The EMBO Journal vol.14 no.19 pp.4763-4772, 1995

The T protein encoded by *Brachyury* is a tissue-specific transcription factor

Andreas Kispert[1], Birgit Koschorz[2] and Bernhard G.Herrmann[2,3]

Max-Planck-Institut für Entwicklungsbiologie, Abteilung Biochemie, Spemannstrasse 35/II, D-72076 Tübingen, Germany
[1]Present address: The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
[2]Present address: Max-Planck-Institut für Immunbiologie, Stübeweg 51, D-79108 Freiburg, Germany

[3]Corresponding author

The mouse *Brachyury* (*T*) gene is required for differentiation of the notochord and formation of mesoderm during posterior development. Homozygous embryos lacking *T* activity do not develop a trunk and tail and die *in utero*. The *T* gene is specifically expressed in notochord and early mesoderm cells in the embryo. Recent data have demonstrated that the T protein is localized in the cell nucleus and specifically binds to a palindrome of 20 bp (the T site) *in vitro*. We show that the T protein activates expression of a reporter gene in HeLa cells through binding to the T site. Thus T is a novel tissue-specific transcription factor. It consists of a large N-terminal DNA binding domain (amino acids 1–229) and two pairs of transactivation and repression domains in the C-terminal protein half. T can also transactivate transcription through variously oriented and spaced T sites, a fact that may be relevant in the search for genes controlled by T protein and important in mesoderm development.

Keywords: *Brachyury*/embryogenesis/mesoderm/mouse/transcription factor

Introduction

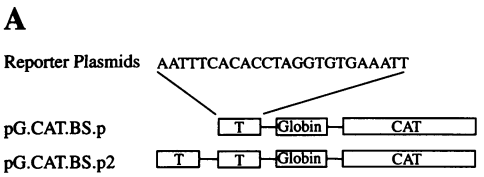
The *Brachyury* (*T*) gene was identified in 1927 through the effect of a loss of function mutation on embryonic development (Dobrovoskaia-Zavadskaia, 1927). Heterozygotes do not complete tail formation, resulting in a short tail phenotype of the adult mouse, while homozygous embryos die early during gestation (Chesley, 1935). The latter lack the trunk and tail as a result of an arrest of mesoderm formation in the primitive streak during posterior development. They also lack a notochord due to disintegration of the head process/notochordal plate, which is formed initially (for review see Herrmann and Kispert, 1994).

Cloning and analysis of the *T* gene revealed a highly restricted expression pattern (Herrmann *et al.*, 1990; Wilkinson *et al.*, 1990; Herrmann 1991). In wild-type embryos *T* is transcribed in the head process and notochord and in the nascent and early migrating mesendoderm. *T* has not been detected in adult mouse tissues. An immunohistochemical analysis of mouse embryos showed that T protein is undetectable after formation of the embryonic axis has been completed (Kispert and Herrmann, 1994). The phenotype of mutant embryos combined with the expression analyses suggest that T protein plays two roles; it is indispensable for differentiation of notochord cells and it is required in formation of posterior mesoderm. In the latter T is probably important for induction, rather than differentiation, of mesoderm (Herrmann and Kispert, 1994).

Recently T homologues have been isolated from the frog *Xenopus laevis* (Smith *et al.*, 1991), the zebrafish *Danio rerio* (Schulte-Merker *et al.*, 1992), the chick (Kispert *et al.*, 1995), the ascidian *Halocynthia roretzi* (a urochordate; Yasuo and Satoh, 1993, 1994) and the insect *Drosophila melanogaster* (Kispert *et al.*, 1994). A comparison of the amino acid sequences revealed a strong conservation of the N-terminal half of these proteins. In addition, the T expression pattern of all vertebrate embryos is conserved, while the ascidian embryo transcribes *T* only in the notochord. The fly embryo, devoid of a notochord, requires the *T*-related gene for differentiation of the hind gut, suggesting that the insect hind gut and the notochord of chordates may have a common origin in evolution.

To determine the function of the T protein in the notochord and in mesoderm development we and others have identified its subcellular location. T is situated in the nucleus, suggesting a role in the control of gene transcription (Schulte-Merker *et al.*, 1992; Cunliffe and Smith, 1994; Kispert and Herrmann, 1994). To substantiate this point we have utilized a strategy for the selection of target sites *in vitro* and have isolated DNA fragments specifically binding to T protein (Kispert and Herrmann, 1993). This established that T protein is indeed a DNA binding protein and revealed the sequence specifically recognized by T. The T consensus site identified *in vitro* is a nearly perfect palindrome of 20 bp. *In vitro* experiments showed that both half sites are required for binding of a T protein monomer. The DNA binding activity is conferred by the N-terminal half (amino acid residues 1–229) of the T protein, termed the T domain, which was found to be strongly conserved in evolution.

In this work we have investigated whether the T protein is able to activate *in trans* expression of a reporter gene containing a T binding site upstream of a minimal promoter with a TATA box. We demonstrate that T is indeed a transcription factor. A detailed analysis of the C-terminal half of T protein identified two positively (transactivation) and two negatively (repression) acting domains. The transactivation domains work in a heterologous context. We also show that T is able to transactivate the transcription of target genes not only from the palindromic target site, but to various extents also from differently spaced directly repeated and inverted target half sites. This finding is important with respect to the search for genes controlled by T protein.



B

Reporter Plasmids	Expression Plasmids	Fold Activation
pG.CAT.BS.p	pSG5.T	17 +/-4
	pSG5	1
pG.CAT.BS.p2	pSG5.T	79 +/-4
	pSG5	1

Fig. 1. Expression of the *T* gene in HeLa cells transactivates a reporter gene construct via the T binding site BS.p. (**A**) Schematic representation of T reporter plasmids and T expression plasmids used in the co-transfection experiments. Either one (pG.CAT.BS.p) or two copies (pG.CAT.BS.p2) of the perfectly palindromic T binding site BS.p (T) were placed upstream of the minimal promoter of the rabbit β-globin gene which confers only a very low basal activity to the CAT reporter gene. In the expression plasmid pSG5.T the promoter of the immediate early gene of SV40 drives T expression. The parental control plasmid pSG5 contains a multiple cloning site (MCS) and a polyadenylation signal [poly(A)]. T, palindromic T binding site; globin, β-globin minimal promoter; CAT, chloramphenicol acetyltransferase. (**B**) HeLa cells were co-transfected either with the reporter plasmid pG.CAT.BS.p2 or with pG.CAT.BS.p and the expression plasmid pSG5.T or pSG5, as described in Materials and methods, and assayed for CAT activity.

Results

The T protein acts as a transcriptional activator

We have recently shown that the protein encoded by the *Brachyury* (*T*) gene binds specifically to the 24 bp palindrome BS.p *in vitro* (Kispert and Herrmann, 1993). To investigate whether the T protein is a regulator of transcription we developed a co-transfection assay in HeLa and Cos7 cells. The *T* cDNA, pme75, was subcloned into the eukaryotic expression vector pSG5, downstream of the promoter of the immediate early gene of SV40. This allows constitutive expression of T protein in HeLa or Cos7 cells. The T expression plasmid is termed pSG5.T. Two reporter plasmids, pG.CAT.BS.p and pG.CAT.BS.p2, containing one or two copies respectively of the palindromic T binding site BS.p upstream of the rabbit β-globin minimal promoter and the coding sequence for the bacterial chloramphenicol acetyltransferase (CAT) gene in the plasmid pG.CAT were constructed (Figure 1A). The pSG5 plasmid without the *T* cDNA served as a control for basal CAT activity in the absence of T protein. Either reporter plasmid was co-transfected into HeLa cells together with pSG5 or pSG5.T. Cytoplasmic extracts were prepared 48 h after transfection and assayed for CAT enzyme activity. Both CAT reporter genes were indeed expressed in cells co-transfected with pSG5.T, but not in cells receiving the control plasmid pSG5. One BS.p binding site resulted in a 17-fold activation of CAT expression, while two BS.p sites led to a 79-fold activation over the control (Figure 1B). These results show that the T protein can act as a transcription factor upon the palindromic binding site BS.p. The presence of a second binding site leads to a more than additive increase in transcriptional activation.

