

1991

# Structure and Function Analysis of a Plant Transcription Activator

Fumiaki Katagiri

Follow this and additional works at: [http://digitalcommons.rockefeller.edu/student\\_theses\\_and\\_dissertations](http://digitalcommons.rockefeller.edu/student_theses_and_dissertations)



Part of the [Life Sciences Commons](#)

---

## Recommended Citation

Katagiri, Fumiaki, "Structure and Function Analysis of a Plant Transcription Activator" (1991). *Student Theses and Dissertations*. 375.  
[http://digitalcommons.rockefeller.edu/student\\_theses\\_and\\_dissertations/375](http://digitalcommons.rockefeller.edu/student_theses_and_dissertations/375)

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact [mcsweej@mail.rockefeller.edu](mailto:mcsweej@mail.rockefeller.edu).

LD4711.6  
K1901  
c.1  
RES



THE LIBRARY



LD 4711.6 K1901 1991 c.1 RES  
Katagiri, Fumiaki.  
Structure and function  
analysis of a plant

Rockefeller University Library  
1230 York Avenue  
New York, NY 10021-6399



STRUCTURE AND FUNCTION ANALYSIS OF  
A PLANT TRANSCRIPTION ACTIVATOR

Submitted to the faculty of the Rockefeller University  
in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy

by

Fumiaki Katagiri

February, 1991

The Rockefeller University

New York



## Acknowledgments

I would like to express my sincere appreciation to my advisor, Prof. Nam-Hai Chua for his excellent scientific guidance throughout my graduate studies. I was always impressed with his enthusiasm and deep scientific insight. His attitude toward science was to enjoy doing science and this attitude has been of great influence to me. I have enjoyed the atmosphere established in his laboratory which allows for discussing even the craziest of ideas and I have learned that such discussions could be a source of great ideas.

I thank all current and former members of the Chua lab. I enjoyed the atmosphere in the lab very much. Particularly, I would like to thank Dr. Eric Lam, Katja Seipel, Dr. Gunter Strittmatter, Dr. Toan Hanh Trinh, and Dr. Anil Kush. My thesis work is based on Eric's work on *as-1* and ASF-1. I learned how to deal with nuclear factors from him and enjoyed my collaboration with him. Katja was a summer student who worked with me. She patiently pushed the project of TGA1a deletion analysis during the days when we were confused with unexpected results. Gunter and Hanh were the first people I worked with at Rockefeller. I feel very lucky to have met nice people like them in those early days when I was having difficulty including a language problem for I had just arrived in the US. I enjoyed Anil's company, especially over at the Faculty Club where science mixed well with beer.

I would like to express my appreciation to my collaborators in addition to Eric and Katja. I worked with Dr. Ken-ichi Yamazaki, Dr. Masami Horikoshi, and Prof. Robert G. Roeder on the *in vitro* transcription studies. I learned great deal through this collaboration. Being surrounded by such great scientists and having the opportunity to work with them is just one of the advantages of studying at Rockefeller. Recently I have begun a collaborative project with Dr. Gunther Neuhaus and Dr. Gabriele Neuhaus-Url on microinjection experiments.

I wish to thank my friends at Rockefeller. It was great fun to play soccer on the tennis court. I thank the people who played with (and against) me. Among my friends I especially thank David Sternberg for his friendship. He made my life at Rockefeller easier and more enjoyable.

And last but not least, I thank my family, namely my Father, Mother, and Sister. They have allowed me to do whatever I wanted to do in my life, and joining the graduate program at Rockefeller was one of the most important. This was definitely not a Japanese tradition.

This work was supported by a grant from Monsanto Co. to Prof. Nam-Hai Chua. I was supported by a graduate fellowship provided by The Rockefeller University.

## Table of contents

Acknowledgments .....	i
Table of contents .....	ii
List of figures .....	iii
List of tables .....	v
Abbreviations .....	vi
 Abstract .....	 1
Chapter 1: Introduction .....	3
Guide to the chapters .....	13
Chapter 2: Materials and methods .....	15
Chapter 3: Cloning of the DNA-binding protein gene .....	35
Introduction .....	35
Results .....	37
Discussion .....	44
Chapter 4: TGA1a can stimulate transcription in a human <i>in vitro</i> system. ....	49
Introduction .....	49
Results .....	50
Discussion .....	58
Appendix to chapter 4: TGA1a can stimulate transcription in a plant <i>in vitro</i> system. ....	64
Introduction .....	64
Results and Discussion .....	65
Conclusions .....	68
Chapter 5: TGA1a can stimulate transcription in leaf cells. ....	70
Introduction .....	70
Results .....	73
Discussion .....	76
Chapter 6: Analysis of functional domains of TGA1a .....	83
Introduction .....	83
Results .....	84
Discussion .....	91
Chapter 7: Summary and future studies .....	98
References .....	102

## List of figures

page followed by the figure

Fig. 1.1	Regulation in transcription initiation .....	3
Fig. 1.2	Comparison of the bZIP domain .....	8
Fig. 1.3	CaMV and its expression products .....	10
Fig. 1.4	Regulation of gene expression from the CaMV 35S promoter .....	11
Fig. 3.1	Binding site probes used in gel retardation assays .....	37
Fig. 3.2	Tobacco nuclear extracts contain DNA-binding proteins specific for the three TGACG-containing elements. ....	37
Fig. 3.3	DNA binding specificities of TGA1a and TGA1b .....	39
Fig. 3.4	DNA and deduced amino acid sequences of TGA1a .....	40
Fig. 3.5	DNA and deduced amino acid sequences of TGA1b .....	41
Fig. 3.6	Northern blot analysis of TGA1a and TGA1b genes .....	43
Fig. 3.7	Southern blot analysis of TGA1a and TGA1b genes .....	43
Fig. 3.8	Comparison of amino acid sequences of TGA1a and TGA1b with those of c-Jun, GCN4, and CREB .....	44
Fig. 4.1	Overproduction and purification of TGA1a .....	50
Fig. 4.2	TGA1a activates transcription in a HeLa cell <i>in vitro</i> system. ....	52
Fig. 4.3	Effects of increasing amounts of TGA1a on transcription activation ...	53
Fig. 4.4	Transcription initiation sites of the CaMV 35S promoter in the HeLa cell <i>in vitro</i> system .....	53
Fig. 4.5	TGA1a stimulates a single round of transcription <i>in vitro</i> . ....	54
Fig. 4.6	TGA1a-dependent preinitiation complexes remain committed to templates. ....	56

Fig. 4.7	Transcription activation by TGA1a	59
Fig. 4.8	<i>In vitro</i> transcription stimulation dependent on TGA1a and its cognate binding site, <i>as-1</i>	65
Fig. 4.9	Effect of TGA1a on single- and multiple- round transcriptions	67
Fig. 4.10	Effect of TGA1a concentration on a single round of transcription	68
Fig. 5.1	Injection of TGA1a induces expression of -90 35S promoter in leaf cells	73
Fig. 5.2	Many copies of -90 35S/GUS plasmid are required to show the expression in leaf cells.	74
Fig. 5.3	A similar concentration of the competitor plasmid effectively competes with expression of the -90 35S promoter.	75
Fig. 6.1	Deletion mutants of TGA1a used in this study	84
Fig. 6.2	The bZIP domain is sufficient for sequence-specific DNA binding.	85
Fig. 6.3	Heterodimer formation between TGA1a derivatives	86
Fig. 6.4	DNA-binding assays of TGA1a deletion mutants	87
Fig. 6.5	<i>In vitro</i> transactivation of TGA1a deletion mutants	90
Fig. 6.6	A model of the dimer stabilization (DS) domain	91

## List of tables

Table 4.1	Purification of TGA1a .....	50
Table 6.1	Heterodimer formation .....	89



## Abbreviations

AGEP	agarose gel electrophoresis
AP-4	activator protein 4
<u>as-1</u>	activation sequence-1
ASF-1	activation sequence factor-1
ATP	adenosine 5'-triphosphate
bZIP domain	basic and leucine-zipper domain
CTP	cytidine 5'-triphosphate
DS domain	dimer stabilization domain
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside 5'-triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
G-free	guanine-free
GTP	guanosine 5'-triphosphate
GUS	$\beta$ -glucuronidase
Hepes	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HLH domain	helix-loop-helix domain
HSBF	<i>hex-1</i> -specific binding factor
IPTG	isopropyl $\beta$ -D-thiogalactopyranoside
MS medium	Murashige-Schoog medium
NTP	nucleoside 5'-triphosphate
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
Pipes	piperazine-N,N'-bis[2-ethanesulfonic acid]
PMSF	phenylmethylsulfonyl fluoride
rbcS	small subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase
RNA	ribonucleic acid
sarkosyl	<i>N</i> -lauroylsarcosine
SDS	sodium lauryl sulfate
Tris	tris(hydroxymethyl)aminomethane
USF	upstream stimulating factor
UTP	uridine 5'-triphosphate
WGE	wheat germ extract

## Abstract

The cauliflower mosaic virus 35S promoter is composed of several cis-elements, each of which confers a different expression pattern in transgenic tobacco. One such cis-element, activation sequence (*as*)-1 (-83 to -63) mediates preferential expression in root. The element contains a tandem repeat of the motif TGACG, that is crucial for *as*-1 function and binding of a tobacco nuclear factor, activation sequence factor (ASF)-1. Two other TGACG-containing cis-elements from plant promoters were identified by a computer-assisted search of a gene bank. These cis-elements, *hex*-1 from the wheat histone H3 promoter and *nos*-1 from the nopaline synthase promoter of T-DNA, also bind ASF-1. Another tobacco nuclear factor, designated *hex*-1-specific binding factor (HSBF), binds only to *hex*-1. A tobacco cDNA library was screened for DNA-binding proteins using *hex*-1 as a binding probe and two types of clones encoding *hex*-1-binding proteins were isolated. Based on their binding specificities to *as*-1, *hex*-1, and *nos*-1, the encoded DNA-binding proteins, named TGA1a and TGA1b, appear to correspond to ASF-1 and HSBF, respectively. Both TGA1a and TGA1b are bZIP proteins. The TGA1a mRNA level is much higher in root than in leaf. TGA1a can function as an *as*-1-specific transcription activator in a HeLa cell *in vitro* system as well as a wheat germ *in vitro* system. In both systems, TGA1a stimulates transcription by increasing the number of preinitiation complexes. When microinjected into cotyledon cells of transgenic tobacco plants carrying a promoter linked to *as*-1, TGA1a can induce the expression of the promoter. Judging from these results, TGA1a is likely to be the factor that is responsible for the *as*-1 function *in vivo*. Such a factor is present in leaf cells in a limiting concentration. Although the bZIP domain of TGA1a is sufficient for the specific binding to its target site, another domain (DS

domain) increases the apparent DNA-binding affinity by stabilizing the dimeric form of TGA1a. The N-terminal region of TGA1a, which is enriched in acidic residues, appears to be essential for transactivation *in vivo*. However, when assayed in a HeLa cell *in vitro* system, most parts of the protein except the bZIP domain seem dispensable for activity.

## **Chapter 1: introduction**

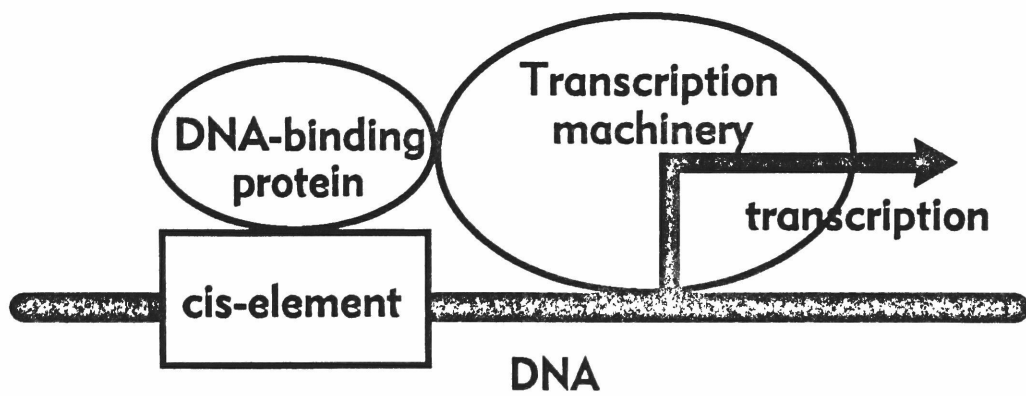
A major interest in the studies of multicellular organisms is to understand how cells that contain the same genetic information become differentiated into distinct cell types. It is known that during the acquisition of cell type-specific characteristics, regulation of gene expression plays a central role. For example, in plants many of the genes that are involved in photosynthesis are exclusively expressed in photosynthetic tissues (Kuhlemeier *et al.*, 1987). This differential gene expression confers distinctive characteristics on photosynthetic cells not found in other cell types. There are many ways that eukaryotic cells regulate gene expression: transcription, RNA processing, RNA transport to the cytoplasm, RNA turnover, translation, protein modification, protein degradation, etc. In many cases, however, transcription initiation is the major regulatory step. Therefore, studies on the regulation of transcription initiation are crucial to understanding how cells of a multicellular organism become differentiated into different cell types.

### **How is regulation of transcription initiation studied in eukaryotes?**

Figure 1.1 illustrates how transcription initiation may be regulated in eukaryotes. A promoter site contains a specific DNA sequence, referred to as a cis-element (usually upstream to the promoter site), that mediates expression of the promoter. A cis-element is recognized by a DNA-binding protein specific to the element. This specific recognition, together with other factors (eg. factors involved in the transcription machinery), results in a positive or negative regulation of the promoter. Although often the real situations are more complicated (eg. interaction with other cis-elements or with other factors is crucial), this simple model provides a good starting point to study how a promoter is regulated. In

Figure 1.1 Regulation in transcription initiation.

A mechanism of transcriptional regulation is schematically illustrated. A specific DNA sequence that mediates a specific expression of promoter (cis-element) is recognized by a sequence-specific DNA-binding protein. This recognition somehow confers an effect on the general transcription machinery, to result in a specific regulation of transcription initiation (arrow).



fact, a common strategy to study regulation of transcription initiation is to identify and characterize such cis-elements and their cognate DNA-binding proteins. Three general approaches have been employed: 1) genetic analysis of regulatory mutants, 2) isolation of genes encoding proteins with homology to well-characterized DNA-binding proteins, and 3) identification of cis-elements.

The genetic approach involves the isolation of mutants that are altered in gene expression and subsequent cloning of the corresponding wild type alleles by molecular techniques. Either when the encoded proteins are found to contain motifs characteristic of DNA-binding domain, eg. bZIP domain, or when they are expected to regulate certain genes directly based on the genetic study, target binding sites of the encoded proteins may be investigated in the regulated genes. Such binding sites are expected to be the cis-elements mediating the regulation. In the case of plants, the genetic approach is often used in maize, *Arabidopsis*, and *Antirrhinum* (snapdragon). In addition to classical genetics (Coe *et al.*, 1988; Meyerowitz, 1989; Schwarz-Sommer *et al.*, 1990), the availability of gene-tagging techniques (Coen *et al.*, 1990; Fedoroff, 1989; Feldmann *et al.*, 1989) and/or detailed genetic/physical maps (Chang *et al.*, 1988; Nam *et al.*, 1988) in these plant species greatly facilitate this approach.

The second approach involves cloning of DNA-binding protein genes based on sequence homology to other known DNA-binding proteins (see below). After cloning of the genes, their functions would have to be investigated. It is not, however, a trivial task to identify the functions of such a DNA-binding protein unless alterations of its expression pattern in transgenic plants (ideally, through a disruption of the corresponding endogenous gene although such technique is not yet available in plant systems) result in a phenotype.

The third approach, which starts with cis-element analysis, is described below in

more detail, because this approach was used in the present study. To delineate a cis-element, synthetic promoters containing wild type and mutated upstream sequences are generated and the effects of such mutations on the expression pattern of interest are analyzed. Several pieces of information are useful for the construction of such wild type and mutant promoters: conserved sequences among promoters with similar expression patterns, sequences that are recognized by nuclear proteins *in vitro*, and sequences similar to the target binding sites of known DNA-binding proteins.

The synthetic promoters are often connected to the coding region of a reporter gene, which facilitates assay of the promoter expression pattern. Four reporter genes are commonly used, chloramphenicol acetyltransferase (CAT) gene (Gorman *et al.*, 1982),  $\beta$ -galactosidase ( $\beta$ -Gal) gene (Guarente and Ptashne, 1981),  $\beta$ -glucuronidase (GUS) gene (Jefferson *et al.*, 1987), and luciferase (LUC) gene (de Wet *et al.*, 1985; Ow *et al.*, 1986). The activity of each of these reporter gene products is readily assayed and correlates with their mRNA abundance. Because the cellular distribution of the enzyme activities of  $\beta$ -Gal, GUS, and LUC can be detected by histochemical analysis, these reporter genes can also be used to investigate the expression pattern at the single cell level. In plants, the GUS gene is most commonly used for this purpose because most tissues of non-transformed plants contain little or no GUS activity. Moreover, the GUS activity can be measured with a very high sensitivity by a fluorometric assay.

When an expression pattern of interest is cell type-specific, the activities of cis-elements should be assayed in a system that allows such specificity to be manifested. In the case of mammals, cell lines that maintain their differentiated states in culture are available and they are often used to assay the activities of cis-elements that may confer cell-specific expression. In plants, it is usually difficult to maintain the differentiated characteristics of



cells in tissue culture because they become "dedifferentiated" rapidly. Therefore, transgenic plants are generally used for the analysis of gene expression. Using staggered deletions, a part(s) of the upstream region is defined to be necessary and sufficient for the expression pattern. This part(s) should contain the specific cis-element responsible for the expression pattern. The function of the cis-element could be ascertained by testing whether it can confer the same expression pattern when linked to another promoter and whether site-specific mutations within the cis-element can abolish its function.

After identification of cis-element(s), DNA-binding proteins specific for the element(s) can be investigated. Such proteins are initially characterized *in vitro* as binding factors in nuclear extracts. The important question is whether the detected DNA-binding protein is indeed responsible for the *in vivo* function of the cis-element. As a first step, it would be important to establish a correlation in sequence-specificity between the *in vitro* DNA binding and the *in vivo* function of the cis-element. However, considering the promiscuous relationships between DNA-binding proteins and cis-elements (discussed in chapter 3), this criterion alone is not sufficient.

Cloning of the genes encoding the DNA-binding proteins is one of the most effective ways for further characterization of the proteins. Strategies for gene cloning and uses of cloned genes are described in chapter 3. If disruption of the corresponding endogenous gene severely affects the expression pattern of interest, this would constitute good evidence that the gene product is involved in the regulation. However, as mentioned above, gene disruption techniques are not yet available in plants. Therefore, it is important to accumulate as many lines of supporting evidence as possible. For example, a correlation between the expression pattern of the DNA-binding protein itself and the expression pattern conferred by the cis-element could suggest that the protein plays a role in the expression

pattern. Moreover, if the DNA-binding protein is expected to be a transcription activator, demonstration of its transactivation activity would provide additional supporting evidence.

### **Domain structure of eukaryotic DNA-binding proteins**

Early studies of eukaryotic DNA-binding proteins, GAL4 (Keegan *et al.*, 1986; Ma and Ptashne, 1987) and GCN4 (Hope and Struhl, 1986), revealed that they are comprised of separable polypeptide regions that mediate different functions. For example, GCN4 (281 amino acid residues) contains a region involved in DNA-binding (amino acid numbers 222-281; Hope and Struhl, 1986) and a region mediating transactivation (amino acid numbers 107-144; Hope *et al.*, 1988). These functionally defined polypeptide regions are not only separable from each other but also portable, ie. the function can be transferred to another protein molecule when the relevant polypeptide region is fused to the latter. For instance, a chimeric protein comprising of a transactivation domain of GCN4 and the DNA-binding domain of LexA can activate transcription from a promoter with LexA-binding sites in yeast (Hope and Struhl, 1986). With the isolation and characterization of more eukaryotic DNA-binding protein genes and their products, it became clear that the structural feature first described for GAL4 and GCN4 is rather common among many DNA-binding proteins. In this thesis, I use the term "domain" to refer to a functionally defined polypeptide region (eg. DNA-binding domain, transactivation domain). Note that the definition of these functionally defined domains is different from that of structurally defined domains.

As information on the structures of DNA-binding domains accumulated, it became clear that most of them contain specific primary sequences that can be classified into a limited number of groups (Mitchell and Tjian, 1989; Johnson and McKnight, 1989). The motifs for DNA-binding domain so far characterized are helix-turn-helix (including homeo

domain; Steitz *et al.*, 1982; Laughon and Scott, 1984), zinc-finger (Miller *et al.*, 1985), basic-leucine repeat (bZIP domain; Landschulz *et al.*, 1988), helix-loop-helix (HLH; Murre *et al.*, 1989), SRF-related (Norman *et al.*, 1988), myb-related (Biedenkapp *et al.*, 1988), rel-related (Ghosh *et al.*, 1990; Kieran *et al.*, 1990), and ETS-domain (Karim *et al.*, 1990) motifs. Usually a relatively small polypeptide region including one of these motifs is sufficient for sequence-specific DNA binding, suggesting that the primary structure contained within the region is able to fold properly. In fact, 3-D structures of polypeptides containing helix-turn-helix (Anderson *et al.*, 1981; McKay and Steitz, 1981; Kissinger *et al.*, 1990) and zinc-finger (Schwabe *et al.*, 1990) motifs have been determined and the structural information has provided clues as to how these folded polypeptides may interact with double-stranded DNA. A 3-D model of the bZIP domain has also been proposed (Vinson *et al.*, 1989). Because of the close relationship between the primary structures of polypeptides and their DNA binding function, a functionally uncharacterized protein is expected to bind DNA if it contains homology to one of these motifs. Moreover, genes encoding DNA-binding proteins have been isolated from libraries by virtue of their homology to these motifs (see above). Most of the DNA-binding domain motifs appear to be prevalent among different eukaryotes. All of the motifs have been found in animals, all except the rel-related and the ETS-domain motifs have been found in plants, and all except the rel-related, the ETS-domain, and HLH have been found in yeast. The helix-turn-helix motif was originally found in prokaryotic DNA-binding proteins (Anderson *et al.*, 1981; McKay and Steitz, 1981).

Since the DNA-binding proteins described in this thesis contain the bZIP domain, this domain is described in more detail in the following. The bZIP domain is comprised of a region rich in basic amino acid residues (basic region) abutting a "leucine-zipper" region (Landschulz *et al.*, 1988). Figure 1.2 shows a comparison of some bZIP domains (cited from

Figure 1.2 Comparison of the bZIP domain (Vinson *et al.*, 1989).

Amino acid sequences corresponding to the DNA binding domains of 11 bZIP proteins. Numbers preceding each sequence correspond to the distance, in amino acid residues, from the N-terminus of the respective protein. Amino acid sequences are shown in the single letter code. The 11 proteins contain a cluster of similar residues in areas designated basic region A (BR-A) and basic region B (BR-B). This cluster of similar residues is designated as a consensus and shown below the 11 protein sequences. See Vinson *et al.* (1989) for the reference of each sequence.

Protein	BR-A	BR-B	Leucine zipper
C/EBP	278-DKNSNEYRVRRRERNNI	AVRKS	RDKAKQRNVETQOKY
Jun	257-SQERIKAEKRMRNRRI	AASKCRKRKL	ERIALREEKYKT
Fos	233-EERRRIIRIRRRERNK	MAAAKCRNR	RRRLTDTLQAETDQLEDKKS
GCN4	221-PESSDPAALKRARNTE	AAARRSRARK	LQRMKQLEDKVEELLSKNYH
YAP1	60-DLDPETKQKRTAQNRA	AQRAFRENK	ERKMKLEKKVOSLESIQOONEVEAT
CREB	279-EAAARKREVRLMKNR	EAARECRKKK	EYVKLENRVAVLENONKTLIEELKALKDLYCHKSD
Cys-3	95-ASRLAAEEDKRKRNT	AASARFRIKKK	KOREQALEKSAKEMSEKVTQLEGRIQALETENKYLKG
CPC1	211-EDPSDVMKRRARNT	LAARKSRERKA	ORLEELAKIEELIAERDRYKNLALAHGASTE
HBP1	176-WDERELKKQKRLSN	RESARRSRLR	KQAECEELGQRAEALKSENSSLRIELDRIKKEYEELLS
TGA1	68-SKPVEKYLRLRLAQN	REARKSRLR	KKAYVQOLENSKCLKLIQLEQELERARKQGMVGGGYDA
Opaque2	223-MPTEERVVRKRKES	NRESARRSRYR	KAHLKELEDQVAQLKAENSCLLRRI AALNQKYNDANV
Consensus	-----BB-BN--AA-B-R-BB-----		L <sup>E</sup> <sub>O</sub> -----L-----L-----L-----L--

Vinson *et al.*, 1989). Note that some residues in this domain are very well conserved. The "leucine-zipper" region contains leucine residues at every seventh amino acid. It can fold into an amphipathic  $\alpha$ -helix and two such helices can dimerize by forming a coiled-coil structure (O'Shea *et al.*, 1989). There is selectivity in dimerization between different "leucine-zippers" (see chapter 6). It has been demonstrated that the basic region alone is required for DNA-binding. When two short polypeptides, each containing only the GCN4 basic region (without "leucine-zipper"), were linked by a disulfide bond at their C-termini, the linked polypeptides were able to bind DNA at a low temperature with the same specificity as that of GCN4 (Talanian *et al.*, 1990). Since the spacing between the basic region and the "leucine-zipper" region is very well conserved (Fig. 1.2), it has been suggested that the "leucine-zipper" region, upon dimerization, presents the basic region in an appropriate configuration for DNA binding (Vinson *et al.*, 1989). Indeed, alteration of this spacing by either the insertion or deletion of five amino acids eliminated DNA binding in GCN4 and C/EBP (Agre *et al.*, 1989).

Unlike DNA-binding domains, transactivation domains do not show very clear homologies with one another. The acidic amino acid residue-rich region (acidic domain) is the best characterized motif for transactivation (Ptashne, 1988). In addition to its negative charge, the ability of this domain to form an amphipathic helix was thought to be important for the transactivation activity (Giniger and Ptashne, 1987). The acidic domain of GAL4 can function in mammals, insects, and plants, as well as in yeast (Ptashne, 1988). Other transactivation domain motifs have only been identified in animal cells. They are the glutamine-rich (Courey and Tjian, 1988) and proline-rich (Mermoud *et al.*, 1989) motifs. Although these motifs were named after their distinctive structural features, it has not been proven that these features are crucial for the function of these domains. There are also

transactivation domains that do not show any distinctive structural features (Friedman and McKnight, 1990).

### **Cauliflower mosaic virus (CaMV)**

Cauliflower mosaic virus (CaMV) is the best studied member of the caulimovirus group. The virus infection is mainly limited to members of the family *Cruciferae* (eg. cauliflower and cabbage) where it causes a systemic mosaic symptom (Hohn *et al.*, 1982).

The genome of CaMV consists of a circular double-stranded DNA of ~8 kb (Fig. 1.3, cited from Pfeiffer and Hohn *et al.*, 1983). This circular DNA contains at least 8 open reading frames (ORFs) and the gene products of six of them have been identified (Hohn *et al.*, 1985). Two transcripts, the 35S and 19S RNA, which were named after their sedimentation coefficients, from the virus genome have been characterized (Fig. 1.3). The 35S RNA, corresponding to the entire genome plus a 180-b terminal repeat, is used not only as a polycistronic mRNA but also as a template for the replication of the virus genome by reverse transcription (Guilley *et al.*, 1982; Pfeiffer and Hohn, 1983), while the 19S RNA is used as a mRNA for ORFVI (Hohn *et al.*, 1985). The 35S and 19S promoters are responsible for the transcription of the 35S and 19S RNA, respectively. Compared with the 19S promoter, the 35S promoter has been studied in much more detail.

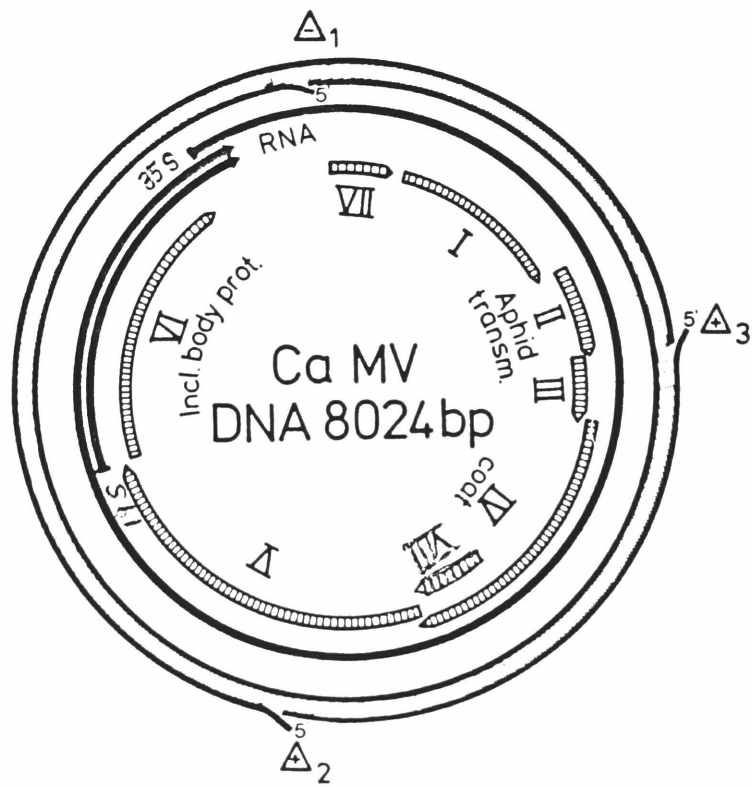
### **Combinatorial structure of cis-elements in the 35S promoter**

The 35S promoter shows strong and constitutive expression in various plants, including non-host species, even in monocots, in the absence of any viral gene products (eg. Odell *et al.*, 1985; Sanders *et al.*, 1987). This characteristic makes the 35S promoter ideal for the expression of foreign genes in plants. Its high level of expression in all plant organs

Figure 1.3 CaMV and its expression products (Pfeiffer and Hohn, 1983).

The outer ring symbolizes the double-stranded DNA as present in virus particles and with its S1 nuclease-sensitive sites:  $\Delta 1$  (minus strand);  $\Delta 2$  and  $\Delta 3$  (plus strand). Bold curved arrows represent the major transcripts detected, the 35S and 19S RNA. Hatched arrows represent the 8 ORFs located in different reading phases (ORF VIII is added to the original figure in Pfeiffer and Hohn, 1983).





has generated considerable interest in the analysis of its cis-elements.

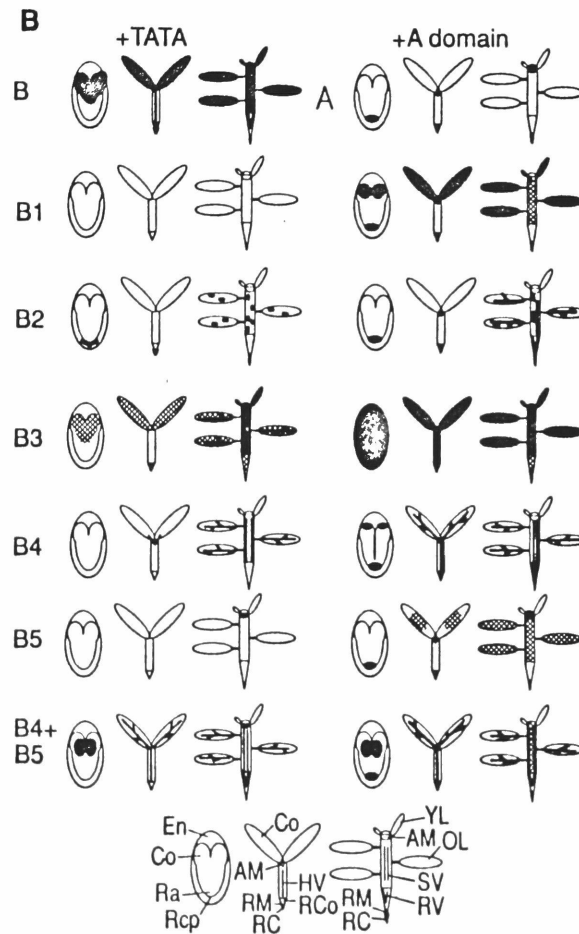
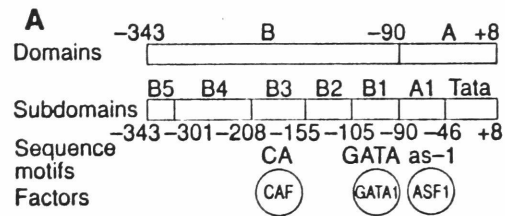
Transient expression assays in protoplasts (Ow *et al.*, 1987) and analysis in transgenic plants (Odell *et al.*, 1985; Fang *et al.*, 1989) have shown that the 35S promoter can be dissected into functional domains. It has been observed that a 35S deletion mutant with only 90 bp of 5' sequence shows preferential expression in root (Poulsen and Chua, 1988). Recently, these analyses have been further refined by the use of the GUS gene as a reporter gene which enables the monitoring of promoter expression at the cellular level in transgenic plants (Jefferson *et al.*, 1987). Using this technique in transgenic tobacco, Benfey *et al.* (1989) have clearly demonstrated that the 35S promoter contains at least two domains (A and B domains), each of which has a different cell type-specific expression pattern during development. The A domain (-90 to +8) confers preferential expression in root, while the B domain (-343 to -90) confers expression in leaf and vascular tissue. More recent studies by Benfey *et al.* (1990a,b) have shown that the B domain can be divided into at least five subdomains, each of which confers a different expression pattern on the promoter. They suggested that the apparent constitutive expression of the 35S promoter is likely to be generated by the combinatorial and synergistic effects of cis-elements residing within these domains and subdomains (summarized in Figure 1.4, cited from Benfey and Chua, 1990).

The combinatorial cis-element structure of the 35S promoter offers a unique opportunity to isolate cis-elements that confer various expression patterns. After characterization of the cis-elements in the 35S promoter, cellular promoters with homologous cis-elements can be examined to determine whether they are under similar controls. Since the amount of available information on functionally defined cis-elements is still quite limited in plants, studies of the constituent cis-elements of the 35S promoter can greatly contribute to our understanding of transcriptional regulation in plants.

Figure 1.4 Regulation of gene expression from the CaMV 35S promoter (Benfey and Chua, 1990).

(A) Domains and subdomains of the CaMV 35S promoter. Two domains (A and B) and 5 subdomains (B1 through B5) have been defined by their ability to confer cell-specific gene expression (Benfey *et al.*, 1989, 1990a,b). The positions of domains and subdomains are shown in respect to the transcription initiation site. *in vitro* binding analyses have identified three sequence motifs within the promoter [*as-1* (Lam *et al.*, 1989), GATA (originally called *as-2*; Lam and Chua, 1989), and CA (Kush and Conner, unpublished)], which are the binding sites for three trans-acting factors [ASF-1, GATA1 (ASF-2), and CAF, respectively].

(B) Schematic representation of expression patterns conferred by the B domain (top left) and the individual B subdomains placed upstream of the minimal TATA region (-46 to +8). Expression patterns conferred by the A domain alone (top right) or the B subdomains in combination with the A domain are shown on the right side of the figure. Crosshatching represents low amounts of expression. Expression is depicted in (from left to right) seeds, seedlings, and mature plants. Expression in the root cap is indicated by an additional set of lines for seeds and seedlings, and by additional filled triangle in mature plants. The tissues represented schematically are identified in the last row. AM, apical meristem; Co cotyledon; En, endosperm; HV, hypocotyl vascular tissue; OL, older leaf; Ra radicle; Rcp, root cap primordia; RC, root cap; RCo, root cortex; RM, root meristem; RV, root vascular tissue; SV, stem vascular tissue; YL, younger leaf.



## The *as-1* element and ASF-1

One of the well studied cis-elements in the 35S promoter is activation sequence-1 (*as-1*), which is located in the -83 to -63 region (within the A domain; Fig. 1.4). This element serves two roles in the 35S promoter expression: it is responsible for the expression pattern of the A domain (preferential expression in root) and it interacts synergistically with cis-elements in the B domain (Benfey et al., 1990a,b; Benfey and Chua, 1990; Fang *et al.*, 1989; Lam *et al.*, 1989; Lam and Chua, 1989; Poulsen and Chua, 1988). The *as-1* element is a portable cis-element. When inserted into a green-tissue specific promoter *rbcS-3A*, *as-1* can cause the expression of the hybrid *rbcS-3A* promoter in root (Lam *et al.*, 1989). The *as-1* element can also interact synergistically with other cis-elements that are not derived from the 35S promoter. Because the synergy with other cis-elements is the more complicated function, in this thesis, the *in vivo* function of *as-1* is defined as the activity that confers preferential expression in root, unless indicated.

Tobacco nuclear extracts contain a DNA-binding protein, designated as activation sequence factor (ASF)-1, which binds to *as-1* *in vitro*. *as-1* contains a tandem repeat of the sequence motif, TGACG and each TGACG-motif serves as a single binding site for ASF-1 (Lam *et al.*, 1989,1990a; Prat *et al.*, 1989). Mutations in both TGACG motifs of *as-1* cause a drastic decrease in ASF-1 binding *in vitro* and abolish the ability of *as-1* to confer root-specific expression on the *rbcS-3A* promoter *in vivo* (Lam *et al.*, 1989). This correlation between the ability to bind ASF-1 *in vitro* and the functional activity of *as-1* *in vivo* suggests that ASF-1 is responsible for the *in vivo* function of *as-1* (Lam *et al.*, 1989).

## Guide to the chapters

In this thesis, I mainly describe studies on a cloned DNA-binding protein, designated as TGA1a, which is considered to correspond to the nuclear factor ASF-1.

Chapter 2 describes the materials and methods used in the experiments presented in chapters 3-6.

Chapter 3 describes the cloning of DNA-binding protein genes from tobacco. This was the first report on the primary structures of sequence-specific DNA-binding proteins from plants. The cDNA-derived DNA-binding proteins, TGA1a and TGA1b, are thought to correspond to tobacco nuclear factors, ASF-1 and HSBF, respectively. Both proteins contain the bZIP domain. Based on the conservation of the bZIP domain as a DNA-binding domain in various eukaryotes, it is likely that this motif appeared early in evolution. Most of the results described in this chapter were reported in Katagiri *et al.* (1989). Other results either were reported in Lam *et al.* (1990b) or are unpublished.

Chapter 4 describes experiments that show that TGA1a can function as a sequence-specific transcription activator in a HeLa cell *in vitro* system. This was the first report demonstrating that a plant transcription factor can function in an animal system. The report also shows that TGA1a increases the number of preinitiation complexes and that the TGA1a-dependent preinitiation complex remains committed to the promoter site even after transcription initiation. The results described in this chapter appeared in Katagiri *et al.* (1990).

The appendix to chapter 4 describes a study similar to that described in chapter 4 but with a plant *in vitro* transcription system. This was the first demonstration of a plant *in vitro* transcription system that is responsive to an exogenously added transcription factor. The report also demonstrates that TGA1a is a transcription activator and that the

transactivation is achieved by facilitating the preinitiation complex formation in this system. The results shown in this chapter were reported in Yamazaki *et al.* (1990).

Chapter 5 describes experiments demonstrating that TGA1a can also function as a transcription activator *in vivo*. The combination of microinjection technology and transgenic plants carrying the appropriate transgene as a reporter gene has enabled us to investigate the *in vivo* activity of TGA1a. The injection of a defined number of a reporter genes into single tobacco cells has shown the factor mediating the *in vivo* function of *as-1* is limiting in leaf cells. The results described in this chapter are being prepared for publication (G.Neuhaus, G.Neuhaus-Url, F.Katagiri, K.Seipel, and N.-H.Chua).

Chapter 6 describes a deletion analysis of the functional domains of TGA1a. The analysis on DNA binding has uncovered a domain (DS domain) in TGA1a that stabilizes dimer formation of the bZIP protein. The transactivation analysis of various TGA1a mutants has produced contradictory results between *in vivo* and *in vitro* assays. The *in vivo* result shows that the N-terminal region of TGA1a is essential for transactivation, while the *in vitro* result shows most of the protein is dispensable for transactivation except a small region containing the bZIP domain. The results shown in this chapter are being prepared for publication (F.Katagiri, K.Seipel, and N.-H.Chua).

