

Designing Transcription Factor Architectures For Drug Discovery

Pilar Blancafort, David J. Segal and Carlos F. Barbas III

Department of Molecular Biology, The Scripps Research institute, 10550 North
Torrey Pines Road, La Jolla, California 92037; P.B. & C.F.B.

Department of Pharmacology and Toxicology, University of Arizona College of
Pharmacy, 1703 East Mabel Street, Tucson, Arizona, 85721; D.S.

Running Title: Designer Transcription Factors

Minireview

Corresponding author: Carlos F. Barbas III. Department of Molecular Biology, The Scripps Research institute, 10550 North Torrey Pines Road, La Jolla, California 92037 Phone: (85) 784-9098; Fax: 784-2583; e-mail: carlos@scripps.edu

Abbreviations: siRNA, small interfering RNAs.

Pages

2 Tables

4 Figures

100 References

Words

Abstract

Recent advances in the design, selection, and engineering of DNA binding proteins have led to the emerging field of designer transcription factors (TFs). Modular DNA-binding protein domains can be assembled to recognize a given sequence of a DNA in a regulatory region of a targeted gene. TFs can be readily prepared by linking the DNA-binding protein to a variety of effector domains (ED) that mediate transcriptional activation or repression. Furthermore, the interaction between the TF and the genomic DNA can be regulated by several approaches, including chemical regulation by a variety of small molecules. Genome-wide single target specificity has been demonstrated using arrays of sequence-specific zinc finger (ZF) domains, polydactyl proteins. Any laboratory today can easily construct polydactyl ZF proteins by linkage of predefined ZF units that recognize specific triplets of DNA. The potential of this technology to alter the transcription of specific genes, to discover new genes, and to induce phenotypes in cells and organisms is now being applied in the areas of molecular therapeutics, pharmacology, biotechnology and functional genomics.

The anatomy of a transcription factor

In complex organisms phenotypic diversity is primarily achieved by transcriptional and post-transcriptional regulation of gene expression. Large families of transcription factors (TFs) are responsible for the regulation of specific genes at the proper time, developmental stage, and tissue location. Furthermore, TFs orchestrate regulatory networks that ultimately dictate complex phenotypic programs (Elkon et al., 2003).

TFs are multidomain proteins typically composed of a DNA binding domain (DBD), responsible for specific contacts with DNA bases, and an effector domain (ED) that mediates activation or repression of targeted genes. Some TFs contain additional postranscriptional regulatory elements, such as dimerization domains and phosphorylation sites (Ciarapica et al. 2003). TFs exert their action by binding to specific DNA sequences in chromatin and recruiting appropriate global co-activator and co-repressor regulatory complexes. TF activator complexes include Mediator, which interacts with core promoter factors P/CAF and CBP/p300 that contain histone acetyltransferases that modify nucleosomes to a transcriptionally active state, and the SWI/SNF chromatin remodeling complex, which modifies the position of nucleosomes enabling additional TF binding (Gebuhr et al., 2003). Examples of TF repressors are Sin3-HDAC and NuRD, which contain histone deacetylases that modify nucleosomes to a transcriptionally inactive state (Ansari and Mapp, 2002).

Given their pivotal role in controlling cell fate, aberrant expression or incorrect processing of TFs contributes to the progression of a variety of diseases, including developmental abnormalities and cancer. The TF p53 is the most commonly mutated gene in human cancer (Harms et al., 2004). Several chromosomal translocations in acute myeloid leukemia (AML) generate chimeric TFs by linking a DBD from one TF with a repression domain of another; the chimera triggers abnormal target gene regulation (Steffen et al., 2003). Another example is the altered regulation of the STAT (signal transducer and activator of transcription) family of TFs. Constitutive activity of STAT proteins or expression of C-terminal mutated STATs, particularly STAT3 and STAT5, contribute to malignancy and cellular transformation (Benekli et al., 2003).

Artificial TF design

Given the ability of TFs to regulate genes in a sequence-specific manner, an enormous effort has been devoted to engineering artificial TFs that are able to bind and regulate specific target genes. Like natural TFs, artificial TFs are composed of a DBD that can recognize a specific DNA sequence (typically near the transcription start site of the targeted gene) and an ED that mediates transcriptional activation or repression (Fig. 1). Activation EDs that have been used on artificial TFs include the herpes simplex virus VP16 (Sadowski et al., 1988), the engineered VP64 (Beerli et al., 1998), and the NF- κ B subunit p65 (Liu et al., 2001). Repression EDs have included KRAB (Krüppel associated box; Margolin et al., 1994), SID (mSin3 interaction domain; Ayer et al., 1996), ERD

(ERF repressor domain; Sgouras et al., 1995) and HMT (histone methyltransferase; Snowden et al., 2002).

Various scaffold molecules have been used for the generation of DBDs. Specific DNA recognition has been successfully achieved using several synthetic approaches: polyamides, triple-helix forming oligonucleotides (TFO) and peptide nucleic acids (PNA) (reviewed by Uil et al., 2003). These DBDs are connected to short activation or repression domains via flexible or rigid linkers (Arora et al., 2002). The advantage of these synthetic approaches is the small size of the TF, which can facilitate both synthesis and cellular uptake. In addition, synthetic DBDs have demonstrated a high level of affinity, permitting not only regulation of targeted promoters but also specific competition with endogenous TFs (Bremer et al., 2001; Chiang et al., 2000; Ehley et al., 2002; Coull et al., 2002; Wurtz et al., 2002; Stanojevic and Young, 2002; Yang et al., 2003; Fechter and Dervan, 2003). Recently, DNA microarray experiments have shown that polyamides seem to be able to regulate a limited number of genes in lymphoid cells (Dudouet et al., 2003).

Using protein scaffolds: Zinc Finger Domains

The classical protein scaffold used for targeting gene expression is the C₂H₂ zinc finger (ZF) domain. It is estimated that the human genome encodes more than 900 C₂H₂ ZF proteins (Tupler et al., 2001; Venter et al., 2001) In a C₂H₂ ZF domain, an α -helix is packed against two anti-parallel β -strands and additional stability is provided by the coordination of a zinc ion by the side-chains

of two cysteine and two of histidine residues (Miller et al., 1985). Amino acids in the N-terminus of the α -helix make specific contacts with DNA bases in the major groove. ZF domains have been useful for the construction of specific DNA binding proteins primarily because of two properties: sequence specificity and modularity (Fig. 2A). The structure of the Zif268-DNA complex (Pavletich and Pabo, 1991; Elrod-Erickson et al., 1996) revealed that the ZF domains interact primarily with three base pairs of DNA (called the recognition triplet). Each ZF interacted with the DNA using the same contact positions in a quasi-independent mode. The residue at position +6 in the ZF helix interacted with the 5' base of the DNA recognition triplet, residue +3 contacted the middle base of the triplet, and the residue at helical position -1 (just before the start of the α -helix) contacted the 3' base. These base contacts are made with only one strand of the DNA duplex. A cross-strand contact involves position +2 in the α -helix (Asp²) and a cytosine or adenine base in the adjacent complementary triplet on the opposite DNA strand. This interaction has been shown to restrict the modularity of this family of ZF proteins (Isalan et al., 1997).

Isolation of sequence-specific ZFs

Because of the simplicity of ZF-DNA contacts, many laboratories have searched for the molecular rules governing the specificity of the interactions between ZFs and DNA. The ultimate goal of many of these studies was to design specific DNA-binding proteins that could bind desired genomic sequences and regulate endogenous genes. Phage display has represented a pivotal tool for the

selection of ZF helices able to bind defined DNA triplets. In early experiments, the three-ZF protein Zif268 was displayed on the surface of the filamentous bacteriophage. The middle helix of Zif268 was randomized to create library of ZF2s and the flanking ZF1 and ZF3 were unchanged to anchor the protein appropriately on the DNA. Phage display experiments were performed to select ZF2 helices from the library that were able to interact specifically with a new DNA triplet. The results of these experiments proved that a correlation could be established between the nature of the bases of the DNA and the identity of the residue selected at the contact positions -1, +3 and +6 (Rebar and Pabo, 1993; Jamieson et al., 1994; Wu et al., 1995; Choo and Klug, 1994; Segal et al., 1999; Dreier et al., 2000). These correlations were rationalized according to the structure of the Zif268-DNA complex. Molecular modeling, mutagenesis experiments, and structural analysis of these novel ZF proteins-DNA complexes provided a collection of ZF domains able to specifically recognize a wide range of different DNA triplets (Dreier et al., 2000; 2001; Elrod-Erickson et al., 1998).

Specific ZF sequences were obtained that displayed good specificity for purine-rich triplets (Segal et al., 1999; Dreier et al., 2000; 2001). However, selection of ZF sequences capable of binding specifically to C- or T-containing triplets, especially those with C or T as the 5' nucleotide, has been more challenging. This is due partly to the fact that purine bases (G and A) offer more hydrogen-bonding possibilities than pyrimidines, and partly because the amino acids that would recognize C and T typically have short side chains and thus cannot easily span the distance from position 6 in the α -helix to the 5' DNA base.

