MINIREVIEW

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, La Jolla, California (P.B., C.F.B.); and Department of Pharmacology and Toxicology, University of Arizona College of Pharmacy, Tucson, Arizona (D.S.)

Received May 14, 2004; accepted August 25, 2004

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 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.

In complex organisms, phenotypic diversity is achieved primarily 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 post-transcriptional 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 corepressor and corepressor regulatory complexes. TF activator complexes include Mediator, which interacts with core promoter factors p300/CREB-binding protein-associated factor and p300/CREB-binding protein, which 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 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

ABBREVIATIONS: TF, transcription factor; DBD, DNA binding domain; ED, effector domain; bp, base pair(s); VEGF, vascular endothelial growth factor; RNAi, RNA interference; siRNA, small interfering RNA; miRNA, micro-RNA.
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 nuclear factor-kB subunit p65 (Liu et al., 2001). Repression EDs have included KRAB (Krüppel-associated box; Margolin et al., 1994; Urruta, 2003), 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, and peptide nucleic acids (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 (Chiang et al., 2000; Bremer et al., 2001; Coull et al., 2002; Ehley et al., 2002; Stanojevic and Young, 2002; Wurtz et al., 2002; Fechter and Dervan, 2003; Yang et al., 2003). DNA microarray experiments have shown that polyamides seem to be able to regulate a limited number of genes in lymphoma cells (Dudouet et al., 2003).

**Using Protein Scaffolds: Zinc Finger Domains**

The classic protein scaffold used for targeting gene expression is the C2H2 ZF domain. The human genome is estimated to encode 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 antiparallel β-strands, and additional stability is provided by the coordination of a zinc ion by the side chains of two cysteine and two 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; Eldred-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 (Asp) 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 (Choo and Klug, 1994; Jamieson et al., 1994; Rebar and Pabo, 1994; Wu et al., 1995; Segal et al., 1999; Dreier et al., 2000). These correlations were rationalized according to the struc-
tecture 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 (Elrod-Erickson et al., 1998; Dreier et al., 2000, 2001).

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 true in part because purine bases (G and A) offer more hydrogen-bonding possibilities than pyrimidines and in part 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 methods have 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 with that of their duplex-DNA-binding counterparts.

Other methods 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). In contrast, we have adopted a helix-grafting strategy, based on the modular property of ZF proteins (Fig. 2, B and C) (Beerli et al., 1998). With the caveat that an overlap contact would need to be accommodated between certain subsets of ZF units, multifinger 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 method for construction of polydactyl ZF proteins without the need for phage display and selection for each new DNA target. Recently Segal et al. (2003a,b) have developed a bipartite complementary method. In this strategy, two phage display libraries were used to select for ZFs binding 5 bp. The ZF proteins from these two selections were recombined to generate a single protein recognizing 9 bp.

In contrast, from 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 multifinger 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 multimodular 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 multifinger 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 9 base-pair (bp) sequence, optimizing one ZF at a time. On the other hand, Isalan et al. (2001a) developed a bipartite complementary method. In this strategy, two phage display libraries were used to select for ZFs binding 5 bp. The ZF proteins from these two selections were recombined to generate a single protein recognizing 9 bp.

In contrast, from 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 multifinger 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.
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 (which naturally binds GC-rich regions) to recognize an AT-rich element. The helices used for grafting were derived from the Drosophila melanogaster CP2-II protein, which recognizes AT-rich sequences.

By increasing the number of ZF units in the multifinger 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 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 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. 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 in which 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), mifepristone (ZF-progesterone receptor fusions), or Ponasterone A (ZF-ecdyson receptor fusions), and were shown to activate reporter constructs by up to 3 orders of magnitude. An inducible system was constructed by Pollock et al. (2002) to regulate the endogenous vascular endothelial growth factor-A (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 analog 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 impaired DNA binding. Several heterocycle-containing compounds were able to rescue the mutations and activate a reporter construct up to 100-fold.

### 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 these genes are related functionally and participate in the same functional pathway.

Beerli and coworkers (Beerli et al., 1998, 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'-untranslated 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 to a transcriptional repression domain, KRAB. It is noteworthy that cancer cell lines expressing these regulators by retroviral delivery recapitulated the cell cycle alterations induced by gain 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 Duchenne muscular dystrophy. 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. Therefore, these genes have become attractive targets for therapeutic regulation. ZF-based TFs have recently been targeted to the pro-
moter of 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. It is noteworthy that 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 organ.

