

Regulating transcription: a chemical perspective

Anna K. Mapp*

Department of Chemistry and Department of Medicinal Chemistry, University of Michigan, MI, USA

Received 7th March 2003, Accepted 21st May 2003

First published as an Advance Article on the web 5th June 2003

Many human diseases are characterized by altered gene expression patterns caused by malfunctioning transcriptional regulators. This has spurred new efforts in the development of artificial transcription factors that regulate the expression of specific genes either positively or negatively. Despite impressive advances in the last decade, artificial transcription factors that reconstitute all of the functions of natural regulators are not yet a reality. Such factors will be powerful chemical tools for unraveling the mechanisms by which gene expression is regulated and in the long term offer considerable therapeutic potential.

Introduction

Nature employs a combinatorial approach to gene regulation, relying upon various assemblies of regulatory proteins to control the patterns of gene expression that dictate the fate and function of each cell. This strategy enables virtually all cells in an organism to carry the same genetic information yet display widely varying function simply by employing different expression patterns. Intra- and extracellular signals direct gene expression patterns by calling into action specific transcriptional regulatory factors.¹ Once mobilized, these factors act as constituents of multiprotein assemblies that cooperatively recruit various cellular machines to the relevant genes and regulate the expression of those genes either positively or negatively (Fig. 1). In yeast, for example, addition of galactose to the growth medium mobilizes multiple copies of the activator protein Gal4, resulting in a 1000-fold increase in transcription of the genes required for galactose metabolism.¹ Similarly, viral infection in humans stimulates the assembly of a complex array of activators onto DNA to form an 'enhanceosome' that up-regulates the interferon- β gene almost 100-fold.² In both of these examples, the DNA-bound activators

initiate transcription by interacting directly or indirectly with chromatin-remodeling enzymes, RNA polymerase II, and associated transcription factors.

Pioneering genetic, structural, and biochemical efforts combined with recent technological breakthroughs in genome-wide expression profiling have done much to advance our understanding of the sequence of events outlined above. Despite such advances, a complete picture of the protein-protein, protein-nucleic acid, and protein-small molecule interactions governing transcription remains elusive. Some of the most compelling questions in the field surround the mechanisms by which activators regulate transcription. There is little doubt that activator proteins interact with the transcriptional machinery to up-regulate transcription, but the relevant activator binding sites within the machinery have for the most part not yet been identified. Again using yeast as an example, the well-characterized activator Gal4 interacts *in vitro* with more than 10 components of the transcriptional machinery, and there is conflicting evidence as to how many of these interactions play a functional role *in vivo*.³ It remains an open question whether activator targets change depending upon the signals received or the specific gene being regulated or if there are privileged targets that are contacted by most if not all activators. When you consider that it is ensembles of activators that regulate the levels and time course of a gene's expression, the picture becomes even more complex. Furthermore, we are only at the beginning of an understanding of the factors that control the lifetime of an activator and initiate the cascade of events that lead to repression of a gene.⁴

The importance of addressing such mechanistic questions comes into focus when one considers the growing number of human diseases that are linked to malfunctioning transcriptional regulators and are characterized by altered patterns of gene expression.⁵ This has inspired increasing efforts in the chemical, biological, and medical community towards the design and synthesis of artificial transcription factors (ATFs), molecules that target specific genes and regulate their expression either positively or negatively.⁶ Such factors would be powerful chemical tools for defining the macromolecular interactions that dictate gene expression patterns. Perhaps even more exciting, fully functional ATFs would have significant therapeutic potential as agents that could be used to restore normal patterns of gene expression in diseased cells.

An historical perspective

The first artificial transcription factors were generated in the early 1980s in experiments aimed at defining the key structural and functional modules of endogenous transcriptional activators. In these so-called 'domain-swapping' experiments, the

Anna Mapp received an A.B. in chemistry from Bryn Mawr College in 1992 before moving on to the University of California, Berkeley to obtain the Ph.D. degree in organic chemistry with Professor Clayton Heathcock. She then moved to the California Institute of Technology in 1997 to complete postdoctoral work with Professor Peter B. Dervan. In the fall of 2000, she joined the faculty at the University of Michigan where her research group is focused upon using chemical approaches to better understand how genes are regulated.