Phage display methodology has also been applied to select for domains or proteins that recognize specialized nucleic acid structures, such as methylated DNA (Choo, 1998), quadruplex DNA (Isalan et al., 2001b), and noncanonical duplex RNA (Blancafort et al., 1999). However, proteins recognizing such exotic structures have yet to find utility comparable to that of their duplex-DNA-binding counterparts.

Other methodologies to screen for novel DNA binding ZF domains take advantage of the yeast one-hybrid system, as shown for the cell-based selection of ZFs that bind sequences in the MDR-1 promoter (Cheng et al., 1997; Bartsevich and Juliano, 2000). A cell-based ZF selection system was also established in bacteria to optimize multi-finger proteins (Hurt et al., 2003). The ZF library used in these studies combined cassette mutagenesis in the ZF helix followed by domain shuffling. Specific ZF proteins were selected in a bacterial two-hybrid system.

Building Polydactyl ZF proteins

The engineering or isolation of sequence-specific ZF domains led to a variety of strategies for building multi-modular ZF proteins. In principle, the modified ZFs could be assembled in modular tandem arrays, much like naturally-occurring ZF proteins. However, concerns about modularity led some to explore methods for selecting domains directly in a multi-finger context. For example, Pabo and collaborators devised a strategy for sequential selection of ZF domains (Greisman and Pabo, 1997). This approach was designed to overcome the

contact overlap between ZF domains involving residue +2 described earlier for Zif268. The technique involved consecutive library construction and phage display against a predefined nine base-pair (bp) sequence, optimizing one ZF at a time. Alternatively, Isalan et al. (2001a) developed a bipartite complementary method. In this strategy, two phage display libraries were used to select for ZFs binding five bps. The ZF proteins from these two selections were recombined to generate a single protein recognizing nine bps.

In contrast, we have adopted a helix grafting strategy, based on the modular property of ZF proteins (Fig. 2B and 2C; Beerli et al., 1998). With the caveat that an overlap contact would need to be accommodated between certain subsets of ZF units, multi-finger proteins were constructed by replacing or “grafting” the helix regions of modified ZFs onto the scaffold of a highly regular, existing zinc finger protein (Sp1C; Shi and Berg, 1995). The use of a highly regular scaffold ensured that each domain would be displayed on the protein in the same way, allowing assembly of the modified ZF domains in nearly any order. This strategy provided an extremely rapid methodology for construction of polydactyl ZF proteins without the need for phage display and selection for each new DNA target. Recently Segal et al. (2003) have used this modular strategy to construct more than 80 engineered three-ZF proteins and have shown that these novel proteins were able to interact with their predicted DNA binding sites. Nagaoka et al. (2002) have used helix grafting to change the specificity of the SP1 protein (that naturally binds GC-rich regions) to recognize an AT-rich

element. The helices used for grafting were derived from the *Drosophila* CF2-II protein, which recognizes AT-rich sequences.

By increasing the number of ZF units in the multi-finger protein, the number of bases targeted can be expanded. It should be possible to target low-frequency, potentially unique, sites in the human genome using six or more ZF units (Liu et al., 1997). Barbas and collaborators have constructed polydactyl six-ZF proteins using the five amino acid (aa) canonical linker (TGEKP) between the ZF units (Liu et al., 1997; Segal and Barbas, 2001; Beerli and Barbas, 2002). Kim and Pabo (1998) reported the construction of a six-ZF protein with high specificity and affinity made by joining two three-ZFs (2 x 3ZFs) with a longer, 9-aa linker. The increased affinity and specificity provided by the longer linker was attributed to an increased flexibility in the six-ZF-DNA complex. Indeed, it has been suggested that smaller linkers could generate a loss of entropy resulting in loss of affinity (Peisach and Pabo, 2003). Moore et al. (2001) described the construction of highly specific six-ZF proteins built by three groups of two-ZF units (3 x 2ZFs). The authors modified the linker sequence between the two-ZF units by insertion of additional Gly or Ser residues into the canonical linker sequence.

Regulating TF expression

Several groups have engineered artificial TFs whose expression was tightly regulated, thereby inducing targeted gene expression in a controlled manner. Beerli et al. (2000b) fused ZF domains with the modified ligand binding

domains of steroid hormone receptors (estrogen, progesterone, or ecdysone). Upon ligand binding, hormone receptors dissociated from an inactive complex in the cytoplasm, entered the nucleus, dimerized and bound DNA. The engineered TF regulators were chemically induced by the drugs 4-hydroxytamoxifen (ZF-estrogen receptor fusions), RU486 (ZF-progesterone receptor fusions) or Ponasterone A (ZF-ecdysone receptor fusions), and were shown to activate reporter constructs up to three orders of magnitude. An inducible system was constructed by Pollock et al. (2002) to regulate the endogenous VEGF-A gene. The TF was induced by an analog of the small molecule rapamycin. This compound was able to reconstitute an active TF by inducing dimerization of separate DBD and ED subunits. The rapamycin analogue was able to activate endogenous gene expression in a dose-dependent manner. In another strategy, Lin et al. (2003) established a screen to select for synthetic small molecules able to regulate DNA binding of ZF-based TFs. The authors used a ZF protein with mutations at His125 and Phe116, which are involved in zinc coordination. The mutations disrupted the ZF structure, creating a cavity, and also impaired DNA binding. Several heterocycle-containing compounds were able to rescue the mutations and activate a reporter construct up to 100 fold.

From genes to phenotypes: towards regulating endogenous gene expression using ZF-based TFs

The newly designed, artificial TFs were next applied to regulate endogenous promoters and modify gene expression. Several genes have been

successfully regulated using designed polydactyl ZF proteins in many different organisms (Table 1). Applications of these artificial TFs in areas such as gene therapy, pharmacology, biotechnology, and functional genomics will be described in the following sections. Fig. 3 summarizes the strategic uses of artificial TFs and their possible applications in gene therapy and pharmacology.

Artificial Transcription Factors regulating specific drug-target and disease genes

Many diseases originate because essential genes are mutated, inactivated, or aberrantly expressed. TFs can potentially be used as therapeutic tools to regulate transcriptional levels of genes associated with disease. In addition, they can be used as molecular tools to verify gene function and, therefore, to validate target genes for drug design. As described in some of the applications below, specific TFs directed against different target genes can ascertain whether or not these genes are related functionally and participate in the same functional pathway.

Beerli and coworkers (Beerli et al., 1998; Beerli et al., 2000a) have constructed two six-ZF proteins able to specifically regulate the *erbB-2* and *erbB-3* proto-oncogenes in several cancer cell lines. These oncogenes are overexpressed in a majority of breast cancer tumors and play an essential role in regulating proliferation of breast cancer cells. The two ZF proteins targeted two highly related DNA sequences (15 of 18 bp identity) in the 5'UTR region of *erbB-2* and *erbB-3*. Independent regulation of one gene but not the other

demonstrated that the designed ZF proteins were able to regulate their endogenous target genes with high degree of specificity. These genes were up-regulated by attaching the VP64 activation domain, and down-regulated by linking a transcriptional repression domain, KRAB. Importantly, cancer cell lines expressing these regulators by retroviral delivery recapitulated the cell cycle alterations induced by gain-of-function or loss-of-function of the *erbB-2* and *erbB-3* oncogenes (Beerli et al. 1998; 2000a). Holbro et al. (2003) used these artificial TFs to demonstrate the essential role of *erbB-3* in conjunction with *erbB-2* to regulate breast tumor cell proliferation.

Corbi et al. (2000) constructed a designed TF able to bind and activate a transgene of the Utrophin gene promoter. Up-regulation of this gene would be therapeutic treatment for Duchene Muscular Dystrophy (DMD). Other examples of disease genes targeted with TFs are IGF2 and H19, involved in cancer and Beckwin-Wiedemann syndrome, respectively (Jouvenot et al., 2003). These genes are silenced by natural mechanisms of imprinting in a disease stage, but were reactivated by an artificial TF.

In mammalian cells, several designed three-ZF proteins have been directed to regulate genes controlling angiogenesis. Angiogenesis is the process of new blood vessel formation, which is critical for tumor development. Consequently, these genes have become attractive targets for therapeutic regulation. Recently, ZF-based TFs have been targeted to the promoter of vascular endothelial growth factor-A (VEGF-A; Liu et al., 2001; Rebar et al., 2002). These proteins were able to activate expression of the endogenous gene,

induce angiogenesis, and accelerate wound healing in mouse models. Interestingly, the new vasculature induced by the TFs was not hyperpermeable, a trait not observed after simple cDNA delivery of the gene. These results demonstrated that artificial TFs could efficiently generate physiological effects in the context of the whole organism. Efficient repression of VEGF-A by artificial TFs was recently demonstrated by Snowden et al. (2002; 2003). In these studies, engineered ZFs recognizing the VEGF-A promoter were linked to a minimal histone methyltransferase domain. The authors showed that the ZF-directed local methylation of histone H3 in cells triggered gene repression. The TFs were able to repress the gene in a highly tumorigenic cell line to the levels comparable to a non-angiogenic, low tumorigenic cell line.

