Small-molecule-based strategies for controlling gene expression

Carilee Denison and Thomas Kodadek

A central goal in chemical biology is to gain control over biological pathways using small molecules, and the mRNA-synthesizing machinery is a particularly important target. New advances in our understanding of transcriptional regulation suggests strategies to manipulate these pathways using small molecules.

Address: Center for Biomedical Inventions, Departments of Internal Medicine and Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard., Dallas, TX 75235-8573, USA.

Correspondence: Thomas Kodadek
E-mail: kodadek@ryburn.swmed.edu

Chemistry & Biology June 1998, **5**:R129–R145
http://biomednet.com/elecref/10745521005R0145

© Current Biology Ltd ISSN 1074-5521

Introduction

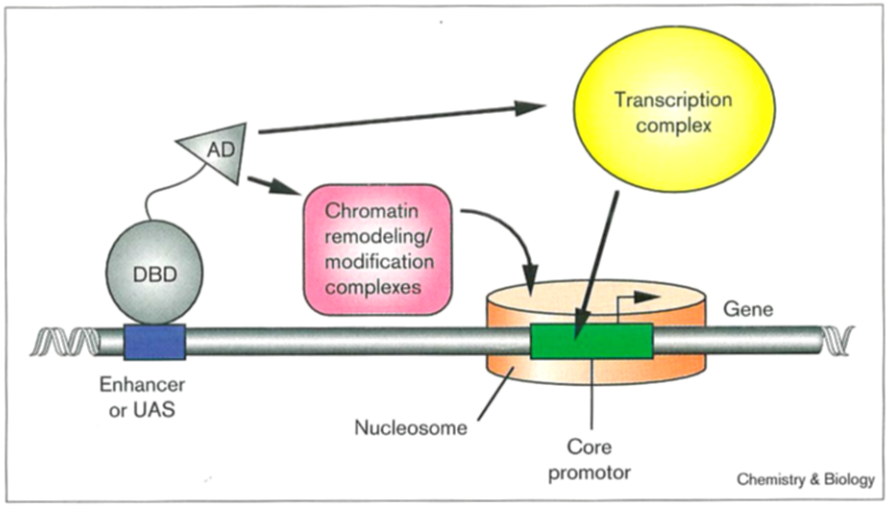
Perhaps the most important regulatory networks in eukaryotic cells are those that control the activity of the mRNA-synthesizing machinery, which includes RNA polymerase II and a host of general transcription factors. This is because the expression of the vast majority of eukaryotic genes is controlled at the level of transcription by activators and repressors. Such proteins function in a gene-specific fashion to up-regulate or down-regulate, respectively, the ability of the transcriptional machinery to synthesize mRNA encoded by the target gene. In turn, the activities of transcriptional activators and repressors are usually regulated by signal transduction cascades that transfer information from the cell membrane to the nucleus. The ultimate goal of molecular medicine is to gain control over these processes and turn particular genes on or off at will. Chemical biology represents one of the two most promising approaches to achieve this goal, the other being gene therapy.

The area of small-molecule-regulated gene expression is reviewed briefly in this article. First, the state of the art of the field is discussed, focusing on work carried out using derivatives of naturally occurring immunosuppressants. Then, the complex protein machinery that mediates mRNA synthesis in eukaryotic cells is discussed, followed by a brief review of the nature of the activators and repressors that regulate the activity of the transcriptional machinery. We then consider potential approaches to finding small molecules that would allow one to manipulate the interactions of native transcription factors and signal-transduction proteins with one another and with nucleic acids. Finally, we consider possibilities for the *de novo* synthesis of small molecules that can regulate transcription directly by acting as activator or repressor mimics.

Natural products as regulators of gene expression

All organisms contain a large number of genes whose levels of expression are controlled by small molecules. For example, many genes involved in biosynthetic pathways are feedback-regulated by a build up of the product of the pathway. Conversely, the expression of catabolic genes, such as those that encode proteins involved in sugar metabolism, are stimulated tremendously when the cell is in a meduim rich in the corresponding sugar. Nature therefore uses small-molecule-controlled gene expression routinely. In most of such cases, binding of the small molecule to a membrane-bound, or sometimes soluble, receptor that is highly specific for recognizing that molecule triggers a cascade that ultimately either stimulates or inhibits the

Figure 1



Schematic view of an activator. The DNA-binding domain (DBD) allows the protein to bind in the vicinity of the target gene. The attached activation domain (AD) makes contacts with chromatin remodeling and modifying complexes as well as the transcription complex, serving to 'open' the chromatin structure and assist association of the transcription complex with the promoter. The activities of the DBD and AD are, to a first approximation, separable and there is not a requirement for a distinct stereochemical relationship between the two domains.

activity of the mRNA-synthesizing machinery at a particular set of genes. In most cases, the ultimate recipients of the signals are transcriptional activators and/or repressors. Activators, which will be discussed in more detail below, generally consist of quasi-separable [1,2] DNA-binding and 'activation' domains (Figure 1). The DNA-binding domain allows the protein to be localized in the vicinity of the target gene, whereas the activation domain binds directly to components of the transcriptional machinery or to chromatin-remodelling complexes, contacts that result in greatly increased transcription of the target genes.

Repressors can block the activity of activators directly, for example by binding to the activation domain and sequestering it from the transcriptional machinery, or they can function indirectly, for example by mediating changes in chromatin structure that block access of transcription proteins to the DNA.

The balance between activators and repressors can be affected by small molecules in a number of ways. For example, the yeast Gal4 protein (Gal4p), a potent activator of genes involved in galactose metabolism [3], is normally muzzled by a specific repressor, Gal80 protein (Gal80p), that binds tightly to the Gal4 activation domain [4]. When the galactose concentration in the medium is increased, the sugar binds to the Gal3 signal-transduction protein, which then acts upon the Gal4–Gal80 complex in an ATP-dependent fashion to relieve the repressive effect of Gal80 and expose Gal4p's activation surface [5]. The result is a huge induction in the expression of the GAL genes.

This theme of small-molecule-dependent release of an activator from a repressive interaction is very common. For example, the important activator NF-κB is sequestered in an inactive complex in the cytoplasm by the protein IκB. Exposure of cells to a variety of stimuli, including small molecules such as phorbol esters, results in the phosphorylation and subsequent proteasome-mediated degradation of IκB. As a result, NF-κB is released and it moves into the nucleus where it binds to DNA and, in concert with other activators, drives the transcription of genes involved in a number of important processes [6], including inflammation and limb development.

Another example of this type of activation mechanism of particular interest to chemical biologists is the action of the clinically used immunosuppressants FK-506 and cyclosporin A (CsA; Figure 2). Although the examples given above were responses to natural stimuli, developmental signals or cellular stress, FK-506 and CsA are produced by soil microorganisms and are not normally present in human cells. These compounds were discovered on the basis of their ability to suppress immune function. Their mechanism of action, which is now understood in considerable detail [7,8], provides an excellent paradigm for the kind of molecular tool for manipulating gene expression that chemical biologists would like to have.

A crucial step in the induction of an immune response is the translocation of the nuclear factor of activated T cells (NFAT) from the cytoplasm to the nucleus of T cells, where it activates the transcription of several genes, including that encoding interleukin-2. The event that triggers nuclear translocation is calcineurin-mediated dephosphorylation of cytoplasmic NFAT (Figure 2). Both FK-506 and CsA interfere with this process by first binding with high affinity to their respective target proteins, FKBP [9] and cyclophilin [10]. Remarkably, the composite surfaces of both the FKBP–FK-506 and cyclophilin–CsA complexes bind calcineurin tightly [11] and inhibit its ability to dephosphorylate NFAT [12,13],

Figure 2

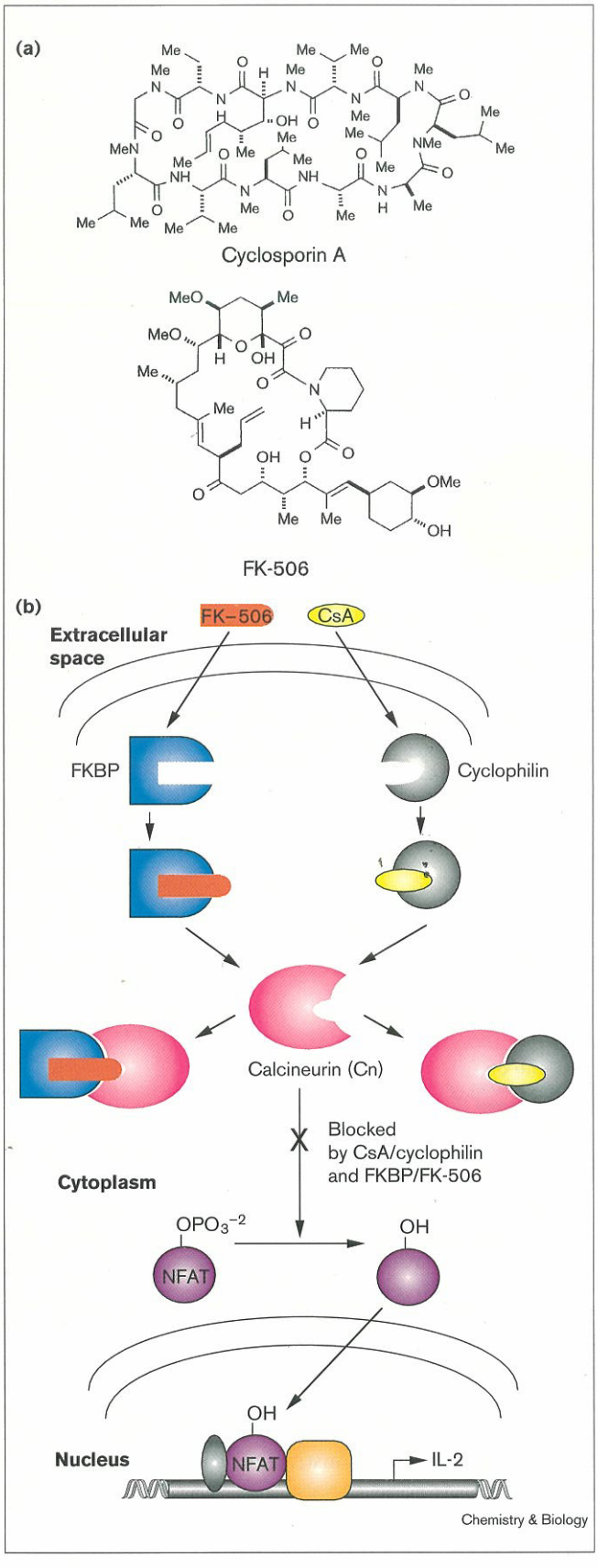


Figure 2

(a) Structures of the immunosuppressants FK-506 and cyclosporin A (CsA). **(b)** Schematic model for the mechanism of action of FK-506 and CsA. Each binds an intracellular receptor (FKBP and cyclophilin, respectively). The protein–small molecule complexes bind and inhibit calcineurin, a phosphatase, which blocks dephosphorylation and nuclear translocation of cytoplasmic nuclear factor for T-cell activation (NFAT). NFAT, along with other gene-specific transcription factors (unlabeled in the figure), is required for transcription of interleukin-2 (IL-2) and other genes involved in generating an immune response.

