

Development of Zinc Finger Domains for Recognition of the 5'-CNN-3' Family DNA Sequences and Their Use in the Construction of Artificial Transcription Factors*

Received for publication, June 20, 2005, and in revised form, August 16, 2005 Published, JBC Papers in Press, August 17, 2005, DOI 10.1074/jbc.M506654200

Birgit Dreier^{†1}, Roberta P. Fuller[‡], David J. Segal[§], Caren V. Lund[‡], Pilar Blancafort[‡], Adrian Huber[‡], Beate Kokschi[¶], and Carlos F. Barbas III^{‡2}

From [†]The Skaggs Institute for Chemical Biology and the Departments of Molecular Biology and Chemistry, The Scripps Research Institute, La Jolla, California 92037, [‡]University of California, University of California Davis Genome Center and Departments of Medical Pharmacology and Toxicology, Davis, California 95616, and [¶]Freie Universität Berlin, Institut für Chemie, Takustrasse 3, 14195 Berlin, Germany

Considerable progress has been made in recent years in the design of transcription factors for the directed regulation of endogenous genes. Although many strategies involve selection methods that must be applied for each new target sequence, we have developed an approach based on linkage of predefined zinc finger domains that each recognize a three-base pair DNA sequence to construct artificial transcription factors that bind to a desired sequence. These domains can be assembled to recognize unique 18-base pair DNA sequences with high specificity. Here we report the development and characterization of zinc finger domains that bind to 15 of the 16 5'-CNN-3' subsites. These domains were created through a combination of phage display selection, site-directed mutagenesis, and *de novo* design. Furthermore, these domains were used to generate a highly specific six-finger protein targeting the *ERBB-2* promoter. When fused to regulatory domains, this protein was capable of up- and down-regulating the expression of the endogenous *ERBB-2* gene. With the addition of this collection of predefined zinc finger domains, most 5'-CNN-3', 5'-GNN-3', and 5'-ANN-3'-containing sequences can now be rapidly targeted for directed gene regulation and nuclease cleavage.

The ability to rapidly prepare proteins with predefined specificities for DNA sequences could enable a wide range of technologies that might be used, for example, to direct the expression of genes or to physically modify genes and genomes. To develop a universal system for gene regulation, much effort has been applied to the development of artificial transcription factors based on polydactyl zinc finger proteins (1–3). Such a system might have considerable impact on biology and biotechnology and offer a new approach for treatment of diseases based on directed gene regulation. It has now been shown that gene expression can be specifically altered using artificial transcription factors based on polydactyl zinc finger proteins that bind to 18-bp target sites (1, 2). Targeting of sites as small as 9 bp can also provide some degree of

regulatory specificity presumably through the aid of chromatin occlusion (4–6). In addition to transcriptional regulation, novel zinc finger DNA binding specificities show tremendous promise in directing homologous recombination through their fusion with the Fok I nuclease domain (7, 8).

Zinc finger domains of the type Cys₂-His₂ are a unique and promising class of proteins for the recognition of extended DNA sequences due to their modular nature. Each domain consists of ~30 amino acids folded into a $\beta\beta\alpha$ structure stabilized by hydrophobic interactions and chelation of a zinc ion by the conserved Cys₂-His₂ residues (9, 10). To date, the best-characterized protein of this family of zinc finger proteins is the mouse transcription factor Zif268. Each of the three zinc finger domains of Zif268 binds to a 3-bp subsite by insertion of the α -recognition helix into the major groove of the DNA double helix (11, 12). To facilitate the rapid construction of DNA-binding proteins and to study protein-DNA interactions, we have previously created domains that bind to the 5'-GNN-3' and 5'-ANN-3' family of DNA sequences (13–15). We demonstrated that these domains function as modular recognition units that can be assembled into polydactyl zinc finger proteins that specifically recognize from 9- to 18-bp target sites. Significantly, an 18-bp site is long enough to potentially be unique within the human or any other genome, and transcriptional specificity of such proteins has been demonstrated in transgenic plants and human cells using array analysis (16, 17). In addition to constitutive regulation, fusion of ligand binding domains from nuclear hormone receptors with specific binding domains provides inducible gene regulation with this class of transcription factors (18). To provide for ultimate freedom in DNA targeting, it is important to identify the 64 DNA binding domains required to target each possible 3-bp subsite.

Due to the limited structural data on zinc finger/DNA interactions (19–24) *de novo* design of zinc proteins that bind with a high degree of specificity to novel sequences has been of limited success (25). Significantly, for the study reported here there is no structural information available on the interaction of natural zinc finger domains with 5'-CNN-3' subsites. Finger 4 of YY1, which contains a DNA recognition helix with the sequence QST-N-LKS (the sequence is given starting from the first residue (–1) proximal to the N terminus of the α -helix of the protein) that binds to the DNA subsite 5'-CAA-3' in the context of the full-length protein does not directly interact with the 5' cytosine and does not bind this site specifically (22). Crystallographic data and mutagenesis studies concerning the mode of interaction of zinc finger domains of the Cys₂-His₂ family has guided us in the construction of phage display libraries for selection of domains that recognize many DNA subsites (13–15). The analysis of the Zif268-DNA complex sug-

* This study was supported by National Institutes of Health Grants CA086258 and GM065059 (to C. F. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) DQ160159 and DQ160160.

¹ Recipient of a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft. Present address: University of Zurich, Dept. of Biochemistry, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

² To whom correspondence should be addressed: The Scripps Research Institute, BCC-550, North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-784-9098; Fax: 858-784-2583; E-mail: carlos@scripps.edu.

gests that DNA binding is predominantly achieved by the interaction of amino acid residues of the α -helix in positions -1, 3, and 6 with the 3', middle, and 5' nucleotides of a 3-bp DNA subsite, respectively (11, 12). Positions 1, 2, and 5 of the α -helix make direct or water-mediated contacts with the phosphate backbone of the DNA and are important contributors to the ultimate specificity of the protein. Leucine is typically found in position 4 and packs into the hydrophobic core of the domain. Position 2 of the α -helix interacts with other helix residues and, in addition, can make contact with a nucleotide outside the 3-bp subsite, resulting in target site overlap (13, 15, 26–28).

Phage display libraries based on Zif268 were suitable for the selection of domains that bound to 5'-GNN-3' motifs (13, 14). Due to target site overlap issues wherein some zinc finger domains interact with extended recognition contacts, the selection of 5'-ANN-3' required refinement of the phage display library. This was achieved by replacement of finger 3 of Zif268 containing an Asp at position 2 of the α -helix with a domain lacking residues mediating interdomain recognition. From this library we selected domains for the 5'-ANN-3' subsites and further refined or designed novel domains through site-directed mutagenesis (15). Other groups have applied different selection strategies toward the development of zinc finger domains with altered DNA binding specificity (1–3).

Here we report a selection approach based on the modularity of zinc finger domains to extend the existing set of predefined domains to domains specific to 5'-CNN-3' target sequences. From phage display libraries, eight zinc finger domains specifically recognizing 5'-CNN-3' target sites were selected. Improvement of the DNA binding specificity for five domains was achieved by site-directed mutagenesis. For six of the 5'-CNN-3' target sites, specific domains were generated by *de novo* design. Resulting proteins were analyzed for DNA binding specificity. Furthermore, we demonstrate that these domains can be used as modules for the construction of artificial transcription factors with DNA binding specificity for an 18-bp target site. When fused to the VP64 activation domain or the KRAB³ repression domain, the six-finger protein E2S targeted at the 5'-untranslated region of the human *ERBB-2* gene was capable of altering the expression of the endogenous gene. The results reported here provide new insight into zinc finger/DNA recognition of 5'-CNN-3' subsites and significantly extends the repertoire of DNA sequences that can be targeted with designed transcription factors and nucleases.

MATERIALS AND METHODS

Construction of Zinc Finger Library and Selection via Phage Display—Construction of the zinc finger library was based on the earlier described C7 protein (29). Finger-3 recognizing the 5'-GCG-3' subsite was replaced by a domain known to bind to a 5'-GAT-3' subsite (13) via a overlap PCR strategy using a primer coding for finger-3 (5'-GAGGAAGTTTGCCACCAGTGGCAACCTGGTGAGGCATACCAAAATC-3') and a pMal-specific primer (5'-GTA-AAACGACGGCCAGTGCCAAGC-3') as previously described (15). Randomization of the zinc finger library by PCR overlap extension was essentially as described (13, 29). The library was ligated into the phagemid vector pComb3H (30). Growth and precipitation of phage were performed as previously described (13, 31, 32). Binding reactions were performed in a volume of 500 μ l of zinc buffer A (10 mM Tris, pH 7.5, 90 mM KCl, 1 mM MgCl₂, 90 μ M ZnCl₂, 0.2% bovine serum albumin, 5 mM dithiothreitol, 1% Blotto (Bio-Rad), 40 μ g of double-stranded, sheared herring sperm DNA containing 100 μ l of precipitated phage (10^{13} colony-forming units). Phage was allowed to bind to non-biotinylated competitor oligonucleotides for 1 h at 4 °C before the biotinylated

target oligonucleotide was added. Binding continued 3 h or overnight at 4 °C. After incubation with 50 μ l of streptavidin-coated magnetic beads (Dyna; blocked with 5% Blotto in zinc buffer A) for 1 h, beads were washed 10 times with 500 μ l of zinc buffer A, 2% Tween 20, and 5 mM dithiothreitol and once with buffer containing no Tween. Elution of bound phage was performed by incubation in 25 μ l of trypsin (10 μ g/ μ l) in Tris-buffered saline (50 mM Tris, 150 mM NaCl, pH 7.5) for 30 min at room temperature.

