This is to certify that the
dissertation entitled

THE ROLE OF THE VP16AD:TBP INTERACTION IN TRANSCRIPTIONAL ACTIVATION

presented by

Dean Daniel Shooltz

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Biochemistry and Molecular Biology

Major Professor's Signature

December 15, 2006

Date

THE ROLE OF THE VP16AD:TBP INTERACTION IN TRANSCRIPTIONAL ACTIVATION

By

Dean Daniel Shooltz

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry and
Molecular Biology

2007

ABSTRACT

THE ROLE OF THE VP16AD:TBP INTERACTION
IN TRANSCRIPTIONAL ACTIVATION

By

Dean Daniel Shooltz

The regulation of transcription is central to development, homeostasis, and disease. Key factors in the regulation of gene expression include transcriptional activator proteins. These proteins provide the functional link between cis-acting DNA regulatory elements and transcriptional responses. While the principles of cis-element recognition are well established, the mechanisms by which transcriptional activation domains function remain elusive. Intense efforts have been directed at delineating the targets of transcriptional activation domains, and many potential targets have been identified. However, the in vivo functional relevance of many of these connections remains poorly characterized. This dissertation describes efforts to characterize the interaction of the VP16 activation domain (VP16AD) with the TATA-binding protein (TBP).

The strength of the VP16AD:TBP interaction correlates with the ability of VP16AD to activate transcription, suggesting that TBP is a relevant target of VP16AD. However, VP16AD, TBP, and DNA appear unable to form a three-way complex, raising the question of how VP16AD can activate transcription through interactions with TBP. This dissertation explored two proposed models for VP16AD-TBP interactions. First, I examined whether VP16AD could increase the orientational specificity of TBP bound to its cognate promoter element. The results of this study indicated that TBP is intrinsically

highly oriented on the TATA DNA sequence, and that no TBP-orienting activity could be detected in VP16AD. The finding that TBP is intrinsically oriented implies that current models of the assembly of the transcriptional machinery need to be updated.

A second proposed mechanism of VP16AD-TBP interaction posits that a cascade of interactions with TBP is involved in transcriptional activation. In this model, VP16AD competes with a negative regulator, binds TBP transiently, and then releases TBP for DNA association. This dissertation describes my efforts to structurally characterize VP16AD when bound to TBP. Although a structural model was not determined, the efforts have resulted in the design and characterization of a protein complex that may serve as a foundation for further study. The experimental efforts have additionally resulted in an improved method of TBP purification.

ACKNOWLEDGMENTS

I would like to thank my mentor, Steve Triezenberg, for sharing with me his seemingly unlimited enthusiasm for science in general and for molecular biology in particular. I appreciate his dedication to training me to think and function as a scientist. I will fondly remember my time as a graduate student in his lab.

I would also like to thank the past and present members of the Triezenberg laboratory. In my early days as a graduate student I relied heavily on the advice of Søren Ottosen, Yaopan Mao, Billy Yang, Kostas Vlachonasios and Francisco Herrera. I benefited from their in-depth answers and guidance, and I fondly remember the times we shared in the lab. I have further enjoyed the camaraderie of the lab as it has continued with Amy Hark, Kanchan Pavangadkar, Sebla Kutluay, Nate Lord and James Doroghazi.

I would also like to thank my family and friends. They have been a constant source of support and encouragement, and they have helped me through the various tough times in my graduate career. I value their understanding of the time that I needed to spend in the lab over the past few years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AdMLP	*Adenovirus* major late promoter
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
BRE	TFIIB recognition element
BSA	Bovine serum albumin
DPE	Downstream promoter element
DTT	Dithiothreitol
FPLC	Fast protein liquid chromatography
GSH	Glutathione
GST	Glutathione-S-transferase
GTF	General transcription factor
GVT	GST-VP16C-TAND2 fusion protein
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum correlation
IAAOP	5-Iodoacetamido-1,10-phenanthroline
Inr	Initiator promoter sequence
IPTG	Isopropyl-beta-D-thiogalactopyranoside
LB	Luria-Bertani medium
NOESY	Nuclear Overhauser enhancement spectroscopy
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PIC	Pre-initiation complex

PMSF	Phenylmethylsulphonylfluoride
RNA pol II	RNA polymerase II
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
TAF	TBP-associated factor
TAND	TAF-1 N-terminal domain
TBP	TATA-element binding protein
TBP-OP	TBP treated with IAAOP
TFII	Transcription factor for RNA pol II
VP16	Virion protein 16 from herpes simplex virus
VP16AD	VP16 activation domain
VP16C	VP16 C-terminal activation subdomain
VP16N	VP16 N-terminal activation subdomain

Chapter 1

Literature Review[1]

Transcriptional regulation in eukaryotes.

The genome of any organism contains not only the information specifying the primary sequence of the many proteins required for growth, development and defense, but also regulatory information for determining when and where a given gene is to be expressed. The link between cis-acting regulatory DNA sequences and the expression of the corresponding genes is made by trans-acting transcriptional activator and repressor proteins. These regulatory proteins must fulfill two functions. They must accurately recognize which genes they are to regulate, and once associated with their target gene they must either stimulate (for activator proteins) or inhibit (for repressor proteins) transcription by RNA polymerase II (RNA pol II). The mechanisms for accomplishing these two functions are built on principles of modular design and combinatorial logic. For any given gene, the cis-regulatory elements comprise a set of short DNA sequences that together dictate the appropriate expression of that gene. The trans-regulatory proteins that bind such elements bring particular domains that interact with either chromatin-modifying enzymes or components of the general transcription machinery. These protein:protein interactions are highly diverse, both in the array of activators and target proteins present in a particular cell and also in that any given activator might bind to several target proteins.

[1] Part of this chapter was published in "Herrera FJ, Shooltz DD and Triezenberg SJ (2004). Mechanisms of transcriptional activation in eukaryotes. Handbook of Experimental Pharmacology. Vol 166:3-31"

Information flow: activating the activators

The specificity of gene regulation in a given cell at different times or in response to various external or internal signals demands that the panoply of transcription factors in that cell should not all be active all of the time. This obvious observation implies that the activity of activators must be regulated, and indeed this principle has been demonstrated in a surprisingly diverse array of mechanisms. Some activators are regulated by control of their cellular localization from cytoplasm to nucleus. For example, NF-κB, an activator of cytokine-induced gene expression, is retained in the cytoplasm by interaction with IκB. Kinases triggered to action by ligand-bound cytokine receptors phosphorylate IκB, which then releases NF-κB to traverse to the nucleus (reviewed by (40)). The sterol regulatory element binding protein (SREBP) proprotein is anchored in the endoplasmic reticulum by two trans-membrane domains. Proteolytic cleavage releases SREBP from the endoplasmic reticulum and allows its translocation to the nucleus where it activates genes involved in sterol biosynthesis (reviewed by (138)). Some nuclear hormone receptors are localized in the cytoplasm until binding to the relevant ligand allows translocation to the nucleus to activate transcription of target genes (reviewed by (110)).

The activities of transcriptional activators can also be regulated by covalent modifications including phosphorylation, methylation, and acetylation, and ubiquitinylation. A particularly impressive example is the tumor suppressor protein p53, which is subjected

to multiple covalent modifications that tightly regulate its function and stability (reviewed by (12)). Ubiquitinylation of p53 by Mdm2 (an E3 ubiquitin ligase) and subsequent destruction in the proteasome is responsible for maintaining p53 protein at low levels. Phosphorylation of p53 in response to ionizing radiation stabilizes the p53 protein. Acetylation of p53 also stimulates its transcriptional activity.

One modification that has garnered much interest lately is the addition of ubiquitin to activators by ubiquitin ligases. Although long known as a "tag" for protein degradation by the proteasome, ubiquitin (and a related polypeptide termed SUMO) is increasingly recognized for non-proteolytic roles as well. Most intriguing has been the suggestion that ubiquitin is required de facto for the function of activators in yeast and mammalian cells (reviewed by (115)). One clue is that the protein motif that targets activators for degradation overlaps with the transcriptional activation domain in proteins including various nuclear hormone receptors, the tumor suppressor p53, the proto-oncogenes c-myc and c-fos, and the VP16 protein (148). Mutations in specific E3 ubiquitin ligases can stabilize an activator protein but reduce its transcriptional activity (147). Moreover, ectopic addition of a single ubiquitin moiety to an activator can enhance its transcription activity in a non-proteolytic manner. Counterpart to these clues regarding ubiquitin is evidence that components of the proteasome can also be found associated with transcriptional promoters in conjunction with gene regulation (36, 42, 128). Although the explicit roles of ubiquitin and the proteasome in the mechanism of transcriptional activation remain uncertain, one model proposes that the addition of ubiquitin may promote interaction of activators with key target proteins. Subsequent ligation of

additional ubiquitin groups may lead to the degradation of the activator by the proteosome, providing a rapid downregulation of the activating signals (reviewed by (97)).

Activator association with promoters

Core promoter architecture and activator specificity

The cis-acting regulatory sequences for genes transcribed by RNA pol II typically comprise a combination of core promoter elements (reviewed in (156)) which serve as the binding site for the basal transcriptional factors and define the translational startpoint. The most obvious feature in many core promoters is the TATA box, located about 30 bp upstream of the transcription start site. Surrounding the start site itself may be found an initiator (Inr) element. A downstream promoter element (DPE) has been defined in Drosophila and in mammalian genes, albeit not in yeast (63). Furthermore, a TFIIB recognition element (BRE), flanking the TATA box, has been described in some organisms (156).

Differences among core promoters of various genes, with respect to the presence and strength of these core elements, have pronounced effects on how those promoters respond to particular transcriptional activators. This specificity may allow a particular enhancer to differentially regulate various target genes, and may allow a particular core promoter

to selectively respond to different enhancers. For example, the Sp1 activation domain strongly activates core promoters containing TATA elements, Inr elements, or both, whereas the Gal4-VP16 activator is most effective at a core promoter with both TATA and Inr (32). In an "enhancer trapping" study in Drosophila, many enhancers were able drive expression from both TATA dependent and DPE dependent promoters, but some enhancers preferentially activated either one or the other core promoter (18). The mechanistic differences leading to enhancer-core promoter specificity may involve the recruitment of transcriptional cofactors that display core promoter specificity. For example, the transcriptional cofactor NC2 represses transcription from TATA dependent promoters, but activates transcription from DPE dependent promoters (176). Thus, either direct or indirect recruitment of transcriptional cofactors by an activator may differentially regulate transcription from different core promoters, leading to activator-promoter specificity.

Enhancers and upstream activating sequence elements

In addition to the core promoter elements, transcription of many genes depends on cis-acting regulatory elements termed enhancers or upstream activating sequences, which provide the binding sites for transcriptional activators. These DNA elements can vary in sequence and affinity for a particular DNA binding domain, and may exist in promoter-proximal locations or hundreds to thousands of basepairs upstream or downstream of a promoter. The consensus sequences of various cis-acting regulatory elements and the

proteins that recognize and bind to those elements have been cataloged in various databases now available at internet websites, including the Eukaryotic Promoter Database at www.epd.isb-sib.ch (132) and the TRANSFAC database at www.gene-regulation.com (108).

Combinations of activator binding sites may be clustered into more complex regulatory elements. In such cases, cooperative binding of activators leads to the formation of large DNA-protein structures termed enhanceosomes, resulting in synergistic effects on transcription. In the prototypical virus-inducible IFN-β enhanceosome (reviewed in (112)), a cluster of three different activator binding sites direct the expression of IFN-β in response to viral infection. However, none of the activator binding sites act alone; only the combination of all three activator binding sites recapitulates the logic necessary to drive proper expression and specificity of the IFN-B gene. In this model system, the enhancer represents not simply the sum of individual activator functions, but rather an integration of inputs from different sources interpreted by the particular combination of transcription factors present at the enhancer.

Taking this organizational theme one step further, many genes may have multiple enhancers, each of which is poised to respond to particular developmental, growth, or environmental signals. Each of these independent enhancers may be simultaneously signaling to the core promoter either to stimulate or repress transcription, depending on the signal inputs received by the regulatory proteins that bind there. This diverse and sometimes conflicting information must be integrated and interpreted at the promoter to

make a final decision on whether or not transcription is to proceed. This "information display" model for genes with multiple enhancers has arisen from studies of developmentally related genes in Drosophila (86), but will likely be relevant for many genes in humans as well.

With the increasing availability of genomic sequences for prominent experimental organisms, computational analysis for identifying cis-acting regulatory sequences has become a growth industry. In some cases, these searches focus on particular cis regulatory elements, such as the estrogen response elements in mammalian genomes (4). Other programs are designed for broader application, searching for sites corresponding to any of the transcription factors in the TRANSFAC database and combining site searches to increase the likelihood of identifying legitimate regulatory regions rather than idiosyncratic consensus sequence matches (10, 47, 69). These computational approaches yield results that still must be validated by direct evidence of the function of putative elements in gene regulation and their interaction with specific transcription factors. Although this is often done on a case-by-case basis, either by mutational analysis of the cis elements or by in vitro binding assays, more global assessments are also now possible. For several transcription factors in yeast (141) and in mammalian cells (140, 172), genomic mapping of transcription factor binding sites has been accomplished by combining chemical crosslinking and immunoprecipitation (so-called chromatin IP or "ChIP" assays) with DNA microarrays comprising intergenic or putative regulatory sequences.

Actions of activators at promoters

Once localized to a promoter, a transcriptional activator can interact with a number of different targets, including RNA pol II, the basal transcription factors, the mediator complex, coactivators, and chromatin-remodeling machinery. A common theme in models of activation is recruitment, where a promoter-bound activator recruits either a component of the transcriptional machinery or a transcriptional cofactor. This model is supported by evidence of direct physical interactions of activators with basal transcription factors, and by activator bypass experiments (reviewed in (134)). In the latter experiments, a component of the transcriptional machinery is fused directly to a DNA binding domain, and this artificial recruitment serves to activate transcription. Alternatively, in a variation of recruitment, a transcriptional activator may modulate the activity of components of the transcriptional machinery, facilitating the assembly of the preinitiation complex.

Stepwise recruitment of basal transcription machinery

Transcription of protein-coding genes requires the assembly of a preinitiation complex (PIC) comprising RNA pol II, the general transcription factors (GTFs), and a number of associated factors. In one stepwise model of PIC assembly, the TATA-binding protein (TBP)-containing TFIID complex binds to a promoter, followed by TFIIA, TFIIB, TFIIF

and RNA pol II, and TFIIE and TFIIH (16). In this model, any step of PIC assembly might be rate limiting, and the recruitment of GTFs by association with activators may facilitate assembly.

TFIID, TBP and TAFs

The TFIID protein complex comprises TBP and several TBP-associated factors (TAFs) (17, 135). TBP binds selectively to the TATA core promoter element, while the TAFs extend the footprint to include the Inr and DPE elements. Studies in vitro show that although TBP is sufficient for basal transcription, the TAFs are required for activated transcription (31, 106, 149).

TBP can bind directly to transcriptional activation domains, as demonstrated by in vitro binding assays using a wide range of activator proteins, and mutations in activators that weaken activation also weaken interactions with TBP (56, 153). While these results suggest that activators simply recruit TBP, other evidence suggests that the mechanism is more complicated. Two "indirect recruitment" models of activator interaction with TBP are described below.

Some TAFs have a direct affinity for certain transcriptional activators, suggesting a role in recruitment or modulation of activity. For example, the glutamine-rich activators Sp1, NFAT, and CREB interact with the TAF4 protein from Drosophila or human cells (21, 35,

41, 71, 144, 178). The acidic VP16 and p53 activation domains can interact with TAF9 (43, 75, 166). However, these interactions cannot always be interpreted to imply an effect on TFIID, since a significant number of these TAF proteins are present in other protein complexes besides TFIID that nonetheless influence transcriptional activation (48, 175).

TFIIA, TFIIB, TFIIF, and TFIIH

TFIIA is a positive cofactor in PIC assembly, as it binds cooperatively with TFIID at TATA DNA elements. TFIIA also functions as an antirepressor, inhibiting the TBP-DNA destabilizing actions of Mot1 and NC2 (reviewed in (135)). The formation of the ternary TFIID-TFIIA-DNA complex is a rate-limiting step in PIC formation, and activators can enhance this step (76, 93). Some evidence points to direct association of the VP16 activation domain (VP16AD) with subunits of TFIIA (76, 77).

TFIIB also stabilizes the TBP-TATA complex, and serves as a docking site for other components of the PIC. Several activators, including VP16AD, have been shown to bind TFIIB with a high affinity (96), and activator-TFIIB connections have been implicated in transcriptional activation (143). Interaction with VP16AD has been shown to alter the conformation of TFIIB, possibly priming it for incorporation into the PIC (52, 142) or altering TFIIB-DNA contacts (33). Although this evidence is points to TFIIB as a potential activator target, other reports have failed to find evidence supporting this

association (43, 153), and thus the role for TFIIB as a target for VP16AD remains controversial.

TFIIH, which contains both protein kinase and nucleic acid helicase activities, also appears as a target for activation domains. The activation domains of VP16, p53, and E2F1 can interact with TFIIH (130, 179) and recruitment of TFIIH may stimulate promoter escape (87), but no clear evidence exists that activators stimulate the enzymatic activities of TFIIH. TFIIF and TFIIE might also be targets for activation domains. The serum response factor (SRF) interacts with the RAP74 subunit of TFIIF (60), and Fos-Jun dimers can interact with both TFIIF and TFIIE (107). Although the mechanistic implications of these interactions are not fully developed, the dual role of TFIIF as both an initiation and elongation factor suggests the possibility that activators modulate promoter escape or elongation in addition to assembly of the preinitiation complex.

Recruitment of RNA pol II holoenzyme

Many of the models described in the preceding sections are predicated on the premise that transcriptional activation involves a sequential recruitment of the basal transcription factors and RNA pol II to form the PIC at the target promoter. This premise was challenged, however, by the biochemical purification from yeast cells of an extraordinarily large protein complex comprising RNA pol II stably associated with a subset of GTFs together with additional polypeptides from the mediator complex, as

described below (74). Certain transcriptional activators were shown capable of recruiting this "holoenzyme" to a promoter in a manner sufficient to achieve transcriptional activation in vitro (54, 74). The model arising from these observations is that rather than separately and sequentially recruiting each general transcription factor, activation might more simply involve recruitment of the distinct TFIID and holoenzyme complexes. Similar RNA pol II holoenzyme complexes have been also purified from human cells (81) suggesting that the recruitment of the holoenzyme by activators might be an evolutionarily conserved mechanism (reviewed by (51)). Kinetic and thermodynamic questions arising from the two competing models have not been fully resolved. How could the stepwise assembly occur quickly enough (given diffusion parameters for each component) for efficient transcriptional activation? And yet, how can a complex the size of the holoenzyme be translocated to specific genes at specific times quickly enough to respond to transcriptional activation?

Recruitment of the mediator complex

The mediator complex, first identified as a component of the yeast RNA pol II holoenzyme (74), comprises about 20 subunits forming three major domains (Gal 11, Med9/10 and Srb modules) that wrap around the RNA pol II (reviewed by (11, 116)). Homologues of the yeast mediator subunits and similar protein complexes have since been identified in a wide range of organisms. Mammalian protein complexes resembling yeast mediator were described independently by several laboratories using biochemical

purifications of proteins stably bound to different activator proteins (reviewed by (103, 136)). These different purifications lead to very similar complexes (variously termed TRAP, DRIP, ARC, CRSP, SMCC and PC2) suggesting that different activators bind the same or highly related human mediator complexes. The slight differences in protein compositions of these human mediators might represent variations due to differences in the biochemical purifications or might represent different forms of the mediator complex that associate with different activators.

Several activators are known to interact physically and functionally with the mediator complex and the particular mediator subunits involved in these interactions are being identified. For example, the p53 tumor suppressor protein interacts with the TRAP80 subunit of the human mediator complex whereas the thyroid hormone receptor and PPARγ2 interact with TRAP220 (38, 57). The VP16AD may associate either with TRAP80 or with ARC92 (57, 114). Interferon-stimulated transcription depends on an interaction of STAT3 with the DRIP150 mediator component (89). These and other examples indicate that the mediator complex can be considered as a modular interface connecting activators with RNA pol II allowing the integration of different signals during transcriptional activation.

Recruitment of chromatin-modifying coactivators

Transcription activation must overcome the physical barriers presented by the packaging of eukaryotic DNA into chromatin. Chromatin compaction can mask activator binding sites, mask core promoter elements, and can impede the actions of RNA polymerase. It is now clear that the histone proteins upon which DNA is wrapped are not only a static scaffold for the compaction of DNA, but rather they participate actively in the regulation of gene expression. Transcriptional activators can affect chromatin by recruiting members from two general classes of chromatin-modifying coactivator proteins. One class comprises enzymes that covalently modify amino acids within the histones themselves. These modifications then either directly alter chromatin structure, or serve as recognition signals for binding additional proteins that modulate that structure (55). Covalent modifications that have been identified in histone proteins include acetylation, methylation, phosphorylation, ubiquitinylation, and ADP- ribosylation (reviewed by (8)). The panoply of such modifications on the various histones has been likened to a "code" (59, 158) that, when deciphered and integrated, signals whether and how strongly a given gene is to be expressed.

A second general class of chromatin-modifying transcriptional coactivators comprises the ATP-dependent chromatin remodeling complexes. The mechanisms of these remodeling activities are not yet fully understood, but include local and stable alterations of the DNA-histone contacts leading to sliding of nucleosomes along DNA or transfer of nucleosomes from one DNA to another in trans. These protein complexes can also alter

the superhelicity of DNA in vitro (reviewed by (118)). Repositioning of nucleosomes may alleviate chromatin-mediated transcription repression, for example by exposing DNA binding sites for additional activators or by exposing core promoter elements that might be critical for the binding of general transcription factors and the formation of the pre-initiation complex.

The handoff mechanism

Some evidence suggests that the interaction of an activator with TBP may involve more than simply recruiting TBP to a promoter. The acidic activator Gal4 binds TBP competitively with TATA DNA (180), and TBP and Gal4 do not bind cooperatively to promoters (181). Additionally, a mutation (L114K) in the DNA binding region of TBP interferes with activator binding (72), indicating that activators and DNA can bind to overlapping regions of TBP. Taken together, these observations suggest that competition for the DNA binding domain of TBP may be involved in some activation mechanisms.

The TFIID complex associates with TATA box DNA more slowly than does isolated TBP, implying that the TAFs contain inhibitory functions (78). Some activators, including VP16 and the Zta protein of Epstein-Barr virus, can stimulate the assembly of a TFIID-TFIIA-DNA complex (76, 93), but do not stimulate a ternary complex when TBP is used instead of TFIID, suggesting that activators may counteract inhibitory functions of the TAFs. This ability of activators to stimulate ternary complex assembly appears to

be relevant for activation, since mutations in an activation domain that reduce transcriptional activation potential in vivo also diminish the in vitro TFIID-TFIIA assembly function (77).

In yeast, two of the TAF inhibitory activities have been mapped to two N-terminal domains in the largest TAF subunit (84). These TAF N-terminal domains, called TAND1 and TAND2, bind to TBP and competitively inhibit the interactions of TBP with DNA and TFIIA, respectively (79, 84). Additionally, the C-terminal subdomain of VP16AD (VP16C) can compete with TAND1 for binding to TBP (124). These observations have led to a handoff model for activation (83) (figure 10, chapter 4), in which VP16C competes with TAND1, in turn loosening the TAND2-TBP connection, which then allows TFIIA to associate with TBP. Then, in a step that is still not understood, TBP is handed-off from VP16C to DNA, leading to TFIID-TFIIA-DNA complex assembly. The handoff step may involve conformational changes in TFIID or the formation of additional contacts between TAFs and core promoter elements.

Interestingly, the handoff mechanism presents a situation where a finely-tuned affinity between an activator and a target may be beneficial. If an activator interacts too weakly with TBP, it will compete poorly with the repressive TAND1, but too much affinity for TBP may interfere with DNA binding. Indeed, this may be the case: *Drosophila* TAND1, which exceeds both VP16C and yTAND1 in affinity for TBP, is nonetheless a weaker activator when fused to a DNA binding domain (83). Thus, the correlation of in vitro

affinity and in vivo function may be broken in mechanisms requiring transient interactions between activators and targets.

Modulation of TBP orientation

In another variation on the theme of recruitment, an activator might function by altering the orientation of the PIC. TBP can bind specifically to the TATA sequence present on some promoters, and thus the TBP-DNA complex is thought to serve as a platform for construction of the PIC. The conserved core region of TBP possesses an approximate 2-fold symmetry, as does the TATA DNA element. However, the TBP surfaces contacting other basal transcription factors (TFIIA, TFIIB, and TAFs) are unique, implying that the orientation of TBP on a promoter defines the orientation of the overall PIC. TBP has been reported to bind the TATA element in vitro with little or no orientational preference, although interactions with other basal transcription factors appear to enforce a productive polarity on TBP orientation (26, 68).

Both VP16AD and a synthetic amphipathic helix have been reported to significantly increase the orientational specificity of TBP on the TATA DNA element in vitro (67). This represents a variation on recruitment which might prime the TBP-DNA complex for productive association with additional general transcription factors, and which might work against the formation of incorrectly oriented PICs.

Leaving a mark: activator effects after initiation

Although most studies of transcriptional activation have focused on recruitment and initiation, subsequent steps including promoter escape and elongation can also be stimulated by activator proteins. Several lines of evidence indicate that activators might also work in post-initiation steps. One well-characterized example corresponds to the human and Drosophila gene encoding heat shock protein 70 (hsp70). The uninduced hsp70 gene contains a paused polymerase near the 5' end of the gene. In response to heat shock, not only does the transcriptional initiation rate increase, but the pausing time is dramatically reduced (13, 145). Transcriptional activators can also stimulate rates of transcriptional elongation. For example, the heat shock factor-1 (HSF-1) involved in the activation of hsp70 gene and the viral activators VP16 and E1A can stimulate elongation by mechanisms that apparently differ from that of stimulation of initiation (14, 183). The interaction of activators including HIV tat, c-myc, and NF-kB with the elongation factor P-TEFb (a kinase that modifies the carboxyl-terminal tail of the largest subunit of RNA pol II) further illustrates this link (5, 22, 65, 66).

Transcription and RNA processing have often been considered as separate and sequential events, but a more recent perspective views these as a single integrated pathway (reviewed by (133)). Capping, splicing and polyadenylation are tightly coupled to RNA pol II through the carboxyl-terminal domain of the largest subunit (reviewed by (104)). Selection of splice sites in the nascent RNA can be influenced by promoter elements in a

manner that is independent of the promoter strength, suggesting that activators might also regulate alternative splicing decisions (2, 27, 125).

Structure of transcriptional activators

The dual functions of a transcriptional activator protein, cis-element recognition and transcriptional activation, are typically fulfilled by distinct regions of the protein's primary structure. For example, the DNA binding domain of the yeast Gal4 protein (a prominent activator in the yeast *Saccharomyces cerevisiae*) resides within the amino-terminal 100 amino acids, whereas the major transcriptional activation domain resides within the carboxyl-terminal 120 amino acids. This modular design seems advantageous both for evolutionary and technological appropriation. In the latter sense, a fusion protein linking the DNA-binding domain of Gal4 with VP16AD (146) is widely used in both in vitro and in vivo investigations into the mechanisms of transcriptional activation. A second example, comprising a fusion of the DNA binding domain of the tetracycline repressor with VP16AD, allows the regulation of DNA binding by the presence or absence of the tetracycline ligand (45). This regulatable artificial activator can function in a wide range of eukaryotes ranging from plants to mammals (44, 46, 173).

DNA binding domains

The DNA binding domains of a large number of eukaryotic transcriptional activator proteins have been extensively characterized by genetic, biochemical, and structural approaches. Recent reviews catalog the known structures and specificities (101) and highlight the common themes in structure and recognition (37). In many cases, the binding activity or specificity of a DNA binding domain may be modulated by ligand binding, dimerization with other DNA binding proteins, or by association with other factors (reviewed by (105)).

The principles of protein:DNA interaction have now been established to a sufficient degree to permit the design of DNA binding modules of engineered specificity (6, 34). This is particularly true for the zinc-finger families of transcription factors (24, 64, 151, 177). This ability to tailor novel chimeric transcriptional activators for recognition of DNA sequences that might not serve as native regulatory elements has profound implications for potential pharmacological application (95, 100, 184).

Activation domains

Features in primary structures

In contrast to DNA binding domains and despite substantial research, relatively little is known about the structures of transcriptional activation domains. By analysis of primary sequence, these domains have been broadly classified based on the abundance of particular amino acids, resulting in acidic, glutamine rich, proline rich, and other classes (113, 163). Despite these amino acid preferences, however, careful mutational analyses have indicated that the most critical elements of activation domains are frequently the patterns of hydrophobic and aromatic amino acids (1, 19, 28, 58, 163). For example, in the acidic VP16AD, mutation of one or several acidic residues has only a modest effect on activity, whereas mutation of a single key hydrophobic residue can severely reduce activity (28, 139, 160).

Secondary and Tertiary Structures

Less is known about the secondary and tertiary structures of activation domains. A number of biophysical analyses of various regulatory proteins have shown that activation domains are largely unstructured in solution under physiological conditions (29, 126, 150, 152). However, key amino acids in an activation domain can become conformationally constrained upon interaction with a target protein, suggesting that the most promising

targets for structural studies will be binary complexes between activators and targets. For example, circular dichroism spectra indicate that the c-Myc transactivation domain is induced to form a helical structure upon binding to TBP (109). The activation domains of VP16 and of the estrogen receptor also become conformationally constrained upon interaction with TBP (153, 171). Furthermore, the VP16AD appears to become helical upon binding human TAF32 (166). An amphipathic helix from the p53 activation domain fills a hydrophobic cleft in the MDM2 oncoprotein (88). An amphipathic helix structure is also seen in the interface between the activator CREB and its coactivator protein CBP (137). These examples support the model that an activator target provides a folding template for an unstructured activation domain, which might allow a particular activation domain to interact with a number of different target proteins.

In some cases the tables may be turned: that is, an activation domain may provide the folding template for a potential target. Nuclear hormone receptors have a conserved C-terminal activation domain, known as AF-2. Ligand binding leads to a conformational change in the activator that opens a hydrophobic groove for interaction with transcriptional co-activators. In this case, the unstructured LxxLL peptide motif present in several coactivators (53) folds into an amphipathic helix upon binding the AF-2 region of a hormone receptor (reviewed in (170)).

Although in many cases activation domains seem to adopt α-helical structures, that rule is not universal. Mutational and biophysical analysis of the Gal4 activation domain suggested that it might form a β-strand instead (168). The activation domain of E2F-2

upon interaction with the Rb tumor suppressor protein assumes a combination of helical and β-strand conformations (90). Together, these observations suggest that activators and their target proteins bind to each other using a highly diverse set of interaction surfaces.

Central Hypothesis

Many potential targets of transcriptional activation domains have been identified, and the detailed interaction of activators and targets may involve a range of mechanisms. However, we do not know which targets and mechanisms are relevant for transcriptional activation in vivo. The strong interaction of VP16AD with TBP and the correlation of this interaction with VP16AD function lead to the hypothesis that TBP is a relevant in vivo target for at least one class of activation domains. Additionally, several lines of evidence suggest that VP16AD and DNA compete for binding to TBP, which argues that VP16AD cannot stably recruit TBP to a promoter. This leads to the hypothesis that the VP16AD:TBP interaction in transcriptional activation involves a variation of direct and stable recruitment. Variations on recruitment include the modulation of TBP orientation on DNA, discussed in chapter 3, and the transient interactions involved in a handoff mechanism, discussed in chapter 4.

Chapter II

Purification of yeast TATA binding protein

Introduction

The TATA binding protein (TBP) plays a central a central role in all classes of eukaryotic transcription (25). TBP comprises a highly conserved C-terminal DNA binding domain and an N-terminal domain that varies considerably among species. TBP specifically binds to the TATA DNA sequence element (156) found about 30 bases upstream from the transcription start site in many promoters. Since TBP can specifically bind to the TATA elements found in some promoters, it is thought to nucleate the assembly of the large protein complexes that poise RNA polymerases at the transcriptional start sites of genes (50).

TBP has been purified by a variety of methods, including bulk fractionation (23), conventional chromatography (9, 20, 23), and fusion to affinity tags. Most large-scale TBP purifications use a combination of these methods; however, each of these purification methods has limitations. Conventional chromatography typically relies on low salt binding conditions, but TBP readily aggregates under low salt conditions (personal observations). Heparin affinity chromatography provides an alternative to ion exchange methods, and since TBP can be bound to heparin columns at 200 mM salt (9, 131), low salt conditions can be avoided. However, heparin affinity chromatography alone is insufficient to purify TBP to homogeneity (131). N-terminal polyhistidine tags

and nickel affinity chromatography have frequently been employed (62, 85, 121, 165), although subsequent dialysis is typically required to remove the elution agent (imidazole or EDTA), and removal of the extrinsic polyhistidine tag requires proteolysis and the subsequent removal of protease activity.

In our efforts to stabilize a complex of *Saccharomyces cerevisiae* TBP bound to the transcriptional activator VP16C (120, 139, 146), we explored a fusion of VP16C to another TBP-interacting peptide, TAND2 (3, 79, 102). This VP16C-TAND2 fusion, based on a precedent by Kotani et al. (83), specifically binds TBP with high affinity in a salt-dependent manner. We have exploited this interaction as a simple, fast, and effective means of purifying untagged yeast TBP.

Materials and Methods

Plasmid construction. To generate an expression plasmid encoding a fusion of glutathione-S-transferase-VP16C-TAND2 (GST-VP16C-TAND2), a DNA fragment encoding TAND2 (TAF1 residues 42 to 78) was PCR amplified from yeast genomic DNA using primers ST805 and ST806. The resulting DNA fragment was tailored by *Eco*RI and *Xho*I and directionally cloned between the *Eco*RI and *Xho*I sites of plasmid pGEX4T-2 (GE Healthcare), yielding plasmid pDS42-5. VP16C (VP16 residues 451-490) was PCR amplified from plasmid pGEX.VP16C (119) with primers ST803 and ST804. The resulting DNA fragment was tailored by *Bam*HI and *Eco*RI and directionally

cloned between the *Bam*HI and *Eco*RI sites of pDS42-5 to yield pDS42-10. Plasmid fidelity was verified by sequencing. Plasmid pKA9, encoding *Saccharomyces cerevisiae* TBP under control of the T7 promoter, was obtained from Dr. Karen Arndt.