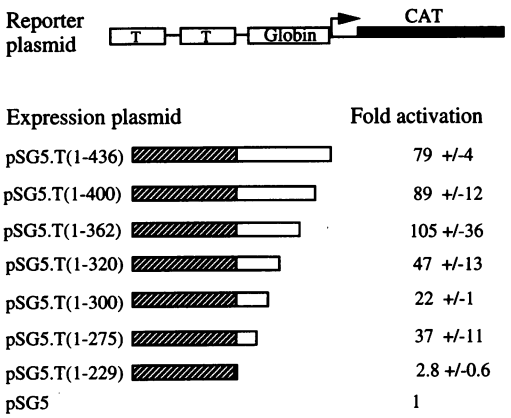


Fig. 2. Deletion constructs indicate a complex domain structure in the C-terminal half of the T protein. Various T deletion constructs were tested by co-transfection with the pG.CAT.BS.p2 reporter plasmid into HeLa cells. CAT enzyme activity was determined as described in Materials and methods. The activation is given as fold activation over the basal value obtained with the non-recombinant pSG5 vector. The hatched box indicates the DNA binding domain of T; T, palindromic T binding site; globin, β-globin minimal promoter; CAT, chloramphenicol acetyltransferase.

Deletion analysis suggests a complex domain structure in the C-terminal half of T protein

To identify the domain(s) of the T protein required for transcriptional activation we co-transfected constructs encoding the full-length or truncated versions of the T protein and reporter plasmid pG.CAT.BS.p2, harbouring two T binding sites, into HeLa cells (Figure 2). The expression plasmids encode T proteins successively truncated from the C-terminus to amino acid residue 229. We have previously demonstrated that the domain from residues 1 to 229, the T domain, is necessary and sufficient for binding to the target site BS.p.

Removal of the entire C-terminal half (residues 230–436) of T protein led to an almost complete loss of transactivation of the reporter gene. This indicated that T, like other transcription factors, may have a modular structure of functional domains. The DNA binding domain and the transactivation domain are encoded by different parts of the gene and may function independently from each other.

Successive truncation of the T protein from the C-terminus led to a variation in the activation values of CAT gene expression in the co-transfection experiments. This indicated a complex nature of the domain structure in the C-terminal half of the protein. Truncation of the C-terminus by up to 74 amino acid residues led to an increase in activation values. Similarly, a further truncation from residues 300 to 275 resulted in an increase in transactivation activity. In contrast, a further removal of residues from positions 362 to 320 and of the region from positions 275 to 229 led to a decrease in CAT gene activation. This deletion analysis suggested that two domains in the C-terminal protein half confer transactivation activity, whereas two different domains, at least in the context of

the T protein, are able to repress transcriptional activation and therefore reduce activity of the activation domains.

However, in this analysis two important factors are not considered. The transactivation values may have been influenced by a variation in protein stability or by a partial or complete loss of nuclear localization. Transfection experiments with various pSG5.T deletion constructs and subsequent detection of the protein with anti-T antiserum, which specifically recognizes the N-terminal portion of the protein, did not reveal any obvious differences in protein stability (not shown). In contrast, our data on mapping of the nuclear localization signal (NLS), described below, showed indeed that T proteins with <320 amino acid residues would not exclusively localize to the nucleus. However, the activation value of T(1–275) suggests that this protein is still able to activate transcription (see next section for comparison) and must occur in the nucleus as well as in the cytoplasm (see Figure 6 for comparison). To obtain more accurate results we analysed the domain structure of the C-terminal half of T protein in a heterologous system.

GAL4–T fusion proteins reveal two transactivation and two repression domains in the C-terminal half of the T protein

The N-terminal 147 amino acid residues of the yeast transcription factor GAL4 contain DNA binding activity and a NLS. This domain can activate the transcription of reporter genes in mammalian cells provided it is fused to an appropriate transactivation domain. A multimer of UAS sequences mediates this activity *in cis*. Thus chimerical GAL4(1–147) fusion proteins can serve as a test system for regulatory domains of transcription factors.

We constructed a CAT reporter gene (UAS-G-CAT) controlled *in cis* by a pentamer of UAS upstream of a globin minimal promoter. Subregions of the C-terminal half of T protein (230–436) were expressed in a C-terminal fusion with GAL4(1–147) and tested for their ability to activate expression of the reporter gene in HeLa cells. The data revealed the presence of two transactivation and two repression domains in the C-terminal half of T protein (Figure 3). This is best demonstrated by the strong activation of the reporter gene by the fusion proteins containing the domains T(230–280) or T(313–380). In our experiments the transactivation activity of these domains is ~3 and ~4 times higher than that of the entire C-terminal half of T. In contrast, a sharp decrease in activity is observed when either domain is further enlarged to position 320 [T(230–320)] or 436 [T(313–436)] respectively. This identifies the domains T(230–280) and T(313–380) as two independent transactivation domains. Both activities are attenuated by adjacent repression domains, the regions T(281–320) and T(401–436) respectively. The present analysis does not exactly define the extent of the transactivation domain T(313–380) and the repression domain T(281–320). The overlap between both, residues 313–320, has not been assigned to either or neither activity.

The existence of a repression domain encoded by residues T(401–436) was also revealed by truncated T proteins T(1–400) and T(1–362). However, the GAL4(1–147)–T(230–400) fusion protein behaves differently. Unexpectedly, it has a lower activity than the GAL4(1–147)–T(230–436) fusion protein. This suggests that in the