Target and competitor oligonucleotides were designed to form hairpins and had the sequence 5'-GGCCGCN'N'N'ATCGAGTTTCTC-GATNNNGCGGCC-3' (target oligonucleotides were biotinylated at the 5' end), where NNN represents the finger-2 subsite, and N'N'N' represents complementary bases. Target oligonucleotides were usually added at 72 nM in the first 3 rounds of selection, then decreased to 36 and 18 nM in the 5th and 6th round. The wild type competitor hairpin in the initial round of selection had the sequence at NNN equal to 5'-TGG-3' at a concentration of 108 nM and was increased in each round to up to 460 nM in the sixth round. For successive rounds of selection, increasing amounts of an equimolar mixture of 15 hairpin oligonucleotides with all finger-2 5'-CNN-3' subsites, except for the target sequence, were usually added in the first round at a 5 M excess of target. Mixtures of each finger-2 subsite sequence of the type 5'-ANN-3', 5'-GNN-3', and 5'-TNN-3' were usually added at a 1.25 M excess of target in the first round and was increased to a 10- or even 40-fold excess of the target sequence depending on the experiment. In addition, competitors with NNN of 5'-CGA-3', 5'-CAG-3', 5'-CGG-3', and/or 5'-CTG-3' (if these were not the target sites) were included at concentrations up to 180 nM to enforce selection for specific recognition of the particular target site.

Multitarget Specificity Enzyme-linked Immunosorbent Assay (ELISA)—The zinc finger-coding sequence was subcloned from pComb3H (30, 31) into a modified bacterial expression vector pMal-c2 (New England Biolabs). The zinc finger-maltose-binding protein fusions were transformed into XL1-Blue (Stratagene) and expressed after the addition of 1 mM isopropyl β -D-thiogalactoside. Freeze/thaw extracts of these bacterial cultures or purified proteins were applied in 1:2 dilutions to 96-well plates coated with streptavidin (Pierce) and were tested for DNA binding specificity against each of the 16 5'-GAT CNN GCG-3' target sites using the hairpin oligonucleotide described above. ELISA was performed essentially as described (13, 14). After incubation with a mouse anti-MBP (maltose-binding protein) antibody (Sigma, 1:1000), a goat anti-mouse antibody coupled with alkaline phosphatase (Sigma, 1:1000) was applied. Alkaline phosphatase substrate (Sigma) was added, and the optical density at 405 nm (OD405) was determined with SOFTMAX2.35 (Molecular Devices).

Gel Mobility Shift and DNase I Footprint Analysis—The coding sequence of pE2S was subcloned into a modified pMAL-c2 (New England Biolabs) bacterial expression vector and transformed into *Escherichia coli* strain XL1-Blue (Stratagene). Protein was purified using the protein fusion and purification system (New England Biolabs) with zinc buffer A, 5 mM dithiothreitol as the column buffer. Protein purity was evaluated by Coomassie Blue-stained 4–12% Novex gels. Concentration was determined by Bradford assay with bovine serum albumin standards. Purified protein was used to perform DNase I footprints and gel mobility shift assays to determine the DNA binding site and affinity.

For DNase I footprints, a DNA fragment of the human *ERBB-2* promoter was generated by PCR using 5'-³²P-labeled E2SF (5'-GGC TGC TTG AGG AAG TAT AAG AAT GAA GTT GTG AAG C-3') and pGLP2 (5'-CTT TAT GTT TTT GGC GTC TTC CA-3', Invitrogen) primers from a genomic fragment inserted into pGL3 (33). This DNA fragment contained 267 bp and included region -209 to +3 of the

³ The abbreviations used are: KRAB, Krüppel-associated box; ELISA, enzyme-linked immunosorbent assay.

Modular Zinc Finger Domains for 5'-CNN-3' Recognition

ERBB-2 promoter. The reaction buffer contained 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, 5 mM CaCl₂, 10 μ M ZnCl₂, pH 7.0. Binding reactions contained 15 kcpm ³²P-end-labeled *ERBB-2* promoter fragment and 5 mM dithiothreitol, and the protein concentration was varied from 0.1 to 100 nM. Reactions were incubated at 4 °C for 12–18 h. Digestion of DNA was performed using DNase I (Roche Diagnostics) as has been described (34). Samples were separated on a 6% acrylamide, 8 M urea gel, exposed on phosphorimaging plates, and recorded by a PhosphorImager SI (Molecular Dynamics). Analysis was performed using ImageQuant (Molecular Dynamics) and KaleidaGraph software (Synergy, Reading, PA) to give *K_d* values. Gel mobility shift analysis was performed with purified protein essentially as described (13).

Site-directed Mutagenesis of Finger-2—Finger-2 mutants were constructed by PCR as described (13, 14). The library clone containing 5'-TGG-3' finger-2 and 5'-GAT-3' finger-3 was used as the PCR template. PCR products containing a mutagenized finger-2 and 5'-GAT-3' finger-3 were subcloned via NsiI and SpeI restriction sites in-frame with finger-1 of C7 into a modified pMal-c2 vector (New England Biolabs).

Construction of Polydactyl Zinc Finger Proteins—Three-finger proteins were constructed by finger-2 stitchery using the SP1C framework as described (33). Six-finger proteins were assembled via compatible XmaI and AgeI restriction sites. Analysis of DNA binding properties was performed from isopropyl β -D-thiogalactoside-induced freeze/thaw bacterial extracts. For the analysis of capability of these proteins to regulate gene expression they were fused to the activation domain VP64 (the tetrameric repeat of herpes simplex virus VP16 minimal activation domain) or repression domain KRAB of Kox-1 as described earlier (33, 35) and subcloned into the retroviral pMX-IRES vector (35, 36) (IRES, internal ribosome-entry site; GFP, green fluorescent protein). The amino acid sequences for the resulting transcription factors KRAB-E2S and E2S-VP64 have been submitted to GenBankTM with accession numbers DQ160159 and DQ160160, respectively.

Retroviral Gene Targeting and Flow Cytometric Analysis—These assays were performed essentially as previously described (35). For production of recombinant retrovirus, 3.5×10^6 293GagPol cells were cotransfected with 3.75 μ g of pMX-IRES encoding each of the zinc finger proteins fused to a regulatory domain and 1.25 μ g of pMDG-VSV plasmid using Lipofectamine PLUS reagent (Invitrogen). Viral supernatant was collected 2 days post-transfection and used to transduce 1×10^5 A431 cells. Two days post-transduction A431 cells were stained with an ERBB2-specific antibody (5 μ g/ml) and analyzed by flow cytometry. As primary antibody, the ERBB2-specific monoclonal antibody FSP77 (a gift from Nancy E. Hynes (37)) was used. Phycoerythrin-labeled donkey F(ab')₂ anti-mouse IgG was used as secondary antibody (Jackson ImmunoResearch).

Computer Modeling—Computer models were generated using InsightII (Molecular Simulations, Inc.). Models were based on the coordinates of the co-crystal structures of Zif268-DNA (PDB accession 1AAY). The structures were not energy-minimized and are presented only to suggest possible interactions. Hydrogen bonds were considered plausible when the distance between the heavy atoms was 3 (\pm 0.3) Å and the angle formed by the heavy atoms and hydrogen was 120° or greater.

RESULTS

Library Construction and Selection—Previously we reported the isolation by phage display selection and characterization of 16 zinc finger domains that each specifically recognize a 5'-GNN-3' type DNA sequence (13, 14). The library used for the selection was based on C7, a high affinity variant of the mouse transcription factor Zif268 (29). Fin-

ger-1 of Zif268 has the amino acid sequence RSD-E-LTR (positions –1 through 6), and C7 has the sequence KSA-D-LKR. Finger-2 of both proteins has the sequence RSD-H-LTT (positions –1 through 6), and finger-3 has the sequence RAD-E-RKR (positions –1 through 6). Both proteins bind to the DNA target site 5'-GCG-TGG-GCG-3'. Attempts to select zinc finger domains from this library that bound to 5'-A/CNN-3' subsites met with no success due to the dominant cross-subsite interaction of an Asp in position 2 of the recognition helix with either adenine or cytosine on the opposite strand of the following 3-bp subsite (11, 12). This interdomain contact has been referred to as target site overlap (13, 26–28).