Expression of GST-VP16C-TAND2. Plasmid pDS42-10 was transformed into *E. coli* strain BL21 (Novagen) and cells containing plasmid were grown in LB medium containing 80 μg/mL ampicillin. One liter cultures were grown at 37 °C with vigorous shaking until an OD_{600} of about 0.7. Expression of GST-VP16C-TAND2 was induced by addition of IPTG to a final concentration of 0.5 mM and growth was continued for an additional 3 hours while the culture temperature was allowed to approach ambient temperature (25 °C). Cells were harvested by centrifugation and resuspended in 1/60 culture volume HEMGT-250 (24 mM HEPES, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% glycerol, 0.1% Tween-20, 250 mM KCl, pH 7.9) supplemented with 1 mM PMSF and 5 mM DTT. Cells were lysed by two passages through a French Press (Aminco) operating at 20,000 psi. Cellular lysates were centrifuged at 10,000 x g at 4 °C for 20 min. The supernatant was recovered, split into 1 mL aliquots, flash frozen in liquid nitrogen, and stored at -80 °C.

Expression of yeast TBP. E. coli strain BL21(DE3) Codon Plus (Novagen) was transformed with plasmid pKA9 which encodes yeast TBP under control of the T7 promoter. Cells containing expression plasmids were grown in TB medium (Terrific Broth) with 80 mg/L ampicillin and 40 mg/L chloramphenicol. One liter cultures were grown at 37 °C with vigorous shaking until an OD_{600} of 1.5. Expression of TBP was

induced by the addition of IPTG to a final concentration of 0.1 mM and growth was continued for 3 h while the culture temperature was allowed to approach ambient temperature (25 °C). Cells were harvested by centrifugation and resuspended in 1/30 culture volume of HEGK-400 (25 mM HEPES, 1 mM EDTA, 10% glycerol, 400 mM KCl, pH 7.5) supplemented with 5 mM DTT. Cells were lysed by sonication (Branson Sonifier 450, 80% duty cycle, power setting 8, 4 cycles of 30 sec with 2 min rest between cycles). Cellular lysates were centrifuged at 9000 x g at 4 °C for 25 min. The supernatant was recovered, split into 1 mL aliquots, flash frozen in liquid nitrogen, and stored at -80 °C.

Quantitation of GST-VP16C-TAND2 recovery. Bacterial lysates (10 µL to 100 µL) were added to 200 µL of a 10% slurry of glutathione (GSH) resin in HEGK-400. Binding reactions were axially rotated for 45 min at 7 °C. Binding reactions were centrifuged at 500 x g for 1 min and the supernatant was removed by aspiration. To wash the resin, 1.5 mL ice-cold HEGK-100 was added with enough vigor to ensure resin resuspension and the resulting diluted slurry was promptly centrifuged at 500 x g for 1 min. After a total of 5 washes, 150 µL of elution buffer (0.9 x HEGK-100, 0.1 x 100 mM reduced GSH pH 7.9) was added, and the resin was axially rotated at 25 °C for 30 min. The resin was centrifuged again and the protein concentration in the supernatant was measured with a Bradford assay relative to BSA standards.

Purification of TBP. All steps except for elution were performed on ice or in a 7 °C environment. 400 µL of GVT-containing lysate was mixed with 10 volumes (4 mL) of a

TBP-containing lysate and the mixture was incubated on ice for one hour. The lysate mixture was added over 300 μL GSH resin (equilibrated in HEGK-100), and the slurry was gently rotated for 45 min at 7 °C. After resin binding, five washes were performed as follows: 75 resin volumes of HEGK-100 were added to the resin, resuspension was ensured, the slurry was promptly centrifuged at 500 x g for 3 min, and the supernatant was removed by aspiration. The washed resin was transferred to a polypropylene chromatography column and the supernatant was drained through the outlet. 2 mL of HEGK-1000 buffer at room temperature (25 °C) was carefully layered onto the resin, and the column was gravity eluted at a flow rate of approximately 1 mL/min. 1 mL fractions were collected into chilled tubes. TBP typically eluted in the first fraction. Typical recovery was approximately 1 mg.

Results and Discussion

In our efforts to stabilize a complex of TBP and VP16C, we noted that a previously developed fusion of VP16C and TAND2 bound tightly to TBP (83). Although the region of TBP bound by VP16C has not been precisely determined, DNA competition experiments and TBP mutations have localized the interaction to the concave DNA binding region of TBP (72, 83, 94). TAND2, an N-terminal domain from the largest TBP associated factor TAF1 (162), interacts with a different region of TBP, on the upper convex surface (102) and in competition with TFIIA (79). Since the previously constructed VP16C-TAND2 was shown to bind TBP with high affinity and also appears

to preserve the VP16C:TBP interface, we constructed and explored a similar fusion protein comprising GST, VP16C (VP16 451-490), and TAND2 (TAF1 42-78).

To form a complex of GST-VP16C-TAND2 (GVT) and TBP, a mixed-lysate approach was tested. GVT and TBP were expressed in separate *E. coli* cultures and the soluble fractions of the cellular lysates were stored at -80 °C. In a first step of lysate characterization, an amount of GVT-containing lysate sufficient to bind 100 μg protein to 20 μL GSH resin was determined. In a second step, increasing volumes of TBP-containing lysate were titrated over the constant volume of GVT lysate, the lysate mixtures were incubated on ice for one hour, and then purified over GSH resin. GVT and any associated proteins were eluted from the resin by boiling in SDS-PAGE loading buffer and then resolved by SDS-PAGE (Fig. 1A). This analysis indicated that GVT could retain TBP though extensive washing of the GSH resin.

In order to quantify the retention of TBP by GVT, high salt conditions were used to selectively elute TBP from the immobilized complex. Lithium sulfate was found to be particularly effective at disrupting the TBP:GVT interaction, and was thus used for analytical-scale elutions. To quantitate purified TBP, 50 μL of 1 M lithium sulfate was layered onto 20 μL of resin with bound protein complexes, the slurry was rotated axially for 20 min at room temperature, and 40 μL of the supernatant was assayed in a 1 mL Bradford assay relative to BSA standards. This method produced a binding curve that

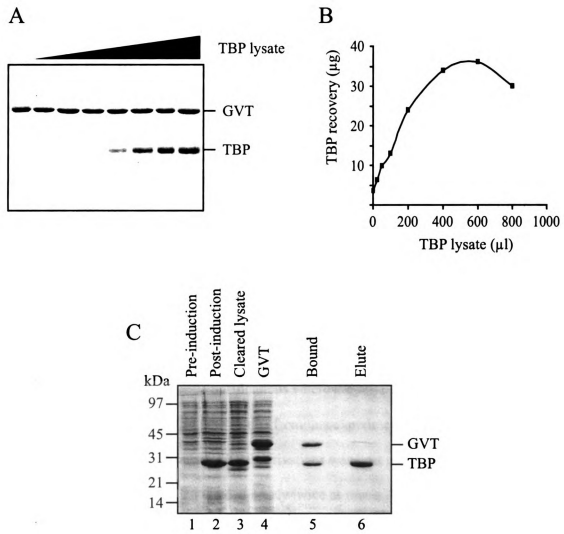


Figure 1: Purification of full-length yeast TBP. (A) Titration of increasing amounts of TBP-containing lysate over a constant amount of GVT-containing lysate. Lysates were mixed, incubated on ice for one hour, and purified over GSH resin. Bound proteins were eluted by boiling in SDS loading buffer, resolved by SDS PAGE, and stained with Coomassie R-250. (B) Quantitation of purified TBP. TBP was released from immobilized GVT:TBP complexes by addition of 1 molar lithium sulfate. Eluted proteins were quantitated by Bradford assay. (C) Purification of isolated TBP monitored by SDS-PAGE. Lanes 1-2, TBP culture before and after induction of TBP expression. Lane 3, lysate containing TBP. Lane 4, lysate containing GST-VP16C-TAND2. Lane 6, proteins retained on GSH resin after washing. Lane 8, TBP released by addition of buffer containing 1 molar potassium chloride.

closely resembled the SDS-PAGE data (Fig. 1B). Notably, the recovery profile of TBP indicated that with low amounts of TBP lysate, the recovery was linear with increasing TBP, while with higher amounts of TBP lysate the recovery approached saturation and then diminished. This indicated that lysate ratio can be tuned for different purposes. To make optimal use of GSH resin, saturating amounts of TBP lysate should be used. Conversely, to make optimal use of TBP lysate, lower amounts should be used.

To optimize the binding reaction, the salt concentration in the lysate mixture was explored. Since TBP readily aggregates under low salt conditions and the GVT:TBP interaction is weakened at high salt concentrations, a moderate salt concentration was expected to be optimal. The optimal salt concentration was determined by performing a parallel series of analytical scale purifications where the salt concentration in the lysate mixture was the only variable. The results from this experiment indicated that the recovery of N-terminally truncated TBP derivatives was best when the lysate mixture contained 400 mM KCl, which yielded nearly twice as much TBP as a lysate mixture containing 100 mM KCl. The salt sensitivity of full-length TBP was less pronounced, but also showed that recovery was optimal from a lysate mixture containing 400 mM KCl. Based on these results, and to simplify subsequent purification, the TBP lysis buffer has been defined to contain 400 mM KCl.

Since the lithium sulfate that was used in the analytical-scale elutions may interfere with downstream applications, potassium chloride was tested as an elution buffer for preparative scale purifications of TBP. To perform this test, 180 μL of elution buffers

with various concentrations of KCl were added to 20 μL beds of protein loaded resin. The samples were rotated axially for 20 min at different temperatures, and the supernatant was assayed for eluted protein. From this analysis, it was determined that 1 M KCl was sufficient to elute TBP at room temperature, while at least 1.25 M KCl was required for elution at 7 °C.

To scale up the purification method for preparative use, a column-based elution format was employed. Lysate mixtures were scaled up to supply protein for 300 μL of GSH resin. The resin was washed as in the small scale purifications, and then transferred to a plastic chromatography column. The resin was settled and the supernatant was drained, and 2 mL of (25 °C) HEGK-1000 was carefully layered onto the resin. The column was gravity-eluted at approximately 1 mL/min flow, and fractions were collected into tubes chilled in ice. TBP eluted primarily in the first fraction, and approximately 1 mg of TBP was recovered (Fig. 1C).

The TBP purified by this method has been successfully used in several experiments. Purified full-length TBP has been found to specifically recognize the TATA DNA sequence (Chapter 3), demonstrating proper function. A purification of the conserved 180 amino acid core region (residues 61-240) was crystallized (Chapter 4), indicating that the protein was properly structured. A preparation of TBP (residues 49-240) has also been used as a standard for gel filtration experiments (Chapter 5), where it eluted with an apparent molecular mass of about 34 kDa, in rough agreement with the expected dimeric state. Additionally, the gel filtration analysis showed no material eluting with a higher

apparent molecular weight, indicating that the preparation was not detectably denatured or aggregated.

This method of TBP purification has several distinct advantages over current methods. Foremost, pure TBP can be obtained in under three hours and without the need for chromatographic equipment. Additionally, the resulting material is monodisperse, structured, and functional. Importantly, TBP can be purified without the need for extrinsic affinity tags or any subsequent proteolysis or cleanup. The method can also be performed at any scale, providing for rapid purification based on need. Finally, since typically over 100 mg of GVT can be purified from a single liter of *E. coli* culture, a single large preparation of GVT-containing lysate (separated into small aliquots and characterized once) can provide for many TBP purifications.

Chapter III

The orientation of TBP on the TATA DNA sequence

Introduction

The core promoter region of a protein coding gene localizes the assembly of a large multiprotein complex comprising RNA polymerase II (polII) and several general transcription factors (GTFs). The resulting pre-initiation complex (PIC) poises polII in the proper location and direction for subsequent transcription of the downstream DNA. In vitro, the PIC can be assembled onto DNA by the ordered addition of GTFs, suggesting a stepwise PIC assembly pathway (16). However, a large multiprotein complex comprising polII and several additional factors can be purified from cells, suggesting that PIC formation may instead involve the recruitment of a larger polII "holoenzyme" (91).

The TATA binding protein (TBP), either alone or with TBP-associated factors (TAFs) in the TFIID complex (162), can bind to the TATA element present in many promoters. TBP is thus thought to mediate the initial protein-DNA contacts during PIC formation. The highly conserved DNA-binding region of TBP comprises two imperfect direct sequence repeats, and the tertiary structure possesses a high degree of pseudodyad symmetry (123). However, the contacts between TBP and the other basal factors are not symmetrical. TFIIA binds specifically to a region in the N-terminal repeat in TBP (39, 161), while TFIIB interacts with residues in the C-terminal repeat (122). The asymmetric

connections between TBP and other basal factors suggest that the orientation of TBP bound to a promoter dictates the orientation of the overall PIC and thus the location and direction of transcription.

TBP specifically recognizes the TATA element located about 30 bases upstream from the transcriptional start site in many promoters. TBP binds in the minor groove of DNA at the TATA sequence, inducing a large distortion in the DNA structure (70, 73). Notably, the minor groove of the TATA sequence provides few structural features to guide TBP toward a particular orientation. Nonetheless, the TATA sequence is found in a preferred orientation in promoters (15, 156), suggesting both that the TATA sequence can orient TBP, and that the TATA-directed orientation of TBP at a promoter serves an important role in directing transcription.

The intriguing symmetry of TBP and the TATA element have led to several studies that attempt to decipher their role in determining transcriptional directionality. All crystal structures of TBP bound to TATA DNA sequences show oriented binding, suggesting that TBP may bind to the TATA element in only one orientation. However, crystal lattice constraints may influence this outcome, possibly by trapping a slightly favored orientation. Efforts to determine the orientation of isolated TBP on DNA in solution by nuclease protection analysis (footprinting assays) are hampered by the symmetry of TBP. However, the larger TFIID complex extends the TBP:DNA complex asymmetrically and can thus be employed in DNA footprinting experiments to determine directionality of the TFIID:DNA complex.

Footprinting experiments with TFIID bound to an isolated TATA sequence revealed asymmetrical protection patterns directed by the orientation of the TATA sequence (16, 117, 167). Additionally, in vitro transcription reactions have indicated that an isolated TATA sequence can direct oriented (though weak) transcription (127). Although this evidence suggests that the orientation of the TATA sequence dictates the orientation of the PIC and transcription, other evidence indicates that this is not a strict rule. For example, an upstream activator can dominantly direct downstream transcription, regardless of the orientation of an intervening TATA element (127).

While these studies indicate that the TATA sequence orientation can orient the PIC and transcription, and also that this TATA-directed orientation can be overridden by an upstream activator, the results come from complicated contexts. The TFIID footprinting assays rely on TAFs, and the transcription reactions involve many additional factors. These additional factors provide additional protein-DNA contacts that may modulate the orientational specificity of TBP, and additional protein-protein contacts that may mediate connections to other factors.

To more directly assess the intrinsic directionality of TBP binding to the TATA sequence, Schepartz and colleagues developed an in vitro system to measure the orientation of TBP in the absence of other factors (26). By attaching the copper chelator 5-Iodoacetamido-1,10-phenanthroline (IAAOP) (129, 155) directly to TBP, they converted TBP into a site-specific nuclease. The IAAOP moiety was positioned such that the orientation of TBP

directed the nuclease activity to positions either upstream or downstream of the TATA sequence. Utilizing this affinity-cleavage methodology, it was determined that isolated TBP binds to the *Adenovirus* major late promoter (AdMLP) TATA sequence with only a modest degree of orientational preference. Inclusion of either TFIIA or TFIIB increased the orientational specificity (26), and inclusion of both TFIIA and TFIIB led to virtually unidirectional TBP binding (68). These results suggested a model of PIC assembly in which isolated TBP is unable to orient itself at a promoter, and thus the orientation of the PIC and the direction of transcription are dependent on other factors.

Subsequent experiments using similar affinity-cleavage assays indicated that the prototypical acidic activator Gal4-VP16 (146) can enhance the orientational specificity of TBP bound to the TATA element (67). This result suggested a novel mechanism of transcriptional activation, in which an activator functions by promoting forward-oriented PIC assembly, or possibly by working against reverse-oriented PIC assembly.

Intriguingly, the TBP-orienting mechanism implies a ternary interaction between VP16AD, TBP and DNA, although several other lines of evidence suggest that acidic activation domains and DNA compete for TBP binding, precluding such a ternary interaction. For example, the activation domains of VP16 and Gal4, when free in solution, disrupt the interaction of TBP and DNA (94, 180, 181). A mutant in the concave DNA binding region of TBP (L114K) abrogates TBP interactions with both VP16AD and the Gal4 activation domain (72), suggesting that activators and DNA bind to overlapping regions of TBP. Additionally, VP16AD has been shown to compete with

a TAF1 inhibitory domain (TAND1) for binding to the DNA binding region of TBP (124). Since VP16AD and DNA apparently compete for binding to TBP, it is unclear how VP16AD might modulate the orientation of the TBP-DNA complex.

To further explore the proposed TBP-orienting activity of VP16AD, we have replicated and extended the affinity cleavage experiments performed in the Schepartz lab. Unexpectedly, and contrary to the results from the Schepartz lab, we found that TBP binds to the TATA DNA element with a high degree of orientational specificity. We have also found that Gal4-VP16AD does not alter the intrinsic orientational specificity of TBP. In addition to our studies of wild-type TBP, we examined the orientational specificity of a TBP mutant that can direct forward transcription from a reversed TATA sequence in vivo (Dr. K. Arndt, pers. comm.). This mutant TBP has a slightly relaxed orientational specificity, which may explain the in vivo phenotype. Taken together, our affinity cleavage results indicate that the current model of TBP orientation needs to be revised.

Materials and Methods

Cleavage probe sequences. Plasmid pG5MLT, encoding five Gal4 binding sites fused to bases -50 to +10 of the *Adenovirus* major late promoter (AdMLP) and a 390 bp G-free sequence, was obtained from Dr. M. Carey, UCLA. The sequence of the TATA element in pG5MLT was altered by site-directed mutagenesis (QuikChange method, Stratagene). Oligonucleotides ST1130 and ST1131 were used to create plasmid pDSB51-1, encoding the symmetric TATA sequence (TATATATA). Oligonucleotides ST1132 and ST1133 were used to create plasmid pDSB51-5 encoding the reverse TATA sequence (CTTTTATA). Plasmid sequences were verified by sequencing.

Cleavage probe generation. Fluorophore-labeled oligonucleotides ST903-6FAM and ST904-HEX (synthesized by Integrated DNA technologies, www.idtdna.com) were used to prime PCR amplification of a 304 bp DNA segment from plasmid pG5MLT or derivative plasmids. 6FAM denotes 6-carboxyfluorescein attached to the 5' position of the oligonucleotide, and HEX denotes hexachlorofluorescein attached to the 5' position of the oligonucleotide. The resulting DNA segment, encompassing five Gal4 binding sites, an AdMLP TATA sequence, and a portion of the G-free sequence, was spin-column purified (Qiagen) and stored at – 80 °C.

TBP expression. Plasmids encoding full length *Saccharomyces cerevisiae* TBP and mutant derivatives were obtained from Dr. Karen Arndt (pKA9: wild-type; pJVS56: K97C, F227I; pJVS57: E188C, F227I; pJVS58: K97C; pJVS59: E188C). TBP

expression plasmids were transformed into *E. coli* strain BL21(DE3)codon-plus. Cells containing expression plasmids were grown in LB medium containing 80 mg/L ampicillin and 40 mg/L chloramphenicol. 5 mL cultures were inoculated into 1 L of TB medium (Terrific Broth) containing 80 mg/L ampicillin, and cultures were grown at 37 °C with vigorous shaking to an OD_{600} of about 1.5. Expression of TBP was induced by the addition of IPTG to a final concentration of 0.1 mM, the ambient temperature was shifted to 25 °C, and growth was continued for 3 additional hours. Cells were harvested by centrifugation and resuspended in 30 mL HEGK-400 (24 mM HEPES, 1 mM EDTA, 10% v/v glycerol, pH 7.5) with 400 mM KCl and 5 mM DTT. Resuspended cells were lysed by sonication (Branson Sonifier 450, 80% duty cycle, power level 8, 4 cycles at 30 sec with 3 min rest between cycles). Cellular lysates were centrifuged at 9,000 x g at 4 °C for 25 min. The lysate supernatant was collected, flash frozen in liquid nitrogen, and stored at -80 °C.

TBP purification. TBP was purified using an intrinsic affinity method described in chapter 2. Briefly, bacterial lysates containing TBP and GST-VP16C-TAND2 were mixed and incubated on ice for one hour. Lysate mixtures were added over glutathione resin and the resulting slurry was gently rotated at 7 °C for 45 min. The resin was collected by centrifugation at 500 x g and the lysate mixture was removed. Five washes were performed as follows: 75 bed volumes of ice-cold HEGK-100 were added, resin resuspension was ensured, and the dilute slurry was immediately centrifuged at 500 x g. After the final wash, the resin was transferred to a fritted column, 2 mL of HEGK-1000 (25 °C) was layered onto the resin bed, and column elution fractions were collected.

Eluted material was flash frozen in liquid nitrogen and stored at -80 °C. The concentration of eluted TBP was measured by the Bradford assay relative to BSA standards.

IAAOP preparation. A 3 mg sample of IAAOP was obtained from Dr. A. Schepartz (Yale). IAAOP is light sensitive, and thus was handled under minimal light conditions. The IAAOP sample was suspended in 88 μL dimethylformamide (DMF) and stored at -80 °C as a master stock. Working stocks (10 mM IAAOP in DMF) were diluted from the master stock.

TBP derivitization. TBP derivatization was performed under low light conditions. 230 μg (8.5 nmol) of TBP was diluted into a final volume of 600 μL in HEGK-1000 (pH 8.2). Ten equivalents of IAAOP (85 nmol) in DMF was added and the solution was gently mixed. Derivatization reactions continued at 25 °C for one hour in the absence of light. To remove unreacted IAAOP, the derivatization reactions were buffer-swapped by dilution into 20 volumes of HGK-400 (24 mM HEPES, 10% v/v glycerol, 400 mM KCl, pH 7.9) and concentration with a centrifugal concentration device (Millipore, 10 kDa molecular weight cut-off). After two cycles of buffer swap, the recovered material was split into single-use aliquots, flash frozen in liquid nitrogen, and stored at -80 °C. The concentration of the resulting TBP-OP was estimated by a Bradford assay relative to BSA standards.

Gal4-VP16AD. Gal4-VP16 was purified as described (119).

Cleavage reactions. Cleavage reactions were carried out as previously described (67) with modifications. 100 ng of fluorophore-labeled probe DNA was incubated with 50 ng TBP-OP, 1 μg poly (dG-dC)•poly(dG-dC), and 12 μg BSA in 300 μL binding buffer (4 mM Tris-Cl, 60 mM KCl, 5 mM $MgCl_2$, 4% glycerol, 0.1% NP-40, pH 8) for 35 min at 25 °C while protected from light. To initiate DNA cleavage, 60 μL of binding buffer containing 30 mM 3-mercaptopropionic acid, 0.06% w/v H_2O_2, and 300 μM $CuSO_4$ was added to the TBP:DNA binding reactions. DNA cleavage was continued for 3 h at 25 °C while protected from light. Cleavage reactions were terminated by ethanol precipitation.

Footprinting reactions. A solution comprising 67 μM Methidiumpropyl-EDTA, 133 μM $Fe(NH_4)_2(SO_4)_2$, and 17 mM DTT was prepared immediately before use. 90 μL of this solution was added to 300 μL of the TBP:DNA binding reactions. Cleavage was continued for 5 min at 30 °C and subsequently terminated by ethanol precipitation.

DNA fragment analysis. DNA cleavage and footprinting reactions were ethanol precipitated, washed twice with 70% ethanol, and dissolved in 18-20 μL formamide mixed with in-lane size standards (Applied Biosystems ROX-500 mixture or equivalent). Samples were analyzed by an Applied Biosystems 3100 genetic analyzer (30 cm capillary, 60 sec injection time) at the Research Technology Support Facility, MSU.

Quantitation of DNA fragment data. Raw chromatographic data was extracted from the ABI Prism data files with the program Batchextract.exe (ftp.ncbi.nih.gov/pub/

forensics/batchextract/). Orientation measures were performed by the program B55-3.C as follows. In-capillary size marker positions were used to define the center of the cleavage pattern. From this center position, raw fluorescence intensity data were summed in regions spanning 13 bases upstream or downstream from the center position. A signal baseline was approximated with a straight line connecting minima found in 17 base regions flanking the quantitated regions. Baseline areas were subtracted from the raw data sums. TBP orientation was measured from the baseline-corrected fluorescence data as the summed signal in one quantitation region divided by the sum from both regions.

Results

To explore the TBP orienting activity of the VP16 activation domain we have replicated and extended the affinity cleavage experiments performed by the Schepartz laboratory (26, 67). In these experiments, the DNA cleavage agent IAAOP was attached to an engineered cysteine (K97C or E188C) located in one of the two stirrup loops of TBP, resulting in affinity-cleavage proteins K97C-OP and E188C-OP. In the TBP:DNA complex, the stirrup loops lie in the minor groove of DNA at positions flanking the 8 bp TATA sequence. Thus, attachment of IAAOP to a stirrup loop position on TBP localizes DNA cleavage activity to one side of the TBP:DNA complex. Subsequent analysis of the DNA fragment distribution allows the determination of the orientation of TBP bound to DNA (Fig. 2).

Our TBP orientation measurements utilized the same DNA cleavage methods as the Schepartz group, but we have used a different method of DNA fragment analysis. The Schepartz group used radioactively end-labeled DNA cleavage probes, resolved the affinity-cleaved DNA fragments on polyacrylamide sequencing gels, and then quantified gel images to determine the distribution of DNA cleavage events in the vicinity of the TATA element. In contrast, we used fluorophore end-labeled DNA cleavage probes and resolved affinity-cleaved DNA fragments using an Applied Biosystems 3100 genetic analyzer. This instrument couples capillary electrophoresis with fluorescence detection, allowing the simultaneous separation and detection of different uniquely labeled DNA fragment populations. By attaching different fluorophores to the top and bottom strands

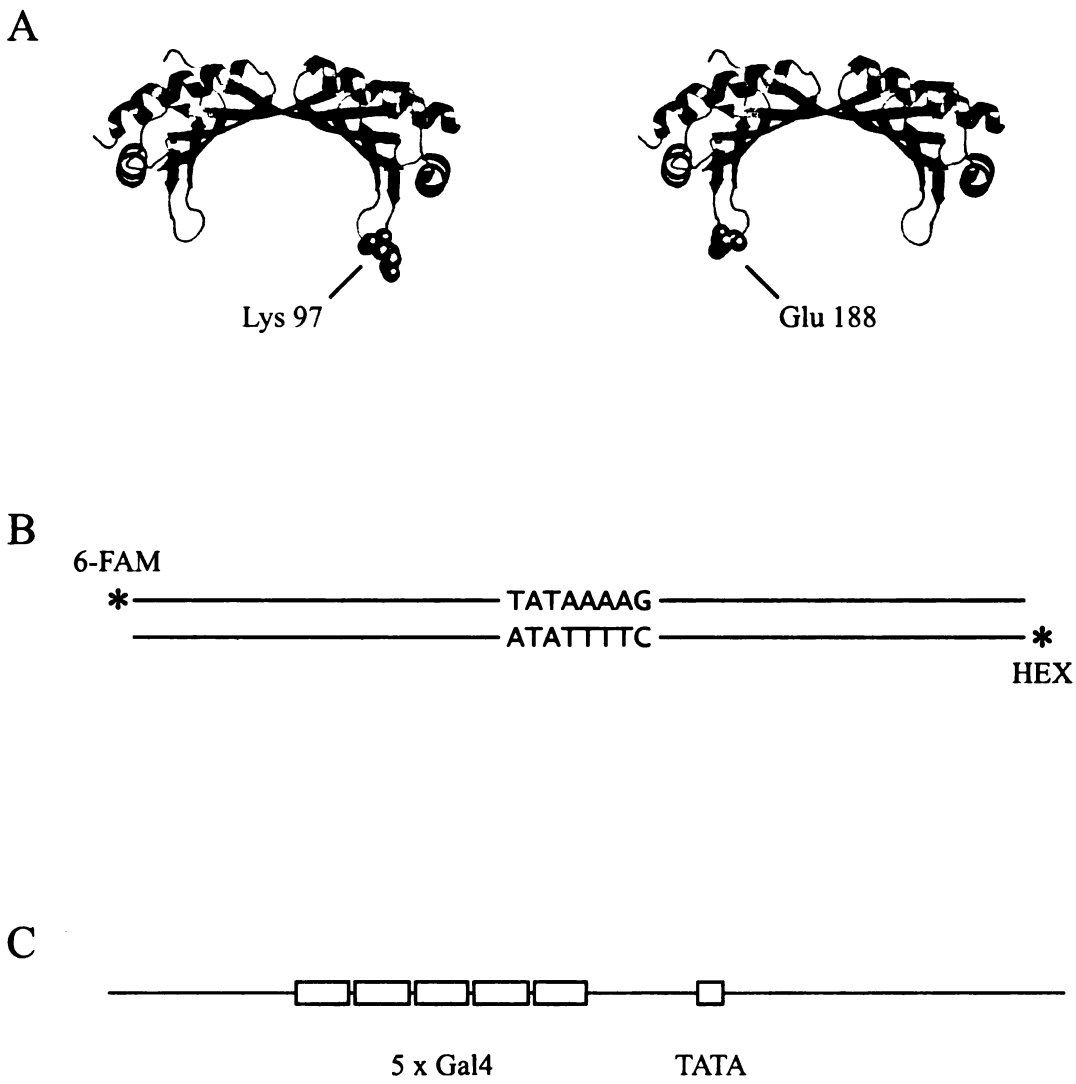


Figure 2: Affinity cleavage methodology. (A) Locations of K97 and E188 shown as spheres mapped onto a ribbon representation of the conserved core of TBP. (B) Location of fluorophores on each strand of the cleavage probe DNA. 6-FAM denotes 6-carboxyfluorescein, and HEX denotes hexachlorofluorescein. (C) Schematic diagram of 304 basepair cleavage probe drawn to scale.

of the cleavage probe DNA, we were able to simultaneously determine the DNA cleavage patterns on both probe DNA strands (Fig. 2B). Since each DNA separation was performed in an individual capillary, the DNA fragment sizes had to be determined relative to an array of in-capillary DNA size standards. To provide for precise fragment sizing, we calibrated a set of generic size standards by comparison to cleavage probe DNA that was digested by restriction enzymes. The restriction enzymes *Bam*HI and *Bst*UI were used to cleave probe DNA at positions 28 bases upstream or 16 bases downstream from the center of the TATA sequence, respectively, providing accurate size standards flanking the region of interest.

To determine the intrinsic orientational preference of TBP binding to the TATA element, K97C-OP and E188C-OP were used to cleave probe DNA encoding the AdMLP TATA element. An analysis of the DNA fragment distribution showed specific cleavage localized to the region of the TATA sequence. The affinity-cleavage patterns were clearly asymmetrical, with more cutting on one side of the TATA sequence. E188C-OP cleaved the top strand of the probe DNA preferentially on the upstream side of the AdMLP TATA sequence (Fig. 3A), while K97C-OP cleaved the top strand preferentially on the downstream side (Fig. 3B). These cleavage preferences match the orientation of TBP that is observed in the crystals structures of all TBP:DNA complexes, where the N-terminal repeat of the TBP core region faces downstream, toward the transcriptional start site. The cleavage patterns on the bottom strand of the probe DNA were consistent with the top strand patterns, again showing that E188C-OP preferentially cleaved upstream from the TATA sequence while K97C-OP cleaved downstream (Figs. 4A, 4B).

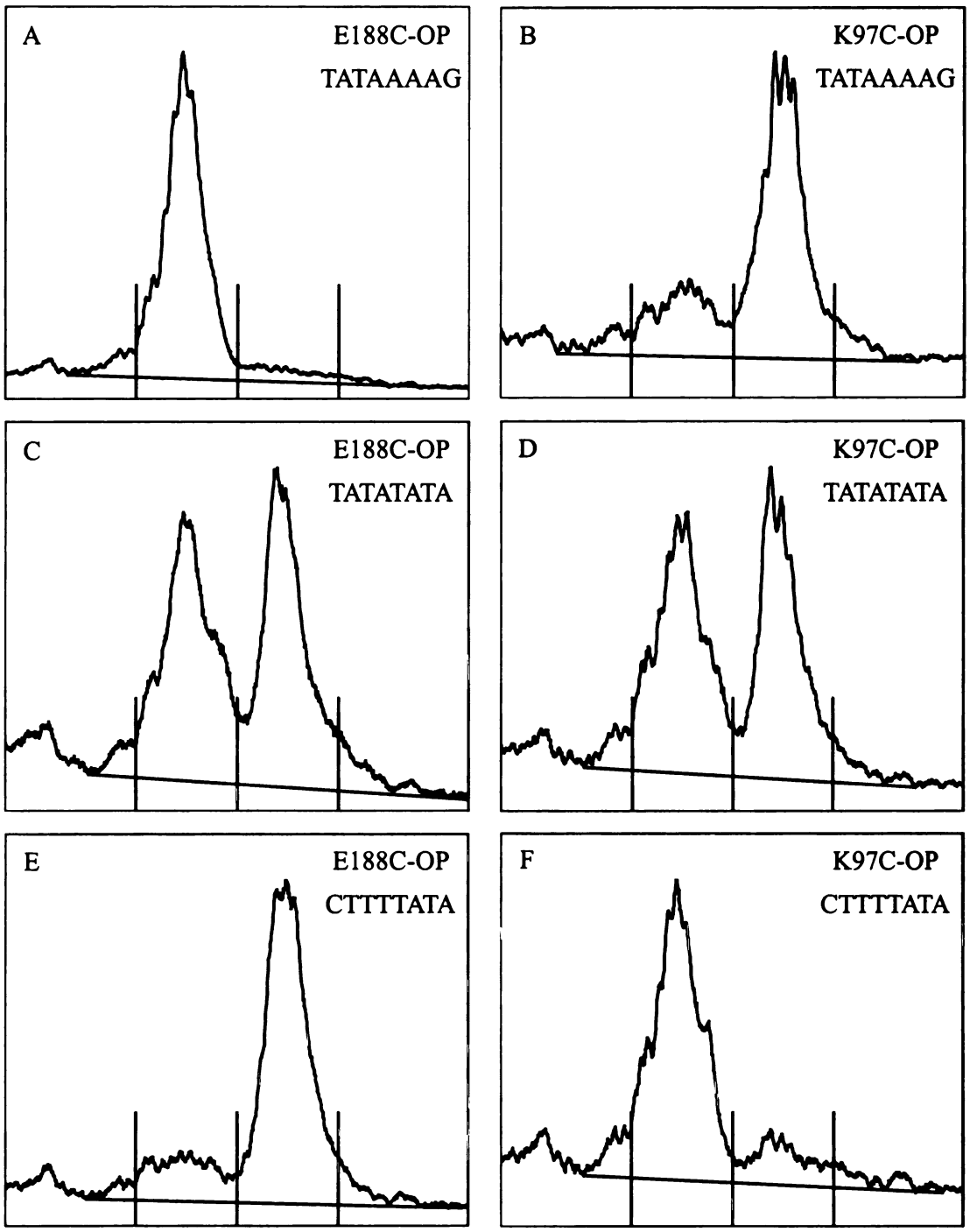


Figure 3: Affinity cleavage patterns on top DNA strand. Representative top strand cleavage data are shown in 60 basepair windows centered on the cleavage patterns. The horizontal axis is DNA position, and the vertical axis is scaled fluorescence units. The sequence of the TATA element is shown on each panel. The vertical bars indicate positions at the center and 13 bases to either side of the cleavage patterns.

