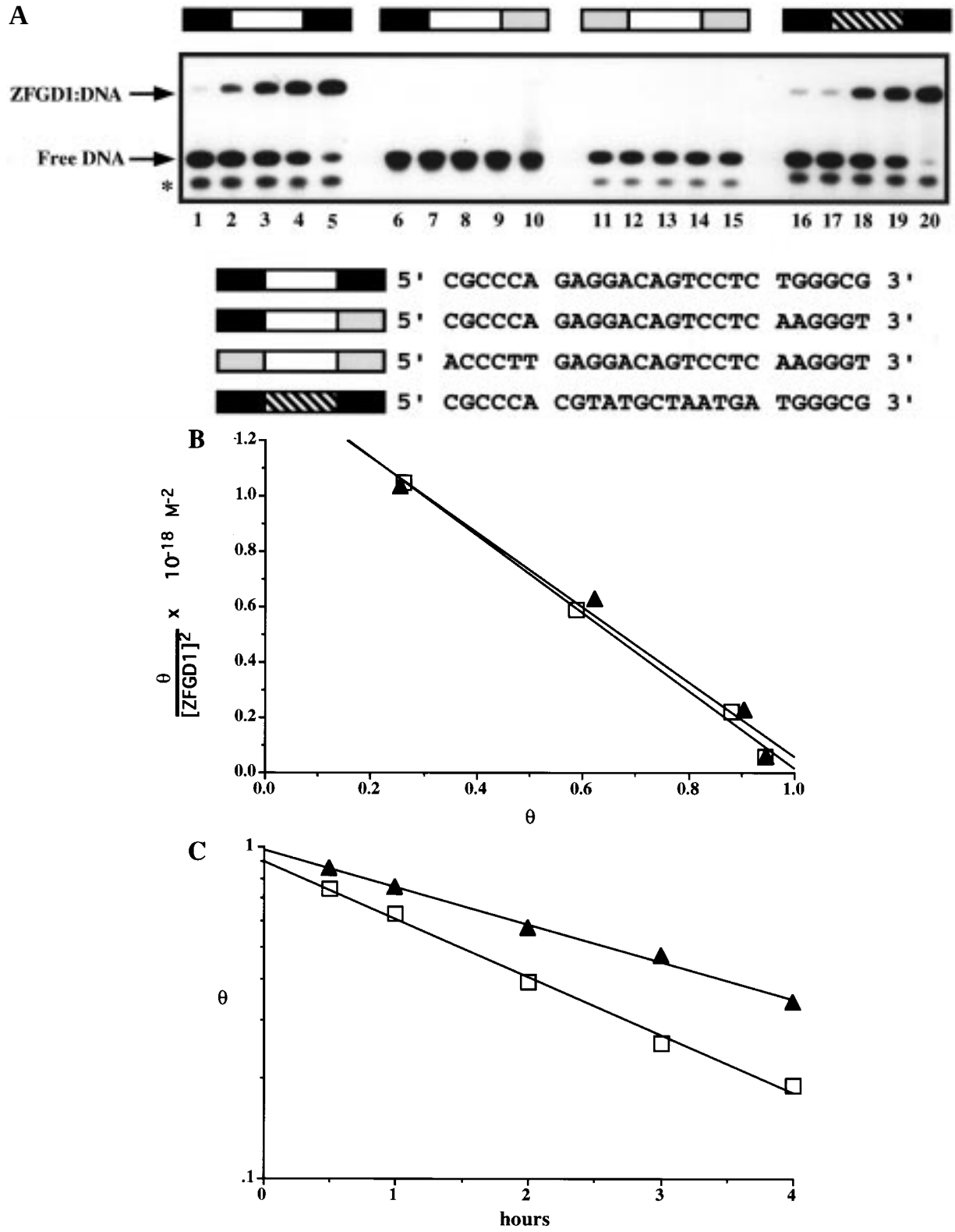

FIGURE 2: ZFGD1 binds as a dimer to the modeled DNA site with specificity determined primarily by the zinc fingers. (A) ZFGD1 was titrated into DNA-binding reactions containing the DNA probe indicated at the top of each set of lanes. Lanes 1, 6, 11, and 16 contain the protein at 2.5×10^{-10} M; the protein concentration increases in 2-fold increments in subsequent lanes of each set. The positions of the protein−DNA complex and free DNA are indicated. The asterisk indicates single-stranded probe DNA. Sequences of the DNA sites are listed below. Solid black bars indicate the 6 base-pair-binding subsite for fingers 1 and 2 of Zif268; gray bars indicate variant subsites; solid white bars indicate the central 13 base-pairs derived from the GAL4 binding site; the hatched bar indicates the substituted intervening 13 base pairs. (B) Representative scatchard plots for the binding of ZFGD1 to the probes 5′- CGCCCAGAGGACAGTCCTCTGGGCG-3′ (original modeled site; solid triangles) and 5′- CGCCCACGTATGCTAATGATGGGCG-3 (site with central 13 base pairs substituted; open squares). θ refers to the fraction of DNA bound by ZFGD1. The data is fully consistent with the underlying equation (Materials and Methods) for binding as a dimer. (C) Representative dissociation plots indicating the stability of the ZFGD1-DNA complex on the probes 5′- CGCCCAGAGGACAGTCCTCTGGGCG-3′ (solid triangles) and 5′- CGCCCACGTATGCTAATGATGGGCG-3 (open squares). The *x*-axis indicates the time elapsed after the addition of excess cold binding site (see Materials and Methods).