## **Chapter 2: materials and methods**

### **Materials**

#### **Plants**

*Nicotiana tabacum* cv. SR1 (tobacco plant) and transgenic plants of this cultivar carrying various reporter gene constructs were used. Construct A contains a chimeric gene comprised of domain A (-90 to +8) of the CaMV 35S promoter (Benfey and Chua, 1989; Benfey *et al.*, 1989), the coding sequence of  $\beta$ -glucuronidase (GUS; Jefferson *et al.*, 1987), and the poly A addition sequence of the pea *rbcS-3C* gene (Fluhr *et al.*, 1986). Construct  $\Delta A$  is the same as construct A except that the CaMV 35S promoter is from -72 to +8 (Benfey *et al.*, 1989). These constructs were introduced into tobacco plants by *Agrobacterium*-mediated transformation (Horsch and Klee, 1986) using the pMON505 vector. Because the pMON505 vector carries the kanamycin resistance gene as a selective marker in plants, the transgenic plants are kanamycin resistant.

#### **Plasmids, phages, and bacteria**

The  $\lambda$ gt11 vector was used for the construction of a cDNA expression library from which TGA1a and TGA1b genes were isolated (Huynh *et al.*, 1985). *Escherichia coli* strain Y1090 was used as the host for the expression library and the strain Y1089 was used to generate lysogens from  $\lambda$  clones. Since the expression of the cDNA inserts are under the control of *lacZ* (the vector is designed to produce a LacZ-fusion protein with the polypeptide encoded by the cDNA insert.), the expression of the fusion proteins are inducible by IPTG in the host bacteria.

To determine the DNA sequences of the TGA1a and TGA1b genes, the M13 single-



stranded DNA phage system, M13mp18 and mp19, were used with *E. coli* strain JM109 as the host (Messing, 1983; Yanisch-Perron *et al.*, 1985). Plasmid vectors of the pBluescript II series were used with the *E. coli* strain XL1-Blue (Stratagene) for various cloning experiments.

The T7 expression system (Studier *et al.*, 1990), comprised of the plasmid vector pET3a and the *E. coli* strain BL21(DE3)/plysS as the host, was used for the overproduction of TGA1a and its derivatives.

The G-free sequence used in the DNA templates for the *in vitro* transcription experiments was derived from the plasmid p(C<sub>2</sub>AT)19 (Sawadogo and Roeder, 1985b).

#### cDNA library

The tobacco root cDNA library from which longer TGA1b clones were isolated was constructed by Rong-Xiang Fang in the Chua Lab. The library contained  $\sim 8 \times 10^4$  recombinants. cDNAs were synthesized using oligo dT primer, and cloned into the  $\lambda$ ZAP II vector (Stratagene). The *E. coli* strain XL1-Blue was used as the host.

#### **Methods**

General techniques in molecular biology were performed according to Ausubel *et al.* (1987) and Sambrook *et al.* (1989).

#### Growth of plants

Tobacco seeds were germinated on MS medium (Murashige and Schoog, 1962) in petri dishes, after sterilization in 10% Chlorox solution (The Chlorox Co.) for 5 to 10 min. Medium containing kanamycin (100  $\mu$ g/ml) was used for germinating seeds of transgenic

plants. The petri dishes were kept at 25°C under a cycle of 16-hour light and 8-hour dark. For the microinjection experiments, 5 to 7 day-old seedlings were used. For other purposes, the seedlings were transferred into Plantcons (Flow Laboratories) containing the same medium. If needed, plants were clonally propagated by cuttings at this stage. Plants were transferred to soil when they reached > 7 cm tall (~1.5 month after germination) and grown to maturity either in a growth chamber (25°C, 16-hour light and 8-hour dark) or the greenhouse (25°C, 16-hour light and 8-hour dark).

### Plasmid construction

#### (1). Construction of TGA1a deletion mutants

A double stranded oligonucleotide shown below was first synthesized and cloned into the *SacI-KpnI* site of pBluescript II KS+ (Stratagene) to obtain pKE1. Unique restriction sites and important codons within the oligonucleotide region of pKE1 are indicated. pKE1 has a T7 promoter site (in the region from pBluescript II KS+) upstream of this cloned fragment.

		Met		3-frame stops	
5'	GAAGCTTCAT	<b>ATGA</b>	ATTCCATCGATG	<b>TAGGTAGGTAG</b>	CTCGAGGATCCAGATCTGTAC 3'
3'	TGCACTTCGAAGTATACTTAAGGTAGCTACATCCATCCATCGAGCTCCTAGGTCTAGA				5'
	HindIII	NdeI	EcoRI	ClaI	XhoI BamHI BglII

The 1.2-kb *EcoRI-XhoI* fragment of pHB1B (a subclone of the TGA1a gene, see below) that contains the entire TGA1a coding region except 2 bp at the 5' end (nucleotide number 3 to 1,156 in Fig. 3.4), was cloned into the *EcoRI-XhoI* site of pKE1 to obtain pKT1A. In this context, the methionine codon in the oligonucleotide region of pKE1 (shown above) became the first methionine codon of the TGA1a reading frame. The open reading frame (ORF) starting from this methionine codon is identical to the authentic TGA1a ORF (see chapter

3).

To generate deletion mutants of TGA1a, the following restriction sites within the coding region were utilized: *Hpa*II [239; the number refers to the nucleotide number (according to Fig. 3.4) of the 5'-end of the cutting sites with this enzyme.], *Aat*II (277), *Nhe*I (431 and 458), *Afl*III (533), *Sac*I (761), and *Bgl*II (946). After pKT1A was digested with one of these restriction enzymes, the staggered ends were, if needed, filled-in with Klenow enzyme (in the case of 5'-hang over) or removed with T4 DNA polymerase (in the case of 3'-hang over). A *Cla*I linker was then ligated to the flushed ends. The following procedures were used to generate either the C-terminal or N-terminal deletion mutants:

1) C-terminal deletion mutants. DNA was digested with *Eco*RI and *Cla*I and the fragments corresponding to the N-terminal regions of TGA1a were purified by agarose gel electrophoresis (AGEP) and cloned into the *Eco*RI-*Cla*I site of pKE1. The C-termini of the deletion mutants were fused to one of the stop codons that are present in three reading frames in pKE1. This method was used to construct the plasmids encoding the C-terminal deletion mutants,  $\Delta$ C108,  $\Delta$ C144,  $\Delta$ C178,  $\Delta$ C252, and  $\Delta$ C316 (the number refers to the last amino acid residue in the deletion mutant). The plasmids encoding these C-terminal deletion mutants were named by adding "p" in front of the deletion mutant names (eg. p $\Delta$ C108).

2) N-terminal deletion mutants. DNA was digested with *Cla*I and *Xho*I and the fragments corresponding to the C-terminal regions of TGA1a were purified by AGEP and cloned into *Cla*I-*Xho*I site of pKE1. After isolation of these plasmids, to obtain in-frame fusions of the initiator methionine codon in pKE1 and the deleted coding region of TGA1a, if needed, the plasmids were redigested with *Cla*I and the staggered ends were filled-in by Klenow enzyme

or by removed with Mung Bean nuclease and then religated. This procedure was used to construct the N-terminal deletion mutants,  $\Delta N80$  and  $\Delta N94$  (the number refers to the position of the N-terminal amino acid residue in deletion mutant). The nomenclature of the plasmids is the same as for the C-terminal deletion mutants (eg. p $\Delta 80$ ).

The 70-bp *EcoRI-AatII* fragment of p $\Delta N80$  purified by polyacrylamide gel electrophoresis (PAGE) was exchanged with the 290-bp *EcoRI-AatII* region of the C-terminal deletion mutant constructs to generate p $\Delta N80\Delta C108$ , p $\Delta N80\Delta C144$ , p $\Delta N80\Delta C178$ , p $\Delta N80\Delta C252$ , and p $\Delta N80\Delta C316$ .

For *in vitro* translation of these TGA1a derivatives (for chapter 6, see below), the plasmids were digested with *HindIII* and *EcoRI* to exchange this region with the synthesized DNA fragment shown below. This fragment mimics the consensus sequence surrounding an eukaryotic translation initiation site (Kozak, 1987). The initiator methionine codon is indicated.

		Met	
5'	AGCTTGCCGCCACCATG		3'
3'	ACGGCGGTGGTACTTAA		5'

## (2) Construction for expression in *E. coli*

The *NdeI-BamHI* fragments of pKT1A, p $\Delta N80$ , and p $\Delta N80\Delta C144$  were cloned into the *NdeI-BamHI* site of pET3a (Rosenberg *et al.*, 1987) to obtain pKT7T1A, pKT7 $\Delta N80$ , and pKT7 $\Delta N80\Delta C144$ , respectively. Each of these plasmids was transformed into the host bacteria BL21(DE3)/plysS and the transformants were used for the overproduction of TGA1a and the derivatives.

## (3) DNA templates used for *in vitro* transcription

A 69-bp DNA was synthesized that contained the -44 to -4 region of the CaMV 35S promoter and a polylinker consisting of *Bgl*III, *Hind*III, *Eco*RI, and *Xho*I sites at the 5'-end and an *Ssp*I site at the 3'-end. This DNA fragment was inserted between the *Eco*RI and *Sac*I sites of the plasmid p(C<sub>2</sub>AT)19 to obtain pP35. Synthetic DNA fragments (44 bp) containing a tandem repeat of either the wild type or the mutant *as-1* sequence were inserted between the *Bgl*III and *Eco*RI sites of pP35. Because it was found that the vector sequence contains several TGA1a-binding sites close to the promoter site, a fragment that does not contain any TGA1a-binding site was inserted to increase the distance between the promoter site and the vector sequences. Between the *Nde*I and *Bgl*III sites of these plasmids, a 268-bp fragment (-343 to -91 of the 35S promoter upstream region plus a linker sequence at -343, the *Eco*RV-*Bgl*III fragment; Fang *et al.*, 1989) was inserted in the reverse orientation. This fragment is known to be devoid of any TGA1a-binding site. The final constructs were named pUWDP and pUMDP, containing the wild type (W) and the mutant (M) *as-1*, respectively (Yamazaki *et al.*, 1990). The plasmids pUWDP and pUMDP (see Fig. 4.2) were used as the DNA templates in the *in vitro* transcription assays. Plasmids identical to pUWDP and pUMDP except lacking the 268-bp inserted region were also constructed and used to confirm the results obtained with pUWDP and pUMDP in chapter 4 (data not shown).

#### (4) Plasmids for microinjection

The following plasmids were used for the microinjection experiments.

- 1) A pUC13 plasmid containing construct A (-90 35S/GUS).
- 2) A pUC13 plasmid containing construct ΔA (-72 35S/GUS).
- 3) A pUC13 plasmid containing construct A-GUSΔ. This is identical to 1) except that a portion of the GUS coding region (*Eco*RV-*Eco*RV fragment) was deleted by digestion with

*EcoRV* followed by religation (-90 35S/GUSΔ).

4) A pUC13 plasmid containing construct A except the promoter contains the -343 to +8 region of the CaMV 35S promoter instead of the -90 to +8 region (35S/GUS).

#### Gel retardation assay

The binding probes were prepared from the plasmids which contain a modified *rbcS-3A* upstream region (shown in Figure 3.1) in pUC or pTZ vectors (Lam *et al.*, 1989; 1990b). The *Bgl*II-*Bst*XI fragments (-166 to -55) were prepared by PAGE and used as competitors. To obtain labeled binding site probe, the fragments were labeled with [ $\alpha$ -<sup>32</sup>P]dNTPs by fill-in reaction with Klenow enzyme and the excess nucleotides removed by gel filtration. The standard binding assay mixture contained DNA-binding protein fractions (lysogen extracts, tobacco nuclear extracts, and purified TGA1a derivatives), 0.1 ng of binding probe (4x10<sup>4</sup> cpm), and 3  $\mu$ g of poly(dIdC) in 10  $\mu$ l of buffer B (20 mM Hepes-KOH, pH 7.5, 40 mM KCl, 1mM EDTA, 10% glycerol, 0.5 mM DTT) supplemented with 0.8 mM PMSF. The mixture was incubated for 20 min at room temperature, before being loaded onto a non-denaturing 0.4% agarose-3% polyacrylamide composite gel or 5% polyacrylamide gel. After electrophoresis, the gel was dried and subjected to autoradiography with an intensifying screen.

#### DNA sequence analysis

Single-stranded DNA was usually used for nucleotide sequence determination. DNA fragments were subcloned into the M13mp18 and 19 vectors and single-stranded DNA was obtained by the M13 phage system (Messing, 1983). Double-stranded DNA sequencing was also carried out using plasmid DNA. Nucleotide sequences of both single- and double-

stranded DNA were determined by the Sequenase sequencing kit (USB), using modified T7 DNA polymerase and deoxyadenosine 5- $\alpha$ -[ $^{35}\text{S}$ ]thiotriphosphate for the reactions, with common primers or synthesized primers. Sequence data were processed by DNASIS and PROSIS (Hitachi) on an IBM PS/2 computer. The homologies shown in Figure 3.8 were detected by visual inspection.

### Methods used in chapter 3

#### Preparation of tobacco nuclear extracts

Tobacco nuclear extracts were prepared essentially according to Green *et al.* (1987). Leaves (~500 g) were collected from 10 tobacco plants (~80 cm tall; ~2.5 month old) grown in a green house. Leaves were washed first with cold distilled water and then with cold homogenization buffer [1M 2-methyl-2,4 pentanediol (hexylene glycol), 10 mM PIPES-KOH, pH 7.0, 10 mM  $\text{MgCl}_2$ , 0.5% v/v triton X-100, 5mM 2-mercaptoethanol, 0.8 mM PMSF]. All subsequent steps were performed at 4°C. The leaf material was disrupted using a Waring Commercial Blender fitted with a tower of new razor blades in 2 l of homogenization buffer. The homogenate was passed through 1000  $\mu\text{m}$  and then 80  $\mu\text{m}$  nylon meshes. The meshes were rinsed with 500 ml of the same buffer and filtrates were pooled. Nuclei in the filtrate were sedimented by centrifugation (3,000xg, 10 min) and the supernatant was decanted. The crude nuclear pellet was gently resuspended with a soft paintbrush in a total volume of 80 ml of nuclei wash buffer (0.5 M hexylene glycol, 10 mM PIPES-KOH, pH 7.0, 10 mM  $\text{MgCl}_2$ , 5mM 2-mercaptoethanol, 0.8 mM PMSF) supplemented with 0.5% v/v triton X-100. Care was taken not to resuspend the white, starch pellet which sedimented faster than nuclear pellet. The suspension was centrifuged at 3,000xg for 5 min. The nuclear pellet was washed with 40 ml of nuclei wash buffer again

and the nuclei repelleted by centrifugation. The pellet was resuspended in 6 ml of nuclei lysis buffer (110 mM KCl, 15 mM Hepes-KOH, pH 7.5, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 5 µg/ml antipain, 5 µg/ml leupeptin) and 0.6 ml of 4M ammonium sulfate was added slowly with gentle mixing. The mixture was agitated gently for 30 min. The chromatin and particulate material were removed from the mixture by centrifugation at 40,000xg for 60 min. Finely ground ammonium sulfate was slowly added to the supernatant with gentle stirring to a final concentration of 0.3 g/ml. After 30-min gentle stirring, the precipitated proteins were collected by centrifugation at 10,000xg for 15 min. The pellet was resuspended in 0.15 ml of nuclear extract buffer (40 mM KCl, 25 mM Hepes-KOH, pH 7.5, 0.1 mM EDTA, 10 % v/v glycerol, 1 mM DTT, 5 µg/ml antipain, 5 µg/ml leupeptin) and the suspension was dialyzed against three changes of the same buffer over a period of 2 hours. After dialysis insoluble material was removed by centrifugation in a microfuge for 10 min. The supernatant fraction was divided into aliquots, frozen in liquid nitrogen, and stored at -80°C. Typically, the nuclear extracts contain ~10 mg/ml of protein.

#### Isolation of cDNA clones encoding TGA1a and TGA1b

PolyA<sup>+</sup> RNA was prepared from leaves of tobacco plants (grown in a green house, ~80 cm tall) adapted in the dark for 2 days. A random-primed cDNA library was constructed in lambda gt11 by using cDNA Synthesis System Plus (Amersham). The library was screened essentially as described by Singh *et al.* (1988). The primary library was plated at a density of ~15,000 pfu/15-cm plate. After six-hour incubation at 37°C, IPTG-impregnated nitrocellulose filters were layered on top of the agar and the plates were further incubated at 37°C for 6 hours. The filters were processed at 4°C as follows. The filters were lifted and incubated in buffer B supplemented with 5% non-fat dry milk for 2



hours. After a brief washing in buffer B the filters were incubated for 4 hours in buffer B containing the labeled binding probe (2.5 ng/ml,  $7 \times 10^5$  cpm/ml) and 5  $\mu$ g/ml sonicated and denatured salmon testis DNA. Synthetic double-stranded *hex-1* oligonucleotides shown below (the TGACG-motif in the bottom strand is underlined, see Fig. 3.1) were concatemerized with T4 DNA ligase until the concatemers contained on the average 8 copies of the sequence.

```

5' TCGACGGCCACGTCACCAATCCGCG      3'
3'      GCCGGTTGCAGTGGTTAGGCGCAGCT 5'

```

The concatemers were labeled by nick translation and used as the binding probe. After incubation with the probe the filters were washed in B buffer for 50 min, briefly dried, and then subjected to autoradiography.

The inserts of the isolated  $\lambda$  clones, hb1, hb2, hb3, hb5, and hb6, were subcloned into plasmid vectors, pBluescript II series. Restriction mapping of the  $\lambda$  clones showed that all except hb6 lost one of the *Eco*RI sites (the inserts were presumed to be cloned into the *Eco*RI site of  $\lambda$ gt11 vector.). To avoid losing any part of the inserts in hb1, hb2, hb3, and hb5, fragments containing the whole insert regions and some vector sequences were subcloned. The 2.2-kb *Eco*RI-*Kpn*I fragment of hb1 (including 1-kb vector sequence on the *Kpn*I side) was cloned into the *Eco*RI-*Kpn*I site of pBluescript II KS- to obtain pHB1B. The *Eco*RI-*Pvu*I fragments of hb3 and hb5 (2.4 and 2.8 kb, respectively; both contained 1.2-kb vector sequence on the *Pvu*I side; the *Pvu*I end was flushed.) were cloned into the *Eco*RI-*Eco*RV site of pBluescript II KS+ to obtain pHB3 and pHB5, respectively. The 8.4-kb *Eco*RI-*Pvu*I fragment of hb2 (including 7.7-kb vector sequence on the *Pvu*I side; the *Pvu*I end was flushed.) was cloned into the *Eco*RI-*Eco*RV site of pBluescript II SK- to obtain

pHB2A. The 1.6-kb *EcoRI* fragment of hb6 (the whole insert) was cloned into pBluescript II KS+ to obtain pHB6A. These subclones were used to prepare DNA fragments for sequence determination, hybridization probes (in chapter 3), and further subcloning of TGA1a (see above).

#### Preparation of lysogen extracts

Lysogens were generated from hb1 and hb2, and the lysogen extracts were prepared, according to Huynh *et al.* (1985) and Ausubel *et al.* (1987).

#### Northern and Southern blot analyses

Preparation of RNA and DNA samples from tobacco plants and Northern and Southern blot analyses were performed essentially as described (Ausubel *et al.*, 1987; Sambrook *et al.*, 1989). The 1150-bp *EcoRI-XhoI* fragment from pHB1B, the 560-bp *EcoRI-FokI* fragment from pHB6A, and the 1.3-kb *HindIII-XbaI* fragment from the  $\beta$ -ATPase gene *atp2-1* (Boutry and Chua, 1985) were labeled by the random primed DNA labeling kit (Boehringer) and used as the TGA1a, the TGA1b, and the  $\beta$ -ATPase probe, respectively. The specific activities of the probes were  $\sim 1.8 \times 10^9$  cpm/ $\mu$ g. The same filter was sequentially hybridized with TGA1a, TGA1b, and  $\beta$ -ATPase probes, after washed off the previous probe by boiling in water. Filters were prehybridized and hybridized in a buffer containing 50% formamide, 10% dextran sulfate, 25  $\mu$ g/ml sonicated and denatured salmon testis DNA, 1xDenhardt's solution, and 5xSSC at 43°C. About  $4 \times 10^6$  cpm/ml of probes were used. The filters were washed in 2xSSC, 0.1% SDS and then in 0.1xSSC, 0.1% SDS at room temperature. They were washed finally in 0.1xSSC, 0.1% SDS at 50°C, briefly dried, and then subjected to autoradiography with an intensifying screen.

## Methods used in chapter 4

### Overproduction and purification of TGA1a

BL21(DE3)/plysS/pKT7T1A was cultured in M9ZB medium (6g Na<sub>2</sub>HPO<sub>4</sub>, 3g KH<sub>2</sub>PO<sub>4</sub>, 0.5g NaCl, 1g NH<sub>4</sub>Cl, and 10g bactotryptone per 1 l) at 37°C until the A<sub>600</sub> reached 0.4. IPTG was added to the culture to a final concentration of 0.4 mM and the bacteria were incubated at 37°C for two hours before harvesting. The following procedure was performed at 4°C. The bacteria (6 g wet weight, from 2-l culture) were suspended in 80 ml of buffer E (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 1 mM PMSF), lysed by two cycles of freeze-thaw and then extracted with 1 M NaCl. The crude extract was adjusted to 40% saturation of ammonium sulfate and the precipitate was collected. The precipitate was suspended in buffer A-0.04 (buffer E plus 20% glycerol, 0.1% Nonidet P-40, 40 mM KCl) and dialyzed against the same buffer. The fraction was applied onto a DE52 (Whatman) column (bed volume 33 ml) equilibrated with buffer A-0.04 and the flow-through fractions were pooled. The DE52 fraction was then applied onto a P11 (Whatman) column (bed volume 5 ml) equilibrated with buffer A-0.04. After consecutive washings with buffer A-0.04 and buffer A-0.2 (same as A-0.04 except 0.2 M KCl), the column was eluted with buffer A-0.5 (same as A-0.04 except 0.5 M KCl) and the peak protein fractions were pooled. The P11 fraction was passed through two DNA affinity chromatography steps to remove a non-specific stimulatory activity of transcription in the fraction, which was observed when analyzed in a wheat germ system. The buffer of the fraction was exchanged to buffer B-0.2 (same as buffer B except 0.2 M KCl) by gel filtration on Sephadex G-25 (Pharmacia) and one third of the total fraction was applied onto the mutant DNA affinity column (bed volume 5 ml) equilibrated with the same buffer. The column was eluted with buffer B-0.4 (same as B except 0.4 M KCl) and the eluant was applied onto the wild type DNA affinity

column (bed volume 5 ml) equilibrated with buffer B-0.4. The column was eluted with buffer B-0.8 (same as B except 0.8 M KCl) and the eluted fractions were pooled as the purified fraction of TGA1a.

The DNA affinity columns were prepared essentially as described (Kadonaga and Tjian, 1986), using HW and HM double-stranded oligonucleotides that contain the wild type and mutant *hex-1* sequences, respectively. Concatemerized HW and HM were coupled to CNBr-activated Sepharose 4B (Pharmacia) to prepare the wild type and mutant DNA affinity columns, respectively. *hex-1* is a cis-element (-180 to -160) of the wheat histone H3 promoter (Mikami *et al.*, 1987). It contains one copy of the TGACG motif that binds to TGA1a (Katagiri *et al.*, 1989). The nucleotide sequences of HW and HM are shown below.

#### HW

```

5' CGAAGCTTCGGCCACGTCACCAATC 3'
3'          CCGTGCAGTGGTTAGGCTTCGAAGC 5'

```

#### HM

```

5' CGAAGCTTCGGCCACGCGTCCAATC 3'
3'          CCGTGCGCAGGTTAGGCTTCGAAGC 5'

```

The TGACG motif of HW is underlined and the mutations in HM are shown in bold letters. HW binds TGA1a whereas HM does not when analyzed as competitors in gel retardation assays (data not shown).

#### In vitro transcription assays using the HeLa cell system

Partially purified fractions of RNA polymerase II, TFIIB, TFIID, and TFIIE/F were

prepared as described (Horikoshi *et al.*, 1988b). The activities of the factors were: RNA polymerase II, 0.7 units/ $\mu$ l; TFIIB, 0.85 units/ $\mu$ l; TFIID, 1.5 units/ $\mu$ l and TFIIE/F, 0.7 units/ $\mu$ l. The reaction mixture (25  $\mu$ l/assay) contained 10 mM Tris-HCl (pH 8.0), 40 mM HEPES-KOH (pH 7.9), 50 mM KCl, 2 mM MgCl<sub>2</sub>, 10% glycerol, 1 mg/ml BSA, 8 mM DTT, 100  $\mu$ M ATP, 100  $\mu$ M UTP, 20  $\mu$ M CTP, 5  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]CTP (800 Ci/mmol), 50  $\mu$ M 3'-O-methyl GTP, 0.15 unit RNaseT1 (Pharmacia), 0.8 unit RNasin (Promega), 0.2  $\mu$ g DNA template, 0.5  $\mu$ l TFIIB, 0.5  $\mu$ l TFIIE/F, 1  $\mu$ l TFIID, 0.25  $\mu$ l RNA polymerase II and the indicated amount of TGA1a. The standard incubation was at 30°C for 20 min. After the incubation, the reaction was terminated by the addition of 75  $\mu$ l of stop solution (0.5% SDS, 10 mM EDTA, 150 mM sodium acetate, 50  $\mu$ g/ml tRNA). After extraction with phenol and chloroform, the transcripts were analyzed on a 6% sequencing gel. For the 'depleted' HeLa general factors, the mixture of TFIIB, TFIID, TFIIE/F and RNA polymerase II was preincubated with 10 ng/assay of HW at 30°C for 30 min before the reaction. In experiments involving a preincubation step, nucleoside triphosphates and the indicated components were omitted from the preincubation mixture. The components indicated in the figures were added according to the time tables. The reactions were stopped and the transcripts were analyzed as described above. Amounts of transcripts were measured by determining the radioactivity in excised gel bands using a scintillation counter.

#### Primer extension assay

*in vitro* transcription with (100 ng/assay) or without TGA1a was carried out under standard assay conditions except that [ $\alpha$ -<sup>32</sup>P]CTP was omitted from the reaction mixture. The transcripts were analyzed by primer extension as described (Hai *et al.*, 1988) using an oligonucleotide primer complementary to +94 to +115 of the template, pUWDP. The

amount of the primer used was in excess to that of the transcripts. For reference, pUWDP was sequenced with the same  $^{32}\text{P}$ -labeled primer using a Sequenase sequencing kit (USB; except using non-radioactive dATP instead).

#### Amino-terminal sequencing of protein

The three polypeptides in the TGA1a preparation were separated by SDS-PAGE, blotted onto polyvinylidene difluoride membrane (Millipore; Matsudaira, 1987) and sequenced by a gas phase (Applied Biosystems Model 470A) or pulsed liquid protein sequencer (Model 477A). Both sequencers were equipped with on-line PTH analyzers (Model 120A) and data analysis modules (Model 900A). The amino-terminal sequencing was carried out by the Protein Sequencing Facility at the Rockefeller University.

#### Southwestern blot

Southwestern blot analysis (Miskimins *et al.*, 1985) was performed as follows. All procedure was performed at 4°C. Protein fractions were separated by 10% SDS-PAGE and the proteins were blotted onto nitrocellulose filter by electroblotting. The filter was incubated in buffer B+M (buffer B supplemented with 0.25% non-fat dry milk) for 2 hours. A 112-bp fragment containing *hex-1* (Fig. 3.1; Katagiri *et al.*, 1989) was labeled by Klenow fill-in and used as the binding probe. The filter was incubated in the same buffer containing 5 ng/ml of the probe ( $7 \times 10^5$  cpm/ml) and 10  $\mu\text{g}/\text{ml}$  of poly(dIdC) for 5 hours. As a competitor, 18 ng/ml of either HW or HM was added to the binding mixture. The filter was washed with buffer B+M twice for a total of 10 min.

Methods used in the appendix to chapter 4

### Preparation of wheat germ extract

Wheat germ used in this study was provided by Nisshin Flour Milling (Japan). The extract was prepared essentially as described (Yamazaki and Imamoto, 1987). All procedure was performed at 4°C. Wheat germ (10g) was suspended in 40 ml of TED (10 mM Tris-HCl, pH7.8, 5mM EDTA, 1 mM DTT) containing 75 mM ammonium sulfate, 1g of polyvinylpolypyrrolidone and 0.5 g of quartz sand, and ground for 5 min in a mortar with pestle. The homogenate was centrifuged at 20,000xg for 15 min. The supernatant was filtered through two layers of Miracloth. 0.075 vol of 10% polymin-P was added to the filtrate and the mixture was centrifuged (20,000xg, 15 min) to precipitate chromatin (RNA polymerase II precipitates together.). The pellet was washed with 20 ml of TED containing 75 mM ammonium sulfate. The pellet was extracted with 20 ml of TED containing 200 mM ammonium sulfate and the suspension was centrifuged (20,000xg, 10 min). Ammonium sulfate was added to the supernatant fraction to 15% saturation. The mixture was centrifuged at 20,000xg for 10 min and the supernatant was saved. The supernatant fraction was adjusted to 33% saturation of ammonium sulfate and the precipitate was collected by centrifugation. The pellet was dissolved in 0.8 ml of 10 mM Hepes-KOH (pH 8.0), 1mM DTT. The insoluble material was removed by centrifugation in a microfuge for 10 min and the supernatant was passed through a gel-filtration column (NAP-10, Sephadex G-25) to remove remaining metal ions and mononucleotides. 0.3 ml of glycerol was added to the eluted fraction. The extracts (WGE) were aliquoted, frozen in liquid nitrogen, and stored at -80°C.

### *In vitro* transcription assays using wheat germ extract

The standard assay mixture (50 µl) contained 20 mM Hepes-KOH (pH 8.0), 1 mM

MnCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 1 mM DTT, 20 mM ammonium sulfate, 0.4 mM ATP, 0.4 mM UTP, 0.04 mM CTP, 10 µCi of [ $\alpha$ -<sup>32</sup>P]CTP (specific activity, 400-800 Ci/mmol), 0.4 µg of DNA template, 10% glycerol, and 15 µl of WGE (final concentration, 0.27 mg of protein per ml). The mixture was incubated at 30°C for 20 min with or without the indicated amount of TGA1a. The reaction was terminated by the addition of stop solution (0.5% SDS, 10 mM EDTA, 150 mM sodium acetate, 50 µg/ml of tRNA). The mixture was extracted with phenol/chloroform and RNA was precipitated with ethanol. RNA was suspended in 5M urea/1 mM EDTA/0.1 % bromophenol blue, incubated at 95°C for 5 min, and analyzed on a 4% polyacrylamide gel containing 8M urea. After electrophoresis, the gel was dried and subjected to autoradiography. The amount of transcripts accumulated, corresponding to the specific 380-nt RNA, were measured by densitometric scanning of the autoradiograms.

## Methods used in chapter 5

### Microinjection into plant materials

Microinjection was carried out by my collaborators, Gunther Neuhaus and Gabriele Neuhaus-Url (ETH Zentrum, Zurich, Switzerland). In brief, cotyledons of 5 to 7-day old tobacco seedlings were used for microinjection. Approximately 5 pl/cell of materials (protein, RNA, or DNA solution; the amounts are described in chapter 5.) were injected into nuclear areas of individual cells in epidermal and subepidermal cells of the cotyledons. The injected seedlings were incubated for 24 hours at 23°C in light and then subjected to histochemical staining of the GUS activity.

### Histochemical staining of the GUS activity



Histochemical staining of the GUS activity was carried out as described (Jefferson *et al.*, 1987; Benfey *et al.*, 1989). Tobacco seedlings were placed directly in the substrate solution [1 mM 5-bromo-4-chloro-3-indolyl glucuronide (X-Glc), 50 mM sodium phosphate buffer (pH 7.0)] and incubated for 16 hours in a humidified chamber at 37°C. Coverslips were placed on the slides before viewing.

## Methods used in chapter 6

### DNA binding assays of TGA1a derivatives

The plasmid constructs for TGA1a and its derivatives (pKT1A and pΔ series) were linearized with *Bam*HI (The *Bam*HI site is located 3' to the coding region; see Fig. 6.1) and transcribed with T7 RNA polymerase *in vitro* using an RNA Transcription Kit (Stratagene). Approximately 1.5 μg of RNA was obtained when 1 μg of plasmid was used as a template. The synthesized RNA species were translated *in vitro* with rabbit reticulocyte lysate (Promega) according to the supplier's instructions. ~0.1 μg of RNA was usually used for each 12.5 μl reaction and the synthesized polypeptides were labeled with [<sup>35</sup>S]methionine. After translation the mixture was tested for DNA-binding activity of the derivatives by gel retardation. The binding reaction (15 μl) contained 5 μl of the mixture, 0.6 μg of BSA, *hex-1* binding probe (5x10<sup>4</sup> cpm, ~0.12 ng), and 4.5 μg of poly(dIdC) in buffer B. After 20-min incubation at room temperature, the mixture was analyzed by 5% PAGE. After electrophoresis, the gel was dried on a DE81 paper (Whatman) and subjected to autoradiography. To shield the film from the emission from <sup>35</sup>S incorporated into the translation products and at the same time to intensify the signal from <sup>32</sup>P incorporated in the binding probe, an intensifying screen was inserted between the gel and the X-ray film with the intensifying side facing the film. The *in vitro* translation products (0.5 μl aliquot)

were also analyzed by 15% SDS-PAGE to verify their sizes and amounts. After electrophoresis, the gel was treated with Enlightening (NEN) according to the supplier's instructions before being subjected to autoradiography. The autoradiograms showed that for a given TGA1a derivative, a polypeptide of the expected size was synthesized as the major product of the reaction. Moreover, a similar number of polypeptide molecules were synthesized when similar amounts of RNA were used (taking into account differences in the number of methionines in the derivatives). The only exception was  $\Delta N80\Delta C108$ , for which the corresponding polypeptide was not detected. This was probably because of its low methionine content (only one methionine for starting methionine, which could be removed) and/or its small size (smaller than 5 kd, below the resolution range of the gel).

Equal amounts of RNA were used for cotranslation experiments involving two +DS or two -DS derivatives (see chapter 6 for DS). For experiments involving a -DS and a +DS derivative, 20 times more RNA was used for the -DS derivative than one for the +DS derivatives. In both cases, the total amount of RNA in the translation mixture remained the same. The cotranslation products were analyzed for their DNA-binding activity as described above.

For mixing experiments, the RNAs were translated separately. After translation, 2.5  $\mu$ l each of the translation products were added to the mixture for binding assay except that the binding probe and poly(dIdC) were omitted in the mixture. After 30-min incubation at room temperature, the binding probe and poly(dIdC) were added to initiate the binding reaction.

#### Overproduction and purification of $\Delta N80$ and $\Delta N80\Delta C144$

$\Delta N80$  and  $\Delta N80\Delta C144$  were overproduced in *E. coli* strain BL21(DE3)/plysS as

described for the overproduction of TGA1a. The derivatives were purified from the bacterial extracts as described above with the following modifications. The P11 column step was omitted; the flow-through fractions from the DE52 column were directly applied to the mutant DNA-affinity column after the salt concentration in the buffer was adjusted to 0.1 M KCl. For purification of  $\Delta N80$ , the mutant DNA-affinity column was washed with buffer B-0.1 (same as B except 0.1 M KCl) and then eluted with buffer B-0.4. The eluted fraction was diluted with the same volume of buffer B-0 (same as B except no KCl) and loaded onto the wild type DNA-affinity column. The column was washed with buffer B-0.2 and eluted with buffer B-0.8. This fraction was used as the purified fraction of  $\Delta N80$ .  $\Delta N80\Delta C144$  was purified by the same procedure except that the wild type DNA affinity column was eluted with buffer B-0.6.

### **Chapter 3: cloning of the DNA-binding protein genes**

Further characterization of DNA-binding proteins identified by *in vitro* binding assays would require the isolation of cDNA clones encoding such proteins. The availability of the cDNA clones would allow us to investigate the detailed molecular structures of the proteins, to overproduce the wild type proteins and their mutant derivatives, and to use the proteins for functional studies. A conventional way of cloning genes encoding such proteins entails protein purification as the first step. Information derived from the amino acid sequences of the purified proteins can then be used for gene cloning. Because of their low abundance in nuclear extracts, purification of a DNA-binding protein is, in general, a tedious task. In the case of plant material, this problem is further compounded by the difficulty of obtaining a large amount of nuclear extracts of high quality. Therefore, it is not surprising that so far there has been no report on the purification of any DNA-binding protein from plants.

An alternative method of cloning genes encoding DNA-binding proteins was developed by Singh *et al.* (1988) and Vinson *et al.* (1988). In this method, expression cDNA libraries are screened with radio-labeled binding site sequences and positive cDNA clones are detected subsequently by autoradiography. This method, however, has several limitations: it cannot be applied to DNA-binding proteins that are composed of heterologous subunits (eg. CP1A and CP1B in human, Chodosh *et al.*, 1988a; HAP2 and HAP3 in yeast, Chodosh *et al.*, 1988b) or require eukaryote-specific post-translational modification for binding to their cognate sites. Notwithstanding these limitations, this screening strategy can circumvent the difficulty of protein purification and facilitate studies of DNA-binding proteins, particularly from plant sources.

One of the well-defined functional cis-elements in the 35S promoter is *as-1*, which

confers preferential expression in root (see chapter 1). This 21-bp element is located in the -83 to -63 region and contains a tandem repeat of the sequence motif, TGACG. A DNA-binding protein, designated ASF-1, has been detected in tobacco nuclear extracts, that specifically binds to *as-1* *in vitro*. Mutations in the TGACG motifs of *as-1* drastically decrease the ASF-1 binding *in vitro* and abolish the ability of *as-1* to confer root-specific expression on the *rbcS-3A* promoter *in vivo* (Lam *et al.*, 1989). This correlation between the ASF-1 binding ability *in vitro* and the functional activity of *as-1* *in vivo* suggests that ASF-1 is responsible for the function of *as-1* *in planta* (Lam *et al.*, 1989).

Because *as-1* is derived from a plant viral promoter, the question arises which cellular genes as well as other genes of pathogen origin are regulated by ASF-1. Two other sequences that contain the TGACG-motif have been characterized as functional cis-elements in plants. They are *hex-1* from the wheat H3 promoter (Mikami *et al.*, 1987) and *nos-1* from the nopaline synthase promoter of the *Agrobacterium* Ti-plasmid (An *et al.*, 1986; Depicker *et al.*, 1982). The histone H3 element, *hex-1* (-180 to -160), contains one copy of the TGACG motif, whereas the nopaline synthase element, *nos-1* (-131 to -111), contains a TGACG motif and a TGAGC motif.

In this chapter, I first show that the nuclear factor ASF-1 binds to all three elements, *as-1*, *hex-1*, and *nos-1* whereas another protein in tobacco nuclear extracts, called *hex-1*-specific binding factor (HSBF), binds only to *hex-1*. Second, I describe the isolation of tobacco cDNA clones encoding two TGACG-specific DNA-binding proteins (TGA1a and TGA1b). Nucleotide sequence analysis of the clones show that both encoded proteins contain a bZIP domain, which is a motif for DNA-binding common among eukaryotic transcription factors (see chapter 1). The basic regions of TGA1a and TGA1b show high sequence homology to those of other bZIP class of DNA-binding proteins. Based on their

binding specificities to the three TGACG-containing elements, I suggest that TGA1a is a good candidate for ASF-1 and that TGA1b is a candidate for HSBF.

## RESULTS

### TGACG-binding proteins in tobacco nuclear extracts

The tobacco nuclear factor ASF-1 binds to *as-1*, which contains a tandem TGACG motif (Lam *et al.*, 1989; Prat *et al.*, 1989). A computer search of the data bank revealed that this motif is present in the upstream sequences of several plant genes and genes encoded by Ti-plasmids. The wheat histone H3 promoter (Mikami *et al.*, 1987) and the nopaline synthase promoter (An *et al.*, 1986) were investigated further. The histone H3 promoter contains a single TGACG motif in the -180 to -160 region (*hex-1*), while the nopaline synthase promoter contains a TGACG motif and its variant, TGAGC, between -131 and -111 (*nos-1*). Complementary oligonucleotides corresponding to the sequence elements as well as mutant sequences with mutations in the TGACG motif were synthesized (Fig. 3.1). The sequences between -109 and -89 of *rbcS-3A* were replaced by the *as-1*, *hex-1*, *nos-1*, and their mutant derivatives (Fig. 3.1). The resulting chimeric promoter fragments (-166 to -55) were used as binding site probes and competitors in gel retardation assays with nuclear extracts from tobacco leaf. Figure 3.2 shows that the tobacco ASF-1 binds to all three TGACG-containing elements, *as-1*, *hex-1* and *nos-1*. Formation of the DNA-protein complex corresponding to ASF-1 (marked as ASF-1) is detected with each of the three elements. In each case, the complex formation is sensitive to competition by the wild type element but not by the mutant derivative. When *hex-1* was used as a binding site probe, an additional DNA-protein complex, that migrates faster than the complex with ASF-1, was observed (Fig. 3.2, marked as HSBF). The DNA-binding protein corresponding to this

Figure 3.1 Binding site probes used in gel retardation assays.

