From carpet bombing to cruise missiles: the 'second-order' mechanisms used by transcription factors to ensure specific DNA binding *in vivo*

Transcription factors generally have only modest specificity for their target sites, yet must find them in a sea of non-specific DNA. Some transcription factors are expressed at very high levels, to ensure that, despite losses to non-specific binding, the promoter is still occupied (the carpet-bombing strategy). Others increase their binding specificity by collaborating with other factors in a variety of ways.

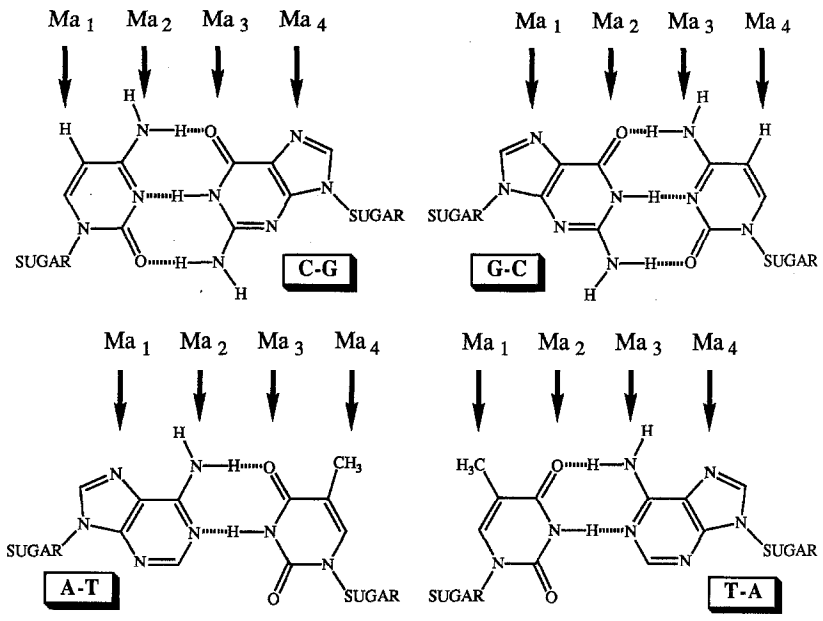
Chemistry & Biology May 1995, **2**:267–279

The genomes of eukaryotic organisms are vast. Even as simple an organism as the yeast *Saccharomyces cerevisiae* has a genome that contains ~1.4 x 10^7 base pairs (bps). The human genome consists of about 10^9 bps and that of lillies contains over 10^{12} bps. If one considers overlapping sites, there are essentially as many potential protein-binding sites on chromosomes as there are base pairs. Yet transcriptional activators and repressors, the proteins that regulate the expression of most eukaryotic genes, are somehow able to find and occupy a few particular sites in this sea of DNA. To achieve this spectacular feat of molecular recognition, the difference between the equilibrium dissociation constant for specific binding (K_d(specific)) and the average dissociation constant for non-specific binding (K_d(non-specific)) must be large. The size of the ratio between these two constants that is required to give a particular level of occupancy is determined by the size of the genome and the number of protein molecules in the cell. For example, in the hypothetical case of a human cell containing a single molecule of an activator that must bind to a single site in the 10^9-bp human genome, K_d(non-specific)/K_d (specific) would have to be 10^9 for the target site to be occupied 50 % of the time. To achieve 90 % occupancy, a ratio of 10^{10} would be required. At 30 °C, this would require a very large free energy difference ($\Delta\Delta G$) of about 13.9 kcal mol^{-1} between specific and non-specific binding.

Transcription factors must also be able to distinguish their target sites from very similar sequences in promoters that they do not control. For example, the yeast GAL4 [1] and PUT3 [2] proteins bind to closely related sites with the consensus sequences 5'-$CGGN_{11}GGC$-3' and 5'-$CGGN_{10}GGC$-3', respectively, but activate genes that are involved in completely different metabolic pathways (galactose and proline metabolism). Clearly, it would be highly disadvantageous to the organism if these activators frequently stimulated transcription from the wrong site. Similarly, there are several hormone receptors in mammalian cells, which recognize the appropriate steroid hormones, then bind to specific promoters to elicit a transcriptional response. The DNA sites to which these activators bind also differ only subtly.

A central goal in chemical biology is to understand the molecular basis of DNA-binding specificity. Most efforts so far have focused on elucidating the nature of the interface of various protein–DNA complexes, through both structural and biochemical analyses. These studies have led to a picture of DNA–protein interactions in which a few specificity-determining residues are displayed on the surface of a general structural motif (for example, the helix-turn-helix or zinc-finger motifs) in such a way that they can be presented to the major or minor groove of DNA. Each base-pair edge (Fig. 1) in double-helical DNA presents a unique array of opportunities for hydrogen bonding and van der Waals interactions [3]. Thus, when the specificity-determining residues are inserted into a helical groove, their functional groups can chemically 'sense' the sequence of the DNA. The protein may also distort the DNA to increase the number of binding interactions. Since the deformability of DNA is also highly sequence-specific, distortion provides an alternative way for the protein to sense the sequence of DNA. These two mechanisms for proteins to recognize DNA sequence have been termed direct and indirect readout, respectively. In this article, we will call them 'first-order' mechanisms for specific DNA binding, since they involve direct interactions between the DNA-binding protein and its target site.

Note, however, that most of the many molecular contacts between a typical DNA-binding protein and its target site are not sequence specific. For example, extensive contacts with groups on the DNA phosphodiester backbone are seen in all specific protein–DNA complexes, yet these groups will be present in any DNA sequence. These 'non-specific' interactions are important for providing an architectural framework that orients the specificity-determining residues and their target base



Base Pair	Ma_1	Ma_2	Ma_3	Ma_4
C-G	C-H	D	A	A
G-C	A	A	D	C-H
A-T	A	D	A	CH_3
T-A	CH_3	A	D	A

Fig. 1. The four Watson–Crick base pairs and potential positions in the major groove (Ma_1–Ma_4) for interaction with a DNA-binding protein. At the bottom of the figure, the array of binding sites is tabulated (A = hydrogen bond acceptor, D = hydrogen bond donor). It can be seen that to make unequivocal distinctions between the base pairs by binding in the major groove, a protein must make at least two hydrogen bonds with the base edge. Alternatively, van der Waals contacts could be used to detect the methyl group of thymine, a contact that can uniquely specify an A–T or a T–A base pair. This figure reflects proposals made in [3].

pairs in the appropriate fashion. It should not be surprising, therefore, that virtually all sequence-specific DNA-binding proteins exhibit measurable affinities for non-specific sites. The number and magnitude of the interactions with the edges of specific bases largely determine the specificity for the target binding site.

In some prokaryotic organisms with genomes of modest size, these first-order mechanisms may be sufficient for transcription factor function. The classic case is the lac repressor of *Escherichia coli*. Competition experiments in which specific repressor–operator complexes were mixed with non-specific DNA show that the first-order binding specificity of lac repressor for its operator is very high. From these data, Lin and Riggs [4] argued that the specificity of binding is sufficient for the small number of lac repressor molecules in a cell to saturate all of the target sites in the *E. coli* genome (~2 x 10^6 bps). In general, however, eukaryotic transcription factors have more modest first-order binding specificities than the lac repressor (see below), yet must function in the context of much larger genomes. How do these factors stably occupy their target sites?

Strategies for increasing promoter occupancy: carpet bombing and cruise missiles

When trying to understand the strategies that transcription factors use to occupy promoters, an analogy from a different type of strategic thinking may be useful. In military campaigns, it is usually desirable to bomb and destroy industrial targets in enemy territory. This can be done in two ways. The low-tech option, known as carpet bombing, is to send in a huge fleet of bombers and saturate the target area. The idea is that a few of the many thousands of 'dumb' bombs dropped will hit the desired target. Many will be wasted, however, and these may also cause unwanted damage. The other end of the spectrum is to employ cruise missiles or 'smart' bombs, which have highly sophisticated targeting systems. In this case, the chance of a single bomb reaching the target site is high, resulting in little waste and little unwanted destruction.

Given a transcription factor with modest first-order DNA-binding specificity (a not completely stupid to mildly intelligent bomb), there are two types of mechanisms that can be used to increase promoter occupancy. One, analogous to carpet bombing, is to express massive amounts of the protein. As we saw above, a low copy number human transcription factor acting in a large genome must have a very impressive ratio of specific to non-specific binding constants to be able to bind its target. But with enough copies of a protein, one could saturate the target sites even if the 'high-affinity' sequences bind only 100-fold better than bulk genomic DNA. Of course, many thousands of non-specific sites would also be occupied in the process of 'hitting' the

target and this may or may not cause problems. The other way to increase promoter occupancy is to find some way of drastically increasing the binding specificity of the protein (the cruise-missile-like solution). This could be done in two ways. One possibility is to modify the structure of the DNA-binding domain itself, for example through an allosteric transition caused by binding of a cofactor. The other is to use a 'second-order' mechanism, by which I mean a way to increase the binding specificity of the protein without affecting the structure or properties of the DNA-binding domain itself. A simple example of a second-order mechanism is cooperative binding with another transcription factor through protein–protein interactions. As we shall see below, recent work on various eukaryotic transcriptional regulatory proteins has revealed fascinating examples of each type of mechanism.