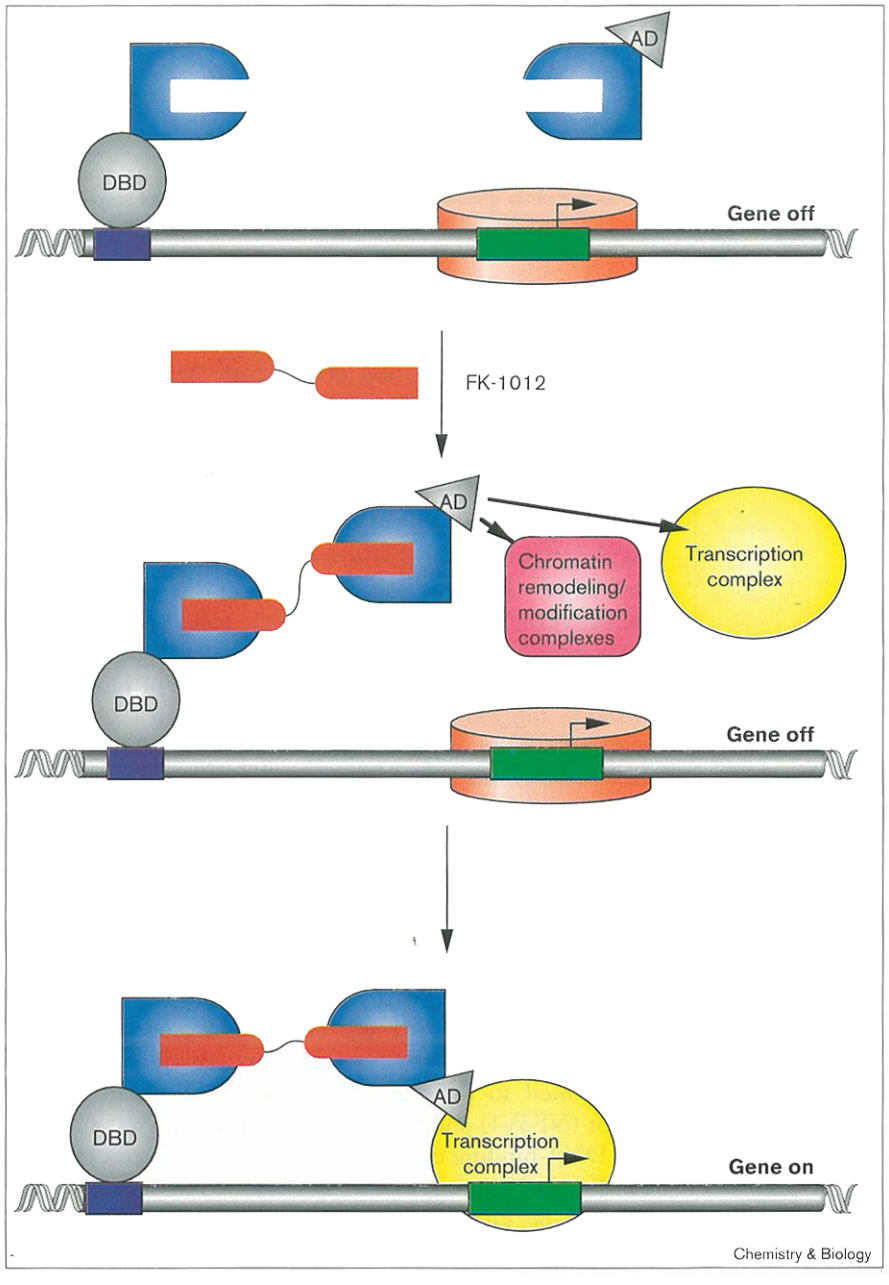
thereby blocking NFAT-activated interleukin-2 gene expression. Thus, CsA and FK-506 can be seen as 'molecular matchmakers' that bring about the association of two proteins that normally do not interact and, in so doing, indirectly down-regulate a particular pathway of gene expression.

Small molecule-mediated control of gene expression in engineered cells

Schreiber, Crabtree and coworkers [14] have used synthetic versions of these remarkable natural products to manipulate protein–protein interactions *in vivo* (Figure 3). The crux of this work is that two proteins of interest are fused to the protein receptors for FK-506 or CsA (FKBP and cyclophilin, respectively) at the DNA level and then expressed in the cell type of interest. The association of the two engineered proteins can then be triggered by the addition of cell-permeable, homodimeric or heterodimeric constructs in which two immunosuppressant molecules have been linked covalently ($(FK\text{-}506)_2$, CsA_2 or FK-CsA) [15]. If mere proximity of the tagged proteins, as opposed to specific interactions, is sufficient to elicit a biological response, then this will occur. For example, an FK-506 homodimer (FK-1012) was introduced into cells engineered to express two fusion proteins, one in which FKBP was linked to the activation domain of a transcription factor (NF3V1) and another in which FKBP was fused to the DNA-binding domain of a transcription factor (GF3 or HF3). As transcriptional activation requires physical linkage of DNA-binding and activation domains (see above), but not any sort of specific interaction between these domains, FK-1012-induced association of these fusion proteins resulted in the transcription of the otherwise silent target genes [16]. Related chemically induced dimerization experiments have been performed using fusions of FKBP or cyclophilin with signal transduction factors, such as the Fas and TCR cell-surface receptors Raf, Src or Sos. These experiments that manipulate signal transduction factors also result in the control of gene expression, but intervene at a point far upstream of the actual transcription machinery.

These experiments illustrate the exciting potential of using small molecules to control gene expression in living cells. They promise to allow scientists and doctors to turn

Figure 3



Manipulation of chimeric transcription factors using a small molecule allows the expression of a reporter gene to be regulated by a cell-permeable small molecule. FK-1012 is a synthetic homodimer containing two tethered FK-506 molecules (see Figure 2a).

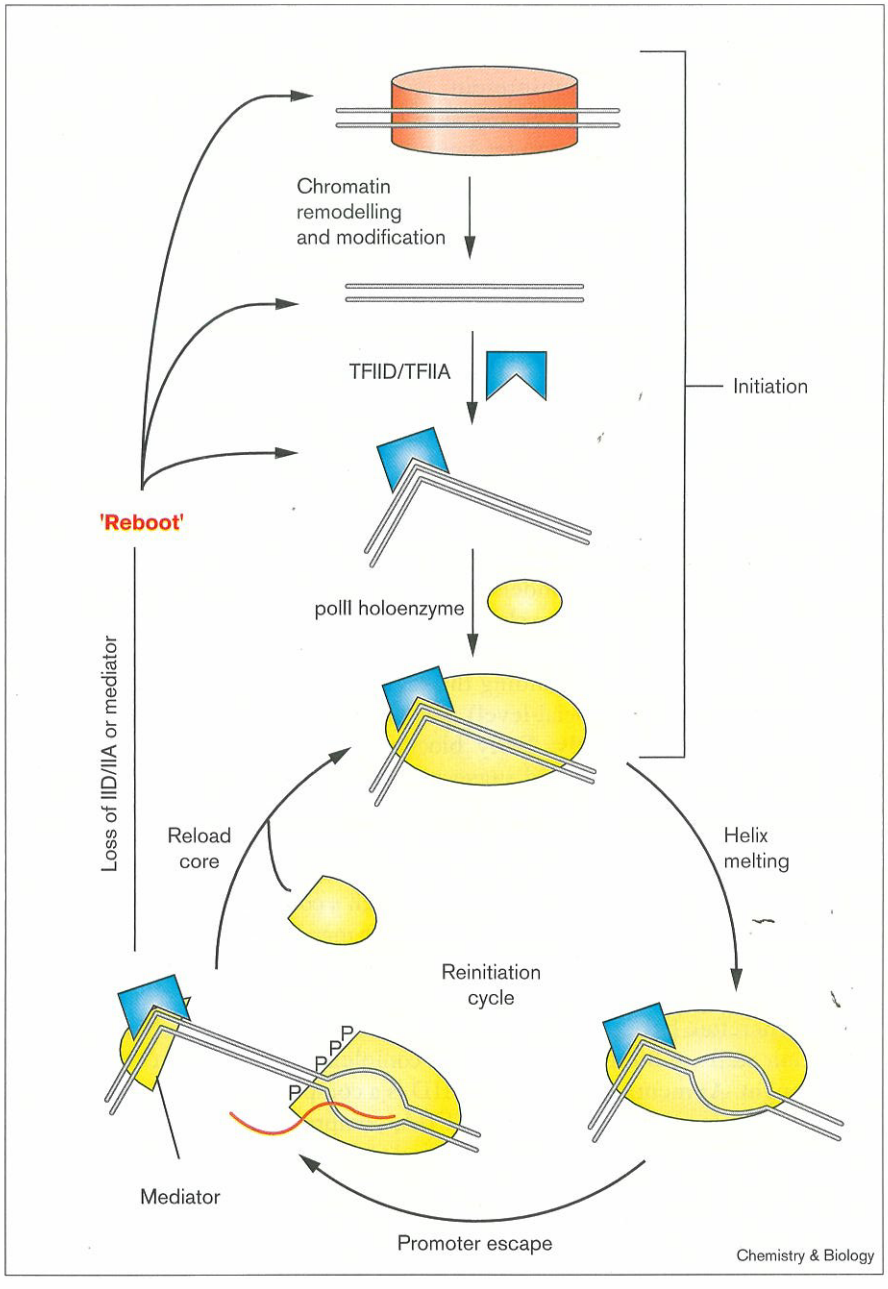
specific genes on and off at will using cell-permeable molecules. As the vast majority of regulation over biological pathways occurs at the level of signal transduction/transcription, such a technology would revolutionize molecular biology and medicine. The limitation with current technology is that immunosuppressant-derived dimerizers can only allow the researcher to manipulate artificial proteins in which an immunophilin or FKBP has been fused to the protein(s) of interest. A key goal is therefore to find compounds that can be used to manipulate the interactions of wild-type macromolecules in nonengineered cells.

The transcription cycle: a complex symphony or a Texas two-step?

To tackle the problem of how to manipulate the transcriptional apparatus using small molecules, it would obviously be helpful to have a sophisticated understanding of how the mRNA-synthesizing machinery in eukaryotic cells works. The properties of eukaryotic transcription proteins have been reviewed exhaustively elsewhere [17,18]. Only a brief overview of the process will therefore be presented here, with a focus on understanding what activators and repressors do to modulate the efficiency of transcription.