To select for zinc finger domains that bound to 5'-CNN-3' subsites, we constructed a phage display library by finger-swapping as reported earlier (15). Finger-3 of C7 (RSD-E-RKR, positions –1 through 6) with specificity for the subsite 5'-GCG-3' was exchanged with a domain previously characterized to bind the 5'-GAT-3' subsite (13), generating the 3-finger protein C7.GAT. This recognition helix (TSG-N-LVR) did not contain Asp at position 2, allowing the selection of zinc finger domains that bound to 5'-ANN-3' or 5'-CNN-3' DNA subsites. Because we had previously argued that recognition of the two target sites, 5'-ACG-3' and 5'-ACT-3', may require aromatic amino acid residues, an additional phage display library was constructed using a NNK codon doping strategy (N, adenine, cytosine, guanine, or thymine; K, guanine or thymine). This library included aromatic amino acid residues but no stop codons. Randomization involved positions –1, 1, 2, 3, 5, and 6 of the α -helix of finger-2 with 32 possibilities for each amino acid position. The library contained 2.4×10^9 members, ensuring representation of almost all amino acid combinations.

Selection of zinc finger-displaying phage was performed using biotinylated hairpin oligonucleotides containing the desired 9-bp binding site. Usually 6 rounds of panning with each of the 16 5'-GAT-CNN-GCG-3' target oligonucleotides were carried out in the presence of non-biotinylated competitor DNA. Stringency of the selection was increased in each round by decreasing the amount of biotinylated target oligonucleotide and increasing the amounts of the competitor oligonucleotide mixtures. In the 6th round the target concentration was typically 18 nM. The competitor mixtures for 5'-ANN-3', 5'-GNN-3', and 5'-TNN-3' finger-2 subsites were in 5-fold excess for each oligonucleotide pool, and the specific 5'-CNN-3' mixture (excluding the target sequence) was present in 10-fold excess of the target. Several specific competitors were added in up to a 20-fold excess depending on the target site. The competitor 5'-TGG-3' was used to reduce selection of the wild type protein. The competitors 5'-CAG-3', 5'-CGG-3', and 5'-CGA-3' were added to reduce selection of proteins binding nonspecifically to all 5'-CNN-3' target sites (see "Materials and Methods"). Phage that bound to the biotinylated target oligonucleotide was recovered by capture with streptavidin-coated magnetic beads.

Clones were sequenced directly after the sixth round of selection or after subcloning into a modified *E. coli* expression vector pMal-c2. The amino acid sequences of selected finger-2 helices were determined by sequence analysis (Fig. 1). Proteins were selected that bound with reasonable affinity and specificity to 8 of the 16 5'-CNN-3' target sites (Figs. 1 and 2). Attempts to select zinc fingers for the other seven subsites by panning using the C7.GAT NNK library or by addition of specific competitors were not successful.

The amino acids composing the helical recognition domain of finger-2 selected by phage display for the 5'-CNN-3' target sites are shown in Fig. 1. Those amino acid sequences with good specificity are indicated by boxes in Fig. 1. More than 50% of selected helices contained a His at position 3. The helices selected for target sites 5'-CNA-3' and

CAA	-1 1 2 3 4 5 6 Q R H H L V E 4 S P G H L I E 1 Q L G H L I Q 1 N T K Q L R N 1 N T T H L E H 1*	CAC	-1 1 2 3 4 5 6 Q R H H L V E 2 S P G H L I E 2 Q L H H L R D 1 N C N H L A E 1 G P A G L I R 2	CAG	-1 1 2 3 4 5 6 Q R H H L V E 1 R L N N L L V 1 D R K V L A T 1 H V I D L D H 1 R A D N L A I 2 N T T H L E H 1* N V N P L P P 1*	CAT	-1 1 2 3 4 5 6 Q R H H L V E 3 S P G H L I E 1 N H G H L T E 1 G R L I L L G 1 A M G H L V E 1 Q W H H L T D 1# S C V H L A E 1# T C G H L V E 1# S C Y K L R E 1# S C G G L T A 2*
CCA	-1 1 2 3 4 5 6 Q R H H L V E 2 S P G H L I E 2 S C G H L T E 1# S C S H L A E 3# S C S H L A E 2* T H R D L R Q 1*	CCC	-1 1 2 3 4 5 6 S P G H L I E 1 S V R N L R E 2 S K K H L A E 1 S K K H L A E 2*	CCG	-1 1 2 3 4 5 6 Q R H H L V E 1 S P G H L I E 1 N H T R L A S 1 R N D T L Q A 4 R S D H L T D 1*	CCT	-1 1 2 3 4 5 6 Q R H H L V E 3 S P G H L V E 1 Q C G H L V A 1# G C S K L R E 1# S C G H L Q E 2# S C G H L Q E 1* C C A H L S E 1*
CGA	-1 1 2 3 4 5 6 Q L A H L K E 3 R S D H L T N 1*	CGC	-1 1 2 3 4 5 6 Q R H H L V E 3 S P G H L V E 1 G L K W L V V 1# D L R W L V V 1# N C F H L T E 1 H T G H L L E 1# H T G H L L E 1* N C F H L T E 1# N V R T L D T 1*	CGG	-1 1 2 3 4 5 6 R S D H L T T 2 I P A D L K Q 1 R R R G W R K 1 E T R T L T A 1 R S D H L T E 1*	CGT	-1 1 2 3 4 5 6 D H Q D L T I 1 D L R S C R S 2 Q R R H L L S 1 L R R Q L A H 1* R M R N L Q K 1* S R R T C R A 1* Q L R H L R E 1*
CTA	-1 1 2 3 4 5 6 Q R H S L T E 4	CTC	-1 1 2 3 4 5 6 Q R H H L V E 4 Q R H H L V E 4#	CTG	-1 1 2 3 4 5 6 H I I D L D H 2 R N D A L T E 1 R N D A L T E 1* T R K E L R S 1*	CTT	-1 1 2 3 4 5 6 Q R H H L V E 3 Q R H H L V E 4# S P G H L I E 1 N L Q H L G E 1*

FIGURE 1. Amino acid sequences of finger-2 recognition helices from selected clones. For each target site several clones were usually sequenced after the sixth round of panning. The amino acid sequence for each clone was determined. The finger-2 subsites are shown on the left side, and the position within the recognition helix is shown above each set. Amino acids likely to make direct contact with DNA are indicated by *bold type*. Boxed sequences represent clones that bound with the best affinity and specificity profiles. DNA binding specificity of these three-finger proteins is shown in Fig. 2, upper panel. Numbers at the right of each helix determined represent the number of occurrences of the particular sequence. The asterisk (*) indicates clones identified after subcloning into a modified pMal-c2 expression vector. # indicates clones selected from the NNK library.

5'-CNG-3' generally showed good conservation in position -1, consistent with previously observed amino acid residues in these positions (13–15). Position -1 was Gln when the 3' nucleotide was adenine (5'-CGA-3' and 5'-CTA-3'). For 5'-CAA-3', Gln, Asn, or Ser were preferred in position -1, whereas for 5'-CCA-3' Ser was selected. The interaction of Ser at position -1 with a 3' adenine had previously been observed for the domain binding 5'-ACA-3' (SPA-D-LTN (15)). Panning against finger-2 target sites containing a 3' guanine strongly selected an Arg at position -1, but Asn, Gln, His, Ser, Thr, and Ile were also observed. The domains binding to 5'-CNG-3' subsites often contained Asp at position 2; Asp likely stabilizes the interaction of the 3-finger protein by contacting the cytosine base paired to the 5' guanine of finger-1. For the target sites 5'-CNT-3' Arg, Asn, Gln, His, Ser, Thr, Ala, and Cys were found in position -1 of the recognition helix. For finger-2 subsites containing a 3' cytosine, Gln, Asn, Ser, Gly, His, or Asp were selected in position -1.

For the recognition of 5'-CAN-3', His, Asn, Gly, Val, Pro, Ile, and Lys were selected in position 3 of the recognition helix. Previously, Thr or Asp was found in position 3 of the helix that recognized 5'-GCN-3' (13, 15). Here, we selected His, Lys, Arg, and Asn as well as Asp and Thr. For the target site 5'-CCC-3', position 3 was Asn or His. For 5'-CCG-3', position 3 was either Thr or His (Fig. 1). For 5'-GGN-3', His was

strongly selected at position 3, consistent with previously observed results (13–15). For the target sites 5'-CGG-3' and 5'-CGT-3', Ser, Asp, Thr, Asn, Gln, and Gly were selected at position 3. For 5'-CGC-3' Trp and Thr were selected in some cases. For helices binding to 5'-CTN-3', position 3 was either Ser or Ala. The exceptions were 5'-CTC-3' and 5'-CTT-3', where His was selected. Positions 1, 2, and 5 of the α -recognition helix were variable. This is not surprising because these residues are usually not involved in direct base contacts with DNA (11, 12). Position 4 was not randomized, but a spontaneous mutation generated a helix with a change from Leu to Cys at position 4 that bound to 5'-CGT-3' (Fig. 1). This type of spontaneous mutation was observed in rare instances in selection for proteins that bound to 5'-ATC-3' and 5'-GCC-3' target sites (13, 15).