**TABLE 1**

<table>
<thead>
<tr>
<th>Targeted Gene</th>
<th>Targeted Gene Function</th>
<th>Phenotype/Detection Assay</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mammalian genes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ErbB-2, ErbB-3</td>
<td>Oncogenic Tyr-Kinase receptors</td>
<td>Cell cycle arrest, Control of breast cancer cell proliferation reporter assays</td>
<td>Beerli et al. (1998, 2000); Holbro et al. (2003); Lund et al. (2004)</td>
</tr>
<tr>
<td>Utrophi promoter</td>
<td>Duchenne muscular dystrophy</td>
<td>Activation of endogenous promoter</td>
<td>Corbi et al. (2000)</td>
</tr>
<tr>
<td>VEGF-A</td>
<td>Angiogenesis</td>
<td>Isofrom-specific Inhibition of adipogenesis</td>
<td>Xue et al. (2002)</td>
</tr>
<tr>
<td>MDR1</td>
<td>Multidrug resistance</td>
<td>Activity on endogenous gene, promotes p53 independent apoptosis</td>
<td>Falke et al. (2003)</td>
</tr>
<tr>
<td>PPARγ</td>
<td>Adipogenesis</td>
<td>Activation and repression of endogenous gene, induces p53 independent apoptosis</td>
<td>Ren et al. (2002)</td>
</tr>
<tr>
<td>CHK2</td>
<td>Cell cycle progression</td>
<td>Suppression p53 activation in response to DNA damage</td>
<td>Tan et al. (2003)</td>
</tr>
<tr>
<td>CD144</td>
<td>Angiogenesis, metastasis</td>
<td>Activation or repression using cell sorting</td>
<td>Blancafort et al. (2003)</td>
</tr>
<tr>
<td>ICAM-1</td>
<td>Cell adhesion</td>
<td>Activation or repression using cell sorting</td>
<td>Magenat et al. (2004)</td>
</tr>
<tr>
<td>OCT-4</td>
<td>ES differentiation</td>
<td>Activation and repression of endogenous gene, regulation of lineage genes, morphological differentiation of ES cells</td>
<td>Bartsevich et al. (2003)</td>
</tr>
<tr>
<td><strong>Viral targets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV-1 (LTR)</td>
<td>Transient Repression of HIV-1 transcription</td>
<td>Reynolds et al. (2003)</td>
<td></td>
</tr>
<tr>
<td><strong>Plant genes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APETALA3 (AP3)</td>
<td>Floral organ identity</td>
<td>Missing petals and petals to sepal transformations in transgenic plants</td>
<td>Guan et al. (2002)</td>
</tr>
</tbody>
</table>

**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.
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 with a nonangiogenic, low tumorigenic cell line.

Bartsevich and Juliano (2000) and Xu et al. (2002) 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 bp in the promoter of the CHK2 gene. The artificial TF was able to repress specifically the CHK2 gene, as determined by DNA microarray experiments. It is noteworthy that 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 cells. TF technology could be used to regulate the cell fate of pluripotent stem cells, perhaps redirecting 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 Antiviral 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 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 studies 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 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). Chimeric recombinases combining mutant variants of Tn3 resolvase and the murine zinc finger protein Zf268 have recently been engineered in bacteria (Akopian et al., 2003). These chimeras were able to catalyze site-specific recombination mediated by Zf268 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 toward 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 (Sanchez et al., 2002; Stege et al., 2002). In Arabidopsis thaliana, 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 A. thaliana 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 biosynthetic pathways, and to repress 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. In general, targeted DNA sequences are localized near 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 (e.g., 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). Detailed knowledge of endogenous factors affecting transcription in cis-regulatory regions of a gene is often limited. Promoters can also be modified epigenetically by specific methylation. Therefore, 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.

To functionally select genomic sequences that can be targeted by TFs and therefore used to efficiently modify endogenous transcription, methods have been developed for screening combinatorial TF libraries in mammalian cells (Blancafort 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, approximately 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 coregulation 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 coworkers have performed selections of TF libraries in several cancer cell lines 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, such angiogenic molecules as CD144 (VE-cadherin; Blancafort et al., 2003), and such cell adhesion molecules as ICAM-1 (Magnevat 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). To select TFs able to regulate directly the erb-B2 gene Lund et al. (2004) developed a novel phage display strategy to select for ZF proteins from combinatorial libraries binding the proximal erbB2 promoter. The authors isolated TFs binding the promoter that were able to regulate the endogenous erb-B2 gene.