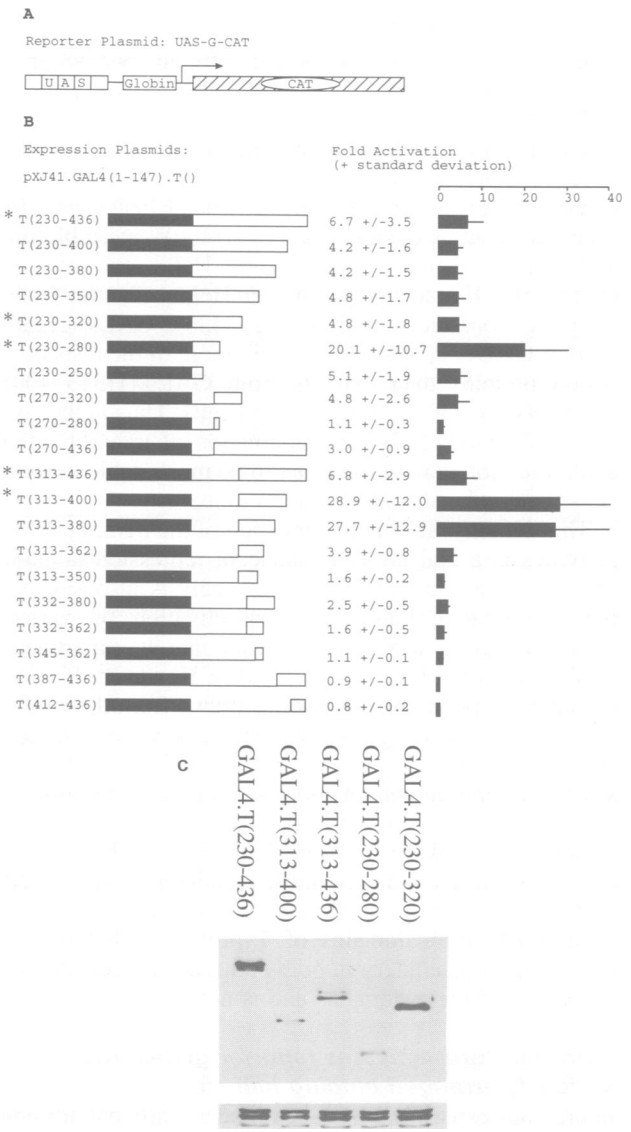


Fig. 3. Fine mapping of the regulatory domains in the C-terminal half of the T protein using GAL4(1–147)–T fusion proteins. (**A**) Schematic representation of the reporter plasmid UAS-G-CAT containing a pentamer of the GAL4 binding site (UAS) and the CAT reporter gene. (**B**) Summary of the transactivation activities of various GAL4(1–147)–T fusion constructs tested in HeLa cell co-transfection assays. The expression constructs are shown schematically. The filled boxes represent the GAL4 DNA binding domain, the open boxes the complete or subregions of the C-terminal half of T protein (drawn to scale). The transactivation values of the various constructs are indicated. A fusion protein of the DNA binding domain of GAL4 with the full-length T protein did not activate transcription, probably due to interference between the two DNA binding domains (not shown). Asterisks indicate the crucial constructs, revealing two transactivation and two repression domains, and the control construct containing the entire C-terminal half of T protein. The marked constructs have also been examined in Cos7 cells. (**C**) Extracts of HeLa cells co-transfected with the indicated expression constructs and the UAS-G-CAT reporter plasmid were separated by electrophoresis, Western blotted and immunochemically stained for GAL4 protein (upper panel). The bands in the lower panel represent loading controls.

context of GAL4(1–147) the presence of the repression domain T(281–320) is alone sufficient to lower the activity of both nearby transactivation domains. On the other hand, it is possible that the various arrangements of binding sites in the two reporter constructs (two sites for T protein,

five sites for Gal4–T fusion proteins) or a difference in protein stability accounts for the different behaviour of transcription factors T(1–400) and GAL4(1–147)–T(230–400).

The effect of the repression domains might also be due to reduced protein stability. To exclude this possibility we determined expression of the five crucial fusion proteins (marked by an asterisk in Figure 3B) by Western blotting and immunochemical analysis of HeLa cell extracts (Figure 3C). Extracts prepared from HeLa cells transfected with the constructs GAL4–T(313–436), GAL4–T(230–320) or GAL4–T(230–436) contain more of the respective fusion proteins than extracts from GAL4–T(313–400)- and GAL4–T(230–280)-transfected cells. Thus in the context of GAL4(1–147) the combination of transactivation and repression domains appears to be more stable in HeLa cells than isolated transactivation domains.

This demonstrates that attenuation of the transactivation activities is a real effect of adjacent repression domains and not an artefact. The Western analysis also suggests that the relative values of transactivation by various GAL4–T fusion proteins measured in our experiments may be influenced by protein stability. However, this does not alter the principal conclusions drawn from the results.

Finally, we investigated whether the behaviour of these fusion proteins was due to artefacts of the HeLa cell system or reproducible in another cell type. Therefore, we tested transactivation activities of the five crucial fusion proteins (marked by an asterisk in Figure 3B) in Cos7 cells. The data confirmed those obtained in HeLa cells (not shown). Thus the interaction of repression and/or transactivation domains of T protein with the basic transcription machinery in both cell types reflects the real properties of these domains and is not due to an artefact.

T protein transactivates reporter genes from variously arranged binding half sites

In previous experiments we identified a nearly palindromic 20mer as the consensus binding site of T protein by target site selection *in vitro*. In this report we have used a perfect 24mer palindrome as a *cis* control element in a reporter gene construct and have shown that T protein transactivates gene expression in HeLa cells through binding to this site (Figure 1). However, genomic target sites of T protein are unknown. Since the *in vitro* target site selection procedure may have favoured the isolation of a palindrome for sterical reasons or due to the relatively small number of nucleotides available for binding (26 bp), we investigated whether T protein could also act by binding to half sites arranged at various distances as direct or inverted repeats. The results show that, indeed, T protein can activate expression of such reporter genes in HeLa cells, but some configurations are preferred (Figure 4). In general, T protein seems to transactivate slightly better through directly repeated dodecamer sites than through inverted repeats. However, the crucial factor is the distance of the dodecamer sites. In our experiments an inverted repeat of the dodecamer sites with a spacer of 24 bp (BS.2-24) gave the best result, while a similar construct with a 21 bp spacer (BS.2-21) had only basal activity. BS.2-24 conferred an up to 2.8-fold higher reporter gene expression than the palindrome BS.p. Only one other construct, BS.1-6, worked slightly better than the palindrome, whereas all

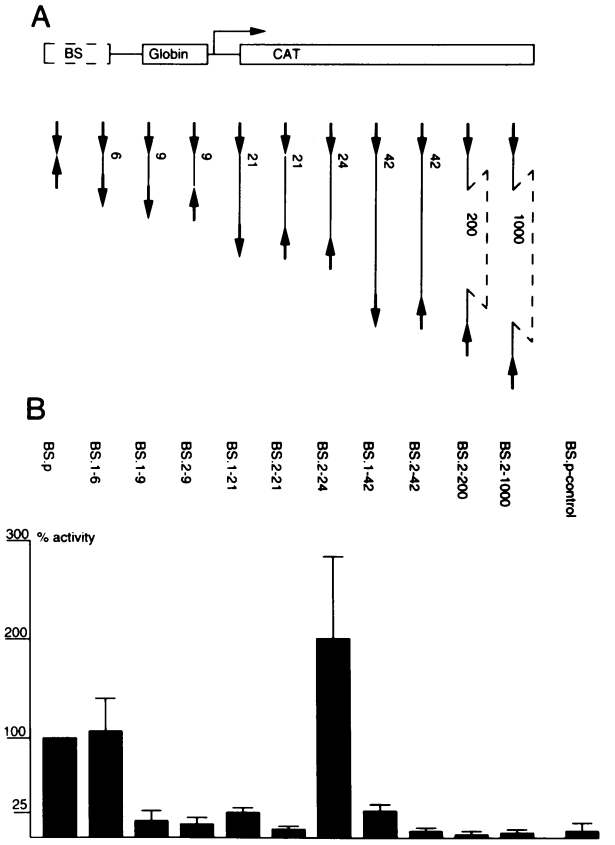


Fig. 4. Analysis of variously spaced and oriented T binding half sites as *cis* control elements. (**A**) Schematic representation of the reporter plasmids tested in HeLa cell co-transfection experiments. T binding sites (BS) inserted upstream of the globin minimal promoter (Globin) and CAT are shown underneath. The arrow represents the orientation of the dodecameric T binding half site AGGTGTGAAATT, bars the spacer between the two half sites and the number above the length of the spacer. (**B**) HeLa cells were co-transfected with the expression plasmid pSG5.T encoding the full-length T protein and with a series of reporter plasmids with various T binding sites. CAT enzyme activity was determined as described in Materials and methods. The activation level of various constructs is shown in a bar diagram relative to that of the reporter plasmid pG.CAT.BS.p harbouring the perfect palindromic T binding site BS.p. To exclude the possibility that polylinker sequences of the donor plasmids of binding sites might confer CAT activity in the reporter constructs the respective regions were cloned into reporter plasmids and tested in parallel (not shown) Control transfection experiments (BS.p-control) were performed with the reporter pG.CAT.BS.p and pSG5, which does not express T protein.

other constructs tested led to much lower gene activation than the palindrome.

