

# GAGA factor and TBF1 bind DNA elements that direct ubiquitous transcription of the *Drosophila* $\alpha 1$ -tubulin gene

Kimberly H.O'Donnell<sup>+</sup> and Pieter C.Wensink\*

Department of Biochemistry and The Rosenstiel Center, Brandeis University, Waltham, MA 02254-9110, USA

Received July 5, 1994; Revised and Accepted October 5, 1994

## ABSTRACT

Three DNA regions (TE1, TE2 and the intron) regulate the ubiquitous expression of the  $\alpha 1$ -tubulin gene of *Drosophila melanogaster*. In this report, we identify two proteins that bind these DNA regions. One is the previously characterized GAGA transcription factor and the other is a newly identified 62 kDa polypeptide, TBF1 (TE1-binding factor 1). Purified GAGA factor binds three sites in TE2 and at least three in the intron. TBF1 was purified from embryos and binds to both TE1 and TE2. Together, the two proteins produce the same DNase I footprints in TE1 and TE2 as does a nuclear extract that transcribes the gene accurately. These footprints cover most of the TE1 and TE2 DNA. Moreover, one binding site for each protein coincides with a site that activates transcription *in vitro*. The characteristics of the GAGA factor and the genes it regulates suggest roles these two proteins are likely to play in regulating ubiquitous expression.

## INTRODUCTION

Many structural proteins and metabolic enzymes are essential for the viability of all cells in a higher organism. The regulatory mechanisms that allow ubiquitous expression of genes encoding such proteins are unknown. Very different regulatory mechanisms could direct this expression. For example, a single transcription factor could activate expression in all tissues and developmental stages. Alternatively, many different factors could regulate expression, each factor occurring and acting in a different tissue and stage. Although the regulation of several ubiquitously expressed genes has been investigated, there is as yet no evidence to support either of these or other mechanisms [for example, (1–7)].

One reason for this lack of evidence is that *in vivo* analyses of ubiquitous expression have been done in cultured cells rather than in transgenic animals. To overcome this limitation, we are investigating regulation of the  $\alpha 1$ -tubulin gene ( $\alpha 1t$ ) from *Drosophila melanogaster*. This is one of the four *Drosophila*

genes that encode a tubulin, a protein essential for cell division and for the motility and structure of cells. Although the level of  $\alpha 1t$  gene expression varies, it is abundant in all tissues and all stages of development (8–11). Studying this gene in *Drosophila* has allowed germline transformation experiments to identify the DNA regions necessary for its expression.

The  $\alpha 1t$  DNA between –157 and +696 is sufficient to direct ubiquitous expression [Fig. 1; (11)]. Three regions within this DNA regulate  $\alpha 1t$  expression both *in vitro* and *in vivo*: TE1 (tubulin element 1; 29 bp), TE2 (68 bp) and the intron (492 bp). *In vitro* transcription experiments have shown that all three activate  $\alpha 1t$  transcription. Germline transformation experiments showed that when any one of the regions is deleted,  $\alpha 1t$  expression *in vivo* becomes highly sensitive to chromosome position effects. At some chromosomal positions the deletions have no effect, whereas at others expression was eliminated in either some or almost all tissues. The elimination of expression in all tissues indicates that each regulatory region functions ubiquitously. Altogether, these findings suggest a simple hypothesis for  $\alpha 1t$  regulation: that a single ubiquitous regulatory protein activates expression from all three regulatory regions.

To begin a test of this hypothesis and investigate alternatives, we have searched for proteins that bind the regulatory regions. Experiments described in this report demonstrate that the *Drosophila* GAGA factor binds to TE2 and the intron, but not to TE1. The GAGA factor is ubiquitous during embryonic development and is present in all subsequent stages tested (12). We also identify and purify another protein, TBF1, which binds to TE1 and TE2. The footprints of these two proteins account for all of the regulatory region footprints observed with transcription extracts. The known properties of the GAGA factor and the genes it regulates indicate the roles these two proteins are likely to play in regulating ubiquitous expression.

## MATERIALS AND METHODS

### DNA constructions

Construction of  $p\alpha 1t(-157/+696)$  and  $p\alpha 1t(\Delta TE1/+696)$  were described previously (11). The  $\Delta TE1$  construct substitutes 5'-GC-

\*To whom correspondence should be addressed

<sup>+</sup>Present address: Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138, USA

TAGCCATATCCCATCCTCCGCACATTG-3' for the entire TE1, 5'-CTTCAGTTATCGGTTATGCGGCGTTTAAAG-3'.

### Nuclear extracts

Nuclear extracts were prepared by standard methods with minor modifications from *Drosophila* embryos (13,14) and from Kc0 cells (15,16). Kc0 cells, a *Drosophila* cell line of embryonic origin, were purchased from M.I.T. Cell Culture Center or from Harvard University. All protein concentrations were determined by the Bradford assay (17), using bovine serum albumin as a standard.

Up to 12 mg of nuclear extract (typically 12 mg/ml) from Kc0 cells were heat-treated at 100°C for 10 min and precipitated proteins were removed by microcentrifugation at 4°C for 5 min. Typically, 98% of the protein was eliminated by this procedure.

### Purification of TBF1

TBF1 was purified from embryonic nuclear extract in two steps. First, 33 ml of embryonic nuclear extract (7.6 mg/ml protein) in 0.1 M KCl, HEMG (25 mM Hepes, pH 7.6, 0.1 mM EDTA, pH 7.9, 12.5 mM MgCl<sub>2</sub>, 10% glycerol and 1 mM DTT), 0.5 mM sodium metabisulphite, 0.1 mM PMSF was heated to 55°C for 10 min. Precipitated material was removed by centrifugation at 20 200 × g for 15 min at 4°C, and soluble protein (~160 mg) was recovered in the supernatant.

