Rational Design of Artificial Zinc-Finger Proteins Using a Nondegenerate Recognition Code Table

Takashi Sera* and Carla Uranga

Torrey Mesa Research Institute, 3115 Merryfield Row, Suite 100, San Diego, California 92121

Received February 1, 2002

ABSTRACT: We have developed a novel and simple method to rationally design artificial zinc-finger proteins (AZPs) targeting diverse DNA sequences using a nondegenerate recognition code table. The table was constructed based on known and potential DNA base−amino acid interactions. The table permits identification of an amino acid for each position (−1, 2, 3, and 6) of the α-helical region of the zinc-finger domain (position 1 is the starting amino acid in the α-helix) from overlapping 4-bp sequences in a given DNA target. Based on the table, we designed ten 3-finger AZPs, each of which targeted an arbitrarily chosen 10-bp DNA sequence, and characterized the binding properties. In vitro DNA-binding assays showed five of the AZPs tightly and specifically bound to their targets containing more than three guanine bases in the first 9-bp region. In addition, 6-finger AZPs, each of which was produced by combining two functional 3-finger AZPs, bound to their 19-bp targets with the dissociation constant of less than 3 pM. The in vivo functionality of the AZP was tested using *Arabidopsis* protoplasts. The AZP fused to a transcriptional activation domain efficiently activated expression of a reporter gene linked to a native promoter containing the AZP target site. Our simple AZP design method will provide a powerful approach to manipulation of endogenous gene expression by enabling rapid creation of numerous artificial DNA-binding proteins.

Zinc-finger DNA-binding proteins have proven to be useful for production of new DNA-binding proteins as novel molecular biological tools (*1*). A variety approaches have been used to alter sequence selectivities of native zinc-finger proteins or create the new selectivities. These approaches include the bioinformatic approach (*2*), site-directed mutagenesis (*3*, *4*), construction of chimeric DNA-binding proteins with a different DNA-binding motif (*5*), assembly of finger proteins through a flexible linker peptide (*6*−*8*) or a dimerizing peptide (*9*, *10*), and the generation and screening of large libraries of mutant zinc-finger proteins by phage display (*11*−*14*). Phage display systems (*11*−*13*) have been used successfully to generate many Zif268 mutants with altered specificities. Zinc-finger domains targeting all possible 16 triples of 5′-GNN-3′ were successfully selected by phage display (*15*). Production of zinc-finger proteins by combination of finger domains selected by phage display was reported (*16*−*19*). Sequential phage selection of all three fingers produced 3-finger proteins with high affinities (*14*). It has been demonstrated that some designed zinc-finger proteins could manipulate endogenous gene expression in cultured cells (*16*, *20*−*22*). However, DNA sequences targeted efficiently using these current approaches are mainly guanine-rich sequences, such as 5′-GNNGNN...-3′, and it appears to be difficult to achieve the high sequence-specificities even by the powerful phage display method (*23*−*25*). To our knowledge, a rational design scheme of zinc-finger proteins for targeting diverse DNA sequences has not been established yet.

We focus on the development of a rapid creation scheme of numerous zinc-finger proteins with satisfactory binding properties rather than production of a limited number of highly optimized zinc-finger proteins using selection approaches. The reasons are the following: (i) ever increasing, already large genome information on various organisms presents us opportunities to manipulate numerous genes, (ii) it appears to be difficult to generate zinc-finger proteins with high sequence-specificities as shown in extensive characterization of designed zinc-finger proteins (*23*−*25*), (iii) the high specificities may not be practically necessary for gene regulation as shown by designed zinc-finger proteins (*16*, *20*−*22*), (iv) it has not been established yet which genomic site in a given promoter should be targeted for the most efficient gene regulation, and, therefore, (v) the ability to rapidly create zinc-finger proteins for multiple sites in a given promoter will enable us to take a shot-gun approach to find the best targeting sites for efficient gene regulation.

Berg's group reported one unique zinc-finger protein composed of identical finger frameworks in the 1st and 2nd fingers and a slightly different framework in the 3rd finger (*26*). The X-ray crystal structural analysis of the DNA complex revealed the following important features: (i) each zinc-finger domain recognizes an overlapping 4-bp DNA sequence, where the last base of each 4-bp target is the first base of the next 4-bp target; (ii) in all three fingers of the protein, amino acids at specific positions contact DNA bases at specific positions in a regular fashion. Namely, amino acids at positions −1, 2, 3, and 6 contact the 3rd, 4th, 2nd,

* Author to whom correspondence should be addressed at Torrey Mesa Research Institute, 3115 Merryfield Row, Suite 100, San Diego, CA 92121. Phone: (858) 812-1088. Fax: (858) 812-1105. E-mail: takashi.sera@syngenta.com.

10.1021/bi020095c CCC: $22.00
Published on Web 05/09/2002

and 1st bases of the overlapping 4-bp DNA targets, respectively (only the 4th base in the antisense strand). These features hinted to us that if we can identify DNA base specificities of amino acids at each of the four critical positions, it should be possible to design zinc-finger proteins targeting diverse DNA sequences simply by a combination of four critical amino acids per finger domain and the assembly of finger domains.

Here, we report a new design scheme of artificial zinc-finger proteins (AZPs)[1] using a nondegenerate recognition code table. We have designed the nondegenerate recognition code table based on known amino acid−DNA base interactions, observed in X-ray crystal structures of DNA complexes with various DNA-binding proteins, and potential amino acid−DNA base interactions. The table allows identification of one specific amino acid at each of positions −1, 2, 3, and 6 of the α-helical region of the zinc-finger domain from the overlapping 4-bp target sequence in a given DNA target, thereby rapidly producing numerous AZPs by PCR assembly of the corresponding finger domains. Using the table, we have designed ten 3-finger AZPs targeting arbitrarily chosen 10-bp DNA sequences, and characterized the in vitro binding properties. Then, we have generated several multi-finger AZPs using the 3-finger AZPs and characterized the binding properties. We also have tested the in vivo functionality of the multi-finger AZP using *Arabidopsis* protoplasts.