The Even-skipped (Eve) and Fushi tarazu (Ftz) homeodomain-containing proteins: the 'dumb' solution

In the fruit fly *Drosophila melanogaster*, homeodomain-containing transcription factors have a central role in establishing the body plan in the embryo [5]. Since homeodomain-containing proteins are found in all eukaryotic cells, *Drosophila* has become a key model system for experiments aimed at understanding the role of transcriptional regulation in development and differentiation in eukaryotes. A particularly interesting class of these factors all contain glutamine at position 50, and so are called the Q50 homeodomain proteins (Q50 HDPs). Structural and biochemical investigations of Q50 HDPs have suggested that they have very similar DNA-binding properties, primarily recognizing the sequence element 5'-TAA-3' [6]. If this is all that homeodomains recognize, how do these factors bind sites in promoters and ignore the many thousands of TAA sites that will be randomly scattered throughout the genome, let alone ignore non-specific sequences?

In an interesting recent paper, Biggin and coworkers [7] used photochemical crosslinking to examine the DNA-binding properties of two *Drosophila* Q50 HDPs, Ftz and Eve. The question they wished to address was whether the *in vivo* DNA-binding properties of these transcription factors mirrored those observed *in vitro*. They therefore irradiated *Drosophila* embryos with light of 254 nm to effect DNA–protein crosslinking [8], then isolated total genomic DNA and cleaved it with restriction enzymes. The protein of interest was immunoprecipitated using monoclonal antibodies, together with any DNA fragment covalently crosslinked to it, and the DNA fragments were identified using Southern hybridization. Since the amount of crosslinked product obtained was directly proportional to the amount of DNA–protein complex formed, this method provides a sensitive and quantitative measure of transcription factor–DNA binding *in vivo*.

In the early embryos used in this study, high levels of Eve protein binding to a 7.3-kbp fragment containing the *eve* promoter were detected. This was expected because Eve protein regulates its own expression. However, when the binding sites in this fragment were mapped more precisely by carrying out further restriction digests, it was found that the Eve protein was widely distributed throughout the DNA and was even bound to sequences in the coding region. This contrasts sharply with the common picture of a transcription factor bound to one or a few specific sites in a target promoter. Even more surprisingly, Eve protein could be detected binding to promoters which it does not regulate. These sequences were crosslinked to Eve at levels that were on average 10-fold lower than that observed for the *eve* promoter, but far above background. Ftz protein yielded a similar result. Indeed, these proteins appear to bind to a huge number of sites scattered throughout the *Drosophila* genome. The Eve and Ftz proteins are transiently expressed at extremely high levels (~50 000 molecules per cell) in early embryos. Some idea of how much of this very high level of expression is required to overcome poor specificity of binding can be gained from the fact that Eve and Ftz are at least 10-fold more abundant than the Zeste transcription factor, which displays much higher binding specificity *in vitro* and gives detectable crosslinking *in vivo* only to the promoters which it regulates.

These results suggest that the Eve and Ftz proteins have adopted the simplest solution to the binding-specificity problem. Since their first-order binding specificity is extremely modest, a huge amount of protein is expressed, allowing the protein to bind to its *bona fide* target sites despite the fact that it also binds to many thousands of other sites. This is the first well characterized example of the carpet bombing mechanism for eukaryotic transcription factor–DNA interactions and has many interesting implications. For example, it would appear that Eve and Ftz must function in concert with other, more highly specific, transcription factors to regulate gene expression. Otherwise, the molecules scattered throughout the genome would spuriously regulate the expression of many genes.

Enhancement of DNA-binding specificity through heterodimer formation

The yeast *S. cerevisiae* has three cell types: **a**, α and **a**/α. The first two are normally haploid, while the latter results from cell and nuclear fusion and is therefore diploid. In each of these cell types, different genes are expressed. In **a**/α cells, haploid-specific genes are repressed; in α cells, **a**-specific genes are silenced. This relatively simple regulatory network has been studied as a model for the more complex developmental circuits of higher organisms [9].

In α cells, the repression of five **a**-specific genes (asg) requires the MATα2 and MCM1 gene products, which bind to asg operators. These consist of two inverted repeat half-sites with the consensus sequence 5'-CATGT-AATT-3' separated by 2.5 helical turns. The product of the MATα2 gene, α2 repressor, is a homeodomain

protein whose domain structure is shown schematically in Figure 2a. The homeodomain is located in the carboxy-terminal region of the molecule and is connected to a dimerization domain by an unstructured linker of about 40 residues. The extreme amino-terminus of the protein carries a determinant required for transcriptional repression, but not for DNA-binding. There are also 22 residues at the carboxy-terminus of the protein that are outside the homeodomain and have no role in asg operator binding, but, as we will see later, are crucial for repression of haploid-specific genes in **a**/α cells.

Though not technically a member of the Q50 HDP family (the corresponding residue is a serine), the α2 homeodomain exhibits strong structural similarity with Eve, Ftz and other homeodomains bound to their targets. Figure 3a,b shows a comparison between the structure of α2 and the structure of the Engrailed homeodomain [10]. The base-specific contacts made by α2 are mainly to the 5'-TGT-3' sequence contained in all known asg operators (Fig. 3c). As would be expected from the small size of this core recognition site, α2 has a modest affinity for its binding site. A curious feature of α2 repressor is that *in vitro* it binds with similar affinities to asg analogs in which the spacing between the half sites has been changed dramatically, and binding is even insensitive to the orientation of the two half-sites [11]. This argues that the linker between the homeodomain and the dimerization region is extremely flexible and accommodates almost any DNA–protein geometry. Thus, α2 repressor has both low affinity and low specificity for its target sites. Yet it very efficiently represses transcription from genes containing asg operators in yeast, despite the fact that it is not produced at extremely high levels, unlike Eve and Ftz. It therefore seems that α2 does not use a carpet bombing strategy to ensure operator occupancy. It also seems unlikely that the cell could tolerate α2 repressor scattered throughout the genome. The amino-terminus of α2 serves to recruit the mobile repressors TUP1 and SSN6 [12], which are thought to inhibit the transcriptional machinery directly. Inappropriate recruitment of these proteins to genes lacking asg operators would cause widespread chaos and prevent the expression of many genes that are essential for cell survival. Thus, since α2 is clearly not acting as a 'dumb' bomb, and since its intrinsic properties do not qualify it as a 'smart' bomb, some second-order mechanism presumably provides a helping hand to target α2 to the appropriate binding sites.

In α cells the helping hand is provided by the MCM1 protein, a member of a superfamily that also includes mammalian transcription factors such as the serum response factor (SRF). MCM1 binds to DNA as a dimer; all asg operators contain high-affinity binding sites for dimeric MCM1 in their central region. When MCM1 and α2 repressor bind cooperatively to asg operators, both show high-affinity binding [13]. Furthermore, only operators with a spacing of 2.5 helical turns between inverted half-sites are recognized [11]. For example, an insertion of even three bps