Little was known about the recognition of a 5' cytosine by Cys₂-His₂ zinc finger domains. The recognition of a 5' guanine has been well characterized and is achieved by either Arg or Lys in position 6 of the helix (12, 13, 19–24). Selection against 5'-ANN-3' subsites and refinement by site-directed mutagenesis also suggested that recognition can be achieved by the amino acid in position 6; for recognition of a 5' adenine, the amino acid can be Asn, Ala, Val, Asp, Arg, or Glu (15). By analogy, one could assume that the recognition of 5' cytosine is accomplished by the amino acid residue in position 6 of the α -recognition

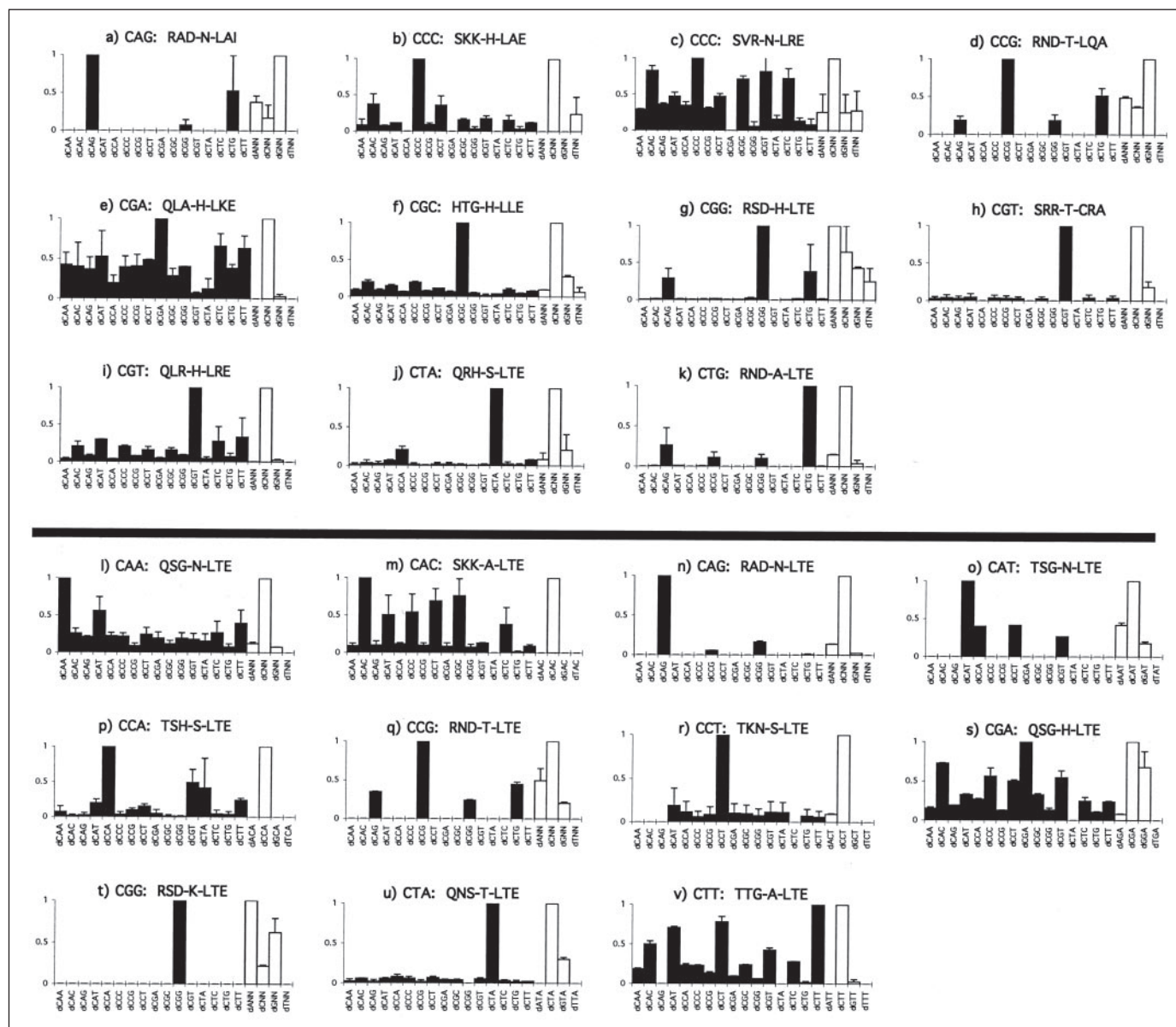


FIGURE 2. **Multitarget specificity assay to study DNA binding properties of selected domains and domains derived by site-directed mutagenesis.** At the *top of each graph* is the oligonucleotide sequence recognized (pNNTN) and the amino acid sequence of the finger-2 domain (positions -1 to 6 with respect to the helix start) of the 3-finger protein analyzed. *Black bars* represent binding to target oligonucleotides with different finger-2 subsites: 5'-CAA-3', 5'-CAC-3', 5'-CAG-3', 5'-CAT-3', 5'-CCA-3', 5'-CCC-3', 5'-CGC-3', 5'-CCT-3', 5'-CGA-3', 5'-CGC-3', 5'-CGT-3', 5'-CTA-3', 5'-CTC-3', 5'-CTG-3', and 5'-CTT-3'. *White bars* represent binding to a mixture of the 16 oligonucleotides of each of the subsites 5'-ANN-3', 5'-CNN-3', 5'-GNN-3', and 5'-TNN-3' to evaluate the 5' recognition. The height of each *bar* represents the relative affinity of the protein for each target, averaged over two independent experiments and normalized to the highest signal among the black or white bars. *Error bars* represent the deviation from the average. *Upper panel*, proteins analyzed correspond to the *boxed helix* sequences from Fig. 1 obtained from panning. *Lower panel*, proteins analyzed were derived by site-directed mutagenesis. Proteins were generally cloned and analyzed in the Zif268 or Sp1C framework. Proteins constructed utilizing the SP1C framework generally bind their targets with better specificity as compared with those made using the Zif268 framework.

helix. Phage display selection of domains binding the 5'-CNN-3' finger-2 subsites resulted in the selection of Glu, Asn, Ile, Asp, Ala, Ser, and Val at position 6. Strikingly, Glu was present at position 6 in 65% (82 of 127) of the sequenced proteins (Fig. 1).

DNA Binding Specificity of Selected Domains Recognizing 5'-CNN-3' Target Sites—To evaluate the DNA binding specificity of the zinc finger domains selected by phage display, each was tested using multitarget specificity ELISA against all 16 sites of the type 5'-GAT-CNN-GCG-3' (Fig. 2, black bars). In addition, the 5' specificity was analyzed by evaluating the binding to an equimolar mixture of the 16 oligonucleotides for each of the 5'-ANN-3', 5'-CNN-3', 5'-GNN-3', and 5'-TNN-3' target sites (Fig. 2, white bars). Results for clones derived from panning

are summarized in Fig. 2, *upper panel*. In general, the selected helices showed exquisite DNA binding specificity to a 5' cytosine when position 6 of the recognition helix was Glu. The only exception was for target site 5'-CGG-3' (protein sequence RSD-H-LTE), where binding was tighter to the oligonucleotides with a 5' adenine (Fig. 2g). Clones containing Ile (RAD-H-LAI; Fig. 2a) or Asp (RSD-H-LTD; not shown) at position 6, selected for binding to 5'-CAG-3' and 5'-CCG-3', respectively, preferentially bound a 5' guanine. Two of the helices had Ala at position 6. The zinc finger with sequence RND-T-LQA, selected for binding to 5'-CCG-3', showed preferential binding to a 5' guanine (Fig. 2d), whereas the protein with sequence SRR-T-CRA, selected for binding to 5'-CGT-3', bound with high specificity to a 5' cytosine (Fig. 2h).

Previously, an Ala at position 6 had been found to mediate recognition of a 5' adenine (15). These results may indicate that recognition of the 5' nucleotide of the triplet is not only dependent on the amino acid residue in position 6 of the α -helix but also on the neighboring amino acids, possibly through coordinated interactions.