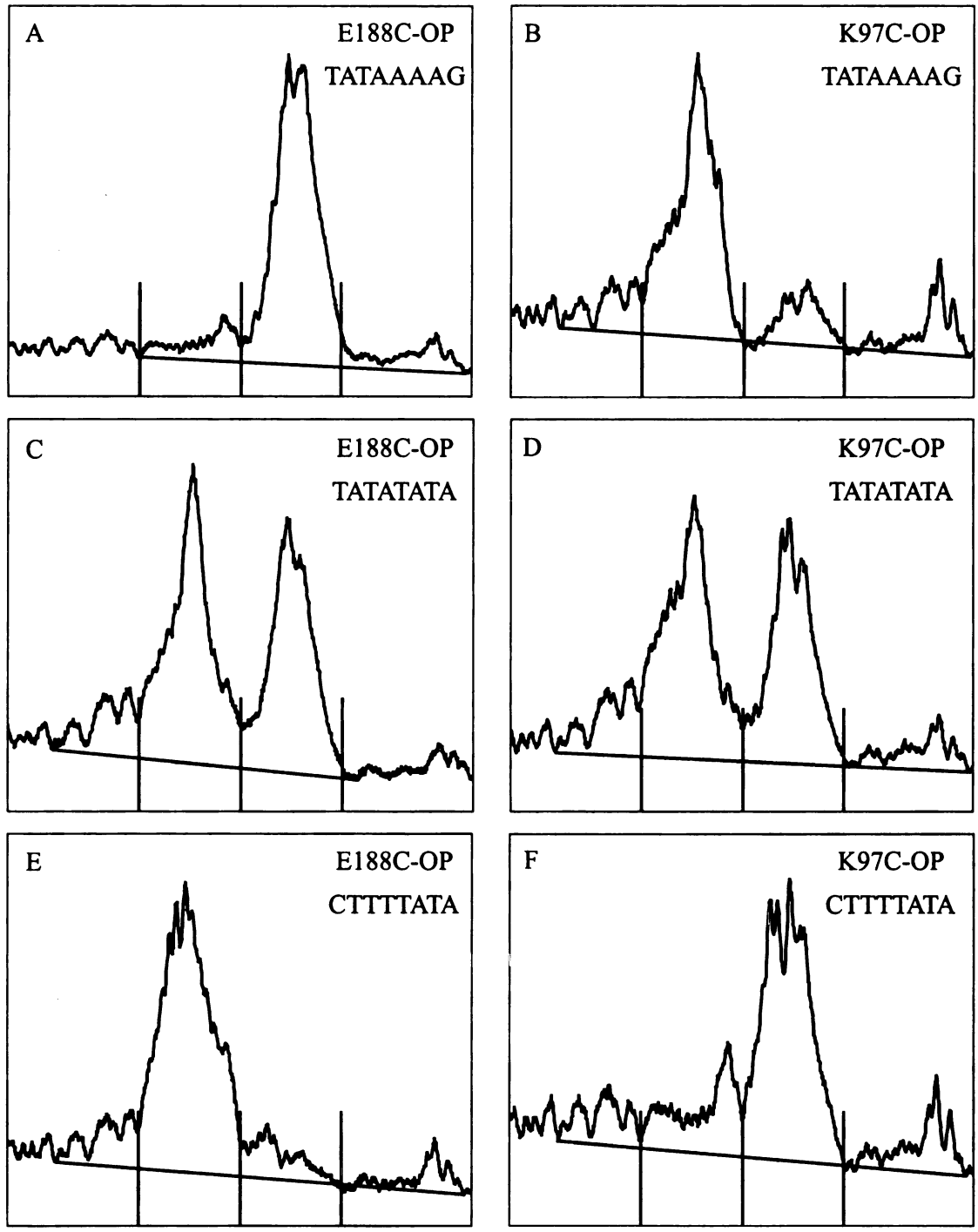


Figure 4: Affinity cleavage patterns on bottom DNA strand. Representative bottom strand cleavage data are shown in 60 basepair windows centered on the cleavage patterns. The horizontal axis is DNA position, and the vertical axis is scaled fluorescence units. The top strand sequence of the TATA element is shown on each panel. The vertical bars indicate positions at the center and 13 bases to either side of the cleavage patterns.

Importantly, the cleavage patterns determined from the top DNA strand and the bottom DNA strand qualitatively agreed. Since the fluorophore label on each strand was located at the 5' position, the agreement between top strand and bottom strand cleavage patterns required that shorter DNA fragments in one strand correspond with longer fragments in the other strand, indicating that the asymmetrical cleavage patterns were not a result of over-cutting of the probe DNA.

The highly oriented binding of TBP to the AdMLP TATA sequence prompted further exploration using variant TATA sequences. The AdMLP TATAAAAG sequence was replaced with the symmetrical TATATATA sequence. TBP-OP affinity cleavage directed by the symmetrical sequence led to equally distributed amounts of cleavage on either side of the TATA sequence (Figs. 3C,D; 4C,D), providing validation of the experimental techniques and indicating that the sequences flanking the AdMLP TATA sequence do not dominantly orient TBP. When the TATA sequence was reversed to CTTTTATA the DNA cleavage patterns were correspondingly reversed (Figs. 3E,F; 4E,F), indicating that the 8 basepair AdMLP TATA sequence is sufficient to direct highly oriented binding of TBP to DNA.

To quantify the degree of TBP orientation, raw fluorescence data were summed in adjacent 13 basepair regions centered on the cleavage patterns. The 13 basepair quantitation regions are sufficiently wide to capture the specific cleavage signals. It should be noted that the cleavage patterns were not precisely centered on the TATA sequence, but were shifted in the 5' direction on each strand (Fig. 5), and thus the centers

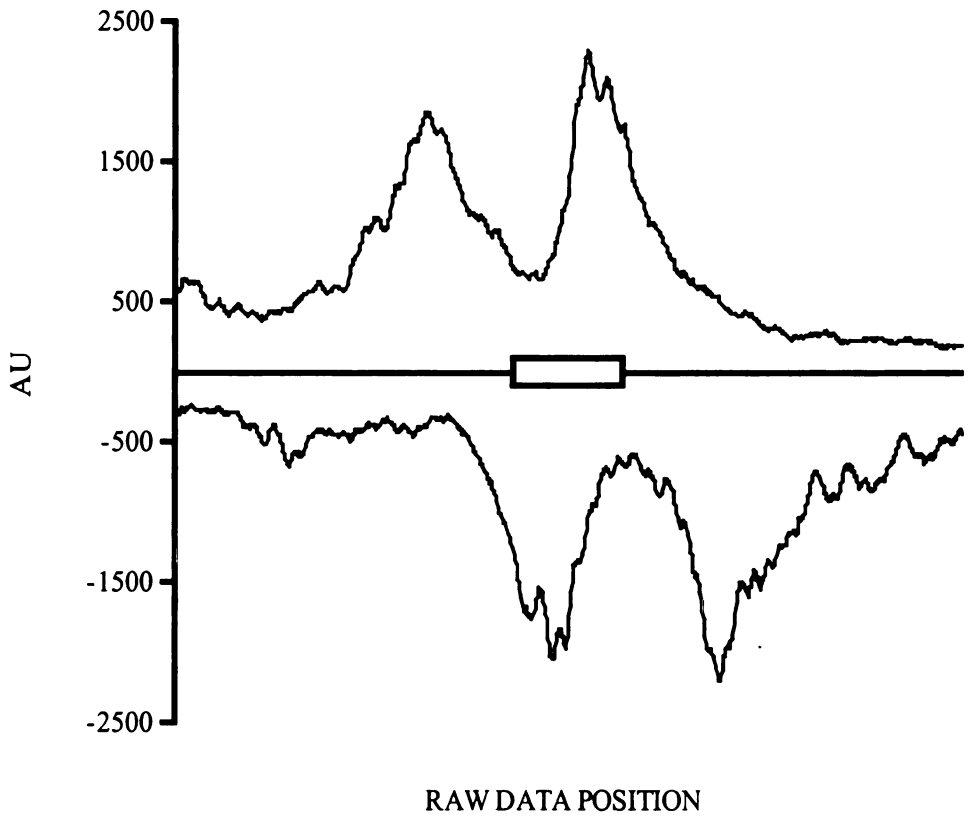


Figure 5: DNA cleavage patterns mapped onto the cleavage probe sequence. Probe DNA encoding a symmetrical TATA sequence was cleaved with K97C-OP, and the resulting capillary electrophoresis data was aligned to the probe DNA sequence. The eight basepair TATATATA sequence is represented by the box on the horizontal axis. top strand data is plotted as positive arbitrary units (AU) and bottom strand data is plotted as negative AU.

of the cleavage patterns were determined by direct inspection. To correct for background DNA cleavage, a signal baseline was approximated by a straight line connecting fluorescence minima in regions upstream and downstream from the quantitated regions. TBP orientation was defined as the baseline-corrected signal in one quantitation region divided by the total baseline-corrected signal (Fig. 6). Orientation comparisons were performed using top strand data, due to the occasional presence of an anomalous signal in the bottom strand cleavage pattern. The source of this anomalous signal has not been determined. Indeed, in side-by-side affinity cleavage reactions, built from common stocks and processed in parallel, the anomalous signal was prominent in one sample but absent in the other (Fig. 7), precluding any obvious correlation with the affinity cleavage experimental procedures.

Using a quantitation of top strand cleavage patterns, E188C-OP was found to be 90% oriented on the forward TATA sequence, but only 78% oriented on the reversed TATA sequence (Table 1). This difference may indicate over-digestion of the probe DNA, since a bias toward shorter top strand fragments would increase the apparent orientation of E188C-OP on the forward TATA sequence and correspondingly reduce the apparent orientation on the reversed TATA sequence. A similar shift in the K97C-OP orientation measures can be explained by a bias toward shorter fragments. Additionally, the cleavage patterns from a TBP mutant (F227I, discussed below) are also consistent with a bias toward shorter DNA fragments. If the cleavage reactions have a slight bias toward shorter DNA fragments, then actual measure of TBP orientation will lie between the positively and negatively shifted orientation measures. With this reasoning, the degree of

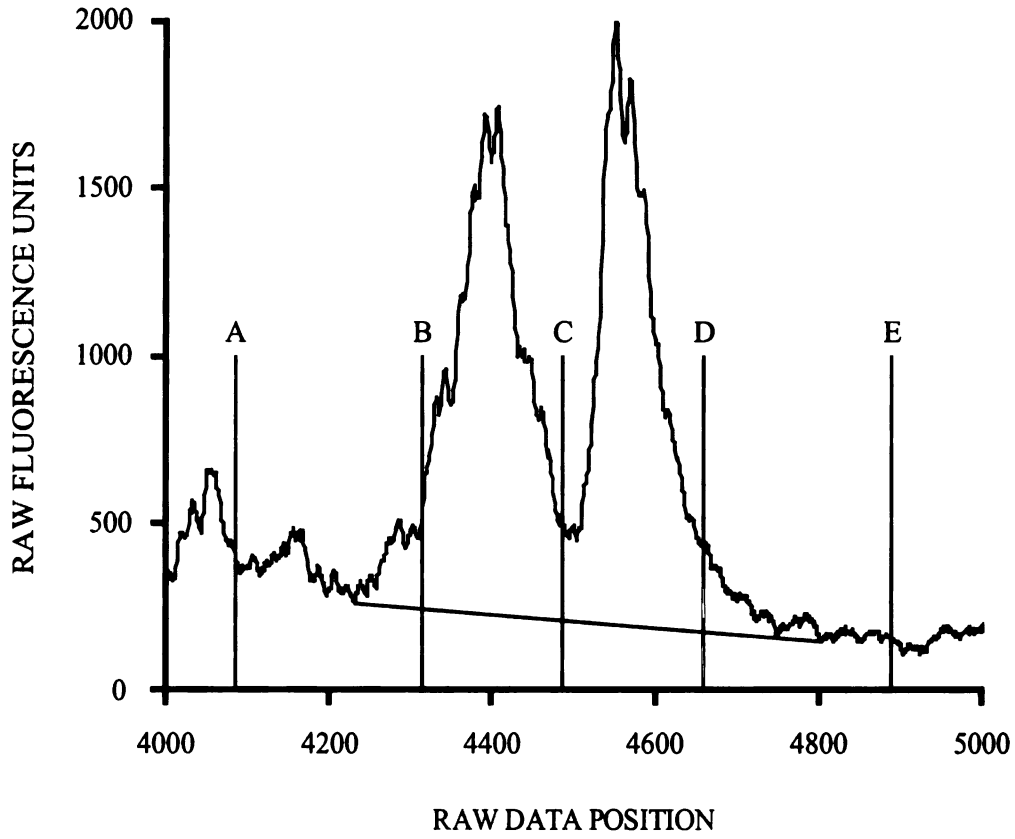


Figure 6: Method of DNA fragment pattern quantitation. Raw fluorescence units were summed in 13 bp regions B-C and C-D. A signal baseline (slanted line) was approximated by a straight line connecting minima found in 17 bp regions A-B and D-E. The area below the baseline was subtracted from the raw data sums from regions B-C and C-D. Orientation was calculated as the ratio of the baseline-corrected signal in B-C or C-D divided by the total baseline corrected signal in B-D.

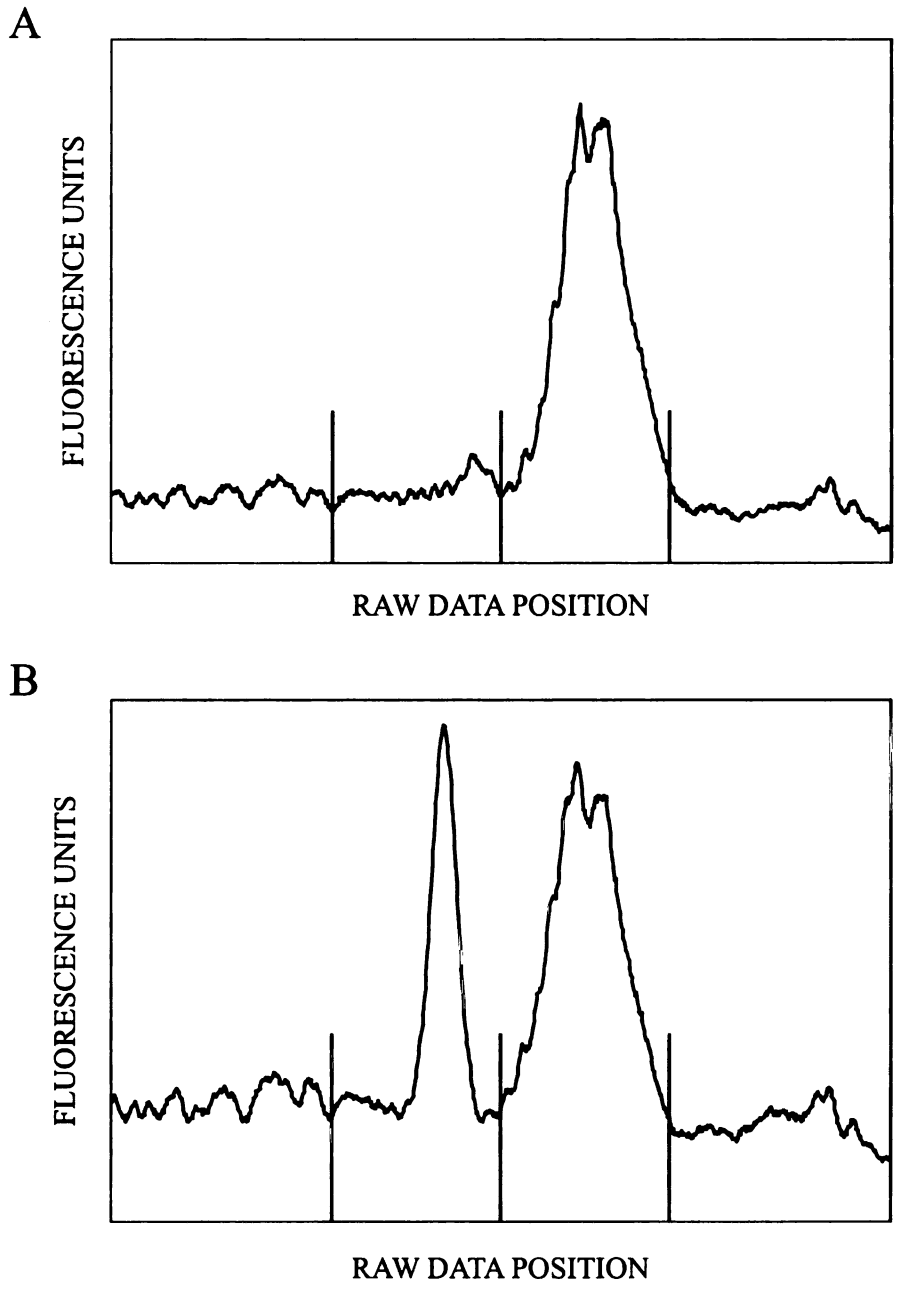


Figure 7: Anomalous peak in bottom strand DNA fragment analysis. Vertical bars indicate 13 base quantitation regions centered on the cleavage patterns. Cleavage reactions (A) and (B) were built from common stocks and performed in parallel. In (B), an extra peak is present in the first quantitation region.

Table 1: Affinity cleavage measurements.

TBP orientation measurements					
TBP version	**TATA version**	**Top strand orientation** [a]	**Samples**	**Standard deviation**	**Calculated orientation** [b]
K97C-OP	Forward	71%	10	2.4%	75%
	Reverse	79%	3	4.3%	
	Symmetric	50%	3	1.6%	
E188C-OP	Forward	90%	10	2.2%	84%
	Reverse	78%	3	3.6%	
	Symmetric	50%	3	0.9%	
K97C-OP (F227I)	Forward	67%	7	3.5%	69%
	Reverse	71%	3	0.7%	
	Symmetric	47%	3	0.2%	
E188C-OP (F227I)	Forward	85%	6	3.5%	81%
	Reverse	77%	4	2.9%	
	Symmetric	53%	3	1.2%	

[a] Orientation relative to the direction of the TATA sequence. For symmetrical TATA sequences, orientation is relative to TATA flanking regions.

[b] Orientation calculated as the average of orientation measures obtained from forward and reversed TATA sequences.

TBP orientation on the AdMLP TATA sequence has been defined as the average of the top-strand orientation measures from forward and reversed TATA sequences. Accordingly, E188C-OP has been determined to be 84% oriented on the AdMLP TATA sequence, and K97C-OP has been determined to be 75% oriented on the AdMLP TATA sequence (Table 1). Our measures of TBP orientation are significantly different from the values of 60%-64% (26) and 51% (67) reported by the Schepartz lab.

Since our measurements indicate that TBP can bind the TATA element with a high degree of orientational specificity, there would appear to be little opportunity for TBP orientational enhancement by Gal4-VP16. Nonetheless, the report that Gal4-VP16 can significantly increase the orientational specificity of TBP orientation (67) led us to replicate these reactions. In these experiments, Gal4-VP16 was bound to the probe DNA prior to addition of TBP-OP. To ensure Gal4-VP16 association with the cleavage probe DNA, a single preparation of DNA, TBP, and Gal4-VP16 was split into separate affinity cleavage and footprinting reactions. DNA footprinting with methidium-propyl EDTA-iron clearly showed protection of the Gal4 binding sites, indicating the presence of Gal4-VP16 on the probe DNA (Fig. 8A). An analysis of affinity-cleaved probe DNA indicated reduced amounts of cleavage at the TATA sequence, but the ratio of upstream and downstream cutting was not significantly altered (Fig. 8B). These results indicated that Gal4-VP16 does not alter the orientation of TBP binding to the AdMLP TATA sequence.

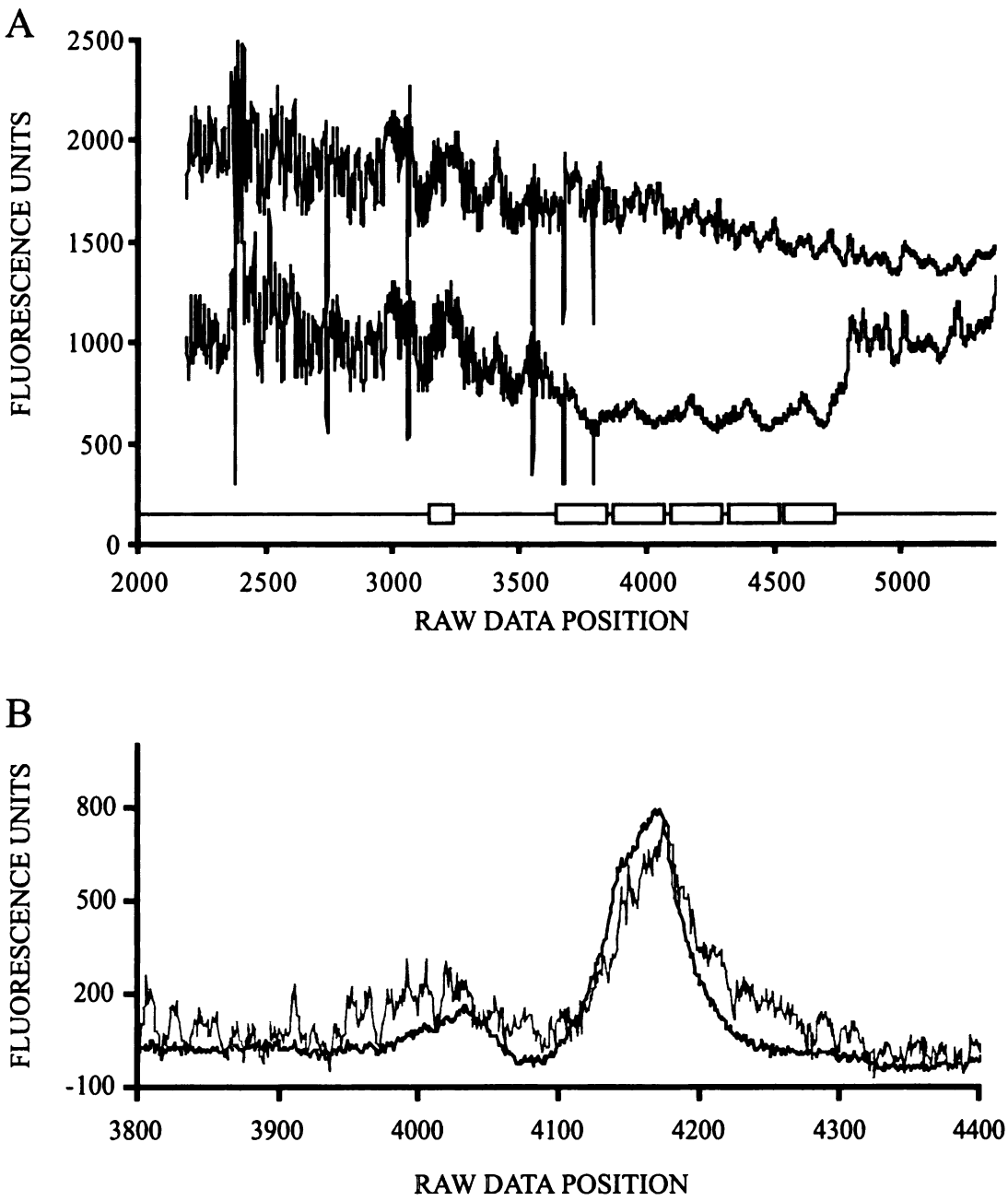


Figure 8: Affinity cleavage pattern in the presence of Gal4-VP16. (A) Footprinting of cleavage probe bound to Gal4-VP16. Bottom strand data is shown. Top trace, absence of Gal4-VP16. Bottom trace, presence of Gal4-VP16. The location of the Gal4 binding sites is indicated by the cluster of five boxes on the graph. The smaller, isolated box indicates the position of the TATA sequence. (B) Affinity cleavage patterns in the absence (black line) and presence (grey line) of Gal4-VP16. The cleavage data from a reaction including Gal4-VP16 was scaled by a factor of three and overlayed with data from a reaction lacking Gal4-VP16.

In an extension of our use of the affinity-cleavage method, we have participated in a collaborative effort to study a TBP mutation that may lead to altered orientational specificity. The laboratory of Dr. Karen Arndt has performed a yeast genetic screen that selected for TBP variants supporting elevated levels of forward transcription from a reversed TATA sequence in vivo. This screen identified the TBP mutant F227I. Three simple models can be put forth to explain the ability of F227I to exceed the ability of wild type TBP at directing forward transcription from a reversed TATA sequence. First, the F227I mutation may lead to an increased affinity for DNA, but not to changes in orientational specificity. In this case, the amount of TBP facing forward on the reversed TATA sequence would be increased simply due to more TBP occupancy at the promoter. Second, the DNA affinity of F227I may be unaffected, but orientational specificity may be relaxed or reversed. In this case the fraction of TBP facing forward would be increased, again allowing increased forward transcription from a reversed TATA sequence. Third, F227I may not affect either DNA affinity or orientational specificity, suggesting that the effects of F227I are mediated by altered interactions with other components of the transcriptional machinery. The Arndt lab has determined that F227I has a normal overall affinity for DNA, implying either that the orientational specificity of F227I is altered or that interactions between F227I and other factors are altered.

To further explore the properties of the F227I TBP variant, we have used the affinity-cleavage method to assess the orientational specificity of F227I bound to the AdMLP TATA sequence. Affinity-cleavage reactions were performed using K97C-OP (F227I) and E188C-OP (F227I) as the cleavage agents. The resulting DNA cleavage patterns

were quite similar to those from the wild type versions of TBP-OP, indicating a similar high degree of orientational specificity. However, a careful comparison of the DNA cleavage data showed that the F227I mutation conferred a slightly relaxed orientational preference. This relaxed orientation was most visible in an overlay of top-strand affinity cleavage results from wild-type and F227I TBP-OP (Fig. 9). Several repetitions of the affinity cleavage reactions have indicated that the difference, although small, is reproducible. A quantitation of the wild type and F227I cleavage patterns indicated that F227I reduced the orientational specificity of E188C-OP bound to the AdMLP TATA sequence from 84% to 81% (Table 1). A Student's t-test of the orientation measures derived from forward TATA sequences indicated that the two sets of orientation measures did not likely overlap ($p<0.01$). Similarly, the F227I mutation reduced the forward orientation of K97C-OP from 75% to 69% and again the measurements from the forward TATA sequence did not likely overlap ($p<0.05$).

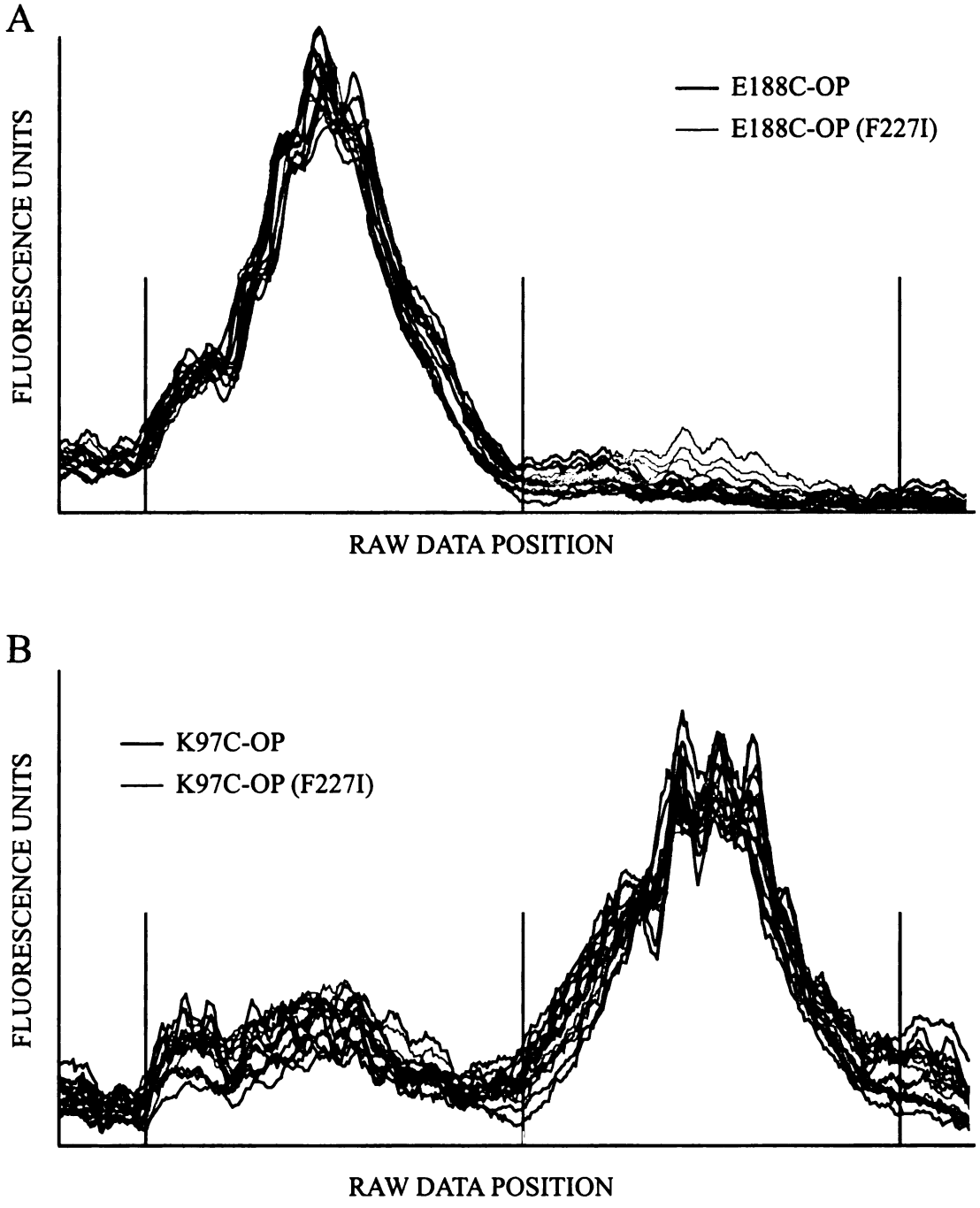


Figure 9: Overlay of TBP-OP and TBP-OP (F227I) top strand cleavage patterns. TBP-OP and TBP-OP (F227I) were used to cleave a forward-oriented AdMLP TATA sequence. DNA cleavage patterns from several independent experiments were normalized to an equal baseline-corrected area in the predominant cleavage region, and the resulting plots were overlayed. Vertical bars indicate the upstream and downstream cleavage regions. (A) E188C-OP top strand cleavage patterns. (B) K97C-OP top strand cleavage patterns.

Discussion

The previous report of a TBP-orienting activity residing in the VP16 activation domain raised questions about the mechanistic details and potential relevance of this novel mechanism of activation. For instance, given a panel of activation domain mutants with a variable degree of functionality, how well would the in vitro orienting activity correlate with in the in vivo transcriptional activation function? And how could the implicit ternary character in the TBP orientation function be reconciled with other evidence indicating that VP16AD, TBP and DNA could not form a ternary complex? To approach these questions we have replicated and extended the affinity cleavage reactions initially developed by the Schepartz lab. Unexpectedly, and contrary to the report from the Schepartz lab, we found that isolated TBP bound to the AdMLP TATA sequence with a high degree of orientational specificity. A review of the published data from the Schepartz lab suggests that the precedent measures of TBP orientation may have been affected by a significant and variable background contribution to the cleavage patterns, leading to an incorrect conclusion that TBP is only modestly oriented on the AdMLP TATA sequence. The clear signals and internal consistency in our affinity cleavage data suggests that TBP in fact binds to the AdMLP TATA sequence with a high degree of orientational specificity.

Although the affinity cleavage patterns produced by K97C-OP and E188C-OP qualitatively agree, these two versions of TBP-OP appear oriented to different degrees. This disagreement raises the possibility that neither version of TBP may provide an

accurate measure of TBP orientation. Since the stirrup loops lie in the minor groove of DNA, the cysteine mutation and OP derivitization may interfere with DNA binding or the intrinsic directionality of TBP. Indeed, other mutations in the stirrup loops have been shown to alter the DNA binding properties of TBP. For instance, in the N-terminal stirrup loop of TBP, an A100P mutation conferred increased DNA affinity to TBP, and also lent TBP the ability to recognize a reversed TATA sequence in vivo (157). A substitution in the symmetrically related position in the C-terminal stirrup loop (P191A) also altered DNA recognition, but in this case DNA binding was reduced (157). Additionally, mutations of L189 and F190 abrogated DNA binding, although position E188 tolerated mutations without a loss of TBP:DNA binding ability (131). These results indicate that TBP is sensitive to mutation at positions located only 1 to 3 residues away from both the K97C and the E188C. Since the current affinity cleavage methods attach a bulky (~300 Da) moiety to a cysteine substitution flanking critical residues, it is possible that the resulting TBP-OP proteins suffer from defects in DNA association. Nonetheless, both K97C-OP and E188C-OP produce highly asymmetric and qualitatively similar DNA cleavage patterns on the forward and reversed TATA sequences, suggesting that any TBP-OP defects are modest, and that TBP binds to the AdMLP TATA sequence with a high degree of orientational specificity.