Of course, not all possible combinations of target dodecamer orientations and distances could be tested in this experiment. Thus the results presented probably do not identify the optimal configuration of target dodecamers. It is conceivable that multimers of target dodecamers might result in even better transactivation. The results show, however, that the target sites selected *in vitro* are not the only possible sites through which T protein can act and that two spaced dodecamer T sites can mediate even better transactivation of reporter genes by T protein. Therefore, variously spaced and oriented dodecamer T sites may also be expected in the promoters of genomic target genes of T.

To confirm that T protein actually binds to the dodecamer target sites in the reporter constructs BS.1-6

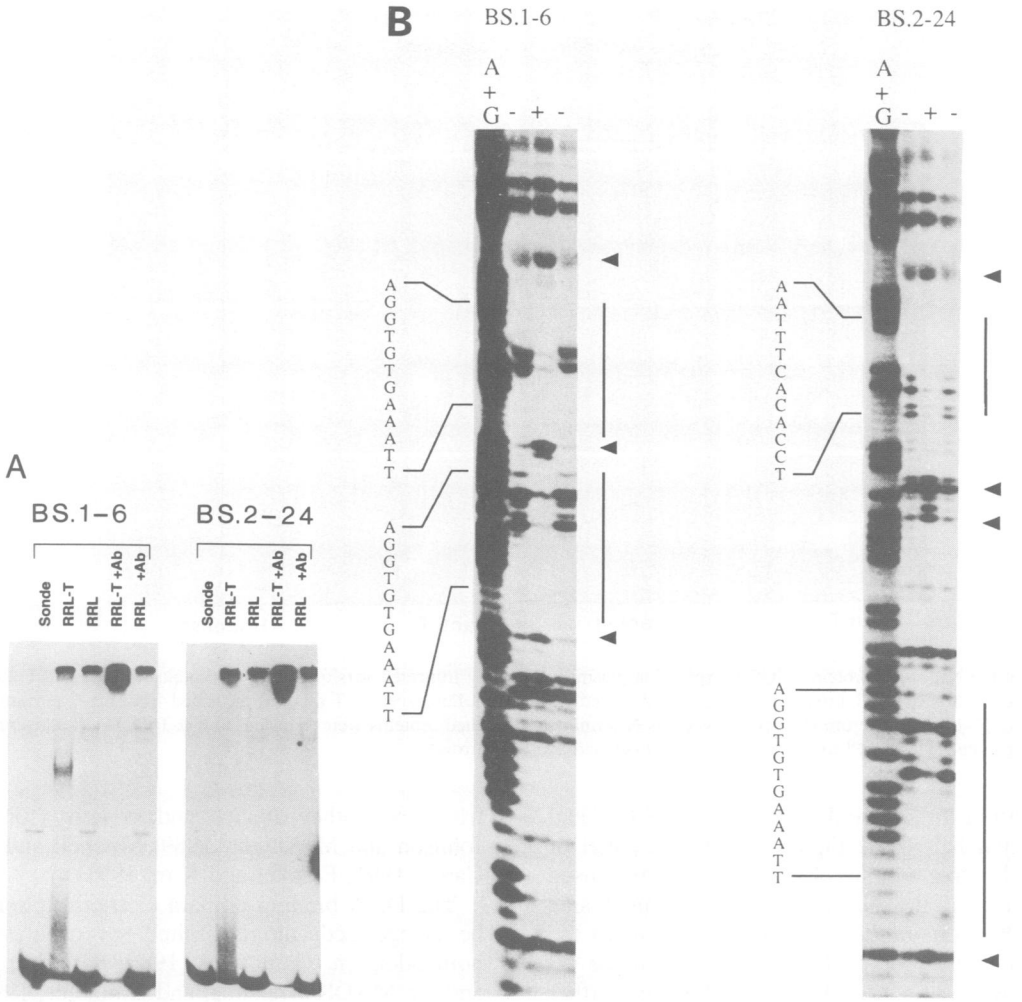


Fig. 5. T protein transactivates upon binding to variously spaced and oriented dodecameric T half sites. (**A**) EMSA of complexes of T protein with the binding sites BS.1-6 or BS.2-24 respectively (see Figure 4). Binding reactions included probe only (Sonde), unprogrammed rabbit reticulocyte lysate (RRL) and rabbit reticulocyte lysate programmed with T RNA encoding the full-length T protein with (+Ab) and without anti-T antiserum α-TC328-420. Addition of the antiserum significantly improves complex stability. (**B**) Characterization of T protein binding by DNase I protection analysis performed on *Xho*I–*Sac*I fragments cut out of the plasmids pBS.1-6 and pBS.2-24 respectively. Fragments of the DNase I digests were co-electrophoresed on a sequencing gel with a chemical cleavage reaction for A and G residues (A+G) of the same fragments. The position of the dodecamer binding sites is indicated on the left. +, DNase I digest of a fragment incubated with *in vitro* translated T protein; –, the fragment was digested after incubation with reticulocyte lysate (without T). Regions protected from DNase I digestion are indicated by bars, hypersensitive nucleotides by arrowheads. T protein binds to both half sites, irrespective of their orientation and spacing.

and BS.2-24 were assayed for specific binding to T protein (Figure 5). In the electrophoretic mobility shift assay (EMSA) a BS.1-6–T protein complex is visible on the gel, whereas a BS.2-24–T protein complex cannot be detected (Figure 5A). It might be that although complexes are formed, they do not focus into a single band or are not sufficiently stable under the *in vitro* conditions used. However, in combination with an antiserum against the C-terminal half of T protein a DNA–T protein–anti-T antibody complex is readily detectable as a supershift. This may be due to the fact that very low mobility complexes are formed which cannot be resolved by the gel and thus appear as one band. On the other hand, we have previously observed that binding of antibody to the DNA–T protein complexes stabilizes binding. This also appears to be the case here, since almost the entire target probe is shifted into the high molecular weight band.

DNase I protection experiments confirmed that T protein binds to the two dodecamer sites in both constructs (Figure 5B). Up to 3 nt adjacent to the dodecamer, that is up to 15 nt in total, were found to be protected from degradation by Dnase I. Interestingly, neighbouring nucleotides become hypersensitive to DNase I, suggesting structural constraints in the DNA–T protein complex that expose nearby DNA regions.

In summary, transactivation of the reporter genes is associated with a specific DNA–T protein interaction. A correlation between the stability of target site–T protein complexes *in vitro* and the transactivation activity *in vivo* is not evident. The stabilization of the complexes by antibody binding rather suggests that *in vivo* an accessory protein might be involved in the activation of reporter gene expression by T protein.