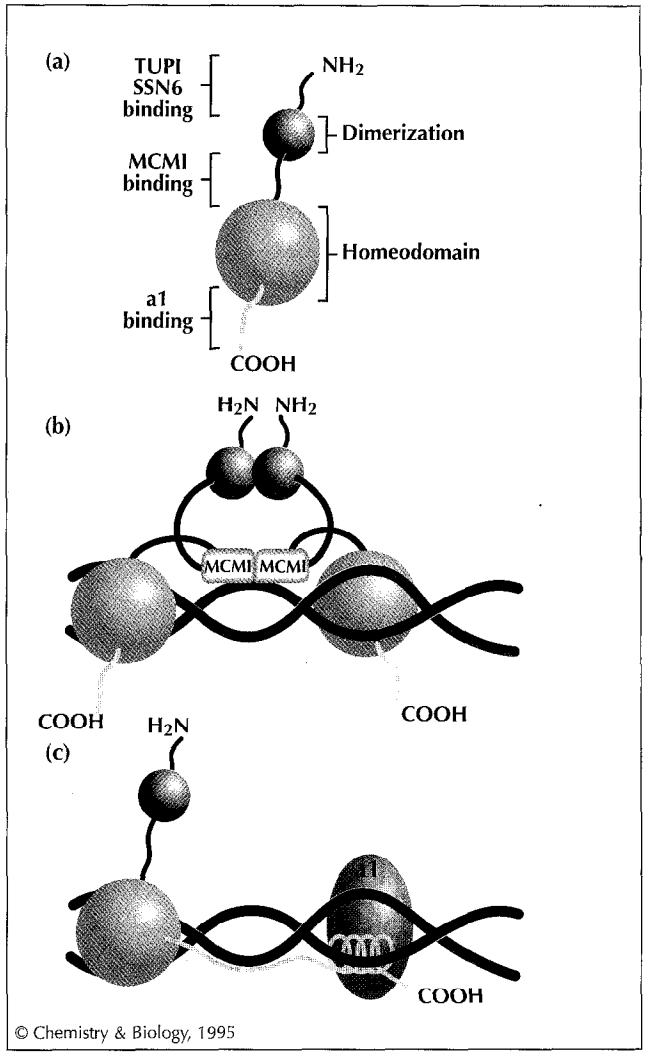


Fig. 2. Schematic representations of the α2 structure and its complexes with MCM1 and **a**1. **(a)** Functional regions of the α2 protein. TUP1 and SSN6 are repressors; MCM1 is a transcription factor specific for the asg operator (see text). **(b)** The α2–MCM–DNA complex. The structure of the linker between the α2 homeodomain and dimerization domain in the complex is unknown. **(c)** The α2–**a**1–DNA complex. The carboxy-terminal tail of the α2 protein becomes helical in this complex.

between the two half-sites, which has no effect on the affinity of either protein individually for the operator, abolishes cooperative binding. Thus, interaction with the MCM1 protein increases not only the affinity of α2 for its *bona fide* target sites, but also its specificity. Vershon and Johnson [14] have mapped the region of α2 that contacts MCM1 protein to the unstructured linker connecting the homeodomain with the dimerization domain. It is believed that, in the complex, this region acquires a well ordered structure that enforces a precise spacing and geometry between the two DNA-binding domains of α2 protein (Fig. 2b). The protein–protein interaction is completely independent of the homeodomain or any other α2 sequences, as was demonstrated by grafting the α2 linker region onto the *engrailed* homeodomain protein. This chimera bound cooperatively with MCM1 protein to sites containing

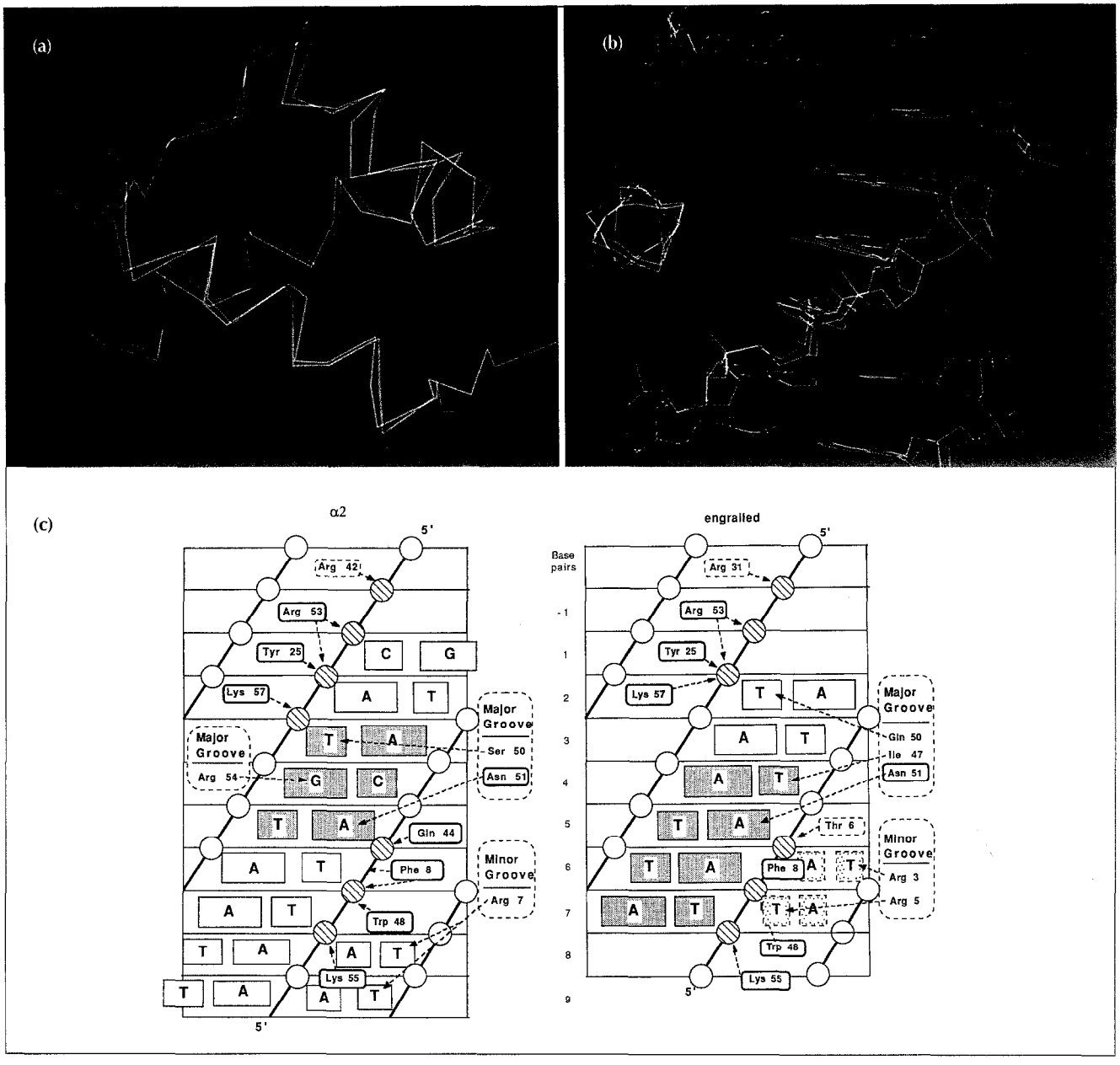


Fig. 3. X-ray crystal structures of the α2–DNA and engrailed–DNA complexes show the highly conserved superstructure of homeodomain–DNA complexes. **(a)** Superposition of the α2 (red) and engrailed (blue) α-carbon chains. **(b)** Superposition of the protein–DNA complexes (same color coding). **(c)** Schematic representation of the molecular contacts made in each complex. Reprinted with permission from [10].

engrailed and MCM1 recognition sequences. The α2/MCM1 complex is an excellent example of the cooperative binding of two different transcription factors, a very common second-order mechanism for facilitating promoter occupancy.

As MCM1 protein is relatively abundant and has a much higher affinity for the operator than α2, but does not by itself repress transcription, Johnson [9] has suggested that one should think about the MCM1–DNA complex as the true operator. In this view, binding of the α2 repressor is a highly specific operator-binding protein which employs two domains (the homeodomain and the linker) to recognize its target (the MCM1–DNA complex) through a combination of DNA–protein and protein–protein interactions.

The a1/α2 complex: allosteric control of DNA-binding activity through protein–protein interactions

In **a**/α cells, several haploid-specific genes are repressed. The promoters of these genes all contain haploid-specific gene (hsg) operators, which resemble asg sites. The major difference is that hsg operators are smaller and lack the central MCM1 binding sites. Indeed, deletion of the central 13 bps from a naturally occurring asg operator turns it into a weak hsg operator [9]. The only other important difference is that the first base pair in the consensus sequence is a G–C, instead of a C–G.

As might be expected from the sequence of the hsg operators, repression again requires the α2 repressor, but does not require MCM1. Instead, another homeodomain protein, the MAT**a**1 gene product, is involved. The binding of **a**1 and α2 to hsg operators appears to have a completely different architecture from that of MCM1 and α2 to asg operators; footprinting studies suggest that the **a**1 protein is bound to one of the half sites and the α2 repressor to the other (Fig. 2c). As is the case with the asg operator, the α2 repressor alone does not bind the hsg operator well. A dissociation constant of more than 10^{-7} M has been reported and only a 30-fold excess of non-specific DNA is required to disrupt 50 % of a repressor–operator complex. But when both purified proteins are added to an hsg-operator-containing DNA fragment *in vitro*, high affinity binding is observed ($K_d \approx 10^{-9}$ M) and more than a 30 000-fold excess of non-specific DNA is required to abolish 50 % of the protein–operator complex. There is evidence that the two proteins form a weak heterodimer in solution and that this complex then binds tightly to hsg operators [9]. The above data strongly suggest that the DNA-binding specificity of the α2/**a**1 complex is due largely to the **a**1 protein, which would appear to have an unusually high level of DNA-binding specificity for a homeodomain protein. This view is also supported by the observation that the α2/**a**1 complex can repress transcription *in vivo* even when three key specificity-determining residues in the homeodomain of α2 are mutated [15].