Bartsevich et al. (2000) selectively down-regulated the MDR1 multidrug resistance gene with an artificial TF. In another recent report, Falke and Juliano targeted the pro-apoptotic Bax gene and showed that a designed five-ZF protein was able to induce apoptosis in p53 deficient cell lines (Falke et al., 2003). This suggests that designed ZF proteins may be used to induce apoptosis in cancer cells that have mutated or inactivated p53.

Tan et al. (2003) targeted CHK2, a key gene regulating cell cycle progression. This protein kinase phosphorylates several substrates, including the tumor suppressor protein p53. The authors targeted a six-ZF protein recognizing 18 bps in the promoter of the CHK2 gene. The artificial TF was able to repress specifically the CHK2 gene, as determined by DNA microarray experiments.

Importantly, the TF-induced repression elicited loss of phosphorylation of p53 in human cells.

In another recent report, Bartsevich et al. (2003) targeted the mouse Oct-4 gene, which is involved in differentiation of embryonic stem (ES) cells. TF technology could be used to regulate the cell fate of pluripotent stem cells, perhaps re-directing specific differentiation programs. These TF could be used as therapeutic tools to regulate tissue regeneration from stem cells.

Another important functional application of designed TFs is described by Ren et al. (2002). The authors used specific TFs targeted against two different promoters to identify the functionally relevant isoform for the gene PPAR, involved in adipogenesis.

Artificial TFs as anti-viral tools

Several groups have targeted viral replication with artificial TFs recognizing viral DNA sequences. Reynolds et al. (2003) reported the construction of TFs targeting Sp1 binding sites in the promoter of HIV. One TF was able to inhibit viral replication in engineered cancer cells by 75%. Segal et al. (2004) reported a TF capable of achieving 100-fold repression of transcription from the HIV promoter, as assessed by reporter assays. Furthermore, this TF was able to repress the replication of several HIV strains in the biologically relevant T cells and primary blood mononuclear cells (PBMC) with no observable cytotoxicity. Repression in primary human cells was maintained for an extended period. Papworth et al. (2003) designed TF repressors able to bind the Herpes

simplex virus 1 promoter. One six-ZF containing TF was able to inhibit the viral replication cycle and reduce the viral titer by 90%. These papers demonstrated the use of artificial TFs for inhibition of viral replication.

Targeting genes with DNA-modifying enzymatic domains

A growing number of studies have used ZF domains linked to enzymatic domains in order to direct modifications in specific sequences of DNA (Fig. 3A). Catalytic activities targeted with ZF include endonucleases (reviewed by Carroll, 2004), recombinases (reviewed by Collins et al., 2003), and integrases (Tan et al., 2004). Recently, chimeric recombinases combining mutant variants of Tn3 resolvase and the murine zinc finger protein Zif268 have been engineered in bacteria (Akopian et al., 2003). These chimeras were able to catalyze site-specific recombination mediated by Zif268 binding sites. Tan et al. (2004) presented in vitro studies showing that HIV integrase tethered to an engineered six-ZF protein could direct integration events to a 10-bp region immediately flanking the ZF binding site. These engineering projects are directed towards potential therapeutic modifications of disease genes in mammalian cells.

Regulating gene expression in plants

TF targeting in plants has been reviewed by Segal et al. (2003). TF regulation has been demonstrated in transgenic plant cells using a variety of reporter assays (Stege et al., 2002; Sanchez et al., 2002). In Arabidopsis, designed ZF proteins were able to alter genes involved in the formation of floral

organs, indicating that artificial ZF proteins could be used in plant biotechnology to induce complex plant phenotypes by altering transcription of specific genes (Guan et al., 2002). Microarray analysis indicated that only the single targeted *Arabidopsis* gene was modulated by the designed TF. Stable expression and transgene control over multiple generations has also been demonstrated in transgenic tobacco plants (Ordiz et al., 2002).

Another emerging application of TF technology is the production of proteins and pharmacologically active plant metabolites. TFs constitute new tools to increase the production of metabolites, such as flavonoids and alkaloids, by activating multiple enzymes involving in biosynthetic pathways and repressing others (Gandet and Memelink, 2002).

TF-based, genome-wide strategies to regulate gene expression: TF libraries for the modification of phenotypes and gene discovery

A persistent challenge in TF engineering has been the selection of appropriate genomic DNA target sequences that will enable potent transcriptional regulation. Typically, targeted DNA sequences are localized in close proximity to the transcription start site. Transactivation analyses of cloned promoters in reporter systems have shown a direct relationship between distance to the transcription start site and transactivation potential (Stege et al., 2002). However, in the context of a living cell, endogenous promoter sequences are packed into defined chromatin structures. The structure of chromatin in regulatory regions is controlled by chromatin remodeling factors. Moreover, a given targeted sequence

might not be accessible for TF interaction. Although the accessibility of a given promoter can be approximated by mapping of DNase I-accessible chromatin regions (Liu et al., 2001), these studies are cumbersome when high numbers of promoters must be regulated, for example, for genome-wide studies. Moreover, DNase I accessibility may be necessary, but is not sufficient, to identify a productive regulatory site (Zhang et al., 2000). Often, detailed knowledge of endogenous factors affecting transcription in cis-regulatory regions of a gene is limited. Promoters can also be modified epigenetically by specific methylation. As a consequence, a given promoter can be "transcriptionally open" in one cellular background but silent or inactive in a different cell line. Finally, genomic sequences possessing potent transcriptional regulatory capabilities might be located kilobases away from the transcription start site, in intergenic sequences, introns, or even in coding regions.

In order to functionally select genomic sequences that can be targeted by TFs and therefore used to efficiently modify endogenous transcription, methodology has been developed for screening combinatorial TF libraries in mammalian cells (Blancafert et al., 2003). Such TF libraries are composed of modified ZFs domains for every targetable 3-bp site, randomly assembled into three- and six-ZF TFs. When delivered into a population of mammalian cells, large TF libraries have the possibility to interact with many different regions of genomic DNA sequence: on the order of hundreds of unique potential binding sites per gene. TF library members "scan" the genome for accessible, transcriptionally open DNA sequences. A variety of assays can be applied to

identify cells displaying a phenotypic change, such as induced expression of a surface marker or altered cell morphology, resulting from the TF activation or repression of one or more genomic loci. TFs inducing the phenotype of interest can then be used as molecular probes to isolate relevant regulatory regions, to discover genes, and to provide insights into the co-regulation of genes in a given pathway. In this sense, TF libraries can be regarded as a functional genomics tool, linking functional regulatory sequences in complex genomes with cellular phenotypes. Barbas and co-workers have performed selections of TF libraries in several cancer cell lines in order to regulate genes crucial to tumor biology and tumor progression. Selections were performed by cell sorting using antibodies recognizing specific antigens that were differentially regulated on the surface of tumor cells. TFs have been isolated from TF libraries that specifically up- and down-regulate many important molecules, such proto-oncogenes *erbB-2*, angiogenic molecules such as CD144 (VE-cadherin; Blancafort et al., 2003) and cell adhesion molecules such as ICAM-1 (Magenat et al., 2004).

TFs selected from combinatorial libraries are able to regulate a given target gene directly (by interacting with the promoter) or indirectly (by regulating upstream genes controlling target gene transcription). In order to select TFs able to regulate directly the *Erb-B2* gene Lund et al. developed a novel phage display strategy to select for ZF proteins from combinatorial libraries binding the proximal *Erb-B2* promoter. The authors isolated TFs binding the promoter that were able to regulate the endogenous *Erb-B2* gene. (Lund et al., 2004).

Recently, our studies have isolated artificial TFs modulating complex phenotypes in cancer cells, such as cell growth, proliferation, resistance to drugs and metastasis (Blancafort, manuscript in preparation). These investigations have discovered and regulated genes involved in tumor progression. Therefore, artificial TFs have demonstrated their potential for therapeutic re-programming of cancer cell phenotypes.

Bae et al. (2003) have produced similar TF libraries by PCR-amplification of endogenous human ZFs. These TF libraries could be used to modulate cellular phenotypes, such as yeast drug resistance and mammalian cell differentiation. In combination with other genomic approaches, such as DNA microarray and chromatin immunoprecipitations, genomic-wide strategies could provide candidate genomic targets that are relevant for drug discovery in complex diseases such as tumor progression. Another functional application of combinatorial TF libraries was described by Lee et al. (2003). A TF library expression combined with cDNA microarray technology provided a tool to cluster and classify groups of genes that are actively transcribed in many different cellular backgrounds.

TFs versus RNA-based methodologies to regulate gene expression.