Figure 4

A model for the transcription cycle in eukaryotic cells. When a gene is first induced, the chromatin structure must be 'loosened' (the nucleosome is shown as disappearing for simplicity; this is probably not the case), TFIID/TFIIA must associate with the core promoter and an intact holoenzyme must bind to the TFIID/TFIIA–DNA complex. This provides the first complete preinitiation complex. This species must then open the double helix to expose the template strand, then move away from the promoter and initiate mRNA synthesis. Promoter escape involves hyperphosphorylation of the carboxy-terminal domain of the polII largest subunit. TFIID/TFIIA and the mediator fragment of the holoenzyme are thought to remain at the promoter after polII and associated factors leave. To rebuild another transcription complex, only the core fragment of the holoenzyme must add. High-level transcription is the result of many reinitiation cycles. If TFIID/TFIIA or the mediator is lost before a new holoenzyme core fragment can associate, the system must fall back to some step in the initiation cycle, which is much slower. It is proposed that a major role of activators is to stabilize mediator at the promoter to facilitate multiple rounds of reinitiation.



RNA polymerase II (polII), the enzyme responsible for the transcription of all mRNA-encoding genes, is comprised of 12 polypeptides and operates in concert with a large number of general transcription factors (TFIIA, TFIIB, TFIID, etc.). Perhaps the most important of these is TFIID [19], a complex of about 13 proteins that includes the TATA-binding protein (TBP) [20,21] and TBP-associated factors (TAFs) [22]. TBP is a sequence-specific DNA-binding protein that recognizes the so-called TATA boxes (consensus: 5′-TATAAAA) present in the promoters of many genes [23,24]. One or more of the TAFs might also have DNA-binding properties [25–27]. TFIID, all of the general transcription factors and polII must assemble on the promoter to form a preinitiation complex in order to begin a transcription cycle (Figure 4). This is followed by an ATP-dependent melting of the promoter region, allowing the polymerase to associate with the template strand. Many of the protein–protein and protein–DNA interactions used to form the complex in the first place must then be severed to allow polII and

some fraction of the preinitiation complex to escape the promoter and begin their march down the gene. In this process, the elongation complex picks up a number of specialized elongation accessory factors [28] and also associates with other multiprotein complexes, for example the splicosome and the excision repair machinery. The transition from a promoter-bound to an elongating polymerase complex involves covalent modifications, in particular multiple phosphorylations of the carboxy-terminal domain (CTD) of the polII largest subunit [29]. Finally, whatever vestige of the preinitiation complex remains at the promoter must accept a new, hypophosphorylated polymerase and its attendant factors to rebuild a new preinitiation complex (Figure 4). This cycle must occur many times, because a highly active gene fires approximately every five seconds.

The complexity of the transcription machinery is daunting. The fully formed preinitiation complex has a mass greater than that of a ribosome, the network of protein–protein interactions is only partially understood, and any or all of the steps in the transcription cycle could be regulated by activators or repressors. Fortunately, recent advances in this field suggest that understanding the regulation of this process (at least at a superficial level) may not be as difficult as was feared originally. Early biochemical experiments using purified factors had suggested that formation of a preinitiation complex required a host of sequential general transcription-factor-binding events [30], leading to an almost palpable depression in the field regarding the prospects of understanding the regulation of such a complicated pathway in detail. This view has now changed dramatically with the realization that the vast majority of transcription factors travel as large, stable complexes. One does not therefore have to think about dozens of individual association steps to build a preinititation complex, as was once thought. As mentioned above, TFIID is a complex of about 13 proteins. It associates with TFIIA, comprised of three polypeptides, which helps TFIID bind to DNA, possibly by competing repressors from TBP [31,32]. Most or all of the rest of the general transcription factors, polII, and a class of proteins known as coactivators (see below) then associate in one step as parts of a huge complex known as the RNA polymerase II holoenzyme [33,34]. TFIIB, a holoenzyme component, binds to TBP [35,36], locking the components of the machine together into a single piece. It now seems likely that assembly of the preinitiation complex may require only two steps: TFIID/TFIIA–DNA binding, followed by association of the holoenzyme with this complex [37] (Figure 4).

How do activators and repressors work?

The holoenzyme is comprised of two parts. One is the so-called 'core' that includes RNA polymerase and all of the other proteins required for synthesizing mRNA. The other is the mediator [38], a complex of ~20 proteins that is required for the holoenzyme to respond to activators *in vitro* and *in vivo*. The mediator is linked to the holoenzyme through the CTD [39]. There is circumstantial evidence that this association is lost after the first firing of the promoter; polII and many associated factors move down the gene, whereas the mediator and TFIID are thought to stay behind [40,41]. This probably makes reinitiation (synthesis of transcripts 2-n) much more facile than initiation, because a stable base for formation of subsequent preinitiation complexes is present and only a fragment of the holoenzyme must reassociate. It is reasonable to assume that for highly active genes the level of mRNA synthesis is closely correlated with the number of reinitiation events for each initiation event. Once the system drops out of the reinitiation cycle as a result of loss of TFIID or mediator from the promoter, it must 'reboot' completely (Figure 4), which is probably slow. Activators clearly play an important role in reinitiation [16] and it is therefore reasonable to suggest that the major role of activators is to help to retain the mediator at the promoter during reinitiation. Some activators may also help to retain TFIID [42].

This mechanistic picture suggests that the level of transcription can be modulated by the lifetimes of the TFIID–DNA and activator–mediator complexes to give greater or lesser amounts of gene expression. The longer-lived these complexes, the more rounds of reinitiation that will occur prior to rebooting. This view is consistent with the current literature. For example, activators, such as Gal4p, that have very high affinities for mediator (C.-J. Jeong, L. Sun, S.-H. Yang, T.K. and S.A. Johnston, unpublished observations) are unusually potent activators, but only on promoters with high affinity TATA boxes. Mutations in the TATA box (Y. Xie, S.-H. Yang, L. Sun and T.K., unpublished observations) or TBP [43] that reduce the lifetime of the complex correlate directly with reduced levels of activator-mediated gene expression *in vivo*. This type of information is very important to the chemical biologist. In addition to substantiating the view of activator function presented above, the effect of point mutations provides an excellent signpost indicating which steps in a biological process should be able to be manipulated using small molecules.

Some activators may also play a role in recruiting TFIID and/or holoenzyme to promoters through direct contacts with TBP during the first initiation event [44–46]. A number of papers have also argued that activators play a major role in recruiting TFIID to promoters or maintaining it there during reinitiation through interactions with TAFs [47–49]. Several recent studies indicate that this is not a major pathway of activation *in vivo* [50–53]. Mutations that affect the rate of TFIID association with the promoter during initiation could increase the lag time between the time of induction and the onset of transcription, but would not have corresponding effects on steady-state transcription

levels. Of course, if initiation were very severely crippled, transcription would be abolished. Thus, the transcription process can be likened to a light controlled by a dimmer switch, in which the circuit must be opened for any light to be produced (initiation), but the overall output of light is controlled by a knob that can be set to any desired level (number of reinitiation events per initiation).

Of course, there are many exceptions to the above picture and no single model will describe the mechanism of action of all activators. For example, genes such as *Drosophila* Hsp70 and HIV genes activated by Tat [54,55] are clearly regulated at the level of promoter escape [56], possibly through activator-mediated recruitment of a kinase that phosphorylates the CTD and severs the association of polII with promoter-bound proteins. Small-molecule inhibitors of this kinase have been identified [57].

Activators and repressors also function at the level of chromatin structure. Chromatin is a repressed template and the first order of business in transcribing a gene must be to 'loosen up' the chromatin structure in order to promote transcription-factor binding [58–60]. The loosening of chromatin structure occurs through activator-mediated recruitment of two types of chromatin modification/remodeling complexes. One class is the histone acetyl transferases (HATs), which contain proteins that acetylate key lysine residues in the amino-terminal tails of certain histones [61]. In some way that is not yet understood, the covalent modification renders the DNA in a nucleosome far more accessible to DNA-binding transcription factors. The importance of HATs in gene activation is underscored by the recent demonstration that certain gene-specific transcriptional repressors act by recruiting a histone deacetylase complex to the target promoter, thus shutting down transcription [62,63]. This is a result that is particularly satisfying to chemical biologists, because histone deacetylase was first purified and characterized on the basis of its binding to trapoxin, a small molecule that blocks histone deacetylation *in vivo* [64]. A different type of complex that also functions at the level of chromatin is typified by the SWI/SNF chromatin remodeller in yeast [65–68] that somehow 'jiggles' core nucleosomes using energy derived from ATP hydrolysis to facilitate transcription-factor binding.

Activators can recruit these complexes in two ways. One is through direct binding. For example, the activator VP16 has been shown to bind to a protein called ADA2 [69] which is part of a multiprotein complex that also includes GCN5, a potent HAT [70]. Alternatively, there are indications that TFIID and the holoenzyme may have associated with them proteins that have HAT and/or chromatin-remodelling activity [71,72], so binding of the activation domain to one or both of these complexes may automatically recruit these activities.

As an aside, understanding how chromatin structure influences transcription is an area of tremendous opportunity for chemical biologists. No one has even the first clue regarding the structural changes in chromatin structure brought about by acetylation or ATP-dependent remodelling. Also, there are so many different HATs it seems likely that different HATs may alter chromatin structure in different ways, so the situation is almost certainly far more complicated than current models would suggest. This area of research is crying out for new probes that will allow investigators to ask and answer more detailed questions.

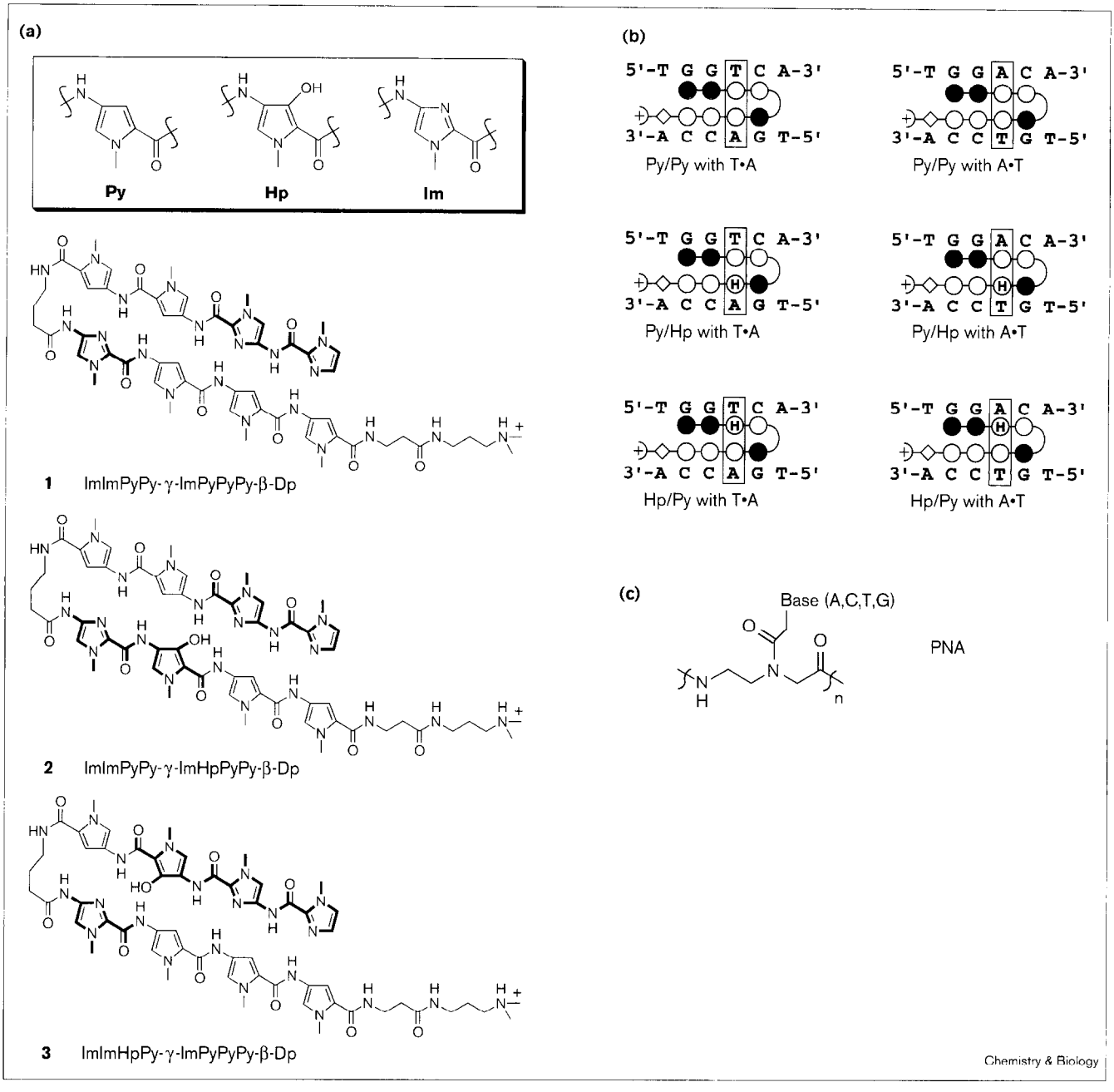
Finally, it is important to point out that the promoters of most eukaryotic genes have binding sites for more than one activator. These different proteins often interact with one another in synergistic fashion [73] and little or no gene expression results unless all of the activators are bound. Very often, this is because the proteins bind to the promoter cooperatively [74–76]. Thus, another tempting target for manipulating transcription are the activator–activator complexes that support cooperative binding.