On the face of it, the repression of haploid-specific genes would appear to be another straightforward example of cooperative binding between two different transcription factors, in which the sophisticated targeting chemistry of the **a**1 protein converts the α2 repressor from a dumb bomb to a cruise missile. However, this simple model is difficult to reconcile with the fact that purified **a**1 protein fails to bind to hsg operator-containing DNAs *in vitro* even at concentrations of 10^{-5} M. One possibility is that, in the absence of α2, the **a**1 protein exists in a conformation that is inappropriate for operator binding. There is considerable precedent for this type of model, in which ligand binding results in large changes in the DNA-binding properties of a transcription factor. Some of the better known examples include bacterial proteins such as the trp repressor, which bind the metabolic product of the operons they regulate, forming a feedback loop, and eukaryotic steroid response factors such as the glucocorticoid receptor. In all of these cases, however, the ligand is a small molecule.

Recently, Stark and Johnson [16] reported a striking result that strongly supports the allosteric control model. It was known from previous work that cooperative binding of **a**1 and α2 required only the homeodomain of **a**1, but requires the carboxy-terminal 22 amino acids of α2 as well as its homeodomain. This raised the possibility that this 22-residue peptide is the ligand that regulates **a**1–operator interactions. To test this idea, a chimeric protein was made which contained the carboxy-terminal tail of α2 fused to the homeodomain of **a**1. This chimera was found to bind tightly to a hsg operator as a homodimer, showing clearly that the α2 peptide had restored the DNA-binding activity of the **a**1 homeodomain. In addition, when this 22-residue sequence was grafted onto the Engrailed homeodomain, cooperative binding of **a**1 and the fusion protein to an artificial operator containing the binding sites for both Engrailed and **a**1 was observed.

The α2/**a**1 complex provides an interesting twist on the general theme of cooperative binding to promoters by different transcription factors. Instead of forming a protein–protein complex with a DNA-binding specificity that is a synthesis of the binding properties of the individual proteins, as was the case for the α2-MCM1 complex, binding of α2 and **a**1 creates a novel, potent DNA-binding activity from components that bind the target site poorly (α2) or undetectably (**a**1) on their own. In the complex, the partner that binds less well alone (**a**1) provides most of the binding specificity and affinity.

A number of questions remain. For instance, why did this mechanism evolve? One possibility is that it is somehow disadvantageous for diploid yeast to allow **a**1 to bind to hsg promoters in the absence of α2, though it is not clear why this would be so. A structural rationalization of the biochemical data is also eagerly awaited. NMR studies have shown that the α2 peptide acquires a helical structure in the complex [17], but little is known about the structure of **a**1 protein, either alone or bound to the α2 peptide. Finally, it will be interesting to see how general this type of mechanism will prove to be.

Promoter recognition by TFIID

Transcription factor IID (TFIID) is a multiprotein complex [18] that is absolutely required for transcription of all mRNA-encoding genes *in vivo*. One of its component proteins, the TATA-binding protein (TBP), binds directly to sequences with the consensus 5'-TATA-AAA-3', though substitutions can be tolerated at many positions [19]. These 'TATA boxes' are found in many eukaryotic promoters. *In vitro*, TBP will support basal, but not activated, transcription. The latter also requires the TBP-associated factors (TAFs) that make up the TFIID complex. The carboxy-terminal 180 residues of TBP, which includes the DNA-binding domain, are highly conserved across all eukaryotic species.

TBP is a very unusual DNA-binding protein. The crystal structures of TBPs from *Arabidopsis* and yeast are virtually identical and reveal a molecular 'saddle' which was originally thought to sit astride a DNA double helix. However, the crystal structure of a TBP–TATA complex [20–22] revealed an unexpected and completely novel architecture in which the protein splays open the minor groove and bends the DNA substantially (Fig. 4). In addition to the bizarre structure, the biochemistry of TBP–DNA binding is atypical. Although the K_ds of complexes of TBP with a consensus TATA box are not

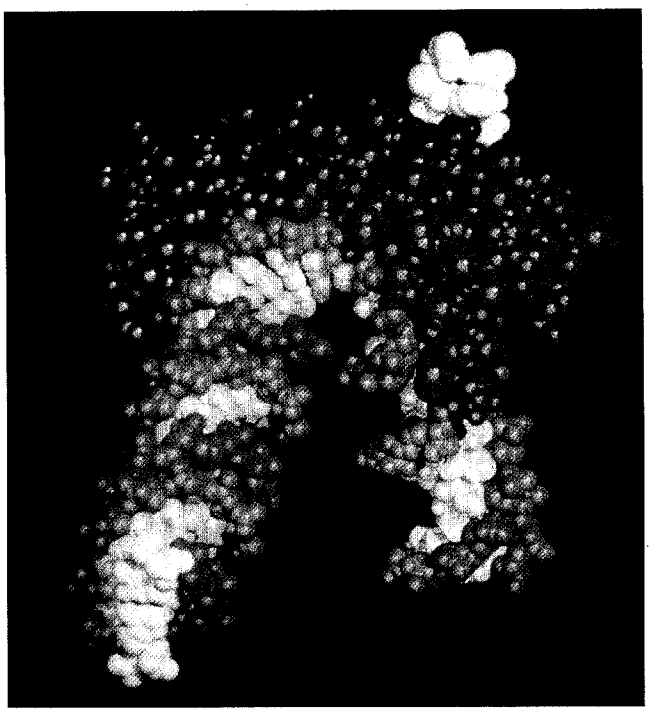


Fig. 4. X-ray crystal structure of the TBP–TATA complex showing the highly distorted DNA structure (shown in green and yellow) caused by insertion of the TBP 'saddle' (shown in blue and red) into the minor groove of the DNA. Reprinted with permission from [20].

remarkable ($K_d \approx 10^{-9}$ M), the kinetic association and dissociation rate constants that contribute to this equilibrium value are both unusual. Such complexes can have half-lives of several hours and are also slow to form. Kinetic studies suggest that the association reaction is at least a two-step process [23]. One possibility is that TBP first associates with the TATA sequence in a specific but weak fashion, and subsequently twists the DNA to form the highly stable structure shown in Figure 4. It has also been suggested that TBP may form dimers which cannot bind DNA, even at relatively low protein concentrations [24]. In this view, slow TBP dimer dissociation would limit the association rate.

The binding specificity of TBP does not appear to be very high, and even non-specific complexes have relatively long half-lives. For example, when probing TBP binding to a 300-bp TATA-containing fragment, Coleman and Pugh [25] observed protection at several sites in the DNA. Surprisingly, complexes of TBP with non-specific DNA have half-lives very similar to those of specific TBP–TATA complexes ($t_{1/2} \approx 100$ min), as determined by challenging the TBP–DNA complex with a large excess of TATA-containing oligonucleotide. The similar off rates argue that the structure of even non-specific TBP–DNA complexes are probably similar to those of TBP–TATA complexes. The association rate (k_{on}) of TBP–DNA complexes was found to depend on the DNA sequence, which must be the case if the equilibrium dissociation constant (K_d) is sequence-dependent but the kinetic dissociation rate (k_{off}) is not. Evidence was also presented which suggests that, although dissociation into bulk solution is slow, TBP easily slides along the DNA. For example, when a TATA-containing DNA fragment nearly saturated with TBP (and therefore containing TBP bound to many non-TATA sites) was challenged with a TATA-containing oligonucleotide, the TATA box continued to be protected from digestion until all of the lower affinity sites had lost protein. This is consistent with the idea that the TBP bound to the original DNA can rapidly redistribute itself along the same DNA molecule. If a TBP molecule dissociated from the TATA box, it was rapidly replaced by a protein that moved over from a lower affinity site. This leads to the notion that TBP can easily move about on a single DNA molecule when it is weakly associated with the DNA, but that when it is locked into the type of structure shown in Figure 4 it is stuck for quite some time.

Some of the biological aspects of TBP function are also curious. TBP is absolutely required for the transcription of all mRNA-encoding genes. Yet many such promoters lack TATA boxes [26]. In addition, TBP is also required for transcription catalyzed by RNA polymerases I and III. These promoters also do not contain TATA boxes. What is TBP doing at these promoters if there is no TATA binding site? Finally, there are uncertainties as to the mechanism of TBP action even on TATA-containing promoters. At the molecular level (see Fig. 4), the TBP–TATA complex looks quite symmetrical. Yet it must serve an asymmetrical function, which is to mark the promoter site and initiate the assembly of a transcription complex that will point RNA polymerase in the direction of the gene. There is no evidence to suggest that the polymerase frequently sets off in the wrong direction. How is this asymmetry engendered from a highly symmetric DNA–protein complex?