Several successful gene regulatory technologies that target RNA are in common use. RNA silencing is a novel regulatory mechanism involving either specific cleavage of mRNA (RNA interference or RNAi), translational repression or chromatin silencing (Morris, et al, 2004, Fukagawa et al, 2004). It is mediated

by small ~ 22 bp dsRNAs called MicroRNAs or small interfering RNAs (siRNAs). Since the discovery of the efficacy of this approach in *C. elegans* in 1998 (Fire et al., 1998), RNAi has become extremely popular and methodologies have been developed by which any laboratory can induce gene specific knock-downs using this technology. Use of siRNAs consisting of fewer than 30 nucleotides in length is preferred since these RNAs do not elicit interferon responses in many cell types and organisms (Agrawal et al, 2004). MicroRNAs are endogenously encoded, ~22-mer single-stranded RNA molecules that effect gene expression using many of the same protein components of the siRNA pathway. The main difference between miRNA and siRNA is that miRNA causes translational repression of the gene target without cleavage of the mRNA. This appears to occur because the miRNA binds to its gene target with imperfect complementarity(Bartel, 2004). Other types of antisense technologies uses single-stranded nucleic acids, typically modified RNA or RNA-DNA hybrids, to specifically base pair with target mRNAs, resulting in translational blockade or RNaseH-mediated degradation of the target (Crooke, 2004). In 1998, the antisense drug fomivirsen became the first gene regulatory agent to gain FDA approval. Many other antisense agents are in clinical trials.

Several interesting features can be compared between DNA-targeting artificial TFs and RNA-targeting methods such as RNAi and antisense, particularly regarding delivery, specificity and function (Table 2). Delivery remains a formidable obstacle to the use of these technologies in humans. TFs, siRNAs and antisense agents can be delivered to target cell types transiently, using

transfection reagents, or stably using retroviral, adenoviral or lentiviral vectors. Stability and half-life *in vivo* of antisense and siRNAs is a primary concern and can be improved by chemical synthesis using modified nucleotides. However, synthesis of such compounds can be expensive and many labs today use vector-derived transfections. For efficient gene knock-down, RNAi requires high expression levels that can be achieved with polymerase III promoters. TF proteins are not naturally taken up by cells so transient delivery requires transfection of TF-encoded cDNA. However, the stability of artificial TFs is comparable to natural occurring ones and since TFs act directly as transcriptional regulators, they do not require high-level expression to achieve biological effects. For all these technologies, the use of tissue specific promoters is perhaps the system of choice in order to express these artificial regulators in the proper target organ or tissue. However, additional mechanisms to control regulator function may be available to artificial TFs, such as activation by small molecule ligands as described earlier.

Specificity for both TFs and RNA strategies is achieved through base contacts. More base contacts generally provides better specificity. The upper limit of this reasoning is reached when the binding energy becomes so strong that a mismatched base contact can no longer sufficiently destabilize the binding complex. However in practice, pragmatic concerns usually govern the size of the binding site. For example, extending the number of ZF units can extend the number of specific TF contacts with the DNA. Since a 6ZF TF can potentially recognize a unique 18-bp site in the human genome, there is little practical

reason to exceed this binding site size. 6ZF TFs have been shown to have higher affinities and better discrimination than 3ZF TFs (Beerli et al., 1998; Blancafort et al. 2003; Lund et al. 2004), and certain designed 6ZF TFs were found to regulate only their single targeted genes based on microarray analysis (Guan et al., 2002; Tan et al. 2003). The specificity of siRNA is governed by Dicer and associated proteins that function optimally with ~22 bp molecules. Although a site of this length should provide unique targeting in the human genome, recent expression profiling has demonstrated off-target gene regulation with siRNA indicating that the full 22 bp specificity is not expressed (Jackson et al. 2003). It should be emphasized, however, that more studies of this type will be required for a proper evaluation of specificity for any of these regulatory methodologies, and investigators should be encouraged to perform such studies. As far as being able to actually build a regulator that can bind an optimal binding site using present-day technology, it might seem, *a priori*, that siRNA and antisense have an advantage. The spectrum of sequences that can be targeted by artificial TFs is somewhat limited by the existing lexicon of zinc finger domains. While the current technology is still sufficient to create over 1,000,000,000 proteins, with the potential to recognize a targetable sequence every 32 nucleotides, recognition of C- and T-rich sequences remains problematic. In the case of the RNA technologies, simple Watson-Crick base pairing rules allow recognition of any sequence. However, in practice, the primary technical barrier limiting the success of both the TFs and RNA technologies *in vivo* is not the number of targetable sites but the accessibility of those sites. Target sites may be blocked by

endogenous binding factors, such as RNA-binding proteins or DNA-binding nucleosomes. siRNA and antisense strategies are additionally susceptible to unfavorable 3-D structures, which occur far more frequently in RNA than DNA. As described earlier, a practical approach has been to construct regulatory agents to several target sites and determine empirically which function best. Combinatorial libraries of agents offer an alternative solution. Finally, it is instructive to consider that some antisense agents have been shown to exert non-sequence-dependant effects through interaction with other macromolecules (Khaled et al., 1996). This example should serve as a caveat to all investigators when considering how to evaluate specificity in their experiments.

The function of artificial TFs and RNA technologies differ significantly once they arrive at their specific target site. Most obviously, TFs target DNA sites, of which there are only two or fewer copies in the cell. In contrast, there will be many more copies of mRNA produced from each DNA gene. For highly expressed genes (for example oncogene overexpression in cancer cells originated by multiple gene duplications) RNAi might not eliminate the total population of target RNA and substantial protein product could elicit some residual phenotype. In this case, TFs and siRNAs might be contemplated as companion technologies that could work in synergy to down-regulate gene expression by both reducing the rate of RNA production and by increasing specific degradation. Another significant difference is that artificial TFs have the ability to both up- and down-regulate transcriptional levels of a given gene (depending on the effector domain), and thus either gain-of-function phenotypes

or loss-of-function phenotypes are accessible. In contrast, RNAi and antisense can only be applied to negatively regulate RNA levels, at least in the direct sense. This difference is important in the context of molecular therapeutics since transcriptional levels of a given gene can oscillate depending on the cell type and the disease stage. Therapeutic application or drug target validation studies may also require up-regulation rather than down-regulation of a particular gene.

Summary and Outlook

This review has described several approaches for design of artificial TFs able to regulate endogenous gene transcription. In the most basic TF design, a DBD, providing specific DNA recognition, is linked to an ED, responsible for gene activation or repression. Several groups have successfully used TFs composed of multimodular ZF domains to regulate cellular genes of interest to the fields of biotechnology and molecular therapeutics. This "explosion" of artificial TF regulators demonstrates the power of this technology in regulating transcription in diverse genomes from plants to humans. Compared to other methodologies for regulating gene expression, such as siRNA and antisense, TFs have the unique ability to induce both gain of function and loss of function mutations, and also the capability to target and modify genomic DNA. In the later case, TFs can incorporate more complex catalytic domains that confer the ability to methylate, cut, and recombine DNA. In addition, TF-mediated regulation of gene expression can be tightly regulated by a variety of small chemical molecules that control TF-dimerization or DNA binding.

Two general strategies have been developed to generate TFs that can regulate endogenous genes. The first is a *de novo* targeting strategy by which a particular DNA binding site is chosen in a promoter of interest close to the transcription start site of a gene (Fig. 4A). Information regarding chromatin-accessibility and endogenous TF binding sites is required in order to choose accessible sites. Target sites are chosen and polydactyl proteins constructed based on the existing lexicon of modified ZF domains and the “targetable” DNA triplets available in the accessible region. Binding and specificity of these custom-designed proteins are verified first with DNA binding assays *in vitro*, then with reporter gene assays, and finally in the chromatin context with assays measuring specific endogenous gene regulation.

The second strategy involves the creation of combinatorial TF libraries for functional screening in living cells or organisms (Fig. 4B). In this case, TF library members eliciting the highest biological effect can be selected in the first step. Subsequently, information regarding the bound DNA sequence (based on the known DNA-recognition domains of the selected DBDs), the binding location (based on chromatin immunoprecipitation assays), and specificity of regulation (based on DNA microarray and genomic search assays) can be integrated to determine the putative genes targeted directly by the TF. Combinatorial TF libraries have been built using both synthetic and natural ZF domains. Such libraries have been shown to be powerful tools for modification of phenotypes and have opened new pharmacogenomic approaches to the discovery of genes and regulatory regions involved in disease. Both approaches, *de novo* design of

TFs and selection of TFs from combinatorial libraries, represent powerful complements to existing methods for genetic manipulation. Artificial TFs exploit the inherent transcriptional capabilities of cells to modify cellular functions. This is especially interesting in the context of diseases that are able to progress or evolve by changes in transcriptional programs in a given cellular type. Another advantage of TFs is their ability to activate specific promoters within a transcriptional unit and generate the transcript isoforms that are relevant in a particular cellular background. As described elsewhere (Segal, 2002), artificial TFs can be constructed easily by any investigator using published information, without the need to employ exotic techniques such as phage display or to collaborate with a specialized zinc finger laboratory. Finally, artificial TFs can be used to overcome existing cDNA patents (Jamieson et al. 2003).