MATERIALS AND METHODS

Design and Construction of AZP-Coding DNA. The design and construction of AZP-coding DNA will be described elsewhere; however, our strategy for AZP DNA synthesis is outlined as follows. First a 10-bp DNA target was divided into three 4-bp DNA segments with a single bp overlap between two 4-bp segments, and four amino acids per finger were chosen from our nondegenerate recognition code table based on the target sequence. DNA sequences encoding each finger were carefully designed to assemble three finger domains in the correct order by PCR. After a Klenow fill-in reaction of synthetic oligonucleotides, the resulting duplex DNA fragments were assembled by PCR to produce an entire AZP-coding DNA.

AZP Expression and Purification. The coding regions of AZPs were cloned into the *Eco*RI and *Hin*dIII sites of pET-21a (Novagen). The resulting plasmids were then introduced into *E. coli* BL21(DE3)pLysS for protein overexpression. Protein expression/purification was performed essentially as described (*27*). All purified proteins were >95% homogeneous, as judged by SDS/PAGE. Protein concentration was determined using Protein Assay ESL (Roche Molecular Biochemicals).

DNA-Binding Assays. The 26-bp oligonucleotides (see legend, Table 1), labeled at the 5′-end with [γ-^{32}P]ATP, were used in gel shift assays. Highly radioactive probes for "multi-finger" AZPs (designated as AZPs containing more than four fingers) were labeled by a Klenow fill-in reaction with [α-^{32}P]dATP and [α-^{32}P]dTTP. Purified AZPs were incubated on ice for 1 h in 10 mM Tris-HCl, pH 7.5/100 mM NaCl/5 mM $MgCl_2$/0.1 mM $ZnCl_2$/0.05% BSA/10% glycerol containing an end-labeled probe (1 fmol per 10 μL of buffer) and 1 μg of poly(dA-dT)$_2$ before loading onto a 6% nondenaturing polyacrylamide gel (45 mM Tris−borate) and electrophoresing at 140 V for 2 h at 4 °C. For multi-finger AZPs, 0.03 fmol of radiolabeled probes (the lowest amount used in our assays) was used. The radioactive signals were quantitated with PhosphorImager (Molecular Dynamics). The dissociation constants (K_D) were determined by curve-fitting the data to the equation: $F = [P]/([P] + K_D)$, where F and [P] represent the fraction of the DNA probe bound by the protein and the total protein concentration, respectively, using the KALEIDAGRAPH (Synergy Software).

Protoplast Transient Expression Assays. The NIM1-luciferase (LUC) reporter plasmid was constructed by cloning an *Arabidopsis thaliana* NIM1 promoter fragment (*28*) containing oligonucleotides −766 to −1, relative to the ATG start codon, in front of the LUC coding sequence. An expression plasmid of an artificial transcription factor (ATF) encodes a nuclear localization signal from the SV40 large T antigen, an AZP for the NIM1 promoter targeting, a herpes simplex virus VP-16 activation domain comprising amino acids 415−490, and a FLAG epitope tag (in this order from the amino terminus) under a cestrum yellow leaf curling virus promoter. *Arabidopsis* protoplast preparation and PEG transformation with the ATF expression plasmid (or a control plasmid without an AZP-coding DNA, 10 μg), the NIM1-LUC reporter plasmid (2 μg), and a GUS reference plasmid (1 μg) were performed as described (*29*). After incubation for 40 h, protoplasts were processed for LUC and GUS assays as described (*30*).

Immunoblot Analysis. Arabidopsis thaliana protoplasts were transformed with the ATF expression vector containing an AZP targeting the NIM1 or the *Arabidopsis thaliana* DREB1A promoter (*31*). After incubation for 1 day, protoplast pellets were directly resuspended in 2 × SDS−PAGE sample buffer. Western blotting with an anti-FLAG M2 monoclonal Ab (Sigma) was performed as described (*32*).

RESULTS

Nondegenerate Recognition Code Table Design. The X-ray crystal structure of the DNA complex of the zinc-finger protein recognizing 5′-GAG GCA GAA C-3′ (*26*), which we designated as the Berg's protein in this paper, was used as the basis for our AZP design. In all three fingers of the protein, (i) each zinc-finger domain recognizes an overlapping 4-bp DNA sequence, and (ii) amino acids at specific positions contact with DNA bases at specific positions in a regular fashion as described in Figure 1A. These features hinted to us that knowing the DNA base specificities of amino acids at four critical positions (−1, 2, 3, and 6) is sufficient for design of zinc-finger proteins targeting diverse genomic DNA sequences simply by assembly of finger domains identical except for these four amino acids.

To test this hypothesis, based on known and potential DNA base−amino acid interactions (Figure 1B), which are derived mainly from information on X-ray crystal structures as described below, we constructed a nondegenerate recognition code table (Figure 1C). Based on the structure of the Berg's protein−DNA complex, amino acids with shorter and smaller side chains were used for recognition of the 2nd and 4th DNA bases, and amino acids with longer and larger side chains were used for recognition of the 1st and 3rd bases.

[1] Abbreviations: AZP, artificial zinc-finger protein; ATF, artificial transcription factor.