In light of the above facts, it should not be surprising that TBP–DNA interactions appear to be influenced and regulated by a number of other proteins. These will be discussed below. Some of the mechanisms involved are reminiscent of examples discussed above, but some are quite novel and many are not understood.

Regulation of TBP–TATA interactions by TAFs

First, we will consider possible roles of the TAFs, which, together with TBP, form the TFIID complex. This complex is the biologically relevant form of TBP in polymerase II transcription. It has long been known that the footprint of TFIID on promoters is much larger than that of TBP alone. Until recently, however, there was no evidence that the extra contacts that result in the larger footprint were specific in nature and it was therefore believed that TBP was the only sequence-specific DNA-binding protein in the TFIID complex. Recent work from the Smale [27] and Tijan [28] laboratories has changed this view. Verrijzer *et al.* [28] purified intact *Drosophila* TFIID using antibody affinity chromatography with an antibody raised against either TAF80 or TAF150. TFIID was then mixed with a radioactive DNA fragment containing the

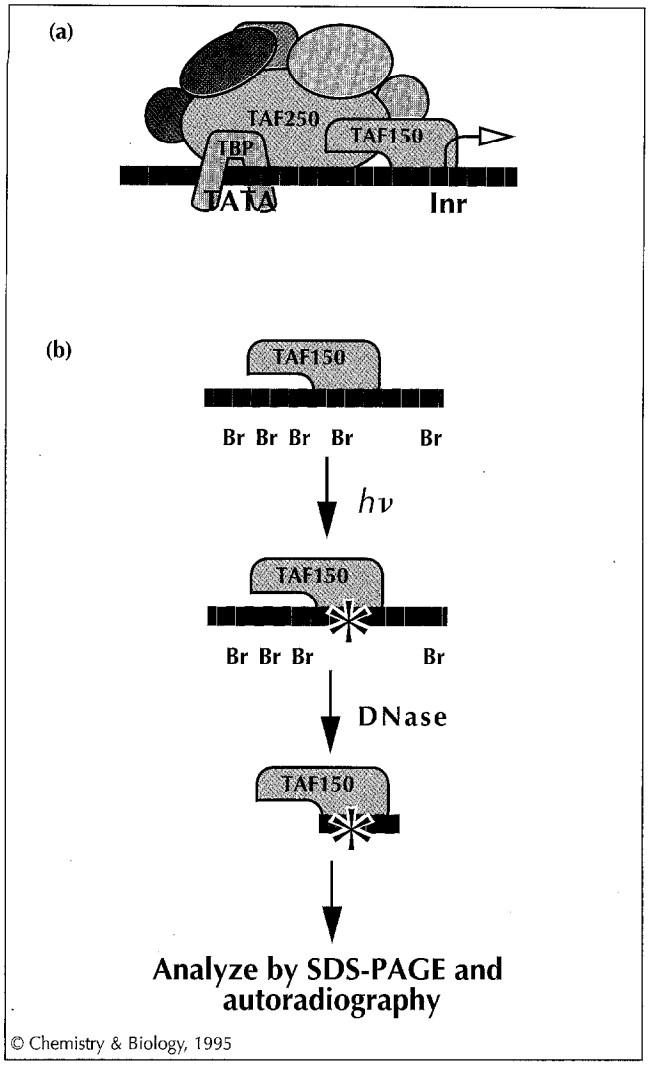


Fig. 5. Schematic representation of a model for the TFIID–promoter complex. **(a)** The proposed interactions between TAF150 and the Inr region. **(b)** Diagram of the photochemical crosslinking experiment used by Verrijzer, *et al.* [28] to detect these interactions. In the crosslinking experiment, the DNA was radioactively labeled, so that any protein that was crosslinked picked up a radioactive tag. Br, bromodeoxyuridine. Asterisk, photo-induced crosslink between DNA and TAF150.

very potent adenovirus major late promoter (AdMLP) substituted with bromodeoxyuridine, which gives high yields of covalent protein–DNA crosslinks on UV irradiation of complexes. When the TFIID–AdMLP complex was irradiated and the product treated with nucleases and analyzed by denaturing gel electrophoresis, it was found that TAF150 had acquired a covalently attached radioactive label (Fig. 5). TBP was not expected to be crosslinked in this assay because the protein does not closely approach C5 of the bromodeoxyuridine in the TBP–TATA complex. Furthermore, highly purified recombinant TAF150 could also be crosslinked to the AdMLP even in the absence of TBP or any other TAFs, unequivocally demonstrating that it is a sequence-specific DNA-binding protein. DNaseI footprinting experiments showed that TAF150 protected a region of about 30 bps beginning at the transcriptional start site and extending into the gene.

Finally, TAF150 was shown to bind *in vitro* to TBP and TAF250 (the largest TAF, which is thought to form the structural core of TFIID). TAF250 itself does not apparently have sequence-specific DNA-binding properties.

These results may explain earlier observations that the sequence of the initiator (Inr) region (the region around the transcription start site), of consensus sequence $Py_{3-4}CANTPy_{3-4}$, contributes to the strength of the promoter [29], presumably because it affects the affinity of TAF150 for the site. The most reasonable model for TFIID–promoter binding is therefore one in which the DNA-binding domains of TBP and TAF150 are displayed on the surface of the complex and allow for two-point attachment to a promoter. Indeed, there are very few functional promoters that lack both a TATA box and an Inr sequence. This two-point attachment presumably amplifies the intrinsically modest DNA-binding specificity of TBP. It may also solve, at least partially, the asymmetry question raised above. Assuming that the TFIID holocomplex is asymmetric, TAF150–Inr interactions would orient the TFIID complex appropriately on the promoter.

Given the results discussed above, the TFIID complex would appear to be just another example of two DNA-binding proteins with limited sequence specificity cooperating with one another to form a more highly specific complex. However, the situation is not so simple. The Conaways and coworkers [30] recently reported a study of the general transcription factor requirements for the formation of stable TFIID–promoter complexes. The results were quite surprising. Addition of rat TFIID to the AdMLP resulted in formation of a complex that was resistant to subsequent challenge with an excess of a DNA fragment containing a different high-affinity promoter. This was expected since AdMLP is a very strong promoter with both consensus TATA and Inr regions. However, for all other promoters tested in the same way, the TFIID–promoter complexes readily dissociated even though many contained consensus or near-consensus TATA regions (see Table 1). These results stand in sharp contrast to the results obtained using TBP alone. Consensus TATA box–TBP complexes have half-lives of more than an hour. The simplest interpretation of these data is that one or more of the TAFs negatively regulate the DNA-binding activity of TBP. Roeder and colleagues [31] have suggested that this effect may be due to TAF250. The pre-initiation complexes formed by promoters other than AdMLP only became stable when TFIIA, TFIIB and RNA polymerase II were also added to the reaction.

The picture that emerges from these studies is that the DNA-binding activity of TBP is regulated both positively and negatively through its association (direct or indirect) with TAF150, TAF250, TFIIA, TFIIB and RNA polymerase. At the level of the TFIID complex, TAF150 and TAF250 appear to have opposing effects and the outcome is highly dependent on the promoter sequence.

The AdMLP, which is employed in the vast majority of *in vitro* transcription experiments, binds TFIID avidly. We can now see that this is a very unusual promoter, however. TFIID binds to all other promoters tested with a relatively short half-life and must be part of a larger complex including TFIIA, TFIIB and RNA polymerase II in order to bind stably.

What does this mean in terms of gene regulation? There is considerable evidence that when a transcript is initiated from the AdMLP *in vitro*, TFIID remains stably bound to the promoter even after RNA polymerase and its associated factors leave the promoter (Fig. 6) [32,33]. On the other hand, Kadonaga [34] has demonstrated that TFIID dissociates from the *Drosophila* Krüppel promoter once transcription starts. This is exactly the result predicted by the Conaway study [30]. It is reasonable to assume that the continued occupancy of a promoter by TFIID would accelerate reinitiation and subsequent rounds of transcription from the AdMLP promoter. This would explain the high rate of basal transcription supported by the AdMLP. However, if TFIID must be recruited to the promoter for every round of transcription, one would imagine that RNA synthesis would be less frequent and the level of basal transcription much lower. High levels of transcription from most promoters *in vivo* require activator proteins. Given the results discussed above, stabilization of TFIID–promoter interactions would appear to be an attractive mechanism for activator function.

The role of gene-specific activators in the regulation of TBP–TATA binding

In vivo, the basal level of transcription supported by most promoters is exceedingly low. Transcription is activated by gene-specific activation proteins, which are composed of functionally separable DNA-binding and 'activation' domains [35–37]. The former recognizes DNA sites in the region of the promoter, thus targeting the activator to the appropriate genes. The activation domain contacts one or more transcription factors in a way that strongly stimulates the initiation of transcription. Although the mechanism of this process is unknown, there is circumstantial evidence that transcriptional activators can stabilize TBP–TATA interactions, at least in some cases.