A potential limitation of ZF-based TF design is that structural features of the current zinc finger domains may ultimately impose restrictions on the spectrum of recognizable DNA sequences. Binding specificity is largely determined by the orientation of the α -helix and the amino acids it displays in the major groove. Since all current domains used to construct custom ZF proteins have been based on ZF2 of Zif268, future domains might benefit from experimentation with different ZF frameworks. For example, such new domains might position the recognition helix closer to the second DNA strand to allow additional specific interactions. Currently, specificity can be improved *in vitro* by adding ZF DNA domains, which increases the number of potential specific interactions with the DNA. However, in the context of a complex genome, the

addition of ZF domains decreases the number of potential TF binding sites. Future development of TF technology should additionally take into account the ability of TFs to access and or modify chromatin in silent promoters, perhaps incorporating novel domains able to control these processes.

Figure Legends

Fig. 1. The mechanism of action of artificial transcription factors (TFs). The DNA binding domain (DBD) is designed to recognize a target DNA sequence upstream the transcription start site of the gene of interest. This DBD is linked to an effector domain (ED) that mediates gene regulation. Activation is achieved when the ED is an activator of transcription, such as VP16, VP64, or p65. Specific repression is mediated by coupling the DBD with a repressor domain, such as KRAB or SID domain.

Fig. 2. Modular assembly of custom zinc finger proteins. **A**, Crystal structure of three-ZF protein Zif268 bound to DNA (Elrod-Erickson et al., 1996) shows a relatively simple and regular binding pattern of three primary residues at positions -1, 3, and 6 (white, outlined labels) in each finger (gray ribbons) contacting three bases (highlighted in black, dark gray, and light gray) on one strand of the DNA. Residue numbering is in relation to the start of the α -helix. The residue in position 2 often contacts a target site overlap base on the opposite strand (not shown). Spheres = zinc ions. **B**, In the modular assembly strategy, recognition modules (boxes) consisting of blocks of seven residues (-1 to 6) are grafted into a regular zinc finger scaffold (**C**) using standard PCR methods. Some modules exhibit target site overlap (curved dashed lines), requiring a G or T to be present in the neighboring finger's subsite ("k" following the recognition site [k = G or T]). **D**, The full set of modules can be grafted in a combinatorial fashion to create multifinger libraries of DNA-binding proteins.

Fig. 3. Applications of ZF technology. **A**, A DBD consisting of several ZFs can be linked to an ED mediating transcriptional regulation. The same DBD can be attached to diverse DNA modification domains (DMD), providing targeted DNA cleavage, recombination, integration, or methylation. **B**, Possible applications of designed TF and ZF technology in areas of functional genomics, molecular therapeutics, and biotechnology.

Fig. 4. Strategies to regulate endogenous gene expression using TFs. **A**, *De novo* design of TFs binding specific regulatory regions in the promoter of interest. **B**, Genome-wide screens of TFs. Large combinatorial libraries of TFs are delivered into cells. The TF inducing the phenotype of interest is isolated and used to determine the group of genes responsible for the modified phenotype.

References

- Agrawal N, Dasaradhi PVN, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK (2003) RNA interference: Biology, Mechanism, and Applications. *Microbiol Mol Biol Rev* **67**: 657-685.
- Akopian A, He J, Boocock MR, Stark WM (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc Natl Acad Sci U S A* **100**:8688-91.
- Ansari AZ and Mapp, A (2002) Molecular design of artificial transcription factors. *Current Opinion in Chem Biol* **6**:765-772.
- Arora PS, Ansari AZ, Best TP, Ptashne M, Dervan P (2002) Design of Artificial Transcriptional activators with rigid poly-L-proline linkers. *JACS* **124**: 13067-13071.
- Ayer DE, Laherty CD, Lawrence QA, Armstrong AP, Eisenman RN (1996) Mad proteins contain a dominant transcription repression domain. *Mol Cell Biol* **16**, 5772-5781.
- Bae KH, Kwon YD, Shin HC, Hwang MS, Ryu EH, Park KS, Yang HY, Lee DK, Lee Y, Park J, Kwon HS, Kim HW, Yeh BI, Lee HW, Sohn SH, Yoon J, Seol W, Kim JS (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**:275-80.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**(2):281-97.
- Bartsevich VV and Juliano RL (2000) Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries. *Mol Pharmacol* **58**:1-10.
- Bartsevich VV, Miller JC, Case CC, Pabo CO (2003) Engineered zinc finger proteins for controlling stem cell fate. *Stem Cells* **21**:632-7.
- Beerli RR, Segal DJ, Dreier B, Barbas CF 3rd (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* **95**:14628-33.
- Beerli RR, Dreier B, Barbas CF 3rd (2000a) Positive and negative regulation of endogenous genes by designed transcription factors. *Proc Natl Acad Sci U S A* **97**:1495-500.
- Beerli RR, Schopfer U, Dreier B, Barbas CF 3rd (2000b) Chemically regulated zinc finger transcription factors. *J Biol Chem* **42**:32617-27.
- Beerli RR and Barbas CF (2002) Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* **20**:135-41.
- Benko S, Love JD, Beladi M, Schwade JWR, Nagy L (2003) Molecular determinants of the balance between co-repressor and co-activator recruitment to the retinoic acid receptor. *J Biol Chem* **44**: 43797-43806.

- Benekli, M, Baer MR, Baumann H, Wetzler M (2003) Signal transducer and activator of transcription proteins in leukemias. *Blood* **15**:2940-51.
- Blancafort P, Magnenat L, Barbas CF 3rd (2003) Scanning the human genome with combinatorial transcription factor libraries. *Nat Biotechnol* **21**:269-74
- Blancafort P., Steinberg SV, Paquin B, Klinck R, Scott JK, Cedergren R (1999) The recognition of a noncanonical RNA base pair by a zinc finger protein. *Chem Biol* **6**, 585-597.
- Bremer RE, Wurtz NR, Szewczyk JW, Dervan PB (2001) Inhibition of major groove DNA binding bZIP proteins by positive patch polyamides. *Bioorg Med Chem* **9**:2093-103.
- Carroll D (2004) Using nucleases to stimulate homologous recombination. *Methods Mol Biol* **262**, 195-207.
- Cheng X, Boyer JL, Juliano RL (1997) Selection of peptides that functionally replace a zinc finger in the Sp1 transcription factor by using a yeast combinatorial library. *Proc Natl Acad Sci U S A* **94**:14120-5.
- Chiang SY, Burli RW, Benz CC, Gawron L, Scott GK, Dervan PB, Breerman TA (2000) Targeting the ets binding site of the HER2/neu promoter with pyrrole-imidazole polyamides. *J Biol Chem* **275**:24246-54.
- Choo Y and Klug A (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci U S A* **91**:11163-11167.
- Choo Y. (1998) Recognition of DNA methylation by zinc fingers. *Nat Struct Biol* **5**, 264-265.
- Ciarapica R, Rosati, J, Cesareni, G, Nasi, S (2003) Molecular recognition in helix-loop-helix-leucine zipper domains. *J Biol Chem* **278**: 12182-12190.
- Collins CH, Yokobayashi Y, Umeno D, Arnold FH (2003) Engineering proteins that bind, move, make and break DNA. *Curr Opin Biotechnol* **14**:371-8.
- Corbi N, Libri V, Fanciulli M, Tinsley JM, Davies KE, Passananti (2000) The artificial zinc finger coding gene 'Jazz' binds the utrophin promoter and activates transcription. *Gene Therapy* **7**:1076-1083.
- Coull JJ, Melander C, Rucker VC, Dervan PB, Margolis DM (2002) Targeted derepression of the human immunodeficiency virus type 1 long terminal repeat by pyrrole-imidazole polyamides. *J Virol* **76**:12349-54.
- Crook ST (2004) Antisense strategies. *Curr Mol Med* **4**:465-87.
- Dreier B, Segal DJ, Barbas CF 3rd (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol* **303**:489-502.
- Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF 3rd (2001) 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. *J Biol Chem* **276**:29466-78
- Dudouet B, Burnett R, Dickinson LA, Wood MR, Melander C, Belitsky JM, Edelson B, Wurtz N, Briehn C, Dervan PB, Gottesfeld JM (2003) Accessibility of nuclear chromatin by DNA binding polyamides. *Chem Biol* **10**:859-867.
- Ehley JA, Melander C, Herman D, Baird EE, Ferguson HA, Goodrich JA, Dervan PB, Gottesfeld JM (2002) Promoter scanning for transcription factor inhibition with DNA-binding polyamines. *Mol Cell Biol* **22**:1723-33.
- Elkon R, Linhart C, Sharan R, Shamir R, Shilon Y (2003) Genome-Wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Research* **13**:773-780.
- Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO (1996) Zif 268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* **15**:1171-80.
- Elrod-Erickson M, Benson TE, Pabo CO (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**:451-64.
- Falke D, Fisher M, Ye D, Juliano RL (2003) Design of artificial transcription factors to selectively regulate the pro-apoptotic bax gene. *Nucleic Acids Res* **31**:E10-20.
- Fechter E and Dervan P (2003) Allosteric inhibition of protein-DNA complexes by polyamide-intercalator conjugates. *J Am Chem Soc* **125**: 8476-8485.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**:806-811.
- Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, Takami Y, Nakayama T, Oshimura M. (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat Cell Biol.* **6**(8):784-91.
- Gandet P and Memelink J (2002) Transcription factors: tools to engineer the production of pharmacologically active plant metabolites. *Trens Pharm Sci* **23**:563-569.
- Gebuhr TC, Kovalev GI, Bultman S, Godfrey V, Su L, Magnuson T (2003) The role of Brg1, a catalytic subunit of mammalian chromatin-remodeling complexes, in T cell development. *J Exp Med* **198**:1937-49.
- Greisman HA and Pabo CO (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**:657-61.
- Guan X, Stege J, Kim M, Dahmani Z, Fan N, Heifetz P, Barbas CF 3rd, Briggs SP (2002) Heritable endogenous gene regulation in plants with designed