Anna Mapp

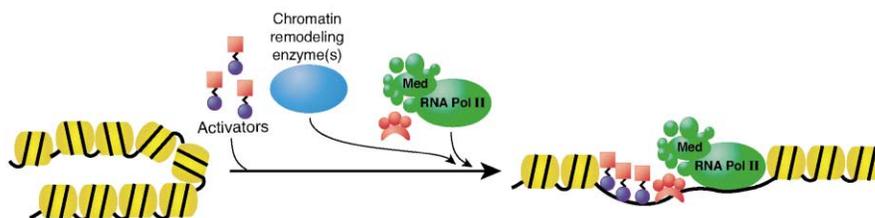


Fig. 1 A general model of initiation of transcription in eukaryotes. In the 'off' state, the DNA (black line) is wrapped around histones (yellow) and is thus largely inaccessible for binding. To initiate transcription, activator proteins bind to specific sites related to the gene and then recruit cellular machines that modify the DNA-histone complexes and enable RNA polymerase II and associated proteins to bind and transcribe.

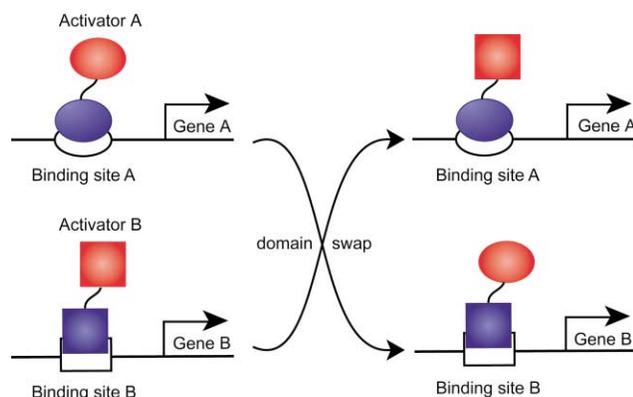


Fig. 2 Domain swapping experiments demonstrate that transcriptional activators are modular and that the domains are exchangeable. The DNA binding domains (DBD) are in blue while the activating domains (AD) are depicted in red; the linking domain connecting the two is in black.

DNA binding domain of one transcriptional activator was switched with that of another to generate a chimeric protein (Fig. 2).¹ When transcriptional activator A had its DNA binding domain replaced with that from activator B, the new hybrid protein bound to DNA binding site B and activated transcription of that gene efficiently; moreover, the new hybrid activator no longer functioned on gene A. These experiments defined the modular nature of transcriptional activators and revealed that most share two key structural features: an activating domain that contacts the transcriptional machinery and a DNA binding domain that localizes the protein to the appropriate site within the genome. The two domains can either reside within the same protein or in separate but interacting proteins. Repressor proteins are similarly modular. However, the DNA binding and repressor domains often do not reside within the same protein but rather are linked through non-covalent interactions.

On the heels of the early domain swapping experiments came detailed studies of the individual functional modules. Many DNA binding domains have since been structurally characterized and a wide variety of structural motifs have been observed in transcriptional activators and repressors.⁷ Activating regions have proven to be much more recalcitrant to structural characterization and are typically categorized by their amino acid sequence content: acid-rich, glutamine-rich, and proline-rich regions are common.⁸ Many activating regions are unstructured in solution and are thought to adopt recognizable secondary structures only when bound to their various partner proteins. The acid-rich activating region of the potent viral protein VP16, for example, becomes helical only upon binding to the human transcriptional machinery protein TAF_{II}31.⁹ These types of studies have been difficult both to carry out and interpret, however, since the relevant targets of most activating regions are unknown. Like activating regions, repressor domains are also often described by their amino acid content, and the structural requirements for repressor function remain relatively poorly defined.^{1,6}

Reconstituting endogenous regulator function

The modular nature of transcriptional regulators often belies the multitude of functions in which they participate. A transcriptional activator does much more than simply bind to a specific site on DNA and recruit the transcriptional machinery; it responds to external stimuli, traffics directly to the nucleus, binds to DNA as part of a multiprotein complex, interacts with many additional proteins in a cooperative manner, is subject to various covalent modifications (e.g., phosphorylation, glycosylation, ubiquitylation) that regulate its overall activity and lifetime, and can participate in more than one of the various steps of transcription. As the impetus for artificial transcription factor design grew, the central question that emerged was whether designed regulators could reproduce the specificity and complexity of function of the endogenous regulators.