Our studies also showed that the transcriptional activator Gal4-VP16 did not enhance the orientational specificity of K97C-OP bound to the AdMLP TATA sequence. This result is in disagreement with the precedent report from the Schepartz lab (67). In efforts to reconcile these different results, the gel images from the precedent study were examined.

From sequencing gel images (figure 2 in (67)), it is apparent that a significant background signal existed in some affinity-cleavage reactions and not others. This background signal, when present, would dilute the specific affinity-cleavage bands, thus reducing the apparent orientation of TBP. Since the background signal was reduced when Gal4-VP16 was present, the apparent orientation of TBP increased upon addition of Gal4-VP16. A visual correction for the background signal in the published gel images suggests that the orientation of TBP is unchanged in the presence of Gal4-VP16. Thus the conclusion from the Schepartz lab that Gal4-VP16 increased the orientation of TBP may have arisen from a significant and variable background contribution to the specific signals.

While our findings indicate that enhancement of TBP orientation is not a function of the VP16 activation domain, the strong correlation of in vivo VP16AD activity with in vitro VP16AD:TBP affinity suggests that TBP is nonetheless a relevant target of VP16AD (120). Other lines of evidence show that VP16AD competes with other factors for binding to the DNA-binding region of TBP. These competitive interactions, while precluding the ternary interaction required for altering TBP:DNA orientation, might allow for activation through a multi-step handoff mechanism (reviewed in Chapter 2).

Overall, the F227I mutation appears to have an effect on TBP orientational specificity, reducing the forward oriented population of TBP by about 4-5%. Although this is a small difference, it was detected when the cleavage activity localized to either stirrup of TBP, and also when using either forward or reversed TATA sequences. Additionally, enough

data was acquired from forward TATA cleavage reactions to show that the effects of F227I on orientational specificity were statistically significant. The modest reduction in F227I orientational specificity raises the question of whether such a small change could lead to the transcriptional differences detected in the in vivo assays performed by the Arndt laboratory. In this respect, an accurate measure of the absolute degree of TBP orientation would allow the determination of the relative change in forward-facing TBP bound to a reversed TATA sequence. For instance, if TBP naturally binds to the AdMLP TATA sequence with 84% forward orientation, then the F227I mutation would be expected to increase the forward-facing TBP on the reversed TATA sequence from 16% to 21%, which may account for the increased transcription from directed by F227I. However, a limitation in this analysis is that we cannot precisely determine the absolute degree of TBP orientation using the current affinity cleavage methodology, and thus we cannot determine the relative change in TBP orientation conferred by the F227I mutation.

Our finding that TBP is highly oriented on the AdMLP TATA sequence has implications for models of PIC assembly. The current view of PIC assembly (reviewed in (50, 156)) assumes that TBP is essentially randomly oriented on the TATA element and that other factors, such as TFIIB, are required to direct the orientation of TBP on a promoter. This model has arisen primarily from two different lines of evidence. First, studies of the relatively simple but homologous transcription apparatus of Archaea found that the highly symmetric archaeal TBP is incapable of establishing transcriptional directionality, and that the TFIIB-related factor TFB is required to define the orientation of transcription (7, 82, 98). Second, the results from the Schepartz lab suggested that eukaryal TBP, like

archaeal TBP, is also incapable of directional recognition of the TATA sequence, and that other factors are required to establish a productive orientation of the eukaryal TBP:DNA complex (26, 67, 68). In the PIC assembly model arising from these observations, the orientation of the TATA sequence is assumed to be a modest factor in overall promoter architecture. Our results, contrary to the report from the Schepartz lab, show that the eukaryal TBP can recognize the TATA sequence in a highly directional manner, suggesting that the archaeal and eukaryal TBP homologs may be functionally divergent, and that the orientation of the TATA sequence in eukaryal promoters may be a more significant variable in promoter architecture than currently believed.

Chapter IV

Attempts to determine the structure of VP16C bound to TBP

Introduction

The transcription of eukaryotic protein-coding genes can be stimulated by cis-acting regulatory sequences that may be located from tens to thousands of bases away from the core promoter region and the transcriptional start site. The link between cis-acting regulatory sequences and transcription is made by trans-acting proteins known as transcriptional activators. The dual functions of transcriptional activators, DNA recognition and transcriptional activation, typically comprise distinct domains of the protein structure. While the principles of DNA recognition by transactivators are well characterized, the mechanisms and targets of activation domains remain elusive.

Studies of transcriptional activation domain function frequently employ the unusually potent activation domain of the VP16 transactivator from the herpes simplex virus as a model system (146). The VP16 activation domain (VP16AD) resides in the C-terminus of the VP16 protein (residues 411-490) (49, 164, 174), and can be further divided into independently acting subdomains VP16N (residues 412 to 456) and VP16C (residues 457-490) (139, 160, 169). NMR analyses of the isolated VP16AD have indicated that it exists in an unstructured state in solution (61, 126), although it can assume a helical conformation under certain conditions (30) and when bound to TAF32 (166) or PC4 (61). The VP16AD has a preponderance of acidic residues, suggesting that the VP16AD may

exist as an "acidic blob" (154), enabling nonspecific interactions with a variety of targets. Thorough mutational analyses of the activation domain have shown, however, that the acidic character of the VP16AD is not critically required for activity. Mutation of one or a few acidic residues to uncharged residues had only a modest effect on in vivo function, whereas bulky hydrophobic residues in a few key positions are critically required for activity (28, 139, 160). The requirement for bulky hydrophobic residues in key positions suggests that a hydrophobic core and tertiary structural elements play a role in VP16AD structure. Whether these bulky hydrophobic residues are involved internally to a VP16AD fold or in interfaces with target proteins remains unknown.

Various studies have identified a wide range of proteins that interact either directly or indirectly with VP16AD. Among these are several of the generally-required basal transcription factors, including TFIIA (76, 77), TFIIB (43, 96, 169), both the TBP (56, 153, 159) and TAF9 (75) subunits of TFIID, and TFIIH (179). Additionally, components of the RNA polymerase II associated mediator complex have been shown to interact with VP16AD (92, 114, 182). Associations with chromatin-modifying coactivators, which alter chromatin structure either covalently or noncovalently, have also been reported. However, in this panoply of possible VP16AD interactions, little evidence establishes which particular interactions are functionally relevant in the complicated process of in vivo transcriptional regulation.

TBP stands out as likely being a relevant target of the VP16AD. TBP, a component of the multisubunit TFIID complex, is a central player in all classes of eukaryotic

transcription, and is thought to mediate the initial protein-DNA contacts in the assembly of the large pre-initiation complex (PIC) that poises RNA polymerase II at the transcriptional start site. A quantitative assessment of TBP interaction with a spectrum of VP16C mutants revealed tight binding of TBP to wildtype VP16C, and a robust correlation between in vitro binding affinity and in vivo function, consistent with the hypothesis that TBP interaction is important for VP16C function (120). Additionally, time-resolved fluorescence anisotropy techniques have shown that critical hydrophobic residues in both VP16 activation subdomains become structurally constrained and protected from solvent when bound to TBP, indicating an acquisition of structural elements in the otherwise unstructured VP16AD (152, 153).

While TBP appears to be a relevant target of the VP16AD, the detailed activation mechanism probably involves more than a simple recruitment of TBP to a promoter. Acidic activation domains in general do not bind cooperatively with TBP at promoter DNA sequences (94, 180, 181). Additionally, a mutation in the concave DNA binding region of TBP (L114K) abrogates interactions with VP16AD (72), indicating that the activator and DNA interact with overlapping regions of TBP. However, despite this competition of acidic activators and DNA for binding to TBP, acidic activators can nonetheless stimulate formation of the TBP-containing TFIID-TFIIA-DNA complex (76, 77). Notably, in the TFIID complex, the intrinsic ability of TBP to bind DNA is inhibited. Two inhibitory activities have been mapped to N-terminal domains of the TFIID component TAF1 (80). These domains, TAND1 and TAND2, have been shown to bind TBP in competition with DNA (80, 84, 99) and TFIIA (3, 79), respectively. The

observation that VP16C binds TBP competitively with the negative regulator TAND1 (124) suggests a cascade of interactions with the concave surface of TBP, with TAND1 being displaced by VP16C, and VP16C subsequently displaced by DNA. In this "handoff model" (83) (Fig. 10), an activator functions as an anti-repressor when interacting with TBP in the context of TFIID. Importantly, the ability of VP16C to stimulate TFIID-TFIIA-DNA assembly in vitro correlates with the ability of VP16C to activate transcription in vivo (77), suggesting that TFIID-TFIIA-DNA assembly is a relevant mode of transcriptional activation in vivo. Furthermore, in support of the model that competition with TAND1 can activate transcription, TAND1 itself functions as an activator in vivo when fused to a DNA binding domain (83).

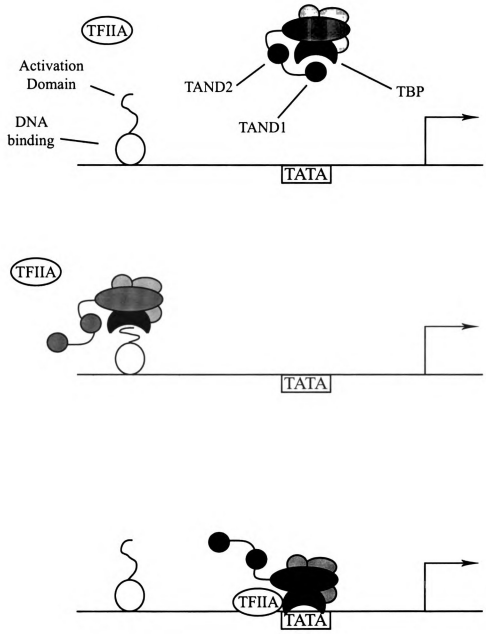


Figure 10: Handoff mechanism of TFIID-TFIIA-DNA assembly. In the upper panel TBP is embedded in the TFIID complex, with TAND1 competing with DNA binding and TAND2 competing with TFIIA binding. In the middle panel, the VP16C has evicted TAND1, which may facilitate TFIIA competition with TAND2 for binding TBP. In the bottom panel, DNA has replaced VP16C, and TFIIA and TFIID are bound to the core promoter.

VP16 complexed with TBP

Rationale

Our current understanding of activation domain structure/function relationships is limited by an absence of structural data. The evidence that intrinsically unstructured activation domains acquire structural features upon binding to target proteins suggests that complexes of activators and targets represent the best candidates for structural analysis. The plausibility of the handoff model and the evidence for structural elements in VP16C when bound to TBP suggests that the VP16C:TBP protein complex represents a structured intermediate in a relevant mechanism of transcriptional activation. A high resolution structural model of this complex would clearly localize the VP16C binding site on TBP, providing insights into the handoff mechanism. It would provide a basis for the prediction and analysis of activation domain mutations, and might guide a precise dissection of the interface of VP16C and TBP. An activation domain structure allows the definition of a structure-based activator motif, and also enables a computational docking approach to search for other targets among known protein structures. Taken together, the wealth of information available from a high resolution structural model warranted our attempts to study the structure of the VP16C:TBP complex by x-ray crystallography.

Materials and methods

Purification of VP16C. The expression vector pGEX.VP16C (119), encoding a fusion of GST, a thrombin cleavable linker, and VP16 residues 452-490 under control of the *tac* promoter, was transformed into BL21(DE3)Codon-plus *E. coli* cells. Cells containing the expression plasmid were grown in LB medium containing 80 mg/L ampicillin and 40 mg/L chloramphenicol. One L cultures were grown at 37 °C with vigorous shaking until reaching an OD_{600} of 0.6-0.9. Expression of the GST-VP16C fusion protein was induced by addition of IPTG to a final concentration of 0.2 mM and growth was continued for an additional 3 h. Cells were harvested by centrifugation and resuspended in 10 mL HEMGT buffer (24 mM HEPES, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% v/v glycerol, 0.1% v/v TWEEN-20, pH 7.9) with 250 mM KCl. Resuspended cells were lysed by three passages through a French Press (Aminco) operating at 20,000 psi. Cellular lysates were centrifuged at 10,000 x g at 4 °C for 10 min to sediment insoluble matter. The soluble lysate fraction was diluted 10-fold with HEMGT-250 containing 5 mM DTT, mixed with 1 mL GSH resin, and gently rotated at 4 °C for 3 h. The resin was collected and washed with two 5 mL aliquots of HEMGT-250, two 5 mL aliquots of HEMGT-100, and two 5 mL aliquots of thrombin digestion buffer (20 mM Tris-Cl, 150 mM NaCl, 2.5 mM CaCl2, pH 8.4). The resin was resuspended in 2 mL thrombin digestion buffer and biotinylated thrombin was added to liberate VP16C from the resin-bound fusion protein. Digestion was performed at 4 °C for 4 h with gentle rotation. The supernatant and a 1 mL wash were collected and incubated with 0.2 mL streptavidin agarose beads to trap the

biotinylated thrombin. The supernatant and a 0.5 mL wash were collected and stored at -80 °C.

The VP16C peptide was further purified by anion-exchange HPLC. VP16C solutions were diluted tenfold with 20 mM Tris buffer (pH 8.5) to lower the salt concentration and bound to a DEAE ion-exchange HPLC column (TSK DEAE-5PW, 2.15 x 15 cm, Beckman). The column was developed with a 275 mL linear gradient from 0 to 750 mM NaCl in 20 mM Tris pH 8.5 at a 5 mL/min flow rate. VP16C typically eluted at about 435 mM NaCl. VP16C concentration was estimated by absorbance at 280 nm and by a bicinchoninic acid assay (BCA assay, Pierce) relative to BSA standards.

Purification of yeast TBP. The plasmid trc-TBP, encoding the fusion of a histidine tag, a thrombin-cleavable linker, and *Saccharomyces cerevisiae* TBP residues 61-240 under control of a T7 promoter, was obtained from Dr. J.H. Geiger. This plasmid was transformed into *E. coli* strain BL21(DE3)Codon-plus. Cells containing the plasmid were grown in LB medium containing 80 mg/L ampicillin and 40 mg/L chloramphenicol. One liter cultures were grown at 37 °C with vigorous shaking to an OD_{600} of about 1.0. Expression of TBP was induced by addition of IPTG to a final concentration of 0.15 mM, cultures were transferred to 18 °C (ambient), and growth was continued for 12-14 h with vigorous shaking. Cells were harvested by centrifugation and stored at -80 °C.

Cell pellets from three liters of culture were resuspended in 50 mL lysis buffer (50 mM sodium phosphate, 20% v/v glycerol, 500 mM NaCl, 1 mM EDTA, 5 mM DTT, 0.1%

Triton X-100, 5 mM imidazole, pH 7.5). The resuspended cells were lysed by 2 passages through a French Press (Aminco) operating at 20,000 psi. Cellular lysates were centrifuged at 45,000 x g at 4 °C for 25 min to sediment insoluble matter. The soluble lysate fraction was collected and mixed with 2.5 mL of nickel-affinity resin (Ni-NTA Superflow, Qiagen) and gently rotated at 4 °C for 1.5 hours. The resin was collected, washed 3 times with 10 mL lysis buffer with no imidazole, and 4 times with lysis buffer containing 40 mM imidazole. Bound proteins were eluted with lysis buffer containing 150-200 mM imidazole and collected in 1 mL fractions. TBP typically eluted in the first 3 elution fractions.

TBP was further purified by ion-exchange FPLC. TBP solutions were mixed with a diluent (20 mM Tris, 10% glycerol, 1 mM DTT, pH 7.5) to lower the salt concentration to 150 mM. Diluted TBP was passed through a 1 mL Q-Fast Flow anion-exchange column (Pharmacia) and loaded onto a 1 mL SP-Fast Flow cation-exchange column (Pharmacia). After loading, the Q column was disconnected, and the SP column was developed with a linear gradient from buffer A (20 mM Tris-Cl, 10% glycerol, 150 mM NaCl, 1 mM DTT, pH 7.5) to buffer B (Buffer A with 1 M NaCl) over 20 column volumes at 1 mL/min flow. TBP typically eluted at around 540 mM NaCl.

Protein crystallization. VP16C and TBP were mixed at a 1.4 to 1 molar ratio, and the mixture was buffer swapped to crystallization buffer (10 mM Tris-Cl, 10% v/v glycerol, 250 mM NaCl, 2 mM DTT, pH 8) and concentrated to 4 mg/mL in a YM-3 centrifugal concentrator (Amicon).

Results and discussion

The VP16C activation domain (VP16 residues 452-490) was expressed as a GST fusion protein, purified over GSH sepharose, cleaved from GST by thrombin digestion, and subsequently HPLC purified over anion exchange resin (Fig. 11A) (119). The yield of purified protein was approximately 1 mg per liter of initial culture. The conserved core region of yeast TBP (residues 61-240) was expressed as a histidine tagged fusion protein, and purified using nickel affinity chromatography (Fig. 11B,C). Notably, the expressed protein was mostly insoluble, but nonetheless about 1 mg of purified TBP could be recovered per liter of initial bacterial culture. The purity of the both proteins was judged as acceptable for attempts at protein crystallization.

Purified VP16C and TBP were mixed at a 1.4:1 molar ratio and concentrated to 4 mg/mL total protein, as assayed by a Bradford dye binding assay. Because low salt conditions lead to TBP instability, a protein buffer containing 250 mM NaCl was chosen. Concentrated protein solutions were mixed with equal volumes of a diverse set of precipitating mixtures, and 2 μL drops were deposited on plastic cover slips, inverted, and sealed over the corresponding precipitating mixtures in a standard hanging-drop vapor diffusion crystallization configuration. Plates were incubated at room temperature and 4 °C and observed for crystal growth.

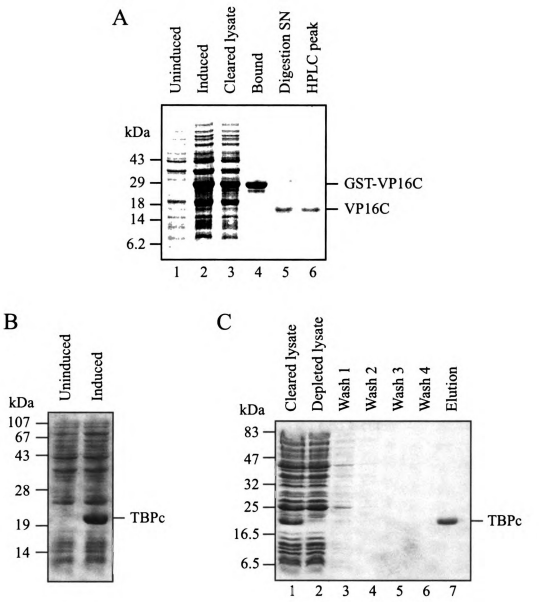


Figure 11: Purification of VP16C and TBP. Protein samples were resolved by SDS-PAGE and stained with Coomassie R-250. (A) Purification of VP16C. Lanes 1-3, cellular extracts. Lane 4, proteins retained on GSH resin. Lane 5, products of thrombin digestion. Lane 6, thrombin-liberated protein after HPLC purification. (B) Expression of the 6xHis-tagged core region of yeast TBP (yTBPc) in *E. coli*, monitored by whole-cell extracts. (C) Nickel-affinity purification of 6xHis-yTBPc. Lanes 1-2, cleared lysate before and after incubation with nickel resin. Lanes 3-6, proteins released by washing with 40 mM imidazole. Lane 7, protein released by 200 mM imidazole.

Short-lived needle-shaped protein crystals were observed in a precipitating condition comprising 100 mM sodium acetate, 200 mM ammonium sulfate, and 25% PEG 4000, pH 4.6. Subsequent refinement of the three components of this crystallization buffer led to a condition comprising 100 mM sodium acetate, 350 mM ammonium sulfate, 25% PEG 4000, pH 5.0. While these conditions produced improved crystal quality and longevity, crystal growth was still primarily one dimensional and the resulting crystals were not suitable for diffraction studies. Notably, a protein solution containing TBP but omitting VP16C led to visually indistinguishable protein crystals under otherwise identical conditions, suggesting that the crystals obtained from the VP16C:TBP mixture were likely to comprise only TBP, probably in a dimerized form (123). At this point crystallization efforts were redirected to a tighter protein complex that apparently preserves the VP16C:TBP interaction, described in the next section.

VP16C-TAND2 bound to TBP

Rationale

TBP exists as a dimer in solution, with the concave DNA binding regions buried in the dimer interface (23). As VP16C likely binds to a region of TBP overlapping the DNA binding region, it is likely that TBP dimerization competes with the VP16C:TBP interaction. Since the dissociation constant of TBP dimers is in the low nanomolar range (23) and the dissociation constant of the VP16C:TBP interaction is about 40 nM (120, 153), TBP dimerization is favored over TBP:VP16C interactions. Increasing the strength of the VP16C:TBP interaction would work to push the equilibrium away from TBP dimerization and toward the desired VP16C:TBP complex.

In domain swap experiments aimed at demonstrating functional equivalence between VP16C and TAND1, Kotani and coworkers (83) constructed a fusion of VP16C and yTAND2. This fusion displayed strong binding to TBP, and was able to bind and retain TBP through extensive washing, whereas neither TAND1, TAND2, nor VP16C alone could retain significant amounts of TBP. The VP16C-TAND2 fusion also appeared to preserve a relevant VP16C:TBP interface. The TBP mutant L114K, which is deficient in interacting with VP16C, did not bind well to the VP16C-TAND2 fusion. Furthermore, VP16C, in the context of the same fusion to TAND2, could complement a TAND1-deletion phenotype in yeast, but a transcriptionally-defective mutant of VP16C failed to substitute for TAND1, indicating a sensitivity to critical mutations of VP16C.

Since the VP16C-TAND2 fusion binds TBP more tightly than VP16C alone, and since this fusion appears to preserve the transcriptionally relevant VP16C:TBP interface, it is an attractive target for structural characterization (Fig. 12A). The following sections describe attempts to crystallize a complex between a VP16C-TAND2 fusion and the conserved core region of yeast TBP.

Materials and methods

Plasmid construction. TAND2 (TAF1 residues 43-73) was PCR amplified from *S. cerevisiae* genomic DNA using adapter-primers ST805 and ST806. The resulting amplified DNA segment was tailored by *Eco*RI and *Xho*I restriction enzyme digestion and directionally cloned between the *Eco*RI and *Xho*I sites of plasmid pGEX4T-2 (Pharmacia), resulting in plasmid pDS42-5. VP16C (VP16 residues 452-490) was PCR amplified from plasmid pGEX-VP16C using primers ST803 and ST804. The resulting amplified DNA segment was tailored by *Bam*HI and *Eco*RI restriction enzyme digestion and directionally cloned between the *Bam*HI and *Eco*RI sites of plasmid pDS42-5 to create plasmid pDS42-10, encoding a direct fusion of VP16 Gly 490 to TAND2 Gly 41 as a GST fusion protein under control of the *tac* promoter. Plasmid sequences were verified by sequencing.

Expression of GST-VP16C-TAND2. Plasmid pDS42-10 was transformed into *E. coli* strain BL21 and cells containing the expression plasmid were grown in LB medium containing 80 mg/L ampicillin. One liter cultures were grown at 37 °C with vigorous shaking until an OD_{600} of about 0.7. Expression of the fusion protein was induced by addition of IPTG to a final concentration of 0.5 mM and growth was continued for an additional 3 h. Cells were harvested by centrifugation and cell pellets were resuspended in 10 mL HEMGT buffer (24 mM HEPES, 0.1 mM EDTA, 12.5 mM MgCl2, 10% v/v glycerol, 0.1% v/v TWEEN-20, pH 7.9) with 250 mM KCl. Resuspended cells were lysed by two passages through a French Press (Aminco) operating at 20,000 psi and

insoluble matter was sedimented by two steps of centrifugation at 10,000 x g at 4 °C for 10 min. The soluble lysate fraction was collected, split into 1 mL aliquots, flash frozen in liquid nitrogen, and stored at -80 °C.

Expression of yeast TBP. 6xHis-tagged yeast TBP was expressed in *E. coli* as described above. The soluble fraction of the bacterial lysate was split into 1 mL aliquots, flash frozen in liquid nitrogen, and stored at -80 °C.

Purification of the VP16C-TAND2:TBP complex. The VP16-TAND2:TBP complex was typically prepared by mixing 6-10 volumes of lysate containing TBP per volume of lysate containing GST-VP16C-TAND2. Lysate mixtures were diluted with ice-cold HEMGT buffer (24 mM HEPES, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% v/v glycerol, 0.1% v/v Tween-20) to reduce the salt concentration to 100 mM and incubated on ice for 1 h. The lysate mixture was added to GSH resin and the mixture was gently rotated for 45 min at 4 °C. The resin was washed 4 times with 75 resin volumes of ice-cold HEMGT with 100 mM KCl. The washed resin was resuspended to a 50% slurry and biotinylated thrombin protease was added. Digestion was allowed to proceed for 14-16 h at 4 °C with gentle rotation. Liberated proteins were collected by column-based elution, including a wash step supplementing the protein solution to final concentrations of 5 mM DTT and 0.05% w/v NaN_3. Eluted proteins were incubated with streptavidin agarose to trap the biotinylated thrombin and the remaining protein was concentrated in a centrifugal concentrator (Amicon) to 5-10 mg/mL total protein.

HPLC analysis and mass spectrometry. To prepare samples for reverse-phase HPLC, purified proteins were diluted in water with 0.1% trifluoroacetic acid (TFA). Alternatively, protein crystals were harvested and dissolved in acetonitrile with 0.1 % TFA. Samples were bound to a C18 reverse-phase HPLC column (Vydac, 5 μm bore, 300 Å particles). The column was developed with a linear gradient from Buffer A (water with 0.1% TFA) to buffer B (90% acetonitrile, 10% water, 0.1% TFA), and eluted material was observed by absorbance at 214 nm. The molecular mass of the proteins in the HPLC peak fractions was determined by matrix-assisted laser desorption/ionization-time of flight mass spectrometry.

Results

The design of the fusion protein comprising VP16C and yeast TAND2 was based on a VP16C-TAND2 fusion developed by Kotani and coworkers (83). The fusion protein incorporated the same junction point as the precedent, but was carefully terminated at the boundaries of the domains interacting with TBP. For VP16C, an alanine scanning study indicated that residues 457 and 458 were involved in activation in both yeast and mammalian activation assays (160), and thus the N-terminus of VP16C was defined as residue 452 to allow a native context for these residues. A deletion analysis of TAND2 (79) indicated that yeast TAF1 residues 10-73 were sufficient for a tight connection to TBP, whereas inclusion of 15 more residues conferred little additional binding strength. Thus, the fusion protein design includes VP16C (452-490) fused directly to TAND2(41-73).

The GST-VP16C-TAND2 (GVT) fusion protein (Fig. 12A,B) was expressed in *E. coli.* Soluble lysate fractions were bound to GSH resin and washed thoroughly. SDS-PAGE analysis of resin-bound proteins indicated recovery of a pure species with little sign of truncation or degradation (Fig. 12C). An elution of bound material with reduced GSH followed by protein quantitation indicated that about 60 mg of GVT could be recovered per liter of bacterial culture.

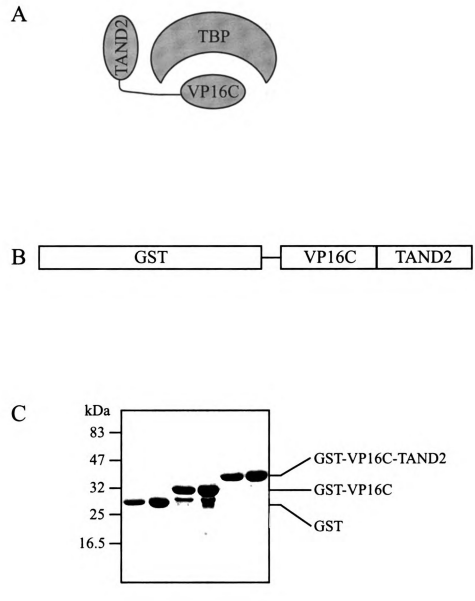


Figure 12: Design and purification of a VP16C-TAND2 fusion protein. (A) Predicted arrangement of VP16C and TAND2 when bound to TBP. (B) Design of the GST-VP16C-TAND2 fusion protein. The link between GST and VP16C encodes a thrombin recognition sequence. (C) SDS-PAGE gel of GST, GST-VP16C, and GST-VP16C-TAND2 proteins purified over GSH resin. Each pair of lanes represents elution of resin-bound proteins (left) by 10 mM reduced GSH and (right) by boiling in SDS-PAGE loading buffer. Proteins stained with Coomassie R-250.

To test the ability of GVT to associate with TBP, a bacterial lysate containing 6xHis-yTBPc was added to the GVT-containing lysate prior to purification of GVT. GVT effectively bound TBP and retained it through the subsequent washing steps, resulting in a clean purification of both GVT and TBP (Fig. 13A). The GVT:TBP association was found to be stable over a range of conditions, but could be weakened by a combination of elevated salt concentration and elevated temperature (Fig. 13B,C). Subsequent experiments have exploited the specificity and salt sensitivity of this interaction as an effective means of purifying untagged TBP (Chapter 2).

To saturate the GVT with TBP, TBP-containing lysate was titrated against a constant volume of GVT-containing lysate, and bound proteins were analyzed by SDS-PAGE. Typically 6 to 10 volumes of TBP-containing lysate were sufficient to saturate the TBP recovery (data not shown). A thrombin digestion was used to liberate the VP16C-TAND2:TBP complex from the resin and also to remove the hexahistidine tag from yTBPc (Fig. 14A). The resulting protein was recovered and concentrated to between 5 and 10 mg/mL total protein. Reverse-phase HPLC resolved the purified protein mixture into two distinct peaks (Fig. 14B). Mass spectrometry of the material eluted in the HPLC peaks indicated masses corresponding to intact VP16C-TAND2 and TBP.

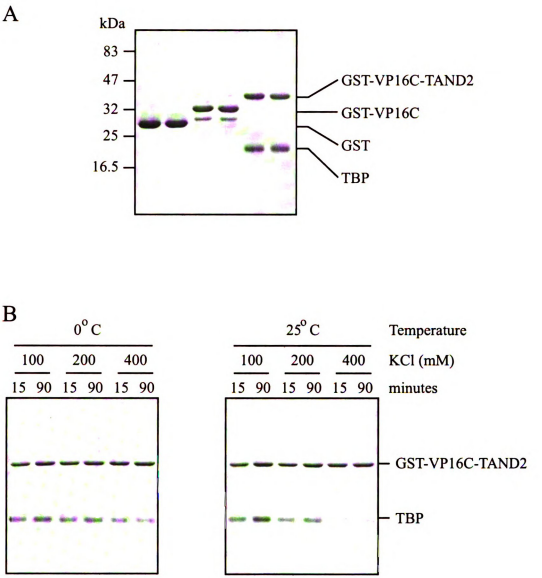


Figure 13: Association of GST-VP16C-TAND2 and TBP. Protein samples were resolved by SDS-PAGE and stained with Coomassie R-250. (A) *E. coli* lysates containing GST, GST-VP16C, or GST-VP16C-TAND2 were incubated with an *E. coli* lysate containing 6xHis-TBPc, and the resulting mixtures were purified over GSH resin. (B) Resin-bound complexes were subjected to washing under different conditions as indicated.

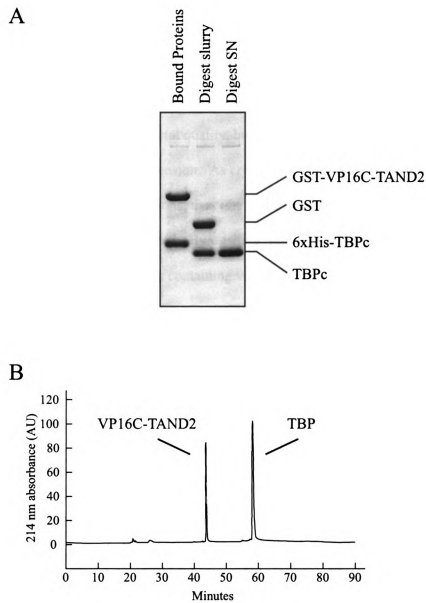


Figure 14: Purification of the VP16C-TAND2:TBP complex. (A) *E. coli* lysates containing GST-VP16C-TAND2 and 6xHis-TBPc were mixed, incubated, and purified over GSH resin. SDS-PAGE was used to resolve the bound proteins, products of thrombin digestion, and liberated proteins. Proteins were stained with Coomassie R-250. VP16C-TAND2, which migrates near the dye front on SDS-PAGE gels, is not shown. (B) Reverse-phase HPLC analysis of proteins liberated by thrombin digestion. HPLC peak fractions were identified by mass spectrometry.

Concentrated VP16C-TAND2:yTBPc was subjected to crystallization trials using both hanging drop and under-oil methods. The growth of needle-shaped crystals was observed with a precipitating mixture comprising 100 mM Tris, pH 8.5, 200 mM lithium sulfate, and 30% PEG 4000 (data not shown). A series of refinements of the crystallization variables led to improvements in crystal quality, but no conditions were found to promote crystal growth in more than one dimension. As crystal growth was promoted by increasing lithium sulfate concentrations, the effects of this salt on the VP16C-TAND2:TBP interaction was examined. By washing immobilized GST-VP16C-TAND2:TBP complexes with buffers containing various salts it was determined that lithium sulfate reduced the amount of retained TBP (data not shown), which suggested that lithium sulfate was effective at disrupting the TBP association. This observation suggested that the crystals obtained in the presence of elevated concentrations of lithium sulfate comprised TBP dimers liberated from the VP16C-TAND2 fusion protein.