- polydactyl zinc finger transcription factors. *Proc Natl Acad Sci U S A* **99**:13296-301.
- Harms K, Nozell S, Chen X (2004) The common and distinct target genes of the p53 family of transcription factors. *Cell Mol Life Sci* **61**:882-42.
- Holbro T, Beerli RR, Maurer F, Koziczak M, Barbas CF 3rd, Hynes NE (2003) The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation. *Proc Natl Acad Sci U S A*. **15**:8933-8.
- Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell based selection. *Proc Natl Acad Sci U S A* **100**: 12271-76.
- Isalan M, Choo Y, Klug A (1997) Synergy between adjacent zinc fingers in sequence-specific recognition. *Proc Natl Acad Sci U S A* **94**: 5617-21.
- Isalan M, Klug A, Choo Y (2001a) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* **19**:656-660.
- Isalan M, Patel SD, Balasubramanian S, Choo Y (2001b) Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. *Biochemistry* **40**, 830-836.
- Jackson AL, Bartz SR, Shelter J, Kobayashi SV, Buchard J, Mao Mao, Li B, Cavet G, Linsley PS (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**:635-7.
- Jamieson AC, Kim SH, Wells J (1994) In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* **33**:5689-95.
- Jamieson AC, Miller JC, Pabo CO (2003) Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov.* **2**:361-8.
- Jouvenot Y, Ginjala V, Zhang L, Liu PQ, Oshimura M, Feinberg AP, Wolffe AP, Ohlsson R, Gregory PD (2003) Targeted regulation of imprinted genes by synthetic zinc-finger transcription factors. *Gene Ther* **10**:513-22.
- Khaled Z, Benimetskaya L, Zeltser R, Khan T, Sharma HW, Narayanan R, Stein CA (1996) Multiple mechanisms may contribute to the cellular anti-adhesive effects of phosphorothioate oligodeoxynucleotides. *Nucleic Acids Res.* **24**:737-45.
- Kim JS and Pabo CO (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc Natl Acad Sci U S A* **95**: 2812-17.
- Lee DK, Park JW, Kim YJ, Kim J, Lee Y, Kim J, Kim JS (2003) Toward a functional annotation of the human genome using artificial transcription factors. *Genome Res* **13**:2708-16.

- Lin Q, Barbas CF 3rd, Schultz PG (2003) Small-molecule switches for zinc finger transcription factors. *J Am Chem Soc* **125**:612-3.
- Liu Q, Segal DJ, Ghiara JB, Barbas CF 3rd (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A* **94**:5525-30.
- Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, Zhong X, Lee YL, Eisenberg SP, Spratt SK, Case CC, Wolffe AP (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem* **276**:11323-34.
- Lund CV, Blancafort P, Popkov M and Barbas CF 3rd (2004) Promoter targeted phage display selections with preassembled synthetic zinc finger libraries for endogenous gene regulation. *J Mol Biol* **340**:599-613.
- Magenat L, Blancafort P and Barbas CF 3rd (2004) In vivo selection of combinatorial libraries and designed affinity maturation of polydactyl zinc finger transcription factors for icam-1 provides new insights into gene regulation. *J Mol Biol* **341**:635-49.
- Margolin JF, Friedman JR, Meyer WK, Vissing H, Thiesen HJ, Rauscher FJ 3rd (1994) Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci U S A* **91**:4509-4513.
- Miller J, McLachlan, AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*. *EMBO J.* **4**:1609-1614.
- Moore M, Klug A, Choo Y (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two finger units. *Proc Natl Acad Sci U S A* **98**:1437-41.
- Nagaoka M, Doi Y, Kuwahara J, Sugiura Y (2002) Novel strategy for the design of a new zinc finger for the AT-rich sequence by α -helix substitution. *JACS* **124**:6526-27.
- Ordix MI, Barbas CF 3rd, Beachy RN (2002) Regulation of transgene expression in plants with polydactyl zinc finger transcription factors. *Proc Natl Acad Sci U S A* **99**:13290-5
- Papworth M, Moore M, Isalan M, Minczuk M, Choo Y, Klug A (2003) Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors. *Proc Natl Acad Sci U S A* **100**:1621-6.
- Pavletich NP and Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**:809-817.
- Peisach E and Pabo CO (2003) Constraints for zinc finger linker design as inferred from X-ray crystal structure of tandem Zif268-DNA complexes. *J Mol Biol* **330**:1-7.

- Pollock R, Giel M, Linher K, Clackson T (2002) Regulation of endogenous gene expression with a small-molecule dimerizer. *Nat Biotechnol* **20**:729-33.
- Rebar, EJ and Pabo PO (1993) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**: 671-673 (1994)
- Rebar EJ, Huang Y, Hickey R, Nath AK, Meoli D, Nath S, Chen B, Xu L, Liang Y, Jamieson AC, Zhang L, Spratt SK, Case CC, Wolffe A, Giordano FJ (2002) Induction of angiogenesis in a mouse model using engineered transcription factors. *Nat Med* **8**:1427-32.
- Ren D, Collingwood TN, Rebar EJ, Wolffe AP, Camp HS (2002) PPAR γ knockdown by engineered transcription factors: exogenous PPAR γ 2 but not PPAR γ 1 reactivates adipogenesis. *Genes Dev* **16**:27-32
- Reynolds L, Ullman C, Moore M, Isalan M, West MJ, Clapham P, Klug A, Choo Y (2003) Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. *Proc Natl Acad Sci U S A* **100**:1615-20
- Sadowski I, Ma J, Triezenberg S, Ptashne M (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563-564.
- Sanchez JP, Ullman C, Moore M, Choo Y, Chua NH (2002) Regulation of gene expression in *Arabidopsis thaliana* by artificial zinc finger chimeras. *Plant Cell Physiol* **43**:1465-72.
- Segal DJ, Dreier B, Beerli RR, Barbas CF 3rd (1999) Controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96**: 2758-63.
- Segal DJ and Barbas CF 3rd (2001) Custom DNA-binding proteins come of age: polydactyl zinc-finger proteins. *Curr Opin Biotechnol* **12**:632-7
- Segal DJ (2002) The use of zinc finger peptides to study the role of specific transcription factor binding sites in the chromatin environment. *Methods*. **26**:76-83.
- Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, Huber A, Koksich B, Lund CV, Magnenat L, Valente D, Barbas CF 3rd (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42**:2137-48.
- Segal DJ, Stege JT, Barbas CF 3rd (2003) Zinc fingers and a green thumb: manipulating gene expression in plants. *Curr Opin Plant Biol* **6**:163-8.
- Segal DJ, Gonclaves J, Eberhardy S, Swan CH, Torbett BE, Barbas CF 3rd (2004) Attenuation of HIV-1 replication in primary human cells with a designed zinc finger transcription factor. *J Biol Chem* **279**:14509-19.
- Sgouras DN, Athanasiou MA, Beal GJ Jr., Fisher RJ, Blair DG, Mavrothalassitis GJ (1995) ERF: an ETS domain protein with strong transcriptional repressor activity, can suppress ets-associated tumorigenesis and is