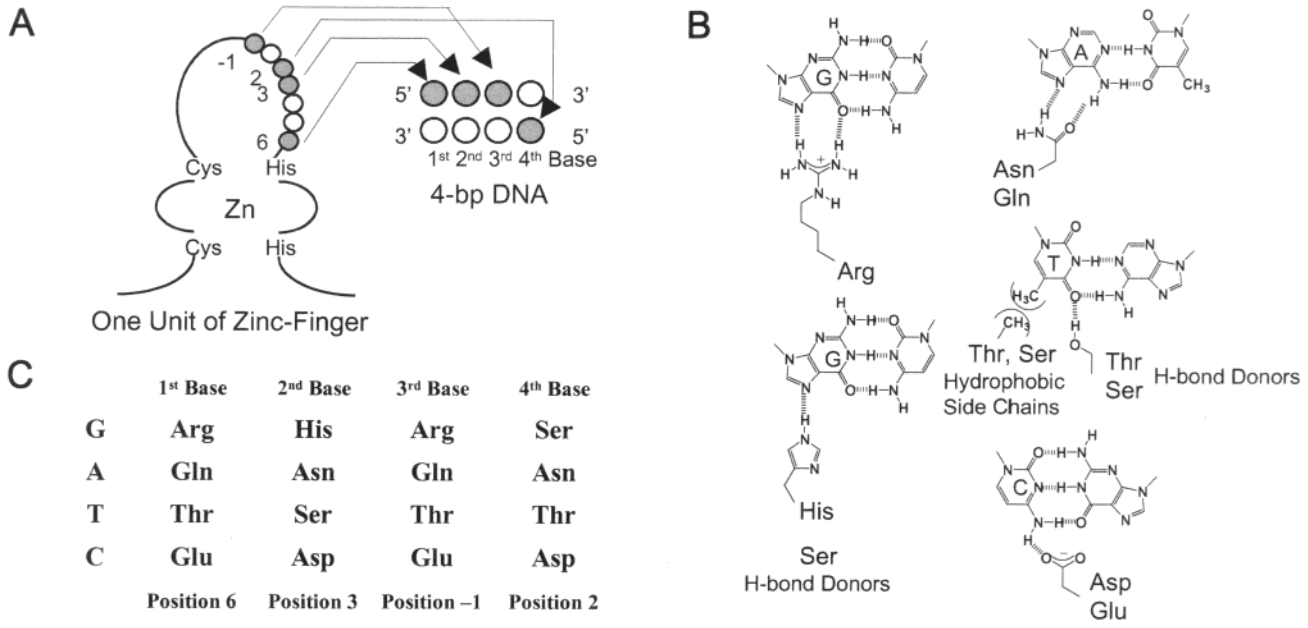

FIGURE 1: Nondegenerate recognition code table design. (A) DNA base−amino acid contacts observed in the Berg's protein. (B) Known and potential DNA base−amino acid interactions. (C) Nondegenerate recognition code table used in this report.

(1) Guanine recognition. Arginine was assigned to the amino acid at positions −1 and 6. Arginine interacts with guanine through divalent hydrogen-bonding as predicted 25 years ago by Rich (*33*). Furthermore, arginine has been selected exclusively against guanine recognition from a random library of a CCAAT/enhancer-binding protein (*37*). Histidine was used at position 3 because its interaction with guanine at the 2nd base was identified in Zif268 (*34*, *35*). Serine was assigned to position 2 for its less bulky side chain. It has been shown that the hydroxyl group of serine can contact the N7 of guanine in the λ repressor−DNA complex (*36*). The interaction with guanine at the 4th base was also observed in the Berg's protein (*26*).

(2) Adenine recognition. Glutamine recognizes adenine through divalent hydrogen-bonding as found in the 434 repressor−DNA complex (*37*). Therefore, the larger residue glutamine was assigned to positions −1 and 6, and the smaller residue asparagine to positions 2 and 3. These interactions were also predicted by Rich (*33*).

(3) Thymine recognition. Hydrophobic amino acid side chains interact with the methyl group of thymine. It is known that the methylene side chain of serine and the methyl group of threonine interact with the methyl group of thymine via hydrophobic interactions (*36*, *37*). These interactions could also be strengthened by the potential hydrogen-bonding between the hydroxyl group of serine/threonine and the O4 carbonyl of thymine. Therefore, threonine was assigned to positions −1, 2, and 6, and serine to position 3.

(4) Cytosine recognition. Cytosine presents an amino group in a DNA major groove. We reasoned that amino acids with a carboxyl group would therefore be advantageous because of potential hydrogen-bonding and electrostatic interaction between the carboxyl group and the amino group. Accordingly, the longer glutamic acid was assigned to positions −1 and 6, and the shorter asparatic acid to positions 2 and 3.

These guanine, adenine, thymine, and cytosine interactions were used to construct the nondegenerate recognition code table as shown in Figure 1C.

AZP Design. To evaluate our approach for zinc-finger protein design using the nondegenerate recognition code table, we designed and synthesized ten 3-finger AZPs (Table 1). Target DNA sites in tomato golden mosaic virus (TGMV) and beet curly top virus (BCTV) genomes are critical cis-elements for the gemini DNA viral replication (*38*, *39*). Other target sites were arbitrarily selected in the region of 50−100-bp upstream from each TATA box in plant gene promoters, *Arabidopsis thaliana* DREB1A (drought tolerance, *31*), NIM1 (systemic acquired resistance, *28*), *Arabidopsis thaliana* and *Oliyz sativa* NHX1 (salt tolerance, *40*). These target DNA sequences are very diverse in terms of base composition.