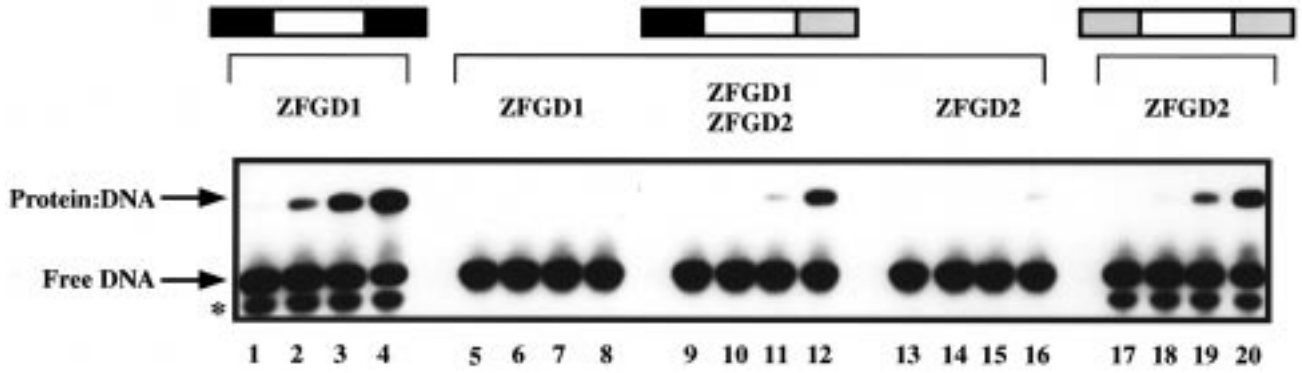

FIGURE 3: Heterodimeric binding by zinc finger-GAL4 fusion proteins. ZFGD1 and/or ZFGD2 were titrated into reactions containing the indicated probes (sequences indicated in Figure 2). As before, black bars indicate the 6 base-pair subsite for fingers 1 and 2 of Zif268, gray bars indicate the 6 base-pair subsite for fingers 2 and 3 of a peptide selected by phage display (6). Lanes 1, 5, and 9 contain ZFGD1 at 1.5×10^{-8} M. Lanes 9, 13, and 17 contain ZFGD2 at 1.5×10^{-8} M. These binding reactions were performed in the presence of sonicated salmon sperm DNA at 25 μg/mL, and the protein concentration was increased in 3-fold increments in subsequent lanes of each set. The positions of the protein−DNA complex and free DNA are indicated. The asterisk indicates single-stranded probe DNA.

should facilitate the optimization of the required two finger units and should account for any context-dependent interactions between adjacent fingers and subsites.

The dimerization interface also provides opportunities for further elaboration and optimization. As demonstrated by many studies, the coiled-coil interaction motif offers the potential to modify the dimerization domain to increase dimerization affinity or to specifically promote heterodimer formation (see refs *19* and *20* for examples). Obligate heterodimer variants of the zinc finger−GAL4 fusion might be constructed to ensure that the proteins could only bind the heterodimer site. It also seems plausible that adjustments in the linker region may give further improvements in the affinity and specificity of the zinc finger−GAL4 proteins.

As zinc finger dimer designs are perfected, it will be interesting to compare the efficacy of these designs with that of covalently linked sets of zinc fingers (*16*). Dimer contacts of modest affinity may allow self-assembly at the appropriate binding site and thereby reduce the risks of nonspecific (kinetic) trapping that may occur with large covalently linked sets of fingers. Cooperative binding, by giving a more dramatic concentration dependence, may also allow more precise on/off switching in targeted gene regulation. Finally, the capacity to homo- and heterodimerize should allow these designed proteins to mimic the combinatorial control that plays such a prominent role in known gene-regulatory systems. The zinc finger−GAL4 fusion protein provides an excellent starting point for examining these potential advantages in the design of site-specific DNA-binding proteins for biological research and gene therapy.

ACKNOWLEDGMENT

We thank E. Fraenkel for helpful discussions and K. Collins and I. Stancovski for critical reading of the manuscript.

REFERENCES

1. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
2. Rebar, E., and Pabo, C. O. (1994) *Science 263*, 671−673.
3. Jamieson, A. C., Kim, S.-H., and Wells, J. A. (1994) *Biochemistry 33*, 5689−5695.
4. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11163−11167.
5. Wu, H., Yang, W.-P., and Barbas, C. F. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 344−348.
6. Greisman, H. A., and Pabo, C. O. (1997) *Science 275*, 657−661.
7. Pavletich, N. P., and Pabo, C. O. (1991) *Science 252*, 809−817.
8. Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Nature 366*, 483−487.
9. Pavletich, N. P., and Pabo, C. O. (1993) *Science 261*, 1701−1707.
10. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure 4*, 1171−1180.
11. Houbaviy, H. B., Usheva, A., Shenk, T., and Burley, S. K. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13577−13582.
12. Kim, C. A., and Berg, J. M. (1996) *Nat. Struct. Biol. 3*, 940−945.
13. Pomerantz, J. L., Sharp, P. A., and Pabo, C. O. (1995) *Science 267*, 93−96.
14. Pomerantz, J. L., Pabo, C. O., and Sharp, P. A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9752−9756.
15. Kim J. S., Kim J., Cepek K. L., Sharp P. A., and Pabo C. O. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3616−3620.
16. Liu, Q., Segal, D. J., Ghiara, J. B., and Barbas, C. F. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5525−5530.
17. Rivera V. M., Clackson T., Natesan S., Pollock R., Amara J. F., Keenan T., Magari S. R., Phillips T., Courage N. L., Cerasoli F. Jr, Holt D. A., and Gilman M. (1996) *Nat. Med. 2*, 1028−1032.
18. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C. (1992) *Nature 356*, 408−414.
19. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) *Curr. Biol. 3*, 658−667.
20. Krylov, D., Mikhailenko, I., and Vinson, C. (1994) *EMBO J. 13*, 2849−2861.

BI972464O