The T protein contains a complex NLS

A series of truncation proteins were tested for translocation to the nuclei of Cos7 cells. The location of the mutated T proteins within the cells was detected with specific

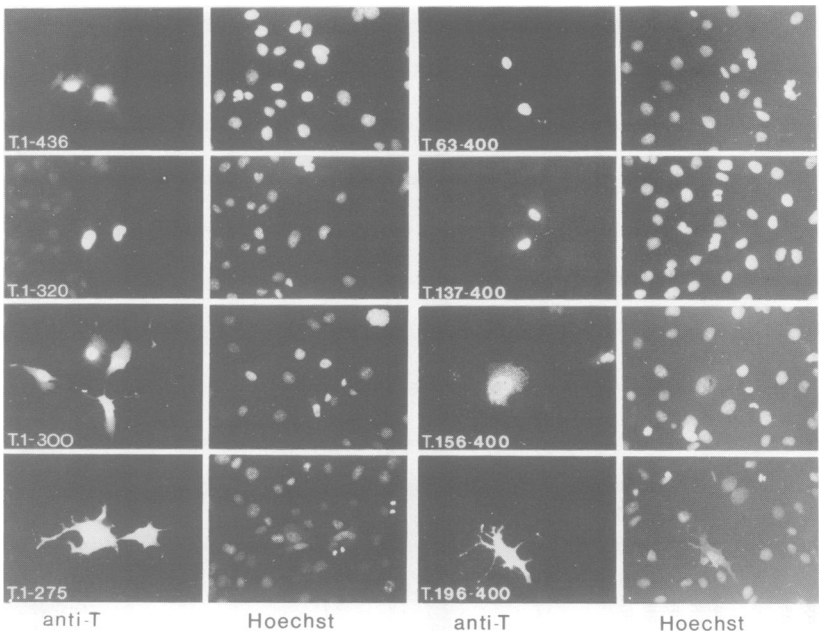


Fig. 6. Mapping of the NLS of T protein. pSG5.T expression plasmids encoding truncated versions of the T protein were transfected into Cos7 cells. The subcellular localization of the T protein was determined by indirect immunofluorescence. T proteins truncated from the C-terminus were detected with the α-TN1-123 antiserum (first row), whereas N-terminally truncated proteins were detected with α-TC328-420 antiserum (third row). Cells were counterstained with DAPI to visualize the nuclei (second and fourth row).

polyclonal antisera directed either against the N-terminal or the C-terminal half of T. Figure 6 shows a panel of results obtained with a number of selected constructs used for fine mapping of the NLS. On the C-terminal side exclusive localization of the truncated T protein to the nucleus is lost in the polypeptide T(1–300), on the N-terminal side when the domain T(137–400) is further truncated to T(156–400). Both polypeptides, T(1–300) and T(156–400), partially stay in the cytoplasm, but are primarily localized to the nucleus. These results suggest the presence of several complex NLSs between residues 137 and 320 which are required in combination for exclusive localization of the T protein to the nucleus. This region overlaps with the C-terminal half of the DNA binding T domain and one each of the N-terminally located transactivation and repression domains.

Discussion

The *Brachyury* (*T*) gene is required in the formation of mesoderm during posterior development of the vertebrate embryo and in the differentiation of the notochord of chordates (for a review see Herrmann and Kispert, 1994), and the hind gut of insects (Kispert *et al.*, 1994). We demonstrate here that the mouse T protein functions as a transcription factor.

T has two transactivation and two repression domains

Transcriptional activators comprise at least two functional features, the DNA binding and the transactivation function(s). Furthermore, they frequently bear a NLS(s) which allows translocation of the protein to the nucleus. DNA binding and transcriptional activation functions reside in different parts of the protein. Due to this modular structure these functions can be separated and studied independently from each other *in vivo* and *in vitro* (for reviews see Johnson and McKnight, 1989; Mitchell and Tjian, 1989; Carey, 1991; Frankel and Kim, 1991).

The DNA binding domain of transcription factors can be categorized into a limited set of classes, e.g. the homeodomain (Qian *et al.*, 1989; Kissinger *et al.*, 1990) and the MADS (Treisman and Ammerer; 1992), ets (Nye *et al.*, 1992; Wasylyk *et al.*, 1992), paired (Treisman *et al.*, 1991; Xu *et al.*, 1995) and forkhead domains (Weigel and Jäckle, 1990; Lai *et al.*, 1991) and others. The members of each class share a high degree of sequence similarity, often across animal and plant phyla, corresponding to a binding motif such as zinc-fingers, helix–loop–helix or others (Harrison, 1991).

We have recently shown that the 229 N-terminal amino acid residues of mouse T protein encode a novel DNA binding domain which we have termed the T domain after the prototypic mouse T protein (Kispert and Herrmann, 1993). This DNA binding domain is highly conserved in the Brachyury homologues of the vertebrates mouse (Herrmann *et al.*, 1990), chick (Kispert *et al.*, 1995), *Xenopus* (Smith *et al.*, 1991) and zebrafish (Schulte-Merker *et al.*, 1992), the insect *Drosophila melanogaster* (Kispert *et al.*, 1994) and the tunicate *Halocynthia roretzi* (Yasuo and Satoh, 1993, 1994). Furthermore, the *Drosophila* gene *omb* encodes a protein with a T domain (Pflugfelder *et al.*, 1992) and a growing family of T-like genes in the mouse and in *Caenorhabditis elegans*, the so-called T box genes, with hitherto unknown developmental functions, employ such a T DNA binding domain (Bollag *et al.*, 1994; Agulnik *et al.*, 1995).

In contrast to the high degree of primary structure conservation in the DNA binding domains, the transcriptional activation functions are characterized by a lack of readily identifiable sequence motifs. Rather, activation functions are often characterized by a distinctive amino

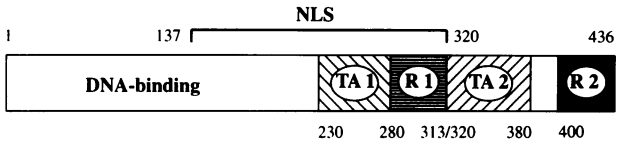


Fig. 7. The domain structure of mouse T protein. TA, transactivation domain; R, repression domain. Numbers indicate amino acid residues in the T sequence. The location of the NLS is indicated by a bracket.

acid composition. Domains rich in acidic residues, in the hydroxylated residues serine, threonine and tyrosine and in proline and glutamine have been described (Hope and Struhl, 1986; Courey and Tjian, 1988; Mermod *et al.*, 1989). In some cases several activation domains can be found in one protein; some function additively, others are redundant (Hollenberg and Evans, 1988; Pani *et al.*, 1992; Hwang *et al.*, 1993; Nagpal *et al.*, 1993).

In this report we developed a co-transfection assay in HeLa and Cos7 cells to investigate the transcriptional activation properties of T. Using a reporter gene construct with the palindromic T binding site BS.p and expression plasmids encoding full-length and truncated versions of T protein we demonstrate that T transactivates a reporter gene upon binding to a target site. The C-terminal half of the T protein mediates these transcriptional activation properties. It contains two alternately arranged transactivation and repression domains. The extents of these domains have been determined relatively precisely. A first activation domain was determined to lie between residues 230 and 280 and a second between residues 313 and 380 (Figure 7). The latter domain is particularly rich in hydroxylated amino acid residues, whereas in the first serine, threonine, glycine and proline residues are prominent. No similarity with other activation domains at the primary structure level has yet been detected. Both activation domains function in a heterologous context as fusions with the DNA binding domain of the yeast transcription factor GAL4. This demonstrates that these activation domains function autonomously, a characteristic of a number of transcription factors (Hollenberg and Evans, 1988; Nagpal *et al.*, 1993).

The deletion analysis also revealed that the activation functions are additive. The loss of one domain does not abolish activation, but decreases activation levels. When stripped of surrounding sequences both domains behave as very strong transcriptional activators. They exceed the activation values of the entire C-terminal half of the T protein several-fold when fused to the heterologous DNA binding domain of GAL4. *In cis*, i.e. in the context of the other residues of the T protein, the activity of these domains is clearly attenuated. The intervening sequence from 281 to 320 and the C-terminal residues from 401 to 436 act as repressing domains. It is not clear, however, whether they only function *in cis*. It seems that at least the C-terminal domain (R2) can repress the already low basal level of the globin promoter further, but a more careful analysis is required to substantiate this point.