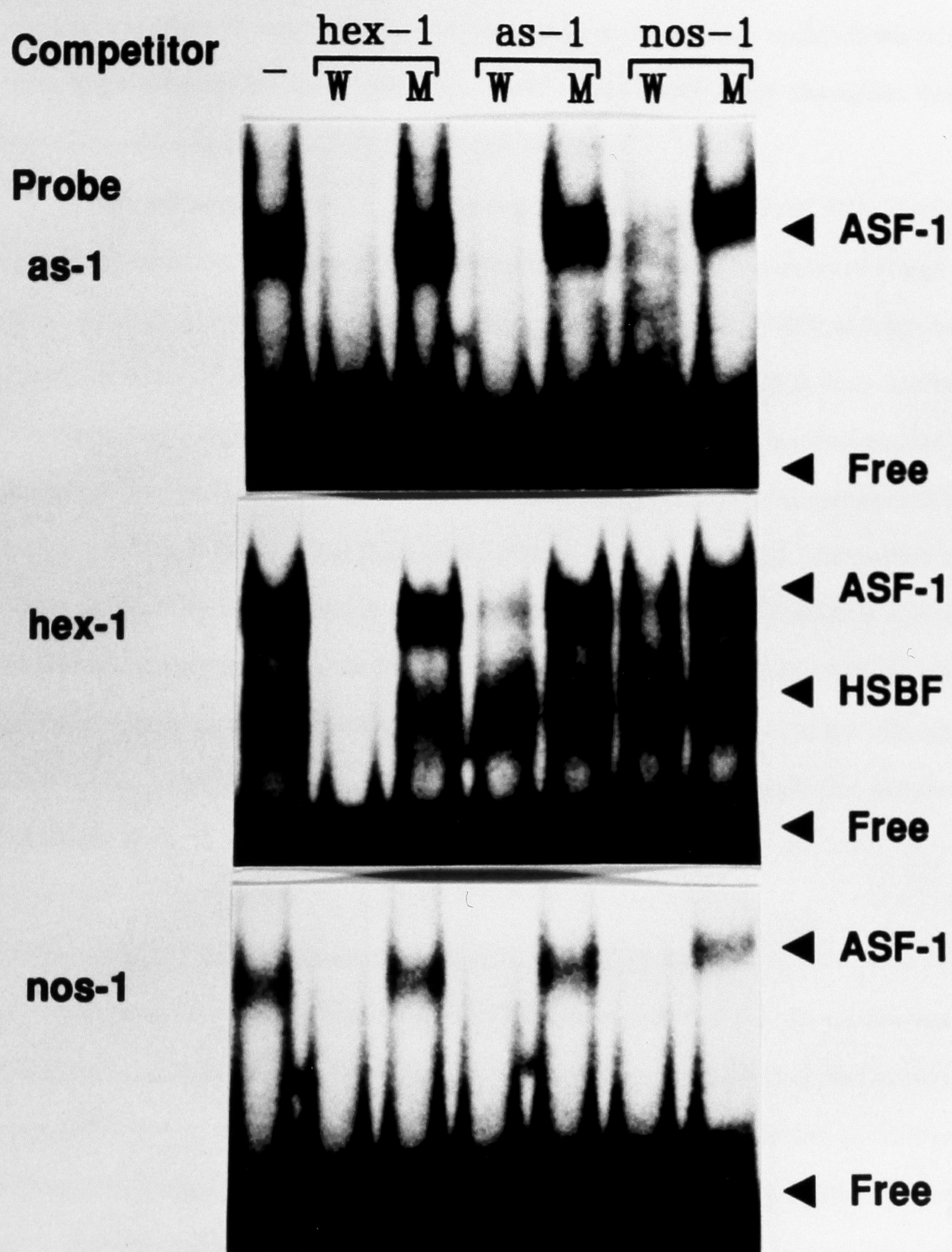
The -109 to -89 region of the *rbcS-3A* upstream fragment (-166 to -55; Fluhr *et al.*, 1986) was substituted with the wild type (W) or mutant (M) sequences. The mutated bases are blocked. Lines over or under the wild type nucleotide sequences indicate TGACG motifs or a similar motif, TGAGC. Each of these motifs represents a single binding site of ASF-1 and TGA1a. The numbers over the wild type sequences show the nucleotide positions in the original promoters (wheat histone H3 promoter, CaMV 35S promoter, and T-DNA nopaline synthase promoter for *hex-1*, *as-1*, and *nos-1*, respectively) (Mikami *et al.*, 1987; Lam *et al.*, 1989; An *et al.*, 1986).





Figure 3.2 Tobacco nuclear extracts contain DNA-binding proteins specific for the three TGACG-containing elements.

Reciprocal competitions of the three TGACG-containing elements were performed in gel retardation assays. The amount of tobacco nuclear extracts used was 7.5  $\mu$ g of protein per lane. The competitors were used in 400-fold molar excess over the binding site probes. Specific DNA-protein complexes corresponding to ASF-1 and HSBF are indicated. F, free DNA; W, wild type binding site; M, mutant binding site.



DNA-protein complex was named *hex-1* specific binding factor (HSBF). Formation of this complex is sensitive to competition by *hex-1* but not by *as-1* and *nos-1*; mutant forms of all these three elements are also ineffective. These results indicate that among the three TGACG-containing elements, HSBF binds only to *hex-1*.

With either *as-1* or *nos-1* as a binding site probe, an additional DNA-protein complex that migrates more slowly than the complex marked as ASF-1 is obtained at higher extract concentrations. *as-1* (Lam et al., 1989) and *nos-1* (Lam et al., 1990b) each has two TGACG-like motifs and therefore two DNA-binding sites for ASF-1. It is likely that the faster migrating complex (marked as ASF-1 in Fig. 3.2) corresponds to occupation of only one of the two binding sites by ASF-1 while the slower migrating complex corresponds to occupation of both binding sites (Lam et al., 1990a). This interpretation is supported by recent studies with *as-1* mutants in which only one of the two TGACG-motifs is altered. At high extract concentrations the slower migrating complex is detected only with the wild type *as-1* but not the single site mutants. By contrast, the mutants as well as the wild type can form the faster migrating complex at low extract concentrations (X.-F.Qin, personal communication).

#### Isolation of cDNA clones encoding DNA binding proteins for *hex-1*

Because *hex-1* binds to both ASF-1 and HSBF, concatemers of this element was used as a probe to screen a tobacco cDNA library in an attempt to isolate cDNA clones encoding these DNA-binding proteins. The cDNA library was constructed in a lambda gt11 vector with random-primed cDNA derived from leaf polyA<sup>+</sup> RNA of adult tobacco plants that had been dark-adapted for two days. The IPTG-induced library was screened with concatemerized, synthetic *hex-1* (average 8 copies of 21-bp *hex-1* shown in Figure 3.1). Five

positive clones were obtained from a primary library consisting of  $6 \times 10^4$  recombinants. All five clones showed binding to *hex-1* reproducibly in the secondary and tertiary screenings. Also, all five clones showed specific binding to the wild type *hex-1* but not to the mutant form of *hex-1* on filters. The five lambda clones, named hb1, 2, 3, 5, and 6, were divided into two groups according to the nucleotide sequences of their cDNA inserts (shown below). The protein encoded by the first group, which includes hb1, 3, and 5, was named TGA1a and the protein encoded by the second group, which includes hb2 and 6, was named TGA1b.

#### Binding specificities of TGA1a and TGA1b

The binding specificities of TGA1a and TGA1b were confirmed by gel retardation assays with the lysogen extract of hb1 and hb2, respectively (Figure 3.3). The lysogen extract of hb1 gives a DNA-protein complex (marked as TGA1a) when *hex-1* is used as a binding site probe. The complex formation is sensitive to competition by the wild type form of either *hex-1*, *as-1*, or *nos-1*, but not by their mutant forms. This result indicates that TGA1a specifically binds to *hex-1*, *as-1*, and *nos-1*, similar to the binding specificity of ASF-1. The lysogen extract of hb2 also gives a DNA-protein complex (marked as TGA1b) when *hex-1* is used as a binding site probe. In contrast to the case of TGA1a, the complex formation is sensitive only to the wild type *hex-1*. This result demonstrates that similar to HSBF, TGA1b binds only to *hex-1*. Differences in the nucleotide sequence flanking the TGACG motif may determine specific binding to either one or both of the proteins.

#### Sequence analysis of TGA1a

The insert DNA sequences of hb1, 3, and 5 were determined. Sequence analysis showed that hb1, 3, and 5 appear to be derived from the same mRNA species. Each of

Figure 3.3 DNA binding specificities of TGA1a, and TGA1b.

Gel retardation assays with lysogen extracts (6.8  $\mu$ g of protein/lane) of hb1 (TGA1a) and hb2 (TGA1b) are shown. The *hex-1* element was used as a binding site probe. The competitors used were wild type (W) and mutant (M) forms of *hex-1*, *as-1*, and *nos-1* and their sequences are shown in Figure 3.1. The specific DNA-protein complexes corresponding to TGA1a and TGA1b are indicated. F, free DNA.

**Probe**

**hex-1**

**Competitor** —

**hex-1**  
W M

**as-1**  
W M

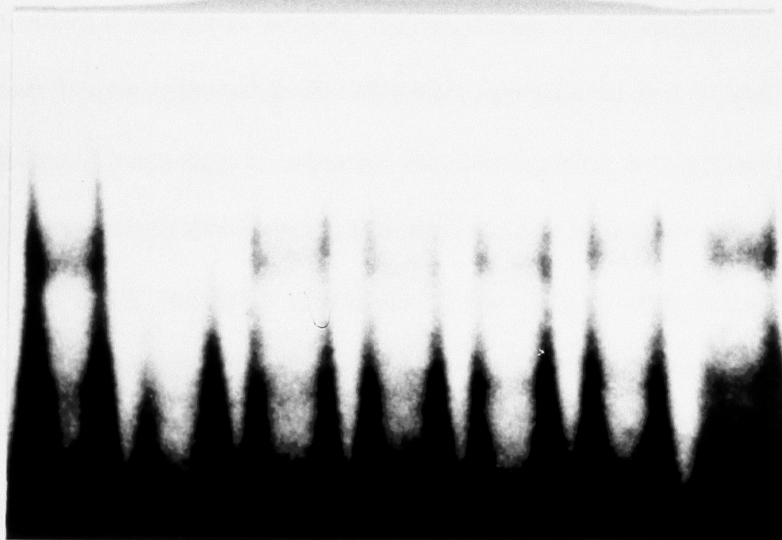
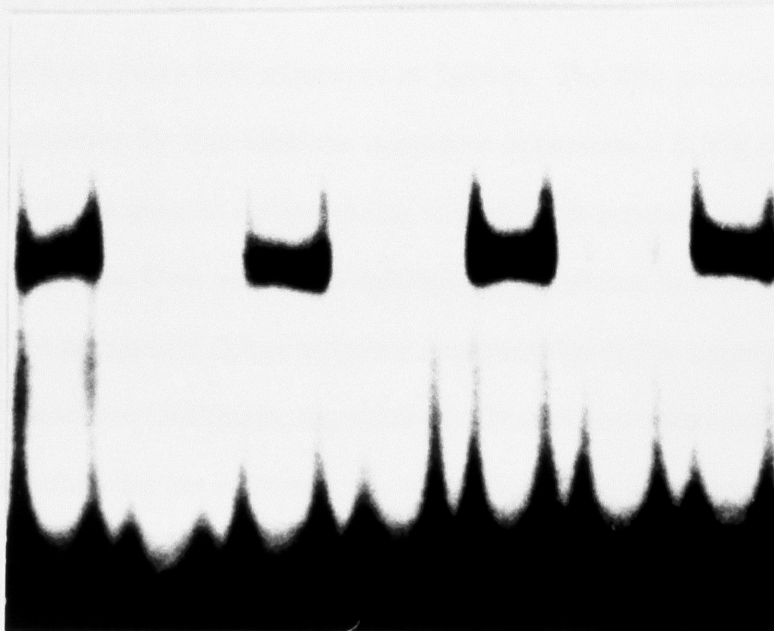
**nos-1**  
W M

**TGA1a** ►

**F**

**TGA1b** ►

**F**



them has a single large open reading frame (ORF) of 1,117 bp. Because no other significantly large ORF was found, it was concluded that this ORF encodes TGA1a. However, the insert sequences do not contain any stop codon upstream of this ORF. To determine the largest ORF starting with a methionine codon, genomic clones of TGA1a were isolated by hybridization using the hb1 insert sequence as a probe. DNA sequence analysis of TGA1a genomic clones revealed that hb1, 3, and 5 lack the first two nucleotides of the first methionine codon of the longest ORF (H.Fromm, unpublished). Figure 3.4 shows the nucleotide and deduced amino acid sequences of TGA1a. The first nucleotide of this ORF and the first methionine for this ORF are numbered as position 1 in Fig. 3.4. Hb1, 3, and 5 contain insert DNA sequences corresponding to nucleotide numbers 3-1264, 3-1278, and 3-1620, respectively. The DNA sequences upstream of nucleotide number 3 is derived from those of the TGA1a genomic clones and were confirmed by cDNA sequence obtained by the 5'-end PCR technique (H.Fromm, unpublished). The longest open reading frame (ORF) starting with a methionine (an in-frame stop codon is located at the nucleotide number -18) is 1,119 bp, corresponding to 373 amino acid residues (molecular weight: 42,301).

The ORF in hb1 was out of frame from the *lacZ* reading frame in the lambda gt11 vector. The ORFs in hb3 and 5 were oriented in the direction opposite to that of *lacZ*. Probably, each of the cDNA inserts has a cryptic sequence that can function as a bacterial promoter site as well as a cryptic sequence downstream to it that can function as a bacterial translation initiation site. The DNA sequence contained in the hb1, 3, and 5 insert sequences has a sequence (nucleotide number 26-TTGCTG...16 bp...TATATG) that is very similar to the consensus sequence of *E. coli*  $\sigma$ 70 promoter (TTGACA...16-19bp...TATAAT; Rosenberg and Court, 1979). The DNA sequence also has four methionine codons (amino

Figure 3.4 DNA and deduced amino acid sequences of TGA1a.

The deduced amino acid sequence for the longest ORF starting with a methionine codon is shown in Triple letter code below the DNA sequences. The nucleotide and amino acid residue corresponding to the presumptive translation initiation site are numbered as 1.



	-18	TAA	ACC	TCA	GTC	CAT	ACA		-
1	ATG AAT TCT TCA ACG TAC ACC CAA TTT GCT GCC TCA AGA AGG ATG GGT ATA TGC GAT CCG ATC CAT CAA CTT GGC	75							
1	Met Asn Ser Thr Tyr Thr Gln Phe Ala Ala Ser Arg Arg Met Gly Ile Cys Asp Pro Ile His Gln Leu Gly	25							
76	ATG TGG GAT GAT TTC AAT AGT AGT TTC CCA AGT ACA TCG GCA ACC ATG ATT TTA GAA GTT GAT AAA TGC CTA GAG	150							
26	Met Trp Asp Asp Phe Asn Ser Ser Phe Pro Ser Thr Ser Ala Thr Met Ile Leu Glu Val Asp Lys Cys Leu Glu	50							
151	GAC CAG ATA CCA ATT ATG GAG AAA AGA CTA GAC AAC GAG ACA GAA GAC ACT TCG CAT GGA ACA GTA GGG ACT TCT	225							
51	Asp Gln Ile Pro Ile Met Glu Lys Arg Lys Arg Leu Asp Asn Glu Thr Glu Asp Thr Ser His Gly Thr Val Gly Thr Ser	75							
226	AAC AGA TAT GAA CCG GAA ACA AGT AAA CCC GTC GAG AAG GTA CTT AGA CGT CTT GCA CAA AAC CGC GAG GCT GCT	300							
76	Asn Arg Tyr Glu Pro Glu Thr Ser Lys pro Val Glu Lys Val Leu Arg Arg Leu Ala Gln Asn Arg Glu Ala Ala	100							
301	CGT AAA AGC CGT TTG CGG AAG AAG GCC TAT GTT CAG CAG TTA GAA AAT AGT AAA TTG AAG CTG ATT CAA CTG GAA	375							
101	Arg Lys Ser Arg Leu Arg Lys Lys Ala Tyr Val Gln Gln Leu Leu Asn Ser Lys Leu Lys Leu Ile Gln Leu Glu	125							
376	CAA GAA CTA GAA CGC GCC AGA AAA CAG GGC ATG TGT GTA GGT GGT GTA GAT GCT AGC CAG CTA AGT TAC TCT	450							
126	Gln Glu Leu Glu Arg Ala Arg Lys Gln Gln Met Cys Val Gly Gly Gly Val Asp Ala Ser Gln Leu Ser Tyr Ser	150							
451	GGA ACC GCT AGC TCA GGA ACT GCT GTA TTT GAT ATG GAG TAT GGT CAC TGG GTA GAA GAG CAA ACT AGA CAA ACA	525							
151	Gly Thr Ala Ser Ser Gly Thr Ala Val Phe Asp Met Glu Tyr Gly His Trp Val Glu Glu Gln Thr Arg Gln Thr	175							
526	AAT GAC TTA AGG ATT GGT TTG CAT TCT CAA ATT GGT GAA GCG GAA TTG CGC ATT ATT GGT GAT GGT TAC CTG AAC	600							
176	Asn Asp Leu Arg Ile Ala Leu His Ser Gln Ile Gly Glu Ala Glu Leu Arg Ile Ile Val Asp Gly Tyr Leu Asn	200							
601	CAC TAC TTT GAT CTC TTC CGC ATG AAA GCT AAA GCT GCT AAA GAA CAA TCC GAG CTT CTA AAG GTT CTC ACA CCG	675							
201	His Tyr Phe Asp Leu Phe Arg Met Lys Ala Thr Ala Lys Ala Asp Val Leu Tyr Ile Met Ser Gly Met Trp	225							
676	AAG ACA TCT GCC GAG CGC TTT ATG TGG ATT GGA GGG TTT CGG CCA TCC GAG CTT CTA AAG GTT CTC ACA CCG	750							
226	Lys Thr Ser Ala Glu Arg Phe Phe Met Trp Ile Gly Gly Phe Arg Pro Ser Glu Leu Lys Val Leu Thr Pro	250							
751	CAT CTT GAG CTC TTG ACA GAA CAA CAA CTT CGA GAG GTT TGT AAC CTG ACC CAA TCA TGT CAG CAA GCA GAA GAC	825							
251	His Leu Glu Leu Leu Thr Glu Gln Gln Leu Arg Glu Val Cys Asn Leu Thr Gln Ser Cys Gln Gln Ala Glu Asp	275							
826	GCC TTG TCA CAA GGA ATG GTA AAA CTC CAC CAG ATT CTT GCC GAG GCT GTT GCA GCT GGC CGA CTA GGA GAA GGA	900							
276	Ala Leu Ser Gln Gly Met Val Lys Leu His Gln Ile Leu Ala Glu Ala Val Ala Ala Gly Arg Leu Gly Gly Gly	300							
901	AAT TAC ACT CTT CCG CAG ATG GGG CCT GCC ATC GAA AAG TTG GAA GAT CTT GTT AGG TTC GTA AAT CAG GCG GAT	975							
301	Asn Tyr Thr Leu Leu Pro Gln Met Gly Pro Ala Ile Glu Lys Leu Glu Asp Leu Val Arg Phe Val Asn Gln Ala Asp	325							
976	CAT CTA CGA CAA GAA ACC CTC CAA CAG ATG TCC CGC ATC CTT AAT ACG TGC CAA GCA GCT CAG GGC TTA CTT GCC	1050							
326	His Leu Arg Gln Glu Thr Leu Gln Gln Met Ser Arg Ile Leu Asn Thr Cys Gln Ala Ala Gln Gly Leu Leu Ala	350							
1051	TTA GGG GAG TAC TTT GAA CGA CTT CGT GTT TTA AGC TCA CAA TGG GCT ACT CGT CTA CGT GAG CCT ACC TAA TGA	1125							
351	Leu Gly Glu Tyr Phe Glu Arg Leu Arg Val Leu Ser Ser Gln Trp Ala Thr Arg Leu Arg Glu Pro Thr ***	373							
1126	AGCACAGAAGATCCGCTGTATTACTCGAGGAGTTTTGCCTTCAGAAGATGATGCTGTGTATGGACCAGAGTAGTGTGCTCACTTGGTATCTAAA	1224							
1225	CCTATAATCAATCAGTGGCGGAGCCACACAGAGTTCAAGGCCAGATGCAAAATTCAGGATTTCCAATGTTATTCGAATCTATTATTCTACTTTACTGGAT	1323							
1324	TTTAAACACATATATGTGATCTGAGCCAAAACTACTAGTTGGATGAAACCATAGTTATACACTAGATCGGCTCCTGCTCAAGGGTGTTCAGTTGA	1422							
1423	GCATCCTTCGTCGMAAATTAGTGTTATATATAAGTCAGATATTATGTGTGAAATCTTTGAGCACACTTAGTGTAATTTCTAGGCTTCGCCATTGAGCT	1521							
1522	ATATTCAATTCAGTTTGTGGTGAATGAATTTTACCATCTTGTCTACTTCTGGTAGGGTCTTGAGAACCTTAAATCAGATTTTACACAAATAGCC	1619							

acid residue numbers 15, 26, 41, and 56) located close to the N-terminus of the ORF. It will be shown later that the methionine codons at positions 41 and 56 can be used as translation initiation codons in *E. coli* (see chapter 4).

The deduced amino acid sequence of TGA1a contains an acidic region (amino acid residues 19-66, 12 acidic and 4 basic residues out of 48 residues), a basic region (84-108, 11 basic and 2 acidic residues out of 25 residues), a putative "leucine-zipper" region (leucine residues 114, 121, and 128; Landschulz *et al.*, 1988), and a relatively glutamine-rich region (258-346, 14 glutamine residues out of 88 residues). The acidic region and the glutamine-rich region may be involved in transactivation by this factor (see DISCUSSION).

#### Sequence analysis of TGA1b

Analysis of the insert DNA sequences of hb2 and 6 showed that they share a very high homology over a 676-bp region (~95% in DNA). Both contain a single large ORF spanning this homologous region. Because no other significantly large ORF was found, it was concluded that this ORF encodes TGA1b. The proteins encoded by the genes corresponding to hb2 and 6 were named TGA1b-2 and TGA1b-6, respectively. Since the insert cDNAs of hb2 and hb6 appear to be partial clones (insert sizes are 726 bp and 1,616 bp, respectively) compared with their mRNA size (2.9 kb; see below), the hb2 insert sequence was used as a hybridization probe to isolate longer cDNA clones. The recovered cDNA clones were, however, still not full length (2,238 bp and 2,129 bp for TGA1b-2 and -6, respectively). Figure 3.5 shows the DNA sequences and the deduced amino acid sequences for the largest ORF encoding TGA1b. The first nucleotide and the first codon of the longest ORF were numbered as position 1 for convenience. The ORF is 1,962 bp long, corresponding to 654 amino acid residues. Because of the difficulty in the sequence

Figure 3.5 DNA and deduced amino acid sequences of TGA1b.

DNA sequence of the TGA1b-2 gene is shown. Nucleotide differences found in the TGA1b-6 gene are indicated over the TGA1b-2 DNA sequence. The deduced amino acid sequence of TGA1b-2 for the longest ORF is shown in triple letter code below its DNA sequence. Amino acid changes in the deduced sequence of TGA1b-6 are indicated below the TGA1b-2 amino acid sequence. The first nucleotide and the first amino acid residue in the insert are numbered as 1 for convenience. Undetermined nucleotides and amino acids are indicated as ? and Xxx, respectively. Nucleotide sequence is spaced by -, to align the sequences for TGA1b-2 and TGA1b-6.

1	GAA	TTC	TGT	<sup>A</sup> GAT	TTT	TCC	GGA	AAT	CAA	GCA	GCT	GGA	GGC	GTT	<sup>A</sup> ATG	<sup>A</sup> GTT	ATG	GAT	<sup>A</sup> ACT	TCA	TCG	CCG	GAG	CTT	CGA	75	
1	Glu	Phe	Cys	Asn	Phe	Ser	Gly	Asn	Gln	Ala	Ala	Gly	Gly	Val	Met	Val	Met	Asp	Thr	Ser	Ser	Pro	Glu	Leu	Arg	25	
76	CAG	AGC	TCA	AGC	GGC	TCA	<sup>C</sup> GAT	GTT	TTG	AAT	GCA	ACG	TCG	TCG	<sup>A</sup> ACG	TCG	TCC	CAC	CAG	GTT	TCT	GGC	GAT	GTC	<sup>T</sup> GCC	150	
26	Gln	Ser	Ser	Ser	Gly	Ser	Asp	Val	Leu	Asn	Ala	Thr	Ser	Ser	Thr	Ser	Ser	His	Gln	Val	Ser	Gly	Asp	Val	Ala	50	
151	GGG	TAC	CTG	AAC	GTG	CCA	TCG	CCG	GAG	TCC	AAT	GGA	TCC	AAC	CAT	GAG	GGT	TCT	CGG	GAG	TCT	<sup>G</sup> GCT	AAT	GAC	AAC	225	
51	Gly	Tyr	Leu	Asn	Val	Pro	Ser	Pro	Glu	Ser	Asn	Gly	Ser	Asn	His	Glu	Gly	Ser	Arg	Glu	Ser	Ala	Asn	Asp	Asn	75	
226	AAG	GGT	TTG	GGT	GAT	<sup>A</sup> GCT	AGG	GTT	TTG	AAT	TGC	CAT	TCG	CCG	GAG	TCG	CAG	GGT	<sup>C</sup> TCA	GGC	<sup>C</sup> AAT	<sup>G</sup> TAT	GGT	TCA	AAT	300	
76	Lys	Gly	Leu	Gly	Asp	Ala	Arg	Val	Leu	Asn	Cys	His	Ser	Pro	Glu	Ser	Gln	Gly	Ser	Gly	Asn	Tyr	Gly	Ser	Asn	100	
301	GTC	TCA	GAA	GGG	CTG	AAT	<sup>A</sup> TAT	<sup>C</sup> TCG	GAT	TCG	AAC	AAA	TCG	GTA	CAT	TCT	TCT	CCT	AAT	TTT	<sup>G</sup> GAA	<sup>C</sup> AAT	AAT	TCA	375		
101	Val	Ser	Glu	Gly	Leu	Asn	Tyr	Pro	Ser	Asp	Ser	Asn	Lys	Ser	Val	His	Ser	Ser	Pro	Asn	Phe	Glu	Asn	Asn	Ser	125	
376	ATA	AAA	AAT	<sup>T</sup> GGA	GCT	GTA	GAA	GAG	AAA	ATC	AAA	TTA	GAG	GGT	GTC	AAT	GCT	AAT	ATA	AGT	<sup>C</sup> AAA	TGT	AGC	TCC	TTG	450	
126	Ile	Lys	Asn	Gly	Val	Val	Glu	Glu	Lys	Ile	Lys	Leu	Glu	Gly	Val	Asn	Ala	Asn	Ile	Ser	Asn	Cys	Ser	Ser	Leu	150	
451	TTG	AAG	AGG	AAA	AAA	AGT	AGT	GAA	GAT	TCT	AAT	AAC	ATA	<sup>GG</sup> AAC	ATA	CAC	CAA	AAA	TTG	ACT	AAT	<sup>A</sup> GTT	<sup>G</sup> GCA	TTG	<sup>T</sup> AGT	525	
151	Leu	Lys	Arg	Lys	Lys	Ser	Ser	Glu	Asp	Ser	Asn	Asn	Ile	Asn	Ile	His	Gln	Lys	Leu	Thr	Asn	Val	Ala	Leu	Ser	175	
526	GAC	AAT	GTT	AAT	AAT	GAT	GAG	GAT	GAA	AAG	AAG	AGA	GCT	AGA	TTG	GTT	AGG	AAT	AGG	GAA	AGT	GCT	CAA	CTG	TCA	600	
176	Asp	Asn	Val	Asn	Asn	Asp	Glu	Asp	Glu	Lys	Lys	Arg	Ala	Arg	Leu	Val	Arg	Asn	Arg	Glu	Ser	Ala	Gln	Leu	Ser	200	
601	AGG	CAA	AGA	<sup>A</sup> AAG	AAG	CAC	TAT	GTT	GAG	GAA	TTA	<sup>G</sup> GAA	GAT	AAA	GTT	AGA	ATA	ATG	CAT	TCA	ACA	ATT	CAA	GAT	TTG	675	
201	Arg	Gln	Arg	Lys	Lys	His	Tyr	Val	Glu	Glu	Leu	Glu	Asp	Lys	Val	Arg	Ile	Met	His	Ser	Thr	Ile	Gln	Asp	Leu	225	
676	AAT	GCT	<sup>A</sup> AAG	<sup>T</sup> GTA	GCT	TAT	ATA	ATT	GCG	GAA	AAT	GCT	ACT	CTA	AAG	ACG	CAG	TTG	GGG	GGT	GCT	GGT	GTA	CCT	TCG	750	
226	Asn	Ala	Lys	Val	Ala	Tyr	Ile	Ile	Ala	Glu	Ala	Thr	Leu	Lys	Thr	Gln	Leu	Gly	Gly	Ala	Gly	Val	Pro	Ser	250		
751	CAG	GTG	CCA	CCA	CCA	CCC	CCT	GGG	ATG	TAT	<sup>G</sup> CCA	CAT	CAT	CCT	GTG	ATG	TAT	CCG	TGG	ATG	CCG	TAT	ACA	CCA	CCT	825	
251	Gln	Val	Pro	Pro	Pro	Pro	Pro	Gly	Met	Tyr	Pro	His	His	Pro	Val	Met	Tyr	Pro	Trp	Met	Pro	Tyr	Thr	Pro	Pro	275	
826	TAT	ATG	GTG	AAA	CCG	CAA	GGA	TCA	CAA	GTG	<sup>G</sup> CCA	TTG	GTT	CCC	ATT	CCT	AAG	TTG	AAA	CCA	CAG	GGT	<sup>A</sup> GCA	GCG	CCA	900	
276	Tyr	Met	Val	Lys	Pro	Gln	Gly	Ser	Gln	Val	Pro	Leu	Val	Pro	Ile	Pro	Lys	Leu	Lys	Pro	Gln	Gly	Ala	Ala	Pro	300	
901	GCG	CCA	AAG	AGT	AGC	AAG	AAA	GTG	GAG	AAA	<sup>T</sup> AAG	AGG	ACT	<sup>G</sup> CAG	GTG	AAA	ACT	AAG	AAG	GTT	GCA	AGT	GTT	AGT	TTC	975	
301	Ala	Pro	Lys	Ser	Ser	Lys	Lys	Val	Glu	Lys	Lys	Arg	Thr	Gln	Val	Lys	Thr	Lys	Lys	Val	Ala	Ser	Val	Ser	Phe	325	
976	CTT	GGC	TTG	TTG	TTC	TTC	ATG	CTG	CTA	TTT	GGT	GGG	TTG	GTT	CCT	TTA	TTA	AAT	GTG	AGA	TAT	GGA	GGA	ACG	AGG	1050	
326	Leu	Gly	Leu	Leu	Phe	Phe	Met	Leu	Leu	Phe	Gly	Gly	Leu	Val	Pro	Leu	Leu	Asn	Val	Arg	Tyr	Gly	Gly	Thr	Arg	350	
1051	GAA	CCA	TTC	ATG	AGT	GGA	GAT	TCT	TTT	<sup>G</sup> TGG	AGT	GGG	TCT	TAT	GAG	AAA	CAT	CAG	GGA	AGA	GTC	TTG	GCT	GTT	GAC	1125	
351	Glu	Pro	Phe	Met	Ser	Gly	Asp	Ser	Phe	Trp	Ser	Gly	Ser	Tyr	Glu	Lys	His	Gln	Gly	Arg	Val	Leu	Ala	Val	Asp	375	
1126	<sup>G</sup> GGA	CCT	GTG	AAT	GGG	ACT	GGT	<sup>T</sup> CAC	<sup>C</sup> TCT	GGA	<sup>A</sup> AAG	TAC	<sup>A</sup> GGT	GGA	<sup>G</sup> AAT	GAT	TTT	AGC	TCA	CAT	TGT	<sup>G</sup> AGT	<sup>C</sup> CGG	AGA	GGT	1200	
376	Gly	Pro	Val	Asn	Gly	Thr	Gly	His	Ser	Gly	Lys	Tyr	Gly	Gly	Asn	Asp	Phe	Ser	Ser	His	Cys	Ser	Arg	Arg	Gly	400	
1201	CAG	GAT	GAG	AGC	GCT	CAG	CCA	<sup>G</sup> AAT	GTC	AAT	GGC	AGT	GAC	CCT	CTA	GCT	GCA	TCC	TTG	TAT	GTC	CCA	AGG	AAT	GAT	1275	
401	Gln	Asp	Glu	Ser	Ala	Gln	Pro	Asn	Val	Asn	Gly	Ser	Asp	Pro	Leu	Ala	Ala	Ser	Leu	Tyr	Val	Pro	Arg	Asn	Asp	425	
1276	AAA	CTT	GTG	AAG	ATT	GAT	GGA	AAC	TTG	ATA	ATT	CAT	TCT	GTA	TTG	GCA	AGT	GAA	AAA	GCC	ATG	GCA	TCT	CAT	GGA	1350	
426	Lys	Leu	Val	Lys	Ile	Asp	Gly	Asn	Leu	Ile	Ile	His	Ser	Val	Leu	Ala	Ser	Glu	Lys	Ala	Met	Ala	Ser	His	Gly	450	
1351	GGT	GCT	GAT	AAG	AAC	AAT	AGA	GAG	ACA	GGC	CTC	GCA	GTT	CCT	GAG	GAT	TTA	GCC	CCT	GCT	ATC	GCA	GGA	AGC	CAT	1425	
451	Gly	Ala	Asp	Lys	Asn	Asn	Arg	Glu	Thr	Gly	Leu	Ala	Val	Leu	Pro	Glu	Asp	Leu	Ala	Pro	Ala	Ile	Ala	Gly	Ser	His	475
1426	CCT	CAC	CTT	<sup>C</sup> TAT	CGA	AGT	ACG	GCA	GTG	GGA	CAA	AGG	<sup>C</sup> GTT	<sup>C</sup> CTT	GGA	ACT	GGA	GAG	GAG	AAT	GGA	<sup>G</sup> AAG	TCA	ACT	GTG	1500	
476	Pro	His	Leu	Xxx	Arg	Ser	Thr	Ala	Val	Gly	Gln	Arg	Val	Leu	Gly	Thr	Gly	Glu	Glu	Glu	Asn	Gly	Lys	Ser	Thr	Val	500
1501	CAA	CAG	TGG	TTC	CTT	GAA	GGT	GTT	GCT	GGG	CCT	TTG	TTG	AGT	TCA	GGC	ATG	TGT	ACA	GAA	GTG	TTC	<sup>G</sup> CA?	<sup>C</sup> TTC	GAT	1575	
501	Gln	Gln	Trp	Phe	Leu	Glu	Gly	Val	Ala	Gly	Pro	Leu	Leu	Ser	Ser	Gly	Met	Cys	Thr	Glu	Val	Phe	Xxx	Phe	Asp	525	

1576	T	G7G	TCA	TCT	TCT	GCT	CCA	GGA	GCC	GTA	GTT	CCT	GC?	T	ACC	AAT	G	ACG	AGG	AAT	TTA	T	TC?	ATG	GAA	GAA	AGG	CAG	AAT	1650
526		Xxx	Ser	Ser	Ser	Ala	Pro	Gly	Ala	Val	Val	Pro	Xxx	Ala	Thr	Asn	Thr	Arg	Asn	Leu	Xxx	Met	Glu	Glu	Arg	Gln	Asn	550		
		Val															Ala				Ser									
1651		GCT	ACA	A	CGC	GTC	CAC	AAG	GTT	AGA	AAT	AGA	AGG	ATC	CTC	AAT	GGT	C	CAT	TCC	GTT	T	ACC	CTT	TCT	AGA	CCC	TCC	CAC	1725
551		Ala	Thr	Arg	Val	His	Lys	Val	Arg	Asn	Arg	Arg	Ile	Leu	Asn	Gly	Pro	His	Ser	Val	Thr	Leu	Ser	Arg	Pro	Ser	His	575		
				His																										
1726		AAC	ATT	TCT	GAA	GAA	CAA	ACT	GGA	AAG	CAA	GAG	AAC	A	TTC	AGT	GGA	AAT	AAG	TCA	CTT	TCT	TCC	ATG	GTT	GTA	TCT	1800		
576		Asn	Ile	Ser	Glu	Glu	Gln	Thr	Gly	Lys	Gln	Glu	Asn	Leu	Phe	Ser	Gly	Asn	Lys	Ser	Leu	Ser	Ser	Met	Val	Val	Ser	600		
1801		GTG	CTT	GTT	GAT	CCA	AGA	GAG	GGA	GGT	GAT	GGT	GAT	GTT	GAT	GGC	ATG	ATG	GGT	CCA	AAG	TCC	CTC	TCT	CGG	ATA	1875			
601		Val	Leu	Val	Asp	Pro	Arg	Glu	Gly	Gly	Asp	Gly	Asp	Val	Asp	Gly	Met	Met	Gly	Pro	Lys	Ser	Leu	Ser	Arg	Ile	625			
1876		TTC	GTC	GTT	GTG	CTG	ATT	GAC	AGT	GTC	AAG	TAT	GTC	ACC	TAT	TCT	TGT	ATG	CTT	CCA	TTT	AAA	GGA	TCT	GTT	CCT	1950			
626		Phe	Val	Val	Val	Leu	Ile	Asp	Ser	Val	Lys	Tyr	Val	Thr	Tyr	Ser	Cys	Met	Leu	Pro	Phe	Lys	Gly	Ser	Val	Pro	650			
1951		TTA	GTG	ACT	ACT	TGA	AGAT	GGAGCATT	CAT	GTAGCAGAGGCAT	CAT	GGACAACCAATTTT	GTAGATAGCATAAGAT	CAGATGGTCCTC	CACAGT	2043														
651		Leu	Val	Thr	Thr	***									654															
2044		TCAC	TTAGCTT	CAATA	GTA	CATTAACTTTTAAACAGCTAT	CATCATTATATAGGGCGACGGTCTTCTTCTTA	-----TGTGAT?TATCTGCA	2129																					
2133		ATA	ATAGTT	ATCGCTCTTCAATTATCAAAGAAGAAGGCAACGCTCCTTTAAGTTCTCTGTA	ACTGTGGATGTATGCTTTTCCCATCT	(A)x16	2238																							

analysis, certain parts of the DNA sequence are ambiguous. The uncertain nucleotides and corresponding amino acid residues are indicated as ? and Xxx in Fig. 3.5, respectively. TGA1b-2 and -6 share a high degree of homology (94% in the amino acid sequence level) over their coding regions. The whole insert sequence of hb2 corresponds to nucleotides 1-726 of TGA1b-2. The insert sequence of hb6 corresponds to nucleotide 1-676 of TGA1b-6. The remaining hb6 insert sequences are unrelated to TGA1b and appear to have resulted from an artifact during the library construction. Two independent cDNA clones isolated by hybridization screening have the same sequence in the region downstream of nucleotide 676, as deduced from restriction enzyme mapping; the hb6 insert, however, has a completely different sequence in this region. The ORFs of hb2 and hb6 are in frame with the reading frame of *lacZ*, suggesting that TGA1b is produced as a *lacZ*-fusion protein in the lysogen. This result is consistent with the observation that the DNA-protein complexes formed by the lysogen-produced TGA1b (hb2) and *hex-1* have a very slow migration rate (not shown).

Although full-length clones encoding TGA1b have not yet been isolated, the deduced amino acid sequence already shows some interesting characteristics. It contains a serine-rich region (amino acid residues 20-160, 33 serine residues out of 141 residues), a basic region (185-206, 11 basic and 1 acidic residues out of 22 residues), a putative "leucine-zipper" region (leucine, methionine, or isoleucine residues at positions 211, 218, 225, 232, and 239), a stretch of five consecutive proline residues (253-257), and another basic region (303-319, 9 basic and 1 acidic residues out of 17 residues). There is no significant homology between TGA1a and TGA1b other than the basic region and the "leucine-zipper" region (bZIP domain, see DISCUSSION).