Recognition of a middle adenine (5'-CAN-3') was observed when the helix contained Asn at position 3, as in binding of 5'-CAG-3' by RAD-N-LAI (Fig. 2a), consistent with previously reported results (13–15). A middle cytosine (5'-CCN-3') was specifically recognized when position 3 of the helix was His, as in the case of SKK-H-LAE (Fig. 2b). Asn was also fairly specific for recognition of a middle cytosine (SVR-N-LRE, Fig. 2c). Also consistent with previous observations was recognition of 5'-CCG-3' by RND-T-LQA (Fig. 2d) with Thr at position 3 (13–15). A middle guanine, 5'-CGN-3', was recognized by a His at position 3 as previously reported (13–15). Although HTG-H-LLE (Fig. 2f) showed excellent DNA binding specificity for 5'-CGC-3', other proteins (QLA-H-LKE, Fig. 2e; RSD-H-LTE, Fig. 2g; QLR-H-LRE, Fig. 2i) that recognized 5'-CGN-3' have some cross-reactivity to other sites. QLA-H-LKE (Fig. 2e) is representative of a low affinity and specificity clone. The helix SRR-T-CRA (Fig. 2h) exhibited specific binding to a middle guanine in 5'-CGT-3' despite the Thr at position 3. A middle thymine was recognized specifically by the helices QRH-S-LTE (Fig. 2j) and RND-A-LTE (Fig. 2k) containing either Ser or Ala at position 3, consistent with previous findings (13–15).

Recognition of a 3' adenine was observed for the helices containing Gln at position -1 (QLA-H-LKE, Fig. 2e; QRH-S-LTE, Fig. 2j) as reported earlier (13–15). For the recognition of a 3' cytosine, Ser and His were observed at position -1 (HTG-H-LLE, Fig. 2f; SKK-H-LAE, Fig. 2b; SVR-N-LRE, Fig. 2c). Recognition of 3' guanine was achieved by an Arg at position -1 (Fig. 2, a, d, g, and k). These data are consistent with previous reports (13–15). Recognition of a 3' thymine was reported to be mediated by Ser, Thr, or His in position -1 (13–15). In this study helices that recognize 3' thymine had Ser (SRR-T-CRA, Fig. 2h) or Gln (QLR-H-LRE, Fig. 2i) in position -1.

In summary, phage display selection yielded domains for eight 5'-CNN-3' target sites, including 5'-CAG-3', 5'-CCC-3', 5'-CCG-3', 5'-CGC-3', 5'-CGG-3', 5'-CGT-3', 5'-CTA-3', and 5'-CTG3'. For the other target sites no domains with reasonable specificity resulted from panning. For 5'-CAG-3' (Fig. 2a) and 5'-CCG-3' (Fig. 2d), improvement of specificity in the binding of the 5' cytosine was necessary. The helix binding to 5'-CGA-3' (Fig. 2e) showed little specificity and required improvement by site-directed mutagenesis. For domains binding the 5'-CNG-3' type of DNA sequences, specificity for the middle nucleotide was usually insufficient. We have previously described this phenomenon in our studies of domains binding to 5'-GNG-3' and 5'-ANG-3' DNA sequences (13–15).

Improvement of the Specificity of Domains Binding to the 5'-CNN-3' Family of Sequences by Site-directed Mutagenesis and de Novo Design—Phage display selections did not generate zinc finger domains that bound specifically to 5'-CAA-3', 5'-CAC-3', 5'-CAT-3', 5'-CCA-3', 5'-CCT-3', 5'-CTC-3', 5'-CTT-3', or 5'-CGA-3'. In some cases selected helices bound their cognate target site but with some cross-reactivity to other sites (for example, 5'-CAG-3', Fig. 2a; 5'-CCG-3', Fig. 2d; 5'-CGA-3', Fig. 2e; 5'-CGG-3', Fig. 2g). Site-directed mutagenesis was used to improve DNA binding specificity. Results are shown in Fig. 2, lower panel. The DNA binding motif Leu-Thr-Glu in positions 4, 5, and 6 of the α -helix was found in numerous sequences selected during panning to mediate 5' cytosine recognition. The helix selected to recognize 5'-CAG-3' (RAD-N-LAI; Fig. 2a) was changed to this consensus sequence (RAD-N-LTE). DNA binding specificity was greatly improved (Fig. 2n) not only for 5' cytosine but also for the middle adenine. Use of the Leu-Thr-Glu

motif also improved binding to 5'-CCG-3' (compare RND-T-LQA, Fig. 2d, to RND-T-LTE, Fig. 2g). In an attempt to improve 5' cytosine recognition in the target site 5'-CGG-3', His at position 3 was replaced with a Lys (RSD-H-LTE, Fig. 2g versus RSD-K-LTE, Fig. 2t). This did not improve 5' cytosine recognition but did result in exclusive recognition of the middle guanine. The amino acid sequence selected for recognition of 5'-CGA-3' was changed from QLA-H-LKE (Fig. 2e) to QSG-H-LTE based on a sequence from Segal *et al.* (13) (QSG-D-LRR). This improved DNA binding specificity and affinity, but the net specificity remained lower than most other domains (Fig. 2s).

Rational design was also applied to generate a domain for recognition of 5'-CAA-3'. The finger-2 helix QSG-N-LTE (derived from a zinc finger that recognized 5'-GCA-3', QSG-D-LRR (13)) bound its target site with good specificity (Fig. 2l). The helix SKK-A-LTE bound preferentially to its target site 5'-CAC-3' with excellent 5' cytosine recognition (Fig. 2m) but weak middle base specificity. For the recognition of 5'-CAT-3', the helix TSG-N-LTE was generated based on a helix that recognized 5'-GAT-3' (TSG-N-LVR) (13). Multitarget ELISA showed that this helix bound preferentially to its target site with excellent recognition of the 5' cytosine (Fig. 2o). Changing the 5'-CTA-3' helix of QRH-S-LTE (Fig. 2j) to QNS-T-LTE (Fig. 2u) reduced the nonspecific binding to other targets. With the exception of 5'-CCC-3' (Fig. 2b), specific helices that targeted sites containing only pyrimidine nucleotides were not generated from phage display. The helix TSH-S-LTE was designed to bind the subsite 5'-CTC-3' but, surprisingly, bound preferentially to 5'-CCA-3' with excellent 5' cytosine recognition (Fig. 2p). These data then allowed us to design a highly specific recognition domain for 5'-CCT-3', TKN-S-LTE (Fig. 2r), and the helix TTG-A-LTE for 5'-CTT-3' (Fig. 2v). For the subsite 5'-CTC-3' no specific binder could be identified after testing of multiple helices.

Generation of Polydactyl Zinc Finger Proteins Containing 5'-CNN-3' Domains—We have previously demonstrated that exogenous and endogenous genes can be regulated with six-finger proteins containing zinc finger domains specifically recognizing 5'-(GNN)₆-3' DNA sequences (33, 35, 36). In addition, we showed that 6-finger proteins containing varying numbers of domains recognizing 5'-GNN-3', 5'-ANN-3', and 5'-TNN-3' target sites were capable of endogenous gene regulation (15). To investigate whether the domains described here that recognize the 5'-CNN-3' family of DNA sequences are suitable for the construction of artificial transcription factors, the 6-finger protein pE2S was constructed. We chose the human *ERBB-2* gene as our model system since we have previously reported specific gene regulation with the six-finger proteins E2C and E2X targeted against the 5' untranslated region of the *ERBB-2* gene (15, 35). The 6-finger pE2S was designed to bind the target site 5'-CGG-GGG-GCT-CCC-CTG-GTT-3' at position -137 to -154 within the 5'-untranslated region (38). This target site contains recognition sites for three 5'-GNN-3' domains previously identified (13) and three 5'-CNN-3' domains described here (indicated by underlining).

Two three finger-coding regions were generated using a rapid PCR overlap extension method and the Sp1C framework (33). These three-finger proteins were then fused to create a six-finger protein that was cloned into the bacterial expression vector pMal-c2. This six-finger protein was expressed in *E. coli* as a maltose-binding fusion protein and was purified. The affinity of purified pE2S was measured using an electrophoretic mobility shift assay, and a dissociation constant of 3.25 nM was determined. Characterization of the binding specificity of the six finger protein E2S for its DNA target site within a 267-bp fragment of the *ERBB-2* promoter containing the pE2S binding site was determined using DNase I footprinting (Fig. 3). E2S protein was titrated over a range of 100 to 0.1 nM to provide a dissociation constant. The average of three