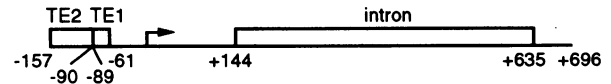
In the second step, heat-stable proteins were subjected to two cycles of DNA affinity chromatography as described previously (18), except that poly [(dI:dC):(dI:dC)] was added to the heat-treated extract to a final concentration of 15 µg/ml for the first cycle of affinity chromatography, and 1 µg/ml for the second. NP-40 was included in all buffers at a concentration of 0.05%. The affinity column was prepared by annealing and ligating oligonucleotides:

5'-AGCTTGCATCTTCAGTTATCGGTTATGCGGCGTTTAAAG-3' and 5'-TCGACTTAAACGCCGCATAACCGATAACTGAAGATGCA-3' (TE1 sequence underlined).

In the first cycle, the extract was applied to a 0.5 ml affinity column equilibrated in 100 mM KCl, HEMG, NP-40. Bound proteins were washed with 0.1 M KCl in HEMG, NP-40 and eluted with a step gradient of 0.2 M KCl, 0.6 M KCl, and 1 M KCl in HEMG, NP-40. The high specific activity 0.6 M fraction from the first cycle was subjected to a second cycle of DNA affinity chromatography. This was diluted to 0.1 M KCl in HEMG, NP-40 with HEMG, NP-40 and applied to a 0.2 ml affinity column, washed and eluted as described for the first cycle.

### DNase I footprinting

DNase I footprinting (19) was performed as described previously (16) with minor modifications (20). Between 0.5 and 25 µl of GAGA factor, heat-treated extract or TBF1 was incubated with 6 to 15 fmoles of <sup>32</sup>P-labeled fragments. Purified GAGA factor (14) was generously provided by A. TenHarmsel and M. Biggin. Reactions containing crude or heat-treated Kc0 nuclear extract included 1 µg poly [(dI:dC):(dI:dC)]. With the exception of the -157 to +39 substrate used in Figure 4, <sup>32</sup>P-labelled DNA fragments from -157 to +39, from -157 to +88, and the ΔTE1 fragment from -157 to +39 were synthesized by the polymerase chain reaction using unlabeled and <sup>32</sup>P-labeled primer pairs, as described previously (21). The <sup>32</sup>P-labelled DNA fragment from -157 to +39 used in Figure 4 and the <sup>32</sup>P-labelled intron



**Figure 1.** Regulatory regions of  $\alpha It$ . Open boxes show the three regulatory regions in the  $\alpha It$  segment used for germline transformation experiments. The arrow shows the transcription initiation site and the numbers indicate the endpoints of regions in the segment.

fragment from +333 to +696 were prepared from  $\alpha It(-157/+696)$  and  $\alpha It(\Delta TE1/TE2/+696)$ , respectively (11), using standard methods (22) with enzymes from New England Biolabs or Boehringer Mannheim.

### UV cross-linking

UV cross-linking was performed as described previously (23) with minor modifications. The TE1 substrate was prepared by hybridizing the coding strand TE1 oligonucleotide (-89 to -61, described above) to a 10 nucleotide primer (5'-CTTAAACGCC-3') and by polymerizing the hybrid with the Klenow fragment of *E. coli* DNA polymerase I in the presence of  $\alpha^{32}P$ -dCTP, nucleotides, and bromodeoxyuridine (BrdU).

Binding reactions included 300 ng of heat-treated Kc0 nuclear extract or 8 ng of the 1 M fraction of purified TBF1, second cycle (Table I), 0.67 nM TE1 substrate, and included 0.02 ng/µl poly [(dI:dC):(dI:dC)] as nonspecific competitor unless otherwise indicated. As indicated in figure legend, heat-treated Kc0 extract was treated with 1 µg of proteinase K for 10 min at 37°C. Synthetic oligonucleotide competitors (see below) were included in binding reactions where indicated. The sizes of the cross-linked products were estimated relative to protein molecular weight standards (Bio-Rad) included in the gels.

The hybridized synthetic oligonucleotides used to form duplex competitors are:

T2, with  $\alpha It$  DNA from -117 to -90 (underlined):  
 5'-AGCTGAGAGACCAAGTGCCATTACCTCTCCCAGCTAGC-3' and  
 5'-TCGAGCTAGCTGGGAGAGGTAATGGCACTTGGTCTCTC-3';  
 T3, with  $\alpha It$  DNA from +409 to +446 (underlined):  
 5'-CCATGATACTTCGACGCATAACTGTAGATTTGGATAG-3'  
 5'-CTATCAAAAATCTACAGTTATGCGTCGAAGTATCATGG-3'; and  
 nonspecific (NS):  
 5'-TCGACCAAGCATTGAAGCCGTCGAGTGGCTCTCCGATCCC-  
 AGC-3' and  
 5'-TCGAGCTGGATCCGGAGAGCCACTGCGACGGCTTCAATGCC-  
 TGG-3'.

