Development of zinc finger domains for recognition of the 5’-ANN-3’ family of DNA sequences and their use in the construction of artificial transcription factors

Birgit Dreier, Roger R. Beerli, David J. Segal, Jessica D. Flippin and Carlos F. Barbas III*

The Skaggs Institute for Chemical Biology and the Department of Molecular Biology
The Scripps Research Institute, La Jolla, CA 92037

Running title: 5’-ANN-3’-specific zinc finger domains

*Corresponding author:
Dr. Carlos F. Barbas III
The Scripps Research Institute, BCC-515
North Torrey Pines Road
La Jolla, CA 92037
Phone: (858) 784-9098
Fax: (858) 784-2583
E-mail: carlos@scripps.edu
Summary

In previous studies we have developed Cys₂-His₂ zinc finger domains that specifically recognized each of the sixteen 5’-GNN-3’ DNA target sequences and could be used to assemble six-finger proteins that bind 18 basepair DNA sequences (Beerli, R. R., Dreier, B. and Barbas, C. F., III (2000), Proc. Natl. Acad. Sci. U. S. A. 97, 1495-1500). Such proteins provide the basis for the construction of artificial transcription factors to study gene/function relationships in the post-genomic era. Central to the universal application of this approach is the development of zinc finger domains that specifically recognize each of the 64 possible DNA triplets.

Here we describe the construction of a novel phage display library that enables the selection of zinc finger domains recognizing the 5’-ANN-3’ family of DNA sequences. Library selections provided domains that in most cases showed binding specificity for the 3 bp target site that they were selected to bind. These zinc finger domains were used to construct 6-finger proteins that specifically bound their 18 bp target site with affinities in the pM to low nM range. When fused to regulatory domains, these proteins containing various numbers of 5’-ANN-3’ domains were capable of specific transcriptional regulation of a reporter gene and the endogenous human erbB-2 and erbB-3 genes. These results suggest that modular DNA recognition by zinc finger domains is not limited to the 5’-GNN-3’ family of DNA sequences and can be extended to the 5’-ANN-3’ family. The domains characterized in this work provide for the rapid construction of artificial transcription factors, thereby greatly increasing the number of sequences and genes that can be targeted by DNA-binding proteins built from pre-defined zinc finger domains.
Introduction

The study of protein-DNA interactions is central to our understanding of the regulation of genes and the flow of genetic information characteristic of life. One practical application of the development of a protein-DNA recognition system is the construction of artificial transcription factors that might be used to purposefully regulate gene expression. We have demonstrated that gene expression can be specifically altered through the use of designed polydactyl zinc finger transcription factors that bind 18 basepairs (bps) of DNA sequence. Because of their extended DNA recognition site, these proteins have the potential to be genome-specific transcriptional regulators (1,2). Targeting of only 9 bps of sequence can also result in gene regulation wherein chromatin structure provides for an additional level of specificity (3,4). Because a universal system for gene regulation would provide many new opportunities in basic and applied biology and medicine, the development of such a system is of considerable interest.

Two key features have made Cys2-His2 zinc finger domains the most promising DNA recognition motifs for the construction of artificial transcription factors, modular structure and modular recognition. Each domain consists of approximately 30 amino acids and folds into a ββα structure stabilized by hydrophobic interactions and the chelation of a zinc ion by the conserved Cys2-His2 residues (5,6). DNA recognition of typically 3 bps is provided by presentation of the α-helix into the major groove of DNA. Binding of longer DNA sequences is achieved by covalent tandem repeats of these domains. We have previously reported the phage display selection of zinc finger domains that recognize each of the 5’-GNN-3’ DNA subsites and the refinement of these domains by site-directed mutagenesis (7,8). These domains can be assembled to create polydactyl zinc finger proteins that recognize extended 18 bp DNA sequences (1,2). DNA addresses of this length have the potential to be unique within any genome. In addition to imposing constitutive transcriptional regulation on endogenous genes, these transcription factors can be made hormone-dependent by fusion to designed ligand-binding domains prepared from a variety of nuclear hormone receptors (9). To allow for the rapid construction of zinc finger-based transcription factors that bind any DNA sequence, it is
important to extend the existing set of modular zinc finger domains such that recognition of all 64 DNA triplets is possible. Phage display selection and/or rational design are approaches that might be used to achieve this aim.

While there are now numerous structural studies of zinc finger proteins in complex with DNA, the data is insufficient to allow for the rational design of zinc finger domains that bind any given trinucleotide subsite (10-18). Further, only few zinc finger domains that bind to sequences of the type 5’-ANN-3’ are found in naturally occurring proteins, like finger 5 (5’-AAA-3’) of Gfi-1 (19), finger 3 (5’-AAT-3’) of YY1 (20), fingers 4 and 6 (5’-[A/G]TA-3’) of CF2II (21) and finger 2 (5’-AAG-3’) of TTK (14).

Analysis of the reported zinc finger/DNA structures, however, provides insights into the design of libraries of zinc fingers that might bind any given DNA subsite. To date, the best-characterized member of the Cys2-His2 family of zinc finger proteins is the mouse transcription factor Zif268. The molecular interaction of Zif268 with its target DNA 5’-GCG T/GGG GCG-3’ has been characterized in great detail (10,11). Analysis of the Zif268/DNA complex and related complexes suggests that DNA binding of each domain is achieved predominantly by interaction of the side-chains of the amino acid residues displayed on its α-helix at positions -1, 3, and 6 (AA⁻¹, AA³, AA⁶) with the 3’, middle, and 5’ nucleotide of a 3 bp DNA subsite, respectively (Fig. 1A). Direct interaction of the helix with all three bases of the DNA subsite has not been observed. Typically, only 1 or 2 bases of the 3 bp subsite are specified by residues in the α-helix (11-18,22). While these studies have revealed a variety of direct interactions between residues at helical positions −1 and 3 and the 3’ and middle nucleotides, respectively, the only direct interaction observed with an amino acid residue in position 6 of the α-helix and the 5’ nucleotide is an Arg⁶ or Lys⁶ to 5’ guanine. Therefore, the optimal residue(s) that should be used at position 6 for recognition of DNA target sequences of the type 5’-ANN-3’, 5’-CNN-3’ or 5’-TNN-3’ is not known. While positions −1, 3, and 6 play critical roles in DNA recognition, the other
residues of the helix are important in domain stability and in some cases for DNA recognition. Leucine, for example, is the relatively conserved amino acid residue typically found at helical position 4 in zinc fingers of this type. The side chain of this residue packs into the hydrophobic core of the domain and is believed to be key in stabilizing the domain. Positions 1 and 5 of the \( \alpha \)-helix have been shown to make direct or water-mediated contacts with the phosphate backbone of the DNA. A particularly important role in base recognition, as it relates to the development of modular recognition domains that bind non-5’-GNN-3’ sequences, is played by the side chain of residue 2.

In Zif268, aspartic acid is found at position 2 of each \( \alpha \)-helix and plays a role in determining base specificity at two positions. By making a pair of buttressing hydrogen bonds with \( \text{Arg}^{-1} \) within the same helix, this residue acts to stabilize the \( \text{Arg}^{-1} \) to 3’ guanine interaction within the 3 bp recognition site of the domain. This conclusion is supported by structural observations and mutagenesis studies (8,10,11,23). It is the role this residue plays in specifying base identity at another position that limits domain modularity in recognition. In finger 3 for example, the carboxylate of \( \text{Asp}^2 \) can accept a hydrogen bond from the N4 of cytosine or the N6 of adenine that is base-paired to the 5’ guanine or thymine, respectively, of the finger-2 subsite (10,11,23). Consequently, Zif268 does not discriminate well between 5’-GGG-3’ and 5’-TGG-3’ at the finger-2 subsite and recognition of 5’-AGG-3’ or 5’-CGG-3’ is precluded. A similar interaction is seen between the \( \text{Asp}^2 \) of finger-2 and the corresponding base within the finger-1 recognition site. This cross-strand contact to a base outside the canonical three-nucleotide recognition site effectively restricts the identity of the base that can be recognized at the 5’ position of the preceding finger’s binding site. Thus, while modularity of DNA recognition is a key feature of zinc finger domains, domain-independent modular interaction is not always complete. This type of constraint to zinc finger specificity and modularity, referred to as the target site overlap problem, has been the subject of much discussion (24-26).
Therefore, the design of zinc finger domains that bind 5’-ANN-3’ DNA sequences is constrained by a lack of information regarding residues required for specification of a 5’ adenine. Selections with Zif268-derived libraries are constrained by the target site overlap problem engendered by the Asp$^2$ of finger 3 or finger 2. Consequently, our previously constructed libraries are best suited for selection of domains that bind 5’-GNN-3’ or 5’-TNN-3’ sequences (7,27). One approach to overcome the limitations imposed by target site overlap in the selection of novel recognition domains involves the randomization of amino acid residues in two adjacent fingers (28,29). A second approach involves the sequential selection of fingers 1 to 3 for each particular 9 bp target site and has led to the development of several 5’-ANN-3’ recognition domains (30,31). Neither of these approaches, however, provides for the selection of zinc finger domains that act as independent recognition units but rather provides domains that are dependent on inter-domain interactions for their specificity.