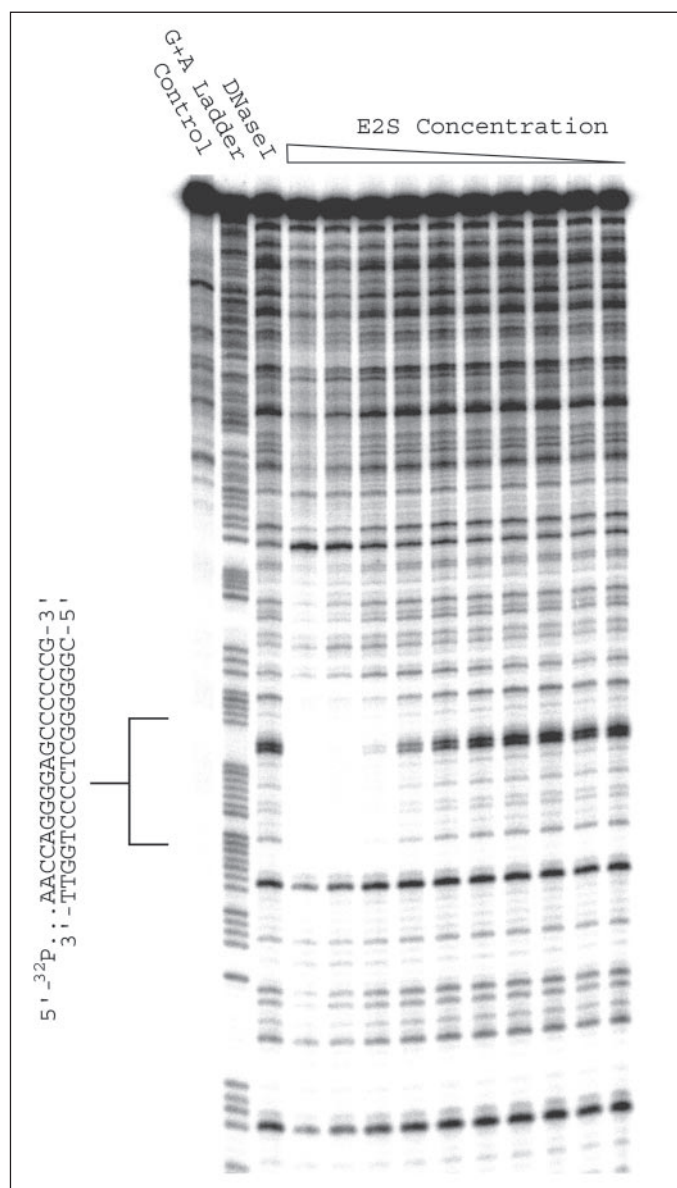


FIGURE 3. DNase I footprint of the six-finger protein pE2S containing domains recognizing 5'-CNN-3' DNA sequences. The six-finger protein pE2S was constructed using the SP1C framework. After bacterial expression and purification, pE2S was incubated with a 267-bp radioactively labeled DNA fragment containing region -209 to +3 of the *ERBB-2* promoter. After DNase I digestion, samples were separated on a 6% acrylamide, 8 M urea gel. Control, DNA fragment without digestion; G+A ladder, sequencing ladder of the 267-bp DNA fragment; DNase I, digestion of the labeled fragment without E2S protein; lanes 4–13 contained 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, and 0.1 nM E2S protein. The location of the E2S binding site is indicated.

independent experiments produced a K_d value of 14 ± 4 nM, consistent with affinity data derived from electrophoretic mobility shift assay. DNA binding of pE2S was observed precisely at the 18-bp target site 5'-CGG-GGG-GCT-CCC-CTG-GTT-3'. This result shows that the 5'-CNN-3' domains characterized here also promote highly specific binding in the context of a six-finger protein.

To investigate the potential of the six-finger protein pE2S to alter expression of the endogenous *ERBB-2* gene, their coding regions were subcloned into the retroviral vector pMX-IRES and fused to either the VP64 activation domain or the KRAB repression domain of Kox-1 (33, 39). The human carcinoma cell line A431 was transduced with the recombinant retrovirus. Two days after transduction, cells were analyzed by flow cytometry using an ErbB-2-specific antibody (Fig. 4).

Strong up-regulation of the endogenous *ERBB-2* gene was observed with pE2S fused to the VP64 activation domain (Fig. 4A), and down-regulation was observed when pE2S fused to the KRAB repression domain was expressed (Fig. 4B). The result for pE2S clearly demonstrates that the 5'-CNN-3' domains described here can serve as modular building blocks for the construction of designed transcriptional regulators.

DISCUSSION

Zinc finger domains of the type Cys₂-His₂ have proven to be versatile tools for the targeted regulation of endogenous genes and more recently as nuclease targeting domains to facilitate gene repair through homologous recombination. Recognition of a 3-bp DNA sequence by a zinc finger is predominantly achieved by the interaction of amino acid residues of the α -helix in positions -1, 3, and 6 with the 3', middle, and 5' nucleotides, respectively, whereas amino acids at other positions within the helix are key to providing elements of fine specificity. Position 2 of the α -helix interacts with other helix residues and, in addition, can make contact with a nucleotide outside the 3-bp subsite (11, 12, 40). This target site overlap is a limitation of the modular approach to generation of transcription factors, but this influence seems to be restricted only to interdomain interactions if position 2 of the α -recognition helix is Asp (40, 41).

In the present study we describe the generation of zinc finger domains binding 5'-CNN-3' subsites by selection from a phage display library based on the three finger protein C7.GAT and refinement and or creation through rational design. This C7.GAT library was constructed with a finger-3 that did not contain an Asp at position 2 to enable the selection of zinc finger domains recognizing finger-2 subsites containing a 5' adenine or 5' cytosine. Phage display selections from this library for domains binding to the 5'-CNN-3' type of DNA sequences was not as successful as for selection of 5'-ANN-3' domains (15). Selections for target sites containing a 5' pyrimidine (cytosine or thymine) may be more difficult since pyrimidines are not as accessible to the side chains of the recognition helix and offer fewer opportunities for hydrogen bonding.

Zinc finger domains that specifically recognized eight of the 16 possible 5'-CNN-3' target sites were selected from phage display (Fig. 1). These included five domains with excellent DNA binding specificity with target sites 5'-CCC-3', 5'-CGC-3', 5'-CGT-3', 5'-CTA-3', and 5'-CTG-3' (Fig. 2, *b, f, h, j*, and *k*) and three domains with good specificity that bound 5'-CAG-3', 5'-CCG-3', and 5'-CGG-3' (Fig. 2, *a, d*, and *g*). DNA binding specificity for target sites 5'-CAG-3', 5'-CCG-3', 5'-CGA-3', 5'-CGG-3', and 5'-CTA-3' were improved by applying rational design to amino acid sequences obtained from phage display (Fig. 2, *n, q, s, t*, and *u*). For the target sites 5'-CAA-3', 5'-CAC-3', 5'-CAT-3', 5'-CCA-3', 5'-CCT-3', and 5'-CTT-3', zinc finger domains with reasonable to excellent DNA binding specificity were obtained through *de novo* design (Fig. 2, *l, m, o, p, r*, and *v*). Despite extensive analysis, we could not identify a zinc finger domain to recognize 5'-CTC-3' with specificity. The most optimal zinc finger domains and their recognition sites are summarized in TABLE ONE.

Nonspecific zinc fingers were also noted in our selections. The same three helices, QRH-H-LVE, SPG-H-LIE, and SCS-H-LAE, were obtained in selections against a number of the target sites: 5'-CAA-3', 5'-CAC-3', 5'-CAT-3', 5'-CCA-3', 5'-CCT-3', 5'-CTC-3', and 5'-CTT-3' (Fig. 1). For these helices no specific target sites were identified. The three helices appear to bind DNA with high affinity but low specificity.

In general, the recognition of the middle and 3' nucleotide of the

FIGURE 4. Retrovirus-mediated gene targeting to study the potential of E2S to regulate the endogenous *ERBB-2* gene. A431 cells were transduced with retrovirus encoding transcription factors consisting of either pE2S fused to the activation domain VP64 or pE2S fused to the repressor domain KRAB. Three days later intact cells were stained with ErbB2-specific monoclonal antibody FSP77 in combination with phycoerythrin-labeled secondary antibody and analyzed by flow cytometry. *Thin line*, staining with secondary antibody alone; *stippled line*, specific staining of mock-transduced cells; *solid lines*, specific staining of cells transduced to express the transcription factors E2S-VP64 (A) or KRAB-E2S (B).

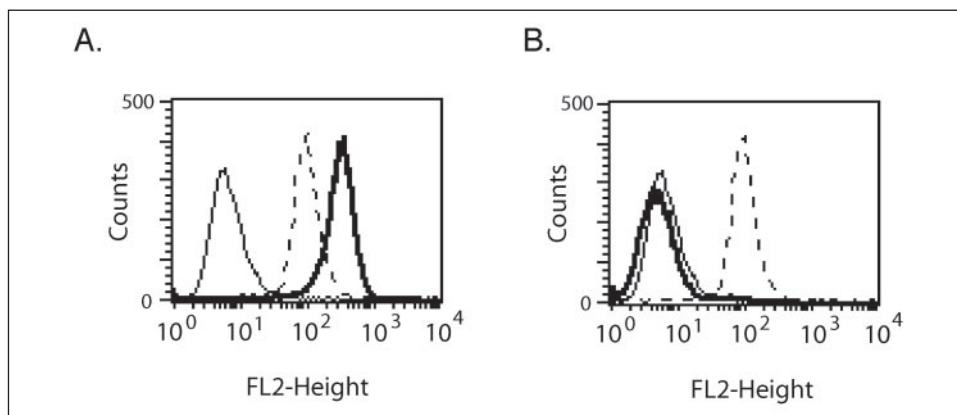


TABLE ONE

Summary of the most optimal zinc finger domain sequences identified in this study for binding to 5'-CNN-3' subsites

The helices that bound to the 5'-CNN-3' target sites are listed. P, helices derived from panning; M, helices derived by site-directed mutagenesis or *de novo* design.