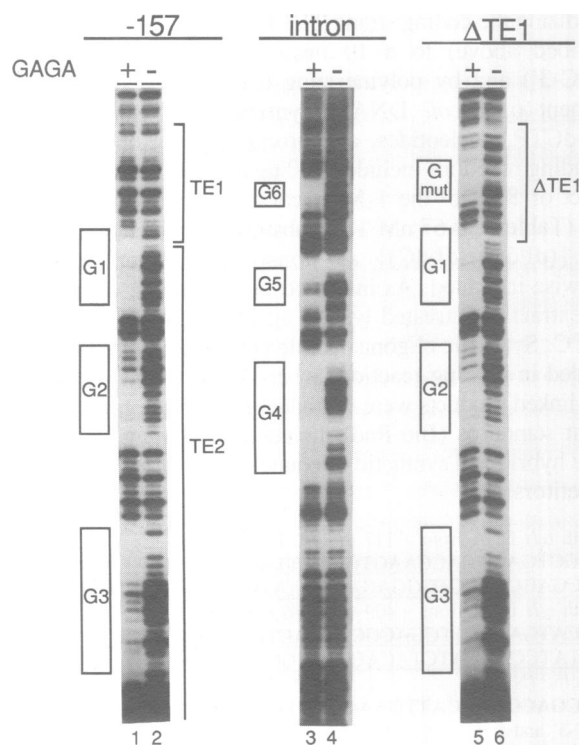
## RESULTS

### GAGA factor binds TE2 and the intron

TE2 and the intron have several sequences that match the consensus binding site for the *Drosophila* GAGA transcription factor (11,14). To test whether GAGA factor binds to these sequences, DNase I footprinting experiments were done with highly purified GAGA protein. As shown in Figure 2 (lanes 1 and 2) and summarized in Figure 3, the protein binds to three sites in TE2 (G1, -101 to -86; G2, -120 to -108; and G3, -148 to -133). It also binds to at least three sites in the intron (G4, +448 to +475; G5, +490 to +501; and G6, +526 to +539), which are shown in Figure 2 (lanes 3 and 4). In contrast to the other GAGA footprints, G3 has no similarity to the GAGA consensus binding site. We speculate that GAGA may have a broad binding specificity or that the G3 footprint may be

produced by a protein co-purifying with GAGA. In either case, the GAGA footprints occupy a substantial portion of the  $\alpha lt$  regulatory regions (Fig. 3).

The G1 footprint is primarily in TE2 but extends a short distance into TE1, suggesting that the previously observed regulatory effects of TE1 may be due to its influence on GAGA factor binding to G1. We examined this possibility by footprinting a DNA fragment containing a TE1 mutation ( $\Delta TE1$ ) that has been shown to dramatically alter regulation *in vitro* and *in vivo*. Results demonstrate that the G1, G2 and G3 footprints are unaffected by this substitution mutation (Fig. 2, lanes 1 and 5). Another GAGA footprint (Gmut) occurred in the substituted DNA, but, like G3, has no similarity to the GAGA consensus binding site. This does not change the conclusion that TE1 is unnecessary for the normal GAGA footprints, including G1. These results lead to the prediction that regulation by GAGA factor will be limited



**Figure 2.** DNase I footprinting of GAGA factor. Footprint experiments used  $\sim 7$  fmoles of a radiolabeled  $\alpha 1$  fragment from  $-157$  to  $+39$  (lanes 1 and 2), an intron fragment from  $+333$  to  $+696$  (lanes 3 and 4), and a fragment from  $-157$  to  $+39$  containing the TE1 substitution mutation,  $\Delta TE1$  (lanes 5 and 6). Reactions contained 120 ng (lanes 1 and 5), 25 ng (lane 3) or 0 ng (lanes 2, 4 and 6) of purified GAGA factor. TE1, TE2 and  $\Delta TE1$  are indicated by the brackets, and GAGA factor (G) footprints are represented by boxes.

to TE2 and the intron and that a different protein mediates regulation by TE1.

#### Identification of TBF1, a TE1-binding factor

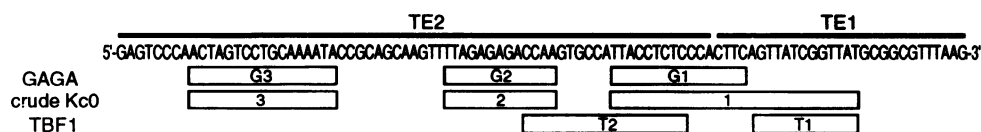
To identify additional  $\alpha lt$  binding proteins, DNase I footprinting assays were done with a nuclear extract from cultured *Drosophila* embryonic cells (Kc0). This extract transcribes  $\alpha lt$  templates accurately and efficiently only if the templates contain both TE1 and TE2 (11). Therefore this extract is likely to have proteins that bind to and regulate from both regions. As shown in Figure 4 (lanes 3 and 4) and summarized in Figure 3, three protected regions were produced. The first, footprint 1 ( $-101$  to  $-73$ ), includes the G1 GAGA footprint, but extends further downstream to the middle of TE1. The other two, footprints 2 and 3, have the same endpoints as GAGA footprints G2 and G3 ( $-120$  to  $-108$  and  $-148$  to  $-133$ ). Since the GAGA factor is known to be present in Kc0 nuclear extracts (14), these results strongly suggest that a portion of footprint 1, and all of footprints 2 and 3 are produced by GAGA factor in the extract. More importantly, they also suggest that the remaining portion of footprint 1 is produced by a different protein.

To separate the second binding activity from GAGA factor, the Kc0 nuclear extract was heated to  $100^\circ\text{C}$  and precipitated proteins were removed. All three GAGA factor footprints were eliminated by this treatment, but the 3' portion of footprint 1 remained (footprint T1;  $-84$  to  $-73$ ; Fig. 4, lanes 7 and 8). We call this heat-stable binding activity TBF1 (TE1-binding factor 1).