DNA–protein interactions as molecular targets

As the examples above should have made clear, there are a number of potential strategies for manipulating the transcription process using small molecules. Of course, compounds that fundamentally alter the activity of the transcriptional machinery itself, for example an inhibitor of polII elongation, would be potent modulators of transcription but would not be gene-specific. Most strategies have therefore focused on compounds that target either the promoter of interest itself, or the activators and repressors that bind to it. The most obvious approach is to develop molecules that block activator–DNA or repressor–DNA interactions and thereby turn genes off or on artificially. Another strategy would be to find small molecules that could promote or antagonize key nuclear protein–protein interactions involved in regulation, for example between cooperating activators, between repressors and histone deacetylases, between repressors and activation domains, and possibly even between activation domains and their targets in the transcription machinery. Finally, for many genes (for example those activated by NFAT) it would be advantageous to manipulate the activity of kinases, phosphorylases or proteases that modulate the activity of an activator or repressor or control its nuclear localization. In any case, the development of protocols to design or discover small molecules that can manipulate protein–protein or protein–DNA interactions is a high priority for chemical biologists interested in manipulating gene regulation. Some particularly interesting recent advances in this area will be discussed below.

By far the most work has been carried out on compounds that bind DNA and that might serve as inhibitors of binding of proteins to overlapping sites. In particular, two types of

Figure 5



DNA-binding synthetic oligomers. **(a)** Structures of polyamides. Hp, 3-hydroxypyrrole; Im, imidazole; Py, pyrrole; β, β-alanine; γ, γ-aminobutyric acid; Dp, dimethylaminopropylamide. **(b)** Binding models for polyamides 1–3 in complex with 5′-TGGTCA-3′ and 5′-TGGACA-3′. Filled and unfilled circles represent imidazole and pyrrole rings respectively; circles containing an H represent 3-hydroxypyrrole; the curved line represents γ-aminobutyric acid; the diamond represents β-alanine; + represents the positively charged dimethyl-aminopropylamide tail group. **(c)** Basic structure of peptide nucleic acid (PNA). Parts (a,b) reproduced with permission from [86].

compounds look very promising in this role. Nielsen [77] has pioneered the development of protein nucleic acids (PNAs), which are oligomers that contain the standard purine and pyrimidine bases of an oligonucleotide but in which the sugar–phosphate backbone is replaced with a simple amide-based chain (Figure 5) [77]. PNAs bind with very high affinities to complementary single-stranded nucleic acids (both DNAs and RNAs), in fact better than standard oligonucleotides because of the lack of repulsive phosphate–phosphate interactions. Indeed, a PNA complementary to one strand of a DNA duplex will invade the double helix, pair with its complement and displace the

'like strand', forming a 'displacement loop' [78], at least in low ionic strength buffers. As might be expected, PNA invasion of a DNA duplex can abolish binding of a protein to an overlapping site and PNAs have also been employed as potent antisense agents. *In vitro*, PNAs have proven to be useful reagents for manipulating transcription and translation. Unfortunately, PNAs are not very cell-permeable, which has greatly limited their use in living cells [79]. Recently, there have been many exciting advances in moving cell-impermeable molecules through cell membranes using special peptides, however [80]. It could be that peptides conjugated to the appropriate PNA could be potent agents for manipulating gene regulation.

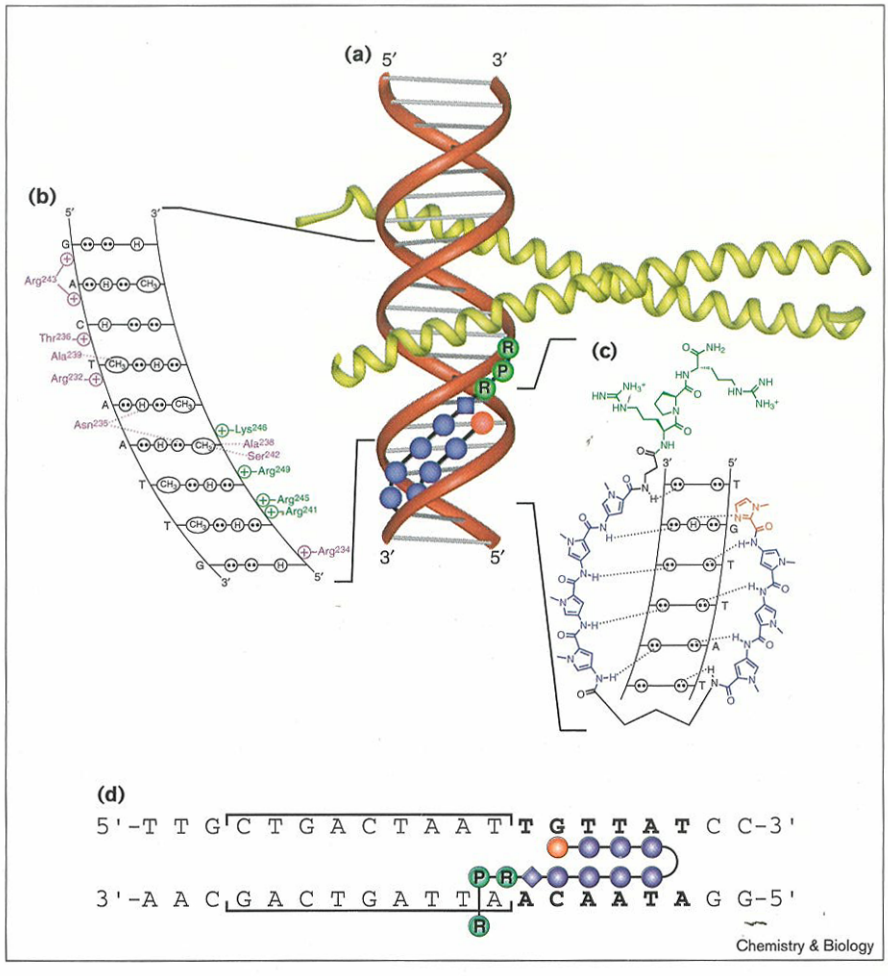
The other class of molecules that shows great promise consists of the remarkable oligomers of modified *N*-methylimidazoles and pyrroles developed by Dervan and coworkers (Figure 5) [81] (also see [82,83] for related work). These compounds were inspired by distamycin, netropsin and other minor-groove-binding natural products. It was hoped that, through both rational and some irrational experimentation, netropsin-like molecules could be made that would have greater sequence discriminatory powers than the natural products, which bind mainly A/T-rich regions. A seminal advance was the realization from nuclear magnetic resonance (NMR) experiments and other biophysical studies that two distamycins were stacked in an 'antiparallel' fashion into the minor groove in these complexes [84]. This insight led to the development of 'hairpin' oligomers that mimicked this 2:1 binding mode, with very high binding constants. Over several years, substituted imidazole and pyrrole compounds were developed that allowed recognition of any the four natural Watson–Crick base pairs in the context of this conserved structural motif [85,86]. In other words, there is now a 'code' by which a given imidazole/pyrrole pair can be selected to bind a particular base pair of DNA. An imidazole/pyrrole oligomer complementary to any sequence of double-stranded DNA can thus be designed [87,88] with little more difficulty than one would have in coming up with an oligonucleotide complementary to a piece of single-stranded DNA. This work represents a major advance in biomolecular recognition.

As one might expect, these compounds are potent inhibitors of protein–DNA interactions when minor groove contacts are critical for protein binding [89]. As most proteins make predominantly major-groove contacts, the imidazole/pyrrole oligomers will probably have to be elaborated to serve as generally useful inhibitors of sequence-specific protein–DNA interactions. There would appear to be many straightforward ways to accomplish this. For example, a recent paper [90] describes the inhibition of binding of a fragment of the yeast GCN4 protein to DNA *in vitro* using a phosphate-interference strategy. It has been known for some time that alkylation of even a single key phosphate can prevent binding of many DNA-binding proteins to their target sites, reflecting the fact that interactions between cationic or polar sidechains and the charged DNA backbone make critical contributions to the negative free energy of protein binding. With this in mind, an imidazole/pyrrole oligomer designed to bind a sequence adjacent to the recognition site of the GCN4 protein [91] was coupled to the tripeptide Arg–Pro–Arg (RPR). The hope was that the cationic arginine sidechains would be brought by the oligomer into close proximity with phosphate groups in the GCN4-recognition site, allowing the formation of strong hydrogen bonds that would occlude binding of GCN4 protein (Figure 6). In fact, this approach worked nicely *in vitro* and should be generally useful for competing the binding of many DNA-binding proteins. Other strategies might have to be explored for applications in the living cell, however, because appending charged groups to the neutral oligomers is likely to reduce their cell permeability. Indeed, the imidazole/pyrrole oligomers are sufficiently new that they have not yet really been subjected to a 'shakedown cruise' in living cells, but initial experiments look very promising [89] and they clearly hold tremendous promise as reagents for the control of gene expression. In fact, it may not be necessary to modify these compounds to manipulate the binding of proteins in the major groove of DNA in order to regulate transcription *in vivo*. This is because stable binding of TBP to TATA boxes is an important event in the transcription of a great many genes, and TBP is a minor-groove-binding protein [92]. Although it was stated above that targeting general transcription factors is a poor strategy to achieve gene-specific regulation, this is an exception because the DNA is the true target. For example, although most TATA boxes more or less resemble the consensus 5′-TATAAAA-3′, an imidazole/pyrrole oligomer could be made that recognizes only part of this site and also binds to a flanking sequence that is unique to the target promoter. As the affinity of TBP for the TATA box is very often correlated directly with transcriptional output (Y. Xie, S.-H. Yang, L. Sun and T.K., unpublished observations, also see [43,93]), manipulation of the TBP–TATA interface using the imidazole/pyrrole oligomers may allow one to modulate, rather than completely abolish, mRNA production in a highly gene-specific fashion.