Target site 5' → 3'	Finger-2 helix	Source	Fig. 2 panel
CAA	QSG-N-LTE	M	l
CAC	SKK-A-LTE	M	m
CAG	RAD-N-LTE	M	n
CAT	TSG-N-LTE	M	o
CCA	TSH-S-LTE	M	p
CCC	SKK-H-LAE	P	b
CCG	RND-T-LTE	M	q
CCT	TKN-S-LTE	M	r
CGA	QSG-H-LTE	M	s
CGC	HTG-H-LTE	P	f
CGG	RSD-K-LTE	M	t
CGT	SRR-T-CRA	P	h
CTA	QNS-T-LTE	M	u
CTC	None		
CTG	RND-A-LTE	P	k
CTT	TTG-A-LTE	M	v

5'-CNN-3' subsite by zinc finger domains was consistent with previous observations (13–15). Middle adenines (5'-CAN-3') are predominantly recognized by Asn at position 3, with the exception of 5'-CAC-3' (SKK-A-LTE, Fig. 2m). Middle guanine was recognized by the His at position 3, with the exception of 5'-CGT-3' (SRR-T-CRA, Fig. 2h), and middle thymine was recognized by either Ser or Ala at position 3. Interestingly, the recognition of middle cytosine was not achieved by Thr, Asp, or Glu at position 3 as previously reported but by Ser or His (the exception was 5'-CCG-3', RND-T-LTE, Fig. 2q). The recognition of the 3' nucleotide of the target was mediated by the amino acid at position –1; 3' adenine was recognized by Gln (the exception was 5'-CCA-3', TSH-S-LTE, Fig. 2p), 3' guanine was recognized by Arg, and 3' thymine was recognized by Thr or Ser. Unusual amino acid residues were found in position –1 for domains that recognized a 3' cytosine, *e.g.* Asn, Ser, Thr, and His. Previously only Asp and Glu were found to mediate specific recognition of a 3' cytosine (13–15).

Many of the selected zinc finger domains had Glu in position 6 of the α -recognition helix (Fig. 1). Although it was tempting to assume that this Glu mediated recognition of the 5' cytosine, analysis of the DNA binding specificity of the selected helices showed that recognition of the cytosine was more complex and dependent on other positions with the

helix. Excellent discrimination of the 5' cytosine was observed for 8 domains with Glu in position 6 for (Fig. 2, b, c, e, f, h, i, j, and k), but RSD-H-LTE (Fig. 2g) was an exception. The most frequent motif selected at positions 5 and 6 was TE. When this motif was used for the rational design of zinc finger domains, domains showed excellent discrimination of a 5' cytosine (Fig. 2, l–v), with the exceptions of QSG-N-LTE and RSD-K-LTE (Fig. 2, s and t).

Molecular modeling was performed to gain insights into how the changes in the zinc finger recognition helix produced the observed DNA binding specificity. Of primary interest was how Glu at position 6 mediated recognition of 5' cytosine. The helix RSD-H-LTE, which recognized 5'-CGG-3' (Fig. 2g), was modeled (Fig. 5B) taking advantage of sequence similarity with finger-2 of the well characterized Zif268 protein, RSD-H-LTT (Fig. 5A). The structure of Zif268 bound to its operator DNA (11) revealed several amino acid/base interactions relevant to the current study. The finger-2 subsite in the Zif268 structure is 5'-TGG-3'. Recognition of 3' guanine was accomplished by two H-bonds from Arg at position –1. The Arg side chain was conformationally constrained by two buttressing H-bonds to the Asp in position 2 (Fig. 5A). Recognition of the middle guanine was accomplished by an H-bond from His at position 3. This His stacked against the 5' thymine, an interaction thought to limit the conformational flexibility of the His and, thus, enhance its specificity for middle guanine. The Thr at position 6 made no hydrogen bonds with the 5' thymine. The specificity for 5' thymine was achieved through a stacking interaction between the His at position 3 and the thymine and a hydrogen bond between the adenine on the opposite strand and an Asp at position 2 in neighboring finger-3 recognition helix (not shown in Fig. 5A). This target site overlap was discussed above.

It seems reasonable *a priori* to expect that the finger-2 helix RSD-H-LTE will bind 5'-CGG-3' in much the same way as the Zif268 finger-2 helix RSD-H-LTT binds 5'-TGG-3'. Substituting the appropriate amino acids and bases onto the coordinates of the Zif268 produced the model for RSD-H-LTE shown in Fig. 5B. This model places the ϵ oxygens of Glu at position 6 within H-bonding distance ($3 \pm 0.5 \text{ \AA}$) of N4 on the 5' cytosine, providing a reasonable explanation for specificity of Glu for this base. However, the multitarget ELISA data demonstrated that RSD-H-LTE actually bound preferentially to a 5' adenine rather than a 5' cytosine (Fig. 2g). A possible explanation for the poor specificity of 5' cytosine by RSD-H-LTE could be a stacking interaction between His at position 3 and the 5' cytosine, similar to the His-5' thymine interaction observed in finger-2 of Zif268. Such an interaction might position the 5' cytosine in a manner that does not favor a hydrogen bond with the position 6 Glu. The phage selection data also supports a potential interaction between His and 5' cytosine (Fig. 1). His appeared at position 3 in

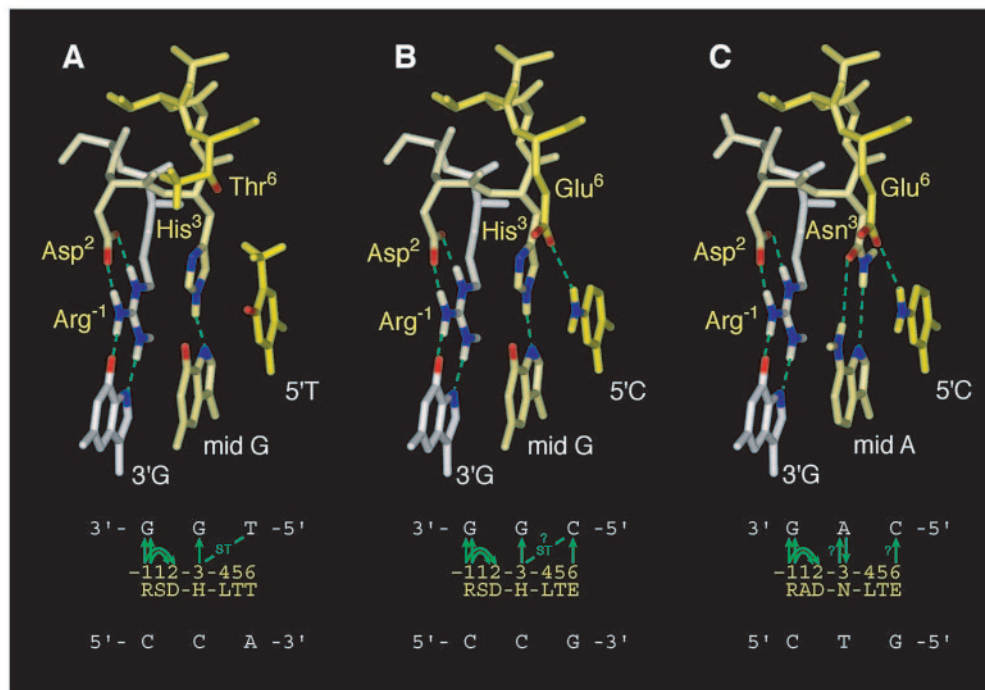


FIGURE 5. **Computer models of finger-2 domains suggesting 5'-cytosine recognition.** Select oxygen (red), nitrogen (blue), and phosphate (purple) atoms are colored for clarity. Green dotted lines indicate suggested hydrogen bonds. The sequence of each helix, the DNA subsequence, and proposed interactions are summarized below each model. Green lines indicate hydrogen bonds. Arrows indicate hydrogen acceptors. A, finger-2 of Zif268; B, pCGG (Fig. 2g); C, pmCAG (Fig. 2n).

56% of selected zinc finger domains targeting 5'-CGN-3', which might be expected due to its established role in the recognition of middle guanine. However, it also appeared in 51% of cases overall (42 of 82 sequences), including most of the domains binding to 5'CNA-3' (with the exception of 5'-CTA-3'), 5'-CNC-3', and 5'-CNT-3' (Fig. 1). Therefore, the presence of 5' cytosine in the target site might have biased the selection of His as the position 3 residue. Other positions in the recognition helix may also affect specificity for the 5' nucleotide. QSG-H-LTE displayed poor 5' specificity (Fig. 2s), but SKK-H-LAE and HTG-H-LLE both strongly specified 5'C (Fig. 2, b and f). Unfortunately, any models based on such helices would be highly speculative due to the lack of sufficient structural data of related sequences.