#### Expression of TGA1a and TGA1b genes

Northern blot analysis of tobacco RNA with the TGA1a cDNA insert showed that the mRNA level (1.8 kb) is approximately 10 times higher in roots than in light-grown or dark-adapted leaves (Fig. 3.6). The insert of hb5 appears near full-length, because the size of the insert (1.6 kb) is close to the size of the mRNA (1.8 kb) although the insert lacks the region encoding 3'-end of the mRNA including the polyA. The level of the TGA1b mRNA (2.9 kb) is 2-fold higher in roots than in dark-adapted leaves and the latter contain three times as much TGA1b mRNA as leaves from light-grown plants (Figure 3.6). As a control, the expression of the gene encoding the  $\beta$ -subunit of mitochondrial ATP synthase (b-ATPase) was found to be about two-fold higher in roots, confirming previous results (Figure 3.6; Boutry and Chua, 1985). The much higher expression of TGA1a in roots than in leaves is correlated with the observation that *as-1* is involved in root expression of the 35S promoter and is consistent with the proposal that the functional form of ASF-1 could be limiting in leaves (Lam *et al.*, 1989).

#### Southern analyses of TGA1a and TGA1b genes

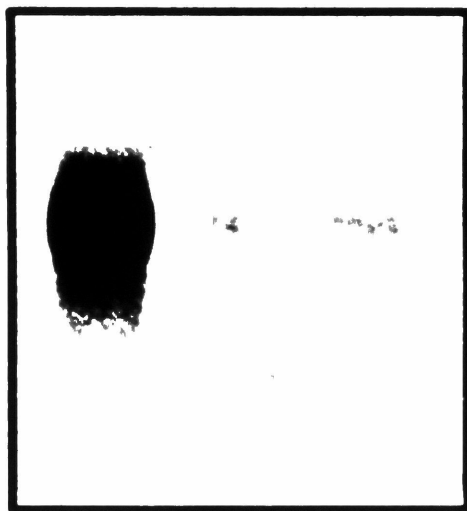
Southern blot analyses of the TGA1a and TGA1b genes showed a few hybridizing bands for each under relatively stringent conditions of hybridization (Figure 5). Because there are no common hybridizing bands, I conclude that under these conditions the two genes do not cross-hybridize. This result suggests that each of them is encoded by a small gene family. Because Northern blot hybridization gives only a single band for each gene (Figure 4), both gene families probably contain very closely related members. In fact, TGA1b has at least two very closely related family members (TGA1b-2 and -6). The isolation of genomic clones and cDNA clones related to TGA1a revealed that it has at least four family members (H.Fromm, F.Katagiri, and N.-H.Chua, in preparation). Whether

Figure 3.6 Northern blot analysis of TGA1a and TGA1b genes.

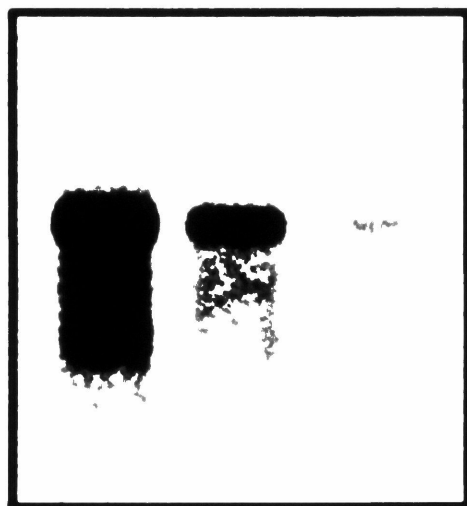
Each lane contained 2  $\mu$ g of polyA<sup>+</sup> RNA. mRNA sizes were determined by comparison to molecular size markers. LEAF(D), leaves from tobacco plants kept in the dark for two days; LEAF(L), leaves from light-grown tobacco plants; ROOT, roots from light-grown tobacco plants.



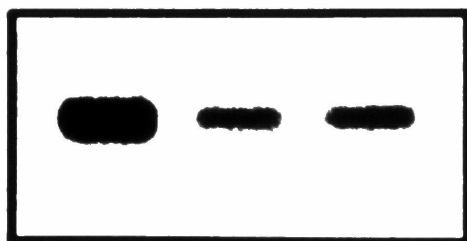
ROOT  
LEAF (D)  
LEAF (L)



TGA1a  
(1.8 kb)



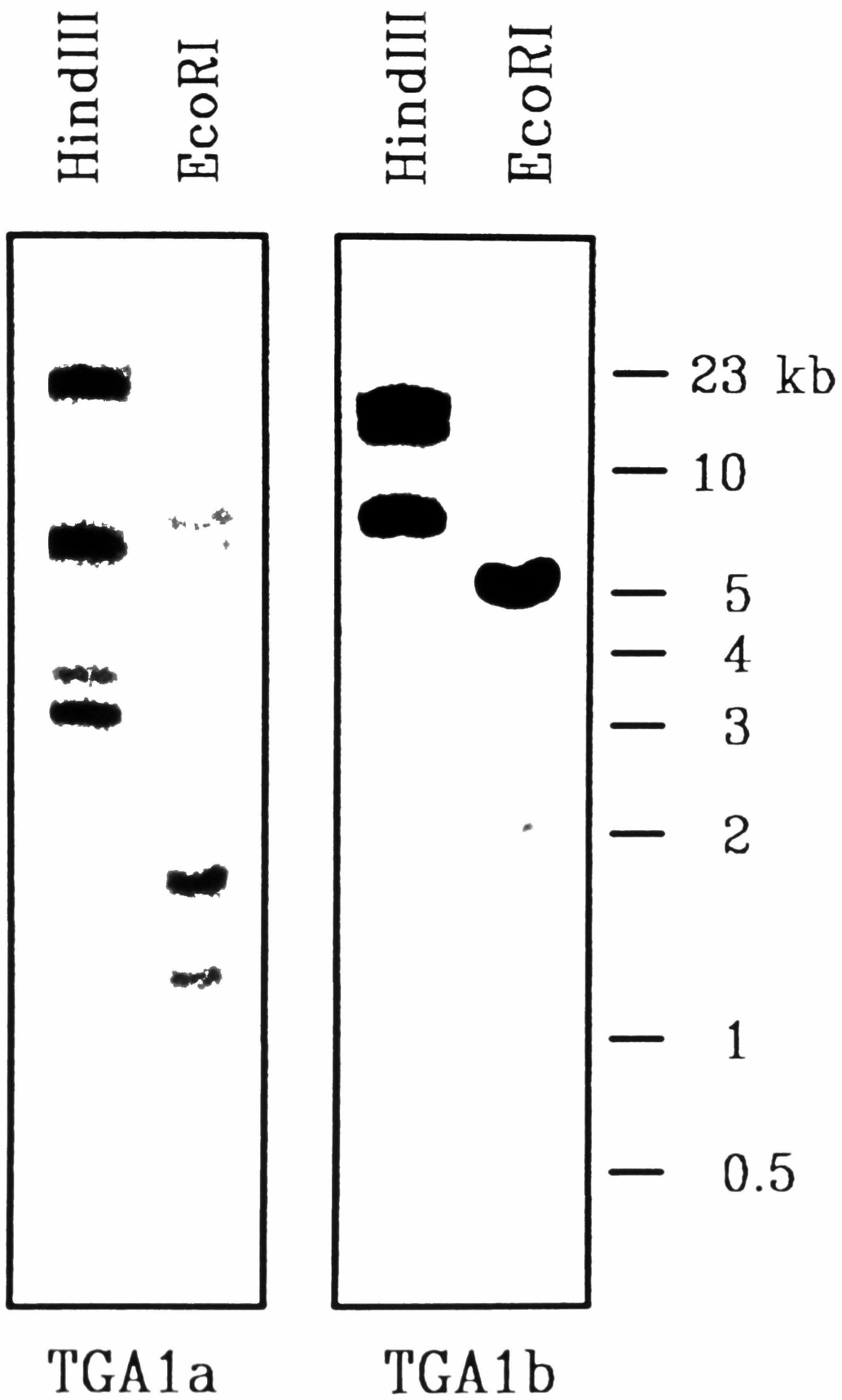
TGA1b  
(2.9 kb)



b-ATPase  
(2.1 kb)

Figure 3.7 Southern blot analysis of TGA1a and TGA1b genes.

Each lane contains 17  $\mu$ g of tobacco DNA digested with either *Hind*III or *Eco*RI. Positions of molecular markers are indicated on the right.



there is any difference in the function and expression pattern among the family members remains to be determined.

## DISCUSSION

In this chapter, I have shown that nuclear extracts from tobacco leaf contain two factors, called ASF-1 and HSBF, that recognize TGACG-containing sequences. Reciprocal competition experiments in gel retardation assays demonstrated that ASF-1 binds to *as-1*, *hex-1*, and *nos-1*, whereas HSBF only binds to *hex-1*. Using *hex-1* as a binding site probe, five cDNA clones that encode two different class of DNA-binding proteins, called TGA1a and TGA1b, were isolated. TGA1a binds to all the three elements, whereas TGA1b binds only to *hex-1*. Based on their differential binding specificities, TGA1a and TGA1b, are considered to be good candidates for ASF-1 and HSBF, respectively.

### Homology with bZIP proteins

The basic regions of TGA1a and TGA1b are homologous to the bZIP DNA-binding domains of other eukaryotic DNA-binding proteins (Landschulz *et al.*, 1988), such as CREB, GCN4, and c-Jun (Figure 3.8; Hoeffler *et al.*, 1988; Gonzalez *et al.*, 1989; Hope and Struhl, 1986; Bohmann *et al.*, 1987). The basic region of the bZIP domain has been demonstrated to determine the DNA-binding specificity: reciprocal exchange of basic regions of two bZIP proteins, GCN4 and C/EBP, shows that the DNA binding specificity is determined by the basic region (Agre *et al.*, 1989). Therefore, it is not surprising that the basic regions of TGA1a and TGA1b are homologous to those of CREB, GCN4, and c-Jun because the target site of CREB contains a TGACG motif (same as *hex-1* and *as-1*) and the target sites of GCN4 and c-Jun contain a similar motif TGACT. I note that among the three proteins,

Figure 3.8 Comparison of amino acid sequences of TGA1a and TGA1b with those of c-Jun, GCN4, and CREB.

The positions of the homologous regions are indicated in parentheses as amino acid residue numbers. The target binding site sequences of the five DNA binding proteins are also indicated. Conserved residues between TGA1a, TGA1b, and the other three factors are boxed in thin lines, whereas the residues conserved only among TGA1a, TGA1b, and CREB are boxed in thick lines. Asterisks indicate the positions of the repeated leucine residues.

BINDING SITE SEQ

ATGAC TCT  
TGAC TCA  
TGACGTCA  
  
TGACG  
TGACG

Jun	(259-317)	QERIKAEEKRMNRHIAASKQPRMRRLERIALEEKVKTKLKAQNSELEASTANMLREQVAQL*	*
GCN4	(223-281)	ESSDPAALKPARNTEAARFSRRKLQRMKQLEDKVEELISKNYHLENEVARLKKLIVGER	*
CREB	(266-324)	EAAPRREVPIAMKNREAAPEQRRKKKEVVKQLENRMVAVLENQNKTLIEELKALKDLYCHK	*
TGA1a	( 84-142)	KPVEKVLFRILAQNREAAARKSRIRKKAMVQQLLEN SKLKLQLECELEPRARKQGMVCVGGV	*
TGA1b	(181-239)	DEDEKRRARILVRNRESAQLSRQRRKHVVEELEDKVRIMHSTIQDLNAKVAYIIAENATL	*

CREB has the highest homology to TGA1a and TGA1b in the basic region. It seems reasonable to predict that the amino acid residues that are conserved only among CREB, TGA1a, and TGA1b (blocked by thick lines in Figure 6) are important in distinguishing the TGACG motif from the TGACT motif. This prediction could be tested by site-directed mutagenesis. The bZIP DNA-binding domain contains another characteristic region, the "leucine-zipper" region, in which leucine residues occur at every seventh amino acid residues (Landschulz *et al.*, 1988). The "leucine-zipper" region is located at the carboxy end of the basic region in the bZIP domain and the distance between the two is very well conserved (Vinson *et al.*, 1989). When the amino acid sequences of these proteins are aligned according to their homology in the basic region, the first three of the four leucine residues (marked by asterisks in Figure 6) in the "zipper" of CREB, GCN4, and c-Jun are also conserved in TGA1a. In TGA1b, the residues in the corresponding positions are leucine, methionine, leucine, isoleucine, and leucine. Because a leucine residue in the "leucine-zipper" of C/EBP can be substituted by either a methionine or an isoleucine residue (although with some reduction in dimerization potential) (Landschulz *et al.*, 1989), we consider this region of TGA1b as a variant of the "leucine-zipper." The "leucine-zipper" region is known to mediate dimerization of proteins (Kouzarides and Ziff, 1988; Landschulz *et al.*, 1989; Turner and Tjian, 1989; Gentz *et al.*, 1989). Studies on the dimerization of TGA1a will be described in chapter 6.

Mikami *et al.* (1987, 1989a) has previously reported that the *hex-1* element binds a DNA-binding protein from wheat germ extract, designated HBP-1. Recently, they reported that HBP-1 in fact contained two different binding proteins, designated HBP-1a and HBP-1b (Mikami *et al.*, 1989b). Among the three TGACG-containing cis-elements, HBP-1a binds only to *hex-1* while HBP-1b binds to *hex-1*, *as-1*, and *nos-1*. Therefore, it is likely that ASF-1

and HSBF are tobacco homologs of HBP-1b and HBP-1a, respectively. Their recent report on the nucleotide sequence of HBP-1 (Tabata *et al.*, 1989; redesignated as HBP-1a), however, show that the homologous regions between HBP-1 and TGA1a or TGA1b are limited to the bZIP domains. Other plant bZIP proteins reported recently (Guiltinan *et al.*, 1990; Hartings *et al.*, 1989; K.Ooeda, J.Salinas, and N.-H.Chua, submitted; Schmidt *et al.*, 1990; Singh *et al.*, 1990) are also homologous to TGA1a or TGA1b only in this domain.

Since plants and animals diverged about 1.5 billion years ago (Dayhoff, 1969), the high homology in the DNA-binding domains of bZIP DNA-binding proteins, including TGA1a, TGA1b, c-Jun, GCN4, and CREB, suggests an ancient origin for the DNA-binding domain and its cognate binding site, the TGACG-like motif. Sequence analysis of TGA1a genomic clones revealed that its DNA-binding domain is encoded by separate exons (H.Fromm, F.Katagiri, and N.-H.Chua, in preparation). This genomic structure raises the speculation that ancient exons for the DNA-binding domain was shuffled with other exons to generate different DNA-binding proteins. Along this line, in chapter 6, I discuss the advantage of having another functional domain of TGA1a, the dimer stabilization domain, encoded by exons separated from those encoding the bZIP domain.

#### Other interesting amino acid sequences in TGA1a and TGA1b

In addition to the bZIP domain, TGA1a and TGA1b have other interesting features in their amino acid sequences. The acidic region of TGA1a (amino acid residues 19-66) may be involved in transcription activation, since this function is carried out by acidic domains of several eukaryotic DNA-binding proteins (Ptashne, 1988). The mechanism of transactivation mediated by acidic domains is considered to be conserved among various eukaryotes, such as yeast, plants, insects, and mammals (Ptashne, 1988). The glutamine-rich



region of TGA1a (amino acid residues 258-346) may also be involved in transcription activation, as is the case for Sp1 (Courey and Tjian, 1988), although the proportion of glutamine residues is not as high as that in the glutamine-rich regions of Sp1. TGA1b has a serine-rich region (amino acid residues 20-160) that contains several potential phosphorylation sites. Because phosphorylation is frequently used to regulate DNA-binding and/or transactivation activities of transcription factors (Mitchell and Tjian, 1989), it will be of interest to see whether TGA1b is phosphorylated. In this region of TGA1b, not only serine residues but also glutamic acid, proline, and threonine residues are also relatively abundant. Regions rich in these amino acid residues (called PEST) are suggested to be involved in protein turnover (Rogers *et al.*, 1986). The turn-over rate of TGA1b might be modulated by these amino acid residues.

#### Promiscuous combinations of cis-elements and DNA-binding proteins

As more information about DNA-binding proteins accumulates, it seems quite common for an organism to have several DNA-binding proteins that can bind to a single binding site. Therefore it is not unusual that both ASF-1 and HSBF can bind to *hex-1*. On the other hand, there are also examples in which a single DNA-binding protein can bind to several cis-elements of different functions. ASF-1 binds to *hex-1*, *as-1* and *nos-1*. In addition, ASF-1 binds to *ocs*, which contains two copies of a TTACG-motif and is located at -193 to -173 region of the octopine synthase promoter (Bouchez *et al.*, 1989; Ellis *et al.*, 1987; Fromm *et al.*, 1989; Singh *et al.*, 1989). In terms of their *in vivo* function, *as-1* (Lam *et al.*, 1989), *nos-1* (Lam *et al.*, 1990b), and *ocs* (Fromm *et al.*, 1989) appear to be similar; all these elements confer preferential expression in roots of transgenic tobacco. By contrast, *hex-1* alone is inactive in transgenic tobacco plants, although it can interact with other cis-

elements, such as *as-1* to give activity (E.Lam, unpublished). The interaction of a DNA-binding protein with a specific cis-element of interest depends on several conditions, such as differential binding affinity of the protein to the element and related sequences, concentrations of the proteins in the nucleus, and interaction of the DNA-binding proteins with other factors. Therefore, it is not easy to address the question as to which DNA-binding protein is really responsible for the *in vivo* function of a particular cis-element, unless genetic data are available (eg. a mutation in the DNA-binding protein gene affects the *in vivo* expression conferred by its cognate binding site.). In the absence of such genetic evidence, it is important to accumulate as many lines of evidence as possible from biochemical and molecular analysis, in addition to DNA-binding specificity. In this regard, *as-1* appears to be a relatively simple system for investigation. This element is comprised of only 21 bp. The *in vivo* expression pattern conferred by this element is distinctive and specific mutations that abolish its function are well characterized. Moreover, interaction of its cognate factor with other factors does not appear to be crucial for its function because *as-1* is a portable cis-element. In the following chapters, I will describe several lines of evidence supporting the notion that TGA1a is indeed responsible for the *in vivo* function of *as-1*

#### **Chapter 4: TGA1a can stimulate transcription in a human in vitro system.**

It has been suggested that unrelated eukaryotes may share a common mechanism for transcription activation (Ptashne, 1988). For example, a yeast transcription activator, GAL4, has been shown to function in mammalian (Kakidani and Ptashne, 1988; Webster *et al.*, 1988), insect (Fischer *et al.*, 1988) and plant cells (Ma *et al.*, 1988). Examples of such transcriptional compatibility among unrelated eukaryotes are, however, still limited; in particular, there has been no report concerning the activity of a plant transcription factor in other organisms.

TGA1a is a tobacco DNA-binding protein encoded by cDNA clones obtained from screening a cDNA expression library with a probe containing the TGACG motif (chapter 3; Katagiri *et al.*, 1989). The high level of TGA1a mRNA in roots is consistent with the notion that this DNA-binding protein is involved in the *in vivo* function of *as-1* that confers preferential expression in roots (Katagiri *et al.*, 1989). Because *as-1* is a positive cis-regulatory element, TGA1a is expected to be a transcription activator.

Compared with *in vivo* assay systems, the analysis of a transcription activator using a well characterized *in vitro* system allows the investigation of the activation mechanisms at the molecular level (Abmayr *et al.*, 1988; Hai *et al.*, 1988; Horikoshi *et al.*, 1988a,b; Klein-Hitpass *et al.*, 1990; Sawadogo and Roeder, 1985a; Workman *et al.*, 1988, 1990). Among the various RNA polymerase II *in vitro* transcription systems, the one based on HeLa cell nuclear extracts is the best characterized (Dignam *et al.*, 1983a,b). In this system, general transcription factors, TFIIB, TFIID and TFIIE/F, are necessary for transcription initiation from the adenovirus major late promoter in addition to RNA polymerase II (for review see Saltzman and Weinmann, 1989).

In this chapter, I describe the function of the plant DNA-binding protein, TGA1a, in a HeLa cell *in vitro* transcription system. The results show that TGA1a can function as a transcription activator in the human *in vitro* transcription system and that it facilitates the formation of active preinitiation complexes.

## RESULTS

### Overproduction of TGA1a in *E. coli*

TGA1a was overproduced in *E. coli* by the T7 expression system (Rosenberg *et al.*, 1987). The coding region of TGA1a was cloned downstream of the T7 promoter in the expression vector pET3a (Rosenberg *et al.*, 1987) to obtain pKT7T1A. The plasmid pKT7T1A was designed to express the full-length TGA1a protein (373 amino acid residues) by utilizing a Shine-Dalgarno sequence in pET3a (see chapter 2 for plasmid construction). The expression of TGA1a can be induced by the addition of IPTG in the medium. Figure 4.1a shows the total protein profile of the bacterial cells before and after IPTG induction (lanes 1 and 2) analyzed by SDS-polyacrylamide electrophoresis (SDS-PAGE). After the induction, the amount of presumptive TGA1a (band I) constitutes about 3% of the total protein. The induced cells were used as the starting material for the purification of TGA1a.

### Purification of TGA1a

TGA1a was purified as described in chapter 2. Table 4.1 shows each step of the purification procedure. From a 2-liter bacterial culture 1.7 mg of TGA1a was obtained at more than 95% purity as assessed by SDS-PAGE (Fig. 4.1b).

Figure 4.1 Overproduction and purification of TGA1a.

(a) Total protein profiles of *E. coli* expressing TGA1a. Total bacteria proteins (40  $\mu$ g/lane) were separated by 10% SDS-PAGE and the gel was stained with coomassie brilliant blue. Lane 1, before induction; lane 2, after induction. In lane 2, the arrowhead indicates the polypeptide that increased in concentration after induction. (b) Purified TGA1a. 10  $\mu$ g of the purified TGA1a was analyzed by 10% SDS-PAGE and the gel was stained with coomassie brilliant blue. The three bands are marked as band I, II, and III in the order of decreasing molecular weight. (c) Southwestern blot analysis of purified TGA1a. The purified TGA1a fraction (10  $\mu$ g/lane) was separated by 10% SDS-PAGE and the proteins were blotted onto a nitrocellulose membrane which was probed with a labeled *hex-1* sequence. Lane 1, no competitor; lane 2, HW was added as a competitor; lane 3, HM was added as a competitor. The positions of the polypeptide bands in (a), (b) and (c) are directly comparable. Migration positions of molecular weight markers are indicated on the right. (d) Amino-terminal sequences of band I, II and III. Amino-terminal sequences of band I, II and III (analyzed) are compared to the corresponding amino acid sequences deduced from the DNA sequence of TGA1a cDNA clone (deduced) (Katagiri *et al.*, 1989). Comparison of the sequence of band I and the deduced sequence suggests that this band likely contains a mixture of the full length TGA1a and the same polypeptide but without the first methionine. The sequence of band II suggests that it contains a truncated TGA1a polypeptide with its N-terminus at position 42. The sequence of band III suggests that it contains a mixture of two truncated TGA1a polypeptides with N-termini at positions 56 and 57. The schematic diagram at the bottom shows the relative positions of the translation start sites of the three bands in the TGA1a protein. The acidic region, the basic region, the leucine (L)-zipper region and the glutamine (Q)-rich region are shown. The amino acid sequences are shown in single-letter code. dS stands for dehydroserine. Amino acids detected in the same sequencing cycles are aligned in perpendicular. However, the first amino acids of the analyzed sequences are not trustworthy, because the first sequencing cycle is often contaminated by irrelevant amino acids. The numbers indicated over the methionine codons of the deduced sequences represent the positions of the methionine codons with respect to the first methionine.

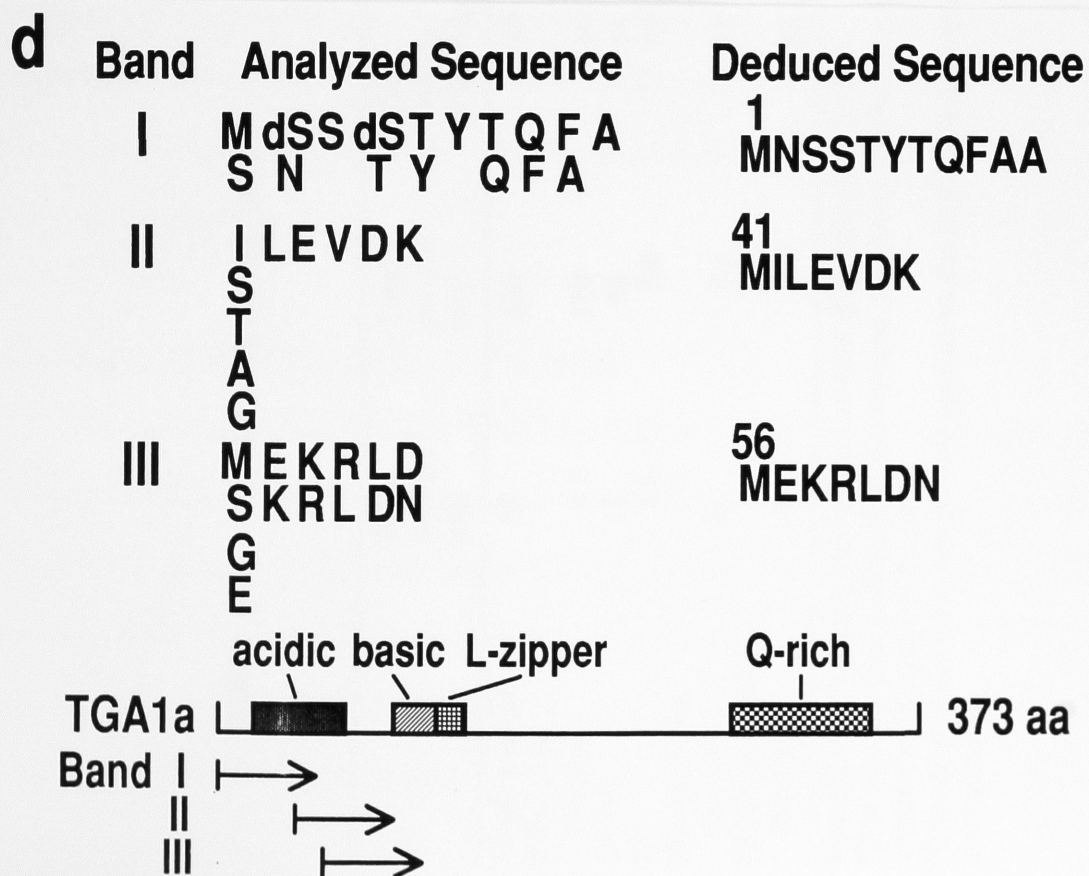
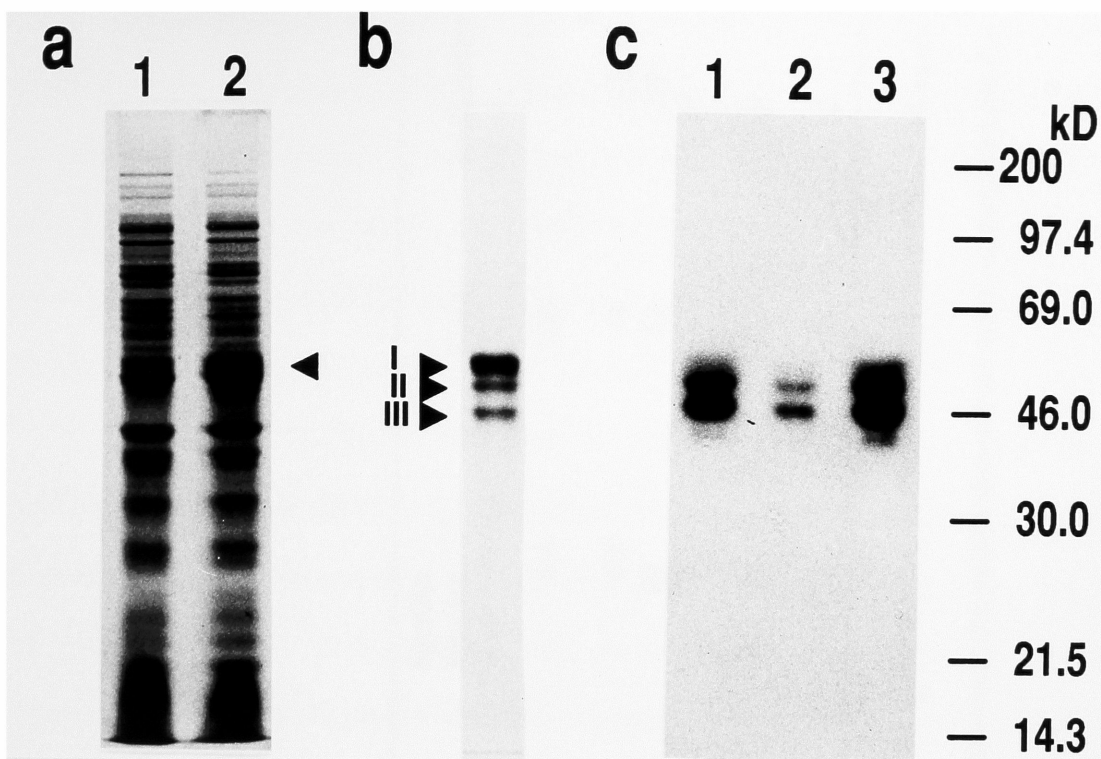


Table 4.1 Purification of TGA1a

	Total volume (ml)	Total protein (mg)	Total activity $\times 10^{-6}$ (unit) <sup>a</sup>	Specific activity $\times 10^{-5}$ (U/mg) <sup>a</sup>	Yield (%)	Purification (-fold)
Crude extract	92	330	180	5.4	100	1
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> fraction	4.7	120	80	6.5	44	1.2
DE-52	13	48	94	20	52	3.7
P11	4.2	8.0	59	73	33	14
Mutant DNA affinity <sup>b</sup>	21	3.3	27	82	15	15
Wild-type DNA affinity <sup>b</sup>	19	1.7	30	170	17	32

Starting from 6 grams (wet weight) of BL21(DE3)/plysS/pKT7T1A.

<sup>a</sup>One activity unit is defined as the amount of activity that binds to 3% of the *hex-1* binding probe (1.4 fmoles) under standard assay conditions (Katagiri et al. 1989).

<sup>b</sup>One-third of the total sample was applied to the DNA affinity columns. However, the values shown here were recalculated for the total sample.

The purified TGA1a fraction contains three molecular species when analyzed by SDS-PAGE (Fig. 4.1b). Amino-terminal sequence analyses of these polypeptides suggested that the three bands correspond to the products of the TGA1a gene with three different amino termini in the same reading frame (Fig. 4.1d). Bands I, II and III correspond to the products that start at Met1, Met41 and Met56, respectively. It is likely that Met41 and Met56 are used as cryptic translation start sites in *E. coli*.

The specific binding abilities of the three polypeptides were examined by Southwestern blot (Fig. 4.1c). All three polypeptides bind to the *hex-1* probe, an element of the wheat histone H3 promoter (Mikami *et al.*, 1987) that contains a specific binding site for ASF-1 and TGA1a (Katagiri *et al.*, 1989). It is not clear why band I gives only a weak signal even though it is the most abundant species (see DISCUSSION). Binding of all three polypeptides is sequence-specific; the binding is competed out by the addition of an oligonucleotide containing the wild-type *hex-1* (HW) but not by an oligonucleotide containing the mutant *hex-1* (HM). HW binds TGA1a but HM does not (see chapter 2). This mixture of three different polypeptides was used as the purified fraction of TGA1a in subsequent experiments.

#### Transcription activation in a human *in vitro* system

The function of the purified TGA1a was assayed in a human *in vitro* transcription system reconstituted with partially purified fractions of TFIIB, TFIID, TFIIE/F and RNA polymerase II from HeLa cell nuclear extracts (Hai *et al.*, 1988; Horikoshi *et al.*, 1988b). Hereafter, these components are referred to as the HeLa general factors. The DNA templates contain two copies of either the wild type or the mutant *as-1* sequence placed upstream of the 35S TATA box region (designated the wild type and the mutant promoters,



respectively; Fig. 4.2a). The mutant derivative of *as-1* has a very low affinity for TGA1a (Katagiri *et al.*, 1989; Yamazaki *et al.*, 1990). To facilitate detection of specific transcripts, a 380-bp G-free sequence (Sawadogo and Roeder, 1985b) placed 3' to the TATA box region was used as the template for RNA polymerase II.

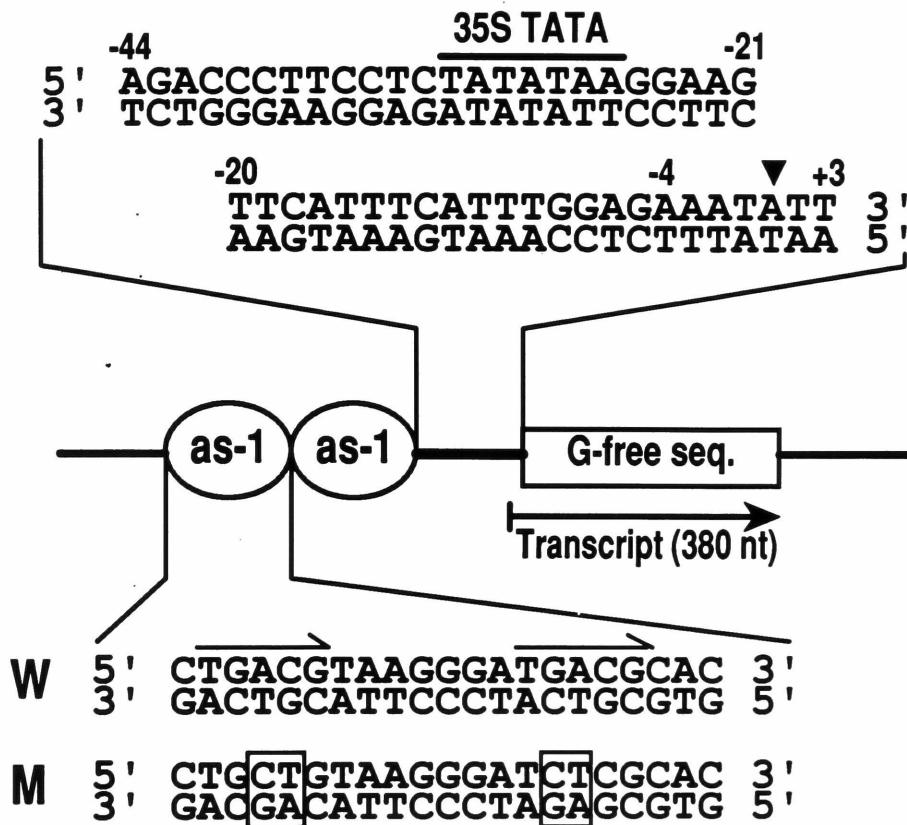
Figure 4.2b shows that with the wild type promoter, addition of 50 ng per assay of TGA1a increases transcription by about 10 times (lanes 1 and 2). Addition of the same amount of TGA1a gives only a small stimulation with the mutant promoter (lanes 3 and 4). This slight stimulation could result from the greatly reduced binding affinity of TGA1a for the mutant *as-1*. Thus, the results clearly demonstrate that TGA1a functions as a sequence-specific transcription activator in the HeLa cell *in vitro* transcription system.

A difference in the basal transcription level is consistently detected between the wild type and mutant promoters (Fig. 4.2b, lanes 2 and 4). The reproducibility of this difference suggests that a TGA1a-like activity is present in the HeLa general factors. In fact, the TFIIB and TFIID fractions used show specific binding activities to the *hex-1* element when analyzed by gel retardation assays (data not shown). To deplete the binding activities, the HeLa general factors were titrated with the oligonucleotide HW. The addition of 10 ng per assay of HW was sufficient to reduce the apparent high basal transcription level of the wild type promoter (data not shown). Figure 4.2c shows the results obtained with the 'depleted' HeLa general factors. With this mixture the basal transcription level of the wild type promoter is reduced to the same level as that of the mutant promoter (Fig. 4.2c, lanes 2 and 4), whereas the transcription stimulation by TGA1a (50 ng/assay) remains at a similar level (compare lanes 1 in Fig. 4.2b and 4.2c). In contrast to the results obtained with the 'untreated' HeLa general factors, no transcription activation by TGA1a is seen with the mutant promoter using the 'depleted' HeLa general factors (compare lanes 3 and 4 in Fig.

Figure 4.2 TGA1a activates transcription in a HeLa cell *in vitro* system.

(a) Structures of DNA templates used for *in vitro* transcription. Two copies of either the wild type (W) or mutant (M) form of *as-1* were placed upstream of the 35S TATA region (-44 to -4). The sequences of the wild type (W) and mutant (M) forms of *as-1* are shown. A tandem repeat of the TGACG motif is featured by arrows in W and the mutations in M are boxed. The sequence of the 35S TATA box region is also shown. The numbers indicated over the sequence represent the nucleotide positions with respect to the determined transcription initiation site (arrowhead; see Fig. 4.4). The 35S TATA box is overlined. G-free sequence placed downstream of the TATA region generates a 380-nt specific transcript. (b) Transcription activation by TGA1a using the 'untreated' HeLa general factors. (c) Transcription activation by TGA1a using the 'depleted' HeLa general factors. W, wild type promoter; M, mutant promoter; +, addition of TGA1a (50 ng); -, no addition of TGA1a. Arrowhead indicates the specific transcript of 380 nt.

**a**

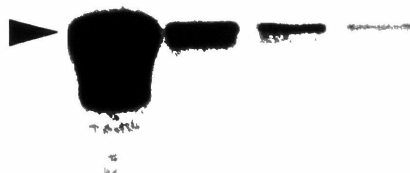


**b**

Promoter	W		M	
TGA1a (50 ng)	+	-	+	-

**c**

W		M	
+	-	+	-



1 2 3 4 1 2 3 4

4.2b and 4.2c). In subsequent experiments, the 'depleted' HeLa general factors were used.

The transcription activation level was measured at different concentrations of TGA1a (Fig. 4.3). With 100 ng per assay of TGA1a, 8 to 20-fold stimulation is reproducibly observed. The amounts of transcript show an approximately linear increase with increasing amounts of TGA1a. This result suggests that within the concentration range examined TGA1a does not appear to activate transcription cooperatively.

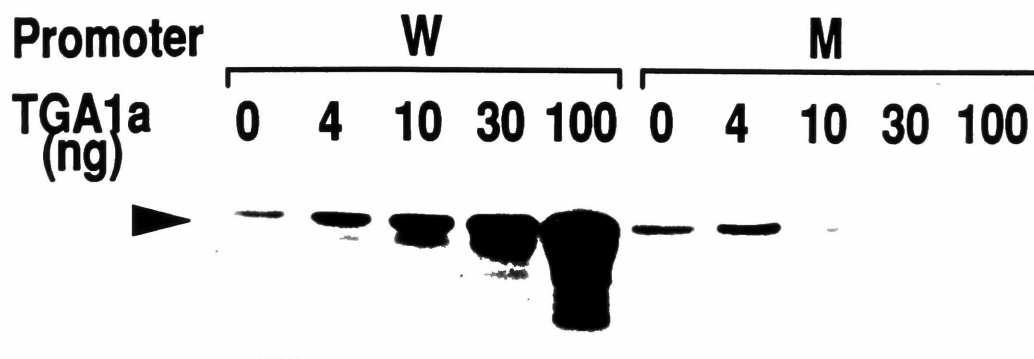
#### Accurate transcription initiation from the 35S promoter in the HeLa cell system.

Because the HeLa cell *in vitro* transcription system is heterologous to the 35S promoter, it was important to ascertain whether accurate transcription initiation is retained in this system. Transcription initiation from the 35S promoter in the HeLa cell system (nucleotide A circled with a thick line in Fig. 4.4) occurs at the same site as in transgenic tobacco (Odell *et al.*, 1985) with respect to the position of the TATA box, although in the DNA template the sequence of the 35S promoter downstream of -4 is different from that of the authentic 35S promoter (Fig. 4.2a). Addition of TGA1a does not alter the major transcription start site; however, a minor initiation site becomes apparent, that is about 30-fold weaker than the major site (Fig. 4.4, lane 5; nucleotide A circled with a thin line). Comparison of the signal strengths (Fig. 4.4, lanes 5 and 6) shows that TGA1a increases transcription by about 20-fold. This amount of stimulation is similar to that obtained using the G-free sequence (Fig. 4.3). Because there are no other significant initiation sites between -60 and +90 (only a part of this region is shown in Fig. 4.4), it is concluded that most transcripts shorter than 380 nt, that were observed in the assays using the G-free sequence (for example, Fig. 2c, lane 1), probably result from premature transcription termination (Carcamo *et al.*, 1989; Sawadogo and Roeder, 1985a).

Figure 4.3 Effects of increasing amounts of TGA1a on transcription activation.