To investigate this question, a modular replacement strategy in which the functional domains of endogenous transcriptional regulators are substituted with artificial counterparts has been adopted (Fig. 3).⁶ This strategy has been most successfully applied to the incorporation of novel DNA binding domains. Perhaps the most important question influencing the choice of DNA binding domain is *specificity* since the artificial regulator must target a specific site within the genome in order to influence only the desired gene, although affinity for the binding site also plays an important role. Given the size of the human genome, a DNA binding molecule must in principle target a sequence at least 15 base pairs in length to ensure recognition of a unique site. This is a truly daunting task since most metazoan transcriptional regulators cannot achieve this level of specificity without relying on cooperative interactions with additional proximally bound proteins.¹ It is likely, however, that recognition of a 15 base pair site will not always be necessary since much of genomic DNA is packaged in chromatin or heterochromatin and is thus less available for binding. Thus far several novel protein-based and nonprotein-based DNA binding modules targeting a range of binding site sizes have been employed in the design of artificial transcriptional activators that function well *in vitro* and/or in cell culture.^{10,11}

It has proven much more difficult to replace activating domains with non-natural counterparts. In most examples of ATF design the activating region used has been taken from an endogenous regulator, typically of the acid-rich variety (Fig. 4). A number of novel peptidic activating regions have been identified, however, utilizing a variety of genetic and biochemical techniques.^{1,6,12} In general, the activating domains identified through these studies are similar in sequence composition to the ubiquitous acid-rich activators and appear to function through similar mechanisms, although there are exceptions.¹³ Another biopolymer, RNA, also functions as an activating region when noncovalently linked to a DNA binding protein (Fig. 3B).¹⁴ There are obvious potential advantages of nonbiopolymer-based activating regions such as increased degradation resistance, cell permeability, target choice, and tunable potency, but despite these compelling incentives, none have been reported. The challenge thus facing chemists is to develop organic molecule scaffolds that incorporate the necessary functional groups