There have been several other scattered reports of DNA-targeted inhibitors of specific protein binding, for example chimeras including carbohydrates and DNA-reactive small molecules [94]. Some of these appear to be quite promising and may emerge as important reagents in the future [95]. But no other class of molecules currently approaches the general utility of the PNAs and imidazole/pyrrole oligomers.

Much less work has been done on the complementary strategy for manipulating DNA–protein interactions: finding molecules that have high affinity for the DNA-binding

Figure 6



(a) A schematic model of Arg–Pro–Arg (RPR) polyamides targeted to the major groove transcription factor GCN4. **(a)** The α-helical GCN4 dimer (yellow) is shown binding to adjacent major grooves [91]. The Arg–Pro–Arg–hairpin polyamide is shown as red, blue and green balls which represent imidazole, pyrrole and Arg–Pro–Arg amino acids, respectively. The blue diamond represents β-alanine. γ-Aminobutyric acid is designated as a curved line. **(b)** The contacts between one GCN4 monomer and the major groove of one half-site of 5′-CTGACTAAT-3′ are depicted (adapted from [91]). Circles with two dots represent the lone pairs of the N7 of purines, the O4 of thymine and the O6 of guanine. Circles containing an H represent the N6 and N4 hydrogens of the exocyclic amines of adenine and cytosine, respectively. The C5 methyl group of thymine is depicted as a circle with CH_3 inside. Protein sidechains that make hydrogen bonds or van der Waals contacts to the bases are shown in purple and connected to the DNA via a dotted line. Green and purple plus signs represent protein residues that electrostatically contact the phosphate backbone. The residues that are predicted to be disrupted by an Arg–Pro–Arg polyamide are shown in green. **(c)** The hydrogen-bonding model of the eight-ring hairpin polyamide ImPyPyPy-γ-PyPyPyPy-β-RPR bound to the minor groove of 5′-TGTTAT-3′. Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. Py and Im rings are represented as blue and red rings, respectively. The Arg–Pro–Arg moiety is green. **(d)** The model of the polyamide binding its target site (bold) adjacent to the GCN4 binding site (brackets). Polyamide residues are as in (a). Reproduced from [90].

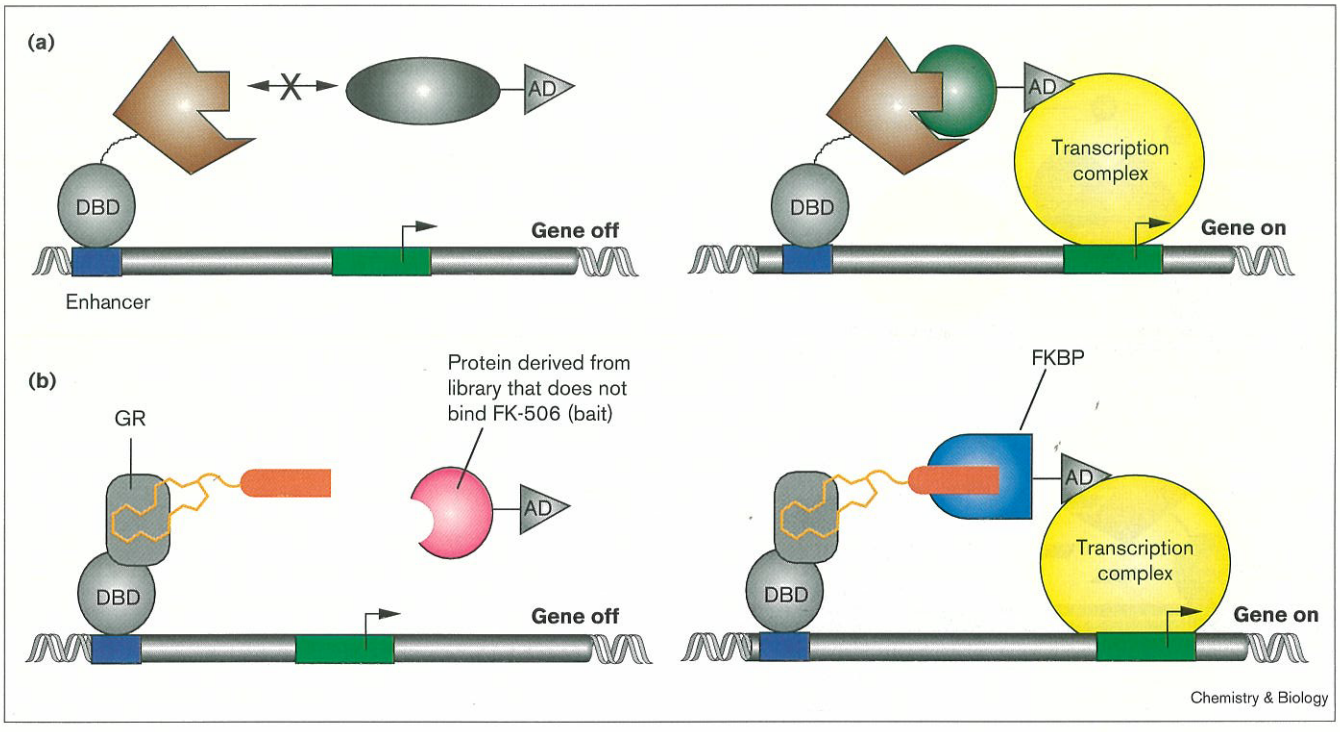
domains of key activators or repressors and therefore block their association with DNA. This is generally considered to be an even harder task than DNA recognition. Although broad structural families of DNA-binding domains certainly exist, polypeptide targets lack a single, well-defined architecture, which is a hallmark of the DNA double helix. Nonetheless, we predict that this approach will be a growth area in the future as chemical biologists begin to learn how to make molecules that bind specific peptide and protein sequences.

Targeting protein–protein interactions: better ways to find a needle in a haystack

Finding new molecular matchmakers or disrupters of protein–protein interactions is a very high priority for chemical biologists. Unnatural molecules that have these properties have been very hard to come by. Part of the reason for this is that pharmaceutical companies, where most protein-binding synthetic molecules come from, have traditionally concentrated on developing enzyme inhibitors rather than manipulators of protein–protein interactions. This will almost certainly change. Once these efforts are brought up to speed it will be critical to already have general assays by which libraries, combinatorial or otherwise, can be screened for molecules that have the property of interest, because rational design is unlikely to succeed in most cases. Although it is increasingly common to screen libraries for molecules that bind a given target protein, finding a matchmaker or disrupter is a difficult process because only a fraction of the molecules that bind a particular protein will influence its interaction with other factors.

Recently, there has been important progress in the design of high-throughput screens or selections designed to identify

Figure 7

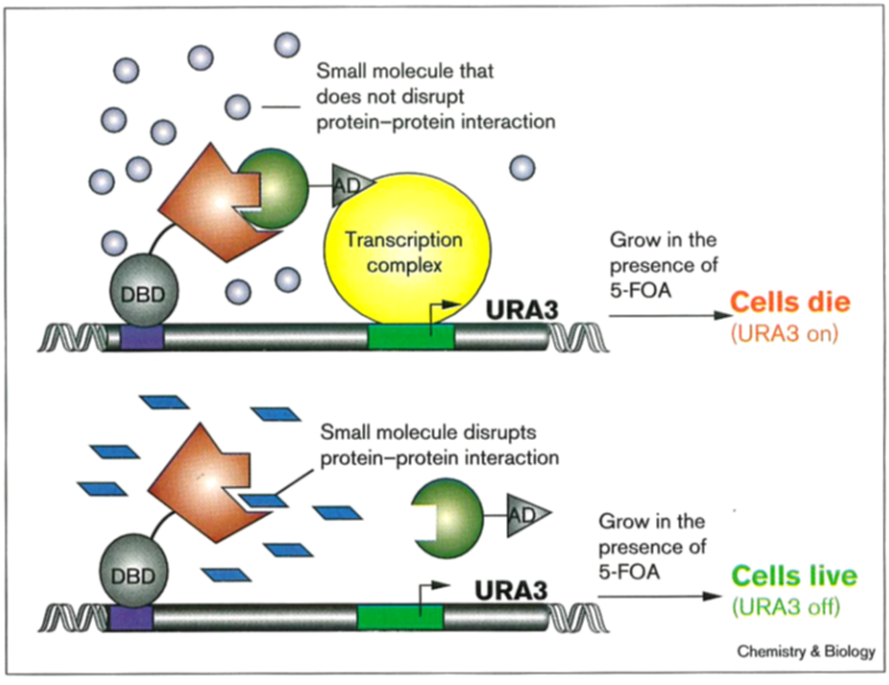


(a). Schematic representation of the two-hybrid system, a genetic method used to detect protein–protein interactions. If the proteins fused to a DNA-binding domain and an activation domain do not interact, then transcription of a reporter gene will be very low. If these proteins do interact, however, then a functional activator will be reconstituted and the reporter gene will be expressed at high levels. **(b).** The three-hybrid assay to detect proteins that bind a given small molecule. The orange steroid-shaped symbol represents dexamethasone. The red bullet represents FK-506. GR, glucocorticoid receptor. See text for details.

those molecules with the desired matchmaker or disrupter function. Much of this work was inspired by systems set up by geneticists (who have been practising combinatorial chemistry of a sort for much longer than chemists have) to identify proteins that interact with one another. Generically known as two-hybrid assays [96], this family of methods takes advantage of the fact that, in many promoter contexts, the DNA-binding and activation domains of an activator function more or less independently of one another (see [97,98] for exceptions), but must be physically connected. For example, if the Gal4-activation and DNA-binding domains are severed and these fragments are expressed in a yeast strain deleted for wild-type GAL4, no transcription of Gal4p-targeted genes will occur. If the genes encoding proteins X and Y are fused to the DNA encoding the severed GAL4 domains, and X and Y bind to one another, Gal4p activity will be reconstituted and transcription of the target genes will occur (Figure 7a). To make this system more convenient, strains have been constructed in which activated transcription of a target gene is essential for cell survival, making the process a straightforward selection for protein–protein interactions. Using this approach, it is now routine to screen genomic cDNA libraries fused to the activation domain for genes or gene fragments that encode polypeptides which bind to a particular 'bait' protein fused to the DNA-binding domain [99].