Our recent studies have isolated artificial TFs modulating complex phenotypes in cancer cells, such as cell growth, proliferation, resistance to drugs, and metastasis (P. Blancafort, manuscript in preparation). These investigations have discovered and regulated genes involved in tumor progression. Therefore, artificial TFs have demonstrated their potential for therapeutic reprogramming 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, genome-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 Methods 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 [RNAi]), translational repression or chromatin silencing (Fukagawa et al., 2004). It is mediated by small ~22-bp double-stranded siRNAs called MicroRNAs or small interfering RNAs (siRNAs). Since the discovery of the efficacy of this approach in Caenorhabditis elegans in 1998 (Fire et al., 1998), RNAi has become extremely popular, and methods 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 long is preferred, because these RNAi 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 seems to occur because the miRNA binds to its gene target with imperfect complementarity (Bartel, 2004). Other types of antisense technologies use single-stranded nucleic acids, typically modified RNA of 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. Cells do not naturally take up TF proteins, so transient delivery requires transfection of TF-encoded cDNA. However, the stability of artificial TFs is comparable with naturally occurring ones, and because 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 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. In practice, however, 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. Because a six-ZF TF can potentially recognize a unique 18-bp site in the human genome, there is little practical reason to exceed this binding site size. Six-ZF TFs have been shown to have higher affinities and better discrimination than three-ZF TFs (Beerli et al., 1998; Blancafort et al., 2003; Lund et al., 2004), and certain designed six-ZF TFs were found to regulate only their single targeted genes based on microarray analysis (Guan 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 sites. 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- or 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, because transcriptional levels of a given gene can oscillate depending on the cell type and the disease stage. Ther-

**TABLE 2**
Comparison of TFs versus RNAi technologies to modulate gene expression

<table>
<thead>
<tr>
<th></th>
<th>TFs</th>
<th>RNAi &amp; Antisense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector-based delivery</td>
<td>Transient</td>
<td>Transient</td>
</tr>
<tr>
<td></td>
<td>Retroviral</td>
<td>Retroviral</td>
</tr>
<tr>
<td></td>
<td>Adenoviral</td>
<td>Adenoviral</td>
</tr>
<tr>
<td></td>
<td>Lentiviral</td>
<td>Lentiviral</td>
</tr>
<tr>
<td>Direct delivery</td>
<td>Protein transduction</td>
<td>Lipids or electroporation</td>
</tr>
<tr>
<td>Optimal expression levels</td>
<td>Low, pol II promoters</td>
<td>High, pol III or pol II promoters</td>
</tr>
<tr>
<td>Chemical synthesis</td>
<td>Yes but not practical</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes, by inducible promoters and post-transcriptional activation of TFs</td>
<td>Yes, by inducible promoters</td>
</tr>
<tr>
<td>Regulation of expression</td>
<td>Limited by number and quality of ZF-DNA contacts</td>
<td>Limited by base pairing and biochemistry</td>
</tr>
<tr>
<td></td>
<td>Limited by accessibility due to endogenous proteins</td>
<td>Limited by accessibility due to endogenous proteins and RNA structure</td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene regulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain of function</td>
<td>Yes, with activator domains</td>
<td>No</td>
</tr>
<tr>
<td>Loss of function</td>
<td>Yes, with repressor domains</td>
<td>Yes</td>
</tr>
</tbody>
</table>
apeutic 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 with other methods 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 latter case, TFs can incorporate catalytic domains that are more complex and 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 near the transcription start site of a gene (Fig. 4A). Information regarding chromatin-accessibility and endogenous TF binding sites is required to choose accessible sites. Target sites are chosen, and polydactyl proteins are 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. Thereafter, 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 (Se-

---

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.
gal, 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. Because 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. At the moment, 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.

References

Agrawal N, Dasaradhi PVN, Mohmmed A, Malhotra P, Bhatnagar RK, and Mukherji SN (2002) Human zinc fingers as building blocks in the construction of artificial TFs can be used to overcome existing cDNA patents. Proc Natl Acad Sci USA 99:1370.