The AZP design and DNA construction strategy are illustrated in Figure 2. We used the 1st finger domain of the Berg's protein as our universal finger framework (*26*) and assembled the finger frameworks with no linker. In this example, the procedure for design and construction of AZP-1 targeting the AL1 site of TGMV is described. First, the 10-bp target was divided into three overlapping 4-bp DNA segments, where the last base of each 4-bp target is the first base of the next 4-bp target (Step 1). Then, four amino acids per finger domain were chosen from our nondegenerate recognition code table (Step 2). For example, in the case of the first finger (designated as Zif1), arginine was selected for the 1st guanine base recognition, serine for the 2nd thymine base, glutamine for the 3rd adenine base, and aspartic acid for the 4th cytosine base in the antisense strand. Based on these selected amino acids, each coding DNA of the 3 fingers was carefully designed for the PCR assembly in the correct order, and 2 DNA oligomers containing 55−60 nucleotides per finger were synthesized (Step 3). Next, each pair of oligomers was annealed and filled-in with Klenow fragment to produce a complete DNA duplex encoding one finger domain (Step 4). Finally, the three coding DNA fragments were assembled in the correct order by PCR (Step 5). The procedure of AZP-coding DNA design/construction we used ensured that only a correct 300-bp

Table 1: Design and Binding Properties of 3-Finger AZPs[a]

			amino acids used for recognition												
			Zif 1				Zif 2				Zif 3				K_D
AZP	target genes	target sequences	−1	2	3	6	−1	2	3	6	−1	2	3	6	(nM)
AZP-1	TGMV	5′-AGT AAG GTA G-3′	Gln	Asp	Ser	Arg	Arg	Asp	Asn	Gln	Thr	Thr	His	Gln	18
AZP-2	BCTV-#1	5′-TTG GGT GCT C-3′	Thr	Ser	Asp	Arg	Thr	Asp	His	Arg	Arg	Asp	Ser	Thr	15
AZP-3	BCTV-#2	5′-CGG ATG GCC C-3′	Glu	Ser	Asp	Arg	Arg	Asp	Ser	Gln	Arg	Thr	His	Glu	NB
AZP-4	DREB1A	5′-TAC GTG GCA T-3′	Gln	Asn	Asp	Arg	Arg	Asp	Ser	Arg	Glu	Asp	Asn	Thr	11
AZP-5	DREB1A-2	5′-ATA GTT TAC G-3′	Glu	Asp	Asn	Thr	Thr	Asn	Ser	Arg	Gln	Asp	Ser	Gln	NB
AZP-6	NIM1	5′-GAT ATA AAT A-3′	Thr	Thr	Asn	Gln	Gln	Thr	Ser	Gln	Thr	Thr	Asn	Arg	NB
AZP-7	NIM1-2	5′-GGA GAT GAT A-3′	Thr	Thr	Asn	Arg	Thr	Asp	Asn	Arg	Gln	Asp	His	Arg	23
AZP-8	At NHX1	5′-ATC GTA GAC G-3′	Glu	Asp	Asn	Arg	Gln	Asp	Ser	Arg	Glu	Asp	Ser	Gln	NB
AZP-9	At NHX1-2	5′-GAC GAT AAA A-3′	Gln	Thr	Asn	Gln	Thr	Thr	Asn	Arg	Glu	Asp	Asn	Arg	NB
AZP-10	Os NHX1	5′-GTT GCG GGA T-3′	Gln	Asn	His	Arg	Arg	Asp	Asp	Arg	Thr	Asp	Ser	Arg	25
Zif268															4

[a] Synthetic DNA duplexes consisted of the sequence 5′-TATATATAN$_{10}$TATATATA-3′. NB: No band shift observed with 128 nM zinc-finger proteins.

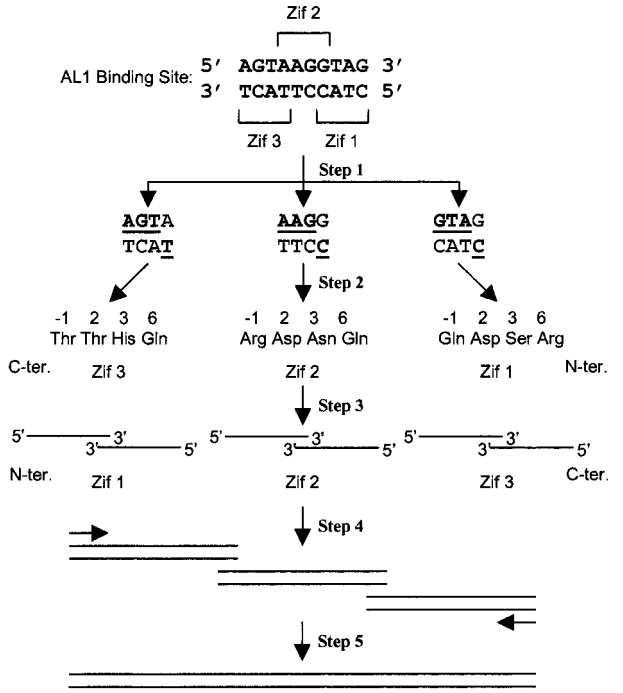

FIGURE 2: Design of an AZP targeting the TGMV AL1-binding site (AZP-1). Step 1: Division of the 10-bp target into three overlapping 4-bp DNA segments, where the last base of each 4-bp target is the first base of the next 4-bp target. Step 2: Assignment of four amino acids per finger domain from our nondegenerate recognition code table. Step 3: Design of the AZP-coding DNA fragments and synthesis of the oligonucleotides. Step 4: Annealing and fill-in reaction to make double-strand DNA fragments encoding each finger domain. Step 5: PCR assembly of three finger-coding fragments to construct the entire AZP-coding DNA.

fragment consisting of the three domains Zif1, Zif2, and Zif3 was produced (data not shown).

AZP Characterization. Each AZP-coding DNA fragment was cloned into the *Eco*RI/*Hin*dIII sites of pET-21a, and the corresponding proteins were expressed in *E. coli* and purified. The interactions of the purified proteins with target DNA were analyzed using gel shift assays. For example, AZP-2 showed band shifts at the protein concentration ranging from 1 to 128 nM (Figure 3A). In AZP-5, however, no band shift was observed at 128 nM protein concentration (Figure 3B). Each dissociation constant was determined by curve-fitting as shown in Figure 3C. All results are summarized in Table 1. These results indicate that the AZPs designed using our nondegenerate recognition code table can recognize target DNA containing more than three guanine bases in the first 9 bp with high affinities. An exception is AZP-3, however, which could not bind to a target containing four guanines and three cytosines in the first 9 bp at 128 nM protein concentration.