Many variations of this basic strategy have been reported for more specialized applications. Most relevant to this discussion are the 'three-hybrid' system and the 'reverse two-hybrid' system. The first, reported by Licitra and Liu [100], is a clever method to identify the protein targets of biologically active natural products. The technique employs the same strategy of reconsituting the activity of a severed transcriptional activator, but is designed such that a small molecule must bridge the interaction between the proteins fused to the DNA-binding and activation domains (Figure 7b). In a proof of principle experiment, the rat glucocorticoid receptor (GR) hormone-binding domain was fused to a sequence-specific DNA-binding domain and a cDNA library was fused to an activation domain. The screen was then carried out in the presence of a chimeric small molecule consisting of dexamethasone (a GR ligand) linked to FK-506. As expected, a screen for cells in which a reporter gene was activated resulted in the isolation of the gene encoding FKBP. This demonstrates the feasibility of using genetic screens for probing small-molecule–protein interactions *in vivo*. The reverse

Figure 8



The reverse three-hybrid system for detecting small molecules that disrupt a protein–protein interaction. See text for details.

two-hybrid system is a method to select for mutations that abrogate protein–protein interactions [101,102]. In this case, the 'reporter' gene targeted by the reconstituted activator is chosen such that its expression is toxic and therefore can be selected against.

Schreiber and coworkers [103] have recently combined elements of the two-hybrid, three-hybrid and reverse two-hybrid systems to create a convenient system for identifying small molecule disruptors of protein–protein interactions. Their approach is shown in Figure 8. As in the reverse two-hybrid system, yeast cells were engineered so that a protein–protein interaction which reconstitutes activator function is conditionally toxic. This was accomplished using a standard yeast genetics trick of placing the URA3 gene under the control of the severed activator and growing the cells in the presence of 5-fluoroorotic acid (5-FOA). When operated on by the URA3 gene product, 5-FOA is transformed into a toxic substance but in the absence of URA3 expression it is harmless. Alternatively, expression of URA3 is nontoxic in the absence of 5-FOA, allowing clones that contain interacting proteins to be grown and propagated easily. The fusion proteins containing the DNA-binding domain and activation domain were placed under the control of the Gal4 protein. Gal4p-mediated expression is essentially zero when the cells are grown in glucose, but occurs at high levels in galactose-containing media. Thus, both the expression of the interacting proteins and the consequences of their interaction can be controlled by the experimenter. The utility of this system was demonstrated by taking advantage of the fact that FK-506 inhibits the binding of FKBP to the transforming growth factor β type I receptor R1 [103]. As expected, growth of yeast containing R1 fused to a DNA-binding domain and FKBP fused to an activation domain was sensitive to the presence of 5-FOA, but this sensitivity was abrogated by FK-506.

Our laboratory has developed a different genetic assay (based on a method originally devised by Hu and coworkers [104]) in which two different fusion proteins, each containing the λ repressor DNA-binding domain, are expressed in *Escherichia coli* equipped with a repressor-controlled green fluorescent protein (GFP) gene. The fusion proteins lack the normal dimerization domain of the Repressor. If the proteins fused to the DNA-binding domain interact and artificially dimerize the repressor DNA-binding domain, GFP expression is therefore blocked (C. Ackerson and T.K., unpublished observations). If the fusion proteins do not heterodimerize, or if a molecule is present that blocks the interaction of the proteins fused to the repressor fragment, however, then GFP is expressed at high levels. These bright green cells are easy to identify in a background of dark cells.

'Spray and pray' and the 'squeege': combining the power of combinatorial libraries and genetic assays

The biological screens and selections described above are ideal for high-throughput screening protocols in which cells are introduced into the wells of 96 well (or denser)

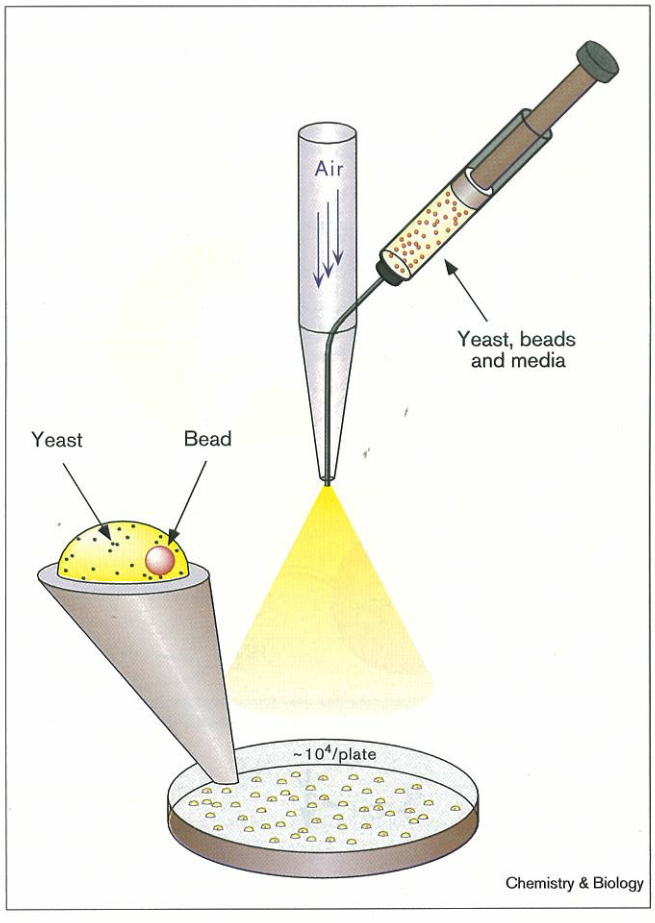
microtiter plates, each of which contains a different chemical. In this way, the entire suite of compounds possessed by a pharmaceutical company could be screened for matchmaker or disrupter activity in a reasonable period of time. Of potentially even greater use, however, would be to apply these techniques to screening combinatorial libraries made by the split and pool method [105] in which each bead is derivatized with many copies of a unique compound. The trick here would be to somehow expose the *E. coli* or yeast reporter strain to many, many different beads in a spatially segregated manner so that ideally one yeast cell sees one bead in some kind of microincubator where the chemical compound can be released from the bead. This knotty problem has been solved elegantly in two ways. One, called the 'stochastic nanodroplet' method [106], employs the simple idea of mixing yeast cells and chemically modified beads together then spraying them as a fine mist onto an agar plate (Figure 9). If the flow and levels of yeast and beads are controlled appropriately, the 'nanodroplets' sprayed onto the plate will contain from zero to a few beads (or 0–1 if bead density is kept very low) per droplet as well as one to a few yeast cells. The nanodroplets are now spatially segregated on the plate. If yeast growth is unimpeded, each nanodroplet will give rise to a yeast colony. Borchardt *et al.* [106] used beads linked to rapamycin via a photocleavable linker to demonstrate that when the plates were photolyzed enough toxic rapamycin was released from the bead to diffuse into the yeast cells and strongly inhibit growth. In theory, the same approach could be employed using combinatorial libraries of compounds attached to the beads by the same photolabile linker and a yeast or *E. coli* reporter strain engineered to report on the state of a particular protein–protein interaction.

The second approach [107] also employs spatially segregated nanodroplets as mini-incubators. In this case, however, the bead/cell mixture is layered onto a plastic plate with small wells that are extremely closely packed. These plates are produced by a photolithographic/imprinting technique and precoated with a substance that makes the bottom of the wells cell-adherent. Again, the amounts of beads and cells are chosen so that after the excess liquid is 'squeegeed' off the plate, the nanodroplets that remain have one to a few beads and a few cells in them. The advantage of this technique is that the squeegee procedure is much more gentle than the spraying technique and even much more fragile mammalian cells can be used in this format. This combination of genetic selection and combinatorial chemistry technologies promises to be an extremely effective route to the discovery of small molecule disrupters and matchmakers.

Synthetic mimics of activators and repressors

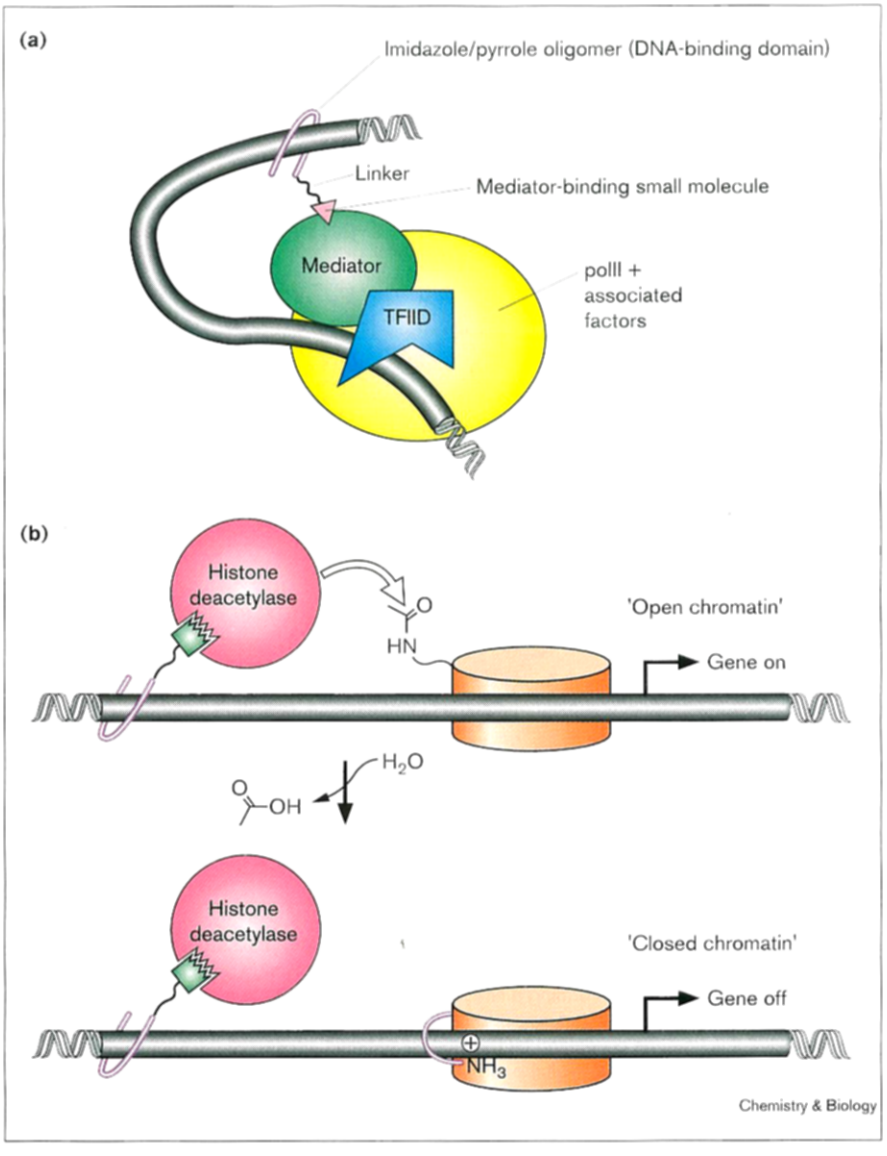
Although almost all of the above discussion has focused on using small molecules to manipulate the interactions of transcription factors with each other and with DNA, perhaps the ultimate goal in this area is to make cell-permeable small molecules that directly mimic the activity of repressors or activators. Such molecules would be extremely valuable research tools and potentially revolutionary drugs. For example, a large percentage of human cancers are associated with a defective p53 gene that encodes a transcriptional activator important in regulating cell-cycle progression [108]. If a nontoxic small molecule could be made that would activate the transcription of the p53 target genes, the impact on human health would be enormous. Although this idea might have seemed to be pure fantasy a decade

Figure 9



Formation of nanodroplets by spraying. A mixture of beads evenly dispersed in medium containing yeast is slowly injected into a stream of air forming a fine mist. When layered on to a surface such as a Petri dish this forms into nanodroplets. The average volume of the droplets is controlled by the amount of liquid applied to the surface. For a droplet volume of 50–200 nl it is possible to deposit 5000–8000 droplets in the area of a Petri dish (80 cm^2). The fraction of droplets containing beads depends on the density of beads in the medium prior to spraying. When a mixture of 80 μm Tentagel beads and medium are sprayed at a density of 14,000 beads/ml, approximately 10% of the droplets contain beads. This results in 1000 bead-containing droplets per Petri dish. Of the bead-containing droplets we find that 88% contain a single bead, 10% contain two beads, 1.3% contain three beads, and 0.7% contain four beads. Reproduced from [106].