Additional screening of crystallization conditions revealed that precipitant mixtures comprising 20% PEG 8000 with either 100 mM HEPES, pH 7.5, or 100 mM CHES, pH 9.5, could produce plate-like crystals. Several rounds of refinement led to precipitant mixtures comprising 100 mM CHES or HEPES, pH between 8.0 and 8.6, and PEG 4000 or PEG 8000 at between 4% and 12.5% v/v. These precipitant mixtures led to crystals that were well-grown in three dimensions (data not shown). These crystals were amenable to cryoprotection by slow stepwise washing into a mother liquor mimic containing 30% glycerol. Initial diffraction data indicated a unit cell of 120x90x60 angstroms and diffraction peaks consistent with crystallized protein. However, reverse-

phase HPLC analysis of dissolved crystals indicated that they comprised TBP and a lesser amount of an unidentified component, but not VP16C-TAND2 (data not shown). In parallel with the peptide analysis, a full set of diffraction data were collected from the crystals and solved by molecular replacement, confirming that the crystals comprised TBP dimers and no VP16C-TAND2.

Since degradation of VP16C-TAND2 may have destabilized the binary complex and released TBP for subsequent crystallization, the mother liquor of a crystal-bearing condition was analyzed for intact proteins. Reverse-phase HPLC indicated a complex mixture, but no clear peak corresponding to VP16C-TAND2 (data not shown). Although VP16C-TAND2 was not detected, it must be noted that the assessment of the mother liquor was made 35 days after the crystallization condition was assembled, and 25 days after TBP crystals were harvested, allowing an extended window for VP16C-TAND2 degradation. A conclusive measure of VP16C-TAND2 integrity during the crystallization period would require further experimentation.

Discussion

The VP16C-TAND fusion, modeled after the precedent from Nakatani and coworkers, binds the core region of yeast TBP with high enough affinity for a co-purification of TBP. The success of the co-purification technique, coupled with the salt-sensitivity of the intermolecular interaction, has been developed as an effective means of TBP purification (Chapter 2).

The TBP-only crystals derived from the VP16C-TAND2 complex indicate that the complex releases TBP over time, which may indicate degradation of VP16C-TAND2. Although VP16C-TAND2 is intact at times soon after preparation, an analysis of stability over long term incubation has not yet been completed. Any further attempts to crystallize this complex should be prefaced by assurances that VP16C-TAND2 is indeed stable over the time periods required for protein crystallization.

Alternatively, the TBP crystals could have arisen from dissociation of the intact VP16C-TAND2:TBP complex. This could occur by a kinetically limited rearrangement of the heterodimeric complex to thermodynamically favored TBP homodimers. Additionally, a TBP crystal lattice may provide a thermodynamically stable "sink" for any free TBP that exists in equilibrium with the heterodimeric complex. In either of these cases, increasing the stability of the complex may be of benefit. One approach to enhancing the stability of the complex, a direct fusion of VP16V-TAND to TBP, is explored in the next section.

VP16C-TAND2-TBP fusion protein

Rationale

Since protein crystals grown from a complex of VP16C-TAND2 and TBP contained only TBP, it may have been that the complex, while stable enough to retain TBP through purification, still allowed enough TBP to escape and for crystallization of TBP dimers to occur. One method for further tightening the complex was to provide a direct tether by encoding it as a single peptide chain, thus ensuring a high local concentration of VP16C-TAND2 and TBP. Other advantages of this approach included a simplified purification, and a resulting complex that would likely prevent the crystallization of dimerized TBP. The risks included the incorporation of a flexible linker that may have masked protein surfaces otherwise useful as lattice contacts, the imposition of steric constraints that may have disrupted proper folding of VP16C or TAND2, and the possibility that the direct fusion would interfere with proper folding of the TBP core domain.

Materials and methods

VP16-TAND2-TBP plasmid construction. A DNA fragment encoding VP16C-TAND2 was PCR amplified from plasmid pDS42-10 using primers ST906 and ST907. The resulting amplified DNA segment was tailored by *Nco*I and *Bam*HI restriction enzyme digestion and directionally cloned between the *Nco*I and *Bam*HI sites of plasmid pET28a to create plasmid pDSE104-1. A DNA fragment encoding TBPc (residues 61-240) was excised from the plasmid TEV-TBP (obtained from J.H. Geiger) by *Bam*HI and *Xho*I restriction enzyme digestion, gel purified, and directionally cloned between the *Bam*HI and *Xho*I sites of plasmid pDSE104-1 to create plasmid pDSE104-5. This plasmid encodes VP16C(452-490) fused to TAND2(41-73), a single glycine, and TBP(61-240), under control of a T7/lac promoter.

Linker extension: Cohesive oligonucleotides ST908 and ST909 were phosphorylated with T4 polynucleotide kinase. Plasmid pDSE104-5 was linearized with *Bam*HI digestion for 1 h, and then T4 DNA ligase, ATP, and phosphorylated oligonucleotides ST908 and ST909 were added directly to the digestion reaction. Extension of the linearized plasmid by linker polymerization was allowed to proceed for 2 h at 12 °C. Subsequently, *Bam*HI and *Mlu*I were added directly to the ligation reaction, and digestion was allowed to proceed at 37 °C for 1 h. The DNA segment corresponding to linker-extended VP16C-TAND2 was gel purified and ligated to a *Bam*HI-*Mlu*I fragment of pDSE104-5 encoding the remainder of the parental plasmid sequence, in order to recapitulate the original plasmid sequence with the linker extensions of TAND2.

Ligated plasmids were transformed into *E. coli* strain DH5α and transformed cells were selected on LB with 50 μg/mL kanamycin. Transformant colonies were screened by colony PCR using primers ST805 and ST925 that flank the extension site, and candidates leading to PCR products of expected length were confirmed by sequencing, resulting in a family of plasmids denoted pDSE106-n, where n represents the number of insertions of the GSGS tetrapeptide sequence between TAND2 and TBP.

From the pDSE104-5 and the pDSE106-n series of plasmids, fragments encoding TAND2-linker-TBP were excised by *Eco*RI and *Xho*I restriction digestion, gel-purified, and directionally cloned between the *Eco*RI and *Xho*I sites of plasmid pDSC2-2 (encoding GST-VP16C-TAND2 under control of a T7/lac promoter), yielding sequences encoding GST-VP16C-TAND2-linker-TBP inserted between the *Nco*I and *Xho*I sites of plasmid backbone pRSF-DUET-1 (Novagen). This plasmid family is denoted pDSE112-n. Subsequently, the *Nco*I-*Xho*I fragments encoding the entire fusion protein were subcloned into the pETDuet-1 (Novagen) plasmid background, resulting in plasmid family pDSE117-n.

Protein expression and purification. Plasmids encoding VP16C-TAND2-linker-TBP were transformed into *E. coli* strain BL21(DE3)-codon plus. Cells containing the plasmid were grown in LB medium containing 40 mg/L chloramphenicol and either 80 mg/L ampicillin or 50 mg/L kanamycin. One liter cultures were grown at 25 °C with vigorous shaking until an OD_{600} of 1.0. Expression of the fusion proteins was induced by

addition of IPTG to a final concentration of 0.1 mM. Cultures were grown an additional 4 h. Cells were harvested by centrifugation and then resuspended in 20 mL wash buffer (20 mM Tris-Cl, 5% v/v glycerol, 0.1 mM EDTA, 500 mM KCl, pH 7.9) with 5 mM DTT. Resuspended cells were lysed by three passages through a French Press (Aminco) operating at 20,000 psi. Insoluble matter was sedimented by centrifugation at 16,000 x g for 30 min at 4 °C. The soluble lysate fraction was recovered and mixed with 0.75 mL GSH resin and rotated gently for 1 h at 4-7 °C. The GSH resin was recovered and washed 5 times with 75 resin volumes of ice-cold wash buffer. Thrombin was added to liberate the fusion proteins from resin-bound GST, and digestion was allowed to proceed for 8 to 14 h at 7 °C with gentle rotation. Eluted proteins were collected and further purified by size exclusion chromatography.

Size exclusion chromatography. Preparative and analytical size exclusion chromatography (SEC) was performed over Superdex S-200 prep grade media in a Pharmacia XK-series 75/100 column using a Pharmacia FPLC system, operating at 7 °C, with 1 mL/min flow rate and 280 nm absorbance detection of eluted material.

Results

In an attempt to stabilize a complex of VP16C-TAND2 and TBP a direct fusion of the two proteins was explored. The fusion protein linkage was designed both to incorporate a minimal length linking peptide segment and to insulate VP16C from linker-imposed steric constraints. Since the TAND2 domain interacts with the convex upper surface of TBP (79, 102), it is likely to be closer to TBP's N-terminus than is VP16C, and thus TAND2 provided a reasonable point for attachment to TBP. Thus, the C-terminus of the VP16C-TAND2 fusion was bridged to the N-terminus of the core region of yeast TBP (Fig. 15A,B). To allow for flexibility and a hydrophilic character to the linker peptide, it was designed to encode a series of glycine-serine repeats. Since the location of the TAND2 C-terminus in unknown, a series of fusion proteins with increasingly longer linker peptides was generated.

To generate this family of peptide linkers of varying length, complementary oligonucleotides encoding Gly-Ser-Gly-Ser were designed (Fig. 15C). The annealed oligonucleotide pairs have 5' GATC overhangs on each end, allowing for polymerization in the presence of DNA ligase activity. The sequences were designed such that only a head-to-head ligation of the annealed oligo pairs creates a *Bam*HI restriction site. In this way, a polymer of randomly-oriented units, when digested by *Bam*HI, results in polymer of forward-oriented units of varying length, terminating with a *Bam*HI-generated 5'-overhang (Fig. 15D). This polymerization strategy was used to extend the DNA

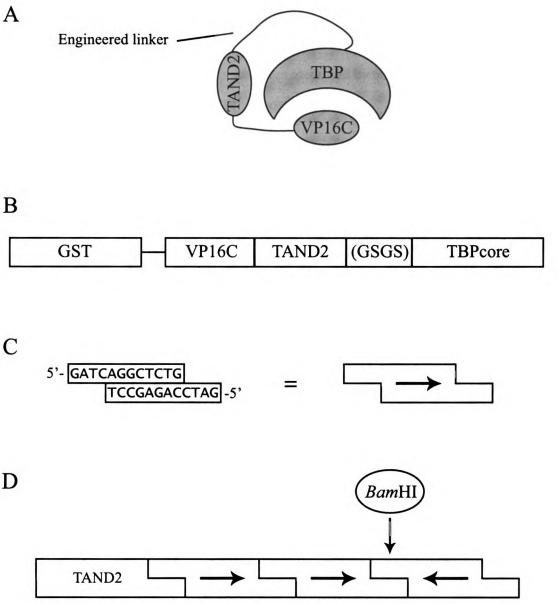


Figure 15: Design and construction of a fusion of VP16C-TAND2-linker-TBP. (A) Predicted arrangement of a fusion of VP16C-TAND2 bound to the core region of yeast TBP, and placement of an engineered linker. (B) Primary structure of the fusion protein. (GSGS) represents a variable number of GSGS tetrapeptide segments. (C) Cohesive oligonucleotides that anneal to form a single extension unit. (D) Extension of the DNA sequence encoding TAND2 by polymerization of multiple randomly-oriented extension units. Only a head-to-head ligation of extension units encodes a *Bam*HI recognition sequence.

sequence encoding TAND2, and the extended sequence was ligated to the sequence encoding the core region of yeast TBP. Resulting plasmids encoding different characteristic lengths of the peptide linker were identified by colony PCR with primers flanking the linker sequence, and candidate clones were verified by sequencing. This procedure generated sequences encoding VP16C-TAND2-TBP fusions with 1 to 7 insertions of the GSGS linker segment, all resulting from a single execution of the oligonucleotide extension reaction.

The fusion protein constructs were initially encoded in pRSF-based plasmids, which confer kanamycin resistance and have a copy number of about 100 per cell. After expression in *E. coli*, GST-VP16C-TAND2-linker-TBP was purified from the bacterial lysates using GSH resin. The overall yield of purified protein was about 20 mg per liter of initial culture. SDS-PAGE analysis of the purified proteins indicated a prominent species of the expected molecular weight, and the molecular weight distribution seen across the family of expression constructs confirmed the identity of the fusion proteins (Fig. 16A). A significant amount of other proteins was present in the purification, and these could not be separated from the intact fusions either by increasing the number of wash steps or by increasing the salt concentration of the wash buffer. Since the contaminating proteins primarily spanned the molecular weight range from GST to full fusion protein, they were likely truncation or degradation variants of the full fusion protein.

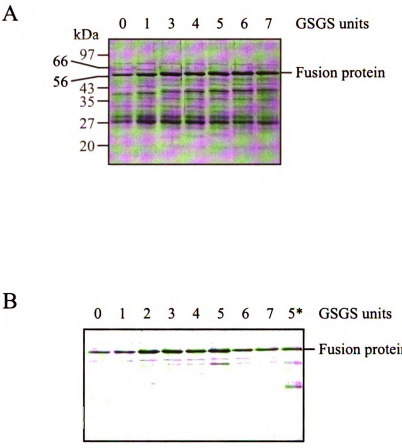


Figure 16: Purification of GST-VP16C-TAND2-linker-TBP fusion proteins. (A) Fusion proteins encoded by pRSF-based vectors were expressed in *E. coli* and purified over GSH resin. Bound proteins were eluted by boiling in SDS-PAGE loading buffer, resolved by SDS-PAGE, and stained with Coomassie R-250. (B) Fusion proteins encoded by pET-based vectors were purified as in (A). Lane 5* denotes protein encoded by a pRSF-based vector, loaded for comparison. GSGS denotes the number of GSGS linker insertions is the fusion protein.

Since the antibiotic kanamycin interferes with protein synthesis, it was possible that expression under a different selection might improve the quality of the purified proteins. With this reasoning, the fusion protein constructs were subcloned into a lower copy pET-based plasmid backbone that conferred ampicillin resistance. After expression under the same conditions, the quality of the purified protein improved significantly (Fig. 16B), although the overall yield diminished to about 3 mg purified protein per liter of initial bacterial culture. Based on the increased quality of the recovered proteins, the pET-based plasmids were used for all subsequent work.

It was hypothesized that if the engineered linker was too short, VP16C and TAND2 would not be able to reach and bind to their cognate surfaces on TBP. Accordingly, it was expected that fusion proteins with engineered linkers below this critical threshold would exist as TBP-linked dimers in solution, whereas fusion proteins with sufficiently long linkers would exist as monomers in solution, with the TBP dimerization surface occluded by VP16C (Fig. 17A,B). The apparent solution size of the series of fusion proteins was resolved by SEC of the purified proteins (Fig. 17C). Interestingly, for fusion proteins with a linker length of 8 or more residues, the apparent molecular weight was approximately 33 kDa, consistent with a compactly folded monomeric species, whereas a shorter linker length led to a larger apparent molecular mass of about 66 kDa, which is consistent with TBP-linked dimers of fusion protein.

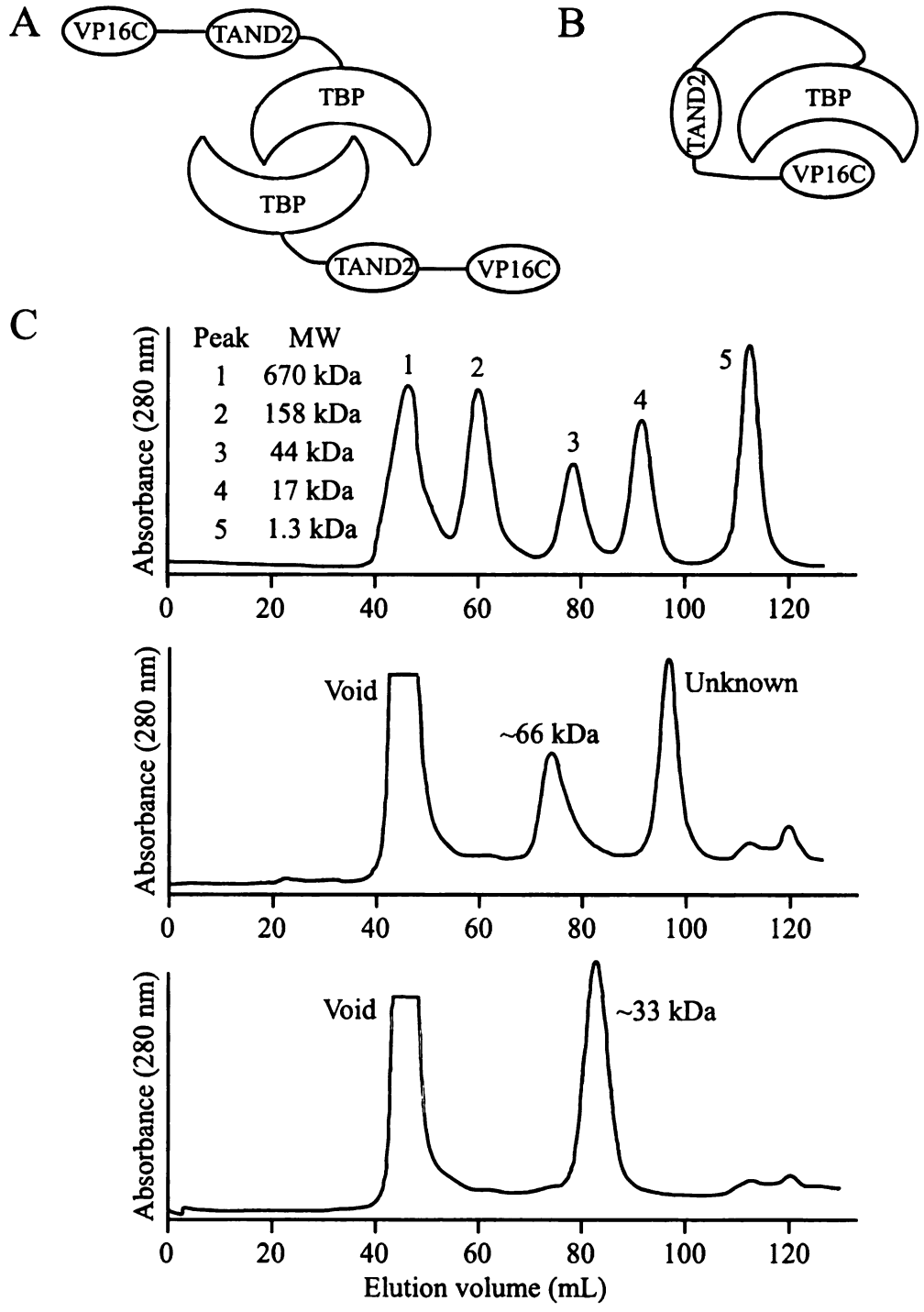


Figure 17: Monomeric and dimeric states of VP16C-TAND2-TBP fusions. (A) Predicted arrangement when the linker is too short to allow intramolecular binding between VP16C-TAND2 and TBP. (B) Predicted arrangement when the linker is long enough to allow intramolecular binding. (C) SEC of purified fusion proteins. Top panel, molecular size standards. Middle panel, fusion protein with 4 amino acid linker. Bottom panel, fusion protein with 8 amino acid linker.

Size-exclusion chromatography effectively separated the intact fusion proteins from a large fraction of material of high apparent molecular weight, and thus this step was included in all subsequent purifications (Fig. 18). The protein in the 33 kDa SEC peak, when recovered, concentrated, and resolved again over the SEC column, eluted almost exclusively at the same position (data not shown), which indicated that the folded, apparently monomeric state of the fusion protein was stable over short times. SDS-PAGE analysis of proteins recovered from subsequent crystallization experiments indicated overall primary structure stability over long incubations at room temperature, although some proteolysis was evident in some cases.

As the size exclusion chromatography fraction was of high purity and consistent with monomeric fusion protein, it was subjected to crystallization trials, using primarily an under-oil approach with either paraffin oil (to maintain mixture concentrations), or a 1:1 mixture of silicone oil and paraffin oil to allow slow evaporation of water from the drops. The length of the engineered linker peptide is a crystallization variable, and thus the GS2, GS3, and GS4 fusion proteins were included in crystallization experiments. Initial trials using commercially-available sparse screens yielded rapidly-forming precipitates under most conditions, with the notable exception of low salt, high pH conditions. It was reasoned that these conditions might be necessary to stabilize the fusion protein, and directed attempts at crystallization were made under these conditions. A diverse series of attempts to precipitate the fusion protein under high pH and low salt conditions led only to clear drops or to phase separation.

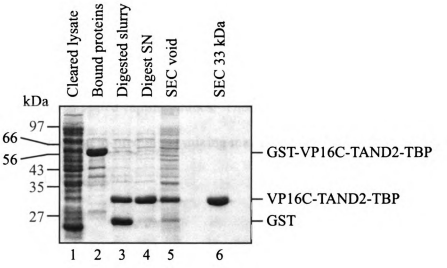


Figure 18: Purification of GST-VP16C-TAND2-TBP fusion protein. Protein samples were resolved by SDS-PAGE and stained with Coomassie R-250. Lane 1, cleared lysate. Lane 2, Proteins bound to GSH resin. Lane 3, products of thrombin digestion. Lane 4, proteins liberated from resin. Liberated proteins were subjected to SEC. Lane 5, SEC void fraction. Lane 6, proteins eluting at approximately 33 kDa.

Based on the high apparent solubility of the fusion proteins in low-salt and high pH conditions, other crystallization methods were employed in efforts to move gently from stabilizing to precipitating conditions. The fusion proteins were concentrated to 40 mg/mL and small aliquots were dialyzed stepwise into lower pH and higher salt conditions. In all cases this method produced slight and then increasing turbidity in the samples, but no crystallization. To attempt crystallization under high concentrations of protein and precipitating agents, free interface diffusion techniques were employed. Fusion protein concentrated to 50 mg/mL was allowed to directly contact a variety of precipitating conditions in sealed capillary tubes. These methods also led to visibly clear protein samples, but the protein-precipitant mixing may have been limited by the high viscosity of the precipitating agents.

Discussion

Approximately 2500 attempts to crystallize the GS2, GS3, and GS4 proteins have failed to produce any observable crystals, suggesting that crystallization of these fusion proteins is unlikely to succeed. Protein crystallization can fail for a variety of reasons, including proteolysis, structural instability, intrinsic disorder, or simply a failure to find (or a complete absence of) successful crystallization conditions. Among these possibilities, degradation may be the easiest to diagnose and control. However, during the course of the fusion protein experiments is the preceding section, the integrity of the proteins was not frequently monitored, thrombin protease was not explicitly removed from the protein preparations, and protease inhibitors were not typically included. While an SDS-PAGE analysis indicated that some conditions preserved the primary structure of the proteins over long term incubation, other conditions showed signs of proteolysis. Therefore some of the crystallization attempts may have failed due to proteolysis. Any further attempts at crystallization of these proteins should incorporate safeguards against proteolysis.

Structural instability can also thwart protein crystallization. While the GSn fusion proteins remained in a folded, apparently monomeric state after purification and concentration, this analysis came only after a short incubation at low temperatures. The long-term structural stability of the fusion proteins remains unexplored. However, a more thorough analysis of the structural stability of a similar VT:TBP crosslinked complex (chapter 5) shows that salt concentrations below 10 mM are required to preserve the compact monomeric state for even a few days at 20 °C. In this respect, most of the

crystallization attempts in the preceding sections may have failed due to structural instability.

Additionally, instrinsic disorder can work against protein crystallization. In the handoff mechanism of activation, the connection of VP16C to TBP is transient, and the transient nature of the interaction may be reflected in intrinsic disorder of the activation domain. Indeed, an NMR analysis of the crosslinked complex (chapter 5) indicated that a significant portion of the VP16C-TAND2 peptide existed in an unstructured state in the complex. While the distribution of structured and unstructured elements is unknown, it is possible that the regions of VP16C and TAND2 included in the fusion protein design encompass a structured core with unstructured flanking regions. In this case, truncations of either or both domains may promote crystallization.

The engineered GSGS linker is also a likely source for disorder. The minimal length of this linker is currently determined as 2 units of 4 residues, but the minimal number of residues may be anywhere from 5 to 8 residues. It is possible that a minimal length linker may assist crystallization. Furthermore, the stabilizing conditions of low salt and low temperature may enable the complex to exist stably without the linker peptide, which is likely to reduce the intrinsic disorder of the complex by exposing more of the well-structured TBP surface. While the long-term integrity of the VP16C-TAND2:TBP heterodimeric complex under stabilizing conditions remains to be tested, this complex may provide the best target for any subsequent attempts at crystallization.

Chapter V

NMR analysis of the VP16-TAND2-TBP complex

Introduction

The efforts to crystallize a protein complex containing the VP16C activation domain complexed with TBP (chapter 4) led to the development of a fusion protein comprising VP16C-TAND2-linker-TBP. This protein could be purified in an apparently compact, monomeric form, suggesting that it was a suitable candidate for crystallization. Although all attempts to crystallize this protein failed, it was noted that this protein could be concentrated to a degree suitable for protein NMR spectroscopy. This chapter describes an initial structural characterization of the VP16C-TAND2-linker-TBP fusion protein using NMR spectroscopy.

NMR can provide a set of through-space distance measures between amide protons in proteins via Nuclear Overhauser Effect spectroscopy (NOESY). NOESY measures span short distances (typically 5 Å or less), and thus are useful for the detection and definition of secondary structural elements and tertiary contacts. A full set of NOESY distance measures can be used to constrain tertiary structural models, providing high-resolution protein structures. NOESY requires an assignment of amide proton resonances to particular residues in a protein's primary structure, both to sort the complex spectra and to apply the resulting information to a structural model. Resonance assignment typically involves a sequential identification of bonds in the protein backbone, requiring the use of

^{13}C isotopic labeling and complex spectral analysis. Resonance assignment typically comprises a large fraction of NMR protein structure determination efforts.

The difficulty of protein NMR increases with the size of the target protein, for a few reasons. With increasing numbers of resonances, the resulting spectra become increasingly dense and more prone to spectral overlap. Larger proteins also have longer rotational correlation times in solution, which leads to inhomogeneous sampling of the NMR magnetic field and thus broadening of the resonance peaks, compounding problems of spectral overlap in the already dense spectra. Additionally, spectral assignment utilizes ^{13}C resonances, which are intrinsically broad, further complicating resonance assignment efforts. These effects suggested that resonance assignment of the 28 kDa VP16C-TAND2-linker-TBP fusion protein would be difficult, although probably tractable.

One technique for simplifying NMR spectra of large proteins is to label only the segments of interest. Since the structure of TBP has already been determined, specific labeling of the VP16C-TAND2 portion of the complex would highlight the unknown regions and greatly simplify the resulting spectra. A drawback to this approach is that specific contacts between the labeled segments and the unlabeled segments would not be determined. The following sections describe the application of NMR techniques to study a cross-linked mimic of the fusion protein that carries isotopic labeling on only the VP16C-TAND2 segment.

Materials and methods

VP16C-TAND2-linker-Cys plasmid construction. Plasmid pDS42-10, encoding a fusion of GST to VP16C (452-490) and TAND2 (41-73), was cleaved with *Eco*RI and *Xho*I and the plasmid backbone was isolated. Cohesive oligonucleotides ST1033 and ST1034 were annealed and phosphorylated with T4 polynucleotide kinase, and directionally cloned into the *Eco*RI and *Xho*I sites of the pDS42-10 plasmid backbone, resulting in plasmid pDSE137-3. This plasmid encodes VP16C (452-490) followed by a sequence of *Eco*RI, *Bgl*II, and *Xho*I recognition sites, encoding a fusion of Gly-Ala-Asn-Ser-Cys-stop to Gly 490 of VP16C. This plasmid serves as a destination plasmid for subsequent cloning.

To create a family of plasmids encoding the fusion VP16C-TAND2-linker-Cys, *Eco*RI-*Bam*HI segments encoding TAND2 fused to a series of Gly-Ser-Gly-Ser repeats were extracted from the plasmid family pDSE106-n (chapter 4), and directionally cloned between the *Eco*RI and *Bgl*II sites of pDSE137-3, resulting in plasmid family pDSE138-n. This plasmid family encodes GST-VP16C-TAND2-linker-Cys, where n denotes the number of GSGS tetrapeptide units comprising the linker.

GST-VP16C-TAND2-linker-Cys expression. The pDSE138-n plasmid family was transformed into *E. coli* strain BL21. Cells containing expression plasmids were grown in LB medium containing 80 mg/L ampicillin. One liter cultures were grown at 37 °C with vigorous shaking until an OD_{600} of about 0.7. Expression of the fusion protein was

induced by addition of IPTG to a final concentration of 0.5 mM and growth was continued for an additional 3-5 h while cultures were allowed to approach ambient temperature (25 °C). Cells were harvested by centrifugation and resuspended in 20 mL HEMGT-250 (24 mM HEPES, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% v/v glycerol, 0.1% v/v Tween-20, pH 7.5) with 250 mM KCl, supplemented with 1 mM PMSF and 5 mM DTT. Resuspended cells were lysed by 3 passages through a French Press (Aminco) operating at 20,000 psi. Cellular lysates were centrifuged at 16,000 x g at 4 °C for 25 min. The supernatant was recovered, flash frozen in liquid nitrogen, and stored at -80 °C.

To obtain VP16C-TAND2-linker-Cys proteins for NMR spectroscopy, the plasmid pDSE138-0 was transformed into *E. coli* strain BL21. Cells containing the plasmid were grown in LB media with 80 mg/L ampicillin. 5 mL of LB-grown culture was inoculated into 1 L of M9 medium prepared with [^{15}N] NH_4Cl and containing 80 mg/L ampicillin. M9 cultures were grown to an OD600 of 0.4, and then shifted to 25 °C. Protein expression was induced by addition of IPTG to a final concentration of 0.5 mM, and growth was continued for 20 h at 25 °C. Cells were harvested, lysed, and the soluble lysate fraction was stored as above.

GST-VP16AD plasmid construction. Plasmid pSJT1193(CRF1) was cleaved by *Bam*HI and *Bgl*II to liberate the VP16 activation domain (413-490). This fragment was cloned into the *Bam*HI site of plasmid pGEX 4T-2 (Amersham), and a forward oriented construct was selected, resulting in plasmid pDSE171-1, encoding the VP16AD as a thrombin-cleavable fusion to GST under control of the *tac* promoter.

GST-VP16AD protein expression. Plasmid pDSE171-1 was transformed into *E. coli* strain BL21. Cells containing the plasmid were grown in LB medium containing 80 mg/L ampicillin. One liter cultures were grown at 37 °C with vigorous shaking to an OD_{600} of 1.0. Protein expression was induced by addition of IPTG to a final concentration of 0.5 mM, and growth was continued 1 h, after which cultures were chilled in an ice-water bath. Cells were harvested by centrifugation and resuspended in 20 mL buffer HEGK-250 (24 mM HEPES, 0.1 mM EDTA, 10% v/v glycerol, pH 7.5) with 250 mM KCl and supplemented with 1 mM PMSF and 5 mM DTT. Resuspended cells were lysed by sonication (Branson Sonifier 450, 80% duty cycle, power level 8, 4 cycles at 30 sec with 2 min between cycles). Cellular lysates were centrifuged at 16,000 x g at 4 °C for 25 min. The lysate soluble fraction was collected, flash frozen in liquid nitrogen, and stored at -80 °C.

TBP expression plasmids. Plasmid trc-TBP, encoding *Saccharomyces cerevisiae* TBP (residues 61-240) (yTBPc) as a 7xHis-tagged thrombin-cleavable fusion protein, was obtained from Dr. J.H. Geiger. Plasmid TEV-TBP, encoding yTBPc as a 7xHis-tagged TEV-protease cleavable fusion protein in a pET21 backbone, was obtained from Dr. J.H. Geiger.

To create a plasmid encoding untagged yTBPc, the TBP coding region was PCR amplified using an upstream primer encoding an *Nde*I sequence followed by a match to TBP Ser 61 and beyond, and a downstream primer matching the plasmid backbone. The

resulting DNA fragment was tailored by *Nde*I and *Xho*I to isolate the TBPc coding sequence and directionally cloned between the *Nde*I and *Xho*I sites of TEV-TBP to create plasmid pDSE148-1 in a pET21 backbone.

To create a plasmid encoding yTBPc with an engineered N-terminal cysteine, plasmid pDSE117-5a, encoding GST-VP16C-TAND2-linker-TBP, was cleaved by *Nco*I and *Bam*HI and the plasmid backbone (carrying the TBP sequence) was isolated. Cohesive oligonucleotides DS1 and DS2 were annealed, phosphorylated with T4 polynucleotide kinase, and directionally cloned into the plasmid backbone, resulting in plasmid pDSE141-1, encoding Met-Gly-Cys-Gly-Ser-Gly-yTBP(61-240).

To create plasmids encoding yTBP (49-240) and yTBP (40-240), the corresponding coding sequences were PCR amplified from a clone of full-length yTBP (ywTBP, Dr. J.H. Geiger) using upstream adaptor-primers DS3 or DS4 and a downstream T7 terminator primer matching the vector sequence. The resulting amplified DNA sequences were tailored by *Nde*I and *Hin*DIII digestion and directionally cloned into the *Nde*I and *Hin*DIII sites of TEV-TBP. The resulting plasmids encode untagged yTBP(49-240) (pDSE162-1) and untagged yTBP(40-240) (pDSE162-5) in a pET21 vector backbone.

TBP expression. His-tagged TBPc and TBPc with the N-terminal engineered cysteine were expressed as described (Chapter 4). For untagged TBP versions (residues 40-240, 49-240, and 61-240), *E. coli* strain BL21(DE3)codon plus was transformed with the expression vectors, and cells containing the plasmids were grown in 50 mL TB (Terrific

Broth) medium with 80 mg/L ampicillin and 40 mg/L chloramphenicol at 37 °C with vigorous shaking for 4-8 h. 7 mL of the starter culture was inoculated into one liter of TB medium with 40 mg/L ampicillin and growth was continued at 37 °C with vigorous shaking until an OD_{600} of 1.7. Protein expression was induced by addition of IPTG to a final concentration of 0.1 mM and growth was continued at 20 °C for an additional 10 h. Cells were harvested by centrifugation and cell pellets were resuspended to 25% w/v in lysis buffer (50 mM sodium phosphate, 20% v/v glycerol, 500 mM NaCl, 1 mM EDTA, 1 mM PMSF, 5 mM DTT, 0.1% Triton X-100, pH 7.5). Resuspended cells were lysed by 3 passages through a French Press (Aminco) operating at 20,000 psi and cellular lysates were centrifuged at 16,000 x g at 4 °C for 35 min. The lysate supernatant was collected, flash frozen in liquid nitrogen, and stored at -80 °C.