Selectivities of the functional AZPs were also examined using mutant probes with one or more altered base pairs in the DNA targets. The assay result of AZP-2 is shown in Figure 3D. Even a single base pair mutation resulted in complete loss of band shift, as did all the more complex mutations we tested. The other four functional AZPs also showed same results, published as Supporting Information Figure S1 on the Internet at http://pubs.acs.org. Thus, the half of AZPs designed using our nondegenerate recognition code table can recognize their target DNA with both high affinity and selectivity.

Multi-Finger AZP with Extremely High Affinity. We also developed a new method for assembly of any number of zinc-finger domains (details of this assembly method will be described elsewhere). In this report, 5- or 6-finger domains were assembled with no linker. The target genes and sequences are shown in Table 2.

The gel shift assays demonstrate that the affinities of all multi-finger AZPs are dramatically improved in comparison with the original 3-finger AZPs (Figure 4). Especially in the 6-finger AZP-A4 (Figure 4B), which was prepared by assembly of two of the functional AZP-2, most of the DNA probe is still bound to the protein even at 3 pM (the lowest protein concentration used in our assay), which means the dissociation constant should be less than 3 pM (e.g., in the femtomolar range). This phenomenon was also confirmed in other 6-finger AZPs [i.e., AZP-A5 and AZP-A6 (Figure 4C, Table 2)], prepared by assembly of two functional 3-finger AZPs. These 6-finger AZPs show the highest affinities among previously reported zinc-finger proteins. Interestingly, the 5-finger AZP-A3 assembled from two nonfunctional 3-finger AZPs, AZP-8 and AZP-9, bound to the target DNA containing three guanines with the dissociation constant of 75 nM.

Efficient Expression of ATF and Gene Activation by ATF in Arabidopsis. In the current version of our AZP/ATF, the codons are optimized for expression in both *E. coli* and a

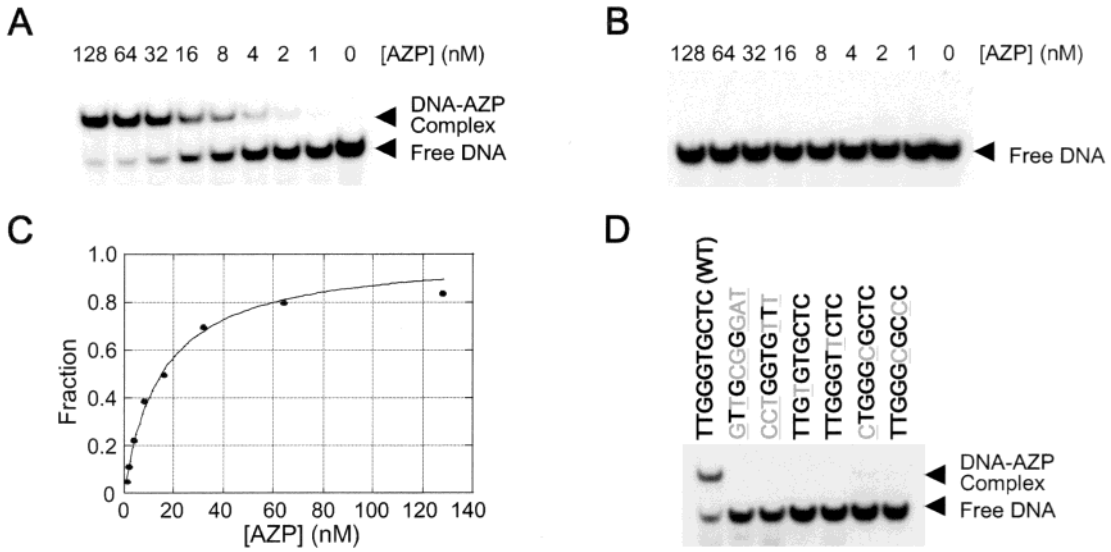

FIGURE 3: Gel shift assays with 3-finger AZPs. (A) Gel shift by AZP-2. (B) Gel shift by AZP-5. (C) Curve-fitting of the result of AZP-2. (D) Sequence-specific binding of AZP-2. Underlining represents mutation.

Table 2: Design and Binding Properties of Multi-Finger AZPs

AZP	target genes	finger no.	target sequences	assembled AZPs	K_D
AZP-A1	DREB1A	5	5′-ATA GTT TAC GTG GCA T-3′	AZP-4 and AZP-5	1.5 nM
AZP-A2	NIM1	5	5′-GGA GAT GAT ATA AAT A-3′	AZP-6 and AZP-7	0.17 nM
AZP-A3	At NHX1	5	5′-ATC GTA GAC GAT AAA A-3′	AZP-9 and AZP-8	75 nM
AZP-A4	BCTV	6	5′-TTG GGT GCT TTG GGT GCT C-3′	AZP-2 and AZP-2	<3 pM
AZP-A5	N/A	6	5′-AGT AAG GTA GGA GAT GAT A-3′	AZP-7 and AZP-1	<3 pM
AZP-A6	N/A	6	5′-TAC GTG GCA TTG GGT GCT C-3′	AZP-2 and AZP-4	<3 pM

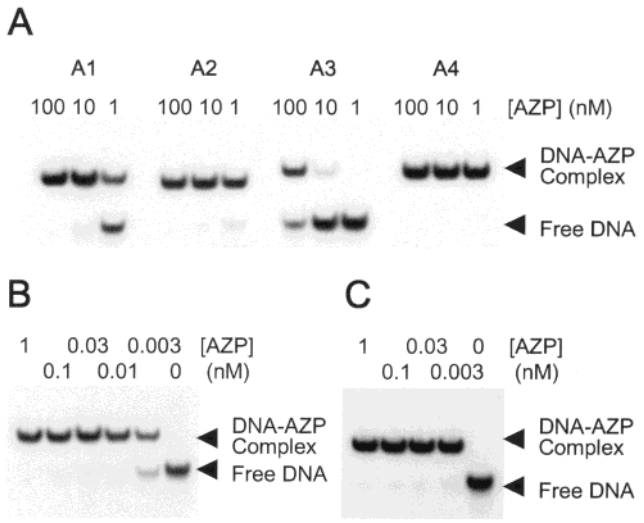

FIGURE 4: Gel shift assays with multi-finger AZPs. (A) Gel shift by AZP-A1, AZP-A2, AZP-A3, and AZP-A4 at 1, 10, and 100 nM protein concentrations. (B) Gel shift by AZP-A4 at 0.003−1 nM protein concentration. (C) Gel shift by AZP-A6.