Both activation domains show some degree of conservation between the vertebrate T proteins. This observation suggests that a similar domain structure, as described here for mouse T, may also be expected for the chick, *Xenopus* and zebrafish T proteins.

The conservation of sequence and functional motifs might also extend to the NLSs. In our deletion analysis we were able to define the region between amino acid residues 137 and 320 as necessary and sufficient to achieve exclusive nuclear localization of the T protein. Further truncation from either side led to partial cytoplasmic retention of the T protein. Nuclear localization signals have been defined in various proteins and are characterized by a high occurrence of basic amino acid residues within short sequence stretches. In the classical case of the SV40 large T NLS a heptamer containing five basic residues has proven to be necessary and sufficient for nuclear localization (Kalderon *et al.*, 1984). The SV40 large T signal has been referred to as a special form of the more general so-called bipartite NLS. In the latter two basic clusters of amino acid residues are separated by a spacer of 8–10 unrelated residues (Dingwall and Laskey, 1991; see Garcia-Bustos *et al.*, 1991 for a review). The large region of T found to be necessary for exclusive nuclear localization points to the existence of multiple NLSs. Candidates for such signals are represented by the clusters of basic residues in the region of amino acid residues 145–151 (three out of seven) and of amino acid residues 278–286 (three out of nine). In accordance with the functional significance of the NLS the basic residues within these regions are unchanged in all vertebrate T proteins.

The regulatory properties of T protein in the embryo might depend on interaction with accessory proteins

We have recently described the molecular structure of two mutant T alleles, T^c and T^{Wis} (Herrmann and Kispert, 1994), which give rise to T proteins truncated at the C-terminus. According to the data shown here the T^c protein, which is truncated after position 387 due to a reading frame shift, would have lost the C-terminal repression domain and therefore would be expected to confer a higher transactivation activity than wild-type T protein. Our preliminary results suggest that this may indeed be so (data not shown). In the case of the T^{Wis} protein, which is truncated presumably after position 345, the C-terminal transactivation and repression domains (TA2 and R2) should be inactivated, resulting in a somewhat reduced transactivation activity.

In terms of understanding the nature of these mutant proteins it is important that either protein is still able to transactivate a reporter construct in cultured cells to a degree that is comparable with wild-type protein. However, none of them has wild-type T activity in embryos or supports mesoderm formation and notochord differentiation. Thus T protein function in the embryo seems to require the interaction of accessory factors with the transactivation and/or repression domains of T and it is this interaction which might be impaired in the mutants.

In addition, wild-type T protein plays different roles in the notochord and in nascent mesoderm, in correlation with a continuous versus a transient expression in the two cell types (Herrmann and Kispert, 1994). This suggests that T controls different target genes in the two embryonic T expression domains, which is explainable by control of accessability of T binding sites or by interference by or interaction with other regulatory proteins.

In vitro data support the idea that interaction of accessory

proteins with T may be important for proper T protein function. Binding of antibodies to DNA–T protein complexes dramatically improves complex stability *in vitro*. The BS.1-6–T protein complex appears more stable than complex BS.2-24–T, whereas after addition of antibodies both binding sites appear to be equally well bound. In cultured cells BS.2-24 is a much better target than BS.1-6, in that it confers approximately twice as much reporter gene expression as BS.1-6 or the palindrome BS.p.

One possible role of accessory proteins might be formation of a bridge between two T proteins bound to neighbouring target sites. The 4-fold enhancement of transactivation of a reporter gene with two T sites compared with that with one by T protein might be the result of such an interaction. Another possible role might be stabilization of target DNA–T protein complexes. A third possibility might be that the type of accessory protein which interacts with T determines whether T functions as an activator or a repressor of a target gene. T^{Wis} or T^{c} would then be expected to misregulate such target genes due to impaired protein–protein interactions.

Recently Rao (1994) generated *XBra* mutant genes whose functions were studied in the classical *Xenopus* animal cap assay. Whilst C-terminal deletions of *XBra* corresponding to the T^{c} and T^{Wis} proteins still had mesodermalizing activity in this ectopic site, the B304 form of *Xbra* encoding only the 304 N-terminal amino acid residues antagonized wild-type protein and exhibited a neuralizing activity. According to our results on mouse T, B304 would have retained the DNA binding and one transactivation domain, provided *Xbra* has the same domain structure as mouse T. This would suggest that B304 might escape regulatory influences imposed by proteins interacting with the transactivation and/or repression domains of wild-type *Xbra* protein. Thus B304 might misregulate target genes of *Xbra*, the simplest assumption being that it might thereby activate expression of neuralizing genes which would normally be repressed by wild-type *Xbra* during early mesoderm formation.

T activates gene transcription through variously arranged and spaced T half sites

We have demonstrated that T protein can bind and transactivate transcription from various binding sites. Two of the constructs tested mediated a higher activity than the palindrome derived from sequences isolated by binding site selection *in vitro*. These data show that the *in vitro* selection procedure was biased due to the fact that only 26 bp were offered for binding, with a minimum of 20 bp required by T protein (Kispert and Herrmann, 1993).

We have identified one particular target site, BS.2-24, which consists of an inverted repeat of two palindromic half sites separated by a 24 bp spacer, which confers better transactivation activity on T than the palindrome. Most other sites tested, including an inverted repeat with a 21 bp spacer, were less effective as *cis* acting elements. This shows that the spacing of binding half sites is important for transactivation and suggests that genomic target genes may have two or multiple, spaced binding half sites for T as *cis* regulatory elements. It is also conceivable that in different target genes different arrangements and spacings of binding half sites of T are utilized as a means of controlling the degree of transcriptional activation of the target gene. The importance of spacing and orientation of binding half sites has clearly been demonstrated in the cognate sites for the transcription factors of the thyroid/steroid receptor superfamily (Stunnenberg, 1993).

Genomic target genes will have to be isolated to study the properties of T protein and its interactions with DNA and protein factors in the embryo. The availability of target genes is also a prerequisite for a better understanding of the role of T in embryogenesis.

Materials and methods

Plasmids and DNA constructs

BS.p was constructed by self-annealing the oligonucleotide 5′-GGGAATTTCACACCTAGGTGTGAAATTCCC-3′, which represents a 24 bp palindromic fragment flanked by *Sma*I half sites. The plasmid pBS.p was made by inserting one BS.p fragment in the *Sma*I site of pBluescript KS, whereas in the case of pBS.p2 two copies of BS.p were inserted. In order to construct the pBS.1-6 plasmid, two T binding half sites BS.p/2, produced by annealing the oligonucleotides 5′-GGGAGGTGTGAAATTCCC-3′ and 5′-GGGAATTTCACACCTCCC-3′, were inserted in tandem into the *Sma*I site of pKS. Therefore the direct repeat of the AGGTGTGAAATT sequence is separated by a *Sma*I restriction site, a 6 bp spacer. For electrophoretic mobility shift analysis a *Bam*HI–*Hin*dIII fragment from pBS.1-6 was isolated and labelled by filling in recessive ends with [α-^{32}P]dCTP and cold dGTP, dATP and dTTP using the Klenow fragment of *Escherichia coli* DNA polymerase. For the DNase I protection analysis the *Xho*I–*Sac*I fragment was employed. To obtain the pBS.2-24 plasmid, harbouring two binding half sites in inverted direction with a 24 bp spacer, plasmid pBS.p/2 with the BS.p/2 fragment inserted in the *Sma*I site of pKS was opened at the *Eco*RV site and another BS.p/2 fragment inserted. As in the case of pBS.1-6, the *Bam*HI–*Hin*dIII fragment was used as a probe in EMSA, whereas the *Xho*I–*Sac*I fragment was employed in the DNase I protection analysis.