In this study we have taken an approach designed to enforce modularity of DNA recognition within zinc finger domains with the goal of pre-defining domains that can be used for the rapid construction of artificial transcription factors without the need for additional phage display selections. Here we describe the application of this new strategy to select zinc finger domains that recognize the 5’-ANN-3’ family of DNA sequences. The specific DNA-binding properties of these domains are evaluated using multi-target-binding assays that employ all sixteen 5’-ANN-3’ triplets. Selection and specificity data coupled with molecular modeling are used to provide new insights into zinc finger/DNA interactions. We demonstrate that these domains can be readily incorporated into polydactyl proteins containing various numbers of 5’-ANN-3’ domains and that these proteins specifically recognize extended 18 bp sequences. Furthermore, we have explored the ability of these proteins to regulate a reporter gene as well as the endogenous human genes erbB-2 and erbB-3. These results underscore the prospects of constructing polydactyl proteins from pre-defined building blocks.
Experimental procedures

Construction of zinc finger library and selection via phage display

Construction of the zinc finger library was based on the earlier described C7 protein (27; Fig 1A). Finger 3 recognizing the 5’-GCG-3’ subsite was replaced by a domain binding to a 5’-GAT-3’ subsite (7) via a PCR overlap strategy using a primer coding for finger 3 (5’-GAGGAAGTTTGCCACCAGTGCAACCTGGTGAGGCATA CCAAATC-3’) and a vector-specific primer (5’-GTAAAACGACGGCCAGTGCCAAGC-3’). Randomization of the zinc finger library by PCR overlap extension was essentially as described (7,27). The library was ligated into the phagemid vector pComb3H (32). Growth and precipitation of phage were performed as previously described (7,33,34). Binding reactions were performed in a volume of 500µl zinc buffer A (ZBA: 10 mM Tris, pH 7.5/90 mM KCl/1m M MgCl2/90 µM ZnCl2)/0.2% BSA/5 mM DTT/1% Blotto (Biorad)/20 µg double-stranded, sheared herring sperm DNA containing 100 µl precipitated phage (10^13 colony-forming units). Phage were allowed to bind to non-biotinylated competitor oligonucleotides for 1 hr at 4°C before the biotinylated target oligonucleotide was added. Binding continued overnight at 4°C. After incubation with 50 µl streptavidin coated magnetic beads (Dynal; blocked with 5% Blotto in ZBA) for 1 hr, beads were washed ten times with 500 µl ZBA/2% Tween 20/5 mM DTT, and once with buffer containing no Tween. Elution of bound phage was performed by incubation in 25 µl trypsin (10 µg/ml) in TBS (Tris-buffered saline) for 30 min at room temperature.

Hairpin competitor oligonucleotides had the sequence 5’-GGCCGCN’N’N’ATCGAGTTTCTCGATNNNGCGGC-3’, where NNN represents the finger-2 subsite oligonucleotides, N’N’N’ its complementary bases. Target oligonucleotides were biotinylated and usually added at 72 nM in the first three rounds of selection, then decreased to 36 nM and 18 nM in the sixth and last round. As competitor a 5’-TGG-3’ finger-2 subsite oligonucleotide was used to compete with the parental clone. An equimolar mixture of 15 finger-2 5’-ANN-3’ subsites, except for the target site and competitor mixtures of each finger-2 subsites of the type
5’-CNN-3’, 5’-GNN-3’, and 5’-TNN-3’ were added in increasing amounts with each successive round of selection. Usually no specific 5’-ANN-3’ competitor mix was added in the first round.

**Multitarget Specificity Assay and Gel mobility shift analysis**

The zinc finger-coding sequence was subcloned from pComb3H into a modified bacterial expression vector pMal-c2 (New England Biolabs) (35). After transformation into XL1-Blue (Stratagene) the zinc finger-maltose-binding protein (MBP) fusions were expressed by addition of 1 nM isopropyl β-D-thiogalactoside (IPTG). Freeze/thaw extracts of these bacterial cultures were applied in 1:2 serial dilutions to 96-well plates coated with streptavidin (Pierce), and were tested for DNA-binding specificity against each of the sixteen 5’-GAT ANN GCG-3’ target sites. ELISA (enzyme-linked immunosorbant assay) was performed essentially as described (7,8). 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. Detection occurred by addition of alkaline phosphatase substrate (Sigma), and the OD405 was determined by a micotiter plate reader with SOFTMAX2.35 (Molecular Devices).

Gelshift analysis was performed with purified protein (Protein Fusion and Purification System, New England Biolabs) essentially as described (7).

**Site-directed mutagenesis of finger 2**

Finger-2 mutants were constructed by PCR as described (7,8). As PCR template the pMal vector encoding for C7.GAT was used. 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 (5’-GCG-3’) 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 (1). The proteins generated in this work contained helices recognizing 5’-GNN-3’
DNA sequences (7), as well as 5’-ANN-3’ and 5’-TAG-3’ helices described here. Six finger proteins were assembled via compatible XmaI and BsrFI restriction sites. Analysis of DNA-binding properties were performed using freeze/thaw extracts from from IPTG-induced bacteria. For the analysis of the capability of these proteins to regulate gene expression they were fused to the activation domain VP64 or repression domain KRAB of Kox-1 as described earlier ((1,2); VP64: tetrameric repeat of the herpes simplex virus VP16 minimal activation domain) and subcloned into pcDNA3 (Invitrogen) or the retroviral pMX-IRES-GFP vector ((36); IRES, internal ribosome-entry site; GFP, green fluorescent protein).

Transfection and luciferase assays

HeLa cells were used at a confluency of 40-60%. Cells were transfected with 160 ng reporter plasmid (pGL3; Promega) containing the promoter sequence with zinc finger-binding sites and 40 ng of effector plasmid (zinc finger-effector domain fusions in pcDNA3) in 24 well plates. Cell extracts were prepared 48 hrs after transfection and measured with luciferase assay reagent (Promega) in a MicroLumat LB96P luminometer (EG & Berthold, Gaithersburg, MD).

Retroviral gene targeting and Flow cytometric analysis

These assays were performed as described (2). As primary antibody an ErbB-1-specific mAb EGFR (Santa Cruz), ErbB-2-specific mAb FSP77 (gift from Nancy E. Hynes; (37)) and an ErbB-3-specific mAb SGP1 (Oncogene Research Products) were used. Fluorescently labeled donkey F(ab’)2 anti-mouse IgG was used as secondary antibody (Jackson Immuno-Research).

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) and QGSR-GCAC (1A1H). 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(+/- 0.3) Å and the angle formed by the heavy atoms and hydrogen was 120° or greater. Plausible van der Waals interactions required a distance between methyl group carbon atoms of 4(+/- 0.3) Å.
Results

Library construction and selection

Selections of one of our previously reported phage display libraries for modular zinc finger domains that bind to 3 bp DNA sites containing 5' nucleotides other than guanine or thymine have met with no success (data not shown). This phage display library (7) was based on C7, a high affinity variant of the mouse transcription factor Zif268 also selected by phage display ((27); Fig. 1A). It is believed that selections of this library for modular recognition domains of these specificities failed due to the cross-subsite interaction from Asp2 of the finger-3 recognition helix RSD-E-LKR (corresponding to helical positions -1, 1, 2, 3, 4, 5, 6)(7). In order to overcome the limitations to selections of novel finger-2 domains imposed by target site overlap, we sought to eliminate the transgressing finger-3 interaction. First, finger-3 of C7 (RSD-E-RKR) that binds to subsite 5’-GCG-3’ was exchanged with a domain which does not contain aspartate in position 2 (Fig. 1A). The helix TSG-N-LVR, previously characterized in the finger 2 position to bind with high specificity to the triplet 5’-GAT-3’, appeared to be an optimal candidate since Gly2 does not possess a side chain capable of the type of interactions observed for Asp2 (7). This 3-finger protein (C7.GAT; Fig. 1A), containing finger 1 and 2 of C7 and the 5’-GAT-3’-recognition helix in the finger-3 position, was analyzed for DNA-binding specificity on target DNAs with different finger-2 subsites by multi-target ELISA and compared with the activity of the original C7 protein (Fig. 1B). Both proteins bound to the 5’-TGG-3’ subsite (note that C7 also binds to 5’-GGG-3’ due to the 5’ specification of thymine or guanine by Asp2 of finger-3 as reported earlier (18)). The recognition of the 5’ nucleotide of the finger-2 subsite was evaluated using a mixture of all 16 5’-XNN-3’ target sites (X = adenine, guanine, cytosine or thymine). Indeed, while the original C7 protein specified a guanine or thymine in the 5’ position of finger 2, C7.GAT did not specify a unique base at this position, indicating that the cross-subsite interaction to the adenine complementary to the 5’ thymine was abolished. A similar effect has been reported for mutants of Zif268 where Asp2 was replaced by Ala2 using site-
directed mutagenesis (8,23). The affinity of C7.GAT, measured by gel mobility shift analysis, was found to be relative low, ~ 400 nM as compared to 0.5 nM for C7 (7).