A second interesting case to consider is that of RAD-N-LTE, which recognizes 5'-CAG-3' (Fig. 5C). The primary difference between this helix and RSD-H-LTE is the Asn in position 3. This helix achieves excellent specificity for 5' cytosine (Fig. 2n). To investigate the basis for the 5' specificity, the appropriate amino acids and bases were substituted onto the coordinates of the structure of finger-2 of Zif268. These simple substitutions placed the δ O and δ N of the position 3 Asn well beyond hydrogen-bonding distance (≈ 4.5 Å) with N6 and N7 of the adenine. However, previous structural data as well as the excellent specificity of this domain for a middle adenine suggests that a hydrogen bond must be present. In principle, this could be accomplished either by repositioning of the adenine or by a rotation of α -recognition helix that would bring Asp closer to the DNA. Both conformational changes are observed in the structure of finger-2 of the protein Tramtrak, RKD-N-MTA, bound to 5'-AAG-3' (21). This conformational change would also be expected to bring the Gln at position 6 closer to 5' cytosine, potentially facilitating an interaction that would produce the observed excellent specificity for 5' cytosine. It should be noted that Ala at position 6 in finger-2 of the Tramtrak protein does not contact the 5' base of the target, and modeling RAD-N-LTE on the coordinates of the Tramtrak helix provided little additional insight (data not shown).

Due to a lack of existing structural data, the specificity of 5' cytosine by many of the zinc finger helices described in this study cannot easily be rationalized by computer modeling. In structures where recognition of 5' adenine, thymine, or cytosine can be rationalized, the specificity has been the result of a target site overlap (11, 41). We expect no target site overlap interactions from the finger-3 helix used in this study, as 5' cytosine recognition was variable despite the presence of a common finger-3 for all proteins in this study. It is clear that more structural studies will be required to understand the parameters involved in DNA base positioning and the docking orientation of the α -helix with respect to the DNA.

Although we do not yet have a detailed understanding of the molecular interactions that underlie the specificity of the zinc finger domains described here for their 5'-CNN-3' target sites, their use in artificial transcription factors has been demonstrated. We have constructed the 6-finger protein pE2S containing three 5'-CNN-3' recognition helices that target the 18-bp sequence 5'-CGG-GGG-GCT-CCC-CTG-GTT-3' within the 5'-untranslated region of the *ERBB-2* gene. Binding of pE2S was specific for this 18-bp recognition site as demonstrated by DNase I footprint analysis (Fig. 3). Furthermore, we have demonstrated that pE2S was capable of up- and down-regulation of the endogenous *ERBB-2* when fused to either the activation domain VP64 or the repression domain KRAB, respectively (Fig. 4). Likewise, we have recently reported the construction of the γ -globin-targeting protein gg1 incorporating the 5'-CTG-3' domain together with 5'-ANN-3' and 5'-GNN-3' domains. This protein binds the sequence 5'-GTC AAG GCA AGG CTG GCC-3' with a 0.7 nM dissociation constant (42). This protein was shown by DNase I footprinting and chromatin immunoprecipitation to bind the targeted sequence *in vitro* and *in vivo*, and transcription factors based on this protein were robust regulators of the endogenous human γ -globin gene. These two cases indicate that the domains selected here are readily combined with the variety of other domains we have reported.

In summary, the zinc finger domains described here that recognize 5'-CNN-3' DNA subsites are suitable for the rapid construction of artificial transcription factors. These 15 5'-CNN-3' zinc finger domains augment the 16 5'-GNN-3', 14 5'-ANN-3', and 2 5'-TNN-3' domains we have previously developed and together provide for the rapid construction of more than 10 billion proteins that bind 5'-[(G/A/C)NN]₆-3' sites. Sites of this type occur approximately once every 6 nucleotides in random sequence. Therefore, the predefined domains disclosed here significantly increase the number of DNA sequences that can be rapidly targeted by artificial transcription factors and nucleases.

Acknowledgments—We thank N. Hynes for providing the ErbB-2-specific antibody and R. Beerli for discussion.

REFERENCES

- Blancafort, P., Segal, D. J., and Barbas, C. F., III (2004) *Mol. Pharmacol.* **66**, 1361–1371
- Beerli, R. R., and Barbas, C. F., III (2002) *Nat. Biotechnol.* **20**, 135–141
- Jantz, D., and Berg, J. M. (2004) *Chem. Rev.* **104**, 789–799
- Zhang, L., Spratt, S. K., Liu, Q., 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
- Liu, P. Q., Rebar, E. J., Zhang, L., Liu, Q., Jamieson, A. C., Liang, Y., Qi, H., Li, P. X., Chen, B., Mendel, M. C., Zhong, X., Lee, Y. L., Eisenberg, S. P., Spratt, S. K., Case, C. C., and Wolffe, A. P. (2001) *J. Biol. Chem.* **276**, 11323–11334
- Blancafort, P., Magnenat, L., and Barbas, C. F., III (2003) *Nat. Biotechnol.* **21**, 269–274
- Bibikova, M., Beumer, K., Trautman, J. K., and Carroll, D. (2003) *Science* **300**, 764
- Porteus, M. H., and Carroll, D. (2005) *Nat. Biotechnol.* **23**, 967–973
- Miller, J., McLachlan, A. D., and Klug, A. (1985) *EMBO J.* **4**, 1609–1614
- Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., and Wright, P. E. (1989) *Science* **245**, 635–637
- Pavletich, N. P., and Pabo, C. O. (1991) *Science* **252**, 809–817
- Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure* **4**, 1171–1180
- Segal, D. J., Dreier, B., Beerli, R. R., and Barbas, C. F., III (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2758–2763
- Dreier, B., Segal, D. J., and Barbas, C. F., III (2000) *J. Mol. Biol.* **303**, 489–502
- Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., and Barbas, C. F., III (2001) *J. Biol. Chem.* **276**, 29466–29478
- Guan, X., Stege, J., Kim, M., Dahmani, Z., Fan, N., Heifetz, P., Barbas, C. F., III, and Briggs, S. P. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13296–13301
- Tan, S., Guschin, D., Davalos, A., Lee, Y. L., Snowden, A. W., Jouvenot, Y., Zhang, H. S., Howes, K., McNamara, A. R., Lai, A., Ullman, C., Reynolds, L., Moore, M., Isalan, M., Berg, L. P., Campos, B., Qi, H., Spratt, S. K., Case, C. C., Pabo, C. O., Campisi, J., and Gregory, P. D. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11997–12002
- Beerli, R. R., Schopfer, U., Dreier, B., and Barbas, C. F., III (2000) *J. Biol. Chem.* **275**, 32617–32627
- Pavletich, N. P., and Pabo, C. O. (1993) *Science* **261**, 1701–1707
- Kim, C. A., and Berg, J. M. (1996) *Nat. Struct. Biol.* **3**, 940–945
- Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Nature* **366**, 483–487
- Houbaviy, H. B., Usheva, A., Shenk, T., and Burley, S. K. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13577–13582
- Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) *J. Mol. Biol.* **273**, 183–206
- Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2938–2943
- Havranek, J. J., Duarte, C., and Baker, D. (2004) *J. Mol. Biol.* **344**, 59–70
- Wolfe, S. A., Greisman, H., Ramm, E. I., and Pabo, C. O. (1999) *J. Mol. Biol.* **285**, 1917–1934
- Isalan, M., Choo, Y., and Klug, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5617–5621
- Pabo, C. O., and Nekludova, L. (2000) *J. Mol. Biol.* **301**, 597–624
- Wu, H., Yang, W.-P., and Barbas, C. F., III (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 344–348
- Rader, C., and Barbas, C. F., III (1997) *Curr. Opin. Biotechnol.* **8**, 503–508
- Barbas, C. F., III, Kang, A. S., Lerner, R. A., and Benkovic, S. J. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 7978–7982
- Barbas, C. F., III, and Lerner, R. A. (1991) *Methods* **2**, 119–124
- Beerli, R. R., Segal, D. J., Dreier, B., and Barbas, C. F., III (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14628–14633
- Trauger, J. W., and Dervan, P. B. (2001) *Methods Enzymol.* **340**, 450–466
- Beerli, R. R., Dreier, B., and Barbas, C. F., III (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1495–1500
- Liu, Q., Segal, D. J., Ghiara, J. B., and Barbas, C. F., III (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5525–5530
- Harwerth, I. M., Wels, W., Marte, B. M., and Hynes, N. E. (1992) *J. Biol. Chem.* **267**, 15160–15167
- Hudson, L. G., Ertl, A. P., and Gill, G. N. (1990) *J. Biol. Chem.* **265**, 4389–4393
- Margolin, J. F., Friedman, J. R., Meyer, W. K.-H., Vissing, H., Thiesen, H.-J., and Rauscher, F. J., III (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4509–4513
- Segal, D. J., Beerli, R. R., Blancafort, P., Dreier, B., Efftz, K., Huber, A., Koksche, B., Lund, C. V., Magnenat, L., Valente, D., and Barbas, C. F., III (2003) *Biochemistry* **42**, 2137–2148
- Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212
- Graslund, T., Li, X., Magnenat, L., Popkov, M., and Barbas, C. F., III (2005) *J. Biol. Chem.* **280**, 3707–3714