We used UV cross-linking assays to examine the binding specificity and size of TBF1. Crosslinking radiolabeled TE1 with protein in the heat-treated extract yielded a predominant product that was not detected when extract or UV light treatment was omitted, or when the extract was pre-treated with proteinase K (Fig. 5A, lanes 1–5). To examine the binding specificity of this TBF1 protein, we tested the effect of competitor DNAs. When binding reactions included an excess of unlabeled TE1, the intensity of the radiolabeled complex was dramatically reduced (Fig. 5B, lanes 1–4). In contrast, an identical excess of a nonspecific competitor DNA resulted in only a very small decrease (Fig. 5B, lanes 7 and 8). These results and the footprinting results indicate that TBF1 binds specifically to TE1. To estimate the size of TBF1, the cross-linked complex was treated with nucleases. Partial nuclease digestion reduced the apparent size to 62 kDa. Further digestion reduced the radioactive signal, but did not change the electrophoretic mobility of the complex (Fig. 5A, lanes 7 to 9). We conclude that TBF1 is a heat-stable protein of approximately 62 kDa that binds specifically to TE1 DNA.

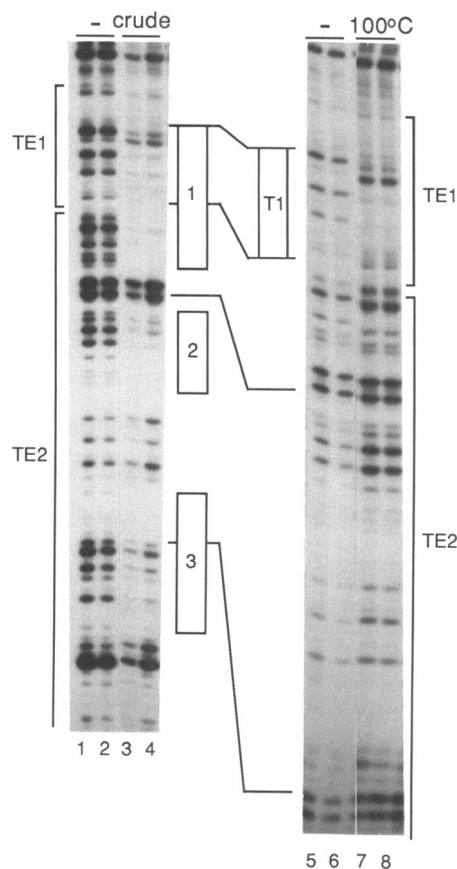
#### Purification of TBF1

To purify TBF1, heat-treated nuclear extract prepared from embryos was fractionated on a TE1 DNA affinity column. Like



**Figure 3.** Summary of footprints in TE1 and TE2. The DNA sequence of TE1 and TE2 are shown. Footprints of purified GAGA factor, Kc0 nuclear extract and purified TBF1 are shown as open boxes.

the Kc0 nuclear extract, embryonic extract required both upstream regulatory regions for efficient transcription of  $\alpha It$  (11). After heating at 55°C, the extract was fractionated over the affinity column and assayed by DNase I footprinting, giving the results shown in Table I. The overall purification after the second cycle of affinity chromatography was ~10 000-fold. As expected, this most highly purified TBF1 fraction produced a cross-linked product with TE1 that was the same size as the product detected in the heat-treated extract (Fig. 5C, lanes 1–3).



**Figure 4.** DNase I footprinting with crude and heat-treated Kc0 nuclear extracts. Radiolabeled fragments from –157 to +39 (lanes 1 to 4) or –157 to +88 (lanes 5 to 8) were incubated with 272  $\mu$ g of crude nuclear Kc0 extract (lanes 3 and 4), 43  $\mu$ g of heat-treated Kc0 nuclear extract (lanes 7 and 8) or no extract (lanes 1, 2, 5 and 6). DNase I was at: 5 ng (lanes 1 and 2), 12.5 ng (lanes 6 and 8), 25 ng (lanes 5 and 7), 400 ng (lane 3) and 800 ng (lane 4). TE1 and TE2 are indicated by brackets, and footprints 1, 2, 3 and T1 are represented by open boxes.

Based on this size (62 kDa, Fig. 5A) and on the specific activity of this fraction (100 000 units/mg, Table I), the purity of active TBF1 binding protein was 10%. Consistent with this estimate, a protein of approximately this size and in this proportion was detected after SDS–PAGE gel electrophoresis of the most highly purified TBF1 (data not shown).

#### TBF1 binds independently to TE1, TE2 and the intron

Like the heat-treated Kc0 extract, purified TBF1 protected the T1 site in TE1 from DNase I digestion and enhanced digestion immediately upstream of T1 (Fig. 6, lane 2). Furthermore, UV cross-linking experiments demonstrated that TE1 alone is sufficient for TBF1 binding. A synthetic TE1 site was both an effective binding substrate for TBF1 (Fig. 5B, lanes 1 and 2) and an efficient binding competitor (Fig. 5B, lanes 3 and 4). Therefore, TBF1 can bind to TE1 independently of TE2 and the intron.

When the TBF1 concentration was increased five-fold, a footprint was observed in TE2 (T2, –111 to –93; Fig. 6, lane 3). In addition to revealing this footprint, the higher TBF1 concentration completely protected T1 and altered the DNase I cleavage pattern at 4 nucleotides between T1 and T2. Binding to T2 persisted when TE1 was replaced by another sequence (Fig. 6, lane 6). At both high and low concentrations of TBF1, a footprint (Tmut) was detected in the DNA used to replace TE1. However, this new footprint was shorter and considerably weaker than the T1 footprint, consistent with independent binding of TBF1 to T2. These observations suggest that TBF1 binds T2 with a lower affinity than and independently of T1. Alternatively, the T2 footprint may be produced by another protein that co-purifies with TBF1. Evidence supporting the first possibility was obtained using UV cross-linking assays with heat-treated Kc0 extract. The presence of 9-fold excess unlabeled T2 site had no effect; however, a much larger excess substantially reduced TBF1 binding to radiolabeled TE1 (Fig. 5B, lanes 1, 2, 5 and 6 and data not shown). That the T2 competitor had a less dramatic effect than TE1 (Fig. 5B, lanes 3 and 4) is consistent with requiring a higher protein concentration for the T2 footprint. We conclude that TBF1 binds T2 independently of TE1 and the intron.