- regulated by phosphorylation during cell cycle and mitogenic stimulation. *EMBO J* **14**:4781-4793.
- Shi Y and Berg J.M (1995) A direct comparison of the properties of natural and designed zinc-finger proteins. *Chem Biol* **2**, 83-89.
- Snowden AW, Gregory PD, Case CC, Pabo CO (2002) Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Curr Biol* **12**:2159-66.
- Snowden AW, Zhang L, Urnov F, Dent C, Jouvenot Y, Zhong X, Rebar EJ, Jamieson AC, Zhang HS, Tan S, Case CC, Pabo CO, Wolffe AP, Gregory PD (2003) Repression of vascular endothelial growth factor A in glioblastoma cells using engineered zinc finger transcription factors. *Cancer Res* **63**:8968-76.
- Snowden AW, Zhang L, Urnov F, Dent C, Jouvenot Y, Zhong X, Rebar EJ, Jamieson AC, Zhang HS, Tan S, Case CC, Pabo CO, Wolffe AP, Gregory PD (2003) Repression of vascular endothelial growth factor A in glioblastoma cells using engineered zinc finger transcription factors. *Cancer Res* **63**, 8968-8976.
- Stanojevic D and Young RA (2002) A highly potent artificial transcription factor. *Biochemistry* **41**: 7209-7216.
- Steffen B, Serve H, Berdel WE, Agrawal S, Linggi B, Buchner T, Hielbert SW, Muller-Tidow., C (2003) Specific protein redirection as a transcriptional therapy approach for t(8;21) leukemia. *Proc Natl Acad Sci U S A* **14**: 8448-8453.
- Stege JT, Guan X, Ho T, Beachy RN, Barbas CF 3rd (2002) Controlling gene expression in plants using synthetic zinc finger transcription factors. *Plant J* **32**, 1077-1086.
- Tan S, Guschin D, Davalos A, Lee YL, Snowden AW, Jouvenot Y, Zhang HS, Howes K, McNamara AR, Lai A, Ullman C, Reynolds L, Moore M, Isalan M, Berg LP, Campos B, Qi H, Spratt SK, Case CC, Pabo CO, Campisi J, Gregory PD (2003) Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc Natl Acad Sci U S A* **100**:11997-2002.
- Tan W, Zhu K, Segal DJ, Barbas CF 3rd, Chow SA (2004) Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. *J Virol* **78**:1301-13.
- Tupler R, Perini G, Green MR (2001) Expressing the human genome. *Nature* **409**:832-833.
- Uil TG, Haisma HJ, Rots MG (2003) Therapeutic modulation of endogenous gene function by agents with designed DNA-sequence specificities. *Nucleic Acids Res* **21**:6064-78.

- Urruta R (2003) Krab-containing zinc finger repressor proteins. *Genome Biology* **4**:231.1-231.7.
- Venter JC et al. (2001) The sequence of the human genome. *Science* **16**:1304-51.
- Xu D, Ye D, Fisher M, Juliano RL (2002) Selective inhibition of P-glycoprotein expression in multidrug-resistant tumor cells by a designed transcriptional regulator. *J Pharmacol Exp Ther* **302**:963-71.
- Wu H, Yang WP, Barbas CF 3rd (1995) Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci U S A* **92**:344-8.
- Wurtz NR, Pomerantz JL, Baltimore D, Dervan P (2002) Inhibition of DNA binding by NF- κ B with pyrrole-imidazole polyamides. *Biochemistry* **41**:7604-7609.
- Yang F, Belitsky JM, Villanueva RA, Dervan PB, Roth MJ (2003) Inhibition of Moloney murine leukemia virus integration using polyamides targeting the long-terminal repeat sequences. *Biochemistry* **42**:6249-58.
- Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP, Case CC (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene *J Biol Chem* **275**:33850-60.

TABLE 1
 Targets regulated with ZF-based TFs

Targeted Gene	Targeted Gene Function	Phenotype/detection assay	Reference
<i>Mammalian genes</i>			
<i>ErbB-2, ErbB-3</i>	Oncogenic Tyr-Kinase receptors Holbro et al.,2003	Cell cycle arrest Control of breast cancer cell proliferation	Beerli et al., 1998 Beerli et al., 2000
<i>Utrophin promoter</i>	Duchenne muscular Distrophy	reporter assays	Corbi et al., 2000
<i>IGF2, H19</i>	Cancer , Beckwin-Wiedemann syndrome	Activation of engogenous promoter	Jouvenot et al., 2002
<i>EPO</i>	Anemia	Endogenous regulation, DNase protection assays	Zhang et al., 2000
<i>VEGF-A</i>	Angiogenesis	Endogenous regulation, DNase protection assays induction of Angiogenesis in mouse	Liu et al., 2001 Rebar et al., 2002
<i>MDR1</i>	Multidrug resistance	Repression of endogenous gene, Sensitivity to Doxorubicin	Bartsevich et al., 2000 Xu et al., 2002
<i>PPARγ</i>	Adipogenesis	Isoform-specific Inhibition of adipogenesis	Ren et al. 2002
<i>BAX</i>	Apoptosis	Activation of endogenous gene Promotes p53 independent apoptosis	Falke et al., 2003
<i>CHK2</i>	Cell cycle progression	Repression on endogenous gene Suppresses p53 activation in response to DNA damage	Tan et al., 2003
<i>CD144</i>	Angiogenesis, metastasis	Activation or repression using cell sorting	Blancafort et al., 2003
<i>ICAM-1</i>	Cell adhesion	Activation or repression using cell sorting	Magenat et al., 2004
<i>OCT-4</i>	ES differentiation	Activation and repression of endogenous gene Regulation of lineage genes Morphological differentiation of ES cells	Bartsevich et al., 2003
<i>Viral targets</i>			
<i>HIV-1 (LTR)</i> al.,2003, Segal et al., 2004		Transient Repression of HIV-1 transcription Stable Inhibition of multiple HIV-1 isolates	Reynolds et
<i>HSV-1 (ICP4 promoter)</i>		Inhibition of HSV-1 infection	Papworth et al., 2003
<i>Plant genes</i>			

<i>APETALA3</i> (<i>AP3</i>)	floral organ identity	Missing petals and petals to sepals transformations in transgenic plants	Guan et al., 2002
--------------------------------	-----------------------	--	-------------------

TABLE 2

Comparison of TFs versus RNAi technologies to modulate gene expression

	TFs	RNAi & antisense
Vector-Based Delivery	Transient, Retroviral, Adenoviral, Lentiviral.	Transient, Retroviral, Adenoviral, Lentiviral.
Direct Delivery	Protein transduction	Lipids or electroporation
Optimal Expression levels	low , pol II promoters	High, pol III or pol II promoters
Chemical Synthesis	yes but not practice	Yes
Regulation of Expression	Yes, by inducible promoters and post-transcriptional activation of TFs	Yes, by inducible promoters
Specificity	Limited by number and quality of ZF-DNA contacts Limited by accessibility due to endogenous proteins	Limited by base pairing and biochemistry Limited by accessibility due to endogenous proteins and RNA structure
Gene regulation Gain-of-function Loss-of-function	Yes, with Activator domains Yes, with Repressor domains	No Yes