Purification of protein complexes. Protein complexes comprising GST-activation domain fusions and TBP versions were purified using a mixed lysate approach (chapter 2 and chapter 4). Briefly, a saturating amount of soluble lysate containing TBP was added to soluble lysate fractions containing GST-activation domain variants. Lysate mixtures were incubated on ice for 1 h, and subsequently purified over GSH resin. Thrombin protease was used to cleave activator:TBP complexes from immobilized GST. In certain cases liberated proteins were further purified by size exclusion chromatography (see below).

Crosslinking of protein complexes for NMR. 1,8-Bis-maleimidotriethyleneglycol ($BM(PEO)_3$) (Pierce) was added to VP16C-TAND2:TBP complexes at a final

concentration of 0.25 mM and samples were incubated on ice for 30 min. Subsequently 2-mercaptoethanol was added to a final concentration of 2 mM to quench unreacted $BM(PEO)_3$. Crosslinked proteins were separated from the crosslinking reaction buffer components by gel filtration.

Reduced IgG. Reduced IgG was prepared from a rabbit IgG preparation by addition of DTT to a final concentration of 50 mM and incubation at 25 °C for 30 min. Reduced IgG was dialyzed into a storage buffer (20 mM Tris, 2 mM EDTA, 100 mM KCl, pH 7), and stored at -80 °C.

Size exclusion chromatography. Preparative and analytical size exclusion chromatography (SEC) was performed over Superdex S-200 prep grade media in a Pharmacia XK-series 75/100 column using a Pharmacia FPLC system, operating at 10 °C, with 1 mL/min flow rate and 280 nm absorbance detection of eluted material. SEC buffers typically comprised 10 mM Tris, 0.1 mM EDTA, and 10 mM KCl, pH 8.

NMR spectroscopy. ^{15}N-^{1}H heteronuclear single quantum correlation (HSQC) spectra were acquired at 20 °C using a Varian Unity 600 MHz spectrometer equipped with a triple resonance probe (Max T. Rogers NMR facility, MSU) with the assistance of Dr. Aizhou Liu.

Results

In an effort to determine structural properties of VP16C bound to TBP we have attempted to characterize a VP16C-TAND2-TBP fusion protein using NMR spectroscopy. An initial 1-dimensional proton NMR spectra obtained from the VP16C-TAND2-linker-TBP fusion protein indicated resolvable proton peaks (data not shown), which suggested that the fusion protein was amenable to NMR structural analysis. However, analysis of this protein faced two different hurdles. First, expression of the fusion protein in M9 minimal media (required for isotopic labeling of the protein) was poor. Under optimized conditions, the overall yield of fusion protein was only 0.5 mg per liter of culture, implying that about 30 liters of culture would be required to produce a single 0.5 mL, 1 mM protein sample for NMR. The cost of 30 liters of [^{13}C] containing M9 media, required for backbone resonance assignment, was prohibitive. Second, the 28 kDa fusion protein is near the upper range for successful NMR structure determination, implying that the spectral analysis would be difficult. To circumvent these weaknesses, it was noted that the expression of the GST-VP16C-TAND2 fusion protein in M9 media was sufficient to provide for an NMR sample from less than one liter of M9 minimal media. Additionally, a separate preparation of VP16C-TAND2 and TBP would allow isotopic labeling of VP16C-TAND2 only, which would result in 76 labeled residues (instead of 256), greatly simplifying the NMR spectral density.

To recapitulate the structure and solubility properties of the single-peptide fusion protein, it was reasoned that a disulfide bridge might be able to connect an engineered cysteine in VP16C-TAND2 to TBPc. Yeast TBP has two cysteine residues, and one, Cys 78, is surface exposed. Cys 78 has been used previously to tether TBP in surface plasmon resonance assays (120). Cys 78 is near the region of TBP that interacts with TAND2 (102), and thus this residue was a good candidate for disulfide bridging to TAND2. Since the structure of TAND2 bound to TBP is unknown, a cysteine was engineered onto the C-terminus of a family of VP16C-TAND2-linker proteins, where the linker comprises a number of GSGS tetrapeptide repeats. With this system, a long linker could be used to explore the cross linking reaction, and subsequently the shortest linker allowing intermolecular disulfide formation could be selected for further study.

To promote disulfide formation in the complex between VP16C-TAND2-linker-Cys and TBP, oxidized glutathione (2 mM) was used as a disulfide shuffling reagent. However, under conditions that promoted disulfide formation in a reduced IgG control reaction, very little VP16C-TAND2 was bridged to TBP. Since little disulfide-linked complex was formed, it was suspected that TBP Cys 78 might not be available for disulfide formation in the heterodimeric complex. Cys 78 is near the TAND2 binding site on TBP, and thus TAND2 binding might occlude Cys 78, preventing access by the C-terminal cysteine of TAND2. To work around this possibility, an additional cysteine was engineered onto the N-terminus of TBPc to provide an alternative crosslinking site. While this Cys-TBPc could be disulfide bridged to VP16C-TAND2-linker-Cys more readily, the crosslinking reaction failed to go to completion (Fig. 19, left panels).

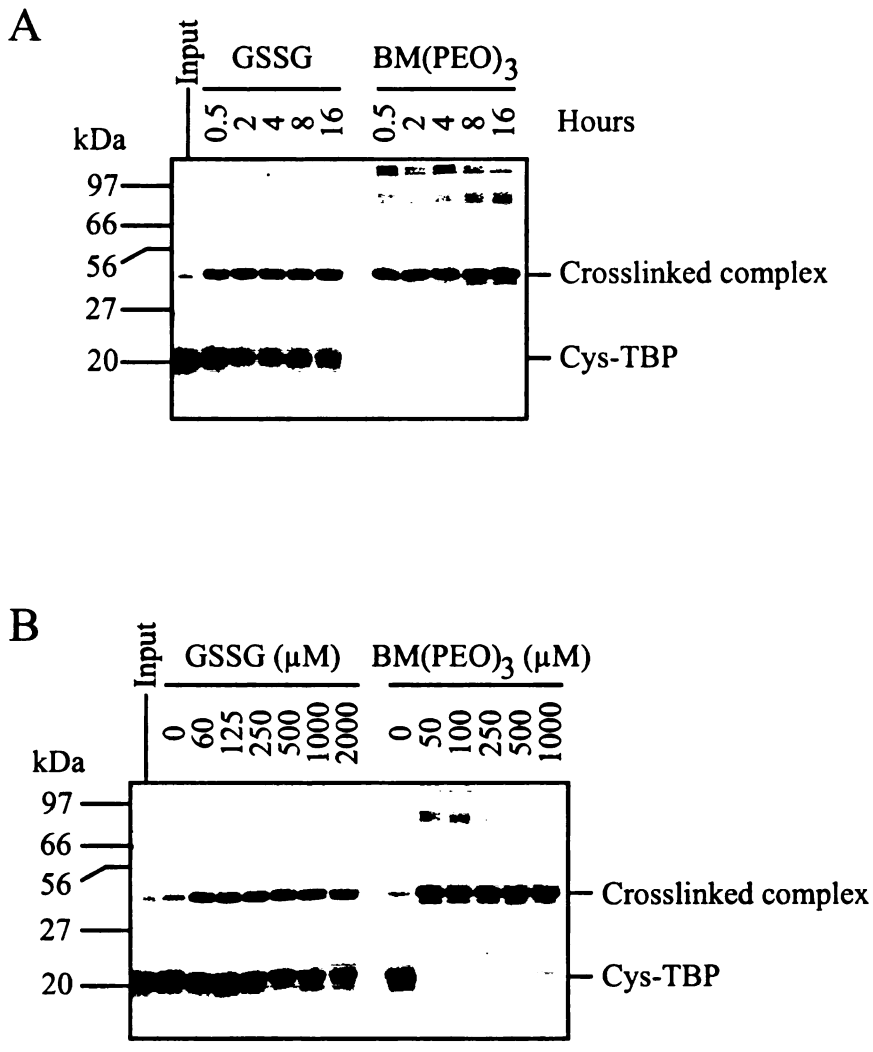


Figure 19: Crosslinking of VP16C-TAND2-Cys to Cys-TBP. (A) Purified VP16C-TAND2-Cys and Cys-TBP were incubated for the specified times at 4°C in the presence of either 2 mM oxidized glutathione (GSSG) or 50 μM $BM(PEO)_3$ and subsequently resolved by SDS-PAGE and stained with Coomassie R-250. Free VP16C-TAND2 is not resolved on the gel. (B) VP16C-TAND2-Cys and Cys-TBP were incubated in the presence of the indicated concentrations of either GSSG or $BM(PEO)_3$. Samples were incubated at 4°C for 16 h (GSSG) or 8 h ($BM(PEO)_3$) and subsequently processed as in (A).

As an alternative to disulfide bridging, the homobifunctional crosslinker $BM(PEO)_3$ was tested for its ability to crosslink VP16C-TAND2-linker-Cys to Cys-TBP. Over a range of conditions, $BM(PEO)_3$ cross-linking of the complex was essentially complete, resulting in primarily a single species (Fig. 19, right panels). The presence of the secondary crosslinked species may reflect a less-favored connection to Cys 78, which is exposed on the surface of TBP. Further exploration of $BM(PEO)_3$ crosslinking showed that the engineered N-terminal Cys on TBP was not required for crosslinking (Fig. 20A), which suggested that Cys 78 of TBP was the point of attachment to the crosslinker. It was also determined that even the shortest GSGS linker supported crosslinking (Fig. 20A). Additionally, the molecular weight ladder seen across the cross-linked species provided an unambiguous measure of VP16C-TAND2-linker-Cys incorporation into the crosslinked complex (Fig. 20A).

When the resulting crosslinked complexes were resolved by size-exclusion chromatography (SEC), they behaved similarly to single-peptide fusion proteins (Chapter 4), with roughly half of the material eluting in the void fraction and the other half eluting at a low molecular weight position corresponding to a compactly folded monomer (data not shown). SEC was included as a final step in purification of the complex (Fig. 20B).

To determine an appropriate buffer system for NMR analysis, the purified and concentrated cross-linked complexes were dialyzed into a variety of conditions and visually monitored over several days. Using this approach, it was determined that Tris buffer was preferred over phosphate or glycine, higher pH values were better, and low

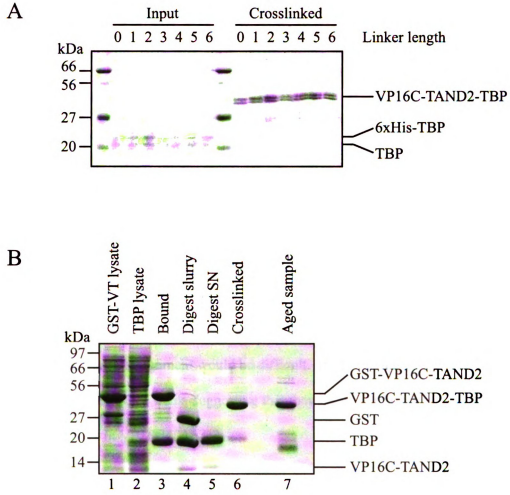


Figure 20: Crosslinking and purification of VP16C-TAND2-TBP complexes. Protein samples were resolved by SDS-PAGE and stained with Coomassie R-250. (A) Analysis of crosslinking reactions. Complexes of VP16C-TAND2-n-Cys and TBP were purified and treated with 250 μM $BM(PEO)_3$ for 45 min at 0 ºC. The 6xHis-tag on TBP was partially cleaved in this experiment, resulting in a pair of TBP bands. VP16C-TAND2-n-Cys is present, although it stained poorly with Coomassie R-250. (B) Preparation of the crosslinked complex. Lanes 1-3, lysates containing both proteins were mixed and purified over GSH resin. Lanes 4-5, thrombin digestion products and liberated proteins. Lane 6, products of crosslinking with $BM(PEO)_3$. Lane 7, crosslinked complex after size-exclusion purification and 5 day incubation at 20 ºC.

salt was required for sample clarity. Using this information, an initial NMR sample buffer was defined (10 mM Tris, 0.1 mM EDTA, 10 mM KCl, 8 mM NaN_3, pH 8.5).

To perform an initial structural analysis of the complex, a sample of cross-linked complex was prepared with ^{15}N labeling of the VP16C-TAND2 segment, the complex was concentrated to 0.64 mM, and 2-dimensional ^{1}H-^{15}N HSQC spectra were acquired from a freshly-prepared sample (Fig. 21). The spectra showed low overall dispersion in the proton dimension, suggesting the presence of some, but not many, structured elements in VP16C-TAND2 segment. Non-uniform peak intensities indicated different peptide backbone dynamics at different positions, suggesting a loose or incompletely folded structure. The combination of peak broadness and overall low dispersion indicated that resonance assignment would be difficult. Additionally, about 10 of the expected 70 peaks were very weak, suggesting that they might not be able to be assigned to particular residues in VP16C-TAND2.

To asses the stability of the NMR sample, it was incubated at 20 °C for 5 days, and the 2-dimensional HSQC spectrum was acquired again for comparison (Fig. 22). The spectra from the aged sample showed a loss of the broad, shifted peaks, and the acquisition of sharp, poorly dispersed peaks, indicating a loss of the structural elements seen in the fresh sample and a transition to an unfolded state. An SDS-PAGE analysis of the aged sample indicated that the primary structure of the complex was largely intact (Fig. 20B, lane 7). The aged sample was observed to scatter light, suggesting that larger particulates had formed, but attempts to sediment the sample at 16,000 x g did not diminish the protein

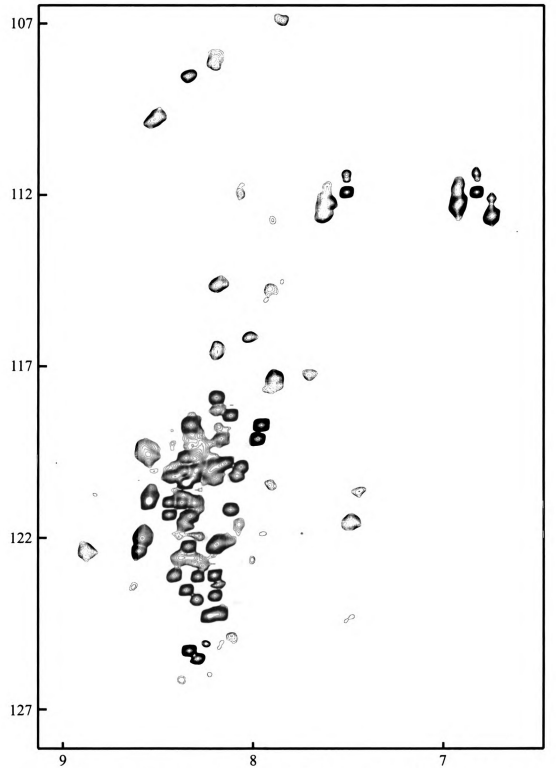


Figure 21: HSQC spectrum of fresh sample at 10 mM KCl. ^{15}N-labeled VP16C-TAND2 crosslinked to unlabeled TBP. Horizontal axis: proton chemical shift. Vertical axis: ^{15}N chemical shift.

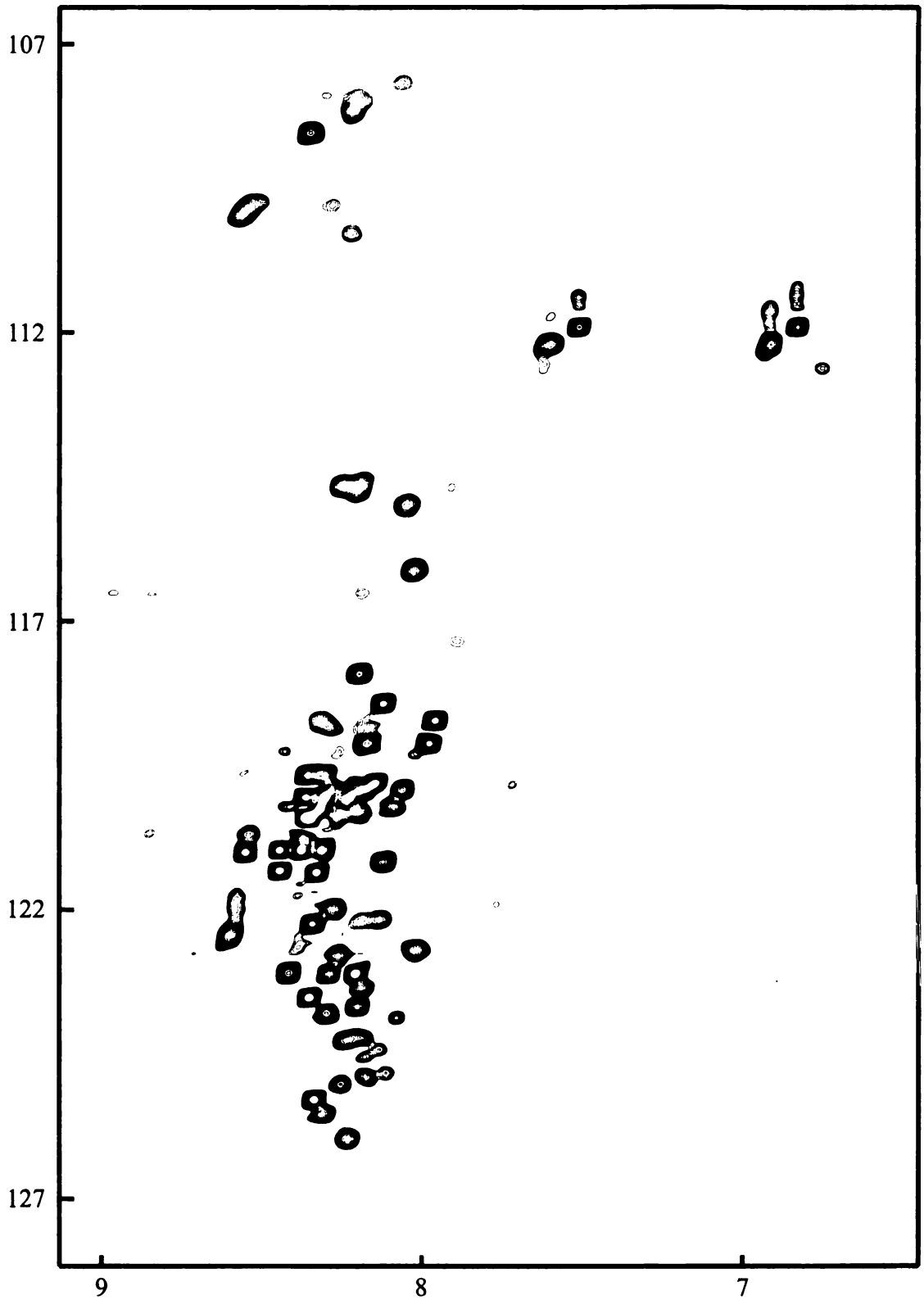


Figure 22: HSQC spectrum of aged sample at 10 mM KCl. ^{15}N-labeled VP16C-TAND2 crosslinked to unlabeled TBP. Horizontal axis: proton chemical shift. Vertical axis: ^{15}N chemical shift.

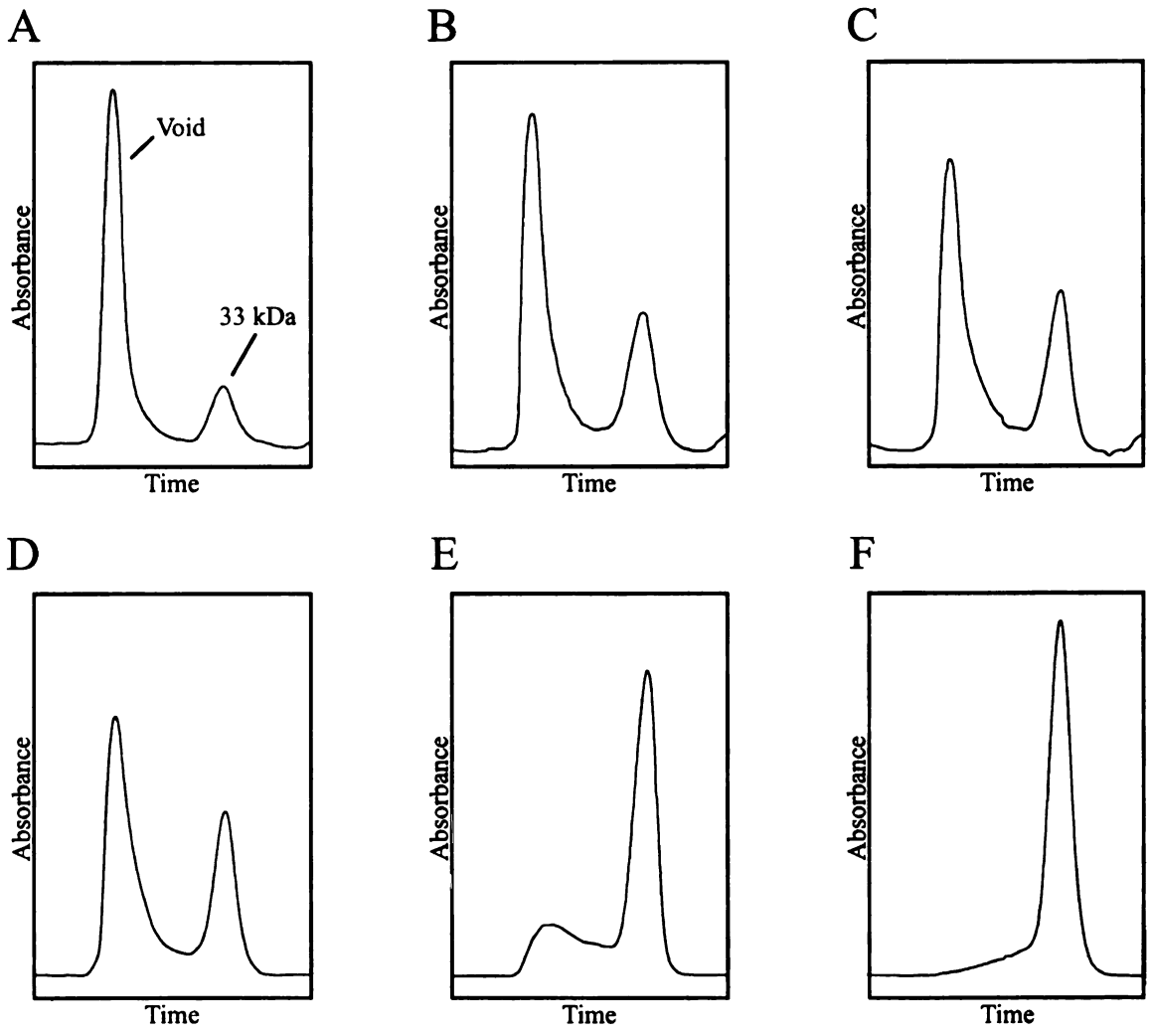
concentration. The aged sample was resolved by size exclusion chromatography, and about 75% of the protein was found to elute in the void fraction, indicating that most the apparently compact, monomeric prepared complex had decayed into either an unfolded or aggregated state (data not shown). The remaining 25% of the protein eluted at a position corresponding to TBP-linked dimers (Chapter 4), which indicated that the prepared monomeric complex had evolved into dimeric complexes. These results suggested that the sample may have followed a decomposition pathway involving a transition from a monomeric to a dimeric state, followed by subsequent unfolding or aggregation.

In attempts to stabilize the complex, the NMR buffer system was optimized. To assess sample stability, concentrated cross-linked protein complexes were dialyzed into different buffers and the protein stability was monitored by visual inspection of precipitation and turbidity. An NMR buffer successfully used for analysis of a similar complex of TBP and TAND1-TAND2 (20 mM Tris-HCl, 150 mM KCl, 10 mM $MgCl_2$, 5% glycerol, 0.5 mM AEBSF, 10 mM DTT, 0.05 mM NaN_3, pH 7.5) (102) was tested and found to not support stability of the VP16C-TAND2-TBP cross-linked complex. A family of buffers rooted on the precedent buffer showed that both lower salt and lower pH significantly improved sample stability. Additionally, removal of $MgCl_2$ greatly improved the stability of the sample. As the buffer optimization converged on improved conditions, the samples became visually indistinguishable, and thus size exclusion chromatography was employed to resolve the apparent molecular weight distributions of the complexes after incubation in different buffers. This analysis indicated that lowering the KCl from 40 mM to 10 mM led to significant improvements in maintenance of the compact,

monomeric state, and lowering the pH slightly also improved stability (Fig. 23A-C). However, at this stage of refinement, the complexes were not fully stable over the course of several days, precluding a practical NMR analysis.

In further attempts to stabilize the complex, adjustments to the protein design were considered. It was noted that two successful NMR structural analyses of TBP-containing complexes (99, 102) incorporated more of the N-terminus of TBP. Accordingly, the previous version of TBP (residues 61-240) was substituted with TBP (residues 40-240) and TBP (residues 49-240). SEC stability analysis of crosslinked complexes built with the longer versions of TBP indicated an improved maintenance of the compact, monomeric state, again with 10 mM KCl preferred over 40 mM (Fig. 23D,E). The results from TBP (residues 40-240) and TBP (residues 49-240) were indistinguishable, and thus TBP (residues 49-240) was utilized for further optimization. Since buffers with 10 mM KCl were significantly better than buffers with 40 mM KCl, buffers lacking KCl were tested, and found to be optimal (Fig. 23F). These results led to a final NMR buffer definition of 1 mM Tris, 0.1 mM EDTA, 0.02 mM KCl, 0.05 mM NaN3, pH 8.

A 2-dimensional HSQC spectrum was obtained from a cross-linked complex comprising TBP(residues 49-240)and ^{15}N-labeled VP16C-TAND2 (Fig. 24). A second spectrum acquired after the sample had been incubated for 2.5 days at 20 °C was essentially the same, indicating that the sample was structurally stable. A comparison to previous spectra (Figs. 21,22; Appendix A) showed new peaks dispersed from random-coil positions, indicating that the new conditions led to more structured elements. However,



Incubation conditions				
Panel	TBP	Buffer	pH	[KCl]
A	61-240	20 mM glycine	9.3	40 mM
B	61-240	20 mM glycine	9.3	10 mM
C	61-240	20 mM glycine	8.8	10 mM
D	49-240	20 mM glycine	9.0	40 mM
E	49-240	20 mM glycine	9.0	10 mM
F	49-240	1 mM Tris	8.0	0 mM

Figure 23: Optimization of buffer conditions. Samples of VP16C-TAND2-TBP cross-linked complex were purified, concentrated to approximately 1 mM total protein, and dialyzed into the indicated buffers. After 2.5 days of incubation at 20 °C, samples were resolved by size exclusion chromatography. The intact complex eluted at approximately 33 kDa.

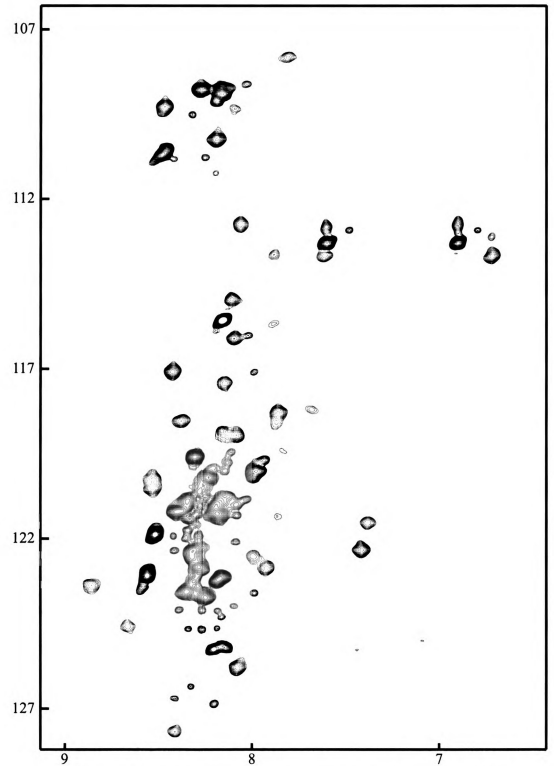


Figure 24: HSQC spectrum of crosslinked complex in low salt conditions. ^{15}N-labeled VP16C-TAND2 crosslinked to unlabeled TBP. Horizontal axis: proton chemical shift. Vertical axis: ^{15}N chemical shift.

the spectral quality was still low, with overall low dispersion, peaks of varied intensity, and spectral overlap. Based on the predicted difficulties in resonance assignment and structure determination, the pursuit of the NMR structure of the complex was not attempted.

A tabulation of HSQC peaks in the spectra acquired from the different samples indicated the existence of two distinct classes of structural elements (Appendix A). Among the shifted peaks in the stable (no salt) sample, a subset was still present in the folded but unstable sample (10 mM KCl). This result suggested that a core set of structural elements existed in the VP16C-TAND2 segment at 10 mM KCl, and that an extended set of structural elements was acquired at the lower-salt condition. The core set comprised about one third of the total HSQC peaks, and the extended set comprised another third of the peaks. While the sensitivity of the complex to salt suggested that the extended set of structural features arose from the lower salt concentration, it must be noted that stable sample conditions also differed by 0.5 pH units and the inclusion of 12 more N-terminal residues of TBP. Therefore, the extended set of structural features may also have arisen from pH dependence or interactions with TBP residues 49-61.

Since the NMR data obtained from the stable VP16C-TAND2-TBP cross-linked complex indicated few well-structured elements, different alterations to the structure of the complex were considered. It was recognized that the fusion of VP16C to TAND2 may have imposed unnatural constraints on VP16C. It was also possible that the fusion to TAND2 tethered VP16C in an unnatural orientation relative to TBP, which could have

led to binding imperfections and structural defects. As VP16C is normally attached to VP16N, an inclusion of both VP16N and VP16C was considered as an alternative to the VP16C-TAND2 fusion. VP16N has been shown to interact with TBP, and mutations that compromise VP16N transcriptional activation function also compromise the association with TBP (56). Additionally, although VP16N does not carry a TFIID-TFIIA-DNA assembly function, it does appear to contribute somewhat to this function when coupled with VP16C (76), suggesting that VP16N doesn't interfere with VP16C-TBP interactions. Taken together, these data suggested that VP16N may help VP16C interact with TBP, and a resulting complex might be more likely to preserve the native VP16C-TBP interface.

GST-VP16NC was tested as a substitute for GST-VP16C-TAND2 in the co-purification of TBP. GST-VP16NC retained TBP through binding and washing steps, allowing the purification of the immobilized complex (Fig. 25A). Thrombin digestion was used to liberate VP16NC and TBP from immobilized GST and the resulting proteins were further purified using size exclusion chromatography (Fig. 25B). The SEC peak fraction comprising VP16NC and TBP was concentrated and subjected to a second round of SEC, which resolved two peaks (Fig. 26A, first panel). This result indicated either that a complex of VP16NC and TBP was decomposing into separate proteins or that VP16NC and TBP existed as separate proteins that were not well resolved in the preparative SEC step.

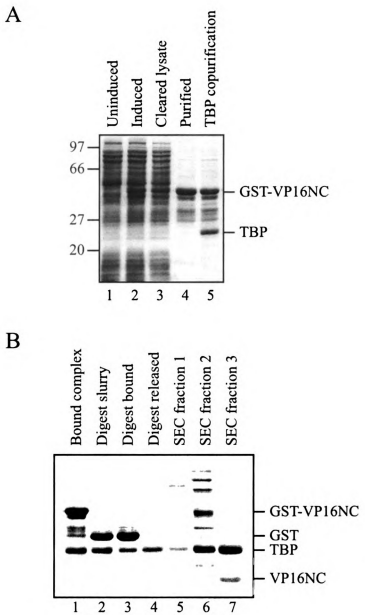
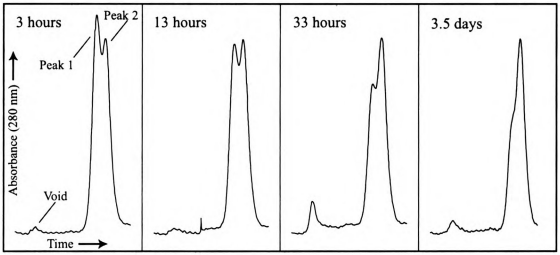


Figure 25: Co-purification of VP16NC and TBP. Protein samples were resolved by SDS-PAGE and stained with Coomassie R-250. (A) Lane 1-3, expression and solubility of GST-VP16NC. Lane 4, purification of GST-VP16NC over GSH resin. In lane 5, an *E. coli* lysate containing TBP was added to the lysate containing GST-VP16NC, and TBP was copurified. (B) Purification of VP16NC and TBP. Lane 1, lysates containing GST-VP16NC and TBP were mixed, incubated, and purified over glutathione resin. Lanes 2-3, thrombin digestion products and liberated proteins. Lanes 5-7, size exclusion chromatography fractions.

A



B

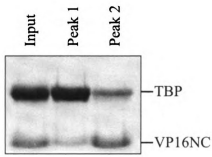


Figure 26: Analysis of the stability of a mixture of VP16NC and TBP. (A) A preparation of VP16NC and TBP was concentrated and incubated at 20°C for the indicated times, and then resolved by size exclusion chromatography. The protein mixture resolved into two peaks, one of which diminished over time. (B) Peak 1 and Peak 2 were isolated from (A), resolved by SDS-PAGE, and stained with Coomassie R-250.

To analyze the stability of the purified VP16NC and TBP proteins, purified samples were swapped into different buffers, incubated for 3.5 days at 20 °C, and resolved using SEC. Interestingly, all samples showed a dramatic loss of one of the two peaks and a gain in the void fraction, indicating that one of the two proteins was unstable. Additionally, low-salt conditions could not stabilize the mixture. SDS-PAGE analysis of the incubated samples showed that the primary structure of both VP16NC and TBP was intact, ruling out proteolysis and implying that either unfolding or aggregation events were affecting one of the proteins. An SEC analysis of samples as a function of time indicated that one of the peaks was stable, and that the other diminished significantly over a period of hours (Fig. 26A). The peaks were isolated and identified by SDS-PAGE (Fig. 26B), indicating that TBP was the unstable protein.

Discussion

Difficulties in the preparation of a fully-labeled VP16C-TAND2-linker-TBP complex, coupled with potential difficulties in NMR analysis of such a large protein, have led to the exploration of a segmentally-labeled, cross-linked complex. Conditions of very low salt concentration were found to stabilize this complex on time scales sufficient for NMR data acquisition. However, the peaks in the resulting HSQC spectra were poorly dispersed from random-coil positions, suggesting few well-structured elements in the VP16C-TAND2 segment. Additionally, variable peak intensities indicated different peptide backbone dynamics in different regions, suggesting that different regions underwent conformational changes on different time scales. This result is not surprising, particularly since TAND2 and VP16C are separate domains, and as such may fold independently. Overall, the combination of poor dispersion, broad peaks, and spectral overlap implied that the sequential assignment of the stabilized, cross-linked complex using ^{13}C labeling would be rather difficult and likely incomplete.