pBS.1-x and pBS.2-x plasmids were constructed in a stepwise manner. First oligonucleotide 5′-CTAGAAGGTGTGAAATTG-3′ was annealed with oligonucleotide 3′-TTCCACACTTTAACCTAG-5′ and the double-stranded fragment with overhanging *Xba*I and *Bam*HI sites was ligated with *Xba*I–*Bam*HI-cut vector pBluescript KSII. In order to introduce a second T binding half site into this construct the plasmid was opened with *Cla*I and *Sal*I and a double-stranded fragment with compatible sites was inserted. In the case of the pBS.1-42 plasmid, having a direct repeat of the half sites, this fragment was produced by annealing oligonucleotide 5′-CGATAGGTGTGAAATTG-3′ with 3′-TATCCACACTTTAACAGCT-5′, in the case of pBS.2-42, which has an inverted repeat of the half sites, oligonucleotides 5′-CGATAATTTCACACCTG-3′ and 3′-TATTAAAGTGTGGACAGCT-5′ were annealed. To obtain pBS.1-9 and pBS.2-9, plasmids that harbour binding half sites with a 9 nt spacer, pBS.1-42 and pBS.2-42 respectively were cut with *Bam*HI and *Cla*I, the recessive ends filled in with the Klenow fragment of *E.coli* DNA polymerase and the fragments re-ligated. In the case of pBS.1-21 and pBS.2-21, plasmids with 21 bp spacing of the half sites, the parental plasmids pBS.1-42 and pBS.2-42 respectively were cut with *Sma*I and *Eco*RV and re-ligated. The spacing in the plasmid with an inverted repeat, pBS.2-42, was increased to ~200 bp by inserting a 200 bp *Eco*RI fragment into the *Eco*RI cloning site of the vector. To separate further the inverted half sites a 1000 bp *Bam*HI fragment was cloned into the *Bam*HI site. The pBP4.T plasmid for *in vitro* transcription of full-length T RNA was described earlier (Kispert and Herrmann, 1993).

The expression vector pSG5.T was constructed by inserting the *Eco*RI fragment of cDNA pme75 (Herrmann *et al.*, 1990), harbouring the full-length *Brachyury* cDNA, into the *Eco*RI cloning site of eukaryotic expression vector pSG5 (Stratagene). For expression of C- or N-terminally truncated T proteins the respective region was amplified from the *T* cDNA pme75 by polymerase chain reaction using primers containing *Eco*RI sites for insertion into the *Eco*RI cloning site of pSG5. In all sense primers an endogenous ATG is adjacent to the *Eco*RI restriction site, in the antisense primers a stop codon was introduced 3′ of the *Eco*RI restriction site. In the N-terminal deletion constructs an *Eco*RI site in the 3′ untranslated region of pme75 was utilized.

The expression vector pXJ40-GAL4(1–147) (Webster *et al.*, 1988; Xiao *et al.*, 1991; Hwang *et al.*, 1993) was used to express chimeric

proteins of the DNA binding domain of GAL4 (residues 1–147) and the full-length or truncated T proteins in experiments aimed to determine the extent of the transactivation domains in the T protein. The pXJ40.GAL4(1–147)–T series of expression constructs was produced by amplifying the respective region of the *T* cDNA pme75 by polymerase chain reaction using primers with *Cla*I and *Bam*HI restriction sites respectively to insert into the respective sites in the pXJ40.GAL4(1–147) vector. The sense primers (with a *Cla*I restriction site) were designed to create an in-frame fusion of T and GAL4, the antisense primers harboured an in-frame stop codon 3′ to the *Bam*HI restriction site. pCMVb (McGregor and Caskey, 1989) was used as a reference plasmid in co-transfection experiments.

All reporter plasmids used in the co-transfection experiments are derivatives of the pG.CAT vector (Klein-Hitpass *et al.*, 1986; Webster *et al.*, 1988). In this vector multiple cloning sites for the insertion of putative enhancer fragments are located upstream of a rabbit β-globin minimal promoter, which confers only basal transcriptional activity. For pG.CAT.BS.p, pG.CAT.BS.p2, pG.CAT.BS.1-6 and pG.CAT.BS.2-24 *Bam*HI–*Hin*dIII fragments were released from plasmids pBS.p, pBS.p2, pBS.1-6 and pBS.2-24 respectively and ligated with the *Bam*HI–*Hin*dIII-cut vector pG.CAT. To test other spaced binding half sites in the reporter gene assay *Xba*I–*Sal*I fragments were isolated from the pBS.1-x and pBS.2-x cloning vectors and inserted into the compatible sites of pG.CAT to create the reporter plasmids pG.CAT.BS.1-x and pG.CAT.BS.2-x respectively.

UAS-G-CAT, the reporter plasmid used in the co-transfection experiments with GAL4–T expression constructs was made by isolating a *Sau*3A fragment containing the pentameric GAL binding site UAS from UAS-tk-CAT (Webster *et al.*, 1988; Hwang *et al.*, 1993) and inserting it into the *Bam*HI cloning site of pG.CAT. This essentially replaces the thymidine kinase promoter by the globin promoter and results in a higher basal activity.

In vitro transcription, in vitro translation and electrophoretic mobility shift assay

In vitro transcription of T RNA and translation in rabbit reticulocyte lysates to obtain full-length T protein, as well as EMSA, were performed as described previously (Kispert and Herrmann, 1993). For antibody shift analysis 1 μl α-TC328-420 antiserum raised against a C-terminal portion of the T protein was added to the samples and incubation continued for another 15 min.

DNase I footprinting analysis

DNase I footprinting analysis was performed exactly as described for the BS.p binding site (Kispert and Herrmann, 1993). The DNA fragments used in this assay were derived from pBS.1-6 and pBS.2-24 respectively. The plasmid was cut with *Xho*I and the recessive ends filled in with [α-^{32}P]dCTP and cold dGTP, dATP and dTTP using the Klenow fragment of *E.coli* DNA polymerase. The fragment with the binding site was released by digestion with *Sac*I, gel purified and isolated using the Mermaid-Kit (Bio101).

Transient transfection, CAT assays and Western analysis

HeLa or Cos7 cells, maintained in Dulbecco's modified Eagle's medium containing 10% fetal calf serum, were transfected using the calcium phosphate precipitation procedure (Graham and van der Eb, 1973). In all experiments cells were grown in 60 mm dishes to 70–80% confluence and transfected with 3 μg reporter plasmid, 2 μg expression plasmid, 1 μg internal *lacZ* reference plasmid pCMVb and 2 μg pBluescript (Stratagene) as a carrier. The medium was changed after 6 and 24 h incubation (37°C, 5% CO_2). Cell extracts were prepared 48 h after transfection (Hall, 1982). Transfection efficiency was first standardized by β-galactosidase assays and the appropriate amount of cell extract was then used in a CAT assay as previously described (Gorman *et al.*, 1982) or using the CAT ELISA kit (Boehringer Mannheim). Radiolabelled CAT assays were quantified following thin layer chromatography with scintillation counting. Each assay was performed in duplicate and repeated at least three times. Standardized (by β-galactosidase assay; transfection efficiencies varied by a factor of two) amounts of cell extracts were separated by SDS–PAGE, Western blotted and incubated with a polyclonal antibody against GAL4 (Santa Cruz Biotechnology) according to standard protocols. Immunodetection was performed with peroxidase-coupled goat anti-rabbit antibodies and the ECL system (Amersham).