(a) An autoradiogram showing the transcripts synthesized at different concentrations of TGA1a. Arrowhead indicates the position of the specific transcript. (b) The gel bands corresponding to the specific transcripts were excised and their radioactivities were determined by a scintillation counter. The amount of radioactivity was used to represent the relative transcription rate. W, wild type promoter; M, mutant promoter.

**a**



**b**

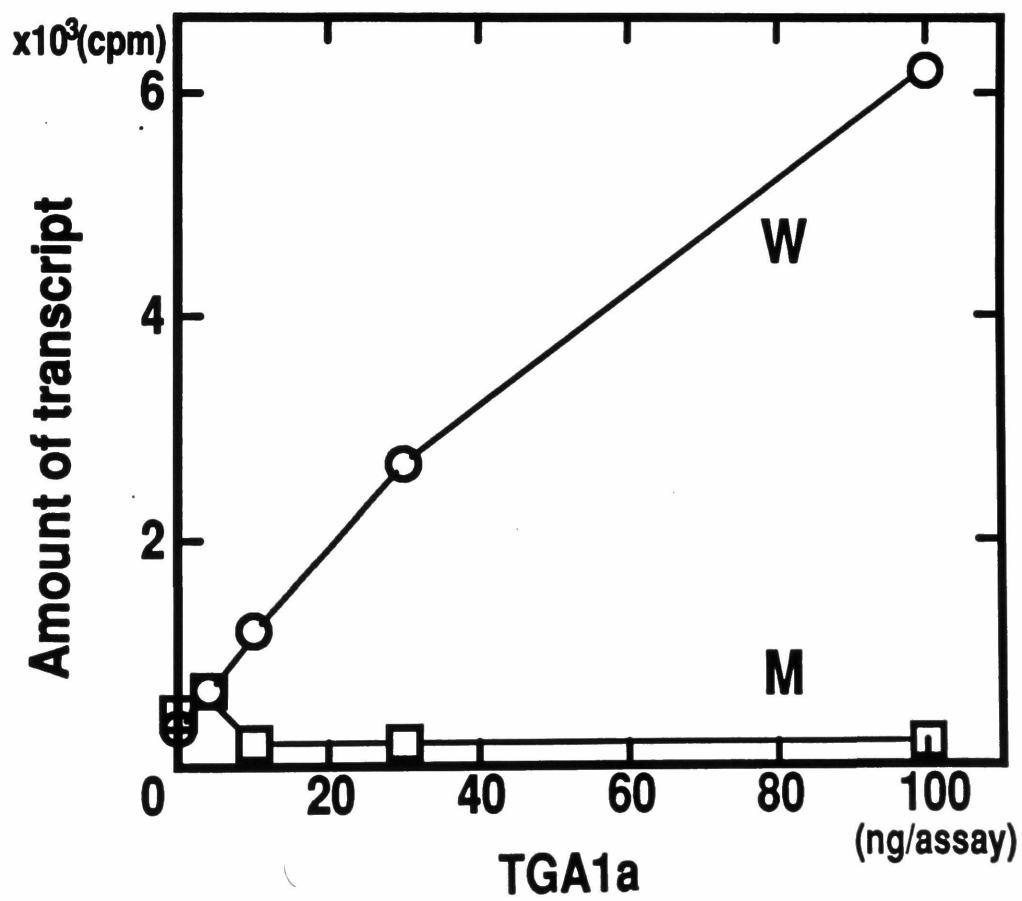
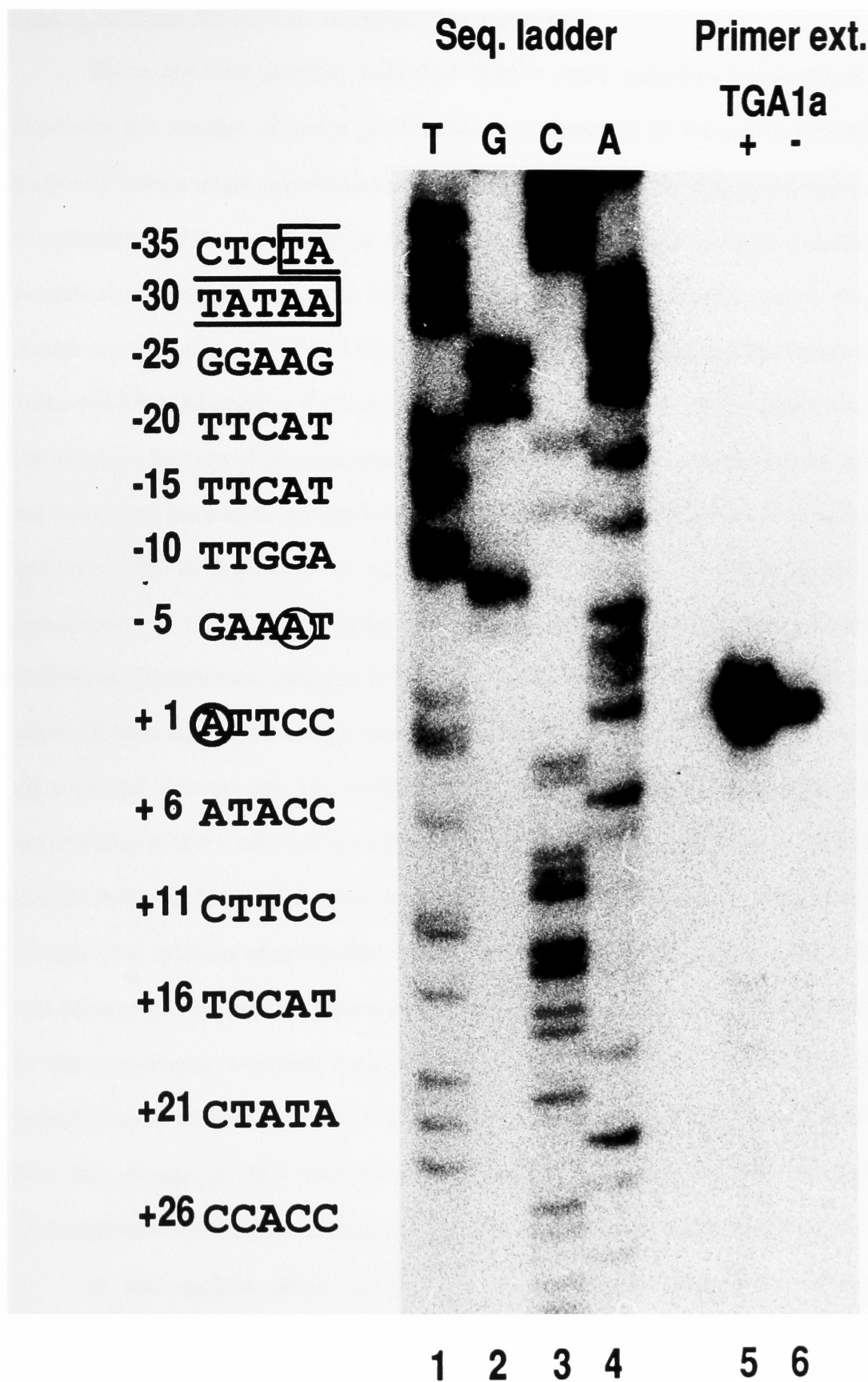


Figure 4.4 Transcription initiation sites of the CaMV 35S promoter in the HeLa cell *in vitro* system.

Transcription initiation sites in the *in vitro* system were determined by primer extension. The results of transcription reactions with (100 ng/assay) or without TGA1a are shown in lanes 5 and 6, respectively. A sequence ladder of the wild type promoter obtained with the same primer used in the primer extension is shown in lanes 1-4 as a reference. The sequence of the top strand shown in Fig. 4.2a is indicated on the left. The major and minor transcription initiation sites are circled with a thick line (+1) and a thin line (-2), respectively. The 35S TATA sequence is boxed.





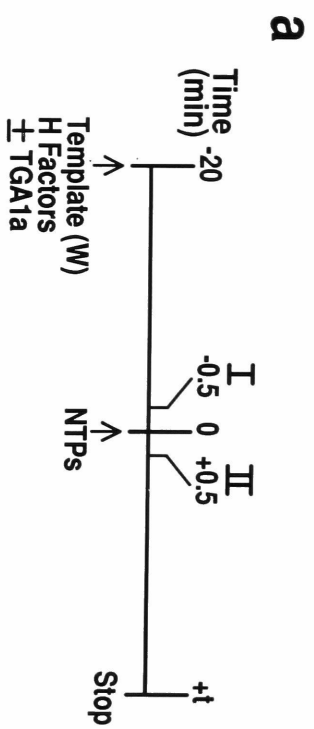
### TGA1a increases the number of preinitiation complexes.

There are four possible ways that TGA1a could stimulate transcription: 1) by increasing the number of active preinitiation complexes, 2) by increasing the initiation frequency from a single preinitiation complex, 3) by increasing the elongation rate or 4) by a combination of these possibilities. I examined whether TGA1a increases the number of preinitiation complexes under the condition of limited transcription reinitiation (Fig. 4.5). A high concentration of KCl (0.3 M) inhibits initiation but does not inhibit elongation in a HeLa cell RNA polymerase II transcription system (Cai and Luse, 1987). Thus when added just after the first round of initiations, 0.3 M KCl inhibits further initiation events, such that the number of transcripts corresponds to the number of active preinitiation complexes (ie. one transcript is derived from one preinitiation complex.). I define single round transcription as that observed under conditions in which one initiation event from each preformed preinitiation complex is allowed, and multiple round transcription as that observed when more than a single round of transcription initiation is allowed. Although the experimental systems are not completely the same as those in this study, previous observations with the adenovirus major late promoter in HeLa cell *in vitro* systems showed that the first round of initiation from preformed preinitiation complexes is completed within 0.5 min after addition of nucleoside triphosphates (NTPs; Hawley and Roeder, 1985) and that subsequent rounds of initiations do not occur within 5 min (Hawley and Roeder, 1987). In the experiments described here, the concentration of KCl was raised from 50 mM (standard assay conditions) to 0.3 M 0.5 min after the initiation of the reaction. I assumed that this change in KCl concentration occurs early enough to exclude significant contributions from the second round of initiation to the final amount of transcript.

In the analysis shown in Fig. 4.5, preinitiation complexes were formed by

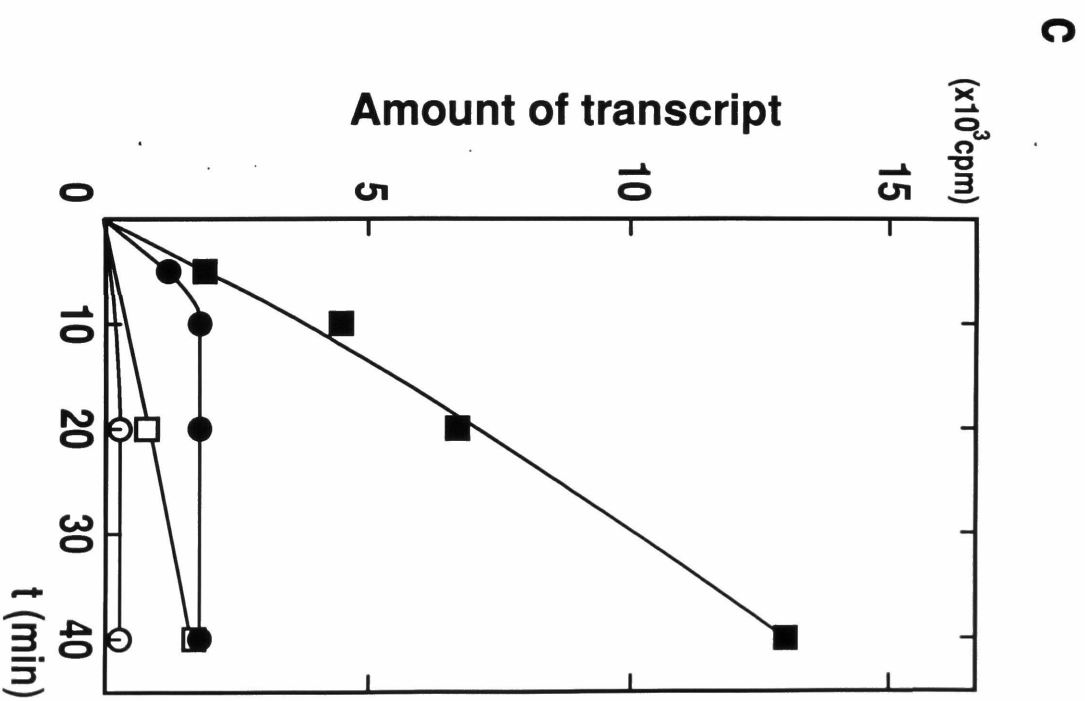
Figure 4.5 TGA1a stimulates a single round of transcription *in vitro*.

(a) The time table for the experiments. H Factors, HeLa general factors; NTPs, ribonucleoside triphosphates. (b) Transcription stimulation by TGA1a is observed in both single and multiple round transcriptions. The transcription reaction was carried out with (+; 100 ng/assay) or without (-) TGA1a. The final concentrations of KCl and the times when the concentration was changed from 50 mM KCl (standard conditions) to 0.3 M KCl are indicated. t, incubation time as indicated in (a). Arrow heads indicate the 380-nt specific transcript. (c) The transcripts shown in (b) were quantified by measuring the amounts of radioactivity of the corresponding bands by a scintillation counter. The amount of radioactivity was used to represent the relative transcription rate. Open circle, 0.3 M KCl at II, -TGA1a; closed circle, 0.3 M KCl at II, +TGA1a; open square, 50 mM KCl, -TGA1a; closed square, 50 mM KCl, +TGA1a. Results obtained in the same experiment are shown.



**b**

	-TGA1a	+TGA1a
0.3 M KCl at I	<p>t (min)</p> <p>40</p> <p>lane 1</p>	<p>40</p> <p>lane 2</p>
0.3 M KCl at II	<p>t (min)</p> <p>20 40</p> <p>lane 3 4</p>	<p>5 10 20 40</p> <p>lane 5 6 7 8</p>
50 mM KCl	<p>t (min)</p> <p>20 40</p> <p>lane 9 10</p>	<p>5 10 20 40</p> <p>lane 11 12 13 14</p>



preincubation of the DNA template with the HeLa general factors in the presence and absence of TGA1a. The transcription reaction was then initiated by the addition of NTPs (Fig. 4.5a). Addition of KCl (final 0.3 M) at 0.5 min before the addition of NTPs (time point I in Fig. 4.5a) almost completely inhibits subsequent transcription initiation (lanes 1 and 2 in Fig. 4.5b), whereas addition of the same concentration of KCl at 0.5 min after the addition of NTPs (time point II in Fig. 4.5a) does not. In the latter case, accumulation of the transcript terminates within 10 min (lanes 3-8 in Fig. 4.5b). The only transcript that accumulates was 380 nt (lanes 3, 4 and 6-8). The appearance of only this full length transcript suggests that under these conditions the second round of transcription is indeed inhibited (Carcamo *et al.*, 1989; Sawadogo and Roeder, 1985a). 0.3M KCl appears to affect the elongation rate slightly (compare lanes 5 and 11 in Fig. 4.5b). After a 5-min incubation at 0.3 M KCl, the average length of the transcripts is about 310 nt (Fig. 4.5b, lane 5). On the other hand, after a 5-min incubation at 50 mM KCl (standard assay conditions) most of the transcripts are full-length (380 nt; Fig. 4.5b, lane 11). Considering this difference in the transcript size (shorter transcripts contain less radioactivity), the number of transcripts produced after a 5-min incubation at 0.3 M KCl is approximately the same as that produced at 50 mM KCl. This similarity in the number of initiated transcripts at early times of the transcription reaction also supports the notion that a single round of transcription is observed in these experiments. At both KCl concentrations, the addition of TGA1a causes about an 8-fold stimulation of transcription (compare solid and open squares for 50 mM KCl or solid and open circles for 0.3 M KCl in Fig. 4.5c). Thus the degree of stimulation by TGA1a is almost the same under single and multiple round transcription conditions. This indicates that TGA1a stimulates transcription principally by increasing the number of preinitiation complexes rather than by increasing either the initiation frequency from a

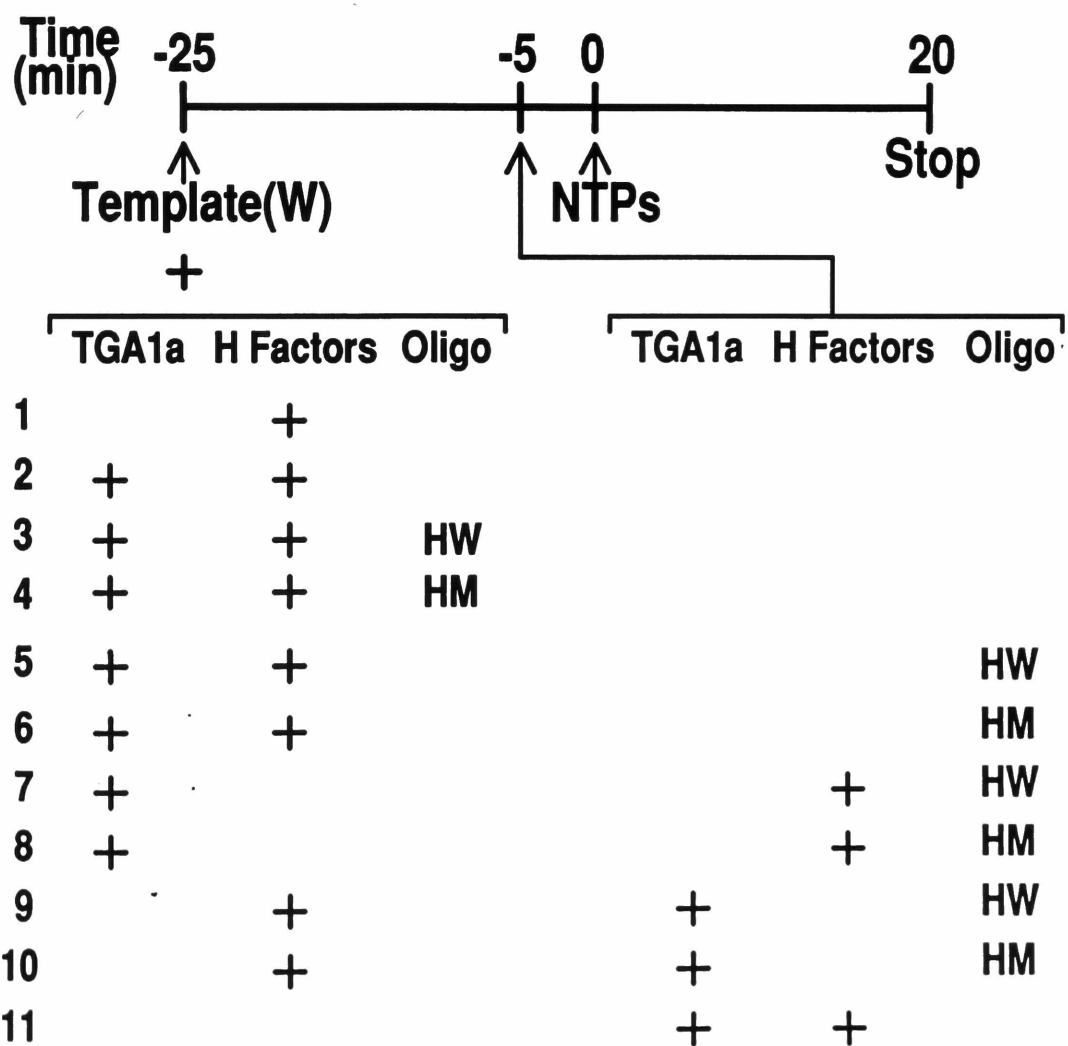
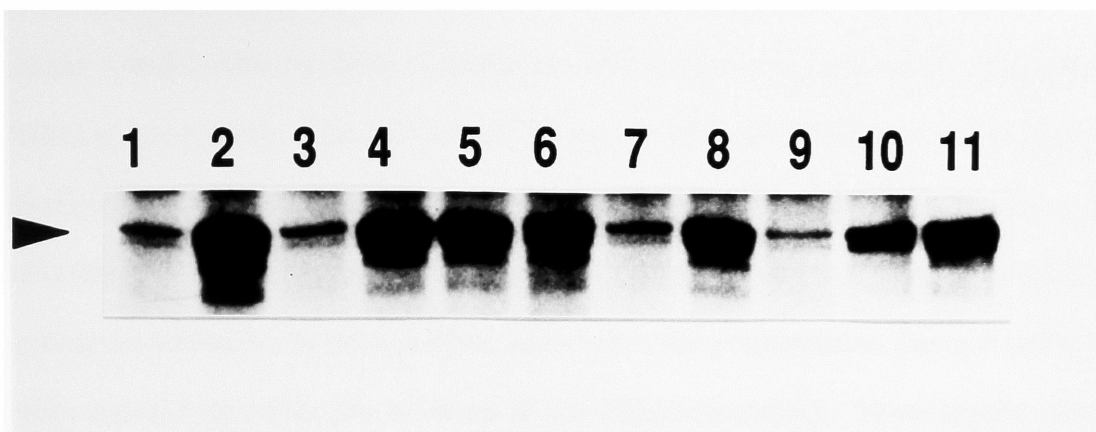
single preinitiation complex or the elongation rate.

TGA1a-dependent preinitiation complexes remain committed during multiple round transcription.

Relevant to the mechanism of action of TGA1a, I investigated whether the TGA1a-dependent preinitiation complexes remain committed to the initial templates during multiple rounds of transcription. If so, it would raise the possibility that TGA1a is required only for establishment of the activated state of transcription. For the study of template commitment of preinitiation complexes, it is desirable that the transcription system be capable of efficient reinitiation. As shown in Fig. 4.5c, the absolute level of transcripts at 40 min post-initiation is 8-fold higher under multiple round transcription conditions than under single round transcription conditions, in either the presence or absence of TGA1a (compare closed square and circle or open square and circle). This indicates that efficient reinitiation occurs under multiple round transcription conditions (standard assay conditions). The occurrence of large amounts of discrete shorter transcripts under multiple round transcription conditions (for example Fig. 4.5, lane 14) also suggests efficient reinitiation, because these shorter transcripts are thought to represent premature terminated products of latter rounds of transcription on the same G-free sequence template (Sawadogo and Roeder, 1985a; Carcamo *et al.*, 1989).

First, the requirement of TGA1a was investigated before and after initial preinitiation complex formation by the addition of either an oligonucleotide (HW) that contains a TGA1a-binding site or a mutant oligonucleotide (HM) that does not bind TGA1a (Fig. 4.6). Preinitiation complexes were formed and the reaction was arrested at this stage by preincubation in the absence of NTPs. Transcription was initiated by the addition of

Figure 4.6 TGA1a-dependent preinitiation complexes remain committed to templates. The time table of the transcription reaction is shown below the autoradiogram. The program for each lane on the top is shown at the bottom. The program indicates the time point at which the indicated components were added to the reaction mixture. 100 ng/assay of TGA1a was used. HW, the oligonucleotide containing the wild type *hex-1* (200 ng/assay); HM, the oligonucleotide containing the mutant *hex-1* (200 ng/assay). The other abbreviations are the same as in Fig. 4.5. Arrowhead indicates the specific transcripts.



NTPs and the transcripts generated in 20 min were analyzed (see the time table in Fig. 4.6). Lanes 1 and 2 show the level of transcript without and with transcription stimulation by TGA1a, respectively. The addition of 200 ng per assay of HW before the preincubation abolishes the transcription stimulation by TGA1a (lane 3) but the same amount of HM decreases the transcript only slightly (lane 4). By contrast, neither HW nor HM significantly affects the activation by TGA1a when added after the preincubation (lanes 5 and 6; NTPs were added 5 min after the addition of the oligonucleotides.). These results show that TGA1a is required during the formation of preinitiation complex to stimulate transcription, as expected from the above described action of TGA1a in increasing the number of preinitiation complexes. They also raise the possibility that TGA1a may not be needed in subsequent steps, since the addition of HW after the preinitiation complex formation does not significantly affect the overall level of transcription activation. However, we cannot exclude the possibility that TGA1a may persist and/or be required after preinitiation complex formation because it could form a very tight complex (with the HeLa general factors) that becomes resistant to competition by HW. To assemble the HW-resistant complex, both TGA1a and the HeLa general factors are needed during the preincubation. Omission of either component prevented the formation of the HW-resistant complex (lanes 7-10).

In addition to showing a requirement of TGA1a during preinitiation complex formation, this experiment indicates that HW does not significantly affect the level of TGA1a-dependent transcription even under conditions of multiple rounds of transcription, once preinitiation complexes for the first round are formed. This suggests that some part of the preinitiation complex remains committed to the promoter site, even after transcription initiation, and that this complex is preferentially used for the secondary



initiations by RNA polymerase. For reasons stated above these experiments do not establish whether template commitment during multiple round transcription occurs in the absence of TGA1a.

## DISCUSSION

In this chapter, I demonstrate that TGA1a can function as a transcription activator in a human *in vitro* transcription system. I also show that TGA1a activates transcription by increasing the number of preinitiation complexes on the promoter. These results provide strong evidence that plants and animals share at least one common transcription activation mechanism.

### The purified TGA1a fraction contains three related polypeptides

The purified fraction of TGA1a contained three polypeptides with different translation start sites (Fig. 4.1b). The two shorter polypeptides probably resulted from use of two other methionine codons (Met41 and Met56) as translation start sites (Fig. 4.1d) in *E. coli*. All three polypeptides are likely to be active in DNA binding, because they were purified by two different DNA affinity chromatography steps. The full-length TGA1a (band I), however, showed only weak binding activity in Southwestern blot analysis (Fig. 4.1c). This result could be explained if there is differential renaturation of the three polypeptides after denaturation by SDS during SDS-PAGE. After denaturation in guanidine hydrochloride, the renatured full-length TGA1a barely recovered its binding activity. By contrast, with the same treatment a shorter derivative of TGA1a (containing only the DNA binding domain,  $\Delta N80\Delta C144$  in chapter 6) readily regained most of its binding activity (K.

Seipel, unpublished observation).

I do not know whether all three TGA1a derivatives are equally active in transcription activation. Although the two shorter derivatives lack a portion of the acidic region that is presumably an activation domain, they retain the entire region of another putative activation domain, the glutamine-rich region (see below). Investigation of the transactivation domain of TGA1a is described in chapter 6.

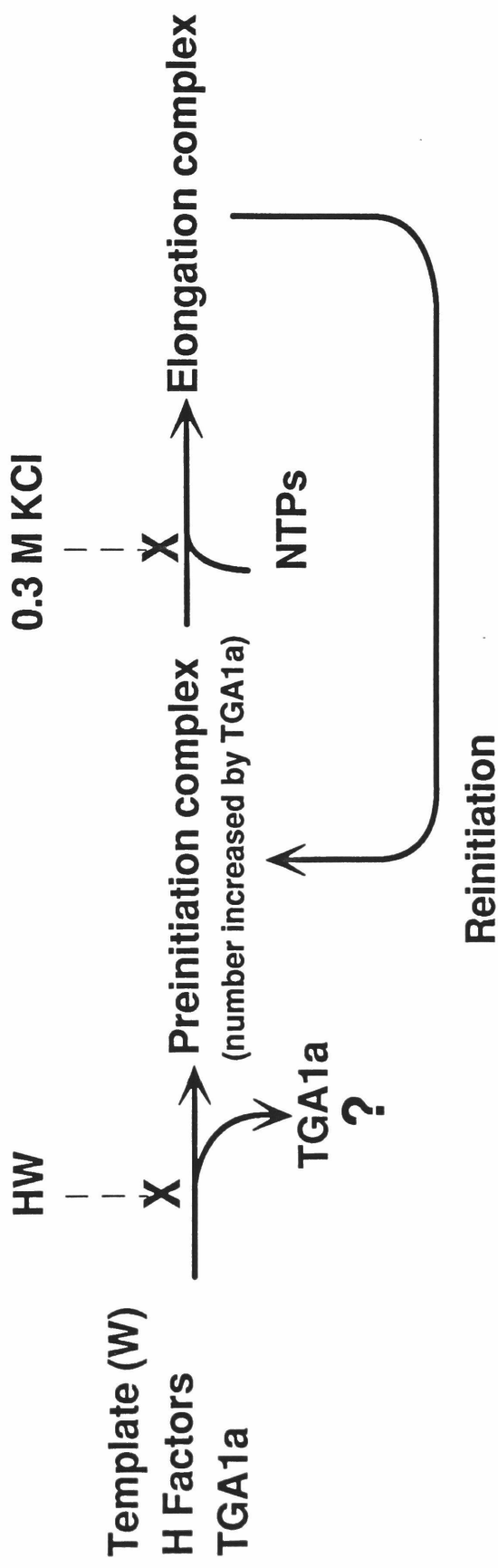
### Transcription activation mechanism of TGA1a

Taking advantage of an *in vitro* transcription system, I demonstrated that TGA1a increases the number of preinitiation complexes and that the TGA1a-dependent preinitiation complexes remains committed to the templates under multiple round transcription conditions. Figure 4.7 summarizes the activation mechanism of TGA1a. When preinitiation complexes are formed with HeLa general factors (H factors) on the template with the wild type TGA1a-binding site (Template (W)), the number of complexes is increased by TGA1a. This step of activation is inhibited by oligonucleotide HW, that contains a TGA1a-binding site. It is not known whether TGA1a remains associated with the preinitiation complexes or not. Transcription is initiated by the addition of ribonucleoside triphosphates (NTPs), and the preinitiation complex is converted into an elongation complex. This initiation step is inhibited by 0.3 M KCl. For reinitiation of transcription, some part of the preinitiation complex that remains committed to the template after the first round of initiation is reused for the next round of initiation. This step circumvents the HW-sensitive step.

It is of interest to know which factors are involved in the part of the preinitiation complex that remains committed to the template and is reused for transcription reinitiation.

Figure 4.7 Transcription activation by TGA1a.

A proposed model showing the sequence of events in transcription initiation (Hawley and Roeder, 1987) and the activation mechanism of TGA1a. The oligonucleotide HW inhibits the increase in the number of the preinitiation complexes. 0.3 M KCl inhibits the conversion of the preinitiation complex into the elongation complex. The reinitiation of transcription involves a complex which bypasses the HW-sensitive step. It has not been ascertained whether TGA1a is released after formation of the preinitiation complex. Abbreviations are the same as in Fig. 4.5.



The absence of TGA1a in this complex would indicate that TGA1a is required only transiently to increase the number of initial preinitiation complexes and that the resulting preinitiation complexes can reinitiate transcription several times. In previous studies on the adenovirus major late promoter, preferential utilization of templates used in the first round of transcription was reported (Hawley and Roeder, 1987). When analyzed by footprinting assays in the presence of its sequence-specific activator USF, some complex minimally containing USF and TFIID was found to remain at the promoter site after initiation (Van Dyke *et al.*, 1988). In the case described in this chapter, a similar complex including TGA1a might remain at the promoter site although this complex is resistant to HW. To address this question, it may be possible to isolate the template-associated complex for further biochemical characterization, because of the apparent stability of the complex during several rounds of initiations. The notion that sequence-specific DNA-binding proteins might be needed for the establishment but not the maintenance and continued utilization of active preinitiation complexes was also suggested from studies of adenovirus E4 promoter activation by ATF (Hai *et al.*, 1988; Horikoshi *et al.*, 1988b) but was best demonstrated for yeast 5S and tRNA genes which are transcribed by RNA polymerase III (Kassavetis *et al.*, 1990).

What could "increasing the number of preinitiation complexes" mean in *in vivo* situations?

In the HeLa cell *in vitro* system, only a limited number of the total templates are utilized. Addition of TGA1a increases the number of preinitiation complexes; this means an increase in the number of promoters that are used for the transcription reaction. When extrapolated to *in vivo* conditions, this situation is easy to interpret for genes that occur in multiple copies in the genome (eg. 5S DNA for RNA polymerase III transcription; Brown

and Gurdon, 1978). Most genes transcribed by RNA polymerase II, however, have only one or a few copies per genome. This is also the case for the -90 35S/GUS transgene in the transgenic tobacco plants that are often used for investigation of the *as-1* function. It is reasonable to interpret the statistical change observed in a large population (a change in the number of promoters used among the large number of promoters) as the change in probability for a single member in the population. Based on this interpretation, I offer two explanations for the *in vivo* transcriptional regulation by the mechanism which is observed as increasing the number of preinitiation complexes *in vitro*. In the following discussion, for the sake of simplicity, I assume that only one promoter of interest exists in a single cell. This assumption would not compromise the fundamental conclusions of the discussion when a single cell contains a few, but not many, promoters of the same kind.

The probability in which a promoter is utilized can be directly correlated with the probability in which a single cell expresses the promoter. When a population of cells are observed, the number of cells expressing the promoter increases upon activation. This situation is observed during the induction of a promoter with GRE (glucocorticoid responsive element) by dexamethasone in clonal Ltk<sup>-</sup> cells stably transformed with the promoter-reporter gene (Ko *et al.*, 1990). The addition of dexamethasone increases the number of cells expressing the promoter in a dose-dependent manner. A similar situation is suggested by the induction of a promoter with NF-AT binding sites (originally from the IL-2 promoter) during T-cell activation (Fiering *et al.*, 1990). In this system single cells show either very low or high expression level of the promoter with very few cells expressing the promoter at an intermediate level. This observation indicates that there are only two stages in terms of the promoter expression; for a given cell the expression is either on or off. This observation could be interpreted to mean that induction of the promoter linked to NF-AT

binding sites is achieved by increasing the number of cells expressing the promoter.

In the *in vitro* system, the TGA1a-dependent preinitiation complexes remains committed to a promoter during several transcription initiation events (Fig. 4.6). However, this observation does not exclude the possibility that the commitment is not strong enough to last for hours or that the commitment is not strong at all *in vivo*. If a weak commitment *in vivo* is assumed, it is possible to obtain an increased expression level of a promoter in a single cell in a factor-dose dependent manner. The probability of preinitiation complex formation is directly correlated with the association rate of the complex formation. If the preinitiation complex dissociates with a certain rate, the average time in which a promoter has the preinitiation complex formed on it (ie. the average time the promoter is activated) is determined by the ratio of the association to dissociation rates (ie. equilibrium). Because an increasing amount of transcription activator results in an increase in the association rate, in a relatively long time range, a single cell expresses the promoter at a level that is correlated with the amount of transcription activator.

#### Transcription activation mechanisms in plants versus animals

There is evidence that transcription activation mechanisms are conserved among unrelated eukaryotes because a yeast transcription activator GAL4 can function in various eukaryotes (Ptashne, 1988), including plants (Ma *et al.*, 1988). The work described here demonstrates for the first time that a plant transcription factor can function in an animal system. Because the activity of GAL4 in plants was examined by an *in vivo* transient expression system, it was difficult to investigate the mechanisms by which the yeast factor stimulated transcription in plants. By contrast, using an *in vitro* system, I was able to demonstrate that the plant factor, TGA1a, activates transcription by increasing the number

of preinitiation complexes and that these complexes remain active and committed during several rounds of transcription.

The transcription activation mechanism that enhances the preinitiation complex formation appears to be common among animal transcription activators so far examined, such as ATF (Hai *et al.*, 1988; Horikoshi *et al.*, 1988b), USF (Workman *et al.*, 1990) and the progesterone receptor (Klein-Hitpass *et al.*, 1990). Preinitiation complexes formed in the presence of TGA1a were resistant to the oligonucleotide containing TGA1a-binding site (Fig. 4.6). Similar stable preinitiation complexes were obtained with ATF (Hai *et al.*, 1988), USF (also known as MLTF; Carcamo *et al.*, 1989), and the progesterone receptor (Klein-Hitpass *et al.*, 1990). The demonstration that TGA1a also acts in the same manner to stimulate transcription, therefore, strengthens the notion that transcription activation mechanisms are conserved between plants and animals.

Although I have shown in this chapter that TGA1a functions as a transcription activator, the protein domain responsible for this activation remains to be identified. As proposed in chapter 3, an acidic region (amino acid residues 20-67) and a relatively glutamine-rich region (amino acid residues 259-347) of TGA1a may serve as activation domains (Katagiri *et al.*, 1989), because such motifs are known to activate transcription in yeast and animal cells (Mitchell and Tjian, 1989; Ptashne, 1988). In the case of GAL4, an acidic region also functions as an activation domain in plant cells (Ma *et al.*, 1988). Analysis of the activation domain of TGA1a is described in chapter 6.



**Appendix to chapter 4: TGA1a can stimulate transcription**  
**in a plant in vitro system.**

In chapter 4, the function of TGA1a was studied using a HeLa cell *in vitro* transcription system. This system was chosen because it is a well characterized *in vitro* system for promoters that are transcribed by RNA polymerase II. Moreover, many studies on sequence-specific transcription activators have been carried out with this system, including those of heterologous origins, eg. GAL4 from yeast (Lin *et al.*, 1988) and Ftz from *Drosophila* (Ohkuma *et al.*, 1990).

It has been suggested that transcriptional regulation mechanisms are conserved among various eukaryotes (chapter 4; Ptashne, 1988). On the other hand, there could also be some differences in regulation mechanisms among different eukaryotes. For example, different results were obtained when the transactivation domains of bicoid protein, a *Drosophila* transcription activator, were analyzed in stably transformed yeast as compared to transiently transformed *Drosophila* cultured cells or *Drosophila* embryos (Driever *et al.*, 1989). These differences emphasize the need to analyze a transcription factor in a homologous system. In this respect, transcription studies in plants have been hindered by the lack of a homologous *in vitro* system that can reconstruct some features of transcription regulation.

Wheat germ extract (WGE) is enriched in RNA polymerase II (Jendrisak and Burgess, 1975) and, therefore, serves as a good source of this enzyme. Yamazaki and Imamoto (1987) have reported accurate transcription initiation from the TC7 promoter of T-DNA using a WGE *in vitro* system. Therefore, WGE likely contains all the general transcription factors required for transcription initiation, in addition to RNA polymerase II.

The extract also recognizes the TATA box of the CaMV 35S promoter and initiates transcription at the correct start site (K.Yamazaki, unpublished). However, it is not known whether the WGE *in vitro* transcription system can respond to exogenously added transcription factors.

Here I show that the WGE *in vitro* system can respond to the addition of TGA1a, which functions as a sequence-specific transcription activator. Also, the activation mechanism of TGA1a in this system is the same as that in the HeLa cell system. In both systems, TGA1a stimulates transcription by increasing the number of preinitiation complexes. So far, the WGE system is the only plant *in vitro* transcription system that has been reported to respond to an added transcription factor.

## RESULTS AND DISCUSSION

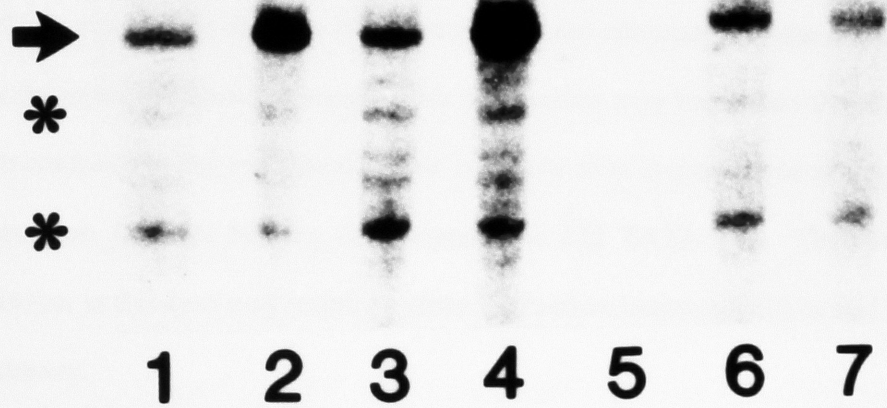
### TGA1a functions as a sequence-specific transcription activator in the WGE *in vitro* system.

The same DNA templates were used in this study as in chapter 4. When the wild type promoter was used as a template, the addition of TGA1a (0.15 µg/assay) gave ~3-fold transcription activation (compare lanes 1 and 2, Fig. 4.8). The specific 380-nt transcript is marked by an arrow in the figure. By contrast, when the mutant promoter was used as a template, the addition of TGA1a (0.15 µg/assay) gave a slight repression of transcription rather than stimulation (compare lanes 6 and 7, Fig.4.8). This result clearly shows that TGA1a stimulates transcription in a sequence-specific manner in the WGE *in vitro* system, as well as in the HeLa cell *in vitro* system (chapter 4). The addition of a relatively low concentration of  $\alpha$ -amanitin (1 µg/ml) inhibited the synthesis of the transcript (lane 5). This sensitivity to  $\alpha$ -amanitin indicates that the transcription is dependent on RNA polymerase II. The effect of wheat germ RNA polymerase II on the WGE *in vitro*

Figure 4.8 *in vitro* transcription stimulation dependent on TGA1a and its cognate binding site, *as-1*.