TBP has been implicated as a target for many activators by the observation of direct binding between it and activation domains (ADs) from several different activators *in vitro* [38–43]. These results must be viewed with a certain amount of caution, however, as the level of binding specificity has not been adequately demonstrated in some of these experiments and most are done with TBP, not with stable TFIID complexes. Nevertheless, the laboratories of Berk [44] and Prives [45] have examined the effect of the Zta and p53 transcriptional activators, respectively, on TBP binding to a DNA fragment containing both a TATA box and an activator binding site. Under certain conditions, the activator was observed to stabilize the TBP–DNA complex. In the Berk study, a strong effect of the activator was seen only when the promoter contained a poor TATA sequence. The intrinsic half-life of the TBP–consensus TATA sequence was so long that the activator had no measurable effect. A recent *in vivo* study (S. Vashee and T.K., unpublished data) has also shown that the yeast activator GAL4, which binds TBP *in vitro*, binds cooperatively to promoters with a factor that recognizes the TATA box (presumably TBP). The fact that activators and TBP bind cooperatively to promoters is reminiscent of the α2–MCM1 story in some respects. The difference is that the binding sites for activators can be far from the TATA box, so that, for the activators to bind to TBP directly, the

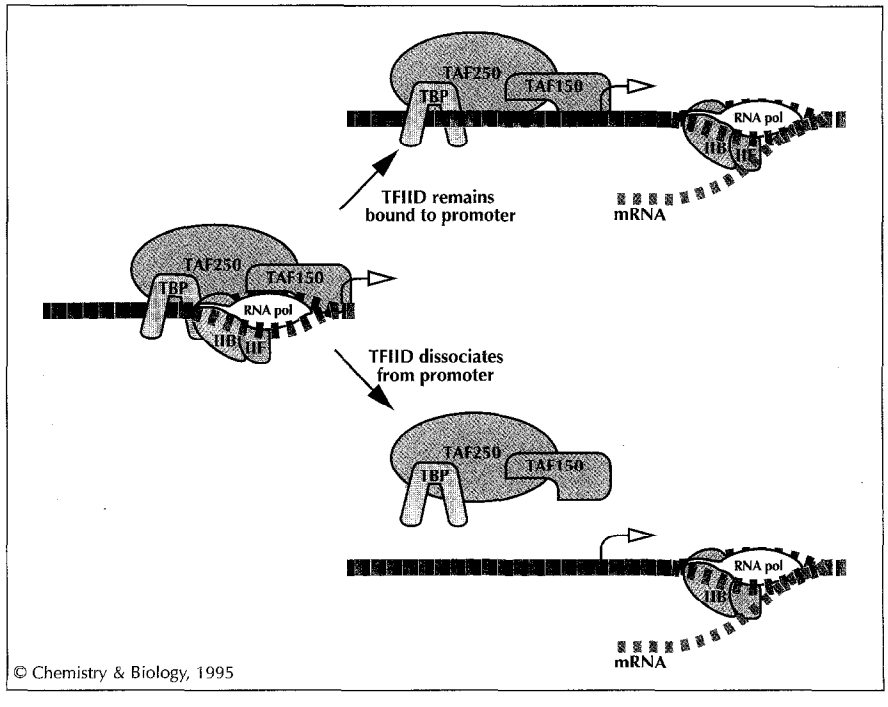


Fig. 6. Two possible models for the structure of the 'postinitiation complex', the group of proteins left behind at the promoter after RNA polymerase II leaves and begins to synthesize a transcript. These differ mainly in whether TFIID remains stably associated or not. The available evidence suggests that for the strong promoter AdMLP TFIID does remain associated with the promoter (upper diagram), increasing the rate of re-initiation, whereas for most other promoters it does not (lower diagram).

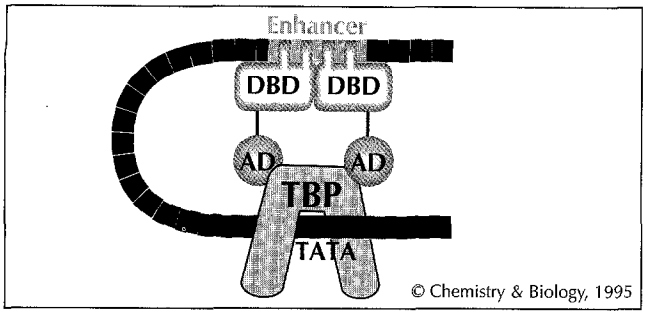


Fig. 7. Looping model for transcriptional activation using TBP as the target of the activator.

intervening DNA must form a loop (Fig. 7). The stability of the activator–TBP–DNA complex may be affected by factors that bind to the looped DNA [46]. Finally, it should be noted that the idea that activators function at least in part by recruitment of TFIID is not universally accepted [47–49]. Indeed, it seems clear that even if this is one of the functions of activators, they must also stimulate other steps in the transcription cycle, at least for some genes [8,50–52]. It would not be surprising if multiple mechanisms for transcriptional activation exist; each promoter, depending on its sequence elements, might then use a different combination of these possible mechanisms.

Global regulators of transcription

Any discussion of protein–DNA interactions in eukaryotic cells would be incomplete if it ignored the effects of chromatin structure. DNA is extensively coated with histones and other proteins *in vivo*, which must certainly affect the affinity of a protein for its target site. In some cases, chromatin structure clearly blocks access of proteins to DNA; this may sometimes, paradoxically, facilitate transcription factor–promoter interactions. For example, it has been argued that if most of the 'junk' DNA present in eukaryotic cells is prevented from binding to a low copy number DNA-binding protein, the level of binding specificity that is required for that protein to efficiently occupy its specific site (which is for some reason more accessible) would be reduced [53]. On the other hand, if a histone octamer or some other higher order chromatin structural element sits on top of the protein's target site, it is easy to imagine that binding would be strongly inhibited. A number of *in vitro* studies, reviewed elsewhere [54], are in accord with this general picture and it is now believed that a major part of transcriptional activation *in vivo* involves overcoming the repressive effects of chromatin structure [55,56]. It is certainly clear that inclusion of a TATA box into a nucleosome core particle can severely inhibit binding of TBP *in vitro* [57]. Therefore, a specific model for transcriptional activation that many investigators in the field find attractive is that binding of TFIID to promoters may be strongly inhibited by nucleosome structure *in vivo* and that this process must be stimulated by activators. One could therefore argue that, while the *in vitro* binding studies using activators, TBP and naked DNA are interesting, they probably grossly underestimate the effect of activators on TFIID–TATA interactions *in vivo*.

This already complex picture is complicated even further by recent work on the SWI/SNF complex. This large complex (MW ≈ 2000 kD) contains at least five proteins, the products of the SWI, SWI2, SWI3, SNF5 and SNF6 genes [58,59]. These genes were originally identified by genetic analysis as global activators of transcription [60–62]; mutations in these genes reduced the expression of a very large number of genes. In yeast cells, the SWI/SNF complex is required for full activation of genes regulated by many activators, including GAL4 protein, *Drosophila* Ftz, mammalian steroid receptors and several others. This suggests that, like so many other aspects of transcriptional regulation, the function of the SWI/SNF complex (whatever it is) is conserved in all eukaryotic cells. It has been suggested that the SWI/SNF complex might be recruited to target genes by gene-specific activators [63]. This idea is supported by the finding that some chimeras containing sequence-specific DNA-binding domains fused to one of the proteins of the SWI/SNF complex, but lacking a *bona fide* activation domain, can activate transcription in yeast. In addition, the mammalian glucocorticoid receptor (a hormone-responsive transcriptional activator) has been shown to associate with SWI3 *in vitro*.

Given that it affects so many genes, it was assumed that the SWI/SNF complex must alter a common aspect of the transcription of most or all genes, perhaps by affecting chromatin structure or by directly stimulating the general transcription complex. There is now good evidence that the former, at least, can occur. For example, Kingston and coworkers [57] recently reported that a partially purified human SWI/SNF complex could facilitate binding of TBP to a core nucleosome particle containing a consensus TATA box. This binding activity also required TFIIA, which has been observed to stimulate TBP–TATA binding in other assays as well. SWI/SNF stimulation was completely ATP dependent. Interestingly, the SWI/SNF complex inhibited binding of TBP to naked DNA. In addition, the SWI/SNF complex has also been shown to stimulate the binding of other transcription factors to nucleosomes [64]. These results argue strongly that the SWI/SNF complex exerts its effect through the chromatin structure rather than by directly targeting TBP or any other particular protein. While the details of this process are not at all clear, the current thinking is that SWI/SNF actively reconfigures the nucleosome structure to promote protein binding (Fig. 8). It will be fascinating to learn exactly how this is accomplished.