Based on the 3-finger protein C7.GAT, a library was constructed in the phage display vector pComb3H (32,34). Gene randomization targeted positions encoding residues -1, 1, 2, 3, 5, and 6 of the α-helix of finger 2 using a VNS codon doping strategy (V = adenine, cytosine or guanine, N = adenine, cytosine, guanine or thymine, S = cytosine or guanine). This allowed for 24 possible amino acids at each randomized position, while the aromatic amino acids Trp, Phe, and Tyr, as well as Cys and all stop codons were excluded in this strategy. Because Leu is predominately found at position 4 of the recognition helices of zinc finger domains of the type Cys2-His2 this position was not randomized. After transformation of the phagemid library into ER2537 cells (New England Biolabs) the library contained 1.5 x 10^9 members. This exceeded the necessary library size by 60-fold and should be sufficient to contain all possible amino acid combinations allowed by the doping strategy. Six rounds of selection of zinc finger-displaying phage were performed for binding to each of the sixteen 5’-GAT-ANN-GCG-3’ biotinylated hairpin target oligonucleotides in the presence of non-biotinylated competitor DNA (see experimental procedures). 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 sixth round the DNA target concentration was usually 18 nM, 5’-CNN-3’, 5’-GNN-3’, and 5’-TNN-3’ competitor mixtures were in 5-fold excess for each oligonucleotide pool and the specific 5’-ANN-3’ competitor mixture (excluding the target sequence) in 10-fold excess. Phage binding to the biotinylated target oligonucleotides were recovered by capture with streptavidin-coated magnetic beads.

Phage clones were typically analyzed after the sixth round of selection. The amino acid sequences of selected finger-2 helices were determined and generally showed good conservation in positions -1 and 3 (Fig. 2), consistent with previously observed amino acid residues in these positions (7). Position -1 was Gln when the 3’ nucleotide was adenine, with the exception of domains binding 5’-ACA-3’ (SPA-D-LTN) where a Ser was strongly selected. Triplets
containing a 3’ cytosine selected Asp\(^1\) (exceptions were domains binding 5’-AGC-3’ and 5’-ATC-3’), a 3’ guanine Arg\(^1\), and a 3’ thymine Thr\(^1\) or His\(^1\). For the recognition of a middle adenine, Asp and Thr were selected in position 3 of the recognition helix. For binding to a middle cytosine an Asp\(^3\) or Thr\(^3\) was selected, for a middle guanine His\(^3\) (an exception was the 5’-AGT-3’ recognition helix, which may have a different binding mechanism due to it’s atypical amino acid residue His\(^1\)), and for a middle thymine Ser\(^3\) and Ala\(^3\). Note also that the domains binding to 5’-ANG-3’ subsites contain Asp\(^2\) which likely stabilizes the interaction of the 3-finger protein by contacting the complementary cytosine base paired to the 5’ guanine in the finger-1 subsite. Even though there was a predominant selection for Arg and Thr at position 5 of the recognition helices, positions 1, 2 and 5 were varied. This is not surprising because these residues are usually not involved in direct base contacts with DNA (10,11). In addition, one domain was selected from this library against a finger-2 subsite 5’-TAG-3’. The amino acid sequence for this helix was identified as RED-N-LHT (Fig. 3z).

The most interesting observation was the selection of amino acid residues in position 6 of the \(\alpha\)-helices since this residue typically specifies binding to the 5’ nucleotide of the 3 bp subsite. In contrast to recognition of a 5’ guanine where a direct base contact is achieved by Arg or Lys in position 6 of the helix, no direct interaction has been observed in protein/DNA complexes for any other nucleotide in the 5’ position (11-18). Selection of domains against finger-2 subsites of the type 5’-GNN-3’ had previously generated domains containing only Arg\(^6\) which directly contacts the 5’ guanine (7). In analogy with guanine specification, one could assume that the recognition of 5’ adenine could be achieved by certain amino acid residues in position 6 of the \(\alpha\)-helix. However, unlike the results for 5’-GNN-3’ zinc finger domains, selections of the phage display library against finger-2 subsites of the 5’-ANN-3’ type identified domains containing a variety of amino acid residues: Ala\(^6\), Arg\(^6\), Asn\(^6\), Asp\(^6\), Lys\(^6\), Glu\(^6\), Thr\(^6\) or Val\(^6\) (Fig. 2).
Characterization of zinc finger domains that bind to the 5’-ANN-3’ family of DNA sequences

Finger-2 variants of C7.GAT were subcloned into a bacterial expression vector as fusions with maltose-binding protein (MBP) and proteins were expressed by induction with 1 mM IPTG (35). Proteins were tested by enzyme-linked immunosorbant assay (ELISA) against each of the 16 finger-2 subsites of the type 5’-GAT-ANN-GCG-3’ to investigate their DNA-binding specificity (Fig. 3, black bars). In addition, 5’-nucleotide recognition was analyzed by exposing zinc finger proteins to the specific target oligonucleotide and three additional oligonucleotides containing subsites that differed only at the 5’-nucleotide of the middle triplet. For example, pAAA (protein AAA) was studied for binding to 5’-AAA-3’, 5’-CAA-3’, 5’-GAA-3’, and 5’-TAA-3’ subsites (Fig. 3, white bars). Many of the 3-finger proteins tested showed exquisite DNA-binding specificity for the finger-2 subsite for which they were selected. The helices that provided for the most specific binding within each set of selected helices are boxed in Fig. 2 and their binding profiles are provided in Fig. 3a-3n. Additional domains tested (marked with an asterisk in Fig. 2) are summarized in Fig. 3o-3y while the lower panel of Fig. 3 (w, x, y, aa, bb) provides specificity data on proteins generated by site-directed mutagenesis (pmNNN). Binding profiles for the pAGC and pATC are not shown since the DNA-binding activity of these proteins was too weak to be detected by ELISA. The most promising helix for pAGC (DAS-H-LHT) contained amino acids Asp\textsuperscript{1} and His\textsuperscript{3} anticipated for specification of a 3’ cytosine and middle guanine. To analyze a larger set of clones selected to bind this site, the pool of coding sequences for AGC was subcloned into the plasmid pMal after the sixth round of selection and 18 individual clones were tested for DNA-binding specificity. None showed measurable DNA binding in ELISA (data not shown). In the case of 5’-ATC-3’ recognition, two helices (RRS-S-CRK and RRS-A-CRR) were selected containing a Leu\textsuperscript{4} to Cys\textsuperscript{4} mutation. Neither protein demonstrated detectable DNA binding. Rational design was applied to find domains capable of binding to 5’-AGC-3’ or 5’-ATC-3’ since no proteins displaying detectable binding in ELISA were generated by phage selections. Finger-2 mutants were constructed based on the recognition helices which were previously demonstrated to bind specifically to 5’-GGC-3’ (ERS-K-LAR,
DPG-H-LVR) and 5’-GTC-3’ (DPG-A-LVR) (7). For 5’-AGC-3’ two proteins were constructed (ERS-K-LRA, DPG-H-LRV) by simply exchanging position 5 and 6 to a 5’ adenine recognition motif RA or RV (Fig. 3a, 3b and 3i). DNA binding of these proteins was also below the level of detection. In the case of 5’-ATC-3’ recognition, two finger-2 mutants containing a RV motif (Fig. 3b) were constructed (DPG-A-LRV, DPG-S-LRV). Both proteins bound DNA with an extremely low affinity (data not shown). In conclusion, these examples demonstrate the difficulty faced in the generation of zinc finger domains by rational design.

Analysis of the 3-finger proteins on the sixteen finger-2 subsites by ELISA revealed that some finger-2 domains bound best to a target DNA that they were not selected to bind. First, the predominant helix sequence selected for 5’-AGA-3’ binding was RSD-H-LTN, which in fact bound 5’-AGG-3’ (Fig. 3t). This can be explained by the presence of Arg in position -1. Further, this protein demonstrated better specification of a 5’ adenine as compared to the helix selected for 5’-AGG-3’ recognition pAGG (RSD-H-LAE; Fig. 3j). Second, a helix binding specifically to 5’-AAG-3’ (RKD-N-LKN; Fig. 3o) was actually selected by panning with the 5’-AAC-3’ oligonucleotide (Fig. 2), and bound more specifically to the finger-2 subsite 5’-AAG-3’ than pAAG (RSD-T-LSN; Fig. 3c), which had been selected in the 5’-AAG-3’ set.

In addition, proteins directed to target sites of the type 5’-ANG-3’ showed cross reactivity with all four target sites of the 5’-ANG-3’type, except for pAGG (RSD-H-LAE, Fig. 3j and RSD-H-LTN, Fig. 3t). Specific recognition of a middle purine within the 5’-ANG-3’ family appears to be more facile than recognition of a middle pyrimidine since pAAG (RSD-N-LKN; Fig. 3o) also displayed minimal cross-reactivity. In comparison, proteins pACG (RTD-T-LRD; Fig 3g) and pATG (RRD-A-LNV; Fig. 3m) demonstrated substantial crossreactivity with all 5’-ANG-3’ subsites. Stringent specification of a middle pyrimidine was also previously found to be difficult within the 5’-GNG-3’ family of zinc finger domains (7,8). To improve the recognition of the middle nucleotide, finger-2 mutants containing different amino acid residues in position 3 were generated by site-directed mutagenesis. Binding of pAAG (RSD-T-LSN, Fig. 3c) was more specific for a middle adenine after a Thr3 to Asn3 mutation (Fig. 3w). Binding to 5’-ATG-
3’ (SRD-A-LNV; Fig. 3m) was improved by a single amino acid exchange Ala³ to Gln³ (Fig. 3bb), while a Thr³ to Asp³ or Gln³ mutation within pACG (RSD-T-LRD; Fig. 3g) abolished DNA-binding activity (data not shown). In addition, the recognition helix of pAGT (HRT-T-LLN; Fig. 3k) showed cross-reactivity for the middle nucleotide that was reduced by a Leu⁵ to Thr⁵ substitution (Fig. 3aa).