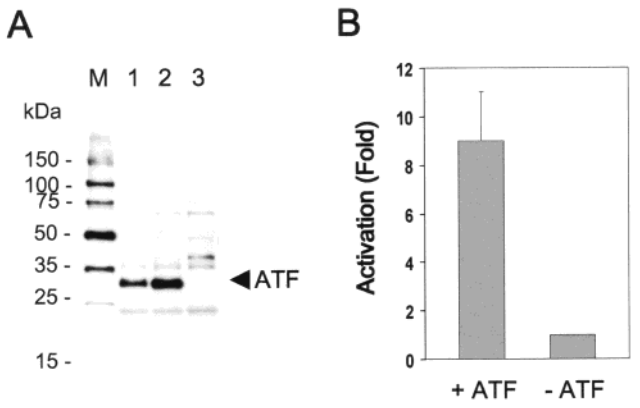

FIGURE 5: ATF expression and NIM1 promoter activation by ATF in *Arabidopsis thaliana* protoplasts. (A) Western blot with an anti-FLAG antibody. Lane 1: ATF for the DREB1A promoter activation. Lane 2: ATF for the NIM1 promoter activation. Lane 3: Mock. (B) *Arabidopsis* NIM1 promoter activation by ATF in protoplast transient assays.

dicot plant, *Arabidopsis thaliana*. As shown in Figure 5A, Western blot analysis with an anti-FLAG antibody showed that ATFs are expressed well in *Arabidopsis* protoplasts.

Quantitative transient gene expression analysis was also performed. In this assay, *Arabidopsis* leaf protoplasts were cotransfected with an ATF expression plasmid for the *Arabidopsis* NIM1 promoter activation, derived from AZP-A2, and a luciferase reporter plasmid under control of the NIM1 native promoter. As shown in Figure 5B, significant gene activation was observed. This activation level was comparable to or higher than that reported for mammalian cells using a VP16 activation domain (*19*, *21*).

DISCUSSION

Three degenerate recognition code tables were previously reported, derived from zinc-finger domains selected by phage display (*23*, *24*) or constructed by the computational approach (*41*). The tables emphasized three amino acids at positions −1, 3, and 6. In the phage display approach, a library randomized at positions −1, 1, 2, 3, 5, 6, and 8 in Zif286 was used, and 3-bp segments in the 5′-GCG XXX GCG-3′ (X: targeted bp) were targeted, where the 4th base (under-

lined) was guanine (*12*, *23*). Since 16 DNA triplets (mainly 5′-GNN-3′) were targeted in the selection, the table constructed from the selected finger domains was partial. It depended on the 3-bp target sequences which amino acid was selected against a DNA base at the specific position. For example, regarding amino acids at position 3 recognizing the 2nd cytosine base, valine and threonine at position 2 were selected toward 5′-ACG-3′ and 5′-GCA-3′, respectively (*12*). And various sets of amino acids at positions −1, 3, and 6 were selected against the single 3-bp target. Therefore, their table produced from the phage display selection was partial and degenerate (*23*). To date, successful rational design of zinc-finger proteins using the degenerate recognition code table has not been reported. Recently, Choo and Isalan have reported that the code table is degenerate and not absolute, and rational design using the table is quite unreliable (*42*). No rational design using the other two degenerate tables has been reported, either.

Construction of zinc-finger proteins by combination of finger domains targeting 3-bp DNA, selected by phage display, was reported by certain groups (*16*−*19*). It is possible to select finger domains targeting all possible 64 triplet DNA sequences by phage display. Actually, selection of finger domains targeting all 16 triplets of 5′-GNN-3′ by the phage display approach was reported (*15*) and showed efficient targeting of 5′-GNNGNNGNN-3′ (*19*). Recently, the same group has also reported selection against 5′-ANN-3′ (*43*). However, since X-ray crystal data of DNA complexes with zinc-finger proteins support direct interaction of amino acids at position 2 with the 4th DNA bases (*26*, *34*, *35*) and the 4th bases in the phage selection are always guanines (*15*, *43*), finger domains available for 4-bp recognition are still limited (256 phage display selection should be done finally).