TBF1 footprints were also observed in the intron in the absence of TE1 and TE2. Two of these, footprints T3 (+409 to +431) and T4 (+509 to +513), are shown in Figure 6 (lane 8), and a third was observed further upstream (T5; +378 to +390, data not shown). A potential consensus binding sequence of 5'-CAG-TTATCGGT-3' as found by inspecting the sequence of all TBF1 footprints (Fig. 7). This consensus suggests that TBF1 is likely to be a novel protein because the sequence has not been previously reported (24–26).

Surprisingly, in the UV cross-linking assay, a DNA segment (+409 to +446) containing T3, a close match to the TBF1 consensus, did not significantly compete with TE1 for TBF1

**Table I.**

	FP activity (units)	Total protein (mg)	Specific activity (units/mg)	Yield (%)	Purification (-fold)
Extract	2600	250	10.4		
Heat-treated	2600	165	16	100	1.5
DNA affinity—cycle I	800	0.108	7400	31	700
DNA affinity—cycle II	400	0.004	100 000	15	9600

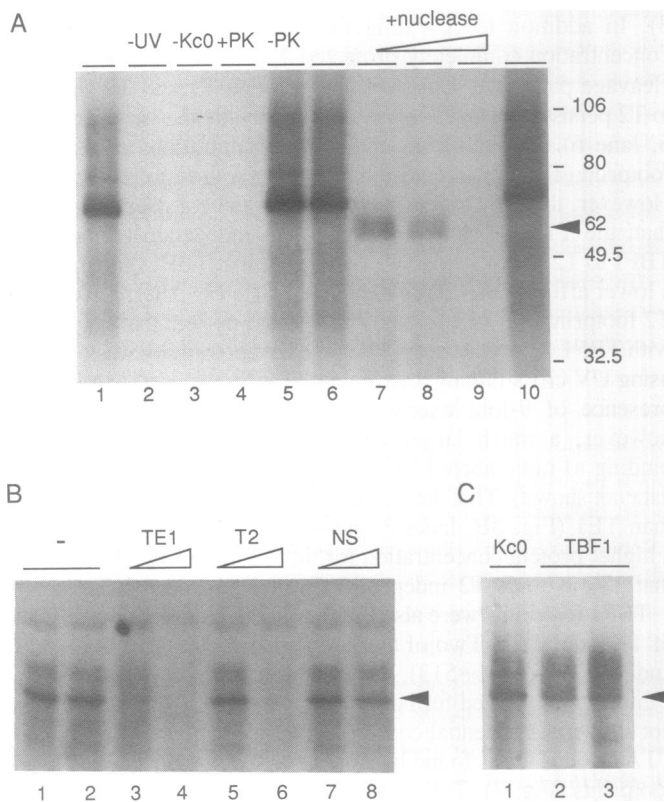
binding (data not shown). This occurred despite the fact that T3 gave a strong DNase I footprint (Fig. 6). A possible explanation is that the T3 oligonucleotide competitor is too short and produces, for example, a secondary structure not occurring in the longer DNA fragment used for footprinting experiments. Alternatively the consensus may be incomplete. Finally, the T3 footprint may be produced by another protein that co-purifies with TBF1. We favor the first or second possibility because the similarity between T3 and T1 sequences suggests that a single protein produced both footprints (Fig. 7).

In summary, TBF1 binds to independent sites in TE1, TE2 and perhaps the intron. This raises the possibility that TBF1 influences transcription from all three  $\alpha 1t$  regulatory regions. We attempted to test this hypothesis by supplementing nuclear extract depleted of TE1-binding proteins with purified TBF1 (27,28). Purified TBF1 had no significant effect on site-dependent *in vitro* transcription of  $\alpha 1t$  (data not shown), perhaps because

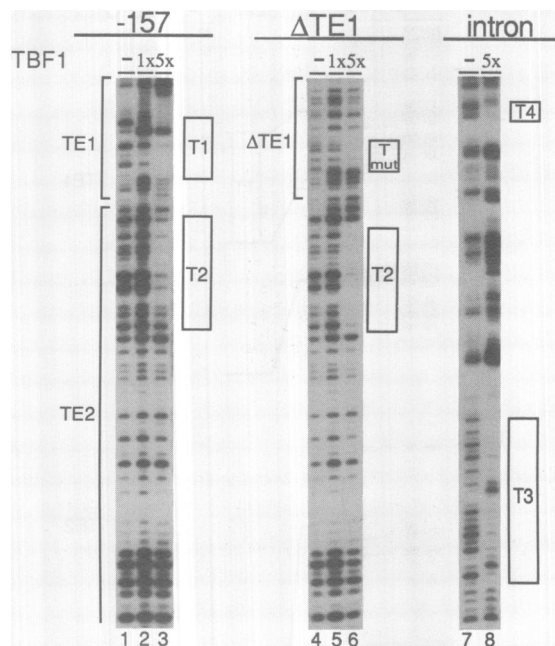
the transcriptional activity of the protein was destroyed during the purification.