A comparison of spectra acquired from different states of the complex indicated two different classes of structure in VP16C-TAND2. Apparently, at low salt a core set of structural elements was present, and at even lower salt an extended set of structural elements formed. However, in the absence of resonance assignment, the distribution of these structural elements in VP16C and TAND2 cannot be determined.

The well-resolved spectrum from the unstructured state of VP16C-TAND2 is amenable to resonance assignment, which suggests an alternative approach to structural analysis. If the resonance peaks in the unstructured state were assigned to particular residues, then an accounting of peaks depleted in the structured state would identify which residues were involved in structural elements. Information obtained in this way could then be compared to data derived from other studies, to help determine if conserved sets of residues are involved in structural elements when VP16C interacts with different targets.

Indeed, there are signs that the regions of VP16AD that interact with TBP differ from those that interact with other targets. Time resolved fluorescence anisotropy and fluorescence quenching experiments have shown that VP16AD residues Phe 442 and Phe 473 become protected from solvent and conformationally constrained upon TBP binding (153). However, NMR studies have shown that the chemical shift of these residues changes little upon TAF9 binding (166), or TFIIB or PC4 binding (61). Additionally, surface plasmon resonance studies have identified VP16C residues critical for TBP interaction (120), but again these residues are only modestly shifted in the TAF9, PC4, and TFIIB NMR studies. These comparisons suggest that VP16 may fold in different ways when interacting with different partners. In this light, a determination of the VP16C residues involved in TBP interactions, as could be determined by assignment of the unstructured state of the VP16C-TAND-TBP complex, would provide valuable data for comparison to the structured regions involved in interactions with other targets.

Chapter VI

Discussion and Prospects for Further Study

Crystallization of a complex of VP16C and TBP

The majority of the research efforts discussed in this dissertation were aimed at determining the structure of the VP16C activation domain when bound to TBP. Since a crystal structure of the VP16C:TBP complex would provide the highest resolution model of VP16C, initial efforts were directed at crystallization of the complex. The apparent crystallization of TBP dimers in initial crystallization attempts led to the reasoning that a tighter complex was needed, and the GST-VP16C-TAND2 fusion protein was adopted for subsequent work. The strong association of this fusion protein with TBP allowed for the co-purification of GVT and TBP directly from mixed bacterial lysates, which simplified the purification scheme and facilitated a series of crystallization attempts. However, even this tight complex was found to eventually release TBP. In attempts to further tighten the complex, a fusion of VP16C-TAND2-linker-TBP was created and characterized. However, no crystals were obtained from this fusion protein.

The continued failure to obtain crystals of a protein complex representing the VP16C:TBP interface led us away from attempts at crystallization, and subsequent efforts were focused on NMR approaches (chapter 5). Importantly, in the efforts to stabilize a crosslinked variation of the VP16C-TAND2-TBP fusion protein, it was discovered that

the protein was unstable under many conditions, and that low salt conditions were required for the maintenance of a folded and monodisperse state. This finding implied that most of the crystallization work described in chapter 4 was performed with unstable protein.

The protein stability characteristics determined in chapter 5 could direct another series of crystallization experiments, with a focus on conditions supporting protein stability. Additionally, the NMR results from chapter 5 indicated that the VP16C-TAND2 peptide was not uniquely structured, but instead was somewhat disordered. The degree of this disorder may be temperature dependent. Furthermore, it has been noted that reduced temperature stabilizes the association of VP16C-TAND2 and TBP (figure C4). Therefore, crystallization attempts employing low salt and low temperatures would focus on the conditions most likely to provide overall stability and reduced structural disorder.

Additionally, the VP16C-TAND2-linker-TBP fusion protein may not represent the best target for continued crystallization trials. The glycine-serine linker peptide probably exists as a poorly structured loop, and as such may not provide for structured crystal lattice contacts. In the heterodimeric VP16C-TAND2:TBP complex this loop is absent and additional structured surfaces of TBP are exposed, suggesting that the heterodimeric complex is a better candidate for subsequent crystallization trails. If this complex is used for trials, it should first be determined whether the heterodimer can be stabilized on time scales suitable for crystallization.

Truncations of the VP16C and TAND2 peptide segments may remove unstructured termini and increase the likelihood of crystallization. It has been reported that a shorter version of VP16C (residues 470 to 490) was as effective as VP16C (residues 457-490) in supporting the tight binding of VP16C-TAND2 to TBP (83). Since VP16C residues 450-470 are not required for the TBP interaction, they may be present as an unstructured peptide segment, which may work against crystallization of the complex. C-terminal truncations of TAND2 should also be considered. Although TAND2 residues 42 to 73 have been included in our version of VP16C-TAND2, shorter versions of TAND2 may still support high affinity interactions with TBP. In this respect, it should be noted that a shorter TAND2 segment (residues 42-64) is sufficient for high-affinity interactions with TBP in the context of TAND1-TAND2 (79).

In the event that crystals of VP16C-TAND2:TBP complexes cannot be obtained, the tightly associated complex may provide for indirect approaches to structure determination. For example, one might search for pairs of cysteine residues spanning the VP16C-TBP interface that could participate in disulfide bonds. A panel of cysteine substitutions in VP16C could be tested with a panel of cysteine substitutions in TBP. Disulfide formation might indicate specific contacts between VP16C and TBP. Additionally, the resulting cross-linked species might be suitable targets for subsequent crystallization trials. A second possible approach is to tether a proteolytic agent such as FeBABE (111) to specific residues of either VP16C or TBP, and then map the resulting cleavage patterns to establish proximity of VP16C residues to TBP residues. However,

the FeBABE approach might be limited by the diffuse cleavage patterns generated by the diffusible hydroxyl radical.

NMR analysis of VP16C-TAND2 bound to TBP

The VP16C-TAND2-linker-TBP fusion protein developed in chapter 4 was found to be soluble enough for NMR analysis. To reduce the cost of the protein and the complexity of the spectra, a segmental labeling system was developed, which allowed selective isotopic labeling of VP16C-TAND2. Initial NMR spectra indicated that the complex had structural elements but was not stable on the several-day time scale required for NMR analysis. Optimization of the buffer conditions and protein design resulted in a protein complex stable enough for NMR analysis. In the resulting NMR spectra, poor dispersion in the proton dimension indicated a low overall degree of structure, overall broad peaks indicated conformational variability, and different peak intensities indicated different backbone dynamics in different regions. Additionally, the broad peaks and variable peak intensities suggested that assignment of the resonances to particular residues would be difficult and likely incomplete.

A comparison of the NMR spectra from structured and denatured complexes indicated that two different classes of structural elements existed, one present both in 10 mM KCl and in a no-salt buffer, and a second class present only in a no-salt buffer. It has not been determined how the residues of VP16C-TAND2 are distributed in these two structural

classes. However, since the NMR spectrum of the denatured state of VP16C-TAND2 is well resolved and thus amenable to resonance assignment, it would be possible to assign the resonances in the unstructured state, and then determine which peaks are absent in the different structured states. Although this approach would not provide details of secondary or tertiary structure, it would delineate the regions of VP16C that are involved in structural elements when bound to TBP. This information could be compared with data from other NMR analyses (61, 166) to determine if conserved regions of VP16C are involved in structural interactions with different binding partners.

TBP orientation

The report that the VP16 activation domain could enforce a productive orientation on an otherwise poorly oriented TBP:DNA complex was intriguing. We were interested in whether the TBP-orienting activity was related to transcriptional activation functions. To address this relationship, we had intended to utilize a panel of VP16AD variants to measure the correlation between in vitro orienting function and in vivo activation function. Additionally, the TBP orienting function implied a ternary interaction between VP16AD, TBP and DNA, but a body of other evidence argued against such a ternary interaction. We sought to explore this apparent contradiction.

Unexpectedly, and contrary to the precedent reports, we found that TBP could bind to the TATA sequence with a high degree of orientational specificity. Clear and internally-

consistent data from both DNA strands and a series of control experiments using reversed and symmetrical TATA sequences validated our experimental methods and indicated that the eight basepair TATA sequence was sufficient to direct highly oriented binding of TBP to DNA.

The finding that TBP can recognize the TATA sequence with a high degree of orientational specificity implies that current models of PIC assembly need to be rethought. In the current models, the ability of TBP to recognize the orientation of the TATA sequence is assumed to be modest, and other factors, such as TFIIA and TFIIB, are required to determine transcriptional directionality (50, 156). Our findings show that, to the contrary, the TBP:TATA connection might define transcriptional directionality in some cases, and thus the orientation of the TATA sequence may be a more prominent variable in core promoter structure than currently assumed.

Since other core promoter elements also can direct transcriptional orientation, questions about the interplay among directional signals arise. For example, if a forward orienting TFIIB recognition element (BRE) flanks a reverse-orienting TATA sequence, is a TBP-TFIIB-DNA complex significantly less likely to form? To what extent will it be oriented? Also, to what extent do the directional signals recognized by TFIID need to cooperate? For example, given an initiator sequence (Inr) and a downstream promoter element (DPE), how much will the orientation of the TATA sequence influence TFIID binding and promoter strength? One approach toward answering these questions is to

monitor PIC assembly on various core promoters in vitro using reconstituted systems (16).

In the interpretation of directional signals by the transcriptional machinery the order of events might be an important factor. For instance, consider TFIID recognition of a hypothetical promoter comprising forward-orienting Inr and DPE elements and a reversed TATA sequence. If TFIID were to recognize the Inr and DPE elements of this promoter before recognizing the TATA sequence, a TFIID:DNA complex might form in the forward-facing direction. Conversely, if the TBP:TATA interaction mediated the initial TFIID:DNA contacts, TFIID might bind the promoter in a reversed orientation. In this case, the reversed orientation of TFIID might prevent the recognition of the Inr and DPE sequences, leading to several possible outcomes, including stalled PIC assembly, reduced PIC stability, or even a fully functional PIC with a reverse orientation. Since core promoter elements may be dynamically masked and exposed by chromatin dynamics in vivo, it is possible that the ordered recognition of core promoter elements is a variable and possibly regulated process in transcription.

In our studies of TBP orientation we also determined that Gal4-VP16 did not noticeably alter the orientation of TPB on the AdMLP promoter sequence. This result is congruent with the observations that VP16AD, TBP, and DNA do not form the ternary connections that would be required for VP16AD to orient TBP on DNA. However, it remains possible that our studies did not detect increased TBP orientation because TBP was already maximally oriented. In this respect, it would be of value to explore the effects of

Gal4-VP16 on TBP orientation using the cleavage probes with reversed and symmetrical TATA sequences. This would allow for a larger analytical window in which to detect changes in TBP orientational specificity. Also, it should be noted that there are two distinct ways an activator might alter TBP orientation: by orienting TBP relative to the position of the activator, or by altering the intrinsic TATA-recognition properties of TBP. In the latter case, TATA sequences that can only partially direct TBP will be required to determine whether an activator can alter the intrinsic DNA recognition properties of TBP.

TBP purification

The high affinity interaction between VP16C-TAND2 and TBP has enabled the development of a novel method of TBP purification (chapter 2). This method is very fast and easy to perform, requires no extrinsic affinity tags, and the resulting TBP appears pure, structured, monodisperse, and functional. So far, this method has been applied to the purification of full-length and N-terminally truncated versions of yeast TBP. However, since both VP16C and TAND2 interact with the highly conserved core region of TBP, it is likely that the purification method will also work with TBP from other species. Indeed, preliminary data (not shown) suggests that this method can be used to purify the conserved core of human TBP. Tests with TBP from other species are planned.

Appendix A

HSQC spectra of VP16C-TAND2 complexed with TBP

Table 2: Intensity of NMR peaks in different spectra

HSQC peak intensity[a]															
Peak	**A**	**B**	**C**	**Peak**	**A**	**B**	**C**	**Peak**	**A**	**B**	**C**	**Peak**	**A**	**B**	**C**
1	3	3	0	26	3	3	1	51	3	3	1	76	3	0	1
2	3	3	1	27	3	3	3	52	0	3	0	77	2	0	1
3	3	3	3	28	3	1	1	53	3	3	2	78	1	0	1
4	3	3	3	29	3	3	1	54	1	3	3	79	2	0	2
5	3	3	0	30	3	3	1	55	1	3	3	80	0	0	3
6	3	3	1	31	3	3	1	56	3	2	1	81	0	0	1
7	3	3	1	32	3	3	1	57	3	3	3	82	0	0	3
8	3	2	1	33	3	3	1	58	3	3	3	83	0	0	1
9	3	3	1	34	3	3	1	59	3	3	1	84	0	0	3
10	3	3	1	35	3	2	2	60	2	2	3	85	0	0	3
11	3	3	1	36	3	3	1	61	0	3	3	86	0	0	3
12	0	3	3	37	3	?	?	62	0	1	3	87	0	0	3
13	2	2	1	38	3	?	?	63	1	2	1	88	0	0	3
14	1	2	3	39	3	3	2	64	0	2	1	89	0	0	3
15	3	3	1	40	3	3	1	65	0	2	1	90	0	0	3
16	1	2	3	41	1	3	3	66	0	2	3	91	0	1	2
17	0	1	0	42	3	1	1	67	1	1	3	92	0	0	3
18	3	1	2	43	3	2	1	68	3	2	3	93	0	0	3
19	2	1	0	44	3	3	0	69	3	1	2	94	0	0	1
20	1	3	3	45	0	1	3	70	3	1	0	95	0	0	3
21	0	1	0	46	0	3	3	71	3	0	0	96	2	1	2
22	0	3	3	47	3	3	1	72	1	0	0	97	3	0	0
23	0	3	3	48	2	3	3	73	1	0	0	98	3	1	1
24	3	3	1	49	3	3	2	74	2	0	0	99	0	2	3
25	3	3	1	50	3	3	1	75	3	1	1	100	2	1	1

A: Intensity of peak in HSQC spectrum of denatured sample at 10 mM KCl (Fig. 27).

B: Intensity of peak in HSQC spectrum of folded sample at 10 mM KCl (Fig. 28).

C: Intensity of peak in HSQC spectrum of folded sample at 0 mM KCl (Fig. 29).

[a] Intensity estimated by visual inspection of HSQC plots (Figs 27-29) placed on a scale from 0 (peak not present) to 3 (most intense).

Table 3: Comparison of NMR spectra

Comparison of NMR peak intensities in different spectra [a]					
10 mM KCl structured relative to 10 mM KCl unstructured			Low salt structured relative to 10 mM KCl unstructured		
Strong new peaks	Stronger common peaks	Weak new peaks	Strong new peaks	Stronger common peaks	Weak new peaks
12	14	17	12 85	14	64
22	16	21	22 86	16	65
23	20	62	23 87	20	81
46	41	64	45 88	41	83
52	48	65	46 89	48	91
61	54	66	61 90	54	94
	55	91	62 92	55	
	63	99	66 93	60	
			80 95	67	
			82 99		
			84		

[a] See Table 2 for tabulation of NMR peak intensities, and Figures 27-29 for peak locations.

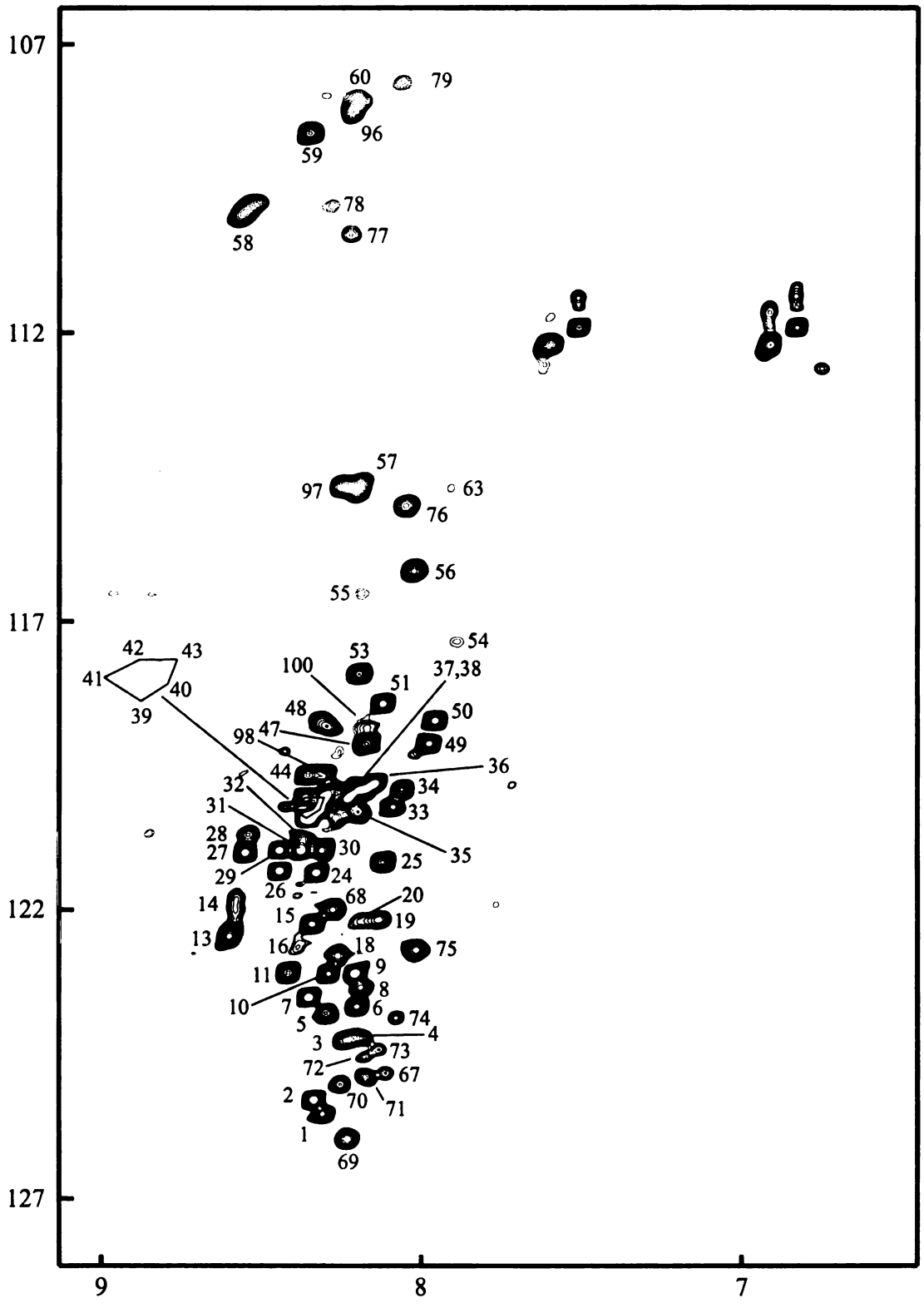


Figure 27: HSQC spectrum of aged sample at 10 mM KCl. ^{15}N-labeled VP16C-TAND2 crosslinked to unlabeled TBP. Horizontal axis: proton chemical shift. Vertical axis: ^{15}N chemical shift.

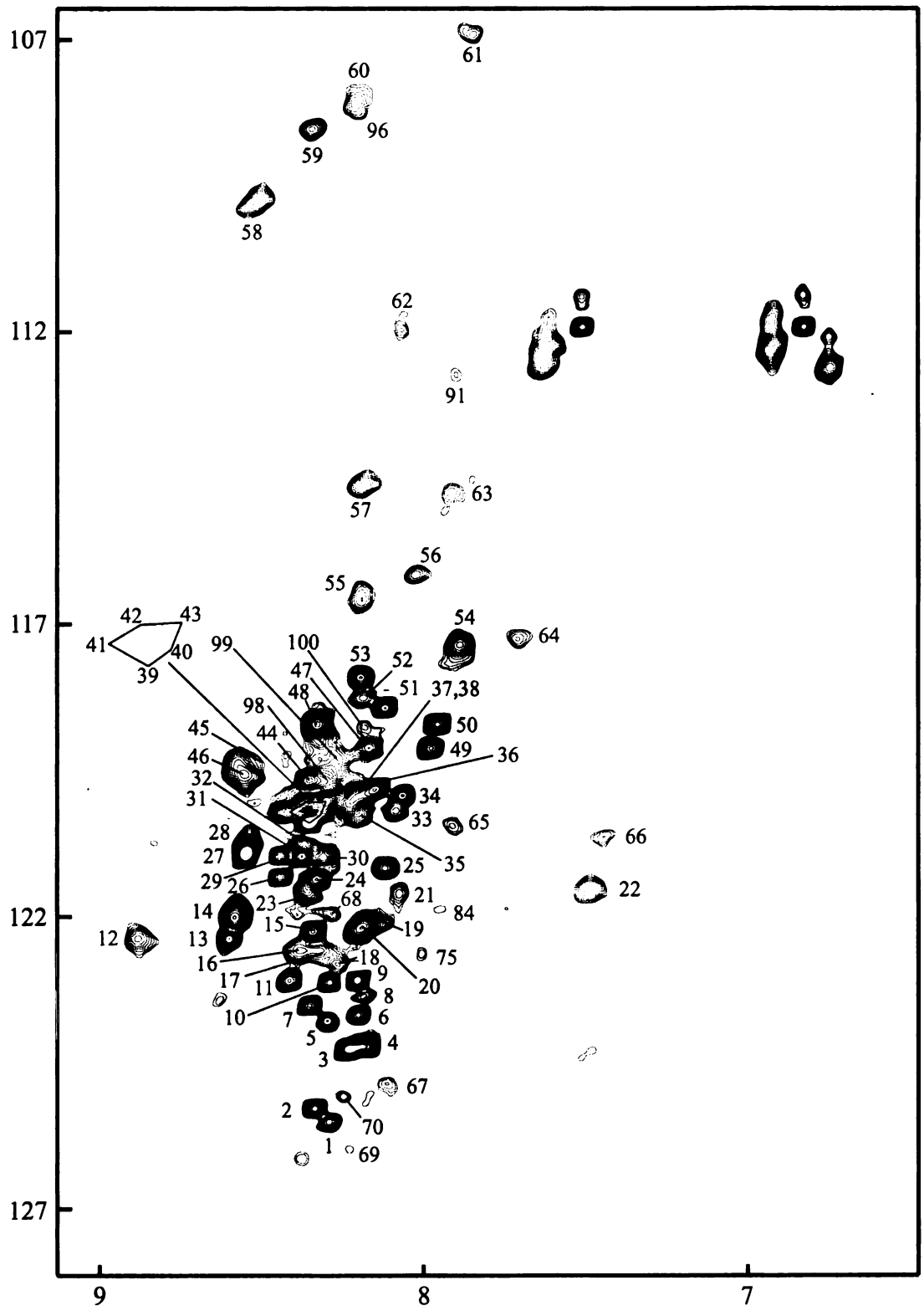


Figure 28: HSQC spectrum of fresh sample at 10 mM KCl. ^{15}N-labeled VP16C-TAND2 crosslinked to unlabeled TBP. Horizontal axis: proton chemical shift. Vertical axis: ^{15}N chemical shift.

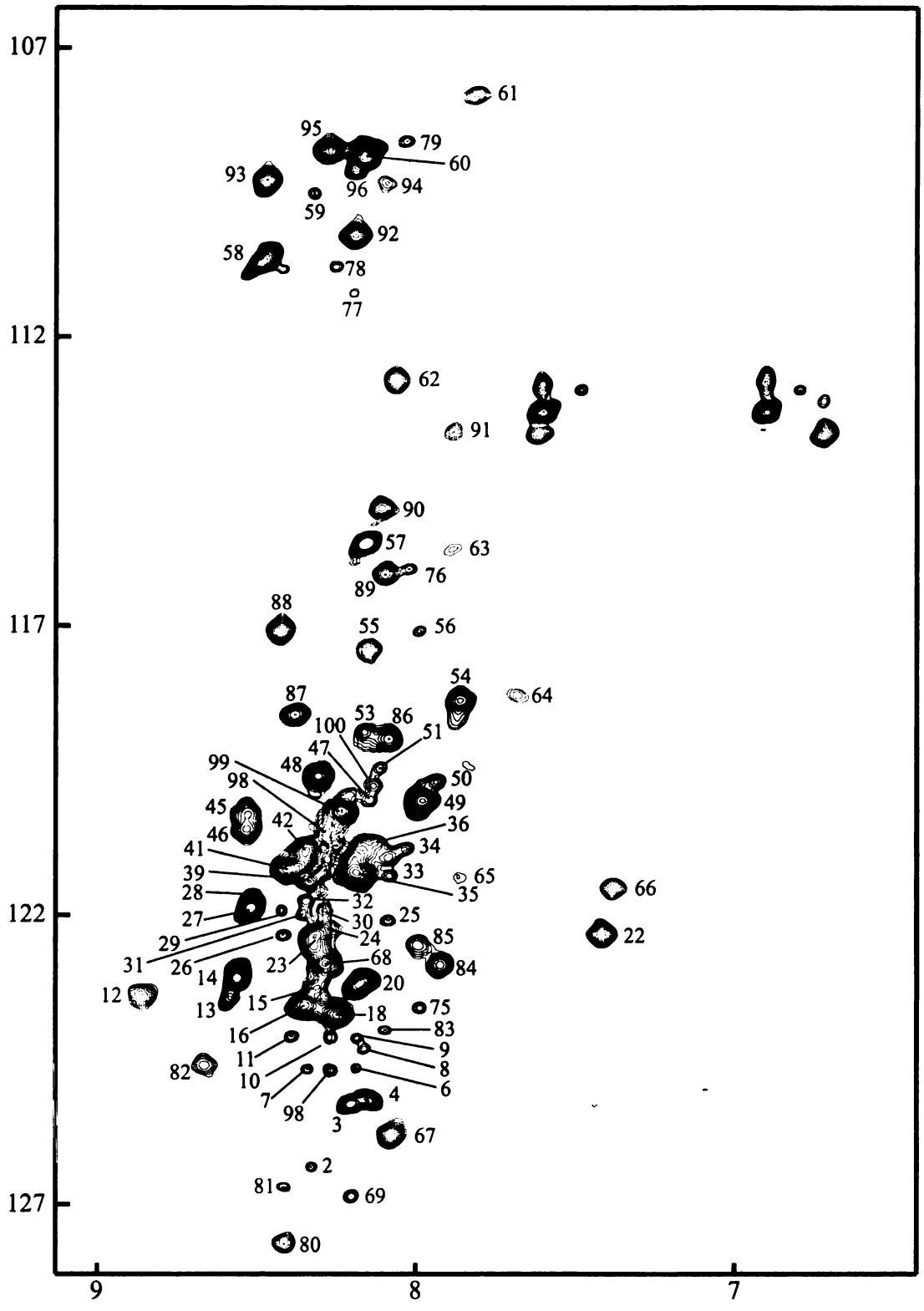


Figure 29: HSQC spectrum of crosslinked complex in low salt conditions. ^{15}N-labeled VP16C-TAND2 crosslinked to unlabeled TBP. Horizontal axis: proton chemical shift. Vertical axis: ^{15}N chemical shift.

Appendix B

Plasmid maps and oligonucleotide sequence

Table 4: Oligonucleotide sequences

Name	Sequence	Matches	Direction
DS1	CATGGGTTGCGGCTCTG		F
DS2	GATCCAGAGCCGCAACC		R
DS3	GCCCCGCCATATGGCTGCCCCAGAATCTG	yTBP 49-54	F
DS4	GCGCCGCCATATGTTCCAGAGTGAAGAGG	yTBP 40-45	F
ST394	see Triezenberg lab folder	pET28a downstream	R
ST803	GCGGGATCCCCGGGTCCGG	VP16 451-456	F
ST804	ATGAATTCGCACCCCCACCGTACTCGTCAATTCCA	VP16 483-490	R
ST805	GGCGGAATTCAAAGGACTATACGGAGCAT	yTAF1 43-50	F
ST806	TATCTCGAGCTATCATTCTTCTGGCAAATCGTCATCGTC	yTAF1 66-73	R
ST903	TCATACACATACGATTTAGGTGACA	pG5MLT upstream	F
ST904	GAAGAGGAGAAGATAATAGGAGGAA	pG5MLT downstream	R
ST903-6FAM	TCATACACATACGATTTAGGTGACA	pG5MLT upstream	F
ST904-HEX	AGAAGAGGAGAAGATAATAGGAGGAA	pG5MLT downstream	R
ST906	GGATACCATGGGTTCCCCGGGTCCGG	VP16 451-456	F
ST907	CGCGCGGATCCTTCTTCTGGCAAATCGTCATC	yTAF1 67-73	R
ST908	GATCAGGCTCTG		F
ST909	GATCCAGAGCCT		R
ST925	GATCTAACCTGCACCCCAAAG	yTBP 75-82	R
ST933	GCCCCGCCATATGTCCGGTATTGTTCCA	yTBP 61-65	F
ST1033	AATTCCTGTTAGAGATCTTGCTGATAGC		F
ST1034	TCGAGCTATCAGCAAGATCTCTAACAGG		R
ST1130	CCTGAAGGGGGGCTATATATAGGGGTGGGGGCGCG	AdMLP TATA	F
ST1131	CGCGCCCCCACCCCTATATATAGCCCCCCTTCAGG	AdMLP TATA	R
ST1132	CCTGAAGGGGGGCCTTTTATAGGGGTGGGGGCGCG	AdMLP TATA	F
ST1133	CGCGCCCCCACCCCTATAAAAGGCCCCCCTTCAGG	AdMLP TATA	R

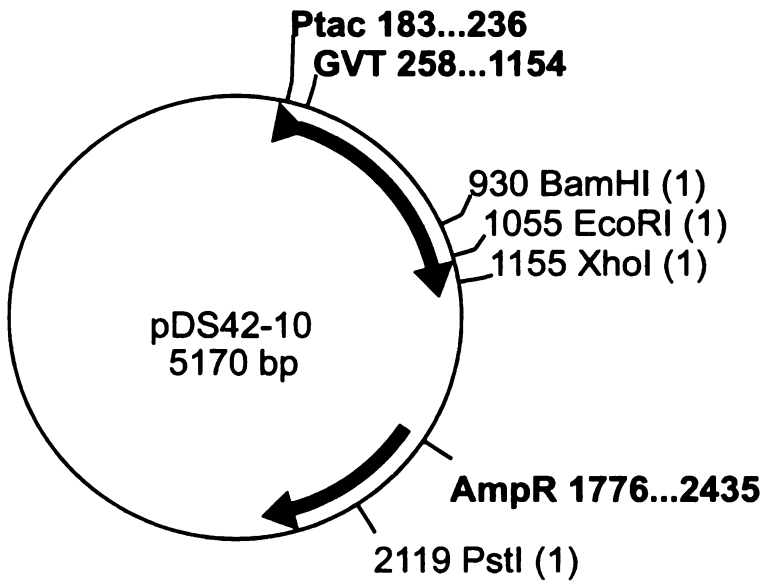


Figure 30: Plasmid pDS42-10.

Parent plasmid: pGEX4T-1 (GE Healthcare)

Inserted sequences: VP16C between *Bam*HI-*Eco*RI
TAND2 between *Eco*RI and *Xho*I

```
          BamHI
            |
CTGGTTCCGCGTGGATCCCCGGGTCCGGGATTTACCCCCCACGACTCCGCCCCCTACGGC
 L  V  P  R  G  S  P  G  P  G  F  T  P  H  D  S  A  P  Y  G

GCTCTGGATATGGCCGACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGGAATTGAC
 A  L  D  M  A  D  F  E  F  E  Q  M  F  T  D  A  L  G  I  D
               EcoRI
                 |
GAGTACGGTGGGGGTGCGAATTCAAAGGACTATACGGAGCATTTGCCGGATGCTGTAGAT
 E  Y  G  G  G  A  N  S  K  D  Y  T  E  H  L  P  D  A  V  D
                                                        XhoI
                                                         |
TTTGAAGATGAAGATGAACTTGCTGATGACGATGACGATTTGCCAGAAGAATGATAGCTC
 F  E  D  E  D  E  L  A  D  D  D  D  D  L  P  E  E  *  *
     NotI
      |
GAGCGGCCGC
```

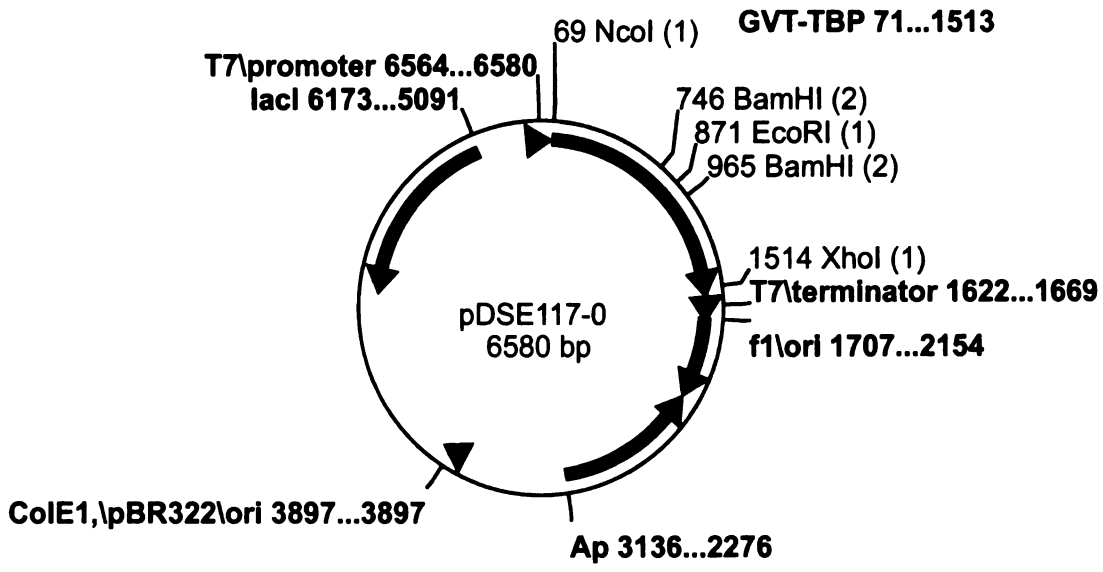


Figure 31: Plasmid family pDSE117-n.