Using our nondegenerate recognition code table, constructed based on known and potential DNA base−amino acid interactions, we showed five designed 3-finger AZPs bound to their targets with high affinities and selectivities, while the other five designed 3-finger AZPs did not under our experimental conditions. As shown in Table 1, the successfully recognized targets have at least three guanines at various positions in the first 9-bp DNA. In contrast, targets not recognized by our AZPs have less than two guanines in the first 9-bp DNA (guanines at the 10th position are not counted because cytosines in antisense strands at the position are recognized by AZP in our model, see Supporting Information Figure S2). In addition, assembly of two nonfunctional 3-fingers, AZP-8 and AZP-9, resulted in production of a 5-finger AZP-A3 with significant affinity (Table 2). The 16-bp target also has three guanines within the 9-bp region (i.e, 5′-ATCGTAGACGATAAAA-3′). These results suggest that three guanines are necessary for successful DNA recognition by our designed AZPs. Since Arg and His are used for the guanine recognition in our design scheme, three basic amino acids may be important for the DNA recognition. If zinc-finger proteins recognize target DNA in the "sliding mechanism" (i.e., first nonspecific DNA binding, then sliding to the target site) as observed in λ Cro repressor and *Eco*RI restriction endonuclease (*44*), basic amino acids could contribute to the nonspecific binding via ionic interactions with phosphate backbones in the first step. The nonspecific binding contributes to the total affinity in this mechanism (*44*). Basic amino acids at positions −1, 3, and 6 of our AZPs may contribute not only to DNA base recognition but also to interactions with phosphate backbones in the first step. To our knowledge, all reported functional zinc-finger proteins recognize guanine-rich sequences (more than three guanines per 9 bp) except one Zif268 derivative targeting 5′-GCTATAAAA-3′ (*14*). Interestingly, the zinc-finger protein also has three basic amino acids in the region of positions −1 to 6 of the three fingers. In our experiments, one example with three cytosines in the first 9 bp does not satisfy the observation (AZP-3, Table 1). In our design scheme, aspartic and glutamic acids are used for cytosine recognition. These acidic amino acids may reduce the ionic interactions with phosphate backbones, thereby reducing the affinity toward the target DNA.

A couple of groups have previously reported multi-finger proteins with no flexible linker peptide; however, no protein exhibited the expected high affinities (*19*, *45*, *46*). For example, two 3-finger Zif268 variants with a dissociation constant of 31.8 nM, which was selected by phage display (*13*), were assembled to produce a 6-finger protein (*45*). Although the dissociation constant of the resulting 6-finger protein was predicted to be 1.0 fM (31.8 nM × 31.8 nM) if there was no structural perturbation, the experimentally determined dissociation constant was only 0.46 nM. Other 6-finger proteins produced by the group also showed similar results (*19*). These suggest that the affinities of their 6-finger proteins are significantly reduced due to structural perturbation caused by the assembly. The finger proteins have three different finger frameworks derived from Zif268, and six amino acid variations at positions −1, 2, 3, 4, 5, and 6 per finger (*19*, *45*, *46*). In contrast, our designed 6-finger AZPs showed much higher affinities than other designed multi-finger proteins. For example, the dissociation constant of our 6-finger AZP-A6 is much lower than 3 pM (see Figure 4C). In our AZPs, identical finger frameworks (i.e., the 1st finger framework of the Berg's protein) differing only at the four amino acids at positions −1, 2, 3, and 6 are assembled. As described already, in the DNA complex of the Berg's protein, all amino acids at positions −1, 2, 3, and 6 contact one specific DNA base, respectively (*26*). In contrast, in the DNA complexes of native Zif268 and Zif268 variants selected by phage display, only subsets of the four amino acids contact their target DNA (*34*, *35*, *47*), except one finger bound to 5′-GACC-3′ (*47*). The 3-D structure of our 6-finger AZP will likely reveal the basis of its high affinity by comparison with the structure of other zinc-finger proteins. Our AZP design scheme also enables the efficient targeting of long genomic sequences.

Our nondegenerate recognition code table and our AZP design method described here enable us to construct a "zinc-finger domain library" for high-through-put production of AZPs targeting diverse DNA sequences. The "zinc-finger domain library" is composed of 3 sets of 256 double-stranded DNA fragments (102 bp), which are synthesized as described for AZP design (Figure 2). Each set is designed for each of the 1st, 2nd, and 3rd fingers to target the 256 different 4-bp sequences. The library allows rapid production of 3-finger AZPs targeting all possible 10 bp (i.e., 1.06×10^6 sequences) from a total of 768 DNA fragments. Even under our current "three guanines" limitation, the library covers 47.4% of all possible 10-bp DNA sequences (4.97×10^5 sequences). By combining with our newly developed assembly method, this

library can potentially cover diverse genomic sequences ranging from 13 bp (by 4-finger AZPs) to over 31 bp (by >10-finger AZPs). Thus, our "zinc-finger domain library" will enable creation of numerous AZPs in a high-through-put way, each of which is designed to target one specific sequence. The current version of our AZP-coding DNA is codon-optimized for expression in *E. coli* and *Arabidopsis thaliana* (*48*). With our method, it is possible to optimize AZP expression for particular organisms by selecting appropriate codon usage as shown in Figure 5A. We are currently constructing "zinc-finger domain libraries" that are optimized for the codon usage of rice (*Oriyz sativa*) and human, respectively.

In the design of transcription factors, the importance of chromatin structure on gene regulation is being realized (*21*). A recent report suggests that one approach for efficient manipulation of gene expression is to design transcription factors targeting nucleosome-free sites in genomes, determined by chromatin structural analysis (*22*). However, we do not have any means to predict or find nucleosome-free sites in a given promoter without performing experimental analysis. Furthermore, there is yet no general rule about which nucleosome-free sites in each promoter should be targeted for optimized gene regulation. A practical solution to these challenges is to design multiple transcription factors that target multiple sites for a given gene and to select one(s) that perform(s) at a satisfactory level. This approach would be facilitated by a rapid and efficient DNA-binding protein design scheme.

In this study, we present a new AZP design scheme using a nondegenerate recognition code table. This design scheme allows rapid creation of numerous AZPs with satisfactory binding properties in a high-through-put manner. Our AZP design scheme will enable manipulation of numerous gene expressions with no chromatin structural information on target genes by screening multiple ATFs per target gene.