**Generation of polydactyl proteins containing 5’-ANN-3’ zinc finger domains**

We have previously demonstrated that endogenous and transgene regulation can be achieved with 6-finger proteins containing zinc finger domains specifically recognizing 5’-(GNN)₆-3’ DNA sequences (1,2). To investigate whether the domains described here for recognition of the 5’-ANN-3’ family of DNA sequences are suitable for the construction of such artificial transcription factors, four 6-finger proteins were assembled containing various numbers of 5’-ANN-3’ domains. For each of the 6-finger proteins, two 3 finger-coding regions were initially generated with a rapid PCR overlap extension method using the Sp1C framework (1). These 3-finger proteins were then fused to create 6-finger proteins via restriction sites (Fig. 4A) and cloned into the bacterial expression vector pMal for analysis of DNA-binding specificity and affinity. First, the 6-finger protein pAart was constructed. This protein was designed to recognize an arbitrary 18 bp target site 5’-ATG-TAG-AGA-AAA-ACC-AGG-3’, and was completely free of any 5’-GNN-3’ triplets. As additional examples, three 6-finger proteins containing both 5’-GNN-3’ and 5’-ANN-3’ domains were constructed. We chose the endogenous human genes erbB-2 and erbB-3 as design targets since we have previously demonstrated specific regulation of these endogenous genes with the 6-finger proteins pE2C and pE3. These proteins had been designed to bind to 5’-(GNN)₆-3’ DNA sequences within the 5’ untranslated regions of these genes (2). In analogy, the 6-finger protein pE2X was constructed for the target site 5’-ACC-GGA-GAA-ACC-AGG-GGA-3’ at position -168 to -151 (with respect to the ATG start codon) in the 5’ untranslated region (UTR) of the erbB-2 gene (Fig. 4a). In addition, two proteins that target the 5’ UTR of the erbB-3 gene were also generated. The
protein pE3Y was designed for the target site 5′-ATC-GAG-GCA-AGA-GCC-ACC-3′ at position -94 to -111 of the 5′ UTR, while pE3Z was constructed for the target site 5′-GCC-GCA-GCA-GCC-ACC-AT-3′ at position -79 to -61 (Fig. 4a). Protein extracts containing the zinc finger-MBP fusion protein were tested for DNA binding in ELISA (Fig. 4b). All four proteins showed exquisite binding specificity for their target DNAs with no cross-reactivity to the other target sites tested. The DNA-binding affinities of the proteins were determined in electrophoretic mobility shift assays using purified proteins. The protein pAart bound its DNA target site with an affinity of 7.5 pM, while proteins pE2X, pE3Y, pE3Z bound their target DNA’s with affinities of 15 nM, 8 nM, and 2 nM, respectively.

To evaluate the potential of a transcription factor based on the pAart protein, the coding sequence of the protein was cloned into the expression vector pcDNA3 and fused to the VP64 activation domain, a tetrameric repeat of the minimal activation domain derived from the herpes simplex virus protein VP16 (1,38). HeLa cells were transiently co-transfected with a luciferase reporter plasmid under the control of a minimal promoter containing multiple zinc finger-binding sites and a TATA-box and the expression constructs coding either for the zinc finger protein alone or the zinc finger protein fused to the VP64 domain. The pAart luciferase reporter construct contained five copies of the pAart-binding site. A luciferase reporter plasmid previously constructed to be responsive to the 2C7 zinc finger protein containing six 2C7-binding sites (35) was used as control. Cotransfection experiments revealed that the expression of luciferase was up-regulated approximately 2000-fold by the pAart-VP64 fusion protein in comparison to the control that expressed the zinc finger protein containing no activation domain (Fig. 5a). Activation was specific and no regulation of the reporter containing 6x2C7-binding sites was observed (Fig. 5b). Further, transfection of a p2C7-VP64 expression construct (35) activated luciferase expression only when the promoter contained 6x2C7-binding sites (Fig. 5b), but not when the promoter contained the 5xAart-binding sites (Fig. 5a). Thus, the zinc finger domains described here can be used to build artificial transcription factors capable of regulating genes in living cells.
To investigate the ability of the 6-finger proteins pE2X, pE3Y, and pE3Z to transcriptionally regulate the endogenous human erbB-2 and erbB-3 genes, their coding sequences were subcloned into the retroviral vector pMX-IRES-GFP and fused to the VP64 activation or the KRAB repression domain of Kox-1 (1,39). Recombinant retrovirus was then used to infect the human carcinoma cell line A431. Three days after infection, cells were subjected to flow cytometry to analyze the expression levels of ErbB-2 and ErbB-3 (Fig. 6). Infection efficiency was determined by measurement of GFP expression. Analysis of all cell pools, with the exception of pE2X-VP64, indicated that greater than 80% of the cells were infected with the retrovirus. To determine the expression levels of ErbB-2 and ErbB-3, cells were stained with specific antibodies, or a control antibody specific for ErbB-1. The fusion protein pE2X-VP64 was capable of up-regulating ErbB-2 expression in ~50% of the cells reflecting the low infection efficiency obtained with its corresponding retrovirus as determined by GFP expression. In contrast, pE2X-KRAB did not cause ErbB-2 down-regulation despite efficient infection (Fig. 6A). This result might be explained by the relatively low affinity, 15 nM, of this 6-finger protein for its target site. In contrast, pE3Y showed specific up- and down-regulation of ErbB-3 expression when fused to VP64 or KRAB, respectively (Fig. 6B). The efficiency of this transcription factor in gene activation and repression was analogous to that previously reported for the pE3 transcription factor (2). No regulation of ErbB-1 and ErbB-2 levels was observed with the pE3Y derived transcription factors. The pE3Z based transcription factors, however, were not capable of regulating endogenous the erbB-3 gene (data not shown), even though pE3Z had an affinity higher than pE3Y or pE2X. The failure of this transcription factor to regulate erbB-3 may be due to a number of factors, such as positioning of the target site within the 5’ UTR, competition with cellular factors that bind to the same site, orientation of the zinc finger fusion protein (note pE3 and pE3Y bind to the minus strand, pE3Z to the plus strand) or chromatin structure. We have not investigated in detail the mechanism responsible for the failure of the pE3Z-based transcription factors. This result, however, emphasizes that it may be essential to explore several zinc finger proteins directed to different target sites in order to
impose regulation on an endogenous gene and is consistent with recently reported studies concerning the regulation of the human *erythropoietin* and *vascular endothelial growth factor* genes (3,4).

The results reported in this study demonstrate that 6-finger proteins containing various numbers of 5’-GNN-3’ and 5’-ANN-3’ zinc finger domains have specific DNA-binding properties and that these proteins are potent artificial transcription factors. The increased targeting potential provided by the 5’-ANN-3’ recognition domains, greatly improves the possibilities for the construction of transcription factors directed to specific genomic sites.
Discussion

Zinc finger proteins of the Cys\textsubscript{2}-His\textsubscript{2} type have shown promise as versatile DNA-binding devices that would be essential components of a universal system for gene regulation (1-4,35). Ideally, zinc finger proteins could be readily constructed to bind any DNA sequence; however, information regarding zinc finger/DNA interactions is constrained to just a few of the 64 possible 5'-NNN-3' DNA subsites. Structural analysis of several domains that may specify a 5' nucleotide other than guanine has not revealed any specific interaction from position 6 of the $\alpha$-helix (11-18,22). Thus at present it is not possible to directly design zinc finger domains that specifically bind any given 3 bp DNA subsite.

While phage display selection coupled with refinement by site-directed mutagenesis has provided domains specifically recognizing each of the 16 DNA triplets of the 5'-GNN-3' type (7,8), selection or design of domains specific for subsites containing a different 5' nucleotide, in particular adenine or cytosine, have proven to be difficult. Recently the rational design of a synthetic zinc finger protein named Mago was reported. This protein was designed to bind 5'-ATG-3' in the finger-3 position with the helix sequence RAA-V-LQN. Characterization of this protein by binding-site selections revealed the preferential recognition of 5' thymine or cytosine by the Asn\textsuperscript{6} of this finger (40). This contrasts the present study where Asn\textsuperscript{6} was frequently selected in finger-2 domains for 5' adenine recognition (Fig. 2). These domains did indeed generally favor binding to a 5' adenine with some cross-reactivity to a 5' guanine as shown by multi-target ELISA (Fig. 3c, 3k, 3n, 3o, and 3t). Zinc finger/DNA recognition, as illustrated by these and many other examples, is more complex than a simple amino acid to base code.