**Footprints from the purified proteins and the transcription extract coincide**

The sum of the footprints produced separately by GAGA and TBF1 proteins is essentially the same as the footprints produced by the Kc0 extract (Fig. 3). These results suggest that the two proteins can bind to TE1 and TE2 simultaneously, despite the fact that the G1 and T1 footprints are separated by only a single nucleotide. To test this possibility directly, both GAGA factor



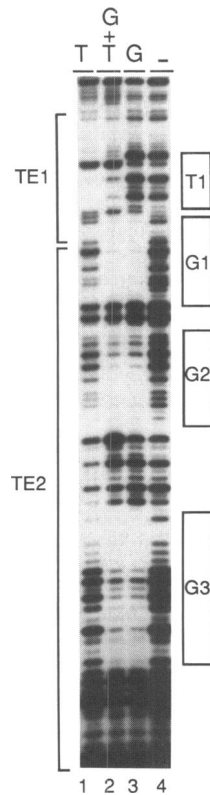
**Figure 5.** UV cross-linking of TBF1 and TE1. (A) Radiolabeled TE1 from a cross-linking reaction with heat-treated Kc0 extract (lane 1, 6 and 10), without UV exposure (lane 2), without extract (lane 3), or with extract pre-treated with (lane 4) or without (lane 5) proteinase K. After cross-linking, reactions in lanes 7, 8 and 9 were treated with 0.15, 0.30 and 0.60 units of micrococcal nuclease, respectively, and 0.12, 0.23 and 0.46 units of DNase I, respectively. Molecular weight standards are indicated on the right, and the arrow shows the 62 kDa product in lane 7. (B) Radiolabeled TE1 was incubated with heat-treated Kc0 extract in the absence (lanes 1 and 2) and presence of the following unlabeled oligonucleotides: TE1 (lanes 3 and 4), T2 (lanes 5 and 6), and nonspecific (NS, lanes 7 and 8). Reactions contained 9-fold (lanes 3, 5 and 7) or 1700-fold (lanes 4, 6 and 8) excess competitor. The arrow on the right indicates the position of cross-linked TBF1. (C) Radiolabeled TE1 cross-linked with heat-treated Kc0 extract (lane 1) or with purified TBF1 in the presence (lanes 1 and 3) or absence (lane 2) of poly [(dI:dC):(dI:dC)] as nonspecific competitor. The arrow on the right indicates the predominant cross-linked band.



**Figure 6.** DNase I footprinting of TBF1. Footprint experiments used approximately 10 femoles of a radiolabeled  $\alpha 1$  fragment from -157 to +39 (lanes 1 to 3), a fragment from -157 to +39 containing the TE1 substitution mutation,  $\Delta$ TE1 (lanes 4 to 6), and an intron fragment from +333 to +696 (lanes 7 and 8). Reactions contained 0 ng (lanes 1, 4 and 7), 5 ng (lanes 2 and 5) or 25 ng (lanes 3, 6 and 8) of the 1 M fraction of purified TBF1, second cycle (Table I). TE1, TE2 and  $\Delta$ TE1 are indicated by the brackets, and TBF1 (T) footprints are represented by boxes.

Region	Sequence	Match
T1	<u>caGTTATCGGT</u> TAT gtcaatagccaata	11/11
T3	<u>caGTTATCGGT</u> CGAAGTATCATGG gtcaatacgcagcttcatagtacc	9/11
T5	<u>tca</u> ttaatcttaa AGTAATTAGAAAT	6/11
T4	<u>ca</u> gtataccat gTCATAtggta	6/11
T2	GAGAGGTAATGGCA <u>CTGG</u> ctctccattaccgtgaacc	5/11

**Figure 7.** Sequence alignment of TBF1 binding sites. Footprint regions are listed in descending order of approximate footprint strength. In the sequences, footprints are capitalized and the best match to the consensus (CAGTTATCGGT) is underlined. Except for T3, the upper strand is the coding strand.



**Figure 8.** Footprints of both TBF1 and GAGA factor. DNase I footprinting analysis with the  $-157$  to  $+39$  radiolabeled fragment ( $\sim 6$  fmole). Reactions contains TBF1 from the  $0.2$  M fraction, first cycle,  $1.7$  footprint units (lanes 1 and 2); purified GAGA factor,  $120$  ng (lanes 2 and 3); or no protein (lane 4).

and TBF1 were included in a single binding reaction and assayed by DNase I footprinting. G1, G2, G3 and T1 footprints were observed, and TBF1 binding did not alter the GAGA footprints (Fig. 8, lanes 2 and 3). However, GAGA protein caused a small but reproducible decrease in TBF1 binding to T1 (Fig. 8, lanes 1 and 2). It is likely that GAGA factor binding eliminates TBF1 binding at T2, since this site overlaps GAGA sites G1 and G2 (Fig. 2). We conclude that purified GAGA factor and TBF1 can bind simultaneously to TE2 and TE1, respectively. Further, the sum of these footprints coincide with those produced by the transcription extract, indicating that GAGA factor binding to TE2 and TBF1 binding to TE1 predominates *in vitro*.

## DISCUSSION

We have identified two proteins that bind to the three  $\alpha It$  regulatory regions. Germline transformation experiments reported in the preceding article showed that these regions have essentially identical effects on the ubiquitous expression of  $\alpha It$  (11). Each region is necessary to obtain position-independent expression in all tissues throughout development and, together with the core promoter, they are sufficient for this expression. Moreover, each region stimulates transcription *in vitro* suggesting that they are positive regulators. One of the binding proteins identified is the GAGA factor. It binds to two of the regulatory regions, TE2 and the intron. The second protein is a new factor, TBF1, which binds all three regions. The sum of the TE1 and TE2 footprints

produced by the combination of these two purified proteins is identical to footprints produced by a nuclear extract that requires these regions for efficient transcription of  $\alpha It$ . These observations suggest that the GAGA factor and TBF1 activate  $\alpha It$  transcription *in vitro* and may be either the predominant or sole regulators of the ubiquitous expression of  $\alpha It$  *in vivo*.