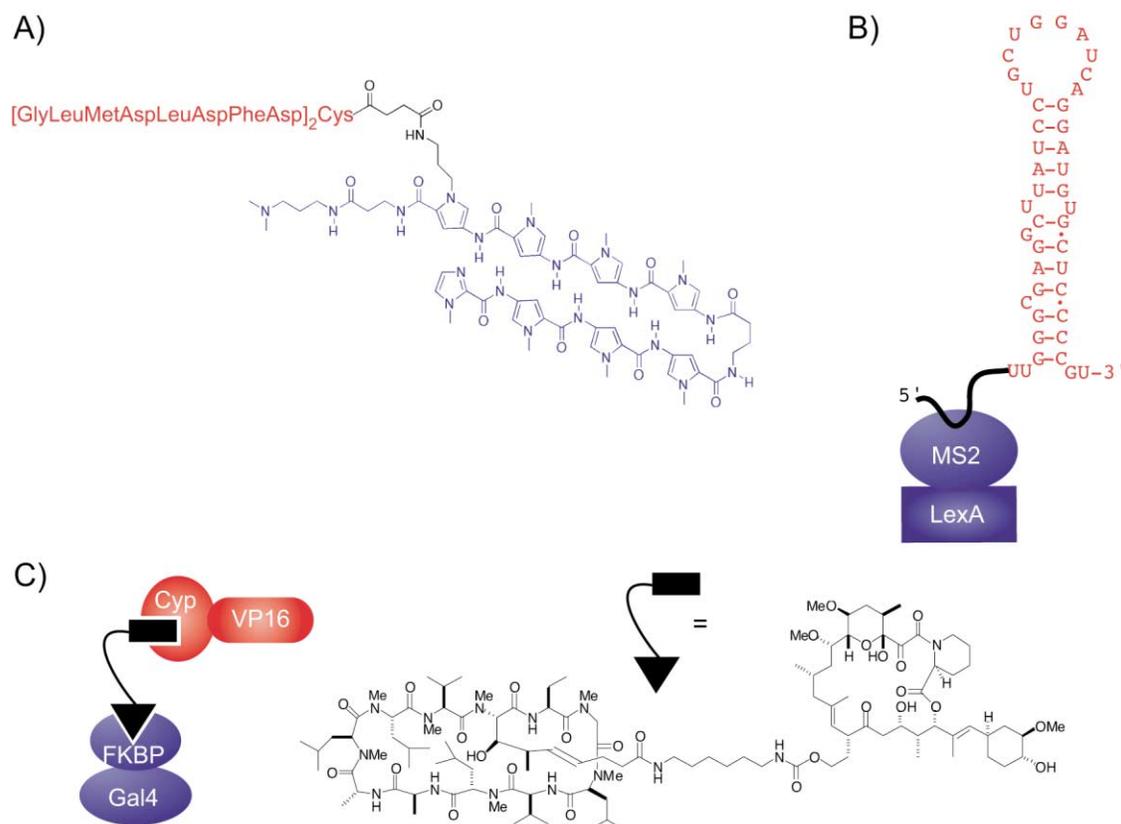


Fig. 3 Examples of artificial transcription factors designed using the modular replacement strategy. Each of these three examples contains a DNA binding domain (blue), an activating domain (red), and a linker connecting the two domains (black) by covalent (3A) or noncovalent (3B and C) interactions. A) In this example, a small molecule DNA binding domain, a hairpin polyamide, replaces the protein DNA binding domain. A peptidic activating domain derived from the potent viral coactivator VP16 is linked to the polyamide *via* a short tether. This ATF activates transcription 18- to 34-fold in an *in vitro* system.¹¹ B) RNA can also function as an activating region.¹⁴ The stem-loop structure shown activates transcription 100-fold in *Saccharomyces cerevisiae* when localized to DNA by binding of a linker RNA sequence to a fusion protein consisting of a protein DNA binding domain (LexA) and a RNA-binding protein (MS2). C) A chemical inducer of dimerization (CID) can also be used to link the DNA binding domain and the activation domain.¹⁹ The linker in this example is the heterodimer of the natural products FK506 and cyclosporin A, each of which bind specifically to the proteins FKBP and Cyp, respectively. In the presence of this linker, the DNA binding domain (a Gal4-FKBP fusion protein) and the activating domain (a VP16-CyP fusion protein) co-localize at a specific DNA binding site and upregulate transcription robustly in cell culture.²⁰

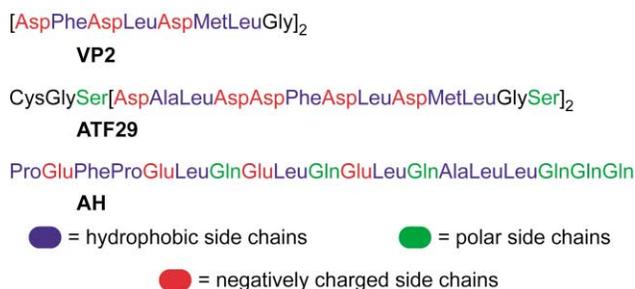


Fig. 4 Three of the most commonly used activating regions in artificial transcription factor design. The first two (VP2 and ATF29) are derived from the potent viral coactivator VP16 while the third sequence (AH) was designed to mimic acid-rich activating regions.¹

in the proper orientation for interaction with the transcriptional machinery; since the smallest peptide activators range from 8 to 16 amino acids in length, this may require an organic molecule with sizeable surface area. Moreover, natural activating regions often contain recognition sites for proteins in addition to their targets in the transcriptional machinery (repressors, kinases, or ligases, for example) and to be fully functional, artificial activating regions may also need to participate in multiple recognition events. ATFs incorporating artificial activating regions will certainly be important tools for addressing the mechanistic questions surrounding transcriptional activator function.