Transcription from either W or M promoter was assayed with (+) or without (-) the addition of TGA1a (0.15  $\mu$ g).  $\alpha$ -amanitin (1  $\mu$ g/ml) was added in lane 5. 1 unit (1.65  $\mu$ g) of RNA polymerase II from wheat germ (Sigma) was added in lanes 3 and 4. The specific transcript (~380 nt) is marked by an arrow. Asterisks indicate transcripts presumably arising from transcription initiation sites within the G-free sequence (see text).

Promoter	W					M	
TGA1a	-	+	-	+	+	-	+
$\alpha$ -amanitin	-	-	-	-	+	-	-



transcription system was examined. It was found that the addition of RNA polymerase II (1 unit/assay) increased the basal transcription level (lane 3), but the effect of TGA1a on transcription activation remained at a similar level (~3-fold; lane 4). This observation suggests that RNA polymerase II in this system is limiting for basal transcription. Since the system retains accurate transcription initiation after addition of RNA polymerase II, the other transcription factors required for accurate initiation appear to be present in excess.

Several RNA products smaller than the specific 380-nt transcript were also detected (Fig. 4.8). These products may be due to initiation at cryptic sites or to premature transcription termination in the G-free sequence. Among them, two major products (~ 320 and 260 nt, marked by asterisks in Fig. 4.8) presumably resulted from cryptic initiation sites (probably TATA-like motifs) in the G-free sequence, because these products were observed even when the 35S TATA-box region was deleted (not shown). These RNA products were also transcribed by RNA polymerase II because their synthesis was sensitive to  $\alpha$ -amanitin (lane 5). Interestingly, production of these RNA species was not stimulated by the addition of TGA1a, unlike the specific 380-nt transcript. This observation may imply that the effect of TGA1a is not transduced to the presumed cryptic initiation sites in the G-free sequence at ~150-bp downstream from its binding sites beyond the 35S TATA box. Therefore, transcription activation in this case may require a close interaction between TGA1a and the transcription machinery.

#### TGA1a increases the number of preinitiation complexes in the WGE *in vitro* system.

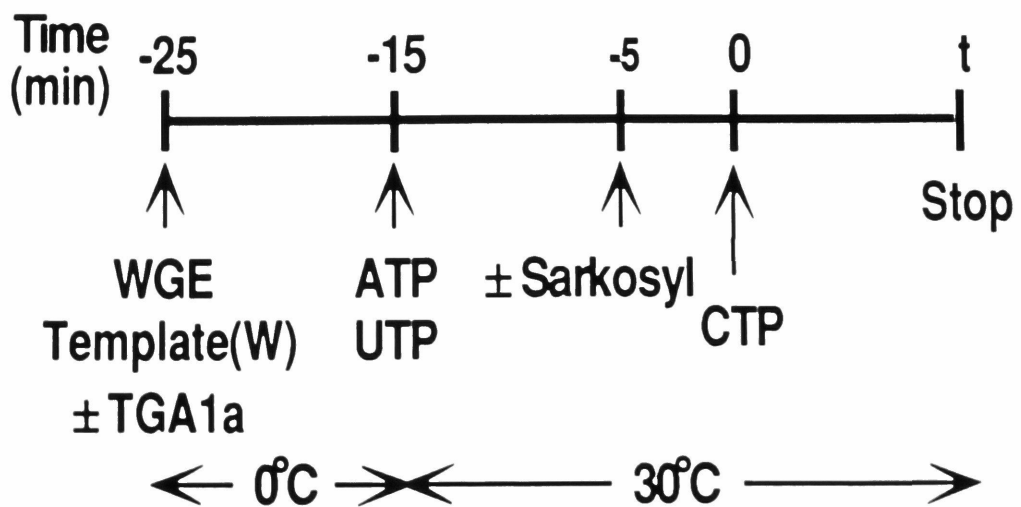
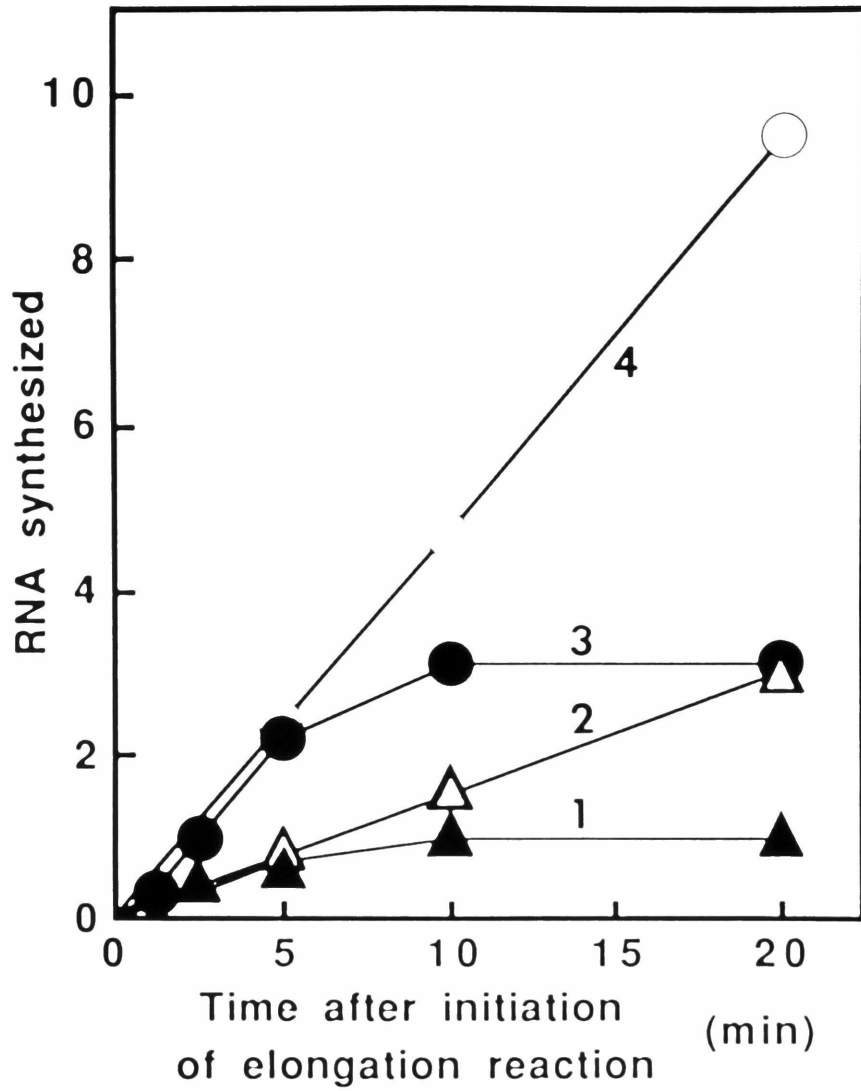
To investigate the transactivation mechanism of TGA1a, conditions were established for a single round of transcription in the WGE system, as it was done in the HeLa cell system (chapter 4). Instead of high salt concentrations, sarkosyl (*N*-lauroylsarcosine; an

anionic detergent) was used to block transcription initiation, because sarkosyl gave better results in this system. Sarkosyl has been successfully used for this purpose in animal *in vitro* transcription systems (Hawley and Roeder, 1985). The time table for single round transcription is shown in Fig. 4.9. After preincubation of WGE and template to allow the formation of preinitiation complexes, the transcription reaction was initiated by the addition of ATP and UTP to give a short transcript comprising of only the first three nucleotides (the first three nucleotides for the specific transcript are AUU.). After the synthesis of the first three nucleotides transcription was arrested because the forth nucleotide is C and CTP was omitted from the reaction mixture. Because the transcription machinery is stacked close to the promoter site, no new transcription initiation can occur from this site at this stage. Transcription elongation was resumed by the addition of CTP to the reaction mixture. Titration with various concentrations of sarkosyl showed that 0.07% sarkosyl was sufficient to inhibit transcription initiation, but not transcription elongation beyond the third nucleotide of the nascent transcript (not shown). Addition of 0.07% sarkosyl before the addition of CTP in this reaction inhibited further transcription initiation, but did not affect the elongation reaction. Using this concentration of sarkosyl, the effect of TGA1a on a single round of transcription was investigated.

In the absence of sarkosyl (ie. when multiple rounds of transcription are allowed), the addition of TGA1a (0.15  $\mu$ g/assay) gave ~3-fold stimulation of transcription (compare graphs 2 and 4, Fig. 4.9). Under conditions for a single round of transcription (ie. when sarkosyl was added before the addition of CTP), a similar level of stimulation was observed by the addition of TGA1a (0.15  $\mu$ g/assay; compare graphs 1 and 3). This result indicates that TGA1a stimulates transcription principally by increasing the number of preinitiation complexes rather than by increasing the initiation frequency at a single promoter site or by

Figure 4.9 Effect of TGA1a on single- and multiple-round transcriptions.

DNA templates (W promoter) were preincubated with WGE in the presence (graph 3 and 4) or absence (graphs 1 and 2) of TGA1a (0.15  $\mu$ g). Sarkosyl (final concentration, 0.07%) was added after transcription initiation as shown in the timetable at the bottom (graphs 1 and 3) or was not added at all (graphs 2 and 4). After addition of CTP, transcription elongation was terminated by adding the stop solution at the times indicated. Amounts of specific transcript synthesized were expressed relative to that synthesized in the reaction with sarkosyl and without TGA1a in 20 min.





increasing the elongation rate. This conclusion for the activation mechanism of TGA1a is the same as that obtained with the HeLa cell *in vitro* system (chapter 4).

#### Effect of TGA1a concentration on transcription initiation

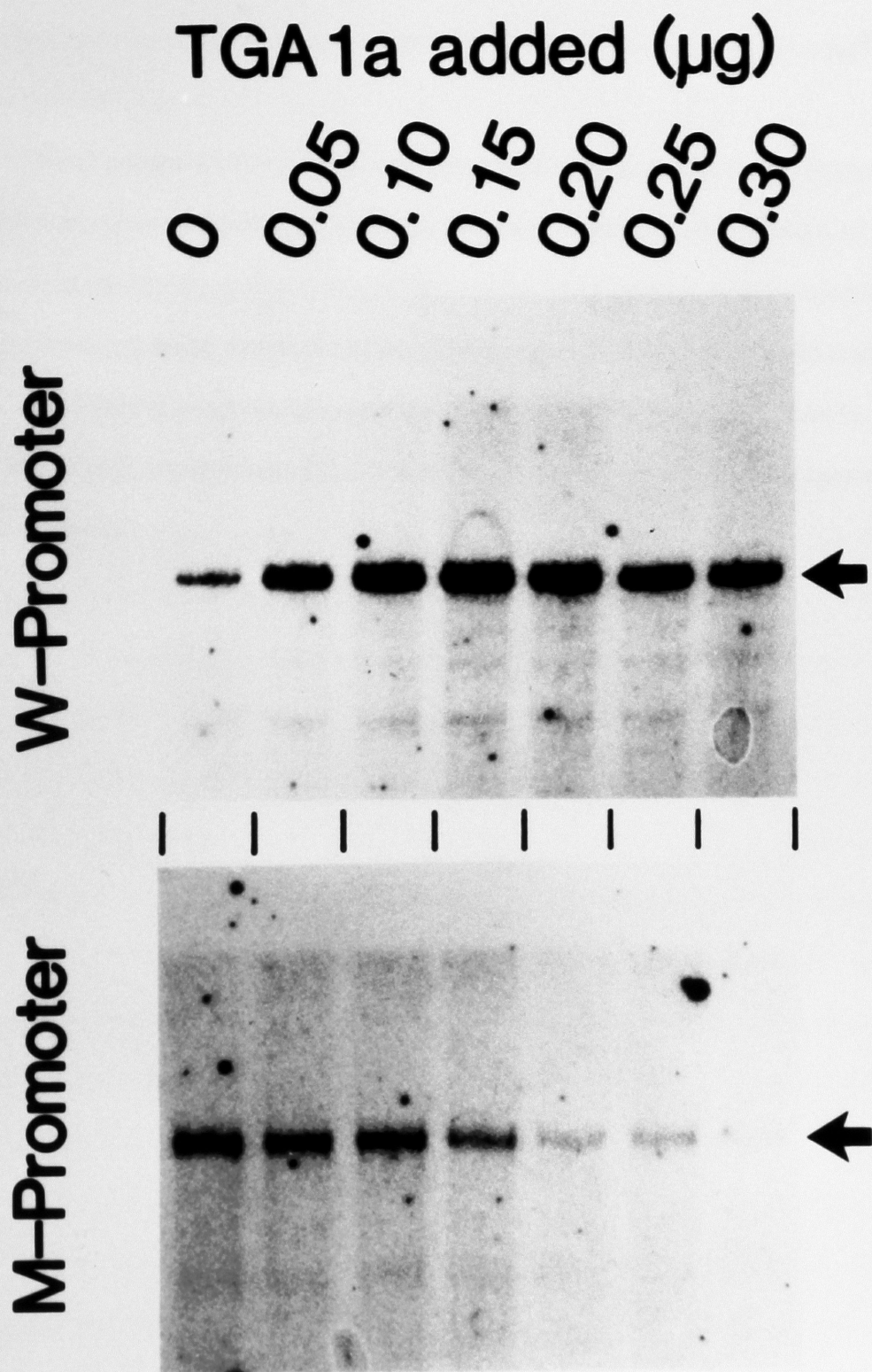
The effect of TGA1a concentrations on transcription was examined under the conditions for single-round transcription (Fig. 4.10). With the wild type promoter, transcription increased with increasing concentration of TGA1a up to 0.15 µg per assay, and a slight decrease rather than further increase was observed at higher concentrations. In the case of the mutant promoter, the transcription level gradually decreased with increasing TGA1a concentration. This decrease of transcription with the mutant promoter at any concentrations of TGA1a and the decrease with the wild type promoter at very high concentrations of TGA1a may be explained by "squenching" (Ptashne, 1988), in which the activation domain of unbound TGA1a titrates out general transcription factors or RNA polymerase II in the reaction mixture.

## CONCLUSIONS

Here I have described that under appropriate conditions transcription in the WGE *in vitro* system is responsive to the addition of the tobacco DNA-binding protein TGA1a. The stimulatory effect of TGA1a is sequence-specific, because only the promoter containing its cognate binding site (*as-I*) can respond. I have also shown that TGA1a activates transcription by increasing the number of preinitiation complexes in this *in vitro* system, because the observed activation level is similar when multiple rounds of transcription are allowed and when the transcription is limited to a single round. These observations are in agreement with the results obtained with the HeLa cell *in vitro* transcription system (chapter

Figure 4.10 Effect of TGA1a concentration on a single round of transcription.

DNA templates (W and M promoters) were incubated with WGE and various concentrations of TGA1a. Sarkosyl (0.07%) was added after transcription initiation as shown in the timetable in Fig. 4.9. Twenty min after addition of CTP, transcription elongation was terminated. Bands corresponding to the specific transcript are marked by arrows.



4). This consistency suggests that TGA1a functions in the same manner in both systems and that at least one transactivation mechanism used by TGA1a is conserved in both plant and animal systems.

The development of such a plant *in vitro* transcription system that is responsive to an added transcription factor is an important step in transcription studies of plants. Extension of the *in vitro* system described here to the investigations of other combinations of DNA-binding proteins and their cognate binding sites would be useful for the elucidation of the mechanism of transcription regulation. This system will enable the direct functional assay of not only sequence-specific DNA-binding proteins but also general transcription factors in plants.

## Chapter 5: TGA1a can stimulate transcription in leaf cells.

*In vitro* transcription systems offer the advantages of investigating directly any possible interactions between a transcription factor and its cognate binding site as well as with other factors and to determine the molecular mechanisms by which a transcription factor modulates transcription. Nevertheless, *in vitro* systems might produce artifactual results. Therefore, *in vivo* studies of a transcription factor are needed to complement results obtained from *in vitro* investigations.

*In vivo* assays of transcription factors are usually performed in cells that do not express the transcription factor of interest. Two constructs, an effector and a reporter constructs, are introduced into such cells. The effector construct is comprised of the coding sequence of transcription factor linked to a promoter that is active in the cells (often inducible promoters are used when the effector construct is stably transformed into the cells.). The reporter construct consists of a reporter gene linked to a promoter containing binding sites for the transcription factor. Methods for the introduction of these constructs into cells vary depending upon the organism in question. For example, in the case of yeast, usually a mutant that does not express the transcription factor of interest is stably transformed with the effector and reporter constructs. In the case of mammals, because various cultured cells are available, the most common strategy is to use cultured cells that do not express the factor for transient expression of the two constructs.

In the case of plants, the transient expression assay of a maize regulatory protein B has been reported recently (Goff *et al.*, 1990). Protein B is a positive regulator of anthocyanin biosynthesis genes in the vegetative parts of maize. Because the protein contains a region homologous to the HLH domain, it was expected to be a transcription

activator. The effector construct contained the 35S promoter linked to the coding sequence of protein B and the reporter construct contained the *Bz1* promoter linked to the GUS coding sequence. These constructs were introduced by microprojectile bombardment into the kernels and embryonic calli of a maize mutant carrying the *b* mutation. The results clearly demonstrated that the protein activates the *Bz1* promoter in a sequence-specific manner.

Although transient expression systems can provide a relatively rapid assay for transcription factors, they suffer from several drawbacks:

- 1) Conditions for the reporter gene is different from those in stably transformed cells. A large number of the reporter construct plasmid is introduced into a single cell. Because the DNA is not integrated into the genome, the higher order structures of the reporter gene are likely to be different from those of cellular genes. I will show in this chapter that a large number of the reporter construct plasmid can be expressed where a single copy of transgene cannot be expressed (see the section, "Multiple copies of -90 35S/GUS can confer expression in leaf cells.").
- 2) The efficiency of transformation varies for each cell. Although difference in transformation efficiency among experiments can be compensated by the use of a reference gene, difference in transformation efficiency for each cell cannot be quantified. When such differentially transformed cells are treated as a single population, the observed quantitative relation between the amount of factor and the reporter gene activity can deviate significantly from the real situation.

Because *as-1* is a positive regulatory element that confers preferential expression in root (Lam et al, 1989), the active form of the factor mediating *as-1* function is likely a positive regulatory factor. Moreover, it is presumably abundant in root cells but absent or

present in very low concentrations in leaf cells. If the factor is indeed present in low concentrations, it should be possible to obtain expression in leaf cells by increasing the copy number of *as-1* linked to the reporter gene. Indeed, it has been shown that four copies of *as-1* are able to confer high level expression in leaf and root, while one copy of the same element is active only in root (Lam and Chua, 1990). These results indicate that the lack of leaf expression of the -90 35S construct can be overcome by increasing the copy number of *as-1* linked to it. A likely interpretation is that the increased copy number of *as-1*, additively or synergistically, increases the probability of binding the factor even when the concentration of the factor is limiting.

In chapter 4, it has been shown that TGA1a can function as an *as-1*-specific transcription activator *in vitro*. Moreover, the low TGA1a mRNA level in leaf is consistent with the hypothesis that the factor occurs in a limiting concentration in leaf cells and consequently *as-1* is not expressed in such cells. Therefore, leaf cells would serve as a good system to assay for the activity of TGA1a *in vivo*.

In this chapter, I describe an *in vivo* assay of TGA1a using microinjection. TGA1a produced in *E. coli* was injected into cotyledon cells of transgenic tobacco plants harboring the -90 35S/GUS reporter gene construct, that contains the *as-1* element in its -83 to -63 region. Because the reporter gene construct is stably integrated into the tobacco genome and the assay is performed in a single cell injected with a known amount of the factor, the problems associated with transient expression assays are obviated in this system. Using this microinjection technique, TGA1a is demonstrated to induce the reporter transgene expression *in vivo* in a sequence-specific manner. Moreover, by injection of varying copy numbers of the reporter gene construct into leaf cells of wild type tobacco plants, evidence is obtained supporting the notion that the factor mediating *as-1* expression *in vivo* occurs in

limiting concentration in leaf cells.

## RESULTS

### Microinjection of TGA1a induces the -90 35S/GUS expression in cotyledon cells.

TGA1a was overproduced in *E. coli* and purified from the bacterial extract as described in chapter 4 (Katagiri et al.,1990). Seven to 10 day-old transgenic tobacco seedlings harboring either the -90 35S/GUS construct (A) or the -72 35S/GUS construct ( $\Delta$ A) were used for injection. The purified TGA1a fraction was injected into epidermal and/or subepidermal cells of the cotyledons. After 24-hour incubation at 22°C to allow expression of the reporter gene, the whole seedlings were processed for GUS activity detection by histochemical staining (see chapter 2). Control experiments confirmed previous observation that transgenic plants carrying construct A or  $\Delta$ A do not show any GUS activity in their cotyledonous cells (Benfey *et al.*, 1989). The promoter of construct A contains the entire *as-1* element, while that of construct  $\Delta$ A has the upstream TGACG-motif deleted.

When purified TGA1a ( $\sim 10^4$  molecules as a monomer) was microinjected into epidermal and subepidermal cells of cotyledons of construct A plants, GUS activity was detected 24 hours after injection in 67.8% (n=648) of the injected cells but not in the neighboring non-injected cells (Fig. 5.1a). Injection of BSA ( $\sim 10$  pg) into other cells of the same transgenic plant did not give any GUS activity (Fig. 5.1b; n=2,205). Therefore, the induction of the GUS activity with TGA1a is not simply a response to wounding induced by the injection procedure. Neither TGA1a (n=603) nor BSA (n=18) (proteins injected are the same amounts as above) gave any GUS expression when injected into the construct  $\Delta$ A plants, although a weak GUS expression was seen in a small number of cells with TGA1a if the incubation time was prolonged to five days from 24 hours ( $\sim 3\%$  of the injected cells).

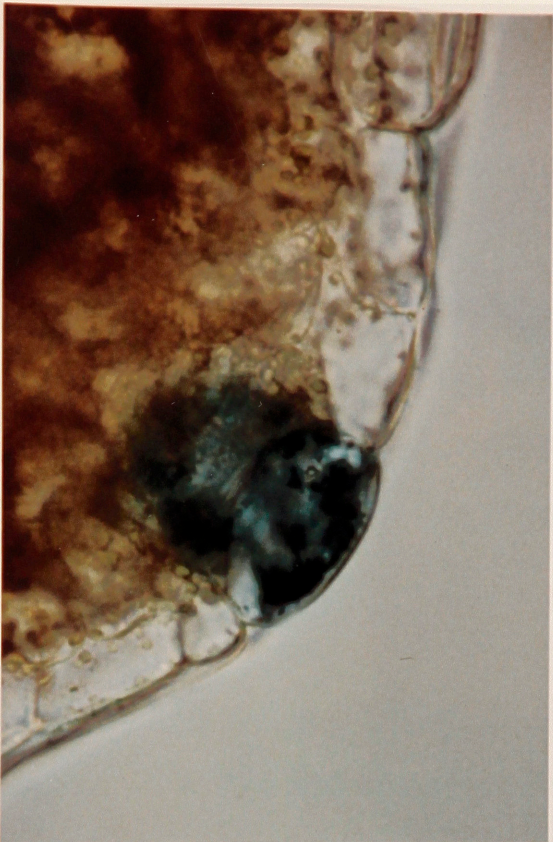


Figure 5.1 Injection of TGA1a induces expression of -90 35S promoter in leaf cells.

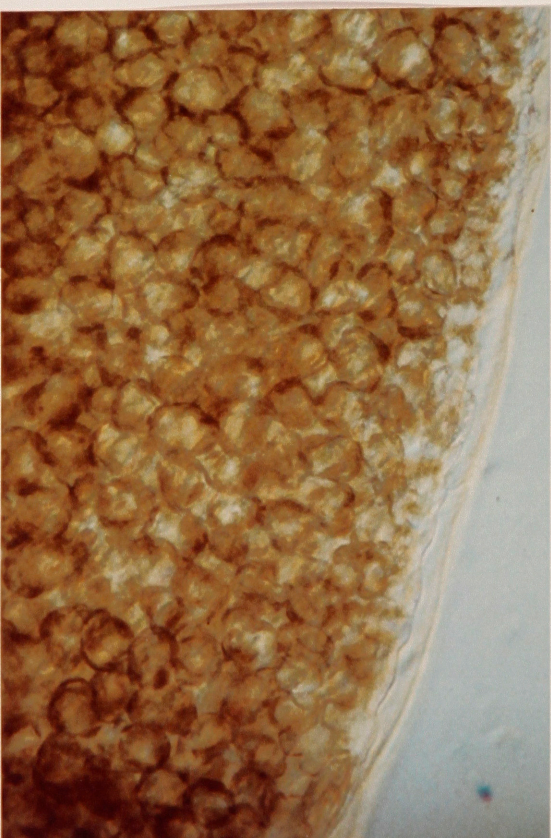
TGA1a ( $\sim 10^4$  molecule; **a**) or BSA ( $\sim 10$  pg; **b**) was injected into epidermal and subepidermal cells of cotyledons of the construct A plants. After histochemical staining of the GUS activity, the injected cells were observed under a microscope. The GUS activity is stained as blue color. Representative results are shown.



**a**



**b**





This weak GUS expression in construct  $\Delta A$  plants injected with TGA1a may result from TGA1a binding to the downstream TGACG-motif (-72 to -63). Taken together, these results demonstrate that TGA1a can also activate expression of the *as-1*-linked promoter in a sequence-specific manner *in vivo*.

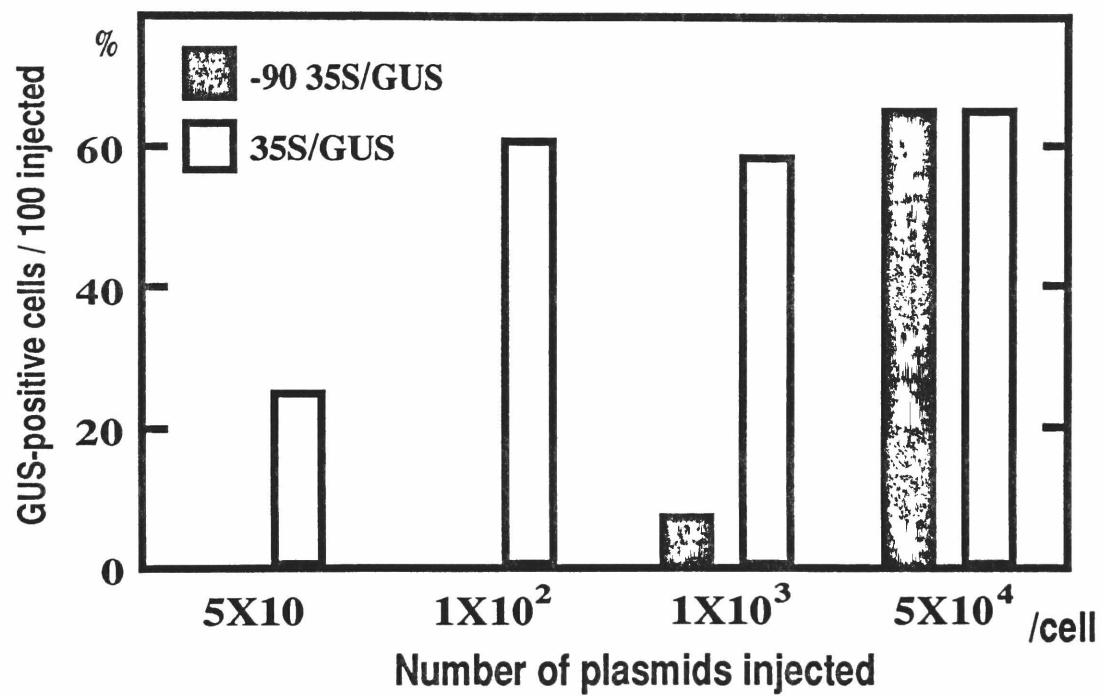
Titration experiments with varying concentrations of TGA1a showed that even ~1000 molecules per cell of TGA1a were sufficient for induction of GUS activity in construct A plants. Since this number of molecules is small compared with the concentrations of various proteins in a cell, TGA1a appears to function at physiological concentration *in vivo*.

#### Multiple copies of -90 35S/GUS can confer expression in leaf cells.

The induction of -90 35S/GUS expression in leaf cells by injection of TGA1a supports the assumption that *as-1*-specific factor is limiting. The results, however, do not exclude the possibility that the factor is totally absent in leaf cells. If the factor is present but in low concentrations, it should be possible to obtain expression of the -90 35S/GUS reporter gene if multiple copies of the gene are introduced into a single leaf cell by microinjection. This prediction was examined by microinjection of a varying number of copies of the reporter gene into wild type tobacco plants. The percentage of injected cells expressing the GUS activity was used as a measure of the expression level. When the 35S/GUS plasmid [the full 35S promoter (-343 to +8) linked to the GUS gene] was injected into leaf cells, 100 molecules of the plasmid per cell were sufficient to confer GUS expression in ~60% of the injected cells (Fig. 5.2, open bars; n=45). It seems inconsistent that more than one copy of the 35S/GUS plasmid were required to be injected for the high level of expression, while one copy of the 35S/GUS transgene is sufficient to confer a

Figure 5.2 Many copies of -90 35S/GUS plasmid are required to show the expression in leaf cells.

Indicated numbers of plasmid (solid bars, -90 35S/GUS plasmid; open bars, 35S/GUS plasmid) were injected into cotyledon cells of wild type tobacco plants. The numbers of GUS-positive cells per 100 injected cells are indicated.



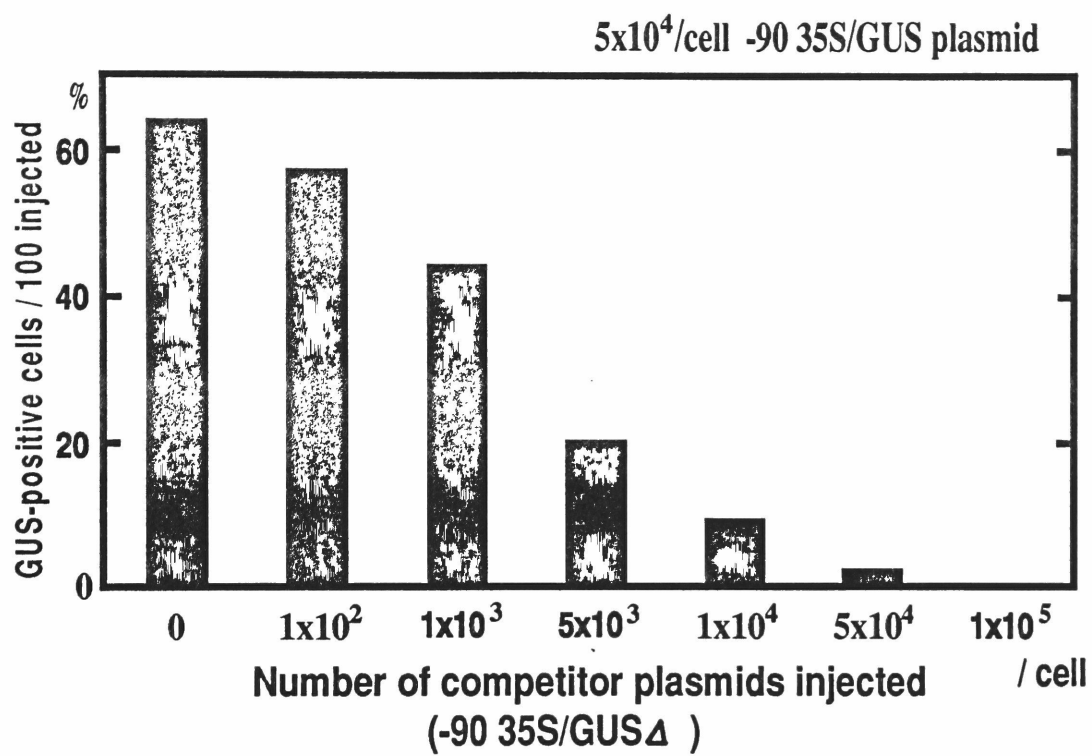
detectable expression. This apparent discrepancy could result from the low efficiency of expression by microinjection (eg. only a small fraction of the injected plasmids are transported into nucleus; plasmids may be degraded in cells.). Factors needed for the full 35S promoter expression are considered not to be limiting in leaf cells. Still further increase in the number of injected plasmids from 100 molecules per cell did not give a higher ratio of GUS-expressing cells. Probably, because of the damage by microinjection, 40% of the injected cells were unable to express the reporter gene. Therefore, I assume that the value, ~60% expressing cells, represents the maximum activity. By contrast, when the -90 35S/GUS plasmid was injected,  $5 \times 10^4$  molecules of the plasmid per cell were required to confer GUS expression in ~60% of the injected cells (Fig. 5.2, closed bars;  $n=45$ ). As a negative control, the -72 35S/GUS plasmid did not give any expression even up to  $5 \times 10^4$  molecules injected per cell ( $n=45$ ). This result is consistent with the hypothesis that the *as-1*-specific factor is limiting in leaf cells. This result also provides an example that a large number of reporter genes can be expressed where a single copy of the corresponding transgene is not expressed.

#### Expression of the -90 35S/GUS can be attenuated by coinjection of TGA1a-binding sites.

One trivial interpretation of the above results is to assume that GUS expression in leaf cells is a consequence of a gene-dosage effect of very weakly expressing -90 35S/GUS. To rule out this possibility, a competition experiment was performed using TGA1a-binding sites. A plasmid that has a deletion in the GUS-coding region of -90 35S/GUS (-90 35S/GUS $\Delta$ ) was used as a competitor and coinjected with the -90 35S/GUS plasmid. It was found that the GUS expression decreased with an increasing number of the competitor (Fig. 5.3). When  $5 \times 10^3$  molecules of the competitor were coinjected with  $5 \times 10^4$  molecules of -90

Figure 5.3 A similar concentration of the competitor plasmid effectively competes with expression of the -90 35S promoter.

Indicated numbers of the competitor plasmid, -90 35S/GUS $\Delta$  plasmid, were coinjected with  $5 \times 10^4$ /cell of -90 35S/GUS plasmid into cotyledon cells of wild type tobacco plants. The numbers of GUS-positive cells per 100 injected cells are indicated.





35S/GUS into leaf cells, only 20% of the injected cells expressed the GUS activity (n=45; cf. ~60% without the competitor). This attenuation of the -90 35S/GUS expression is likely due to a competition between the competitor gene and the reporter gene for the same factor, which is limiting in leaf cells. However, a proper control using -46 35S/GUS $\Delta$ , in which both TGA1a binding sites are deleted, as a competitor should be taken to exclude the possibilities that -90 35S/GUS $\Delta$  competes for general transcription factors and/or that -90 35S/GUS $\Delta$  competes for nuclear transport of the plasmids.

## DISCUSSION

### TGA1a can stimulate transcription *in vivo* in a sequence-specific manner.

Injection of purified TGA1a into construct A plants induces the GUS reporter gene activity while injection of TGA1a into construct  $\Delta$ A plants and injection of BSA into both transgenic plants have no effect. These results clearly demonstrate that the induction of the GUS reporter gene expression by TGA1a is dependent on the presence of an intact *as-1* element. Considering the previous observation that TGA1a serves as a sequence-specific transcription activator *in vitro* (chapter 4), it is likely that this transcription factor also functions similarly *in vivo*. However, in the case of *in vivo* experiments, as discussed in chapter 4, it is not very clear whether TGA1a acts on the *as-1*-linked promoter directly or indirectly. There are at least two formal possibilities to explain the *in vivo* results if TGA1a acts indirectly: (1) TGA1a displaces an unknown transcription factor X from its binding sites on other promoters. This released factor X is responsible for the induction of the GUS activity. (2) TGA1a somehow stimulates the expression of another transcription factor which in turn acts on the *as-1*-linked promoter. The first possibility is unlikely because a deletion mutant of TGA1a which lacks a 79-amino acid residue region at the N-terminus

( $\Delta$ N80) does not induce GUS expression (chapter 6). It should be noted that  $\Delta$ N80, produced in *E. coli* and purified in the similar way as TGA1a (see chapter 2), is as active as TGA1a in DNA-binding *in vitro*. To test the second possibility, it is necessary to investigate the effects of TGA1a on *as-1*-linked expression in the absence of protein synthesis. This would entail the detection of GUS gene expression by in situ hybridization of GUS mRNA rather than by histochemical analysis of the GUS enzyme activity. If TGA1a acts directly on the *as-1*-linked promoter then expression of the GUS mRNA would not require protein synthesis.

#### Commitment of gene expression in plants

The results on the transactivation of the *as-1*-linked promoter by TGA1a could be relevant to the commitment of gene expression during plant cell differentiation. It is striking that although construct A (-90 35S/GUS) is not expressed in cotyledon cells, it is responsive to an increase in TGA1a concentration within the cells. By contrast, animal genes that are not expressed in certain differentiated cells are often inactivated. The inactivation may involve changes in the chromatin structure (Elgin, 1988) and/or methylation of DNA flanking the genes (Ceder, 1988). Reactivation of the silenced genes usually requires cell division. The observation that the -90 35S/GUS transgene can be activated by injection of TGA1a suggests one of the following possibilities: (1) In plant cells, there is no stringent commitment of gene expression and most genes are ready to respond to trans-acting factors. (2) The -90 35S/GUS transgene has a special function to prevent its own inactivation (this may be related with a low level of TGA1a in leaf cells). (3) The factor TGA1a has the ability to overcome gene inactivation. Because of the totipotency of plant cells, the stringency of cell commitment in plants will be an interesting topic for future

studies.

Changes in transcription activator concentration can generate an all-or-none expression pattern.

The *as-1*-specific factor is present in leaf cells but its concentration is not sufficiently high to occupy the *as-1* site and consequently the *as-1*-linked promoter is inactive in such cells. Three lines of evidence supports the notion:

- 1) While the -90 35S promoter which contains only one copy of *as-1* is inactive in leaf cells, the same promoter can be activated by attaching four additional copies of *as-1* to its 5'-end. One interpretation of this result is to assume that a higher copy number of *as-1* would have a greater probability of binding *as-1*-specific factor, even if the factor is limiting in concentration (Lam and Chua, 1990).
- 2) Similarly increasing the copy number of the -90 35S/GUS chimeric gene should also enhance the probability that at least one of the promoters would have its *as-1* site occupied by the factor. Therefore, although one or a few copies of the -90 35S/GUS transgene is not active in leaf cells, microinjection of 50,000 copies of the same gene can confer expression in the same cells (described in this chapter).
- 3) An up-mutant of *as-1*, that has a higher binding affinity for TGA1a as compared to the wild type, should also have a greater probability of interacting with the endogenous *as-1*-specific factor and therefore should confer expression in leaf cells. This was indeed observed (X.-F.Qin, unpublished).

In the study of the expression of a promoter linked to three NF-AT binding sites (originated from IL-2 promoter) during T-cell activation, it was suggested that there is a threshold level for the NF-AT action: no expression of the promoter under the threshold

level of NF-AT and full expression of the promoter over the threshold level (Fiering *et al.*, 1990). The situation of the *as-1*-linked promoter (-90 35S) might be related although the expression level of the promoter is not quantified. A low concentration of the *as-1*-specific factor in leaf cells is not sufficient to give expression of the promoter, while a high concentration of the factor (ie. microinjection of TGA1a) can give expression. Three copies of NF-AT binding sites are needed for good induction of the promoter with NF-AT binding sites during T-cell activation (Fiering *et al.*, 1990). It is interesting that *as-1* has two TGA1a-binding sites and both binding sites appear to be important for the *as-1* function (P.N.Benfey, personal communication).

What is the significance of a low level of the *as-1*-specific factor in leaf cells?

The *as-1*-specific factor occurs at a low concentration in leaf cells and this concentration is insufficient to activate *as-1*. What is the biological significance of this finding? This concentration of TGA1a is sufficient to activate a promoter linked to multiple copies of *as-1* (Lam and Chua, 1990) and a promoter linked to a up-mutant of *as-1* (X.-F.Qin, unpublished). Therefore, if there were cellular homologs of these promoters, this low level of the factor would be significant for the expression of such promoters. As mentioned in chapter 1, *as-1* has the ability to interact synergistically with other cis-elements. The low level of *as-1*-specific factor may be needed for this interaction. This contention is supported by several lines of evidence: 1) Mutations in *as-1* reduce the activity of the full 35S promoter in leaf (Lam *et al.*, 1989). 2) *as-1* interacts synergistically with various B subdomains of the 35S promoter to generate different expression patterns in leaf tissues (Benfey *et al.*, 1990b) 3) The 3AF1 binding site and *hex-1* are inactive by themselves but confer expression in different plant tissues when linked to *as-1* (Lam *et al.*, 1990b; Lam, unpublished).