### GAGA factor and TBF1 are likely to be activators of $\alpha It$ transcription

Several observations suggest that the GAGA factor mediates the positive effect of TE2 on transcription. First, we have shown that purified GAGA factor binds TE2. Furthermore, not only do the GAGA footprints in TE2 predominate in experiments done with transcriptionally competent extract, but they also occupy a substantial portion of TE2 (46 of 68 bp; Figure 3). Second, GAGA factor is known to be an activator of the transcription of many genes. *In vivo* experiments indicate that its binding sites activate the *engrailed*, *hsp26*, *hsp70*, *actin 5C* and *ultrabithorax* genes (7,13,29–33). *In vitro*, GAGA factor activates transcription by counteracting the repressive effects of histone H1, a mechanism called antirepression (14,34,35). Both GAGA factor and histone H1 are present in the nuclear extracts used in our transcription assays. Importantly, the 5' ends of one of the GAGA footprints in TE2 (G2;  $-120$  to  $-108$ ) and of the sequence responsible for the majority of the *in vitro* stimulatory effect of TE2 [ $-120$  to  $-114$ , (11)] coincide. Together, these observations suggest that the GAGA factor mediates the positive effects of TE2.

Several findings indicate that TBF1 is likely to mediate the positive effect of TE1 on transcription. Like the footprints of GAGA factor in TE2, the TBF1 footprint in TE1 is the predominant footprint produced by transcriptionally competent extract and occupies a substantial portion of TE1 (12 of 29 bp; Figure 3). Further, the 5' end of this footprint (T1;  $-84$  to  $-73$ ), colocalizes with those  $\alpha It$  sequences outside of the core promoter that have the largest effect on transcription *in vitro* [ $-89$  to  $-77$ ; (11)]. Since this effect is positive, we propose that TBF1 is an activator. In contrast, the TBF1 footprint in TE2 does not coincide with sequences that stimulate transcription and is not observed in footprinting assays with transcriptionally competent extract. For these reasons we speculate that TBF1 bound to TE1 stimulates transcription, but TBF1 bound to TE2 may play a different role or may stimulate transcription only in some tissues.

The GAGA and TBF1 proteins may also be responsible for the strong regulatory influence of the intron that was observed in germline transformation experiments (11). However, with the intron there is no useful correlation between the binding sites and *in vitro* effects on transcription (11). We speculate that these GAGA and TBF1 binding sites have little activity *in vitro* because of their large distance from the promoter (at least 378 bp). Under the conditions used in these *in vitro* experiments, regulatory effects are typically limited to DNA elements within approximately 200 bp of the promoter, perhaps due to the absence of chromatin structures on the template (35,36).

### Roles for GAGA factor and TBF1 in regulating ubiquitous $\alpha It$ expression

Analyses of the *Drosophila* heat shock genes provide insight into the potential roles of GAGA factor and TBF1 in regulating  $\alpha It$  transcription. Studies of the *hsp26* heat shock gene indicate that GAGA binding sites are necessary for full levels of induction and for open chromatin structures (30). In contrast, the heat shock

elements in *hsp26* are necessary for induction, but have little or no influence on chromatin structure. It was proposed that binding by GAGA protein opens chromatin thereby allowing the heat shock factor, a transcription activator, access to the heat shock elements in the DNA. This interpretation is strongly supported by *in vitro* experiments which demonstrated that the GAGA protein can activate transcription by histone H1 antirepression and can displace bound histones in the presence of ATP (35,37). In further support, a mutation in the gene encoding the GAGA factor enhances position effect variegation [(38); F.Karch, personal communication]. The second function of GAGA binding sites, and by inference of GAGA factor itself, is that they are necessary for the pausing of RNA polymerase II shortly after it has begun transcription of *hsp70* (32). Poised polymerase has also been observed in the *hsp26* gene (39) and in the  $\alpha 1t$  gene (J.T.Lis, personal communication). From these studies, it appears that GAGA factor plays a key role in preparing the heat shock genes and presumably other genes for activation, both by remodeling chromatin and by maintaining a transcriptionally engaged but paused polymerase in the promoter region (30,32).

It seems reasonable to speculate that the GAGA binding sites in  $\alpha 1t$  and the heat shock genes have similar functions. For this reason we propose that GAGA factor acts in TE2 and the intron to open chromatin, making DNA in TE1, TE2 and the intron accessible to activator proteins. Since the  $\alpha 1t$  gene also has a paused polymerase and the GAGA factor is necessary for pausing in *hsp70*, the open chromatin and polymerase pausing may be linked functions (30). We presume that TBF1 functions as an activator of  $\alpha 1t$ , analogous to the role proposed for the heat shock factor in *hsp70*. Like the heat-induced heat shock factor, TBF1 may be a ubiquitous activator or alternatively one of a family of activators that operates at TBF1 binding sites in open chromatin. A continuously open chromatin structure and paused polymerase may be an essential feature of genes like  $\alpha 1t$  that are activated constitutively or like *hsp70* that can be rapidly induced in almost all tissues at any stage of development.