The MOT1 protein of yeast also has global effects on transcription. Like the SWI/SNF complex, the ~175 000 kD MOT1 protein is an ATPase. It also has some homology to SWI2, but this is where the similarity ends. Whereas the SWI/SNF complex is a global activator of transcription, MOT1 is a global repressor [65]. Interestingly, only genes transcribed by RNA polymerase II, not those transcribed by RNA polymerases I or III, appear to be subject to MOT1 repression, suggesting that the protein must target a step that is unique to the synthesis of mRNA. *MOT1* is an

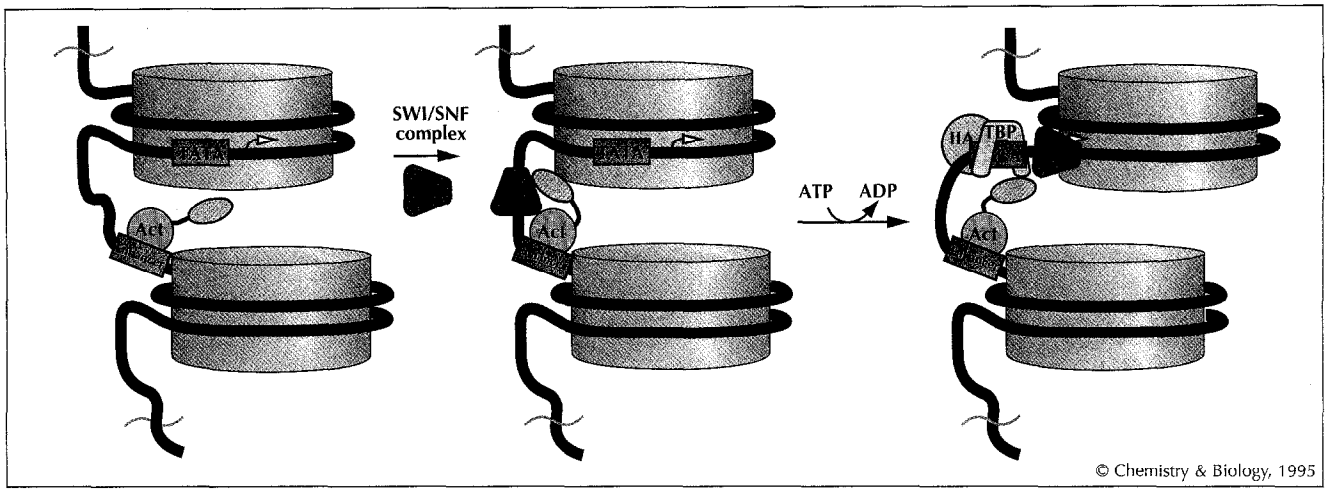


Fig. 8. A model for activator and SWI/SNF-facilitated binding of TBP to a nucleosomal TATA box. The activator recruits the SWI/SNF complex to the promoter. This complex then reconfigures the nucleosome structure in such a way that TBP, along with TFIIA, can gain access to the TATA box. In this figure, I suggest that the SWI/SNF complex moves the histone octamer off of the TATA box. This is an arbitrary aspect of the model. The nature of the structural reconfiguration is not known.

essential gene in yeast and the ATPase activity of the protein is also essential. Certain nonlethal mutations in *MOT1* lead to increased levels of basal transcription of many genes.

It has been suggested that the MOT1 protein might function to remove transcription factors from DNA. A remarkable study by Auble *et al.* [66] confirmed this notion, by showing that the MOT1 protein disrupts TBP–TATA complexes in an ATP-dependent fashion. To the best of my knowledge, this is a completely novel mechanism for controlling transcription factor–DNA interactions. The closest analogy comes from observations that DNA helicases remove proteins from the path of DNA replication forks and homologous strand exchange intermediates in an ATP-dependent fashion [67,68]. These events are not specific for the DNA-binding protein to be removed, however. Current evidence suggests that helicases will pry almost any protein off DNA. In contrast, MOT1 appears to be completely specific for TBP. Direct protein–protein interactions between MOT1 and TBP have been demonstrated *in vitro* and *in vivo* [66,69]. Indeed, MOT1 protein co-immunoprecipitates with TBP from a yeast lysate [70].

The exact mechanism of action of MOT1 is not clear. An attractive model is that MOT1 is associated with TBP in the TFIID complex. While it might seem counterproductive for the TFIID complex to include a protein that can cause dissociation of TBP–DNA complexes, there are reasons that this might be desirable. For example, if TFIID, like TBP, is slow to dissociate from non-specific sites, MOT1 protein might increase the rate of dissociation from non-specific sites, improving the specificity of binding to *bona fide* promoters. Alternatively, it might be important to dissociate TBP from the promoter after every few rounds of transcription so that genes are not overactivated. However, Poon *et al.* [71] have shown that the complex of MOT1 with TBP appears to be distinct from the yeast TFIID complex. This result was unexpected since MOT1 protein seems to be specific for RNA polymerase II transcription. Much more work will be required to deduce the role of MOT1 protein in transcriptional repression. For example, what is the relative effect of MOT1 on specific and non-specific TBP–TATA complexes? What effects do other TAFs have on these events? Is it possible for MOT1 to exchange between different TBP-containing complexes?

Summary and conclusions

Chemists and biochemists interested in sequence-specific DNA–protein interactions have generally focused on understanding the detailed interactions between the DNA-binding domain of the protein and the bases in the target site. This is clearly an important and fascinating avenue of investigation. Understanding the true *in vivo* DNA-binding specificity of a transcription factor, however, involves many other layers of complexity. In addition to relatively straightforward cooperative interactions between different DNA-binding proteins, much more elaborate mechanisms for modulating binding specificity have been unearthed. A spectacular example is the case of TBP–TATA interactions, where there is evidence that TFIIA, TAF150, TAF250, TFIIB, RNA polymerase II and many gene-specific activators all affect the stability of this important complex. Even more remarkable are the active, ATP-dependent mechanisms employed by the SWI/SNF and MOT1 proteins to either facilitate or disrupt TBP–DNA interactions. Although this level of complexity may seem daunting, it also constitutes an exciting challenge for scientists working at the chemistry/biology interface. We will not be bored. Indeed, the only major surprise in this area would be if there are not many more major surprises waiting just around the corner.

Acknowledgements: I would like to thank Professor Stephen Johnston for many helpful discussions and a critical reading of this manuscript and Professors Stephen Burley and Alexander Johnson for providing figures.