Phage display selection of modular zinc finger domains that bind to subsites containing a 5' adenine or cytosine from our previously described finger-2 library based on the 3-finger protein C7 (7) failed due to the limitations imposed by Asp\textsuperscript{2} of finger 3 of this protein which makes a cross-subsite contact to the nucleotide complementary to the 5' position of the finger-2 subsite (Fig. 1A). In the library reported here, this contact was eliminated by exchanging finger
3 of C7 with a domain lacking Asp\(^2\) yielding C7.GAT (Fig. 1A). In most cases, novel 3-finger proteins that bound finger-2 subsites of the 5’-ANN-3’ type were obtained. For subsites 5’-\(\Delta\)GC-3’ and 5’-\(\Delta\)TC-3’, however, no tight binding proteins were identified. This was not expected since domains that bind the subsites 5’-\(\Delta\)GC-3’ and 5’-\(\Delta\)TC-3’ were previously selected and shown to exhibit excellent DNA-binding specificity and affinity for their target sites (7). One potential explanation for this might be the limited randomization strategy used here based on VNS codons that do not encode for the aromatic amino acids. This limited randomization strategy was chosen since within the domains selected for 5’-GNN-3’ recognition, no aromatic amino acid residues were selected even though they were included in the randomization strategy (7). Several naturally occurring zinc finger domains do indeed contain aromatic residues, for example finger 2 of CFII2 (VKD-Y-LTK; (21)), finger 1 of TFIIIA (KNW-K-LQA; (16)), finger 1 of TTK (HIS-N-FCR; (14)) and finger 2 of GLI (AQQ-M-LVV; (12)). It is tempting therefore to speculate that aromatic amino acid residues might be important for the recognition of subsites 5’-AGC-3’ and 5’-ATC-3’. Alternatively, high-affinity interactions with these particular subsites might not be possible with these proteins but at this time we believe this to be an unlikely explanation.

In recent years it has become clear that the recognition helix of Cys2-His2 zinc finger domains can adopt different orientations relative to DNA in order to achieve optimal binding (26). However, the orientation of the helix in this region may be partially restricted by interactions involving the zinc ion, His\(^7\), and the phosphate backbone since these interactions are frequently observed in structural studies (Fig. 7a). Comparative studies of zinc finger/DNA complexes has led to the conclusion that the Cα atom of position 6 is usually 8.8 ± 0.8Å away from the nearest heavy atom of the 5’ nucleotide in the DNA subsite. This distance is most readily bridged by the long side chains presented by the amino acids Arg\(^6\) or Lys\(^6\) that most typically provide for 5’ guanine specification (26). No interaction of any other position 6 residue with a base other than guanine has been observed in zinc finger/DNA complexes. For example, finger 4 of YY1 (QST-N-LKS) recognizes 5’-\(\Delta\)AA-3’ but no contact between Ser\(^6\) and the 5’
cytosine is seen (15). Further, in the case of Thr$^6$ of finger 3 of YY1 (LDF-N-LRT) that recognizes 5’-ATT-3’, and finger 2 of Zif268 (RSD-H-LTT) that recognizes 5’-T/GGG-3’, no contact with the 5’ nucleotide is observed (15). Finally, Ala$^6$ of finger 2 of tramtrack (RKD-N-MTA) that binds to the subsite 5’-ΔAG-3’ does not contact the 5’ adenine (14).

In the present study, eight different amino acid residues were selected at position 6 of finger-2 of the C7.GAT library for recognition of DNA subsites of the 5’-ANN-3’ type; Ala$^6$, Arg$^6$, Asn$^6$, Asp$^6$, Glu$^6$, Lys$^6$, Thr$^6$, Val$^6$ (Fig. 2). Selection of a wide range of residues at this position is consistent with studies from other laboratories where positions within adjacent fingers were randomized (28,29). These studies selected amino acid residues Tyr, Val, Thr, Asn, Lys, Glu and Leu, as well as Gly, Ser and Arg, but not Ala, for 5’ adenine recognition. In addition, a sequential phage display selection strategy identified several 5’-ANN-3’-binding fingers and evaluated their specificity using target site selections. Arg, Ala, and Thr in position 6 of the helix were demonstrated to predominantly specify 5’ adenine recognition (31). Further, Thr$^6$ was identified by target site selections of finger 5 of Gfi-1 (QSS-N-LIT) that binds the subsite 5’-AAA-3’ to specify a 5’ adenine (19). In combination with the data presented here there appears to be a complex but nonrandom relationship between the amino acid residue in position 6 and 5’ adenine recognition. No insight into the nature of this relationship is obtained by analysis of the reported zinc finger/DNA structures since they record only 5’ guanine interactions. Some residues that display short amino acid side chains, like Ala, Val, Thr, or Asn may simply not sterically interfere with 5’-ANN-3’ recognition and thereby play a passive role in base specification. This notion is supported by site-directed mutagenesis studies of position 6 in the helix (QRS-A-LTV) that binds to a 5’-G/ATA-3’ subsite (21). Replacement of Val$^6$ with Ala$^6$ or Lys$^6$, residues also found in the domains described here, had no affect on the binding specificity or affinity of this domain. In general, the data leaves open the question of whether a direct interaction between the residue at position 6 of the recognition helix and a 5’ adenine exists.

In order to obtain insight into potential interactions of the most frequently selected position 6 residues (Ala$^6$, Asn$^6$, and Arg$^6$) and the 5’ adenine, we have turned to computer
modeling. Analysis of the interaction of Ala$^6$ in the helix binding to 5′-AAA-3′ (QRA-N-LRA; Fig. 3a) with a 5′ adenine was based on the coordinates of the protein/DNA complex of finger 1 (QSG-S-LTR) of a Zif268 variant ((18); Fig. 7g). If Gln$^{-1}$ and Asn$^3$ of QRA-N-LRA hydrogen bond with their respective adenine bases in the canonical way, these interactions should fix a distance of approximately 8 Å between the methyl group of Ala$^6$ and the 5′ adenine and a distance of more than 11 Å between the methyl groups of Ala$^6$ and the thymine base-paired to the 5′ adenine. This result suggests that no direct contact can be proposed for Ala$^6$, Val$^6$, or Thr$^6$ and a 5′ adenine or the thymine to which it is paired.

Interestingly, the expected lack of 5′ specification by amino acid residues that present short side chains in position 6 of the α-helix is only partially supported by the binding data. While helices such as RRD-A-LN$^V_\_$(Fig. 3m) and the finger-2 helix RSD-H-LTT of C7.GAT (Fig. 1B) did indeed show essentially no 5′ specificity, helix DSG-N-LRV (Fig. 3b) displayed excellent specificity for a 5′ adenine, while TSH-G-LTT (Fig. 3v) was specific for 5′ adenine or guanine. Other helices with position-6 residues of this type displayed varying degrees of 5′ specificity, but typically excluded 5′ thymine recognition (Fig. 3). Since it is unlikely that the position-6 residue makes a direct base contact, the observed binding patterns must result from other binding mechanisms. Possibilities include the involvement of bound water, local sequence-specific DNA structure changes, and overlapping interactions from neighboring domains. The latter possibility is disfavored, however, because the residue in position 2 of finger 3 (which is frequently observed to contact the neighboring site) is glycine in the parental protein C7.GAT, and because 5′ thymine was not excluded by RRD-A- LN$^V_\_$ (Fig. 3m) or RSD-H-LTT (Fig. 1B).

Asparagine was also frequently selected in position 6 of the recognition helix. Helix HRT-T-LTN (Fig. 3k) and RSD-T-LSN (Fig. 3c) displayed excellent specificity for 5′ adenine. However, Asn$^6$ also seemed to impart specificity for both adenine and guanine in some cases (Fig. 3n, 3o and 3t) suggesting an interaction with the N7 common to both nucleotides. Computer modeling of the helix binding to 5′-AGG-3′ (RSD-H-LTN; Fig. 3t) based on the
coordinates of finger 2 bound to 5’-TGG-3’ in the Zif268/DNA crystal structure (RSD-H-LTT; (18)), suggested that the Nδ of Asn6 would be approximately 4.5Å from N7 of the 5’ adenine. A modest reorientation of the α-helix which is considered within the range of canonical docking orientations (26) could plausibly bring the Nδ within hydrogen bonding distance. This reorientation would be analogous to that observed when glutamate rather than arginine appears in position -1. However, it is interesting to speculate why Asn6 was selected in this 5’-ANN-3’ recognition set while the longer side chain of Gln6 was not. The side chain of Gln6, being more flexible, may have been selected against during phage display or its absence may be an artifact of the selection procedure. Alternatively, the shorter side chain of Asn6 might accommodate an ordered water molecule that could contact the 5’ nucleotide without reorientation of the helix.