Parent plasmid: pETDUET-1 (Novagen)

Inserted sequences: GST between *Nco*I and *Bam*HI
VP16C between *Bam*HI-*Eco*RI
TAND2 between *Eco*RI and *Bam*HI
TBP between *Bam*HI and *Xho*I

Note: In this plasmid family, n indicates the number of insertions of the sequence GGATCAGGCTCT upstream from the BamHI site between TAND2 and TBP.

List of references

1. **Almlof, T., A. E. Wallberg, J. A. Gustafsson, and A. P. Wright.** 1998. Role of important hydrophobic amino acids in the interaction between the glucocorticoid receptor tau 1-core activation domain and target factors. Biochemistry **37:**9586-94.

2. **Auboeuf, D., A. Honig, S. M. Berget, and B. W. O'Malley.** 2002. Coordinate regulation of transcription and splicing by steroid receptor coregulators. Science **298:**416-9.

3. **Bagby, S., T. K. Mal, D. Liu, E. Raddatz, Y. Nakatani, and M. Ikura.** 2000. TFIIA-TAF regulatory interplay: NMR evidence for overlapping binding sites on TBP. FEBS Lett **468:**149-54.

4. **Bajic, V. B., S. L. Tan, A. Chong, S. Tang, A. Strom, J. A. Gustafsson, C. Y. Lin, and E. T. Liu.** 2003. Dragon ERE Finder version 2: A tool for accurate detection and analysis of estrogen response elements in vertebrate genomes. Nucleic Acids Res **31:**3605-7.

5. **Barboric, M., R. M. Nissen, S. Kanazawa, N. Jabrane-Ferrat, and B. M. Peterlin.** 2001. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Mol Cell **8:**327-37.

6. **Beerli, R. R., and C. F. Barbas, 3rd.** 2002. Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol **20:**135-41.

7. **Bell, S. D., P. L. Kosa, P. B. Sigler, and S. P. Jackson.** 1999. Orientation of the transcription preinitiation complex in Archaea. Proc Natl Acad Sci U S A **96:**13662-13667.

8. **Berger, S. L.** 2002. Histone modifications in transcriptional regulation. Curr Opin Genet Dev **12:**142-8.

9. **Bleichenbacher, M., S. Tan, and T. J. Richmond.** 2003. Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. J Mol Biol **332:**783-93.

10. **Boardman, P. E., S. G. Oliver, and S. J. Hubbard.** 2003. SiteSeer: Visualisation and analysis of transcription factor binding sites in nucleotide sequences. Nucleic Acids Res **31:**3572-5.

11. **Boube, M., L. Joulia, D. L. Cribbs, and H. M. Bourbon.** 2002. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. Cell **110:**143-51.

12. **Brooks, C. L., and W. Gu.** 2003. Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Curr Opin Cell Biol **15:**164-71.

13. **Brown, S. A., A. N. Imbalzano, and R. E. Kingston.** 1996. Activator-dependent regulation of transcriptional pausing on nucleosomal templates. Genes Dev **10:**1479-90.

14. **Brown, S. A., C. S. Weirich, E. M. Newton, and R. E. Kingston.** 1998. Transcriptional activation domains stimulate initiation and elongation at different times and via different residues. Embo J **17:**3146-54.

15. **Bucher, P.** 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol **212:**563-78.

16. **Buratowski, S., S. Hahn, L. Guarente, and P. A. Sharp.** 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell **56:**549-61.

17. **Burley, S. K., and R. G. Roeder.** 1996. Biochemistry and structural biology of transcription factor IID (TFIID). Annu Rev Biochem **65:**769-99.

18. **Butler, J. E., and J. T. Kadonaga.** 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev **15:**2515-9.

19. **Candau, R., D. M. Scolnick, P. Darpino, C. Y. Ying, T. D. Halazonetis, and S. L. Berger.** 1997. Two tandem and independent sub-activation domains in the amino terminus of p53 require the adaptor complex for activity. Oncogene **15:**807-16.

20. **Chasman, D. I., K. M. Flaherty, P. A. Sharp, and R. D. Kornberg.** 1993. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. Proc Natl Acad Sci U S A **90:**8174-8.

21. **Chiang, C. M., and R. G. Roeder.** 1995. Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science **267:**531-6.

22. **Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski.** 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev **15:**3319-29.

23. **Coleman, R. A., A. K. Taggart, L. R. Benjamin, and B. F. Pugh.** 1995. Dimerization of the TATA binding protein. J Biol Chem **270:**13842-9.

24. **Corbi, N., M. Perez, R. Maione, and C. Passananti.** 1997. Synthesis of a new zinc finger peptide; comparison of its 'code' deduced and 'CASTing' derived binding sites. FEBS Lett **417:**71-4.

25. **Cormack, B. P., and K. Struhl.** 1992. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. Cell **69:**685-96.

26. **Cox, J. M., M. M. Hayward, J. F. Sanchez, L. D. Gegnas, S. van der Zee, J. H. Dennis, P. B. Sigler, and A. Schepartz.** 1997. Bidirectional binding of the TATA box binding protein to the TATA box. Proc Natl Acad Sci U S A **94:**13475-80.

27. **Cramer, P., C. G. Pesce, F. E. Baralle, and A. R. Kornblihtt.** 1997. Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci U S A **94:**11456-60.

28. **Cress, W. D., and S. J. Triezenberg.** 1991. Critical structural elements of the VP16 transcriptional activation domain. Science **251:**87-90.

29. **Dahlman-Wright, K., H. Baumann, I. J. McEwan, T. Almlof, A. P. Wright, J. A. Gustafsson, and T. Hard.** 1995. Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. Proc Natl Acad Sci U S A **92:**1699-703.

30. **Donaldson, L., and J. P. Capone.** 1992. Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. J Biol Chem **267:**1411-4.

31. **Dynlacht, B. D., T. Hoey, and R. Tjian.** 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell **66:**563-76.

32. **Emami, K. H., W. W. Navarre, and S. T. Smale.** 1995. Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. Mol Cell Biol **15:**5906-16.

33. **Evans, R., J. A. Fairley, and S. G. Roberts.** 2001. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. Genes Dev **15:**2945-9.

34. **Falke, D., and R. L. Juliano.** 2003. Selective gene regulation with designed transcription factors: implications for therapy. Curr Opin Mol Ther **5:**161-6.

35. **Felinski, E. A., and P. G. Quinn.** 1999. The CREB constitutive activation domain interacts with TATA-binding protein-associated factor 110 (TAF110) through specific hydrophobic residues in one of the three subdomains required for both activation and TAF110 binding. J Biol Chem **274:**11672-8.

36. **Ferdous, A., T. Kodadek, and S. A. Johnston.** 2002. A nonproteolytic function of the 19S regulatory subunit of the 26S proteasome is required for efficient activated transcription by human RNA polymerase II. Biochemistry **41:**12798-805.

37. **Garvie, C. W., and C. Wolberger.** 2001. Recognition of specific DNA sequences. Mol Cell **8:**937-46.

38. **Ge, K., M. Guermah, C. X. Yuan, M. Ito, A. E. Wallberg, B. M. Spiegelman, and R. G. Roeder.** 2002. Transcription coactivator TRAP220 is required for PPAR gamma 2-stimulated adipogenesis. Nature **417:**563-7.

39. **Geiger, J. H., S. Hahn, S. Lee, and P. B. Sigler.** 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. Science **272:**830-6.

40. **Ghosh, S., and M. Karin.** 2002. Missing pieces in the NF-kappaB puzzle. Cell **109 Suppl:**S81-96.

41. **Gill, G., E. Pascal, Z. H. Tseng, and R. Tjian.** 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. Proc Natl Acad Sci U S A **91:**192-6.

42. **Gonzalez, F., A. Delahodde, T. Kodadek, and S. A. Johnston.** 2002. Recruitment of a 19S proteasome subcomplex to an activated promoter. Science **296:**548-50.

43. **Goodrich, J. A., T. Hoey, C. J. Thut, A. Admon, and R. Tjian.** 1993. Drosophila TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. Cell **75:**519-30.

44. **Gossen, M., and H. Bujard.** 2002. Studying gene function in eukaryotes by conditional gene inactivation. Annu Rev Genet **36:**153-73.

45. **Gossen, M., and H. Bujard.** 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A **89:**5547-51.

46. **Gossen, M., S. Freundlieb, G. Bender, G. Muller, W. Hillen, and H. Bujard.** 1995. Transcriptional activation by tetracyclines in mammalian cells. Science **268:**1766-9.

47. **Grabe, N.** 2002. AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol **2:**S1-15.

48. **Grant, P. A., D. Schieltz, M. G. Pray-Grant, D. J. Steger, J. C. Reese, J. R. Yates, 3rd, and J. L. Workman.** 1998. A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. Cell **94:**45-53.

49. **Greaves, R., and P. O'Hare.** 1989. Separation of requirements for protein-DNA complex assembly from those for functional activity in the herpes simplex virus regulatory protein Vmw65. J Virol **63:**1641-50.

50. **Hahn, S.** 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol **11**:394-403.

51. **Hampsey, M., and D. Reinberg.** 1999. RNA polymerase II as a control panel for multiple coactivator complexes. Curr Opin Genet Dev **9**:132-9.

52. **Hayashi, F., R. Ishima, D. Liu, K. I. Tong, S. Kim, D. Reinberg, S. Bagby, and M. Ikura.** 1998. Human general transcription factor TFIIB: conformational variability and interaction with VP16 activation domain. Biochemistry **37**:7941-51.

53. **Heery, D. M., E. Kalkhoven, S. Hoare, and M. G. Parker.** 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature **387**:733-6.

54. **Hengartner, C. J., C. M. Thompson, J. Zhang, D. M. Chao, S. M. Liao, A. J. Koleske, S. Okamura, and R. A. Young.** 1995. Association of an activator with an RNA polymerase II holoenzyme. Genes Dev **9**:897-910.

55. **Iizuka, M., and M. M. Smith.** 2003. Functional consequences of histone modifications. Curr Opin Genet Dev **13**:154-60.

56. **Ingles, C. J., M. Shales, W. D. Cress, S. J. Triezenberg, and J. Greenblatt.** 1991. Reduced binding of TFIID to transcriptionally compromised mutants of VP16. Nature **351**:588-90.

57. **Ito, M., C. X. Yuan, S. Malik, W. Gu, J. D. Fondell, S. Yamamura, Z. Y. Fu, X. Zhang, J. Qin, and R. G. Roeder.** 1999. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. Mol Cell **3**:361-70.

58. **Jackson, B. M., C. M. Drysdale, K. Natarajan, and A. G. Hinnebusch.** 1996. Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. Mol Cell Biol **16**:5557-71.

59. **Jenuwein, T., and C. D. Allis.** 2001. Translating the histone code. Science **293**:1074-80.

60. **Joliot, V., M. Demma, and R. Prywes.** 1995. Interaction with RAP74 subunit of TFIIF is required for transcriptional activation by serum response factor. Nature **373**:632-5.

61. **Jonker, H. R., R. W. Wechselberger, R. Boelens, G. E. Folkers, and R. Kaptein.** 2005. Structural properties of the promiscuous VP16 activation domain. Biochemistry **44**:827-39.

62. **Juo, Z. S., T. K. Chiu, P. M. Leiberman, I. Baikalov, A. J. Berk, and R. E. Dickerson.** 1996. How proteins recognize the TATA box. J Mol Biol **261**:239-54.

63. **Kadonaga, J. T.** 2002. The DPE, a core promoter element for transcription by RNA polymerase II. Exp Mol Med **34:**259-64.

64. **Kamiuchi, T., E. Abe, M. Imanishi, T. Kaji, M. Nagaoka, and Y. Sugiura.** 1998. Artificial nine zinc-finger peptide with 30 base pair binding sites. Biochemistry **37:**13827-34.

65. **Kanazawa, S., T. Okamoto, and B. M. Peterlin.** 2000. Tat competes with CIITA for the binding to P-TEFb and blocks the expression of MHC class II genes in HIV infection. Immunity **12:**61-70.

66. **Kanazawa, S., L. Soucek, G. Evan, T. Okamoto, and B. M. Peterlin.** 2003. c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. Oncogene **22:**5707-11.

67. **Kays, A. R., and A. Schepartz.** 2002. Gal4-VP16 and Gal4-AH increase the orientational and axial specificity of TATA box recognition by TATA box binding protein. Biochemistry **41:**3147-55.

68. **Kays, A. R., and A. Schepartz.** 2000. Virtually unidirectional binding of TBP to the AdMLP TATA box within the quaternary complex with TFIIA and TFIIB. Chem Biol **7:**601-610.

69. **Kel, A. E., E. Gossling, I. Reuter, E. Cheremushkin, O. V. Kel-Margoulis, and E. Wingender.** 2003. MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res **31:**3576-9.

70. **Kim, J. L., D. B. Nikolov, and S. K. Burley.** 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature **365:**520-7.

71. **Kim, L. J., A. G. Seto, T. N. Nguyen, and J. A. Goodrich.** 2001. Human Taf(II)130 is a coactivator for NFATp. Mol Cell Biol **21:**3503-13.

72. **Kim, T. K., S. Hashimoto, R. J. Kelleher, 3rd, P. M. Flanagan, R. D. Kornberg, M. Horikoshi, and R. G. Roeder.** 1994. Effects of activation-defective TBP mutations on transcription initiation in yeast. Nature **369:**252-5.

73. **Kim, Y., J. H. Geiger, S. Hahn, and P. B. Sigler.** 1993. Crystal structure of a yeast TBP/TATA-box complex. Nature **365:**512-20.

74. **Kim, Y. J., S. Bjorklund, Y. Li, M. H. Sayre, and R. D. Kornberg.** 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell **77:**599-608.

75. **Klemm, R. D., J. A. Goodrich, S. Zhou, and R. Tjian.** 1995. Molecular cloning and expression of the 32-kDa subunit of human TFIID reveals interactions with VP16 and TFIIB that mediate transcriptional activation. Proc Natl Acad Sci U S A **92:**5788-92.

76. **Kobayashi, N., T. G. Boyer, and A. J. Berk.** 1995. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. Mol Cell Biol **15:**6465-73.

77. **Kobayashi, N., P. J. Horn, S. M. Sullivan, S. J. Triezenberg, T. G. Boyer, and A. J. Berk.** 1998. DA-complex assembly activity required for VP16C transcriptional activation. Mol Cell Biol **18:**4023-31.

78. **Kokubo, T., D. W. Gong, S. Yamashita, M. Horikoshi, R. G. Roeder, and Y. Nakatani.** 1993. Drosophila 230-kD TFIID subunit, a functional homolog of the human cell cycle gene product, negatively regulates DNA binding of the TATA box-binding subunit of TFIID. Genes Dev **7:**1033-46.

79. **Kokubo, T., M. J. Swanson, J. I. Nishikawa, A. G. Hinnebusch, and Y. Nakatani.** 1998. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. Mol Cell Biol **18:**1003-12.

80. **Kokubo, T., S. Yamashita, M. Horikoshi, R. G. Roeder, and Y. Nakatani.** 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 U S A **91:**3520-4.

81. **Koleske, A. J., and R. A. Young.** 1994. An RNA polymerase II holoenzyme responsive to activators. Nature **368:**466-9.

82. **Kosa, P. F., G. Ghosh, B. S. DeDecker, and P. B. Sigler.** 1997. The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. Proc Natl Acad Sci U S A **94:**6042-7.

83. **Kotani, T., K. Banno, M. Ikura, A. G. Hinnebusch, Y. Nakatani, M. Kawaichi, and T. Kokubo.** 2000. A role of transcriptional activators as antirepressors for the autoinhibitory activity of TATA box binding of transcription factor IID. Proc Natl Acad Sci U S A **97:**7178-83.

84. **Kotani, T., T. Miyake, Y. Tsukihashi, A. G. Hinnebusch, Y. Nakatani, M. Kawaichi, and T. Kokubo.** 1998. Identification of highly conserved amino-terminal segments of dTAFII230 and yTAFII145 that are functionally interchangeable for inhibiting TBP-DNA interactions in vitro and in promoting yeast cell growth in vivo. J Biol Chem **273:**32254-64.

85. **Kou, H., J. D. Irvin, K. L. Huisinga, M. Mitra, and B. F. Pugh.** 2003. Structural and functional analysis of mutations along the crystallographic dimer interface of the yeast TATA binding protein. Mol Cell Biol **23:**3186-201.

86. **Kulkarni, M. M., and D. N. Arnosti.** 2003. Information display by transcriptional enhancers. Development **130:**6569-75.

87. **Kumar, K. P., S. Akoulitchev, and D. Reinberg.** 1998. Promoter-proximal stalling results from the inability to recruit transcription factor IIH to the transcription complex and is a regulated event. Proc Natl Acad Sci U S A **95:**9767-72.

88. **Kussie, P. H., S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich.** 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science **274:**948-53.

89. **Lau, J. F., I. Nusinzon, D. Burakov, L. P. Freedman, and C. M. Horvath.** 2003. Role of metazoan mediator proteins in interferon-responsive transcription. Mol Cell Biol **23:**620-8.

90. **Lee, C., J. H. Chang, H. S. Lee, and Y. Cho.** 2002. Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor. Genes Dev **16:**3199-212.

91. **Lee, T. I., and R. A. Young.** 2000. Transcription of eukaryotic protein-coding genes. Annu Rev Genet **34:**77-137.

92. **Lewis, B. A., and D. Reinberg.** 2003. The mediator coactivator complex: functional and physical roles in transcriptional regulation. J Cell Sci **116:**3667-75.

93. **Lieberman, P. M., and A. J. Berk.** 1994. A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA--promoter DNA complex formation. Genes Dev **8:**995-1006.

94. **Liljelund, P., C. J. Ingles, and J. Greenblatt.** 1993. Altered promoter binding of the TATA box-binding factor induced by the transcriptional activation domain of VP16 and suppressed by TFIIA. Mol Gen Genet **241:**694-9.

95. **Lin, Q., C. F. Barbas, 3rd, and P. G. Schultz.** 2003. Small-molecule switches for zinc finger transcription factors. J Am Chem Soc **125:**612-3.

96. **Lin, Y. S., I. Ha, E. Maldonado, D. Reinberg, and M. R. Green.** 1991. Binding of general transcription factor TFIIB to an acidic activating region. Nature **353:**569-71.

97. **Lipford, J. R., and R. J. Deshaies.** 2003. Diverse roles for ubiquitin-dependent proteolysis in transcriptional activation. Nat Cell Biol **5:**845-50.

98. **Littlefield, O., Y. Korkhin, and P. B. Sigler.** 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci U S A **96:**13668-13673.

99. **Liu, D., R. Ishima, K. I. Tong, S. Bagby, T. Kokubo, D. R. Muhandiram, L. E. Kay, Y. Nakatani, and M. Ikura.** 1998. Solution structure of a TBP-TAF(II)230

complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. Cell **94:**573-83.

100. **Liu, P. Q., E. J. Rebar, L. Zhang, Q. Liu, A. C. Jamieson, Y. Liang, H. Qi, P. X. Li, B. Chen, M. C. Mendel, X. Zhong, Y. L. Lee, S. P. Eisenberg, S. K. Spratt, C. C. Case, and A. P. Wolffe.** 2001. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. J Biol Chem **276:**11323-34.

101. **Luscombe, N. M., and J. M. Thornton.** 2002. Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. J Mol Biol **320:**991-1009.

102. **Mal, T. K., Y. Masutomi, L. Zheng, Y. Nakata, H. Ohta, Y. Nakatani, T. Kokubo, and M. Ikura.** 2004. Structural and functional characterization on the interaction of yeast TFIID subunit TAF1 with TATA-binding protein. J Mol Biol **339:**681-93.

103. **Malik, S., and R. G. Roeder.** 2000. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. Trends Biochem Sci **25:**277-83.

104. **Maniatis, T., and R. Reed.** 2002. An extensive network of coupling among gene expression machines. Nature **416:**499-506.

105. **Marmorstein, R., and M. X. Fitzgerald.** 2003. Modulation of DNA-binding domains for sequence-specific DNA recognition. Gene **304:**1-12.

106. **Martel, L. S., H. J. Brown, and A. J. Berk.** 2002. Evidence that TAF-TATA box-binding protein interactions are required for activated transcription in mammalian cells. Mol Cell Biol **22:**2788-98.

107. **Martin, M. L., P. M. Lieberman, and T. Curran.** 1996. Fos-Jun dimerization promotes interaction of the basic region with TFIIE-34 and TFIIF. Mol Cell Biol **16:**2110-8.

108. **Matys, V., E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D. U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, and E. Wingender.** 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res **31:**374-8.

109. **McEwan, I. J., K. Dahlman-Wright, J. Ford, and A. P. Wright.** 1996. Functional interaction of the c-Myc transactivation domain with the TATA binding protein: evidence for an induced fit model of transactivation domain folding. Biochemistry **35:**9584-93.

110. **McKenna, N. J., and B. W. O'Malley.** 2002. Combinatorial control of gene expression by nuclear receptors and coregulators. Cell **108:**465-74.

111. **Meares, C. F., S. A. Datwyler, B. D. Schmidt, J. Owens, and A. Ishihama.** 2003. Principles and methods of affinity cleavage in studying transcription. Methods Enzymol **371:**82-106.

112. **Merika, M., and D. Thanos.** 2001. Enhanceosomes. Curr Opin Genet Dev **11:**205-8.

113. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:**371-8.

114. **Mittler, G., T. Stuhler, L. Santolin, T. Uhlmann, E. Kremmer, F. Lottspeich, L. Berti, and M. Meisterernst.** 2003. A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. Embo J **22:**6494-6504.

115. **Muratani, M., and W. P. Tansey.** 2003. How the ubiquitin-proteasome system controls transcription. Nat Rev Mol Cell Biol **4:**192-201.

116. **Myers, L. C., and R. D. Kornberg.** 2000. Mediator of transcriptional regulation. Annu Rev Biochem **69:**729-49.

117. **Nakajima, N., M. Horikoshi, and R. G. Roeder.** 1988. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. Mol Cell Biol **8:**4028-40.

118. **Narlikar, G. J., H. Y. Fan, and R. E. Kingston.** 2002. Cooperation between complexes that regulate chromatin structure and transcription. Cell **108:**475-87.

119. **Nedialkov, Y. A., D. D. Shooltz, and S. J. Triezenberg.** 2003. Purification and protein interaction assays of the VP16C transcription activation domain. Methods Enzymol **370:**522-35.

120. **Nedialkov, Y. A., and S. J. Triezenberg.** 2004. Quantitative assessment of in vitro interactions implicates TATA-binding protein as a target of the VP16C transcriptional activation region. Arch Biochem Biophys **425:**77-86.

121. **Nikolov, D. B., H. Chen, E. D. Halay, A. Hoffman, R. G. Roeder, and S. K. Burley.** 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. Proc Natl Acad Sci U S A **93:**4862-7.

122. **Nikolov, D. B., H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley.** 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature **377:**119-28.

123. **Nikolov, D. B., S. H. Hu, J. Lin, A. Gasch, A. Hoffmann, M. Horikoshi, N. H. Chua, R. G. Roeder, and S. K. Burley.** 1992. Crystal structure of TFIID TATA-box binding protein. Nature **360:**40-6.

124. **Nishikawa, J., T. Kokubo, M. Horikoshi, R. G. Roeder, and Y. Nakatani.** 1997. Drosophila TAF(II)230 and the transcriptional activator VP16 bind competitively to the TATA box-binding domain of the TATA box-binding protein. Proc Natl Acad Sci U S A **94:**85-90.

125. **Nogues, G., S. Kadener, P. Cramer, D. Bentley, and A. R. Kornblihtt.** 2002. Transcriptional activators differ in their abilities to control alternative splicing. J Biol Chem **277:**43110-4.

126. **O'Hare, P., and G. Williams.** 1992. Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using 1H NMR. Biochemistry **31:**4150-6.

127. **O'Shea-Greenfield, A., and S. T. Smale.** 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J Biol Chem **267:**1391-402.

128. **Ottosen, S., F. J. Herrera, and S. J. Triezenberg.** 2002. Transcription. Proteasome parts at gene promoters. Science **296:**479-81.

129. **Pan, C. Q., R. Landgraf, and D. S. Sigman.** 1994. DNA-binding proteins as site-specific nucleases. Mol Microbiol **12:**335-42.

130. **Pearson, A., and J. Greenblatt.** 1997. Modular organization of the E2F1 activation domain and its interaction with general transcription factors TBP and TFIIH. Oncogene **15:**2643-58.

131. **Poon, D., R. A. Knittle, K. A. Sabelko, T. Yamamoto, M. Horikoshi, R. G. Roeder, and P. A. Weil.** 1993. Genetic and biochemical analyses of yeast TATA-binding protein mutants. J Biol Chem **268:**5005-13.

132. **Praz, V., R. Perier, C. Bonnard, and P. Bucher.** 2002. The Eukaryotic Promoter Database, EPD: new entry types and links to gene expression data. Nucleic Acids Res **30:**322-4.

133. **Proudfoot, N. J., A. Furger, and M. J. Dye.** 2002. Integrating mRNA processing with transcription. Cell **108:**501-12.

134. **Ptashne, M., and A. Gann.** 1997. Transcriptional activation by recruitment. Nature **386:**569-77.

135. **Pugh, B. F.** 2000. Control of gene expression through regulation of the TATA-binding protein. Gene **255:**1-14.

136. **Rachez, C., and L. P. Freedman.** 2001. Mediator complexes and transcription. Curr Opin Cell Biol **13:**274-80.

137. **Radhakrishnan, I., G. C. Perez-Alvarado, D. Parker, H. J. Dyson, M. R. Montminy, and P. E. Wright.** 1997. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell **91:**741-52.

138. **Rawson, R. B.** 2003. The SREBP pathway--insights from Insigs and insects. Nat Rev Mol Cell Biol **4:**631-40.

139. **Regier, J. L., F. Shen, and S. J. Triezenberg.** 1993. Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcriptional activator. Proc Natl Acad Sci U S A **90:**883-7.

140. **Ren, B., H. Cam, Y. Takahashi, T. Volkert, J. Terragni, R. A. Young, and B. D. Dynlacht.** 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev **16:**245-56.

141. **Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young.** 2000. Genome-wide location and function of DNA binding proteins. Science **290:**2306-9.

142. **Roberts, S. G., and M. R. Green.** 1994. Activator-induced conformational change in general transcription factor TFIIB. Nature **371:**717-20.

143. **Roberts, S. G., I. Ha, E. Maldonado, D. Reinberg, and M. R. Green.** 1993. Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. Nature **363:**741-4.

144. **Rojo-Niersbach, E., T. Furukawa, and N. Tanese.** 1999. Genetic dissection of hTAF(II)130 defines a hydrophobic surface required for interaction with glutamine-rich activators. J Biol Chem **274:**33778-84.

145. **Rougvie, A. E., and J. T. Lis.** 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.

146. **Sadowski, I., J. Ma, S. Triezenberg, and M. Ptashne.** 1988. GAL4-VP16 is an unusually potent transcriptional activator. Nature **335:**563-4.

147. **Salghetti, S. E., A. A. Caudy, J. G. Chenoweth, and W. P. Tansey.** 2001. Regulation of transcriptional activation domain function by ubiquitin. Science **293:**1651-3.

148. **Salghetti, S. E., M. Muratani, H. Wijnen, B. Futcher, and W. P. Tansey.** 2000. Functional overlap of sequences that activate transcription and signal ubiquitin-mediated proteolysis. Proc Natl Acad Sci U S A **97:**3118-23.

149. **Sauer, F., S. K. Hansen, and R. Tjian.** 1995. Multiple TAFIIs directing synergistic activation of transcription. Science **270:**1783-8.

150. **Schmitz, M. L., M. A. dos Santos Silva, H. Altmann, M. Czisch, T. A. Holak, and P. A. Baeuerle.** 1994. Structural and functional analysis of the NF-kappa B p65 C terminus. An acidic and modular transactivation domain with the potential to adopt an alpha-helical conformation. J Biol Chem **269:**25613-20.

151. **Sera, T., and C. Uranga.** 2002. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. Biochemistry **41:**7074-81.

152. **Shen, F., S. J. Triezenberg, P. Hensley, D. Porter, and J. R. Knutson.** 1996. Critical amino acids in the transcriptional activation domain of the herpesvirus protein VP16 are solvent-exposed in highly mobile protein segments. An intrinsic fluorescence study. J Biol Chem **271:**4819-26.

153. **Shen, F., S. J. Triezenberg, P. Hensley, D. Porter, and J. R. Knutson.** 1996. Transcriptional activation domain of the herpesvirus protein VP16 becomes conformationally constrained upon interaction with basal transcription factors. J Biol Chem **271:**4827-37.

154. **Sigler, P. B.** 1988. Transcriptional activation. Acid blobs and negative noodles. Nature **333:**210-2.

155. **Sigman, D. S., M. D. Kuwabara, C. H. Chen, and T. W. Bruice.** 1991. Nuclease activity of 1,10-phenanthroline-copper in study of protein-DNA interactions. Methods Enzymol **208:**414-33.

156. **Smale, S. T., and J. T. Kadonaga.** 2003. The RNA polymerase II core promoter. Annu Rev Biochem **72:**449-79.

157. **Spencer, J. V., and K. M. Arndt.** 2002. A TATA binding protein mutant with increased affinity for DNA directs transcription from a reversed TATA sequence in vivo. Mol Cell Biol **22:**8744-8755.

158. **Strahl, B. D., and C. D. Allis.** 2000. The language of covalent histone modifications. Nature **403:**41-5.

159. **Stringer, K. F., C. J. Ingles, and J. Greenblatt.** 1990. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. Nature **345:**783-6.

160. **Sullivan, S. M., P. J. Horn, V. A. Olson, A. H. Koop, W. Niu, R. H. Ebright, and S. J. Triezenberg.** 1998. Mutational analysis of a transcriptional activation

region of the VP16 protein of herpes simplex virus. Nucleic Acids Res **26:**4487-96.

161. **Tan, S., Y. Hunziker, D. F. Sargent, and T. J. Richmond.** 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. Nature **381:**127-51.

162. **Tora, L.** 2002. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes Dev **16:**673-5.

163. **Triezenberg, S. J.** 1995. Structure and function of transcriptional activation domains. Curr Opin Genet Dev **5:**190-6.

164. **Triezenberg, S. J., R. C. Kingsbury, and S. L. McKnight.** 1988. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. Genes Dev **2:**718-29.

165. **Tsai, F. T. F., and P. B. Sigler.** 2000. Structural basis of preinitiation complex assembly on human Pol II promoters. Embo J **19:**25-36.

166. **Uesugi, M., O. Nyanguile, H. Lu, A. J. Levine, and G. L. Verdine.** 1997. Induced alpha helix in the VP16 activation domain upon binding to a human TAF. Science **277:**1310-3.

167. **Van Dyke, M. W., R. G. Roeder, and M. Sawadogo.** 1988. Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science **241:**1335-8.

168. **Van Hoy, M., K. K. Leuther, T. Kodadek, and S. A. Johnston.** 1993. The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. Cell **72:**587-94.

169. **Walker, S., R. Greaves, and P. O'Hare.** 1993. Transcriptional activation by the acidic domain of Vmw65 requires the integrity of the domain and involves additional determinants distinct from those necessary for TFIIB binding. Mol Cell Biol **13:**5233-44.

170. **Warnmark, A., E. Treuter, A. P. Wright, and J. A. Gustafsson.** 2003. Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. Mol Endocrinol **17:**1901-9.

171. **Warnmark, A., A. Wikstrom, A. P. Wright, J. A. Gustafsson, and T. Hard.** 2001. The N-terminal regions of estrogen receptor alpha and beta are unstructured in vitro and show different TBP binding properties. J Biol Chem **276:**45939-44.

172. **Weinmann, A. S., S. M. Bartley, T. Zhang, M. Q. Zhang, and P. J. Farnham.** 2001. Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol **21:**6820-32.

173. **Weinmann, P., M. Gossen, W. Hillen, H. Bujard, and C. Gatz.** 1994. A chimeric transactivator allows tetracycline-responsive gene expression in whole plants. Plant J **5:**559-69.

174. **Werstuck, G., and J. P. Capone.** 1989. Mutational analysis of the herpes simplex virus trans-inducing factor Vmw65. Gene **75:**213-24.

175. **Wieczorek, E., M. Brand, X. Jacq, and L. Tora.** 1998. Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II. Nature **393:**187-91.

176. **Willy, P. J., R. Kobayashi, and J. T. Kadonaga.** 2000. A basal transcription factor that activates or represses transcription. Science **290:**982-5.

177. **Wolfe, S. A., E. I. Ramm, and C. O. Pabo.** 2000. Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. Structure Fold Des **8:**739-50.

178. **Wolstein, O., A. Silkov, M. Revach, and R. Dikstein.** 2000. Specific interaction of TAFII105 with OCA-B is involved in activation of octamer-dependent transcription. J Biol Chem **275:**16459-65.

179. **Xiao, H., A. Pearson, B. Coulombe, R. Truant, S. Zhang, J. L. Regier, S. J. Triezenberg, D. Reinberg, O. Flores, C. J. Ingles, and et al.** 1994. Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. Mol Cell Biol **14:**7013-24.

180. **Xie, Y., C. Denison, S. H. Yang, D. A. Fancy, and T. Kodadek.** 2000. Biochemical characterization of the TATA-binding protein-Gal4 activation domain complex. J Biol Chem **275:**31914-20.

181. **Xie, Y., L. Sun, and T. Kodadek.** 2000. TATA-binding protein and the Gal4 transactivator do not bind to promoters cooperatively. J Biol Chem **275:**40797-803.

182. **Yang, F., R. DeBeaumont, S. Zhou, and A. M. Naar.** 2004. The activator-recruited cofactor/Mediator coactivator subunit ARC92 is a functionally important target of the VP16 transcriptional activator. Proc Natl Acad Sci U S A **101:**2339-44.

183. **Yankulov, K., J. Blau, T. Purton, S. Roberts, and D. L. Bentley.** 1994. Transcriptional elongation by RNA polymerase II is stimulated by transactivators. Cell **77:**749-59.

184. **Zhang, L., S. K. Spratt, Q. Liu, B. Johnstone, H. Qi, E. E. Raschke, A. C. Jamieson, E. J. Rebar, A. P. Wolffe, and C. C. Case.** 2000. Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. J Biol Chem **275:**33850-60.