Since mutation of TE1, TE2 or the intron causes chromosomal position effects (11), we speculate that the genome is punctuated with tissue-specific regulatory elements that can compete with activation directed by TBF1 and GAGA factor. Further,  $\alpha 1t$  may require as few as two ubiquitous factors with different functions, but together directing position-independent expression in all tissues throughout development.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the generous gift of purified GAGA factor from A.TenHarmsel and M.Biggin and the helpful discussions with Mark Biggen and Ranjan Sen. We thank Lark Claassen, Marie Lossky, Wenqian An, Yue Wang, Sayeon Cho and Bob Kolouch for valuable criticism of the manuscript. This work was supported by Public Health Service grants GM21626 and GM31234 from the National Institutes of Health.

## REFERENCES

- Means, A.L. and Farnham, P.J. (1990) *Mol. Cell. Biol.*, **10**, 653–661.
- Stapleton, G., Somma, M.P. and Lavia, P. (1993) *Nucleic Acids Res.*, **21**, 2465–2471.
- Chang, M., Naik, S., Johanning, G.L., Ho, L. and Patel, M.S. (1993) *Biochemistry*, **32**, 4263–4269.
- Luo, X. and Kim, K.-H. (1990) *Nucleic Acids Res.*, **18**, 3249–3254.
- Boyer, T.G. and Maquat, L.E. (1990) *J. Biol. Chem.*, **265**, 20524–20532.
- McBurney, M.W., Sutherland, L.C., Adra, C.N., Leclair, B., Rudnicki, M.A. and Jardine, K. (1991) *Nucleic Acids Res.*, **19**, 5755–5761.
- Chung, Y.-T. and Keller, E.B. (1990) *Mol. Cell. Biol.*, **10**, 206–216.
- Kalfayan, L. and Wensink, P.C. (1982) *Cell*, **29**, 91–98.
- Natzle, J.E. and McCarthy, B.J. (1984) *Dev. Biol.*, **104**, 187–198.
- Matthews, K.A., Miller, D.F.B. and Kaufman, T.C. (1989) *Dev. Biol.*, **132**, 45–61.
- O'Donnell, K.H., Chen, C.-T. and Wensink, P.C. (1994) *Mol. Cell. Biol.*, **14**, 6398–6408.
- Soeller, W.C., Oh, C.E. and Kornberg, T.B. (1993) *Mol. Cell. Biol.*, **13**, 7961–7970.
- Soeller, W.C., Poole, S.J. and Kornberg, T. (1988) *Gene Dev.*, **2**, 68–81.
- Biggin, M.D. and Tjian, R. (1988) *Cell*, **53**, 699–711.
- Parker, C.S. and Topol, J. (1984) *Cell*, **36**, 357–369.
- Heberlein, V., England, B. and Tjian, R. (1985) *Cell*, **41**, 965–977.
- Bradford, M.M. (1976) *Annals of Biochem.*, **72**, 248–254.
- Kadonaga, J.T. and Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5889–5893.
- Galas, D. and Schmitz, A. (1978) *Nucleic Acids Res.*, **5**, 3157–3170.
- Burtis, K.C., Coschigano, K.T., Baker, B.S. and Wensink, P.C. (1991) *EMBO J.*, **10**, 2577–2582.
- Coschigano, K.T. and Wensink, P.C. (1993) *Gene Dev.*, **7**, 42–54.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V. and Ueda, H. (1987) *Science*, **238**, 1247–1253.
- Biggin, M.D. and Tjian, R. (1989) *Trends Genet.*, **5**, 377.
- Faisst, S. and Mayer, S. (1992) *Nucleic Acids Res.*, **20**, 3–26.
- Wingender, E. (1990) *Adv. in Mol. Genet.*, **4**, 95–108.
- Perkins, K.K., Dailey, G.M. and Tjian, R. (1988) *EMBO J.*, **7**, 4265–4273.
- O'Donnell, K.H. (1993) *Regulation of ubiquitous expression: the  $\alpha 1$ -tubulin gene of *Drosophila melanogaster**, PhD Thesis, Brandeis University.
- Glaser, R.L., Thomas, G.H., Siegfried, E., Elgin, S.C.R. and Lis, J.T. (1990) *J. Mol. Biol.*, **211**, 751–761.
- Lu, Q., Wallrath, L.L., Granok, H. and Elgin, S.C.R. (1993) *Mol. Cell. Biol.*, **13**, 2802–2814.
- Lu, Q., Wallrath, L.L., Allan, B.D., Glaser, R.L., Lis, J.T. and Elgin, S.C.R. (1992) *J. Mol. Biol.*, **225**, 985–998.
- Lee, H.-S., Kraus, K.W., Wolfner, M.F. and Lis, J.T. (1992) *Gene Dev.*, **6**, 284–295.
- Laney, J.D. and Biggin, M.D. (1992) *Gene Dev.*, **6**, 1531–41.
- Kerrigan, L.A., Croston, G.E., Lira, L.M. and Kadonaga, J.T. (1991) *J. Biol. Chem.*, **266**, 574–582.
- Croston, G.E., Kerrigan, L.A., Lira, L.M., Marshak, D.R. and Kadonaga, J.T. (1991) *Science*, **251**, 643–649.
- Laybourn, P.J. and Kadonaga, J.T. (1992) *Science*, **257**, 1682–1685.
- Tsukiyama, T., Becker, P. and Wu, C. (1994) *Nature (London)*, **367**, 525–532.
- Dorn, R., Szidonya, J., Korge, G., Sehnert, M., Taubert, H., Archoukieh, E., Tschiersch, B., Morawietz, H., Wustmann, G., Hoffmann, G. and Reuter, G. (1993) *Genetics*, **133**, 279–290.
- Rougvie, A.E. and Lis, J.T. (1990) *Mol. Cell. Biol.*, **10**, 6041–6045.