To further study this question, finger-2 mutants containing Asn6 to Gln6 amino acid exchanges were constructed for pAAG (Fig. 3o), pAGG (Fig. 3t), and pATT (Fig. 3n). Analysis of these proteins in multi-target ELISA studies showed a shift in the recognition of the 5’ nucleotide from adenine towards guanine and cytosine (Fig. 3x and 3y). The mutant pmATT failed to bind DNA at all (data not shown), while the parental protein containing Asn6 showed excellent binding specificity for its target site. These results indicate that the longer side chain of Gln6 may sterically interfere with the binding of the protein, explaining why this residue was not selected by phage display.

The final position 6 residue to be considered is Arg6. It was somewhat surprising that Arg6 was selected so frequently on 5’-ANN-3’ targets since in previous studies it was unanimously selected to recognize a 5’ guanine with high specificity (7). However, in the present study Arg6 primarily specified 5’ adenine (Fig. 3e, f, h and s), with cross reactivity to a 5’ guanine in some cases (Fig. 3q and 3r). Computer modeling of a helix binding to 5’-ACA-3’ (SPA-D-LTR; Fig. 3e) based on the coordinates of finger 1 (QSG-S-LTR) of a Zif268 variant bound to 5’-GCA-3’ (18), suggested that Arg6 could easily adopt a configuration that allowed it to make a cross-strand hydrogen bond to O4 of a thymine base-paired to the 5’ adenine (Fig. 7c and d). In fact, Arg6 could bind with good geometry to both the O4 of thymine and the O6 of a...
guanine base-paired to a middle cytosine. Such an interaction is consistent with the fact that Arg$^6$ was selected almost unanimously when the target sequence was 5'-ACN-3'. The notion that arginine can facilitate multiple interactions is compelling. Several lysines in TFIIIA were observed by NMR to be conformationally flexible (41), and Gln$^{-1}$ also behaves in a manner which suggests flexibility (8). Arginine has more rotatable bonds and more hydrogen bonding potential than lysine or glutamine and it is attractive to speculate that Arg$^6$ is not limited to specification of only a 5’ guanine.

The amino acid residues selected in positions -1 and 3 in the present study were typically analogous to those identified in their 5’-GNN-3’ binding counterparts (7) with two exceptions. Ser$^{-1}$ was selected for pACA, recognizing a 3’ adenine (Fig. 3e and 3q) and His$^{-1}$ was selected for pAGT and pATT, recognizing a 3’ thymine (Fig. 3k, 3n). While Gln$^{-1}$ was frequently used to specify a 3’ adenine in subsites of the 5’GNN-3’ type, a new element of 3’ adenine recognition is suggested in the present study involving Ser$^{-1}$ selected in domains recognizing the 5’-ACA-3’ subsite (Fig. 2). Computer modeling was used to study the interactions of this helix with DNA. Models suggested that Ala$^2$, co-selected in the helix SPA-D-LTR (Fig. 3e), can potentially make a van der Waals contact with the methyl group of the thymine based-paired to 3’ adenine (Fig. 7b and c) and that Ser$^{-1}$ potentially makes a hydrogen bond with the 3’ adenine (Fig. 7b). Additional evidence that Ala$^2$ might also be directly involved in specification of the binding site of this protein is that helix SPA-D-LTR (Fig. 3e) is strongly specific for 3’ adenine while SHS-D-LVR (Fig. 3q) is not. Gln$^{-1}$ is often sufficient for 3’ adenine recognition. However, data from our previous studies suggested that the side chain of Gln$^{-1}$ can adopt multiple conformations, enabling, for example, recognition of 3’ thymine (8,18,42). It is therefore intriguing to speculate that Ala$^2$ in combination with Ser$^{-1}$ may provide an alternative means for specifying a 3’ adenine.

Another interaction not observed in our 5’-GNN-3’ study is the cooperative recognition of 3’ thymine by His$^{-1}$ and the residue at position 2. The finger 1 helix HIS-N-FCR of tramtrack binds the subsite 5’-GAT-3’ (14). The structure of this protein in complex with DNA shows the His$^{-1}$ ring perpendicular to the plane of the 3’ thymine base and approximately 4Å from the
methyl group. Further, Ser$^2$ also makes a hydrogen bond with O4 of 3’ thymine. A similar set of contacts can be envisioned by computer modeling of the recognition of 5’-ATT-3’ by helix HKN-A-LQN (Fig. 3n). Asn$^2$ in this helix has the potential not only to hydrogen bond with 3’ thymine but also with the adenine base-paired to it (Fig. 7e and f). His$^{-1}$ was also found within the helix the selected to bind 5’-AGT-3’ (HRT-T-LLN; Fig. 3k) in combination with a Thr$^2$. Residue Thr$^2$ might be involved in a similar recognition mechanism as Ser$^2$.

The examples discussed above demonstrate that it is difficult at the present time to understand why amino acid residues like alanine, valine, or threonine in position 6 of the $\alpha$-helix assist in the recognition of a 5’ adenine. These helices may simply not sterically exclude an adenine in the 5’ position of the triplet. It is reasonable to consider that the surrounding domains, as well as bound water molecules and local DNA structure might influence the DNA-binding properties of these domains as has been discussed in detail (24-26,43). However, the domains characterized in this study that contain Asn$^6$ and Arg$^6$ more likely specify 5’ adenine recognition by direct interaction with the nucleotide, as discussed above.

While further exploration of the recognition mechanisms used by these 5’-ANN-3’-binding zinc finger domains will likely require structural studies, their immediate application towards the construction of artificial transcription factors has many practical consequences for basic and applied biology. Previously we demonstrated for the first time that polydactyl zinc finger proteins containing six domains of the 5’-GNN-3’ recognition type could be assembled into DNA-binding proteins that bind an 18 bp target DNA sequence (1,2). The set of 16 5’-GNN-3’ domains is sufficient to target 17 million different 18 bp sequences, 16$^6$. Assuming random sequence distribution such sites should appear once in approximately every 4000 nucleotides. Since eukaryotic promoters are not composed of a random distribution of nucleotides and are generally G/C-rich, analysis of the promoter regions of human genes typically reveals the presence of two or more 5’-(GNN)$_6$-3’ sites. Even though such target sites can be found in eukaryotic promoters these sites might not be optimal for imposing regulation on any given endogenous gene since they may be occupied by cellular proteins that bind to GC-rich
sequences, like Sp1 (44) or they may be modified by CpG methylation (45). Sp1 is a natural Cys$_2$-His$_2$ general transcription factor that binds to GC-boxes (consensus sequence 5’-G/TGGGCGG(G/A) (G/A)(C/T)-3’) frequently found in multiple copies in TATA-less housekeeping genes, but also in TATA-box containing promoters. The development of 5’-ANN-3’ recognition domains greatly expands the targeting potential of zinc finger domains as illustrated in Fig. 8 for the erbB-2 and erbB-3 promoters. Target sites with a 5’-(G/ANN) 6-3’ composition are highly abundant and present every 64 nucleotides considering a random base composition of DNA. The tremendous number of possible target sites that are found within the natural erbB-2 and erbB-3 promoters reflect this fact. Thus an expanded basis set of predefined zinc finger domains vastly expands the DNA targeting potential of this class of proteins.

Here we have demonstrated that our 5’-ANN-3’ recognition domains are useful modules for the construction of artificial transcription factors. While we have shown that these domains can be incorporated into 6-finger proteins and are compatible with our 5’-GNN-3’ binding domains in the construction of mosaic proteins, not all of the resulting transcription factors were equally potent in gene regulation. The 6-finger protein pAart was shown to specifically activate a luciferase reporter gene under the control of a minimal promoter containing 5xAart-binding sites (Fig. 5). In addition to transgene regulation, we have demonstrated that proteins containing these domains are suitable for the construction of transcription factors that regulate endogenous human genes. The mosaic transcription factors based on pE3Y specifically up- and downregulated the endogenous human erbB-3 gene (Fig. 6) with an efficiency analogous to the pE3 transcription factor constructed entirely from 5’-GNN-3’ binding domains (2). In contrast, pE2X and pE3Z based transcription factors showed a weaker level of transcriptional regulation or no regulation at all. Affinity seems to play a role in the ability of these artificial transcription factors to impose regulation on endogenous genes. Our previously reported 6-finger proteins pE2C and pE3 recognized 18 bp sites in their respective target genes that differed by only 3 bp. These proteins bound to their target DNAs with affinities of 0.75 nM and 0.35 nM, respectively. However, while pE2C bound to the E3 site with an 11 nM affinity and pE3 to E2C with a 10 nM
affinity they specifically regulated only their respective target genes. In addition, while pE2C was capable of altering gene expression its component 3-finger proteins bound DNA with approximately 50-fold reduced affinity and failed to regulate the endogenous gene. The results of the present study are consistent with our previous report implying an affinity threshold of approximately 10 nM for imposing gene regulation by binding to a single site (2). This might explain why pE3Y binding with an affinity of 8 nM was capable of altering expression of the \(erbB-3\) gene, while pE2X with an affinity of 15 nM showed less efficient regulation of \(erbB-2\).

On the other hand, affinity does not explain why pE3Z was not capable of altering gene expression since it bound its target site with an affinity of 2nM. In addition to affinity, competition of the artificial transcription factor with cellular proteins that bind to the same or an overlapping site, as well as other factors such as accessibility within chromatin, DNA methylation, the strength of the effector domain, and positioning of the binding site within the promoter need to be considered.

A recent report has demonstrated that not all sites within the chromosomal erythropoietin locus are accessible for the binding of artificial transcription factors due to steric constraints imposed by chromatin (3). Chromatin accessible sites as identified by Dnase hypersensitivity studies were suggested to be optimal targets for designed transcription factors. Thus chromatin accessibility may also limit the function of the pE3Z based transcription factors. These results suggest that it may be necessary to investigate several transcription factors for their ability to regulate gene expression in the absence of a detailed description of the gene’s chromatin structure. Since 5’-(GNN)\(_6\)-3’ DNA sequences might not always be present within chromatin accessible sites, the ability to target a wide range of DNA sequences should facilitate this goal.

In conclusion, the results of the characterization of the zinc finger domains reported in this study that bind the 5’-ANN-3’ family of DNA subsites is consistent with the overall view that there is no simple general recognition code that makes the direct rational design of
additional sequence specific domains possible. However, phage display selections can be applied and pre-defined and modular zinc finger recognition domains can be developed that allow for the rapid construction of artificial transcription factors. The zinc finger domains described here for recognition of 5’-ANN-3’ DNA sequences contribute greatly to the number of potentially genome-specific 6-finger proteins that can now be constructed. More than 10⁹ transcription factors with a 5’-(G/ANN)₆-3’ specificity can be rapidly prepared using predefined 5’-GNN-3’ and 5’-ANN-3’ recognition domains. We have demonstrated that the domains identified here are functionally modular and compatible with our previously disclosed 5’-GNN-3’ recognition domains. These pre-defined domains can be incorporated into polydactyl proteins capable of transgene and endogenous gene regulation. Progress towards the development of the remaining zinc finger DNA specificities is ongoing and will be reported in due course.
References

1. Beerli, R. R., Segal, D. J., Dreier, B., and Barbas III, C. F. (1998) *Proc Natl Acad Sci U S A* **95**(25), 14628-14633
2. Beerli, R. R., Dreier, B., and Barbas III, C. F. (2000) *Proc Natl Acad Sci U S A* **97**(4), 1495-1500
3. Zhang, L., Spratt, S. K., Liu, Q., Johnstone, B., Qi, H., Raschke, E. E., Jamieson, A. C., Rebar, E. J., Wolfe, A. P., and Case, C. C. (2000) *J. Biol. Chem.* **275**(43), 33850-33860
4. 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., Spratt, S. K., Case, C. C., and Wolfe, A. P. (2001) *J Biol Chem* in press
5. Miller, J., McLachlan, A. D., and Klug, A. (1985) *Embo J* **4**(6), 1609-14
6. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., and Wright, P. E. (1989) *Science* **245**(4918), 635-7
7. Segal, D. J., Dreier, B., Beerli, R. R., and Barbas III, C. F. (1999) *Proc Natl Acad Sci U S A* **96**(6), 2758-2763
8. Dreier, B., Segal, D. J., and Barbas III, C. F. (2000) *J. Mol. Biol.* **303**, 489-502
9. Beerli, R. R., Schopfer, U., Dreier, B., and Barbas III, C. F. (2000) *J. Biol. Chem.* **275**(42), 32617-32627
10. Pavletich, N. P., and Pabo, C. O. (1991) *Science* **252**(5007), 809-817
11. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure* **4**(10), 1171-1180
12. Pavletich, N. P., and Pabo, C. O. (1993) *Science (Washington, D. C., 1883-)* **261**(5129), 1701-7
13. Kim, C. A., and Berg, J. M. (1996) *Nat Struct Biol* **3**(11), 940-945
14. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Nature (London)* **366**(6454), 483-7
15. Houbaviy, H. B., Usheva, A., Shenk, T., and Burley, S. K. (1996) *Proc Natl Acad Sci U S A* **93**(24), 13577-82

16. Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) *J Mol Biol* **273**(1), 183-206

17. Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998) *Proc Natl Acad Sci U S A* **95**(6), 2938-2943

18. Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) *Structure* **6**(4), 451-464

19. Zweidler-McKay, P. A., Grimes, H. L., Flubacher, M. M., and Tsichlis, P. N. (1996) *Mol Cell. Biol.* **16**(8), 4024-4034

20. Hyde-DeRuyscher, R. P., Jennings, E., and Shenk, T. (1995) *Nucleic Acids Res.* **23**(21), 4457-4465

21. Gogos, J. A., Jin, J., Wan, H., Kokkinidis, M., and Kafatos, F. C. (1996) *PNAS* **93**, 2159-2164

22. Narayan, V. A., Kriwacki, R. W., and Caradonna, J. P. (1997) *J. Biol. Chem.* **272**(12), 7801-7809

23. Isalan, M., Choo, Y., and Klug, A. (1997) *Proc Natl Acad Sci U S A* **94**(11), 5617-5621

24. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (1999) *Annu. Rev. Biophys. Biomol. Struct.* **3**, 183-212

25. Segal, D. J., and Barbas, C. F. (2000) *Curr Opin Chem Biol* **4**(1), 34-39

26. Pabo, C. O., and Nekludova, L. (2000) *J. Mol. Biol.* **301**, 597-624

27. Wu, H., Yang, W.-P., and Barbas III, C. F. (1995) *PNAS* **92**, 344-348

28. Jamieson, A. C., Wang, H., and Kim, S.-H. (1996) *Proc Natl Acad Sci USA* **93**, 12834-12839

29. Isalan, M., Klug, A., and Choo, Y. (1998) *Biochemistry* **37**(35), 12026-12033

30. Greisman, H. A., and Pabo, C. O. (1997) *Science* **275**(5300), 657-661

31. Wolfe, S. A., Greisman, H. A., Ramm, E. I., and Pabo, C. O. (1999) *J Mol Biol* **285**(5), 1917-1934
32. Rader, C., and Barbas III, C. F. (1997) *Curr. Opin. Biotechnol.* **8**(4), 503-508
33. Barbas III, C. F., Kang, A. S., Learner, R. A., and Benkovic, S. J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7978-7982
34. Barbas III, C. F., and Lerner, R. A. (1991) *Methods: Companion Methods Enzymol.* **2**(2), 119-124
35. Liu, Q., Segal, D. J., Ghiara, J. B., and Barbas III, C. F. (1997) *Proc Natl Acad Sci U S A* **94**, 5525-5530
36. Liu, X., Sun, Y., Constantinescu, S. N., Karam, E., Weinberg, R. A., and Lodish, H. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10669-10674
37. Harwerth, I. M., Wels, W., Marte, B. M., and Hynes, N. E. (1992) *J. Biol. Chem.* **267**(21), 15160-15167
38. Seipel, K., Georgiev, O., and Schaffner, W. (1992) *EMBO J.* **11**(13), 4961-4968
39. Margolin, J. F., Friedman, J. R., Meyer, W. K.-H., Vissing, H., Thiesen, H.-J., and Rauscher III, F. J. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4509-4513
40. Corbi, N., Perez, M., Maione, R., and Passananti, C. (1997) *FEBS* **417**, 71-74
41. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) *Nat. Struct. Biol.* **4**(8), 605-608
42. Nardelli, J., Gibson, T., and Charnay, P. (1992) *Nucleic Acids Res.* **20**(16), 4137-44
43. Choo, Y., and Isalan, M. (2000) *Curr. opin. Struct. Biol.* **10**, 411-416
44. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51**, 1079-1090
45. Cedar, H. (1988) *Cell* **53**, 3-4
Footnotes

Abbreviations: bp(s), basepair(s); MBP, maltose-binding protein; ELISA enzyme-linked immunosorbant assay; KRAB, Krüppel-associated box; UTR, untranslated region.

Acknowledgments: We thank M. Elia and D. Valente for their technical assistance, N. Hynes for providing us with the ErbB-2-specific antibody. This work was supported by National Institutes of Health Grant CA86258, Novartis, and the Skaggs Institute for Chemical Biology (to C.F.B.). BD was recipient of a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft.
**Figure legends**

**Fig.1. Construction of the zinc finger phage display library.** A: Finger 3 of the C7 protein (27), recognizing the DNA subsite 5’-GCG-3’, was exchanged with a domain previously characterized to bind 5’-GAT-3’ (7). This new protein scaffold (C7.GAT) was used for the randomization of finger 2. Solid arrows: interactions of the amino acid residues of the α-helices with the nucleotides of their binding site as determined by x-ray crystallography of Zif268 (10,11). Dotted lines: Proposed interactions. B: Multi-target specificity ELISA for the C7 and C7.GAT protein; black bars: target sites of the type 5’-GCG-GNN-GCG-3’ (C7) or 5’-GAT-TGG-GCG-3’ (C7.GAT); gray bar: 5’-GCG-TGG-GCG-3’; white bars: evaluation of the 5’ recognition of finger 2 against each mixture containing 16 5’-XNN-3’ subsites where X represents 5’-adenine, 5’-cytosine, 5’-guanine, or 5’-thymine, respectively. Affinities of the proteins to their target site are given in the right upper corner of each graph.

**Fig.2. Amino acid sequences of finger-2 recognition helices from selected clones.** For each DNA target site several single clones were sequenced after the sixth round of panning and the amino acid determined to evaluate the selection. The DNA recognition subsite of finger 2 is shown on the left of each set, followed by the number of each occurrence. The position of the amino acid residues within the α-helix is shown at the top (residues likely to make direct contact with DNA are bold). Clones with the aminoacid sequence (boxed) were studied in detail and represent the best binders of each set. *: additional clones analyzed.

**Fig.3. Multitarget specificity assay to study DNA-binding properties of selected domains.** At the top of each graph 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 (pNNN) is indicated. Black bars represent binding to target oligonucleotides with different finger-2 subsites: AAA, AAC, AAG, AAT, ACA, ACC, ACG, ACT, AGA, AGC, AGT, ATA, ATC, ATG, and ATT. White bars represent
binding to a set of oligonucleotides where the finger-2 subsite only differs in the 5’ position, for example for the domain binding the 5’-AAA-3’ subsite (Fig. 3a) AAA, CAA, GAA, or TAA 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. 2. Middle panel: proteins marked with an asteriks in Fig. 2. Lower panel: proteins containing a finger-2 domain which was generated by site-directed mutagenesis (pmNNN).

**Fig. 4.** Construction of six-finger proteins containing domains recognizing 5’-ANN-3’ DNA sequences and ELISA analysis. A: The six-finger proteins pAart, pE2X, pE3Y and pE3Z were constructed using the Sp1C framework. Amino acid residues in position -1 to 6 of the α-recognition helix are given for each finger that was utilized. B: Proteins were expressed in *E. coli* as MBP fusion proteins. Specificity of binding was analyzed by measurement of the binding activity from crude lysates to immobilized biotinylated oligonucleotides (E2X, 5’-ACC-GGA-GAA-ACC-AGG-GGA-3’; E3Y, 5’-ATC-GAG-GCA-AGA-GCC-ACC-3’; E3Z, 5’-GCC-GCA-GCA-GCC-ACC-AAT-3’; Aart, 5’-ATG-TAG-AGA-AAA-ACC-AGG-3’). Assays were performed in duplicates, error bars representing the standard deviation. Black bars: pE2X; striped bars: pE3Y; Gray bars: pE3Y; white bars: pAart.

**Fig. 5.** Luciferase reporter assay. HeLa cells were cotransfected with the indicated zinc finger expression plasmid (pcDNA3 as control) and a reporter plasmid containing a luciferase gene under the control of a minimal promoter with TATA-box and zinc finger- binding sites (A: 5 x Aart binding site; B: 6 x 2C7 binding sites). Luciferase activity in cell extracts was measured 48h after transfection. Each bar represents the mean value (+/- standard deviation) of duplicate measurements. Y-axis: light units divided by 10^3. X-axis: constructs coding for zinc finger proteins transfected; control: reporter alone.
Fig. 6. Retrovirus-mediated gene targeting. A431 cells were infected with retrovirus encoding for pE2X (A) or pE3Y (B) fused to either the activation domain VP64 or repression domain KRAB. Three days later, intact cells were stained with the ErbB-1-specific mAb EGFR-1, the ErbB-2-specific mAb FSP77, or the ErbB-3 specific mAb SGP1 in combination with phycoerythrin-labeled secondary antibody. Dotted lines: control staining (primary antibody omitted); dashed lines: specific staining of mock-infected cells; dotted/dashed lines: cells expressing zinc finger protein-VP64 fusions; solid lines: cells expressing zinc finger protein-KRAB fusions.

Fig. 7. Computer models of finger-2 domains explaining novel interactions. Select oxygen (red), nitrogen (blue) and phosphate (purple) atoms are colored for clarity. Green dotted lines indicate suggested hydrogen bonds. Shells surrounding methyl groups indicate van der Waals radii. The sequence of each helix, the DNA subsite, and proposed interactions are summarized below each model. Green lines indicate hydrogen bonds. Arrows indicate hydrogen acceptors, VDW van der Waals interaction. a): Finger 2 of Zif268 (10,11). Hydrogen bond between H⁷ and the phosphate in the DNA backbone is indicated. b) and d) are opposite axial views of c), and e) axial view of f).

Fig. 8. Frequency of 6-finger target sites in the human erbB-2 and erbB-3 promoters. Displayed are 1300 to 1400 bp of the promoter regions including the 5’ UTR. These were searched for 5’-(GNN)₆-3’ (indicated by white circles) and 5’-(RNN)₆-3’ sites (R = adenine or guanine; indicated by gray circles). Light gray triangles: transcription start; dark gray triangles: translation start.
A

FINGER 1     FINGER 2     FINGER 3

Zif 268

\[ -1 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \]

\[
\begin{array}{c}
\text{RSDELTR} \\
3' - \text{GC} \text{G} - \text{GGT} \\
5' - \text{CG} \text{C} - \text{CCA} \\
\end{array}
\]

\[ -1 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \]

\[
\begin{array}{c}
\text{RSDLKT} \\
3' - \text{GC} \text{G} - \text{GGT} \\
5' - \text{CG} \text{C} - \text{CCA} \\
\end{array}
\]

\[ -1 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \]

\[
\begin{array}{c}
\text{RSDELKR} \\
3' - \text{GC} \text{G} - \text{GGT} \\
5' - \text{CG} \text{C} - \text{CCA} \\
\end{array}
\]

C7

\[
\begin{array}{c}
\text{KSADLKR} \\
3' - \text{GC} \text{G} - \text{GGT} \\
5' - \text{CG} \text{C} - \text{CCA} \\
\end{array}
\]

C7.GAT

\[
\begin{array}{c}
\text{KSADLKR} \\
3' - \text{GC} \text{G} - \text{GGT} - \text{TAG} \\
5' - \text{CG} \text{C} - \text{CCA} - \text{ATC} \\
\end{array}
\]
| Sequence | Amino Acid | Sequence | Amino Acid | Sequence | Amino Acid | Sequence | Amino Acid | Sequence | Amino Acid |
|----------|------------|----------|------------|----------|------------|----------|------------|----------|------------|
| AAA      | S T N T K L H A | AAC      | S R K D N L K N | AAG      | S R S D T L S N | AAT      | S T T G N L T V |
|          | 1         | 2         | 3         | 4         | 5         | 6         | 1         | 2         | 3         | 4         | 5         | 6         |
|          | 1         | 2         | 3         | 4         | 5         | 6         | 1         | 2         | 3         | 4         | 5         | 6         |
| ACA      | S S P A D L T R | ACC      | S D K K D L T R | ACG      | S R T D T L R D | ACT      | S T R T D L L R |
|          | 1         | 2         | 3         | 4         | 5         | 6         | 1         | 2         | 3         | 4         | 5         | 6         |
| AGA      | S R S D H L T N | AGC      | S D A S H L H T | AGG      | S R S D H L A E | AGT      | S H R T T L L N |
|          | 1         | 2         | 3         | 4         | 5         | 6         | 1         | 2         | 3         | 4         | 5         | 6         |
| ATA      | S Q A S S L K A | ATC      | S R R S A C R R | ATG      | S R R D A L N V | ATT      | S T S H G L T T |
|          | 1         | 2         | 3         | 4         | 5         | 6         | 1         | 2         | 3         | 4         | 5         | 6         |
|       | F 1          | F 2          | F 3          | F 4          | F 5          | F 6          |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|
| pAart | RRD-A-LNV    | RED-N-LHT    | QLA-H-LRA    | QRA-N-LRA    | DKK-D-LTR    | RSD-H-LAE    |
|       | 3’-GTA-5’    | 3’-GAT-5’    | 3’-AGA-5’    | 3’-AAA-5’    | 3’-CCA-5’    | 3’-GGA-5’    |
| pE2x  | DKK-D-LTR    | QSS-H-LVR    | QSS-N-LVR    | DKK-D-LTR    | RSD-H-LTN    | QSS-H-LVR    |
|       | 3’-CCA-5’    | 3’-AGG-5’    | 3’-AAG-5’    | 3’-CCA-5’    | 3’-GGA-5’    | 3’-AGG-5’    |
| pE3Y  | DPG-A-LRV    | RSD-N-LVR    | QSG-D-LRR    | QLA-H-LRA    | DCR-D-LAR    | DKK-D-LTR    |
|       | 3’-CTA-5’    | 3’-GAG-5’    | 3’-ACG-5’    | 3’-AGA-5’    | 3’-CCG-5’    | 3’-CCA-5’    |
| pE3Z  | DCR-D-LAR    | QSG-D-LRR    | QSG-D-LRR    | DCR-D-LAR    | DKK-D-LTR    | TTG-N-LTV    |
|       | 3’-CCG-5’    | 3’-ACG-5’    | 3’-ACG-5’    | 3’-CCG-5’    | 3’-CCA-5’    | 3’-TAA-5’    |
A

Luciferase activity (x 1000 light units)

control  pcDNA  Aart  Aart-VP64  2C7  2C7-VP64

B

Luciferase activity (x 1000 light units)

control  pcDNA  Aart  Aart-VP64  2C7  2C7-VP64
erbB-2 promoter: 6xGNN: 11
erbB-3 promoter: 6xGNN: 7