Figure 10



Proposed scheme to make completely synthetic activators and repressors. **(a)** A synthetic activator could be constructed from a Dervan-type pyrrole/imidazole oligomer targeted to a sequence just upstream of the target gene. Fused to the artificial DNA-binding domain would be a small molecule selected to bind tightly to the mediator fragment of the polII holoenzyme. This should act as an artificial activation domain. **(b)** A synthetic repressor could be constructed by fusing a pyrrole/imidazole oligomer to a small molecule that binds to, but does not inhibit, histone deacetylase. This would result in a highly inaccessible template in the region around the small-molecule-binding site, thereby strongly repressing transcription.

ago, some of the advances in our understanding of transcriptional regulatory mechanisms suggest that the development of such a mimic is now eminently feasible. Transcriptional regulators appear mainly to be matchmakers between specific DNA sequences (promoters) and either the transcriptional machinery itself or catalytic activities that condense or decondense the chromatin structure. Making synthetic mimics of transcriptional regulatory proteins should therefore be orders of magnitude simpler than making small molecules with catalytic activities comparable to enzymes (for an intriguing study directed towards the creation of an artificial coactivator, see [109]).

As described above, the major role of many activators is probably to recruit the polII holoenzyme to the promoter and, perhaps more importantly, retain the mediator fragment there through many rounds of transcription. In theory, one could therefore make a synthetic activator by linking a sequence-specific DNA-binding molecule, for example the appropriate imidazole/pyrrole oligomer, with a molecule selected to bind to a surface-exposed mediator constituent (Figure 10). At least in yeast, most of the mediator components have been identified and the genes cloned [110], so this is quite feasible. Comparable information on the human mediator will undoubtedly be available in the near future.

Similarly, the recent discovery that many repressors function mainly to recruit a histone deacetylase complex to a given gene suggests a straightforward method to make an

artificial repressor. Again an imidazole/pyrrole oligomer could be used to localize a covalently linked histone-deacetylase-binding molecule isolated from a library. In this case, the mammalian histone deacetylase is known [64] (see discussion above), and it would not surprise us if exactly this sort of experiment is underway is several laboratories. In fact, we predict that, if the imidazole/pyrrole oligonucleotides, or possibly PNAs, prove to be generally useful *in vivo* (i.e., artificial DNA-binding domains are readily available), within 5–10 years biological chemists will have in hand an arsenal of small, cell-permeable molecules with which they can artificially control the expression of a very significant fraction of genes in the human genome. These are exciting times.

Acknowledgements

We thank Stephen Johnston (UT-Southwestern) and members of the Kodadek laboratory for many helpful conversations. Work in this laboratory on *E. coli*-based screening technology was supported by a grant from the Welch Foundation and studies of yeast transcription were supported by grants from the American Cancer Society and the NIH.

References

1. Brent, R. & Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA binding specificity of a prokaryotic repressor. *Cell* **43**, 729-736.
2. Kodadek, T. & Johnston, S.A. (1995). The dangers of 'splicing and dicing': on the use of chimeric transcriptional activators *in vitro*. *Chem. Biol.* **2**, 187-194.
3. Lohr, D., Venkov, P. & Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *FASEB J.* **9**, 777-787.
4. Johnston, S.A., Salmeron, J.M. Jr. & Dincher, S.S. (1987). Interaction of positive and negative regulatory proteins in the galactose regulation of yeast. *Cell* **50**, 143-146.
5. Leuther, K.K. & Johnston, S.A. (1992). Nondissociation of GAL4 and GAL80 *in vivo* after galactose induction. *Science* **256**, 1333-1335.
6. May, M.J. & Ghosh, S. (1997). Rel/NF-kappa B and I kappa B proteins, an overview. *Semin. Cancer Biol.* **8**, 63-73.
7. Ho, S., *et al.* & Crabtree, G.R. (1996). The mechanism of action of cyclosporin A and FK-506. *Clin. Immunol. Immunopathol.* **80**, S40-S45.
8. Schreiber, S.L. & Crabtree, G.R. (1992). The mechanism of action of cyclosporin A and FK506. *Immunol. Today* **13**, 136-42.
9. Van, D.G., Standaert, R.F., Karplus, P.A., Schreiber, S.L. & Clardy, J. (1991). Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science* **252**, 839-42.
10. Ke, H., *et al.* & Walsh, C.T. (1994). Crystal structures of cyclophilin A complexed with cyclosporin A and *N*-methyl-4-[(E)-2-butenyl]-4,4-dimethylthreonine cyclosporin A. *Structure* **2**, 33-44.
11. Liu, J., *et al.* & Schreiber, S.L. (1991). Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807-815.
12. Shaw, K.T.Y., *et al.* & Hogan, P.G. (1995). Immunosuppressive drugs prevent a rapid dephosphorylation of transcription factor NFAT1 in stimulated immune cells. *Proc. Natl Acad. Sci. USA* **92**, 11205-11209.
13. Loh, C., *et al.* & Rao, A. (1996). Calcineurin binds the transcription factor NFAT1 and reversibly regulates its activity. *J. Biol. Chem.* **271**, 10884-10891.
14. Crabtree, G.R. & Schreiber, S.L. (1996). Three-part inventions: intracellular signalling and induced proximity. *Trends Biochem. Sci.* **21**, 418-422.
15. Spencer, D.M., Wandless, T.J., Schreiber, S.L. & Crabtree, G.R. (1993). Controlling signal transduction with synthetic ligands. *Science* **262**, 1019-1024.
16. Ho, S.N., Biggar, S.R., Spencer, D.M., Schreiber, S.L. & Crabtree, G.R. (1996). Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature* **382**, 822-826.
17. Orphanides, G., Lagrange, T. & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657-2683.
18. Conaway, R.C. & Conaway, J.W. (1993). General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161-190.
19. Burley, S.K. & Roeder, R.G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Ann. Rev. Biochem.* **65**, 769-799.
20. Kim, J.L., Nikolov, B.D. & Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527.
21. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA complex. *Nature* **365**, 512-520.
22. Dynlacht, B.D., Hoey, T. & Tjian, R. (1991). Isolation of coactivators associated with the TATA binding protein that mediate transcriptional activation. *Cell* **66**, 563-576.
23. Hoopes, B.C., LeBlanc, J.F. & Hawley, D. (1992). Kinetic analysis of yeast TFIID–TATA box complex formation suggests a multi-step pathway. *J. Biol. Chem.* **267**, 11539-11547.
24. Coleman, R.A., & Pugh, B.F. (1995). A kinetic mechanism by which the TATA-binding protein achieves sequence specific DNA binding. *J. Biol. Chem.* **270**, 13850-13859.
25. Kaufmann, J. & Smale, S. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* **8**, 821-829.
26. Verrijzer, C.P., Yokomori, K., Chen, J.-L. & Tjian, R. (1994). *Drosophila* TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science* **264**, 933-941.
27. Verrijzer, C.P., Chen, J.-L. Yokomori, K. & Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**, 1115-1125.
28. Reines, D., Conaway, J.W. & Conaway, R.C. (1996). The RNA polymerase II general elongation factors. *Trends Biochem. Sci.* 351-355.
29. O'Brien, T., Hardin, S., Greenleaf, A. & Lis, J.T. (1994). Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature* **370**, 75-77.
30. Burtowski, S., Hahn, S., Guarente, L. & Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561.
31. Geiger, J.H., Hahn, S., Lee, S. & Sigler, P.B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830-836.
32. Kokubo, T., Swanson, M.J., Nishikawa, J.-I., Hinnebusch, A.G. & Nakatani, Y. (1998). The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell. Biol.* **18**, 1003-1012.
33. Thompson, C.M. Koleske, A.J. Chao, D.M. & Young, R.A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361-1375.
34. Koleske, A.J. & Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466-469.
35. Nikolov, D.B., *et al.* & Burley, S.K. (1995). Crystal structure of a TFIIB–TBP–TATA element ternary complex. *Nature* **377**, 119-128.
36. Burley, S.K. (1996). Picking up the TAB. *Nature* **381**, 112-113.
37. Stargell, L.A. & Struhl, K. (1996). Mechanisms of transcriptional activation *in vivo*: two steps forward. *Trends Genet.* **12**, 311-315.
38. Bjorkland, S. & Kim, Y.-J. (1996). Mediator of transcriptional regulation. *Trends Biochem. Sci.* **21**, 335-337.
39. Kim, Y.-L., Bjorklund, S., Li, Y., Sayre, M.H. & Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608.
40. Svejstrup, J.Q., *et al.* & Kornberg, R.D. (1997). Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl Acad. Sci. USA* **94**, 6075-6078.
41. Zawel, L., Kumar, K.P. & Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* **9**, 1479-1490.
42. Lieberman, P.M. & Berk, A.J. (1991). The Zta trans-activator protein stabilizes TFIID association with promoter DNA by direct protein-protein interaction. *Genes Dev.* **5**, 2441-2454.
43. Lee, M. & Struhl, K. (1995). Mutations on the DNA-binding surface of TATA-binding protein can specifically impair the response to acidic activators *in vivo*. *Mol. Cell. Biol.* **15**, 5461-5469.
44. Klein, C. & Struhl, K. (1994). Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains *in vivo*. *Science* **266**, 280-282.
45. Barberis, A., *et al.* & Ptashne, M. (1995). Contact with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* **81**, 359-368.
46. Farrell, S., Simkovich, N., Yu, Y., Barberis, A. & Ptashne, M. (1996). Gene activation by recruitment of the RNA polymerase II holoenzyme. *Genes Dev.* **10**, 2358-2367.
47. Chen, J.-L., Attardi, L.D., Verrijzer, C.P., Yokomori, K. & Tjian, R. (1994). Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**, 93-105.