Therefore, the low level of the *as-1*-specific factor in leaf cells may be important for interaction with cognate factors of other cis-element (Lam et al., 1989).

#### Advantages, disadvantages, and possible applications of the microinjection technique

As an assay for transcription factors, transient expression systems suffer from two problems: 1) conditions for the reporter gene (chromatin structure and the copy number) are different from those of the cellular genes in the genome. 2) the transformation efficiency varies among different cells. By contrast, the microinjection technique described in this chapter circumvents these two problems by using transgenic plants that carry the reporter gene as a transgene and by performing the assay in single cells with a predetermined amount of injected factors. Also, because transient expression systems require the synthesis of the effector gene product, it is impossible to inhibit protein synthesis to test whether the factor acts on the promoter directly or indirectly, unless the activity of the factor can be induced post-transcriptionally. Although the detection of the GUS reporter gene expression used in the microinjection experiments described here requires protein synthesis, as mentioned above, it is possible to develop an alternative detection system, such as *in situ* hybridization, that is independent of protein synthesis. Thus it would be possible to use the microinjection technique to determine whether the transcription factor acts on the promoter directly or not.

An intrinsic difficulty of the microinjection experiments is the technique itself. The use of intact plants for these experiments is an advantage because plant cells usually de-differentiate upon culture *in vitro*. Microinjection of plant cells that are enclosed in thick cell walls, however, is a technically demanding task and it is very tedious if injection of a large population of cells is needed. Moreover, cells that can be microinjected are limited

to those located close to the surface of the organs. Because the injection procedure causes damage to cells, this technique cannot be applied to studies of wound-inducible promoters.

Although TGA1a purified from *E. coli* was successfully used in this study, other transcription factors may be inactive when produced in bacteria (eg. they may need plant-specific post-translational modifications.). Therefore, TGA1a-encoding RNA synthesized *in vitro* with T7 RNA polymerase was also tested for activity by injection of the RNA into the construct A plants. It was found that injection of TGA1a RNA indeed induce GUS expression, whereas RNA encoding  $\Delta N80$  (TGA1a derivative in which N-terminal 79-amino acid residues are deleted; see chapter 6) is inactive (not shown). The success with RNA injection will open up the possibility to investigate by microinjection transcription factors that may require plant-specific post-translational modifications (although it is impossible, in this case, to test the direct action of the transcription factors on the promoters by inhibition of protein synthesis). Moreover, tedious protein purification steps can be avoided by the use of *in vitro* synthesized RNA.

Because the assay is performed at a single cell level in the microinjection technique, only a small amount of transcription factor is needed. In the experiments described here, 1,000 molecules per cell of TGA1a are sufficient for transactivation. Therefore, it is possible to use this technique to assay the activity of transcription factors purified by biochemical methods from plant materials, even if the amount available is limited by the low yield. In this case, another advantage is that the factor genes do not have to be cloned.

In this study, the reporter gene activity was detected by histochemical staining of GUS activity. Although this method allows the detection of the reporter gene expression in a single cell, the expression level cannot be quantified. This problem can be resolved, if the luciferase gene (LUC; de Wet *et al.*, 1985; Ow *et al.*, 1986) is used as the reporter gene

instead of the GUS gene. The luciferase reaction emits luminescence which can be detected by a microscope equipped with a video camera. When the video camera is connected to an image analyzing system, the expression of luciferase can be quantified at single cell resolution (Gallie *et al.*, 1989).

The microinjection technique in combination with transgenic plants carrying the appropriate reporter transgene is a powerful tool to investigate general problems on plant gene regulation. For example, the effects of post-translational modification on the activities of transcription factor may be examined. Moreover, putative cellular mediators of signal transduction pathways and reagents that affect these pathways can also be evaluated. These investigations can be further refined by introducing the appropriate reporter transgene into well-characterized mutants that are presumably affected in signal transduction. Microinjection of various signal mediators will aid greatly in the ordering of the mutations. Finally, the microinjection technique can be used to ascertain whether a particular molecular event is cell-autonomous or not.

## **Chapter 6: analysis of functional domains of TGA1a**

Studies with the yeast DNA-binding proteins, GAL4 (Keegan *et al.*, 1986; Ma and Ptashne, 1987) and GCN4 (Hope and Struhl, 1986), showed that their DNA-binding activity and transactivation activity reside in discrete and separable regions of the polypeptide chains. These polypeptide domains are portable, ie. their functional properties are retained when they are fused to another protein. As results from other DNA-binding proteins began to accumulate (Mitchell and Tjian, 1989; Johnson and McKnight, 1989), it became clear that such structural features are rather common among many DNA-binding proteins. This observation has encouraged deletion analysis of DNA-binding proteins to define polypeptide regions that mediate specific functions.

The bZIP domain of DNA-binding proteins forms a dimer through the formation of a coiled-coil structure in the "leucine-zipper" region (O'Shea *et al.*, 1989). This dimer formation is thought to be essential for DNA binding (Gentz *et al.*, 1989; Landschulz *et al.*, 1989; Talanian *et al.*, 1990; Turner and Tjian, 1989). Dimerization of bZIP proteins is a specific process, depending on the specific structure of the leucine repeat region, and in many cases, only homodimers are formed. Although heterodimer formation can provide the opportunity to generate DNA-binding proteins of novel specificities and functions, non-specific heterodimer formation can also disrupt selective regulation mediated by these proteins. Therefore, it is of interest to identify the structural features that determine the dimerization specificity of bZIP proteins.

As described in chapter 3, TGA1a has a bZIP domain structure; however, whether this domain in fact functions as a DNA-binding domain is not known. Also, TGA1a contains an acidic region and a glutamine-rich region. Their potential roles in



transcriptional activation need to be examined. Finally, there is the possibility that other regions of the TGA1a protein may modulate its dimerization and/or transactivation activity.

In this chapter, I first describe results demonstrating that the bZIP domain of TGA1a is sufficient for sequence-specific DNA binding. Besides the bZIP domain, I have identified another functional domain in TGA1a, that increases the apparent DNA-binding affinity of the protein by stabilizing the active dimeric form. I designated this domain as the dimer stabilization (DS) domain. Analysis of deletion mutants of TGA1a by microinjection (see chapter 5) suggests that the 79 amino acid region at the N-terminus, including the acidic region, is required for transactivation *in vivo*. On the other hand, analysis in a HeLa cell *in vitro* transcription system (see chapter 4) showed that the same 79 amino acid region at the N-terminus is dispensable for transactivation *in vitro*. Moreover, the *in vitro* analysis showed that only the bZIP domain can give transactivation *in vitro*, albeit at a lower level as compared to TGA1a. The discrepancy between *in vivo* and *in vitro* results for transactivation domains is discussed.

## Results

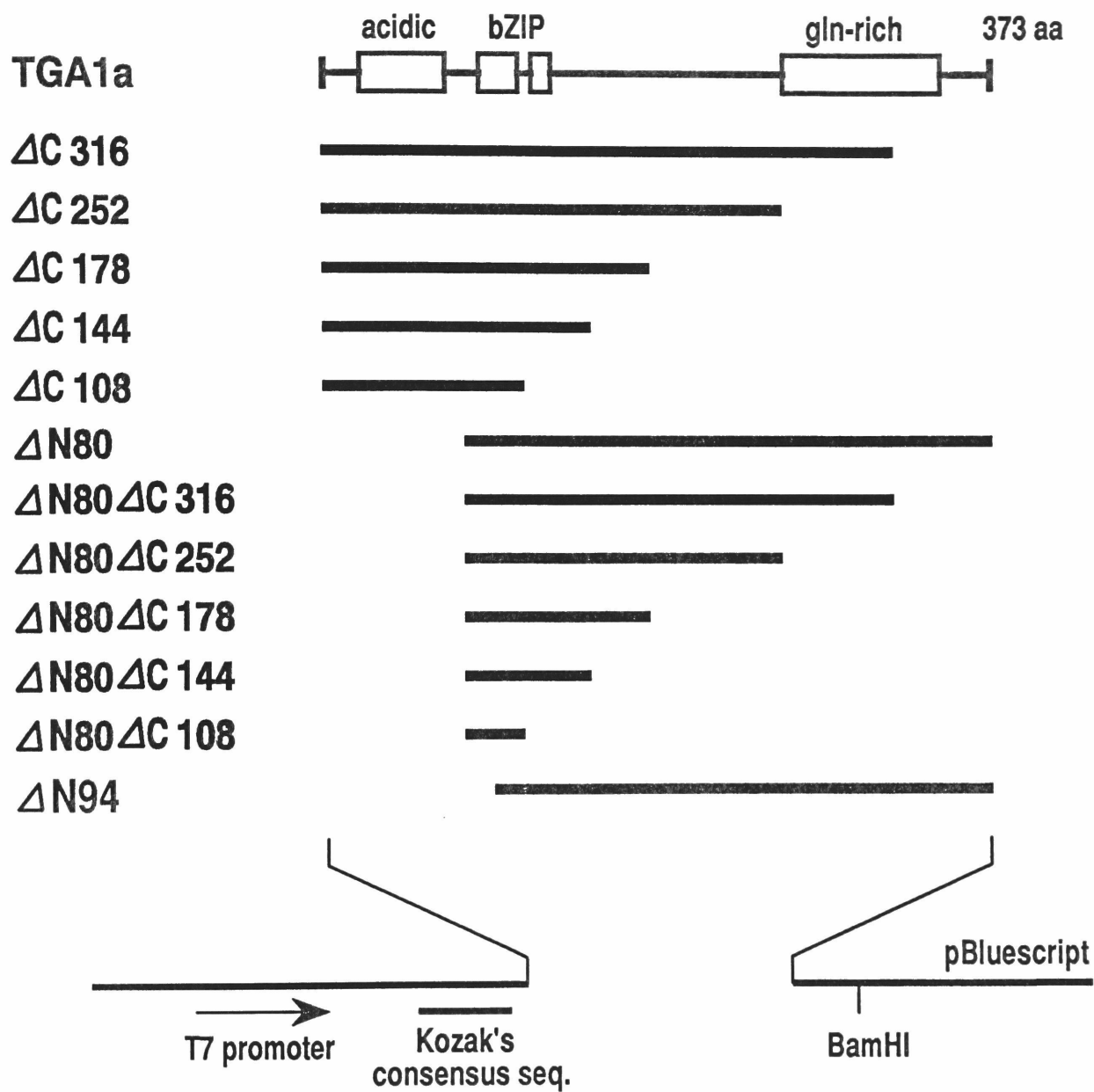
### The bZIP domain is necessary and sufficient for sequence-specific DNA binding.

Various derivatives of the TGA1a cDNA were constructed by deletions (chapter 2). The deletion derivatives (Fig. 6.1) were transcribed *in vitro* with T7 RNA polymerase and the synthesized RNAs were translated in rabbit reticulocyte lysates. DNA-binding activities of the TGA1a deletion derivatives were analyzed by gel retardation assay using *hex-1* (see Fig. 3.1) as a binding probe.

When the full-length TGA1a was analyzed by this method, a prominent DNA-protein complex (Fig. 6.2, lane 2, marked as TGA1a), as well as several minor complexes were

Figure 6.1 Deletion mutants of TGA1a used in this study.

The coding regions the TGA1a derivatives contain are schematically shown by thick lines in respect to the TGA1a coding region (373 amino acid residues) shown at the top. The relative positions of the acidic, bZIP, and glutamine-rich regions of TGA1a are indicated at the top.  $\Delta N$  and  $\Delta C$  represent N-terminal and C-terminal deletions, respectively. The numbers following  $\Delta N$  or  $\Delta C$  represent the positions of the terminal amino acid residues in the derivatives, according to the amino acid numbers shown in Fig. 3.4. The derivatives contain a small number of additional amino acid residues at the deleted termini, due to the plasmid construction (chapter 2). The positions of T7 promoter, which was used for *in vitro* transcription of these derivatives, a consensus sequence of eukaryotic translation initiation sites (Kozak, 1987), and the *Bam*HI site, which was used for the linearization of the plasmids, are schematically indicated. See chapter 2 for details.



obtained. Because the minor complexes were also obtained with the reticulocyte lysate alone (lane 1, marked by asterisks), they likely resulted from interaction of the probe with proteins in the lysate. The prominent complex (marked as TGA1a), therefore, was due to binding of the full-length TGA1a to the *hex-1* probe. The binding is sequence-specific, because the complex formation was abolished by excess wild type *hex-1* but not by the mutant competitor (lanes 3 and 4). Specific complex formation was also obtained with a TGA1a derivative that contains only 65 amino acid residues (amino acid numbers 80-144,  $\Delta N80\Delta C144$ ), including the bZIP domain, although in this case the complex has a faster mobility (Fig. 6.2, lanes 5-7, marked as bZIP). By contrast, a further deletion of 36 amino acids from the C-terminus of  $\Delta N80\Delta C144$  ( $\Delta N80\Delta C108$ , amino acid numbers 80-108) prevented DNA binding (Fig. 6.4b, also see below). Moreover, a derivative lacking the first 91 amino acids, including the basic domain (amino acid numbers 84-108), was also inactive (not shown). Taken together, these results demonstrate that the region between amino acid numbers 80-144 of TGA1a, including the bZIP domain, is necessary and sufficient for sequence-specific DNA binding.

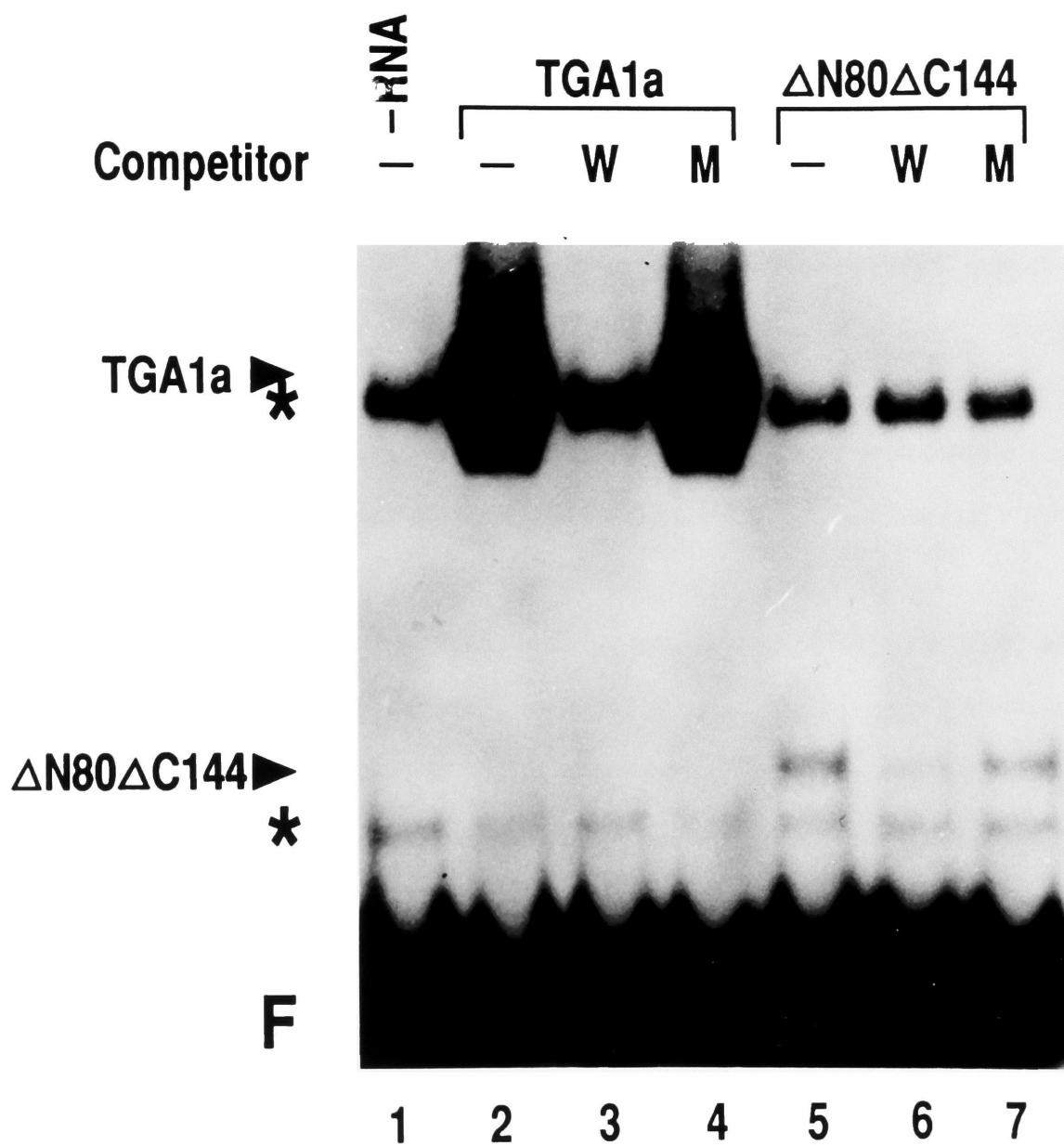
The number of protein molecules added to the binding reaction was almost the same for the full-length TGA1a and the mutant  $\Delta N80\Delta C144$ , as judged from the incorporation of  $^{35}\text{S}$ -labeled methionine into the *in vitro* translated polypeptides (not shown). The amount of the shifted binding probe was, however, much larger with the full-length TGA1a than with the mutant. This decrease in the apparent DNA-binding affinity by deletion is suggestive of a loss of a functional domain that affects DNA binding.

#### TGA1a binds to its target DNA site as a dimer.

All the bZIP proteins investigated thus far bind to their target DNA site as a dimer

Figure 6.2 The bZIP domain is sufficient for sequence-specific DNA binding.

Reticulocyte lysates after translation reaction with no RNA (lane 1), TGA1a RNA (lanes 2-4), and  $\Delta N80\Delta C144$  RNA (lanes 5-7) were analyzed by gel retardation assays. As binding competitors, 2,000-fold molar excess of the wild type *hex-1* oligonucleotide (HW, see chapter 2; lanes 3 and 6) and the mutant *hex-1* oligonucleotide (HM, see chapter 2; lanes 4 and 7) were used. The specific DNA-protein complexes formed with TGA1a and  $\Delta N80\Delta C144$  are indicated. Asterisks, DNA-binding activities originated from reticulocyte lysates; F, free probe.



(Vinson *et al.*, 1989). To see whether TGA1a also binds to DNA as a dimer, derivatives of TGA1a with different lengths were generated and tested for their ability to form heterodimer by gel retardation assays (Hope and Struhl, 1987). Incubation of the full-length TGA1a with the *hex-1* probe produced a major specific DNA-protein complex (Fig. 6.3, lane 1, marked as 1), and a minor specific complex (marked as 1') with a slightly faster mobility. The minor complex 1' is thought to represent the interaction of a shorter derivative of TGA1a with the probe. In fact, analysis of the *in vitro* translated polypeptides by SDS-PAGE showed the presence of a polypeptide species smaller in size than the full-length TGA1a (not shown). This smaller derivative is presumably a product of translation initiation from an internal methionine codon. It is unlikely to be a premature termination product of polypeptide synthesis, because translation of a RNA for  $\Delta$ N80 produced only a single polypeptide in the same *in vitro* system (not shown).

When *in vitro* synthesized RNAs encoding the full-length TGA1a and  $\Delta$ N80 were cotranslated in reticulocyte lysates and then assayed by gel retardation, a new DNA-protein complex intermediate in mobility between complex 1 and 2 (Fig. 6.3, lane cotrl 1+2, marked as 1+2) was obtained. This new complex likely corresponds to the binding of TGA1a and  $\Delta$ N80 heterodimer to the probe. Complexes 1 and 2 are formed by homodimers of the full-length TGA1a and the mutant  $\Delta$ N80, respectively (lanes 1 and 2). The presence of the band 1+2, presumably corresponding to the heterodimer formation between the full-length TGA1a and  $\Delta$ N80, provides strong evidence that TGA1a binds to its target DNA as a dimer.

An N-terminal region has a negative effect on DNA binding in the absence of a C-terminal region.

The effect of C-terminal deletion on the DNA-binding activity of TGA1a was

Figure 6.3 Heterodimer formation between TGA1a derivatives.

Heterodimer formation either between TGA1a (referred to as 1) and  $\Delta$ N80 (2) (left panel) or between  $\Delta$ N80 $\Delta$ C178 (3) and  $\Delta$ N80 $\Delta$ C144 (4) (right panel) were examined. Lane 1, TGA1a; lane 2,  $\Delta$ N80; lane ctrl 1+2, cotranslation of TGA1a and  $\Delta$ N80; lane mix 1+2, mix of reticulocyte lysates for TGA1a and  $\Delta$ N80 after separate translation; lane 3,  $\Delta$ N80 $\Delta$ C178; lane 4,  $\Delta$ N80 $\Delta$ C144; lane ctrl 3+4, cotranslation of  $\Delta$ N80 $\Delta$ C178 and  $\Delta$ N80 $\Delta$ C144; lane mix 3+4, mix of reticulocyte lysates for  $\Delta$ N80 $\Delta$ C178 and  $\Delta$ N80 $\Delta$ C144 after separate translation. The positions of DNA-protein complexes corresponding to homodimers of these derivatives are indicated by the reference numbers. Band 1' corresponds to a shorter derivatives contaminated in the reaction for TGA1a (see text). Heterodimers between TGA1a and  $\Delta$ N80 and between  $\Delta$ N80 $\Delta$ C178 and  $\Delta$ N80 $\Delta$ C144 are indicated as 1+2 and 3+4, respectively. Note the exposure time of the right panel is about 50 times longer than that of the left panel. Asterisks, DNA-binding activities originated from reticulocyte lysates; F, free probe.

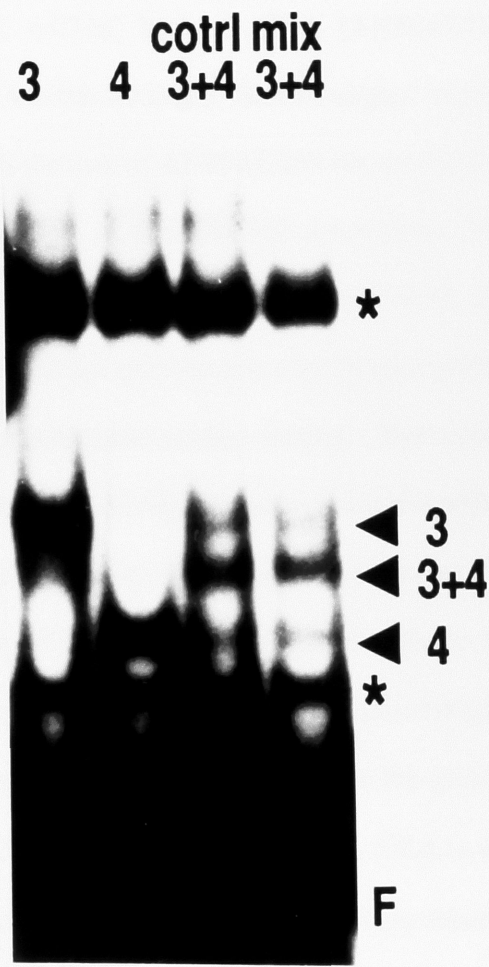
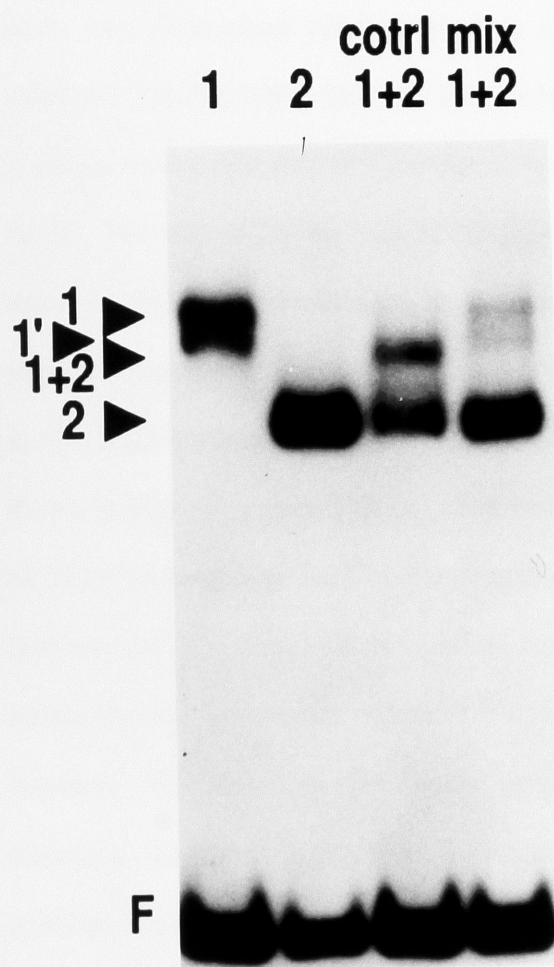


1. TGA1a

2.  $\Delta N80$

3.  $\Delta N80\Delta C178$

4.  $\Delta N80\Delta C144$



investigated. A mutant truncated at amino acid 316 ( $\Delta C316$ ; the region C-terminal to amino acid number 316 is deleted) was able to bind to DNA but further C-terminal deletion ( $\Delta C252$ , 178, 144, 108) abolished the DNA-binding activity (Fig. 6.4a). Two of the C-terminal deletion mutants ( $\Delta C178$ , 144) that were inactive showed a weak but reproducible activity when the N-terminal 79 amino acid residue region was deleted from them ( $\Delta N80\Delta C178$ ,  $\Delta N80\Delta C144$ ; Fig. 6.4b). Removal of the same N-terminal region from the other two C-terminal deletion mutants ( $\Delta C252$ ,  $\Delta C108$ ) had no effect ( $\Delta N80\Delta C252$ ,  $\Delta N80\Delta C108$ ; Fig. 6.4b). Because  $\Delta C108$  does not contain an intact "leucine-zipper" region, it seems reasonable that this mutant as well as its derivative  $\Delta N80\Delta C108$  do not bind to DNA. The reason for the lack of DNA-binding activity of  $\Delta N80\Delta C252$  is not clear. The mutant polypeptide synthesized has the expected size when analyzed by SDS-PAGE (not shown). Moreover, sequence analysis of the mutant did not uncover any accidental mutation in the bZIP coding region, that might have been generated during cloning. The results shown in Fig. 6.4 suggest that the N-terminal 79 amino acid region exerts a negative effect on DNA binding when the C-terminal region is deleted further than amino acid number 316. Because the effect of this N-terminal region on DNA binding is observed only in the absence of the C-terminal region, it is not clear whether this region has any physiological function. This negative effect could simply be explained by a constraint in the protein structure. However, it may have an effect on DNA binding affinity, because TGA1a and  $\Delta N80$  produced in *E. coli* behave differently upon chromatography on DNA-affinity columns containing either *hex-1* or its mutant. TGA1a binds fairly well to the wild type DNA column in a buffer with 0.4M KCl and also to the mutant DNA column in a buffer with 0.2 M KCl. By contrast,  $\Delta N80$  does not bind well to either of the columns under the same conditions.

Figure 6.4 DNA-binding assays of TGA1a deletion mutants.

**a.** C-terminal deletion mutants with the intact N-termini. **b.** C-terminal deletion mutants with the 79-amino acid residue deletion at the N-termini. See Fig. 6.1 for the deletion mutants. Asterisks, DNA-binding activities originated from reticulocyte lysates; F, free probe.

**a**



$\Delta$ C108

$\Delta$ C144

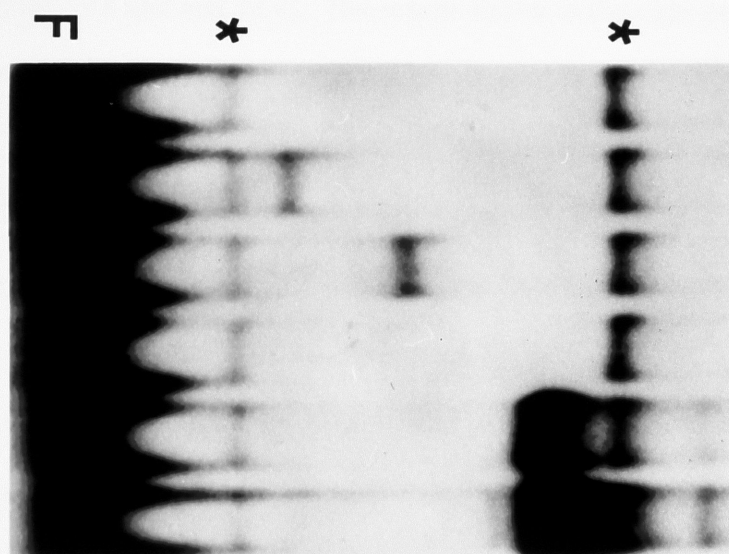
$\Delta$ C178

$\Delta$ C252

$\Delta$ C316

TGA1a

**b**



$\Delta$ N80 $\Delta$ C108

$\Delta$ N80 $\Delta$ C144

$\Delta$ N80 $\Delta$ C178

$\Delta$ N80 $\Delta$ C252

$\Delta$ N80 $\Delta$ C316

$\Delta$ N80

### The C-terminal region stabilizes the dimeric form.

When the full-length TGA1a and  $\Delta N80$  were separately translated and then mixed together and kept at room temperature for 30 min before the binding assay, no heterodimer formation was detected (Fig. 6.3, lane mix 1+2), while cotranslation of the two clearly resulted in heterodimer formation (lane ctrl 1+2). This result suggests that once formed the dimer is so stable such that no subunit exchange is detected within 30 min. However, when the same experiments were performed with  $\Delta N80\Delta C178$  and  $\Delta N80\Delta C144$ , cotranslation of the RNAs for these two derivatives (lane ctrl 3+4) as well as mixing of the separately translated products (lane mix 3+4) both resulted in heterodimer formation (marked as 3+4). In this case, exchange of the subunits was completed within 30 min, because the ratios of the band intensities corresponding to  $\Delta N80\Delta C178$  homodimer (marked as 3), the  $\Delta N80\Delta C144$  homodimer (marked as 4), and the heterodimer between the two (marked as 3+4) are very similar under the two different experimental conditions (Fig. 6.3, lanes ctrl 3+4 and mix 3+4). The results shown in the right panel of Fig. 6.3 indicate that the two C-terminal deletion mutants ( $\Delta N80\Delta C178$  and  $\Delta N80\Delta C144$ ) bind to DNA as a dimer, but the dimeric form is unstable. Alternatively these mutants may not be able to form dimers without the target DNA. Therefore, the bZIP domain itself is not sufficient for the formation of a stable dimer, and the C-terminal region (amino acid numbers 178-316) is needed for stabilization. Hereafter, I will refer to this C-terminal region as the DS domain (dimer stabilization domain) and use the term "stable dimers" to refer to those dimers that do not show any subunit exchange within 30 min under the conditions used.

TGA1a mutants without the DS domain show apparent low DNA-binding affinity (Figs. 6.2 and 6.4b). This phenomenon can be explained by the absence of the DS domain. Mutants without the DS domain (-DS derivatives) exist mainly as monomers, and dimers

that are active in DNA binding are not abundant in solution. On the other hand, mutants with the DS domain (+DS derivatives) exist mainly as active dimers that are capable of binding DNA. Consequently, the -DS derivatives show apparent low DNA-binding affinity as compared to +DS derivatives.

Can the DS domain function when only one of the two subunits contains the domain or can it function only when both subunits have it? If the heterodimer between a +DS and a -DS derivative is as stable as the homodimer of the +DS derivative, heterodimers would be detected by cotranslation of the RNAs encoding these two derivatives. If the molecule number of the -DS derivative is the same or greater than that of the +DS derivative, heterodimer formation would be favored over formation of +DS homodimers. On the other hand, if stable dimer formation requires that both subunits contain the DS domain no heterodimer would be detected upon cotranslation of RNAs for a +DS and a -DS derivative. Immediately after translation, the +DS derivative would form stable dimers which virtually do not dissociate. This rapid dimer formation would drastically decrease the monomeric form in the translation mix. Even if a heterodimer is accidentally formed, it would dissociate within a short time. Consequently, the -DS derivative can only form homodimers. The results in Table 6.1 support the latter possibility. Heterodimer formation between +DS and -DS derivatives (combinations of 1&3, 1&4, 2&3, and 2&4) was not seen, while heterodimer formation between two +DS derivatives (1&2) and between two -DS derivatives (3&4) was detected. It should be pointed out that the experiment to detect possible heterodimerization between +DS and -DS derivatives was carried out under excess concentration of the -DS derivative (confirmed by SDS-PAGE, not shown). Even under these conditions, the formation of heterodimers between +DS and -DS derivatives was not observed. Therefore, both subunits must have the DS domain to form a stable dimer.

**Table 6.1 Heterodimer formation.**

1. TGA1a 2.  $\Delta$ N80 3.  $\Delta$ N80 $\Delta$ C178  
4.  $\Delta$ N80 $\Delta$ C144

	1	2	3	4
1	(+)	+	-	-
2		(+)	-	-
3			(+)	+
4				(+)

+, heterodimer formation

-, no heterodimer formation

(+), supposed to be homodimer

An N-terminal region is necessary for transactivation activity *in vivo*.

To assess the transactivation activities of TGA1a mutants, the full-length TGA1a,  $\Delta$ N80, and  $\Delta$ N80 $\Delta$ C144 were produced in *E. coli* and purified as described in chapter 2. The purified proteins were microinjected into epidermal cells ( $\sim 10^4$  mol/cell) of cotyledons of transgenic tobacco plants harboring construct A (-90 35S/GUS) as a reporter gene (see chapter 5). Transactivation activity was assayed by the induction of GUS activity in the injected cells. While the full-length TGA1a was able to transactivate the *as-1*-linked promoter, both  $\Delta$ N80 and  $\Delta$ N80 $\Delta$ C144 were inactive even at 10 times higher concentration. These results suggest that the N-terminal 79 amino acid residue region (including the acidic region) that is missing in  $\Delta$ N80 is necessary for transactivation.

A small region including the bZIP domain is sufficient for transactivation *in vitro*.

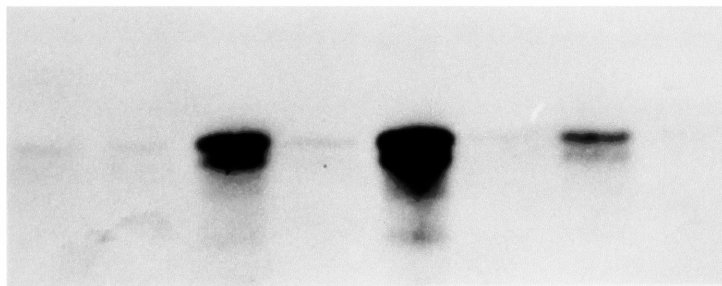
The transactivation activities of TGA1a derivatives were also assayed using a HeLa cell *in vitro* transcription system (see chapter 4). Purified full-length TGA1a,  $\Delta$ N80, and  $\Delta$ N80 $\Delta$ C144 were the same preparations as the ones used for microinjection described in the last section. Figure 6.5 shows that all three TGA1a derivatives can stimulate transcription in this *in vitro* system in a sequence-specific manner. The addition ( $10^{-12}$  mol/assay, equivalent to  $\sim 50$  ng of TGA1a/assay) of either TGA1a,  $\Delta$ N80, or  $\Delta$ N80 $\Delta$ C144 stimulated the accumulation of the specific transcript from the wild type *as-1* promoter (W) but not the mutant promoter (M). It is surprising that  $\Delta$ N80, which was inactive *in vivo*, stimulated transcription *in vitro* at a two times higher level than TGA1a. Moreover,  $\Delta$ N80 $\Delta$ C144, which has only 65 amino acid residues, can stimulate transcription although at a three times lower level when compared to TGA1a. Based on these results, the N-terminal region (amino acid numbers 1-79) is dispensable for transactivation *in vitro* and a



Figure 6.5 *In vitro* transactivation of TGA1a deletion mutants.

Effects of TGA1a,  $\Delta$ N80, and  $\Delta$ N80 $\Delta$ C144 on transcription were analyzed in a HeLa cell *in vitro* system. W, promoter with the wild type *as-1*; M, promoter with the mutant form of *as-1*. Arrowhead indicates the position of the specific transcript.

<b>Added Factor</b>	<b>—</b>		<b>TGA1a</b>		<b>ΔN80</b>		<b>ΔN80ΔC144</b>	
<b>Template</b>	<b>W M</b>		<b>W M</b>		<b>W M</b>		<b>W M</b>	



transactivation domain appears to be located very close to or overlap with the bZIP region.

## Discussion

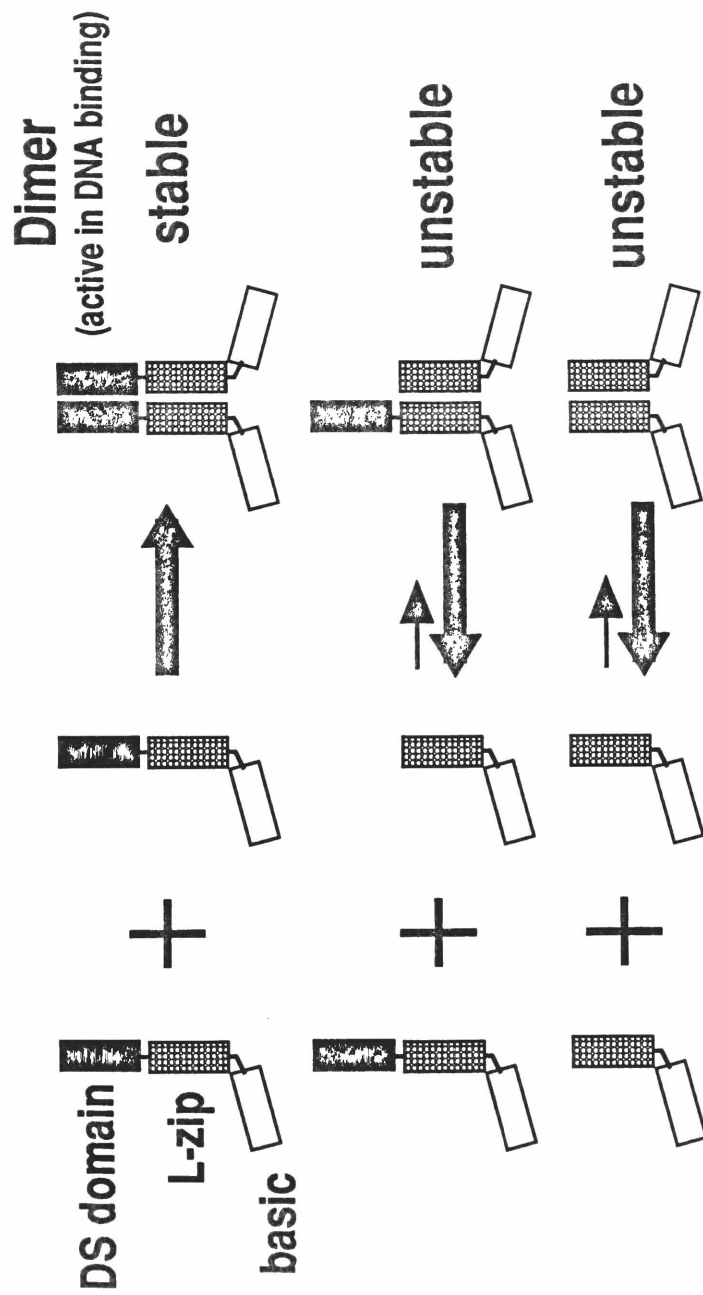
In this chapter, I describe functional mapping of various regions of TGA1a by deletion analyses. I show that the bZIP domain of TGA1a is sufficient for sequence-specific DNA binding, but its apparent DNA-binding affinity is much lower than that of the full-length TGA1a. In the C-terminal region, TGA1a contains a dimer stabilization (DS) domain, that is responsible for the higher apparent DNA-binding affinity of the +DS derivatives. *in vitro* transcription analysis using the microinjection technique suggests that the N-terminal region (amino acid numbers 1-79) is essential for transactivation. On the other hand, results from the HeLa cell *in vitro* transcription system suggest that this N-terminal region is not required for transactivation *in vitro*. The *in vitro* experiment also suggests that a 65-amino acid residue region (amino acid numbers 80-144) that includes the bZIP domain is sufficient for transactivation at a reduced level.

### How does the DS domain stabilize the dimeric form?

Related to the function of the C-terminal region (amino acid numbers 178-316), three phenomena were observed: 1) Deletion of this region decreases the apparent DNA-binding affinity. 2) Deletion of this region causes a frequent exchange of subunits. 3) Derivatives with this region cannot form heterodimers with those lacking this region. The presumed function of the dimer stabilization (DS) domain, summarized in Fig. 6.6, can explain these three phenomena. The decrease in the apparent DNA-binding affinity of the -DS derivatives can be attributed to the decrease of the population of its dimeric form which is the species active in DNA binding (compare the top and bottom panels in Fig. 6.6). The

Figure 6.6 A model of the dimer stabilization (DS) domain.