As of yet few artificial repressor domains have been identified and all have been peptide-based. These function only modestly relative to endogenous repressors when fused to protein DNA

binding domains and have not yet been tested when attached to DNA binding small molecules.⁶ Given the dearth of well-characterized repressor domains, the down-regulation of specific genes has more commonly been accomplished by competition for the DNA binding site of an endogenous transcription factor. In principle the chief advantage of this strategy is that it requires only a single functional module (a DNA binding domain) and can thus be accomplished with a structurally simpler molecule. The three most common classes of DNA binding molecules used for this strategy are triplex-forming oligonucleotides that target the major groove of DNA,¹⁵ peptide–nucleic acids that typically bind cognate sequences through strand-invasion of the DNA duplex,¹⁶ and polyamides composed of heterocyclic amino acids designed to recognize specific sites in the DNA minor groove (Fig. 3A).¹⁷ Indeed, DNA binding molecules from each of these classes have been successfully used to inhibit the transcription of specific genes *in vitro* and in cell culture.¹⁸ This strategy does, however, require one to know exactly which DNA binding sites to target, still not well characterized for many transcription factors. Furthermore, many transcription factors bind to DNA as part of a multi-protein complex and competing with this complex for specific DNA binding sites may be difficult to achieve in many circumstances. It thus remains a desirable goal to develop repressor domains that could be used to actively repress adjacent genes without having to compete with transcription factors.

In designing either a repressor or an activator the various domains must be linked in order for the ATF to be functional. Most commonly this is accomplished with a covalent linkage; a simple polyether tether linking the two domains, for example,

has proven effective in a number of systems.⁶ Non-covalent interactions can also be used, however, and ATFs that respond to external signals have been designed using this strategy.¹⁹ In the simplest case, the DNA binding domain and the activating region co-localize at the DNA binding site only in the presence of a specific small molecule, the external signal (Fig. 3C).²⁰ This provides temporal control over ATF function, certainly an important consideration for therapeutic potential, and it also increases the utility of ATFs as tools for unraveling transcriptional regulatory pathways. Small molecule signals have also been used to control other aspects of ATF function such as DNA binding and conformation and one can imagine future applications in transport and controlled degradation.²¹

Future directions

Enormous strides in artificial transcription factor design and function have been made in the last decade. Due to breakthrough accomplishments in the field of nucleic acid recognition, ATFs that target specific promoters have been successfully employed *in vitro* and in cell culture. An artificial transcription factor that reconstitutes all of the functions of an endogenous regulator remains in the distant horizon, however. Two barriers faced by all ATFs are those imposed by the cell and nuclear membranes, and while a number of innovative strategies for enhancing the transport of small and large molecules into the nucleus have been developed, a general solution is not yet at hand. Of course the most glaring omission in the field is the lack of nonbiopolymer-based activation or repression domains. Small molecules that can effectively compete with protein–protein surface interactions, a long-standing challenge in bioorganic chemistry, will be required to accomplish this goal.²² Success in this area will thus have impact far beyond the development of artificial transcriptional regulators. Given the multitude of challenges and the exciting potential of artificial transcription factors as both chemical tools and therapeutic agents, this will continue to be fertile ground for scientific exploration in the years to come.

Acknowledgements

I am grateful for financial support from the Burroughs Wellcome Fund (New Investigator in the Toxicological Sciences), the March of Dimes (Basil O'Connor Starter Scholar), the NIH (GM65330), and the University of Michigan. I thank Professor A. Ansari for helpful discussions.

References

- 1 M. Ptashne and A. Gann, *Genes & Signals*, Cold Spring Harbor Laboratory, New York, 2001.
- 2 T. Agalioiti, S. Lomvardas, B. Parekh, J. Yie, T. Maniatis and D. Thanos, *Cell*, 2000, **103**, 667.
- 3 K. Melcher and S. A. Johnston, *Mol. Cell Biol.*, 1995, **15**, 2839; Y. Q. Xie, C. Denison, S. H. Yang, D. A. Fancy and T. Kodadek,