**Targeted
DNA Sequence**



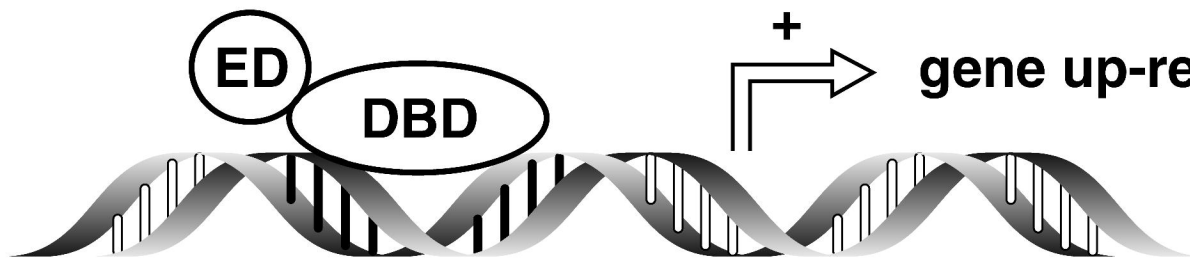
**Promoter
of Interest**

gene X



**Design of TF
interacting specifically
with targeted DNA**

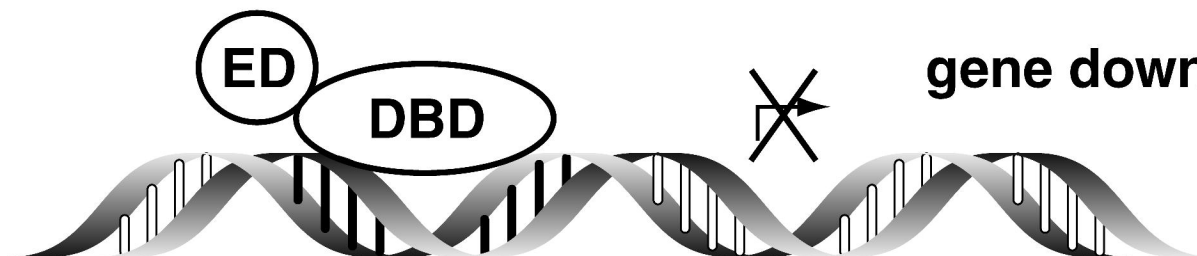
ED = Activator Domain



gene up-regulated

gene X

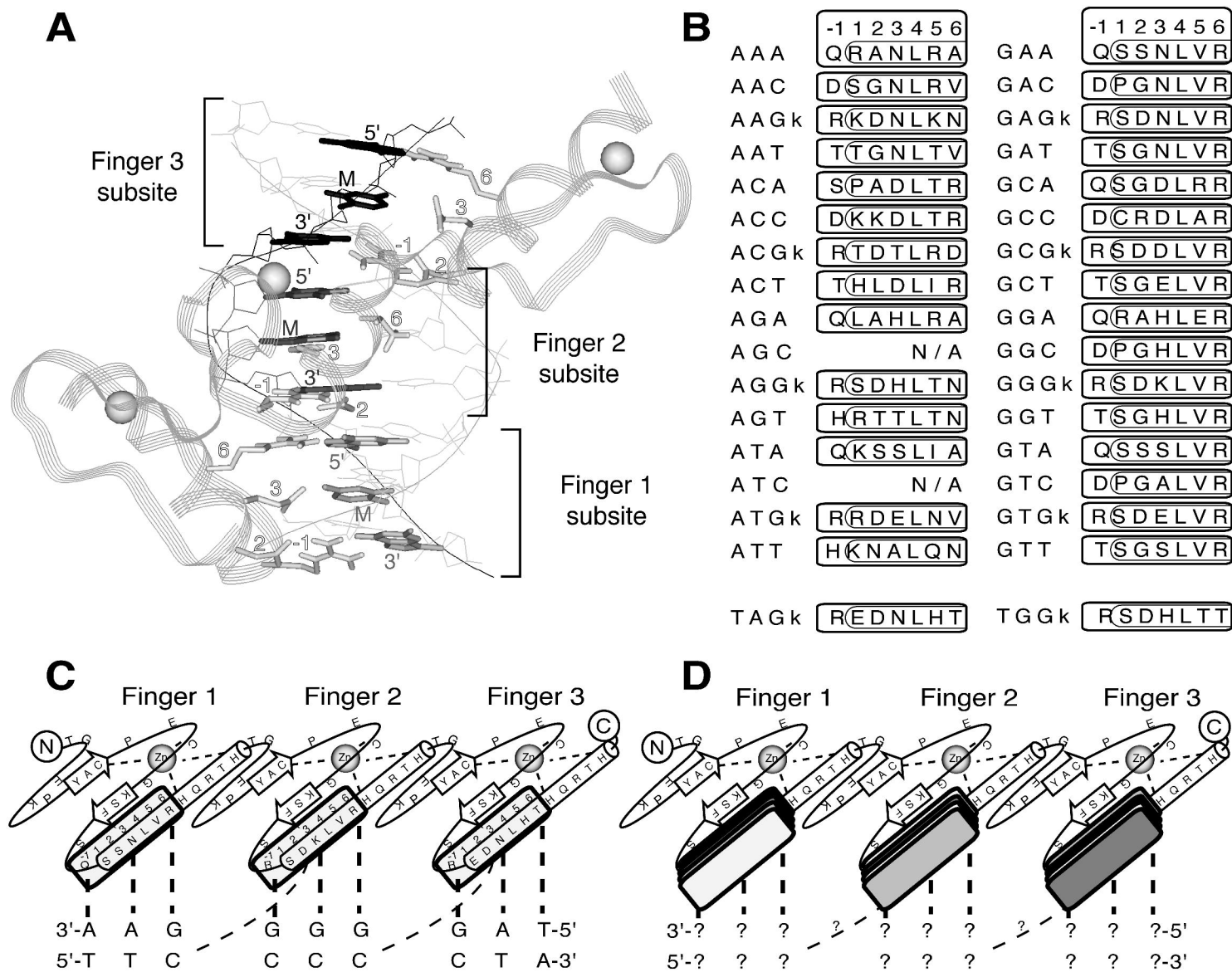
ED = Repressor Domain



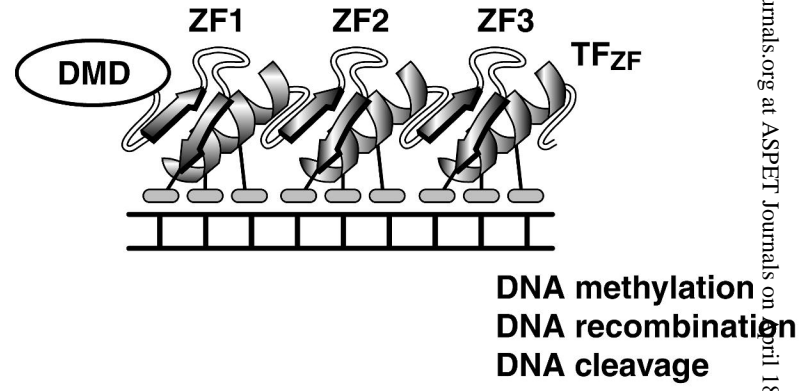
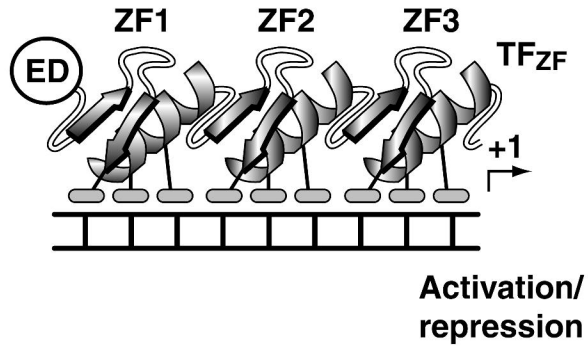
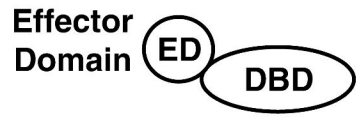
gene down-regulated

gene X

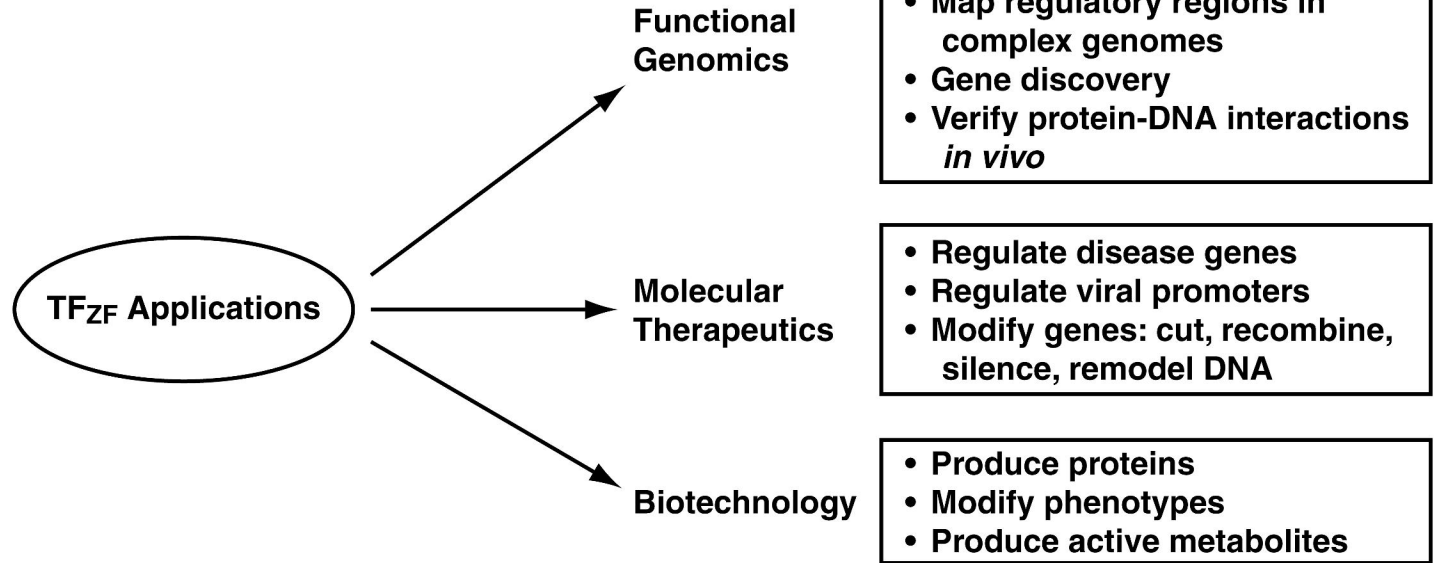
Figure 2



A

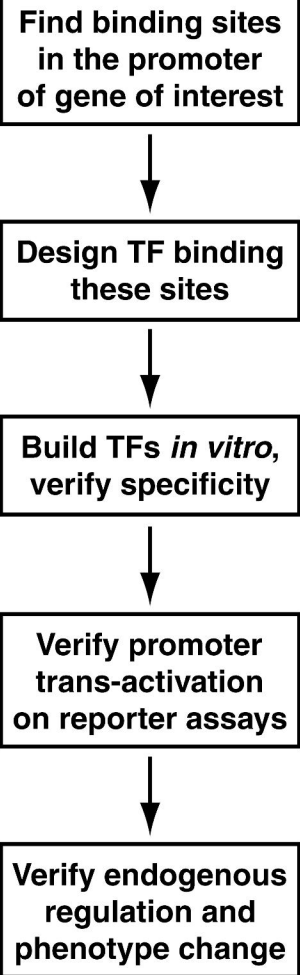


B



A**Site-directed
TFZF-targeting**

- Genome databases, DNA accessibility data

**B****Genome-wide
TFZF-targeting**