Dimer formation of TGA1a derivatives with and without the DS domain is schematically shown. The basic and "leucine-zipper (L-zip)" regions of the bZIP domain are also shown.



higher rate of subunit exchange of the -DS derivatives can be explained by the instability of the dimeric form lacking the DS domain (bottom panel). The absence of detectable heterodimer formation between +DS and -DS derivatives is due to the rapid depletion of the monomers of the +DS derivatives by the stable homodimer formation of the +DS derivatives (top panel) as the heterodimer between +DS and -DS derivatives is unstable (middle panel). However, the results described in this chapter are still preliminary in two aspects. The deletion analysis is not detailed enough to allow the conclusion that all the three phenomena are mediated by a single functional domain. Furthermore, because DNA-binding activity was used to assay indirectly heterodimer formation, the assay cannot distinguish between whether a heterodimer is not formed and whether a heterodimer is formed but unable to bind DNA.

More detailed analyses, including deletion and site-specific mutations, will reveal whether these three phenomena are mediated by a single functional domain, how large a region is required for this function, and whether the domain can function in a position-independent manner. With regards to the last point, it would be interesting to test whether this domain can function in the context of other unstable dimer-forming proteins. For example, C/EBP, which is a bZIP protein, exhibits rapid exchanges of its subunits in solution. It would be important to know whether fusion of the DS domain to C/EBP would stabilize dimer formation and prevent subunit exchange. More generally, it could be tested whether the DS domain has any dimer stabilization activity, when fused to other unstable dimer-forming DNA-binding proteins (eg. some of the HLH proteins; Murre *et al.*, 1989) or even simply proteins that form unstable dimers. That the DS domain must be present in both subunits for the dimer stabilization activity (Fig. 6.5) suggests that interaction between the two DS domains is crucial for this activity. If the DS domain functions in the

context of other unstable dimer-forming proteins, it would confirm the notion that the interaction between two DS domains is intrinsic for dimer stabilization.

A direct dimer-formation assay, such as a protein cross-linking assay, is required to distinguish whether a heterodimer is not formed or whether it is formed but unable to bind DNA. Such a direct assay can also address the question whether the dimer population of a -DS derivative is in fact much smaller than that of a +DS derivative when they are both present at the same concentration. It would also be interesting to investigate conditions (temperature, salt concentration, etc.) that can modulate dimer formation of -DS derivatives. Because dimerization of the bZIP domain is mediated by hydrophobic interaction in the "leucine-zipper" (coiled-coil; Crick, 1952; Landschulz *et al.*, 1988), decreasing temperature and increasing salt concentration are expected to facilitate dimer formation of these derivatives. Moreover, the direct assay can be used to investigate whether -DS derivatives form any significant amount of dimers without DNA and whether addition of the target DNA sequence would facilitate dimerization. Finally, the possibility that the DS domain itself may have dimerization activity can also be tested by the assay.

#### Dimer stabilization domains may be common among DNA-binding proteins.

TGA1a has two structural domains for dimerization, the bZIP domain and the DS domain. So far, this type of structural features has not been reported for other bZIP proteins. The "leucine-zipper" region of GCN4 alone can form stable dimers (Hope and Struhl, 1987). C/EBP cannot form stable dimers (Landschulz *et al.*, 1989). Two HLH proteins, AP-4 (Hu *et al.*, 1990) and USF (Gregor *et al.*, 1990), have been reported to possess domains that are functionally similar to the DS domain of TGA1a. The HLH domain is similar to the bZIP domain in terms of its dimerization and DNA-binding activity.

Both protein domains possess dimerization activity that is required for DNA binding. Both domains by themselves are sufficient for sequence-specific DNA binding. AP-4 contains two leucine-repeats ("leucine-zipper") located at the C-terminus of the HLH domain. AP-4 derivatives without the leucine-repeat regions cannot form dimers in solution when analyzed by a protein cross-linking assay, whereas derivatives with these regions can. AP-4 derivatives without the leucine-repeat regions can form heterodimers with another HLH protein E47, but those containing the regions cannot. USF also has a "leucine-zipper," which is important for efficient DNA binding. No heterodimer formation is detected between USF derivatives with and without the "leucine-zipper." The observation that both AP-4 and USF have domains functionally equivalent to the DS domain suggests that such domains may be quite common among DNA-binding proteins.

#### Advantage of having a dimer stabilization domain separated from a DNA-binding domain

If DS domains (in this section, domains functionally equivalent to the DS domain of TGA1a are also called the DS domains) are common among DNA-binding proteins, there should be some advantages to having a combination of a weak dimerization domain and a DS domain. In the following, I speculate on the possible advantages of DS domains during evolution of DNA-binding proteins. When an organism needs new independent pathways for more complicated regulatory systems, additional regulatory proteins, such as DNA-binding proteins, of different properties are required. How can an organism accumulate genes encoding different DNA-binding proteins rapidly during evolution? The first event for this purpose presumably involves duplication of genes encoding DNA-binding proteins. One of the duplicated genes may then accumulate mutations to produce a new regulatory protein gene. However, in the case of DNA-binding proteins that require dimerization for



activity, such as the bZIP and HLH proteins, mutations in one of the duplicated genes would not lead to the creation of a new independent DNA-binding protein, because the duplicated gene product would interfere with the parental product by heterodimer formation. This problem could be avoided if the duplicated gene acquires specific mutations that prevent heterodimer formation without affecting homodimer formation.

The combinatorial structure for dimerization consists of two separate domains, the DNA-binding domain (bZIP or HLH domain) and the DS domain, and together they determine dimerization specificity. As mentioned above, a bZIP protein can form heterodimers with a limited variety of other bZIP proteins. The results shown in Fig. 6.5 also suggest that proteins with the same bZIP domain but different DS domains (different in dimerization specificity) cannot form heterodimers. Thus, if two duplicated genes with different DS domains exchange their coding sequences for such domains, the products of such shuffled genes would not be able to form heterodimers. By this method, an organism might be able to generate a large variety of DNA-binding proteins that can function independently from one another, within a relatively short period during evolution. In this respect, it is interesting to note that the bZIP domain of TGA1a is encoded by exons separated from the other parts of the gene (H.Fromm, F.Katagiri, and N.-H.Chua, in preparation). Such a genomic organization would certainly facilitate the exchange of DNA-binding domains and DS domains by alternative splicing or exon shuffling (Gilbert, 1978).

#### Is the N-terminal region needed for transactivation?

Contradictory results were obtained between *in vivo* and *in vitro* experiments on the localization of transactivation domains of TGA1a. Results from the microinjection experiments (*in vivo*) showed that the N-terminal region (amino acid numbers 1-79) is

necessary for transactivation. The results obtained from the HeLa cell *in vitro* transcription experiments showed not only that this N-terminal region is dispensable but also that the 65-amino acid residue region (amino acid numbers 80-144) including the bZIP region is sufficient for transactivation although the activation level was three times lower than that of TGA1a. There are two possible interpretations for this discrepancy. 1) In addition to the transactivation activity, TGA1a may require other activities (eg. nuclear localization, Burke, 1990; chromatin interaction, Chasman *et al.*, 1990; etc.) to successfully activate its target genes *in vivo*. In this respect, it is important to point out that both mutants  $\Delta$ N80 and  $\Delta$ N80 $\Delta$ C144 contain a nuclear localization sequence (A. R. van der Krol and N.-H. Chua, submitted); nevertheless, it is unable to transactivate the *as-1*-linked promoter. It is also worth pointing out that the bZIP domain alone could have the ability to interact with other proteins. The bZIP domain of GCN4 interacts with RNA polymerase II (Brandl and Struhl, 1989). The "leucine-zipper" of c-Jun interacts with glucocorticoid receptor (Schule *et al.*, 1990). 2) The transactivation activity observed in the *in vitro* system is irrelevant to the *in vivo* function of TGA1a. The N-terminal region is simply a transactivation domain, as expected from the fact that this region is enriched in acidic amino acid residues. However, gain-of-function experiments are needed to provide conclusive evidence for the role of this domain. For example, the *in vivo* analysis of chimeric proteins in which the TGA1a N-terminal acidic region is fused to other DNA-binding domain would help to ascertain the function of this domain. Similarly, analysis of TGA1a derivatives in which the N-terminal region is replaced by well characterized transactivation domains (eg. GAL4 transactivation domain) would determine whether the function of the N-terminal region can be replaced by other transactivation domains. These analyses can also address the question whether the N-terminal region of TGA1a is simply a transactivation domain or has other unknown

functions.

## Chapter 7: summary and future studies

In this thesis, I have described the characterization of a tobacco DNA-binding protein, TGA1a, that binds to the *as-1* element of the CaMV 35S promoter. TGA1a is likely to be responsible for the function of *as-1*, ie. conferring root expression on a promoter, because of the following lines of evidence:

- 1) Mutations in *as-1* that abolish the function of the element also block TGA1a binding.
- 2) TGA1a mRNA is far more abundant in root than in leaf. This is correlated with the observation that the factor mediating *as-1* function is present in leaf, albeit at a limiting concentration.
- 3) TGA1a can function as an *as-1*-specific transcription activator in a HeLa cell *in vitro* system as well as a wheat germ *in vitro* system. Because *as-1* is a positive regulatory element, the factor mediating its function is expected to be a positive regulatory factor. Therefore, the observation that TGA1a is a transcription activator is correlated with this expectation.
- 4) When microinjected into cotyledon cells of transgenic tobacco plants, TGA1a can induce the expression of a transgene promoter linked to *as-1*. Therefore, TGA1a probably functions as an *as-1*-specific transcription activator *in vivo*.

Although all these observations are in support of a direct relationship between TGA1a and *as-1* function, other possibilities should be considered. For example, there may be more than one *as-1*-specific transcription activators. In the case of maize, the *Lc* gene (a member of *R* genes), which positively regulates anthocyanin biosynthesis in the midrib and other parts of the plant body but not in aleurone cells, can, nevertheless, induce anthocyanin biosynthesis in aleurone cells when it is constitutively expressed (Ludwig *et al.*, 1990). The

TGA1a activity analyzed in leaf may not necessarily represent the physiological function of TGA1a *in vivo*. Without genetic evidence (including gene disruption), it is very difficult to conclude that TGA1a is indeed responsible for the *as-1* function.

The situation is further complicated by the fact that TGA1a is encoded by a multigene family. At present it is not known whether members of the family are functionally interchangeable. If so, genetic studies of TGA1a would be very difficult. It would be important to ascertain if there are any differences among family members in function and expression profile. The possibility of heterodimer formation among the members should also be investigated. If so, it would be of interest whether the heterodimers are functionally distinct from the homodimers. In the case of the Jun family in mammals, c-Jun is a positive regulator while Jun-B is a negative regulator (Chiu *et al.*, 1989; Schutte *et al.*, 1989). Furthermore, the heterodimers of c-Jun and c-Fos bind to DNA stronger than c-Jun homodimers (Kouzarides and Ziff, 1989; Rausher *et al.*, 1988; Turner and Tjian, 1989).

TGA1a itself appears to be regulated at a transcription level (although other possibilities, such as regulation at a mRNA degradation level, are not excluded). Its mRNA is far more abundant in root than in leaf. It would be of interest to know what regulates TGA1a expression. Other factors could determine the expression pattern of TGA1a, or alternatively, TGA1a could be autoregulated. Consistent with the latter idea, it is interesting to note that the TGA1a promoter region has its own binding site (H.Fromm, F.Katagiri, and N.-H.Chua, in preparation). It is possible that TGA1a may require post-transcriptional modification for activity and this possibility may be investigated using monospecific antibodies. Cloning of TGA1a gene made it possible to raise such antibodies without protein purification.

TGA1a stimulates *in vitro* transcription by increasing the number of preinitiation

complexes. If this is also the case *in vivo*, I have proposed two models to explain how TGA1a level can regulate expression of an *as-1*-linked promoter *in vivo*. The first model assumes that an increasing concentration of TGA1a increases the probability that the promoter is expressed in a single cell. This model can be tested because an increasing amount of TGA1a increases the number of cells expressing the promoter. The second model assumes that the preinitiation complexes are unstable *in vivo*. When the assembly and disassembly of preinitiation complexes on promoter sites are at equilibrium, the average time when a preinitiation complex is formed on a given promoter (ie. when the promoter is active) increases with an increasing concentration of TGA1a. As a consequence, the amount of transcription in each cell is regulated in a factor-dose dependent manner over a relatively long time. This model is difficult to test because so far it has not been possible to assay for preinitiation complex formation *in vivo*. If the amount of transcription were measured in a single cell within a short time, this model could be tested indirectly. Under this condition, the results should be similar to those predicted by the first model, ie. the number of cells expressing the promoter increases with an increasing concentration of TGA1a. In this case, however, it could be impossible to rule out the possibility that TGA1a increases the frequency of transcription initiation *in vivo*.

The bZIP domain of TGA1a is sufficient for sequence-specific DNA binding. An additional domain (DS domain) appears to affect the apparent DNA-binding affinity by stabilizing the active dimeric species. To ascertain the function of the DS domain, a direct assay for dimer formation should be established. Domains functionally equivalent to the DS domain could constitute a new family of functional domain among DNA-binding proteins. In AP-4 (Hu *et al.*, 1990) and USF (Gregor *et al.*, 1990), a "leucine-zipper" structure seems to be involved in dimer-stabilizing function. A more refined analysis of the DS domain of

TGA1a might reveal its polypeptide structure. I have proposed that new combinations of dimerization domains (eg. bZIP and HLH) and DS-related domains can generate new selectivity in the formation of dimers. This proposal could be tested by constructing artificial fusion proteins comprising these domains.

The polypeptide segment N-terminal to the bZIP region, including the acidic region, is necessary for transactivation by TGA1a when assayed *in vivo*. However, when assayed in a HeLa cell *in vitro* system, the majority of TGA1a, except a 65-amino acid residue region including the bZIP region, is dispensable for transactivation. This discrepancy between *in vivo* and *in vitro* results casts doubt on the fidelity of the *in vitro* system. Alternatively, the discrepancy could be due to a requirement of some additional functions of TGA1a for the *in vivo* activity (eg. nuclear localization, making the promoter region available). This question can be addressed by a functional comparison of the N-terminal region with other well-characterized transactivation domains.

## References

- Abmayr, S.M., J.L. Workman, and R.G. Roeder. 1988. The pseudorabies immediate early protein stimulates *in vitro* transcription by facilitating TFIID:promoter interactions *Genes Dev.* **2**: 542-553.
- Agre, P., P.F. Johnson, and S.L. McKnight. 1989. Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. *Science* **246**: 922-926.
- An, G., P.R. Ebert, B.-Y. Yi, and C.-H. Choi. 1986. Both TATA box and upstream regions are required for the nopaline synthase promoter activity in transformed tobacco cells. *Mol. Gen. Genet.* **203**: 245-250.
- Anderson, W.F., D.H. Ohlendorf, Y. Takeda, and B.W. Matthews. 1981. Structure of the cro repressor from bacteriophage  $\lambda$  and its interaction with DNA. *Nature* **290**: 754-758.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl (Eds.). 1987. Current protocols in molecular biology. vols. 1-2, John Wiley and Sons, New York.
- Benfey, P.N. and N.-H. Chua. 1989. Regulated genes in transgenic plants. *Science* **244**: 174-181.
- . 1990. The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* **250**: 959-966.
- Benfey, P.N., L. Ren, and N.-H. Chua. 1989. The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J.* **8**: 2195-2202.
- . 1990a. Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* **9**: 1677-1684.



----. 1990b. Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. *EMBO J.* **9**: 1685-1696.

Biedenkapp,H., U.Borgmeyer, A.E.Sippel, and K.-H.Klempnauer. 1988. Viral *myb* oncogene encodes a sequence-specific DNA-binding activity. *Nature* **335**: 835-837.

Bohmann,D., T.J.Bos, A.Admon, T.Nishimura, P.K.Vogt, and R.Tjian. 1987. Human proto-oncogene *c-jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science* **238**: 1386-1392.

Bouchez,D., J.G.Tokuhisa, D.J.Llewellyn, E.S.Dennis, and J.G.Ellis. 1989. The ocs-element is a component of the promoters of several T-DNA and plant viral genes. *EMBO J.* **8**: 4197-4204.

Boutry,M. and N.-H.Chua. 1985. A nuclear gene encoding the beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia*. *EMBO J.* **4**: 2159-2165.

Brandl,C. and K.Struhl. 1989. Yeast GCN4 transcriptional activator protein interacts with RNA polymerase II *in vitro*. *Proc. Natl. Acad. Sci. USA* **86**: 2652-2656.

Brown,D.D. and J.G.Gurdon. 1978. Cloned single repeating units of 5S DNA direct accurate transcription of 5S RNA when injected into *Xenopus* oocytes. *Proc. Natl. Acad. Sci. USA* **6**: 2849-2853.

Burke,B. 1990. The nuclear envelope and nuclear transport. *Curr. Opinion Cell Biol.* **2**: 514-520.

Cai,H., D.S.Luse. 1987. Transcription initiation by RNA polymerase II *in vitro*. *J. Biol. Chem.* **262**: 298-304.

Carcamo,J., S.Lobos, A.Merino, L.Buckbinder, R.Weinmann, V.Natarajan, and D.Reinberg. 1989. Factors involved in specific transcription by mammalian RNA polymerase II. *J. Biol.*

*Chem.* **264**: 7704-7714.

Ceder,H. 1988. DNA methylation and gene activity. *Cell* **53**: 3-4.

Chang,C., J.L.Bowman, A.W.DeJohn, E.S.Lander, and E.M.Meyerowitz. 1988. Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **85**: 6856-6860.

Chasman,D.I., N.F.Lue, A.R.Buchman, J.W.LaPointe, Y.Lorch, and R.D.Kornberg. 1990. A yeast protein that influences the chromatin structure of UAS<sub>G</sub> and functions as a powerful auxiliary gene activator. *Genes Dev.* **4**: 503-514.

Chiu,R., P.Angel, and M.Karin. 1989. Jun-B differs in its biological properties from, and is a negative regulator of, c-Jun. *Cell* **59**: 979-986.

Chodosh,L.A., A.S.Baldwin, R.W.Carthew, and P.A.Sharp. 1988a. Human CCAAT-binding proteins have heterologous subunits. *Cell* **53**, 11-24.

Chodosh,L.A., J.Olesen, S.Hahn, A.S.Baldwin, L.Guarente, and P.A.Sharp. 1988b. A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable. *Cell* **53**: 25-35.

Coe,E.H.,Jr., M.G.Neuffer, and D.A.Hoisington. 1988. The genetics of corn. in: *Corn and Corn Improvement*, Agronomy Monograph No.18 3rd ed.: pp. 81-258.

Coen,E.S., J.M.Romero, S.Doyle, R.Elliott, G.Murphy, and R.Carpenter. 1990. *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**: 1311-1322.

Courey,A.J. and R.Tjian. 1988. Analysis of Sp1 *in vivo* reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**: 887-898.

Crick,F.H.C. 1952. Is  $\alpha$ -keratin a coiled coil? *Nature* **170**: 882-883.

- Dayhoff,M.O. 1969. Computer analysis of protein evolution. *Scient. Am.* **221**: 86-95.
- Depicker,A., S.Stachel, P.Dhaese, P.Zambryski, and H.M.Goodman. 1982. *J. Mol. Appl. Genet.* **1**: 561-573.
- De Wet,J.R., K.V.Wood, D.R.Helinski, and M.DeLuca. 1985. Cloning of firefly luciferase cDNA and the expression of active luciferase in *E. coli*. *Proc. Natl. Acad. Sci. USA* **82**: 7870-7873.
- Dignam,J.D., R.M.Lebovitz, and R.G.Roeder. 1983a. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucl. Acids Res.* **11**: 1475-1489.
- Dignam,J.D., P.L.Martin, B.S.Shastry, and R.G.Roeder. 1983b. Eukaryotic gene transcription with purified components. *Meth. Enzymol.* **101**: 582-598.
- Driever,W., J.Ma, C.Nusslein-Volhard, and M.Ptashne. 1989. Rescue of *bicoid* mutant *Drosophila* embryos by Bicoid fusion proteins containing heterologous activating sequences. *Nature* **342**: 149-154.
- Elgin,S.C. 1988. The formation and function of DNase I hypersensitive sites in the process of gene activation. *J. Biol. Chem.* **263**: 19252-19262.
- Ellis,J.G., D.J.Llewellyn, J.C.Walker, E.S.Dennis, and W.J.Peacock. 1987. The *ocs* element: a 16 base pair palindrome essential for activity of the octopine synthase enhancer. *EMBO J.* **6**: 3203-3208.
- Fang,R.-X., F.Nagy, S.Sivasubramaniam, and N.-H.Chua. 1989. Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *Plant Cell* **1**: 141-150.
- Fedoroff,N.V. 1989. About maize transposable elements and development. *Cell* **56**: 181-191.

- Feldmann, K.A., D. Marks, M.L. Christianson, and R.S. Quatrano. 1989. A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science* **243**: 1351-1354.
- Fiering, S., J.P. Norghrop, G.P. Nolan, P.S. Mattila, G.R. Crabtree, and L.A. Herzenberg. 1990. Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev.* **4**: 1823-1834.
- Fischer, J.A., E. Giniger, T. Maniatis, and M. Ptashne. 1988. GAL4 activates transcription in *Drosophila*. *Nature* **332**: 853-856.
- Fluhr, R., P. Moses, G. Morelli, G. Coruzzi, and N.-H. Chua. 1986. Expression dynamics of the pea *rbcS* multigene family and organ distribution of the transcripts. *EMBO J.* **5**: 2063-2071.
- Friedman, A.D. and S.L. McKnight. 1990. Identification of two polypeptide segments of CCAAT/enhancer-binding protein required for transcriptional activation of the serum albumin gene. *Genes Dev.* **4**: 1416-1426.
- Fromm, H., F. Katagiri, and N.-H. Chua. 1989. An octopine synthase enhancer element directs tissue-specific expression and binds ASF-1, a factor from tobacco nuclear extracts. *Plant Cell* **1**: 977-984.
- Gallie, D.R., W.J. Lucas, and V. Walbot. 1989. Visualizing mRNA expression in plant protoplasts: factors influencing efficient mRNA uptake and translation. *Plant Cell* **1**: 301-311.
- Gentz, R., F.J. Rauscher III, C. Abate, and T. Curran. 1989. Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. *Science* **243**: 1695-1699.
- Ghosh, S., A.M. Gifford, L.R. Riviere, P. Tempst, G.P. Nolan, and D. Baltimore. 1990. Cloning of the p50 DNA binding subunit of NF- $\kappa$ B: homology to *rel* and *dorsal*. *Cell* **62**: 1019-1029.
- Gilbert, W. 1978. Why genes in pieces? *Nature* **271**: 501.

Giniger,E. and M.Ptashne. 1987. Transcription in yeast activated by a putative amphipathic  $\alpha$  helix linked to a DNA binding unit. *Nature* **330**: 670-672.

Goff,S.A., T.M.Klein, B.A.Roth, M.E.Fromm, K.C.Cone, J.P.Radicella, and V.L.Chandler. 1990. Transactivation of anthocyanin biosynthetic genes following transfer of *B* regulatory genes into maize. *EMBO J.* **9**: 2517-2522.

Gonzalez,G.A., K.K.Yamamoto, W.H.Fischer, D.Karr, P.Menzel, W.Biggs III, W.W.Vale, and M.R.Montminy. 1989. A cluster of phosphorylation sites on the cyclic AMP-regulated nuclear factor CREB predicted by its sequence. *Nature* **337**: 749-752.

Gorman,C.M., L.F.Moffat, and B.H.Howard. 1982. Recombinant Genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**: 1044-1051.

Green,P.J., S.A.Kay and N.-H.Chua. 1987. Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the *rbcS-3A* gene. *EMBO J.* **6**: 2543-2549.

Gregor,P.D., M.Sawadogo, and R.G.Roeder. 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev.* **4**: 1730-1740.

Guarente,L. and M.Ptashne. 1981. Fusion of *E. coli lacZ* to the cytochrome *c* gene of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **78**: 2199-2203.

Guilley,H., R.K.Dudley, G.Jonard, E.Balazs, and K.E.Richards. 1982. Transcription of cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* **30**: 763-773.

Guiltinan,M.J., W.R.Marcotte,Jr., R.S.Quatrano. 1990. A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* **250**:267-271.

Hai,T., M.Horikoshi, R.G.Roeder, and M.R.Green. 1988. Analysis of the assembly of a

transcription factor ATF in the assembly of a functional preinitiation complex. *Cell* **54**: 1043-1051.

Hartings,H., M.Maddaloni, N.Lazzaroni, N.Di Fonzo, M.Motto, F.Salamini and R.Thompson. 1989. The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activator. *EMBO J.* **8**: 2795-2801.

Hawley,D.K. and R.G.Roeder. 1985. Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. *J. Biol. Chem.* **260**: 8163-8172.

----. 1987. Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. *J. Biol. Chem.* **262**: 3452-3461.

Hoeffler,J.P., T.E.Meyer, Y.Yun, J.L.Jameson, and J.F.Habener. 1988. Cyclic AMP-responsive DNA-binding protein: structure based on a cloned placental cDNA. *Science* **242**: 1430-1433.

Hohn,T., B.Hohn, and P.Pfeiffer. 1985. Reverse transcription in CaMV. *Trends Biochem. Sci.* **10**: 205-209.

Hohn,T., Richards,K. and Genevieve-Lebeurier. 1982. Cauliflower mosaic virus on its way to becoming a useful plant vector. *Curr. Topics Microbiol. Immun.* **96**: 194-236.

Hope,I.A., S.Mahadevan, and K.Struhl. 1988. Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* **333**: 635-640.

Hope,I.A. and K.Struhl. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**: 885-894.

----. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target

DNA. *EMBO J.* **6**: 2781-2784.

Horikoshi,M., M.F.Carey, H.Kakidani, and R.G.Roeder. 1988a. Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. *Cell* **54**: 665-669.

Horikoshi,M., T.Hai, Y.-S.Lin, M.R.Green, and R.G.Roeder. 1988b. Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell* **54**: 1033-1042.

Horsch,R.B. and H.J.Klee. 1986. Rapid assay of foreign gene expression in leaf discs transformed by *Agrobacterium tumefaciens*: role of T-DNA borders in the transfer process. *Proc. Natl. Acad. Sci. USA* **83**: 4428-4432.

Hu,Y.-F., B.Luscher, A.Admon, N.Mermoud, and R.Tjian. 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Dev.* **4**: 1741-1752.

Huynh,T.V., R.A.Young, and R.W.Davis. 1985. Constructing and screening cDNA libraries in  $\lambda$ gt10 and  $\lambda$ gt11. in: DNA Cloning, vol.1, a practical approach, D.M.Glover (Ed.), IRL Press, Oxford, England.

Jefferson,R.A., T.A.Kavanagh, and M.W.Bevan. 1987. GUS fusions:  $\beta$ -glucuronidase as a sensitive and versatile gene marker in higher plants. *EMBO J.* **6**: 3901-3907.

Jendrisak,J.J. and R.R.Burgess. 1975. A new method for the large-scale purification of wheat germ DNA-dependent RNA polymerase II. *Biochemistry* **14**: 4639-4645.

Johnson,P.F. and S.L.McKnight. 1989. Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58**: 799-839.

Kadonaga,J.T. and R.Tjian. 1986. Affinity purification of sequence-specific DNA binding

proteins. *Proc. Natl. Acad. Sci. USA* **83**: 5889-5893.

Kakidani, H. and M. Ptashne. 1988. GAL4 activates gene expression in mammalian cells. *Cell* **52**: 161-167.

Karim, F.D., L.D. Urness, C.S. Thummel, M.J. Klemsz, S.R. McKercher, A. Celada, C. van Beveren, R.A. Maki, C.V. Gunther, J.A. Nye, and B.J. Graves. 1990. The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes Dev.* **4**: 1451-1453.

Kassavetis, G.A., B.R. Braun, L.H. Nguyen, and E.P. Geiduschek. 1990. *S. cerevisiae* TFIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIA and TFIIC are assembly factors. *Cell* **60**: 235-245.

Katagiri, F., E. Lam, and N.-H. Chua. 1989. Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. *Nature* **340**: 727-730.

Katagiri, F., K. Yamazaki, M. Horikoshi, R.G. Roeder, and N.-H. Chua. 1990. A plant DNA-binding protein increases the number of active preinitiation complexes in a human *in vitro* transcription system. *Genes Dev.* **4**: 1899-1909.

Keegan, L., G. Gill, and M. Ptashne. 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231**: 699-704.

Kieran, M., V. Blank, F. Logeat, J. Vandekerckhove, F. Lottspeich, O. Le Bail, M.B. Urban, P. Kourilsky, P.A. Baeuerle, and A. Israel. 1990. The DNA binding subunit of NF- $\kappa$ B is identical to factor KBF1 and homologous to the *rel* oncogene product. *Cell* **62**: 1007-1018.

Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 1990. Crystal Structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**: 579-590.



Klein-Hitpass,L., S.Y.Tsai, N.L.Weigel, G.F.Allan, D.Riley, R.Rodriguez, W.T.Schrader, M.-J.Tsai, and B.W.O'Malley. 1990. The progesterone receptor stimulates cell-free transcription by enhancing the formation of a stable preinitiation complex. *Cell*, **60**, 247-257.

Ko,M.S.H., H.Nakauchi, and N.Takahashi. 1990. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J.* **9**: 2835-2842.

Kouzarides,T. and E.Ziff. 1989. Leucine zippers of *fos*, *jun* and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature* **340**: 568-571.

Kozak,M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNA. *Nuc. Acids Res.* **15**: 8125-8132.

Kuhlemeier,C., Green,P.J., and N.-H.Chua. Regulation of gene expression in higher plants. *Annu. Rev. Plant Physiol.* **38**: 221-257.

Lam,E., P.N.Benfey, P.M.Gilmartin, R.-X.Fang, and N.-H.Chua. 1989. Site-specific mutations alter *in vitro* factor binding and change promoter expression pattern in transgenic plants. *Proc. Natl. Acad. Sci. USA* **86**: 7890-7894.

Lam,E., Benfey,P., and N.-H.Chua. 1990a. Characterization of *as-1*: a factor binding site on the 35S promoter of cauliflower mosaic virus. in *Plant Gene Transfer*, UCLA Symposia on Molecular and Cellular Biology, New Series, C.Lamb and R.Beachy (Eds.). Alan R. Liss, Inc., New York **29**: pp.71-79.

Lam,E. and N.-H. Chua. 1989. ASF-2: a factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in *Cab* promoters.

----. 1990. GT-1 binding site confers light-responsive expression in transgenic tobacco. *Science* **248**: 471-474.

Lam,E., F.Katagiri, and N.-H.Chua. 1990b. The plant nuclear factor ASF-1 binds to an essential region of the nopaline synthase promoter. *J. Biol. Chem.* **265**: 9909-9913.

Landschulz,W.H., P.F.Johnson, and S.L.McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764.

----. 1989. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243**: 1681-1688.

Laughon,A. and M.Scott. 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature* **310**: 25-31.

Lin,Y.-S., M.F.Carey, M.Ptashne, and M.R.Green. 1988. GAL4 derivatives function alone and synergistically with mammalian activators *in vitro*. *Cell* **54**: 659-664.

Ludwig,S.R., B.Bowen, L.Beach, and S.R.Wessler. 1990. A regulatory gene as a novel visible marker for maize transformation. *Science* **247**: 449-450.

Ma,J., E.Przibilla, J.Hu, L.Bogorad, and M.Ptashne. 1988. Yeast activators stimulate plant gene expression. *Nature* **334**: 631-633.

Ma,J. and M.Ptashne. 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48**: 847-853.

Matsudaira,P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**: 10035-10038.

McKay,D.B. and Steitz,T.A. 1981. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* **290**: 744-749.

Mermod,N., E.A.O'Neill, T.J.Kelly, and R.Tjian. 1989. The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58**:

741-753.

Messing,J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**: 20-78.

Meyerowitz,E.M. 1989. Arabidopsis, a useful weed. *Cell* **56**: 263-269.

Mikami,K., T.Tabata, T.Kawata, T.Nakayama, and M.Iwabuchi. 1987. Nuclear protein(s) binding to the conserved DNA hexameric sequence postulated to regulate transcription of wheat histone genes. *FEBS Lett.* **223**: 273-278.

Mikami,K., A.Sakamoto, H.Takase, T.Tabata, and M.Iwabuchi. 1989a. Wheat nuclear protein HBP-1 binds to the hexameric sequence in the promoter of various plant genes. *Nucl. Acids Res.* **17**, 9707- 9717.

Mikami,K., H.Takase, T.Tabata, and M.Iwabuchi. 1989b. Multiplicity of the DNA-binding protein HBP-1 specific to the conserved hexameric sequence ACGTCA in various plant gene promoters. *FEBS Lett.* **256**, 67-70.

Miller,J., A.D.McLachlan, and A.Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**: 1609-1614.

Miskimins,W.K., M.P.Roberts, A.McClelland, and F.H.Ruddle. 1985. Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. *Proc. Natl. Acad. Sci. USA* **82**: 6741-6744.

Mitchell,P.J. and R.Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**: 371-378.

Murashige,T. and F.Schoog. 1962. *Physiologia Plantarum* **15**: 473-497.

Murre,C., P.S.McCaw, and D.Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. *Cell* **56**: 777-783.

Nam,H.-G., J.Giraudat, B.den Boer, F.Moonan, W.D.B.Loos, B.M.Hauge, and H.M.Goodman. 1989. Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell* **1**: 699-705.

Norman,C., M.Runswick, R.Pollock, and R.Triesman. 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the *c-fos* serum response element. *Cell* **55**: 989-1003.

Odell,J.T., F.Nagy, and N.-H.Chua. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**: 810-812.

Ohkuma,Y., M.Horikoshi, R.G.Roeder, and C.Desplan. 1990. Binding site-dependent direct activation and repression of *in vitro* transcription by Drosophila homeodomain proteins. *Cell* **61**: 475-484.

O'Shea,E.K., R.Rutkowski, and P.S.Kim. 1989. Evidence that the leucine zipper is a coiled coil. *Science* **243**: 538-542.

Ow,D.W., K.V.Wood, M.DeLuca, J.R.de Wet, D.R.Helinski, and S.H.Howell. 1986. Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science* **234**: 856-859.

Ow,D.W., J.D.Jacobs, and S.H.Howell. 1987. Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. *Proc. Natl. Acad. Sci. USA* **84**: 4870-4874.

Pfeiffer,P. and T.Hohn. 1983. Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* **33**: 781-789.

Poulsen,C. and N.-H.Chua. 1988. Dissection of 5' upstream sequences for selective expression of the *Nicotiana plumbaginifolia rbcS-8B* gene. *Mol. Gen. Genet.* **214**: 16-23.

- Prat,S., L.Willmitzer, and J.J.Sanchez-Serrano. 1989. Nuclear proteins binding to a cauliflower mosaic virus 35S truncated promoter. *Mol. Gen. Genet.* **217**: 209-214.
- Ptashne,M. 1988. How eukaryotic transcriptional activators work. *Nature* **335**: 683-689.
- Rauscher III,F.J., P.J.Voulalas, B.R.Franza Jr., and T.Curran. 1988. Fos and Jun bind cooperatively to the AP-1 site: reconstitution *in vitro*. *Genes Dev.* **2**: 1687-1699.
- Rogers,S., R.Wells, and M.Reichsteiner. 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**: 364-368.
- Rosenberg,M. and D.Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**: 319-353.
- Rosenberg,A.H., B.N.Lade, D.Chui, S.-W.Lin, J.J.Dunn, and F.W.Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**: 125-135.
- Saltzman,A.G. and R.Weinmann. 1989. Promoter specificity and modulation of RNA polymerase II transcription. *FASEB J.* **3**: 1723-1733.
- Sambrook,J., E.F.Fritsch, and T.Manias. 1989. Molecular Cloning, a laboratory manual, 2nd Ed., vols. **1-3**, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanders,P.R., J.A.Winter, A.R.Barnason, S.G.Rogers, and R.T.Fraley. 1987. Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucl. Acids Res.* **15**: 1543-1558.
- Sawadogo,M. and R.G.Roeder. 1985a. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**: 165-175.
- . 1985b. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative *in vitro* assay. *Proc. Natl. Acad. Sci. USA* **82**: 4394-4398.

Schmidt,R.J., F.A.Burr, M.J.Aukerman, and B.Burr. 1990. Maize regulatory gene opaque-2 encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc. Natl. Acad. Sci. USA* **87**: 46-50.

Schule,R., P.Rangarajan, S.Kliwer, L.J.Ransone, J.Bolado, N.Yang, I.M.Verma, and R.M.Evans. 1990. Functional antagonism between oncoprotein c-Jun and the glucocorticoid receptor. *Cell* **62**: 1217-1226.

Schutte,J., J.Viallet, M.Nau, S.Segal, J.Fedorko, and J.Minna. *jun-B* inhibits and *c-fos* stimulates the transforming and trans-activating activities of *c-jun*. *Cell* **59**: 987-997.

Schwabe,J.W., D.Neuhaus, and D.Rhodes. 1990. Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* **348**: 458-461.

Schwarz-Sommer,Z., P.Huijser, W.Nacken, H.Saedler, and H.Sommer. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**: 931-936.

Singh,K., E.S.Dennis, J.G.Ellis, D.J.Llewellyn, J.G.Tokuhisa, J.A.Wahleithner, and W.J.Peacock. 1990. OCSBF-1, a maize ocs enhancer binding factor: isolation and expression during development. *Plant Cell* **2**: 891-903.

Singh,H., J.H.LeBowitz, A.S.Baldwin, and P.A.Sharp. 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. *Cell* **52**: 415-423.

Singh,K., J.G.Tokuhisa, E.S.Dennis, and W.J.Peacock. 1989. Saturation mutagenesis of the octopine synthase enhancer: correlation of mutant phenotypes with binding of a nuclear protein factor. *Proc. Natl. Acad. Sci. USA* **86**: 3733-3737.

Steitz,T.A., D.H.Ohlendorf, D.B.McKay, W.F.Anderson, and B.W.Matthews. 1982. Structural similarity in the DNA-binding domains of catabolite gene activator and cro repressor proteins. *Proc. Natl. Acad. Sci. USA* **79**: 3097-3100.

Studier, F.W., A.H. Rosenberg, J.J. Dunn, and J.W. Dubendorf. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**: 60-89.

Tabata, T., H. Takase, S. Takayama, K. Mikami, A. Nakatsuka, T. Kawata, T. Nakayama, and M. Iwabuchi. 1989. A protein that binds to a cis-acting element of wheat histone genes has a leucine zipper motif. *Science* **245**: 965-967.

Talanian, R.V., C.J. McKnight, and P.S. Kim. 1990. Sequence-specific DNA binding by a short peptide dimer. *Science* **249**: 769-771.

Turner, R. and R. Tjian. 1989. Leucine repeats and an adjacent DNA binding domain mediated the formation of functional cFos-cJun heterodimer. *Science* **243**: 1689-1694.

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**: 1335-1338.

Vinson, C.R., K.L. LaMarco, P.F. Johnson, W.H. Landschulz, and S.L. McKnight. 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* **2**: 801-806.

Vinson, C.R., P.B. Sigler, and S.L. McKnight. 1989. Scissors-grip model for DNA recognition by a family of zipper proteins. *Science* **246**: 911-916.

Webster, N., J.R. Jin, S. Green, M. Hollis, and P. Chambon. 1988. The yeast UAS<sub>G</sub> is a transcriptional enhancer in human HeLa cells on the presence of the GAL4 trans-activator. *Cell* **52**: 169-178.

Workman, J.L., S.M. Abmayr, W.A. Cromlish, and R.G. Roeder. 1988. Transcriptional regulation by the immediate early protein of pseudorabies virus during *in vitro* nucleosome assembly. *Cell* **55**: 211-219.

Workman, J.L., R.G. Roeder, and R.E. Kingston. 1990. An upstream transcription factor, USF

(MLTF), facilitates the formation of preinitiation complexes during *in vitro* chromatin assembly. *EMBO J.* **9**: 1299-1308.

Yamazaki,K. and F.Imamoto. 1987. Selective and accurate initiation of transcription at the T-DNA promoter in a soluble chromatin extract from wheat germ. *Mol. Gen. Genet.* **209**: 445-452.

Yamazaki,K., F.Katagiri, H.Imaseki, and N.-H.Chua. 1990. TGA1a, a tobacco DNA-binding protein increases the rate of initiation in a plant *in vitro* transcription system. *Proc. Natl. Acad. Sci. USA* **87**: 7035-7039.

Yanisch-Perron,C., J.Vieira, and J.Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.





THE LIBRARY



19010000065749

**End**