- J. Biol. Chem.*, 2000, **275**, 31914; S. S. Koh, A. Z. Ansari, M. Ptashne and R. A. Young, *Mol. Cell*, 1998, **1**, 895; Y. C. Lee, J. M. Park, S. Min, S. J. Han and Y. J. Kim, *Mol. Cell Biol.*, 1999, **19**, 2967; Y. B. Wu, R. J. Reece and M. Ptashne, *EMBO J.*, 1996, **15**, 3951; J. M. Park, H. S. Kim, S. J. Han, M. S. Hwang, Y. C. Lee and Y. J. Kim, *Mol. Cell Biol.*, 2000, **20**, 8709.
- 4 M. Muratani and W. R. Tansey, *Nat. Rev. Mol. Cell Biol.*, 2003, **4**, 192.
- 5 P. P. Pandolfi, *Oncogene*, 2001, **20**, 3116; C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Aksien, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein, *Nature*, 2000, **406**, 747; X. Chen, S. T. Cheung, S. So, S. T. Fan, C. Barry, J. Higgins, K. M. Lai, J. F. Ji, S. Dudoit, I. O. L. Ng, M. van de Rijn, D. Botstein and P. O. Brown, *Mol. Biol. Cell*, 2002, **13**, 1929.
- 6 A. Z. Ansari and A. K. Mapp, *Curr. Opin. Chem. Biol.*, 2002, **6**, 765.
- 7 C. W. Garvie and C. Wolberger, *Mol. Cell*, 2001, **8**, 937.
- 8 H. P. Gerber, K. Seipel, O. Georgiev, M. Hofferer, M. Hug, S. Rusconi and W. Schaffner, *Science*, 1994, **263**, 808; A. J. Courey and R. Tjian, *Cell*, 1988, **55**, 887; N. Mermoud, E. A. Oneill, T. J. Kelly and R. Tjian, *Cell*, 1989, **58**, 741.
- 9 M. Uesugi, O. Nyanguile, H. Lu, A. J. Levine and G. L. Verdine, *Science*, 1997, **277**, 1310.
- 10 R. R. Beerli, B. Dreier and C. F. Barbas, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 1495; S. Kuznetsova, S. Ait-Si-Ali, I. Nagibneva, F. Troalen, J. P. Le Villain, A. Harel-Bellan and F. Svinarchuk, *Nucleic Acids Res.*, 1999, **27**, 3995; D. Stanojevic and R. A. Young, *Biochemistry*, 2002, **41**, 7209.
- 11 A. K. Mapp, A. Z. Ansari, M. Ptashne and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 3930.
- 12 J. Ma and M. Ptashne, *Cell*, 1987, **51**, 113; J. V. Frangioni, L. M. LaRicca, L. C. Cantley and M. R. Montminy, *Nat. Biotechnol.*, 2000, **18**, 1080; Y. Han and T. Kodadek, *J. Biol. Chem.*, 2000, **275**, 14979.
- 13 Z. Lu, A. Z. Ansari, X. Y. Lu, A. Ogirala and M. Ptashne, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 8591.
- 14 S. Saha, A. Z. Ansari, K. A. Jarrell and M. Ptashne, *Nucleic Acids Res.*, 2003, **31**, 1565; D. J. Sengupta, M. Wickens and S. Fields, *RNA*, 1999, **5**, 596.
- 15 M. Faria and C. Giovannangeli, *J. Gene. Med.*, 2001, **3**, 299.
- 16 P. E. Nielsen, *Methods Enzymol.*, 2001, **340**, 329.
- 17 P. B. Dervan, *Bioorg. Med. Chem.*, 2001, **9**, 2215.
- 18 L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 12890; S. Janssen, O. Cuvier, M. Muller and U. K. Laemmli, *Mol. Cell*, 2000, **6**, 1013; M. Faria, C. D. Wood, L. Perrouault, J. S. Nelson, A. Winter, M. R. H. White, C. Helene and C. Giovannangeli, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 3862; G. Cutrona, E. M. Carpaneto, M. Ulivi, S. Roncella, O. Landt, M. Ferrarini and L. C. Boffa, *Nat. Biotechnol.*, 2000, **18**, 300; R. Besch, C. Giovannangeli, C. Kammerbauer and K. Degitz, *J. Biol. Chem.*, 2002, **277**, 32473.
- 19 H. Lin and V. W. Cornish, *Angew. Chem.*, 2001, **40**, 871.
- 20 P. J. Belshaw, S. N. Ho, G. R. Crabtree and S. L. Schreiber, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 4604.
- 21 R. R. Beerli, U. Schopfer, B. Dreier and C. F. Barbas, *J. Biol. Chem.*, 2000, **275**, 32617; Q. Lin, C. F. Barbas and P. G. Schultz, *J. Am. Chem. Soc.*, 2003, **125**, 612; U. Baron and H. Bujard, *Methods Enzymol.*, 2000, **327**, 401.
- 22 A. G. Cochran, *Curr. Opin. Chem. Biol.*, 2001, **5**, 654; P. L. Toogood, *J. Med. Chem.*, 2002, **45**, 1543.