48. Sauer, F., Hansen, S.K. & Tjian, R. (1995). Multiple TAF_{II}s directing synergistic activation of transcription. *Science* **270**, 1783-1788.
49. Verrijzer, C.P. & Tjian, R. (1996). TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.* **9**, 338-342.
50. Moqtaderi, Z., Bai, Y., Poon, D., Weil, P.A. & Struhl, K. (1996). TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature* **383**, 188-191.
51. Apone, L.M., Virbasius, C.A., Reese, J.C. & Green, M.R. (1996). Yeast TAF_{II}90 is required for cell-cycle progression through G2/M but not for general transcription activation. *Genes Dev.* **10**, 2368-2380.
52. Walker, S.S., Reese, J.C., Apone, L.M. & Green, M.R. (1996). Transcription activation in cells lacking TAF_{II}s. *Nature* **383**, 185-188.
53. Walker, S.S. Shen, W.-C. Reese, J.C. Apone, L.M. & Green, M.R. (1997). Yeast TAF_{II}145 required for transcription of G1/S cyclin genes and regulated by the cellular growth state. *Cell* **90**, 607-614.
54. Cujec, T.P., *et al.* & Peterlin, B.M. (1997). The HIV transactivator Tat binds to the CDK-activating kinase and activates the phosphorylation of the carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**, 2645-2657.
55. Garcia-Martinez, L.F., *et al.* & Gaynor, R.B. (1997). Purification of a Tat-associated kinase reveals a $TF_{II}H$ complex that modulates HIV-1 transcription. *EMBO J.* **16**, 2836-2850.
56. Gilmour, D.S. & Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster*. *Mol. Cell. Biol.* **6**, 3984-3989.
57. Mancebo, H.S.Y., *et al.* & Flores, O. (1997). P-TEF kinase is required for HIV Tat transcriptional activation *in vivo* and *in vitro*. *Genes Dev.* **11**, 2633-2644.
58. Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. *Nature* **355**, 219-224.
59. Kingston, R.E., Bunker, C.A. & Imbalzano, A.N. (1996). Repression and activation by multiprotein complexes that alter chromatin structure. *Genes Dev.* **10**, 905-920.
60. Kadonaga, J.T. (1998). Eukaryotic transcription: An interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**, 307-313.
61. Brownell, J.E. & Allis, C.D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet Dev.* **6**, 176-184.
62. Brehm, A., *et al.* & Kouzardis, T. (1998). Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature* **391**, 597-601.
63. Magnaghi-Jaulin, L., *et al.* & Harel-Bellan, A. (1998). Retinoblastoma protein represses transcription by recruiting a histone deacetylase. *Nature* **391**, 601-605.
64. Taunton, J., Hassig, C.A. & Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-411.
65. Burns, L.G. & Peterson, C.L. (1997). The yeast SWI–SNF complex facilitates binding of a transcriptional activator to nucleosomal sites *in vivo*. *Mol. Cell. Biol.* **17**, 4811-4819.
66. Quinn, J., Fyrberg, A.M., Ganster, R.W., Schmidt, M.C. & Peterson, C.L. (1996). DNA-binding properties of the SWI/SNF complex. *Nature* **379**, 844-847.
67. Cote, J., Quinn, J., Workman, J.L. & Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* **265**, 53-60.
68. Kwon, H., Imbalzano, A.N., Khaviri, P.A., Kingston, R.E. & Green, M.R. (1994). Nucleosome disruption and enhancement of activator binding by a human SWI/SNF complex. *Nature* **370**, 477-481.
69. Berger, S.L., *et al.* & Guarente, L. (1992). Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic actviation domains. *Cell* **70**, 251-265.
70. Candau, R., Zhou, J.X., Allis, C.D. & Berger, S.L. (1997). Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function *in vivo*. *EMBO J.* **16**, 555-565.
71. Mizzen, C.A., *et al.*, & Allis, C.D. (1996). The TAFII250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261-1270.
72. Wilson, C.J., *et al.* & Roung, R.A. (1996). RNA polymerase II holoenzyme contains Swi/Snf regulator involved in chromatin remodeling. *Cell* **84**, 235-244.
73. Herschlag, D. & Johnson, F.B. (1992). Synergism in transcriptional activation: a kinetic view. *Genes Dev.* **7**, 173-179.
74. Kim, T.K. & Maniatis, T. (1997). The mechanism of transcriptional synergy of an *in vitro* assembled interferon-β enhanceosome. *Mol. Cell* **1**, 119-129.
75. Giese, K., Kingsley, C., Kirshner, J.R. & Grosschedl, R. (1995). Assembly and function of a TCRα enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev.* **9**, 995-1008.
76. Molkentin, J.D., *et al.* & Olson, E.N. (1998). A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215-228.
77. Nielsen, P.E. (1997). Design of sequence-specific DNA-binding ligands. *Chem. Eur. J.* **3**, 505-508.
78. Footer, M., Egholm, M., Kron, S., Coull, J.M. & Matsudaira, P. (1996). Biochemical evidence that a D-loop is part of a four-stranded PNA–DNA bundle. Nickel-mediated cleavage of duplex DNA by a Gly–Gly–His–Bis–PNA. *Biochemistry* **35**, 10673-10679.
79. Good, L. & Nielsen, P.E. (1998). Antisense inhibition of gene expression in bacteria by PNA targeted to mRNA. *Nat. Biotechnol.* **16**, 355-358.
80. Rojas, M., Donahue, J.P., Tan, Z. & Lin, Y.-Z. (1998). Genetic engineering of proteins with cell membrane permeability. *Nat. Biotechnol.* **16**, 370-375.
81. Trauger, J.W., Baird, E.E. & Dervan, P.B. (1996). Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559-561.
82. Bruice, T.C., Mei, H.-Y., He, G.-X. & Lopez, V. (1992). Rational design of tripyrrole peptides that complex DNA by both selective minor groove binding and electrostatic interaction with the phosphate. *Proc. Natl Acad. Sci. USA* **89**, 1700-1704.
83. Chiang, S.-Y., Bruice, T.C., Azizkhan, J.C., Gawron, L. & Beerman, T.A. (1997). Targeting E2F1-DNA complexes with microgonotropen DNA binding agents. *Proc. Natl Sci. USA* **94**, 2811-2816.
84. Mrksich, M., *et al.* & Wemmer, D.E. (1992). Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. USA* **89**, 7586-7590.
85. Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1994). Design of a G.C-specific DNA minor groove-binding peptide. *Science* **266**, 646-650.
86. White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E. & Dervan, P.B. (1998). Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-471.
87. Wemmer, D.E. & Dervan, P.B. (1997). Targeting the minor groove of DNA. *Curr. Opin. Struct. Biol.* **7**, 355-361.
88. White, S., Baird, E.E. & Dervan, P.B. (1997). On the pairing rules in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **4**, 569-578.
89. Gottesfeld, J.M., Neely, L., Trauger, J.W., Baird, E.E. & Dervan, P.B. (1997). Regulation of gene expression by small molecules. *Nature* **387**, 202-205.
90. Bremer, R.E., Baird, E.E. & Dervan, P.B. (1998). Inhibition of major groove-binding proteins by pyrrole-imidazole polyamides with an Arg–Pro–Arg patch. *Chem. Biol.* **5**, 119-133.
91. Ellenberger, T.E., Brandl, C.J., Struhl, K. & Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted helices. *Cell* **71**, 1223-1237.
92. Starr, D.B. & Hawley, D.K. (1991). TFIID binds in the minor groove of the TATA box. *Cell* **67**, 1231-1240.
93. Harbury, P.A.B. & Struhl, K. (1989). Functional distinctions between yeast TATA elements. *Mol. Cell. Biol.* **9**, 5298-5304.
94. Ho, S.N., Boyer, S.H., Schreiber, S.L., Danishefsky, S.J. & Crabtree, G.R. (1994). Specific inhibition of formation of transcription complexes by a calicheamicin oligosaccharide: a paradigm for the development of transcriptional antagonists. *Proc. Natl Acad. Sci. USA* **91**, 9203-9207.
95. Kahne, D. (1995). Strategies for the design of minor groove binders: a re-evaluation based on the emergence of site-selective carbohydrate binders. *Chem. Biol.* **2**, 7-12.
96. Fields, S. & Song, O.-k. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246.
97. Vashee, S. & Kodadek, T. (1995). The activation domain of GAL4 protein mediates cooperative promoter binding with general transcription factors *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 10683-10687.
98. Vashee, S., Melcher, K., Ding, W.V., Johnston, S.A. & Kodadek, T. (1998). Evidence for two modes of cooperative DNA binding *in vivo* that do not involve direct protein-protein interactions. *Curr. Biol.* **8**, 452-458.
99. Phizicky, E.M. & Fields, S. (1995). Protein–protein interactions: Methods for detection and analysis. *Microb. Rev.* **59**, 94-123.
100. Lecitra, E.J. & Liu, J.O. (1996). A three-hybrid system for detecting small ligand-protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 12817-12821.

101. Vidal, M., Braun, P., Chen, E., Boeke, J.D. & Harlow, E. (1996). Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. *Proc. Natl Acad. Sci. USA* **93**, 10321-10326.
102. Leanna, C.A. & Hannink, M. (1996). The reverse two-hybrid system: a genetic scheme for selection against specific protein/protein interactions. *Nucleic Acids Res.* **24**, 3341-3347.
103. Huang, J. & Schreiber, S.L. (1997). A yeast genetic system for selecting small molecule inhibitors of protein-protein interactions in nanodroplets. *Proc. Natl Acad. Sci. USA* **94**, 13396-13401.
104. Hu, J., O'Shea, E.K., Kim, P.S. & Sauer, R.T. (1990). Sequence requirements for coiled-coils: Analysis with λ repressor-GCN4 leucine zipper fusions. *Science* **250**, 1400-1403.
105. Borman, S. (1997). Combinatorial chemistry. *Chem. Eng. News* **75**, 43-62.
106. Borchardt, A., Liberles, S.D., Biggar, S.R., Crabtree, G.R. & Schreiber, S.L. (1997). Small molecule-dependent genetic selection in stochastic nanodroplets as a means of detecting protein–ligand interactions on a large scale. *Chem. Biol.* **4**, 961-968.
107. You, A.J., Jackman, R.J., Whitesides, G.M. & Schreiber, S.L. (1997). A miniaturized arrayed assay format for detecting small molecule–protein interactions in cells. *Chem. Biol.* **4**, 969-975.
108. Cox, L.S. & Lane, D.P. (1995). Tumour suppressors, kinases and clamps: how p53 regulates the cell cycle in response to DNA damage. *Bioessays* **17**, 501-508.
109. Nyanguile, O. Uesugi, M. Austin, D.J. & Verdine, G.L. (1997). A nonnatural transcriptional coactivator. *Proc. Natl Acad. Sci. USA* **94**, 13402-13406.
110. Myers, L.C., *et al.* & Kornberg, R.D. (1998). The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **12**, 45-54.