ACKNOWLEDGMENT

We thank Fumiaki Katagiri, Michelle Wood, and Yi Tao for technical advice on protoplast transformation, LUC/GUS analysis, and *Arabidopsis thaliana* plant maintenance; Tomoko Sera for her help in plasmid construction; Maureen Milnamow and Natasha Ginzburg for extensive DNA sequencing; and Fumiaki Katagiri, Peter Heifetz, and Tim Torchia for their critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Figures of gel shift assays of AZP-1, AZP-4, AZP-7, and AZP-10 toward mutant DNA targets and a proposed model of 10-bp DNA recognition by a 3-finger AZP. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Wolfe, S. A., Kekludova, L., and Pabo, C. O. (2000) *Annu. Rev. Biophys. Biomol. Struct. 29*, 183−212, and references cited therein.
2. Krizek, B. A., Amann, B. T., Kilfoil, V. J., Merkle D. L., and Berg J. M. (1991) *J. Am. Chem. Soc. 113*, 4518−4523.
3. Desjarlais, J. R., and Berg, J. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7345−7349.
4. Desjarlais, J. R., and Berg, J. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11099−11103.
5. Pomerantz, J. L., Sharp, P. A., and Pabo, C. O. (1995) *Science 267*, 93−96.
6. Kim, J.-S., and Pabo, C. O. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 2812−2817.
7. Moore, M., Choo, Y., and Klug, A. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 1432−1436.
8. Moore, M., Klug, A., and Choo, Y. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 1437−1441.
9. Pomerantz, J. L., Wolfe, S. A., and Pabo, C. O. (1998) *Biochemistry 37*, 965−970.
10. Wang, B. S., and Pabo, C. O. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 9568−9573.
11. Rebar, E. J., and Pabo, C. O. (1994) *Science 263*, 671−673.
12. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11163−11167.
13. Wu, H., Yang, W.-P., and Barbas, C. F., III (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 344−348.
14. Greisman, H. A., and Pabo, C. O. (1997) *Science 275*, 657−661.
15. Segal, D. J., Dreier, B., Beerli, R. R., and Barbas, C. F., III (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 2758−2763.
16. Choo, Y., Sanchez-Garcia, I., and Klug, A. (1994) *Nature (London) 372*, 642−645.
17. Corbi, N., Perez, M., Maione, R., and Passananti, C. (1997) *FEBS Lett. 417*, 71−74.
18. Corbi, N., Libri, V., Fanciulli, M., and Passananti, C. (1998) *Biochem. Biophys. Res. Commun. 253*, 686−692.
19. Beerli, R. R., Segal, D. J., Dreier, B., and Barbas, C. F., III (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 14628−14633.
20. Beerli, R. R., Dreier, B., and Barbas, C. F., III (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 1495−1500.
21. Zhang, L., Spratt, S. K., Liu, Qiang, Johnstone, B., Qi, H., Raschke, E. E., Jamieson, A. C., Rebar, E. J., Wolffe, A. P., and Case, C. C. (2000) *J. Biol. Chem. 275*, 33850−33860.
22. Liu, P.-Q., Rebar, E. J., Lei, Z., 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., Soratt, S. K., Case, C. C., and Wolffe, A. P. (2001) *J. Biol. Chem. 276*, 11323−11334.
23. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11168−11172.
24. Wolfe, S. A., Greisman, H. A., Ramm, E. I., and Pabo, C. O. (1999) *J. Mol. Biol. 285*, 1917−1934.
25. Dreier, B., Segal, D. J., and Barbas, C. F., III (2000) *J. Mol. Biol. 303*, 489−502.
26. Kim, C. A., and Berg, J. M. (1996) *Nat. Struct. Biol. 3*, 940−945.
27. Sera, T., and Schultz, P. G. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2920−2925.
28. Ryals, J., Weymann, K., Lawton, K., Friedrich, L., Ellis, D., Steiner, H. Y., Johnson, J., Delaney, T. P., Jesse, T., Vos, P., and Uknes, S. (1997) *Plant Cell 9*, 425−439.
29. Leister, T. R., and Katagiri, F. (2000) *Plant J. 22*, 345−354.
30. Leister, T. R., Frederick, A. M., and Katagiri, F. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 15497−15502.
31. Shinwari, Z. K., Nakashima, K., Miura, S., Kasuga, M., Seki, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998) *Biochem. Biophys. Res. Commun. 250*, 161−170.
32. Sera, T., and Wolffe, A. P. (1998) *Mol. Cell. Biol. 18*, 3668−3680.
33. Seeman, N. C., Rosenberg, J. M., and Rich A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804−808.
34. Pavletich, N. P., and Pabo, C. O. (1991) *Science 252*, 809−817.
35. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure 4*, 1171−1180.
36. Jordan, S. R., and Pabo, C. O. (1988) *Science 242*, 893−899.
37. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) *Science 242*, 899−907.
38. Fontes, E. P. B., Gladfelter, H. J., Schaffer, R. L., Petty, I. T. D., and Hanley-Bowdoin, L. (1994) *Plant Cell 6*, 405−416.
39. Stenger, D. C. (1994) *Virology 203*, 397−402.

40. Gaxiola, R. A., Rao, R., Sherman, A., Grisafi, P., Alper, S. L., and Fink, G. R. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 1480−1485.
41. Suzuki, M., and Yagi, N. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12357−12361.
42. Choo, Y., and Isalan, M. (2000) *Curr. Opin. Struct. Biol. 10*, 411−416.
43. Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., and Barbas, C. F., III (2001) *J. Biol. Chem. 276*, 29466−29478.
44. Shimamoto, N. (1999) *J. Biol. Chem. 274*, 15293−15296, and references cited therein.
45. Liu, Q, Segal, D. J., Ghiara, J. B., and Barbas, C. F., III (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5525−5530.
46. Kamiuchi, T., Abe, E., Imanishi, M., Kaji, Tamaki, Nagaoka, M., and Sugiura, Y. (1998) *Biochemistry 37*, 13827−13834.
47. Elrod-Erickson, M., Benson, T. M., and Pabo, C. O. (1998) *Structure 6*, 451−464.
48. http://www.kazusa.or.jp/codon.

BI020095C