References

1. Johnston, M. (1987). A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **51**, 458–476.
2. Axelrod, J.D., Majors, J. & Brandriss, M.C. (1991). Proline-independent binding of PUT3 transcriptional activator detected by footprinting *in vivo*. *Mol. Cell. Biol.* **11**, 564–567.
3. Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804–808.
4. Lin, S. & Riggs, A.D. (1975). The general affinity of lac repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes. *Cell* **4**, 107–111.
5. Gehring, W.J. (1987). Homeo boxes in the study of development. *Science* **236**, 1245–1252.
6. Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. & Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. *Cell* **63**, 579–590.
7. Walter, J., Dever, C.A. & Biggin, M.D. (1994). Two homeo domain proteins with similar specificity to a wide range of DNA sites in *Drosophila* embryos. *Genes Dev.* **8**, 1678–1692.
8. 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.
9. Johnson, A.D. (1992). A combinatorial regulatory circuit in budding yeast. In *Transcriptional Regulation*. (Yamamoto, K. & McKnight, S.L., eds), pp. 975–1006, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. Wohlberger, C., Vershon, A.K., Liu, B., Johnson, A.D. & Pabo, C.O. (1991). Crystal structure of a MATα2 homeodomain–operator complex suggests a general model for homeodomain–DNA interactions. *Cell* **67**, 517–528.
11. Smith, D.L. & Johnson, A.D. (1992). A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an α2 dimer. *Cell* **68**, 133–142.
12. Komachi, K., Redd, M.J. & Johnson, A.D. (1994). The WD repeats of Tup1 interact with the homeo domain protein α2. *Genes Dev.* **8**, 2857–2867.
13. Keleher, C.A., Passmore, S. & Johnson, A.D. (1989). Yeast repressor α2 binds to its operator cooperatively with yeast protein MCM1. *Mol. Cell. Biol.* **9**, 5228–5230.
14. Vershon, A.K. & Johnson, A.D. (1993). A short disordered protein region mediates interactions between the homeodomain of the yeast α2 protein and the MCM1 protein. *Cell* **72**, 105–112.
15. Vershon, A.K., Jin, Y. & Johnson, A.D. (1995). A homeo domain protein lacking specific side chains of helix 3 can still bind DNA and direct transcriptional repression. *Genes Dev.* **9**, 182–192.
16. Stark, M.R. & Johnson, A.D. (1994). Interaction between two homeo domain proteins is specified by a short C-terminal tail. *Nature* **371**, 429–432.
17. Phillips, C.L., Stark, M.R., Johnson, A.D. & Dahlquist, F.W. (1994). Heterodimerization of yeast homeo domain transcriptional regulators α2 and **a**1 induces an interfacial helix in α2. *Biochemistry* **33**, 9294–9302.
18. Dynlacht, D.B., Hoey, T. & Tjian, R. (1991). Isolation of coactivators associated with the TATA-binding protein that mediates transcriptional activation. *Cell* **66**, 563–576.
19. Greenblatt, J. (1991). Roles of TFIID in transcriptional initiation by RNA polymerase II. *Cell* **66**, 1067–1070.
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. Kim, J.L. & Burley, S.K. (1994). 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat. Struct. Biol.* **1**, 638–653.
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., Taggart, A.K.P., Benjamin, L.R. & Pugh, B.F. (1995). Dimerization of the TATA-binding protein. *J. Biol. Chem.*, in press.
25. Coleman, R.A. & Pugh, B.F. (1995). A kinetic mechanism by which the TATA-binding protein achieves sequence specific DNA binding. *J. Biol. Chem.*, in press.
26. Singer, V.L., Wobbe, C.R. & Struhl, K. (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev.* **4**, 636–645.
27. Kaufmann, J. & Smale, S. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* **8**, 821–829.
28. Verrijzer, C.P., Yokomori, K., Chen, J.-L. & Tjian, R. (1994). Drosophila $TAF_{II}150$: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science* **264**, 933-941.
29. Smale, S.T. & Baltimore, D. (1989). The initiator as a transcriptional control element. *Cell* **57**, 103–113.
30. Aso, T., Conaway, J.W. & Conaway, R.C. (1994). Role of promoter core structure in assembly of the RNA polymerase II preinitiation complex. *J. Biol. Chem.* **269**, 26575–26583.
31. Kokubo, T., Yamashita, S., Horikoshi, M., Roeder, R.G. & Nakatani, Y. (1994). Interaction between the N-terminal domain of the 230-kDa subunit and the TATA box-binding subunit of TFIID negatively regulates TATA-box binding. *Proc. Natl. Acad. Sci. USA* **91**, 3520–3524.
32. Hawley, D.K. & Roeder, R.G. (1987). Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. *J. Biol. Chem.* **262**, 3452–3461.
33. Van Dyke, M.W., Roeder, R.G. & Sawadogo, M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* **241**, 1335–1338.
34. Kadonaga, J. (1990). Assembly and disassembly of the *Drosophila* RNA polymerase II complex during transcription. *J. Biol. Chem.* **265**, 2624–2631.
35. 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.
36. Corton, J.C. & Johnston, S.A. (1989). Altering DNA-binding specificity of GAL4 requires sequences adjacent to the zinc finger. *Nature* **340**, 724–727.
37. Brent, R. & Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA binding specificity of a prokaryotic repressor. *Cell* **43**, 729–736.
38. Horikoshi, N., Maguire, K., Kralli, A., Maldonado, E., Reinberg, D. & Weinmann, R. (1991). Direct interaction between adenovirus E1A protein and the TATA box binding transcription factor IID. *Proc. Natl. Acad. Sci. USA* **88**, 5124–5128.
39. Kerr, L.D., *et al.*, & Verma, I.M. (1993). Association between proto-oncoprotein Rel and TATA-binding protein mediates transcriptional activation by NF-κB. *Nature* **365**, 412–419.
40. Lee, W.S., Kao, C.C., Bryant, G.O., Liu, X. & Berk, A.J. (1991). Adenovirus E1A activation domain binds the basic repeat in the TATA box transcription factor. *Cell* **67**, 365–376.
41. Stringer, K.F., Ingles, C.J. & Greenblatt, J. (1990). Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFID. *Nature* **345**, 783–786.
42. Melcher, K. & Johnston, S. (1994). GAL4 interacts with TBP and coactivators. *Mol. Cell. Biol.* **15**, 2839–2848.
43. Truant, R., Xiao, H., Ingles, C.J. & Greenblatt, J. (1993). Direct interaction between the transcriptional activation domain of human p53 and the TATA box-binding protein. *J. Biol. Chem.* **268**, 2284–2287.
44. 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.
45. Chen, X., Farmer, G., Zhu, H., Prywes, R. & Prives, C. (1993). Cooperative DNA binding of p53 with TFIID (TBP): a possible mechanism for transcriptional activation. *Genes Dev.* **7**, 1837–1849.
46. Tjian, R. & Maniatis, T. (1994). Transcription activation: a complex puzzle with few easy pieces. *Cell* **77**, 5–8.
47. Xiao, H., *et al.*, & Greenblatt, J. (1994). Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Mol. Cell. Biol.* **14**, 7013–7024.
48. Lin, Y.-S. & Green, M.R. (1991). Mechanism of activation of an acidic transcriptional activator *in vitro*. *Cell* **64**, 971–981.
49. Roberts, S.G.E. & Green, M.R. (1994). Activator-induced conformational change in general transcription factor TFIIB. *Nature* **371**, 717–720.
50. 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.
51. Rougvie, A.E. & Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804.
52. Giardina, C., Perez-Riba, M. & Lis, J.T. (1992). Promoter melting and TFIID complexes on *Drosophila* genes *in vivo*. *Genes Dev.* **6**, 2190–2200.
53. Pearlmann, T. (1989). Glucocorticoid receptor DNA-binding specificity is increased by the organization of DNA in nucleosomes. *Proc. Natl. Acad. Sci. USA* **89**, 3884–3888.
54. Workman, J.L. & Buchman, A.R. (1993). Multiple functions of nucleosomes and regulatory factors in transcription. *Trends*

Biochem. Sci. **18**, 90–95.
55. Workman, J.L., Taylor, I.C.A. & Kingston, R.E. (1991). Activation domains of stably bound GAL4 derivatives alleviate repression of promoters by nucleosomes. *Cell* **64**, 533–544.
56. Croston, G.E., Laybourn, P.L., Paranjape, S.M. & Kadonaga, J.T. (1992). Mechanism of transcriptional antirepression by GAL4-VP16. *Genes Dev.* **6**, 2270–2281.
57. Imbalzano, A.N., Kwon, H., Green, M.R. & Kingston, R.E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* **370**, 481–485.
58 Peterson, C.L., Dingwall, Å. & Scott, M.P. (1994). Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. *Proc. Natl. Acad. Sci. USA* **91**, 2905–2908.
59. Cairns, B.R., Kim, Y.-J., Sayre, M.H., Laurent, B.C. & Kornberg, R.D. (1994). A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proc. Natl. Acad. Sci. USA* **91**, 1950–1954.
60. Peterson, C.L. & Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2 and SWI3 genes, which encode a global activator of transcription. *Cell* **68**, 573–583.
61. Laurent, B.C., Treitel, M.A. & Carlson, M. (1990). The SNF5 protein of *Saccharomyces cerevisiae* is a glutamine-rich and proline-rich transcriptional activator that affects the expression of a broad spectrum of genes. *Mol. Cell. Biol.* **10**, 5616–5625.
62. Laurent, B., Treitel, C. & Carlson, M. (1991). Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proc. Natl. Acad. Sci. USA* **88**, 2687–2691.
63. Yoshinaga, S.K., Peterson, C.L., Herskowitz, I. & Yamamoto, K.R. (1992). Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors. *Science* **258**, 1598–1604.
64. 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.
65. Davis, J.L., Kunisawa, R. & Thorner, J. (1992). A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in yeast. *Mol. Cell. Biol.* **12**, 1879–1892.
66. Auble, D.T., Hansen, K.E., Mueller, C.G.F., Lane, W.S., Thorner, J. & Hahn, S. (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.* **8**, 1920–1934.
67. Bedinger, P., Hochstrasser, M., Jongeneel, C.V. & Alberts, B.M. (1983). Properties of the T4 bacteriophage DNA replication apparatus: the T4 dda helicase is required to pass a bound RNA polymerase molecule. *Cell* **34**, 115–123.
68. Salinas, F. & Kodadek, T. (1994). Strand exchange through a DNA–protein complex requires a DNA helicase. *Biochem. Biophys. Res. Comm.* **205**, 1004–1009.
69. Auble, D.T. & Hahn, S. (1993). An ATP-dependent inhibitor of TBP binding to DNA. *Genes Dev.* **7**, 844–856.
70. Poon, D. & Weil, P.A. (1993). Immunoprecipitation of yeast TATA-binding protein and associated factors. *J. Biol. Chem.* **268**, 15325–15328.
71. Poon, D., Campbell, A.M., Bai, Y. & Weil, P.A. (1994). Yeast Taf170 is encoded by MOT1 and exists in a TATA box-binding protein (TBP)-TBP-associated factor complex distinct from transcription factor IID. *J. Biol. Chem.* **269**, 23135–23140.

Thomas Kodadek, Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA.