Modular design of artificial transcription factors
Aseem Z Ansari* and Anna K Mapp†

Eukaryotic transcription factors are composed of interchangeable modules. This has led to the design of a wide variety of modular artificial transcription factors (ATFs) that can stimulate or inhibit the expression of targeted genes. The ability to regulate the expression of any targeted gene using a ‘programmable’ ATF offers a powerful tool for functional genomics and bears tremendous promise in developing the field of transcription-based therapeutics.

Introduction: regulation of gene expression
All cells in an organism, with a few exceptions, bear the same genome, yet they specialize to give rise to tissues with diverse morphology and function. Clearly, this diversity is not due to the differences in genomic information between the tissues in an organism; rather, it is the choice of genes that are expressed which governs the fate and function of cells. The recent decoding of the human genome coupled with genome-wide expression profiling is clarifying the relationship between specific gene-expression patterns and cell fate. Patterns of gene expression in cells are triggered by a variety of intracellular and extracellular signals that mobilize specific transcriptional regulatory factors [1••]. Once marshaled, these regulators interact and recruit various cellular machines to directly govern the transcription of the specified genes, prescribing both the structural and control contexts for which they function (Figure 1a). In diseases as diverse as diabetes and cancer, it is often malfunctioning transcriptional regulators that produce the altered patterns of gene expression at the heart of the ailment [2,3,4••].

In this context, molecules that can seek out specific genes and control their expression are attractive targets for study [5–7,8•]. These artificial transcription factors (ATFs) are formally defined as functional entities that regulate transcription either as a single molecule or as an ensemble of interacting molecules (Figure 1b). To achieve this, an ATF may (but does not necessarily) bind to DNA and directly influence the transcription of a specific gene or set of genes either positively or negatively. ATFs do not function only on initiation but may also regulate subsequent steps in transcription. Ideally, ATFs could be designed to regulate any gene or set of genes without influencing the expression of any other gene in the genome. Such factors would be powerful tools for unraveling the key transcriptional events that govern cell fate, and in the long term would have significant therapeutic potential.

Due to the rich cross-fertilization of ideas at the interface of chemistry, biology, medicine and genomics, this is an exciting time in the field of ATF design. Technological advances coupled with new mechanistic insights into gene regulation and in vivo delivery make the creation of ATFs that can reprogram gene expression in living organisms a tangible goal. In this review, we summarize recent advances in ATF design within a historical context and outline a few of the key challenges facing the field.

Modular nature of eukaryotic transcription factors
Natural transcriptional regulators typically bear two essential yet separable modules: the DNA-binding domain (DBD) and the regulatory domain (RD) [1••]. The DBD imparts most of the specificity in targeting the RD to a particular site in the genome. The regulatory modules typically play a less critical role in selecting a gene for regulation; instead, they mediate their effects directly on the gene to which they are delivered. These two functional modules are usually exchangeable (Figure 2). In some cases, more often with repressors, the DBD is not a part of the regulatory factor [1••]. Rather, proteins containing repressor domains are targeted to specific genes by interactions with other DNA-binding proteins already bound to the promoter [9,10].

The DNA-binding modules of transcriptional regulators consist of a wide variety of structural folds and many have been well characterized from both a structural and functional standpoint [11••]. By contrast, regulatory modules are less well-defined and only a few have been characterized structurally [8•]. Whereas DBDs are often described by their structural folds (helix-turn-helix, for example) or by requirements for certain co-factors (zinc fingers), RDs are typically categorized using simpler criteria such as the preponderance of certain residues within the regulatory sequence (acid-rich [1••], glutamine-rich [12,13], proline-rich [14] activating modules and alanine-rich or positive-charge-rich [15] repression modules, for example). RDs are also cataloged by functional context. For example, some activator and repressor modules are categorized according to the distance(s) from the promoter at which they function.

Abbreviations
ATF artificial transcription factor
DBD DNA-binding domain
KRAB Kruppel-associated box
NLS nuclear localization signal
PNA peptide nucleic acid
RD regulatory domain
TFO triplex-forming oligonucleotide

Addresses
*Department of Biochemistry and The Genome Center, 433 Babcock Drive, University of Wisconsin-Madison, Madison, WI 53706, USA
†Department of Chemistry and Department of Medicinal Chemistry, 930 N University Avenue, University of Michigan, Ann Arbor, MI 48109-1055, USA

Current Opinion in Chemical Biology 2002, 6:765–772
1367-5931/02/$ — see front matter © 2002 Elsevier Science Ltd. All rights reserved.
Published online 18 October 2002
or by the specific stages of transcription at which they exert their influence [16–18]. This includes, for example, activating domains that work on transcription initiation, those that work on elongation, and others that stimulate both steps of the process [16–18].

**Modular design of ATFs**

The earliest ATFs were those that were generated in the domain-swapping experiments that defined the modular nature of natural transcriptional regulators [1••]. Thus, subsequent efforts to generate ATFs with broader and ‘programmable’ DNA targeting and regulatory features have replaced the modules present in natural transcription regulators with designed counterparts [8•]. In Figure 3 we summarize the domains that have been used to generate ATFs and discuss their development below.

**DNA-binding domains**

Because the DNA-binding module, in great measure, specifies which genes are to be bound and regulated, it has been the focus of the most attention in the development of ATFs. Protein DBDs have been attractive targets because of the high specificity and affinity that they typically display for their target sequences. Despite a rich base of structural knowledge of protein DBDs, no clear ‘recognition code’ has been deduced for the rational design of new recognition motifs to target desired DNA sequences [19••]. To fill this need, genetic selection approaches have been used to generate protein DNA-binding modules for specific sequences [20,21•,22]. Using these approaches, zinc-finger-based DBDs that target unique promoter sequences have been isolated and shown to function in cultured cell lines [22].
The development of non-peptidic DBDs has provided exciting alternatives for ATF design and function in recent years [23]. As illustrated in Figure 3, these include the triplex-forming oligonucleotides (TFOs), which recognize specific DNA sequences through the formation of Hoogsteen base pairs and have been utilized as a DNA-binding module [24]. Similarly, peptide nucleic acids (PNAs), in which the phosphodiester backbone has been replaced with amide bonds, have been used with some success to target specific DNA sequences [25]. Also noteworthy are the DNA-binding modules inspired by the natural product distamycin. Unlike most DBDs, these ‘polyamides’ bind in the minor groove of DNA; Dervan and co-workers [6] have used rational design to develop molecules that recognize each of the four possible DNA base pairs.

Regulatory domains: activation
Activating modules that bear multiple acidic and hydrophobic residues (‘acidic activators’) function quite robustly in all eukaryotes tested and, as a result, have been frequently used in ATFs. Designed zinc fingers fused to one such acidic activating domain, the potent viral co-activator VP16, have been used to up-regulate the endogenous genes erbB-2 by 6- to 8-fold [26] and VEGF-A by 15- to 30-fold [27] in cultured cell lines. The minimal segment of VP16 that retains activator function [28] has also been used in several ATFs (Figure 3). Recently, Yaghmai and Cutting [29••] demonstrated that the potency of zinc finger ATFs is dependent upon cell type, the distance of the DBD sites from the transcription start site, and the number of VP16 activating domain repeats present. Attachment of one or two repeats of VP16 activating domain to TFOs also provides ATFs with activation levels in cell culture ranging from 5- to 40-fold [30•,31•]. When an analogous VP16 activating region was fused to hairpin polyamides, activation levels of 18- to 34-fold were observed in an in vitro system [32••,33]. Verdine and co-workers [34] have also reported that the minimal VP16 activating region composed of D-amino acids functions almost as well as the counterpart bearing L-amino acids.

The generation of novel peptidic activators has been accomplished by several approaches. One of the earliest studies used genetic screens with random sequences of Escherichia coli DNA attached to natural DBDs. The activating modules identified contained an excess of acidic residues, analogous to natural activating regions [1,35]. A recent study utilizing randomized octapeptides provided several hydrophobic sequences that activated transcription in yeast as robustly as the strongest natural activators when fused to protein DBDs [36•]. More recently, attempts at broadening the repertoire by creating regulatory modules that specifically interact with one or another component of the transcriptional machinery have yielded peptides that interact with a human histone acetyl-transferase [37], the yeast repressor Gal80 [38], or the yeast mediator component Gal11 (AK Mapp, AR Minter, G Belanger and JK Lum, unpublished data). The Gal80-binding peptides, despite targeting a yeast repressor, function as activators both in yeast and in mammalian cells (T Kodadek and Y Han, personal communication).

Genetic screens have also yielded RNA molecules that activate up to 100-fold in budding yeast (Saccharomyces cerevisiae) when tethered to DNA. This is the first reported example of a non-peptidic activator [39] (S Saha, AZ Ansari, K Jarrell and M Ptashne, unpublished data).

Regulatory domains: repression
RDs that repress transcription have not been as clearly defined. A few short peptides that function as repressor
modules have been identified, and these commonly contain a preponderance of alanine and/or basic residues; these, however, function only modestly [15]. One interesting example is the tetrapeptide WRPW that, when fused to a DBD, interacts with the Drosophila repressor Groucho (TLE in humans), which then recruits proteins that modify chromatin structure and actively repress the expression of neighboring genes [40]. In yeast, cysteine-rich peptides have been reported to localize the gene to which they are tethered via a DBD to the nuclear envelope and, by an unknown mechanism, ‘silence’ its expression [41]. Repressive ATFs have also been generated by fusing large segments of repressor proteins to a DBD. For example, when the Kruppel-associated box (KRAB) repressor module is fused to designed zinc fingers, a modest repression of the adjacent gene is observed [26,29••]. It is not yet clear, however, how the various repressor modules described above will function when fused to synthetic DBDs. This remains a relatively unexplored area of ATF design and function.

More typically, inhibition of gene expression with an ATF is achieved by competition for the DNA binding site of an

---

**Figure 3**

<table>
<thead>
<tr>
<th>DBD</th>
<th>Linker</th>
<th>RD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>Gal4</td>
<td>Activation domain</td>
</tr>
<tr>
<td>HTH</td>
<td>PEG</td>
<td>VP16</td>
</tr>
<tr>
<td>Zinc finger</td>
<td>Peptide</td>
<td>[...DFDLDMLG...]</td>
</tr>
<tr>
<td>Non-protein</td>
<td>(GGSGGS)</td>
<td>AH</td>
</tr>
<tr>
<td>* TFO</td>
<td>(PPPP)</td>
<td>PEFPGIELQELQE</td>
</tr>
<tr>
<td>* RNA</td>
<td></td>
<td>LOALQQQ</td>
</tr>
<tr>
<td>Polyamide</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3**

Modules used in the design of ATFs. Representative examples of the DBDs, RDs and linkers used in the design of most ATFs are summarized. Asterisks denote modules for which little structural data are available. Common protein-based DBDs include Gal4 [59], helix-turn-helix proteins (HTH; a generic binding scheme is shown) such as LexA, LacI and TetR, and both natural and designed zinc-finger motifs (structure from [60]) [11••]. The most frequently used non-protein DBDs are TFOs, PNAs and polyamides (for structure see [61]). While the linker is most commonly a flexible (GGSGGS) or rigid (polyproline [50]) peptide, a simple flexible linker derived from polyethylene glycol (PEG) can also be used without loss of function. Small molecules such as the FK506-CsA dimer [49] or the dimer of methotrexate (Mtx) and a synthetic analog of FK506 (SLF) [52] can also bridge the DBDs and RDs by non-covalent interactions [48]. Most activating regions have been derived from natural activator proteins such as VP16 [8••,12] or designed to mimic natural activating regions (AH) [32••,35]. One exception is the use of RNA hairpins [39] (S Saha, AZ Ansari, K Jarrell and M Ptashne, unpublished results). The repression domain most often used is derived from the KRAB protein [26,29••] but small peptides such as WRPW [40] and CCVC [41] can also function as repressors when bound to DNA. In many cases, the modules shown can be readily swapped to generate ATFs that target different DNA sequences or have a different regulatory function.
endogenous transcription factor. A breakthrough in this approach was achieved when polyamides designed to target transcription factor binding sites were successfully used to inhibit the expression of HIV genes in cell culture [42]. Recently, Laemmli and co-workers [43•] designed polyamides to target specific sequences in Drosophila melanogaster. When fed to developing flies, a distinct phenotypic change was caused by the specific polyamides but not by those designed to target a different sequence. Alternatively, PNA-based inhibitors have been used to inhibit transcriptional elongation of the oncogenic c-myc gene in Burkitt’s lymphoma and in prostate cancer cells [44•,45]. In general, the inhibition strategy demands precise knowledge of which DNA binding sites to target in vivo. Moreover, the requirement that the ATF out-compete the endogenous factor for binding to its cognate DNA site may be hard to achieve for every factor [46••], especially in cases where a large number of transcription factors bind cooperatively to adjacent sites [1••]. Thus, it may be more desirable in the long term to generate ATFs bearing repression modules that can actively silence transcription of a gene even when bound at a distal site on the promoter that does not overlap with a binding site of an endogenous transcription factor.

**Linking the DNA-binding and regulatory domains**

To achieve activation or active repression, the DBD and RD must be linked together to function as an ATF [32••]. Most commonly, the two domains are tethered via peptide linkers. However, a covalent linker is not essential; non-covalent interactions that bridge a DBD to an RD suffice to generate an ATF (Figure 3). These bridging non-covalent interactions can be mediated by protein–protein [1••], protein–RNA [47], or protein–small molecule binding events [48,49]. Systematic studies with flexible chemical linkers as well as rigid peptide linkers have shown that the linker length is an important variable in optimizing the function of the regulatory module [32••,33,50].

Some of the small molecule linkers mentioned above have the additional benefit of providing external control over the function of intracellular ATFs. Two pioneering examples of ligand-dependent ATFs were provided by the groups of Bujard and Schreiber. Bujard and co-workers [51] developed ATFs that bind to DNA only in the presence of doxycyclin, whereas Schreiber and co-workers [49] used chemical inducers of dimerization (CID) to mediate the interaction of a DBD and a RD in eukaryotes. In a recent development, Cornish and co-workers [52] reported a novel CID composed of methotrexate and a synthetic analog of the natural product FK506 to manipulate the interaction of a DBD with an activating region that functions robustly in bacteria. Inspired by earlier work on steroid-responsive nuclear hormone receptors, Barbas and co-workers [53•] described an allosteric ligand-responsive ATF constructed by linking a zinc finger DBD to a VP16 AD via the ligand-binding domain of a steroid receptor. These strategies continue to provide powerful tools for dissecting regulatory pathways in transcription and may also prove important for the development of ATF-based therapies.

**Traffic and delivery**

Delivery of ATFs into cells and organisms continues to be a considerable obstacle. For example, a recent report by Dervan and co-workers demonstrates that the trafficking of polyamide ATFs into cell nuclei is strongly dependent upon cell type, with cultured and primary T-cells showing the most nuclear localization [54•]. TFO-based regulators are typically delivered into cells either by electroporation or by cationic liposomes [30•,31•]. This is an unsustainable way to modulate transcriptional regulation and it severely limits the tissues that can be targeted. By contrast, the covalent attachment of a short peptide such as the nuclear localization signal (NLS) or of the steroid dihydroxytestosterone to PNA-containing ATFs results in effective delivery into cell nuclei [44•,45]. In addition, an arginine-rich peptide derived from the HIV transcriptional regulator Tat, peptidomimetic derivatives of Tat, and certain hydrophobic peptides can impart potent cell membrane transducing properties on the molecules to which they are coupled [55–58]. Peptides that serve as delivery modules could also be expressed contiguous with the protein-based

---

**Figure 4**

ATFs that respond to physiological cues. An intriguing goal for the design of future ATFs involves the harnessing of signal transduction pathways to activate dormant ATFs. (a) Physiological signals are often monitored by cell surface receptors. (b) The interaction between a signal and the receptor typically triggers an intracellular signal transduction pathway. (c) A dormant ATF containing a signal-responsive functional moiety would be converted to its functional state upon sensing the signal. (d) The functional ATF would then regulate specifically targeted genes in either a positive or negative manner. ATFs that could integrate instructions from different signal transduction pathways or respond only to signals in certain tissues would provide an even greater level of temporal and spatial control over the regulation of specific genes or gene clusters.
ATFs; this may enable delivery of these molecules into cells without gene therapy. This strategy could make ATFs of all varieties more viable in vivo.

**Future design considerations**

As summarized in Figure 3, most ATFs to date have been generated by mixing and matching the DBD, the RD, and, in some cases, the linker between the two. Although the ability to substitute the DBD and RD has been an essential design feature of the current generation of ATFs, additional features must be included in the coming generations to achieve greater control in regulating targeted genes. To achieve a further degree of subtlety in transcriptional control, the next generation of ATFs may also have to include properties found in natural transcriptional regulators such as the ability to interact only with cell type-specific components of the transcriptional machinery and the ability to respond to physiological cues.

The ability to sustain the desired expression level of targeted genes over a prolonged period also remains a significant hurdle. Small molecule RDs are attractive targets in this context because they would probably be resistant to proteolytic degradation in the cell. Small molecule regulators might also overcome some of the limitations faced by peptidic RDs, such as poor cell permeability, the potential for triggering an immune response, or possible effects on function of other endogenous regulators by competing for common cellular targets. Furthermore, small molecules that interact with cell type-specific or species-specific components of the transcriptional machinery could provide an additional degree of functional specificity.

As mentioned above, a desirable RD feature is the capacity to provide a graded potency in response to physiological or therapeutic cues. Initial steps towards this control were achieved with the development of ligand-responsive ATFs as well as RDs with tunable potency. A more ambitious goal would be to design ATFs that are functionally dormant in cells until activated by physiological cues (Figure 4). Thus, ATFs that respond to natural cellular signal transduction pathways, and perhaps even integrate information from a combination of pathways, to regulate the expression of a set of genes, present an exciting horizon for the future.

**Conclusions**

The modular nature of natural transcriptional regulators has inspired a modular design of ATFs. The success of this approach has allayed early concerns that stereospecific presentation of regulatory modules would be necessary to generate a functional ATF. This modularity allows tremendous flexibility in design where desired features can be appended to the essential functional modules. Current developments in the design of ATFs allow us to probe the mechanistic role of various architectural and temporal features of transcriptional regulators. As our design of ATFs becomes more sophisticated, there is the real potential of developing regulatory molecules that may provide a powerful tool to dissect transcriptional networks (functional genomics) and perhaps create a new class of transcription-based therapeutics.

**Acknowledgements**

AZA thanks the University of Wisconsin for a Steenbock Career Development award. AKM acknowledges financial support for this work from the NIH (GM65330), the Burroughs Wellcome Fund, and the March of Dimes.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

* of special interest
** of outstanding interest


This book summarizes much that is known about transcriptional mechanisms, activators, repressors and the transcriptional machinery in an accessible format with many primary literature citations. The authors propose that 'regulated recruitment' is a guiding principle in most cellular processes.


This recent review discusses the relationship between disruption of a transcription factor and the onset of acute promyelocytic leukemia. It also proposes the use of ATFs as therapeutic agents for treatment of this disease.


This review outlines in detail the structure and function of natural and artificial transcriptional regulators. It contains an overview of the advantages and limitations of both protein and synthetic DBDs and a discussion of the current understanding of the structure-function relationship of RDs.


This recent review outlines the architectural principles governing protein-DNA interaction with relevant citations of the primary literature as well as other reviews on this topic.


**Model systems**

Abstracts and models of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
** of outstanding interest


This recent review discusses the relationship between disruption of a transcription factor and the onset of acute promyelocytic leukemia. It also proposes the use of ATFs as therapeutic agents for treatment of this disease.


This review outlines in detail the structure and function of natural and artificial transcriptional regulators. It contains an overview of the advantages and limitations of both protein and synthetic DBDs and a discussion of the current understanding of the structure-function relationship of RDs.


This recent review outlines the architectural principles governing protein-DNA interaction with relevant citations of the primary literature as well as other reviews on this topic.


36. Lu YX, Ansari AZ, Ptashne M: An artificial transcriptional activating region with unusual properties. Proc Natl Acad Sci USA 2000, 97:1988-1992. The authors report the isolation of ‘non-acidic’ octapeptides that function as well as the strongest natural yeast activators. As such, they are the smallest eukaryotic artificial activators.


43. Janssen S, Cuvier O, Muller M, Laemmli UK: Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to Drosophila melanogaster. Mol Cell 2001, 6:1013-1024. In this elegant study, the authors use polyamide-based ATFs to induce phenotypic changes in D. melanogaster. Here, the ATF functioned in an organism rather than in cultured cell lines.

44. Cutrona G, Carpenato EM, Ulivi M, Roncella S, Landt O, Ferrarin M, Boffa LC: Effects in live cells of a c-myc anti-gene PNA linked to a nuclear localization signal. Nat Biotechnol 2000, 18:300-303. The authors demonstrate that covalent attachment of a nuclear localization sequence to a PNA results in transport into cells and nuclei. The NLS does, however, appear to interfere with the ability of the PNA to target its specific DNA site.


53. Beerli RR, Schooper U, Dreier B, Barbas CF: Chemically regulated zinc finger transcription factors. J Biol Chem 2000, 275:32617-32627. This study represents a first step in designing ATFs that can be used to regulate gene expression in a temporal manner. This approach offers much
promise for the design of ATFs that respond to one or more physiological cues (Figure 4).


In this recent report, the authors examined the cellular uptake of fluorescently labeled polyamides in seven different cell lines and observed that five cell lines showed no discernable nuclear localization of the conjugates. Also noteworthy is the observation that cell fixing caused rapid uptake into the nuclei of all cell types examined, indicating the importance of live cell controls.