Immunocytochemical in situ detection of T proteins

Cos7 cells were grown to 70% confluence in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. The medium was removed and cells incubated in 2.5 mM chloroquine, 500 mg/ml DEAE–dextran, 1 mg/ml pSG5.T expression plasmid in medium for 2.5 h (37°C, 5% CO_2). Cells were subsequently washed with phosphate-buffered saline (PBS) and medium and grown for 24 h in medium before they were split 1:2 and seeded on coverslips. After 24 h the cells on the coverslips were prepared for T protein detection. After two washes with PBS (5 min each) cells were fixed in 4% paraformaldehyde in PBS for 1 h. The paraformaldeyde was removed by two PBS washes and the cells were permeabilized with 50% methanol/PBS for 2 min and 100% methanol for 10 min. Subsequently cells were equilibrated in PBS for 30 min before free aldehyde groups were blocked in 1 M glycine for 30 min. After a 5 min wash in PBS endogenous peroxidases were destroyed by a 5 min incubation in 3% H_2O_2 in PBS. After a 5 min wash in PBS, a 5 min wash in PBST (0.05% Tween-20 in PBS) and a 5 min pre-incubation in PBSTN (10% fetal calf serum in PBST) the primary antibody was applied for 1 h in PBSTN (1:500 diluted α-TN1-123 or α-TC327-420 respectively). Primary antibody was removed by washes in PBST (4×5 min) and PBSTN (1×5 min) before the secondary antibody (FITC-conjugated goat-anti-rabbit; Jackson ImmunoResearch) was applied (1 h in PBSTN). The PBST washes were repeated. Cells were mounted in DPX (Fluka) and photographed on Kodak Ektachrome 320T film.

Acknowledgements

We thank Uli Schwarz for support, Irwin Davidson and Pierre Chambon for the pXJ40-GAL4(1–147) vector and the UAS-tk-CAT reporter plasmid, Heidrun Ellinger-Ziegelbauer for the donation of the pG.CAT, pSG5 and pCMVb plasmids, for her precious help in teaching us cell culture techniques and for many stimulating discussions, Hubert Ortner and Reinhard Korn for helpful discussions and Scott Lee for critically reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

References

Agulnik,S.I., Bollag,R.J. and Silver,L.M (1995) Conservation of the T-box gene family from *Mus musculus* to *Caenorhabditis elegans*. *Genomics*, **25**, 214–219.

Bollag,R.J., Siegfried,Z., Cebra-Thomas,J., Garvey,N., Davison,E.M. and Silver,L.M. (1994) An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T-locus. *Nature Genet.*, **7**, 383–389.

Carey,M. (1991) Mechanistic advances in eukaryotic gene activation. *Curr. Opin. Cell. Biol.*, **3**, 452–460.

Chesley,P. (1935) Development of the short-tailed mutant in the house mouse. *J. Exp. Zool.*, **70**, 429–435.

Courey,A.J. and Tjian,R. (1988) Analysis of SP1 *in vivo* reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell*, **55**, 887–898.

Cunliffe,V. and Smith,J.C. (1994) Specification of mesodermal pattern in *Xenopus laevis* by interaction between *Brachyury*, *noggin* and *XWnt-8*. *EMBO J.*, **13**, 349–359.

Dingwall,C. and Laskey,R.A. (1991) Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.*, **12**, 478–481.

Dobrovolskaia-Zavadskaia,N. (1927) Sur la mortification spontanee de la queue chez la souris nouveau-nee et sur l'existence d'un caractere hereditaire 'non viable'. *C.R. Hebd. Seanc. Soc. Biol.*, **97**, 114–116.

Frankel,A.D. and Kim,P.S. (1991) Modular structure of transcription factors: implications for gene regulation. *Cell*, **65**, 717–719.

Garcia-Bustos,J., Heitman,J. and Hall,M.N. (1991) Nuclear protein localization. *Biochim. Biophys. Acta*, **1071**, 83–101.

Gorman,C.M., Moffrat,L.F. and Howard,B.H. (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.*, **2**, 1044–1051.

Graham,F. and van der Eb,A. (1973) A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, **52**, 456–457.

Hall,C.V. (1982) Expression and regulation of *Escherichia coli* lacZ fusions in mammalian cells. *J. Mol. Appl. Genet.*, **2**, 101–109.

Harrison,S.C. (1991) A structural taxonomy of DNA-binding domains. *Nature*, **353**, 715–719.

Herrmann,B.G. (1991) Expression pattern of the *Brachyury* gene in whole-mount T^{Wis}/T^{Wis} mutant embryos. *Development*, **113**, 913–917.

Herrmann,B.G. and Kispert,A. (1994) The *T* genes in embryogenesis. *Trends Genet.*, **8**, 280–286.

Herrmann,B.G., Labeit,S., Poustka,A., King,T.R. and Lehrach,H. (1990)

Cloning of the *T* gene required in mesoderm formation in the mouse. *Nature*, **343**, 617–622.
Hollenberg,S.M. and Evans,R.M. (1988) Multiple and cooperative transactivation domains of the human glucocorticoid receptor. *Cell*, **61**, 1021–1033.
Hope,I.A. and Struhl,K. (1986) Functional dissection of a eukaryotic transcriptional activator, GCN4 of yeast. *Cell*, **46**, 885–894.
Hwang,J.-J., Chambon,P. and Davidson,I. (1993) Characterization of the transcription activation function and the DNA binding domain of the transcriptional enhancer factor-1. *EMBO J.*, **12**, 2337–2348.
Johnson,P.F. and McKnight,S.L. (1989) Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.*, **58**, 799–839.
Kalderon,D., Richardson,W.D., Markham,A.T. and Smith,A.E. (1984) Sequence requirements for nuclear location of simian virus 40 large T-antigen. *Nature*, **311**, 33–38.
Kispert,A. and Herrmann,B.G. (1993) The *Brachyury* gene encodes a novel DNA-binding protein. *EMBO J.*, **12**, 3211–3220, 4898–4899.
Kispert,A. and Herrmann,B.G. (1994) Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev. Biol.*, **161**, 179–193.
Kispert,A., Herrmann,B.G., Leptin,M. and Reuter,R (1994) Homologs of the mouse *Brachyury* gene are involved in the specification of posterior terminal structures in *Drosophila*, *Tribolium*, and *Locusta*. *Genes Dev.*, **8**, 2137–2150.
Kispert,A., Ortner,H., Cooke,J. and Herrmann,B.G. (1995) The chick *Brachyury* gene: developmental expression pattern and response to axial induction by localized activin. *Dev. Biol.*, **168**, 406–415.
Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. *Cell*, **63**, 579–590.
Klein-Hitpass,L., Schorpp,M., Wagner,U. and Ryffel,G.U. (1986) An estrogen-responsive element derived from the 5′ flanking region of the *Xenopus* vitellogenin A2 gene functions in transfected human cells. *Cell*, **46**, 1053–1061.
Lai,E., Prezioso,V.R., Tao,W., Chen,W. and Darnell,J.E.,Jr (1991) Hepatocyte nuclear factor 3a belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *forkhead*. *Genes Dev.*, **5**, 416–427.
MacGregor,G.R. and Caskey,C.T. (1989) Construction of plasmids that express *E.coli* β-galactosidase in mammalian cells. *Nucleic Acids Res.*, **17**, 2365.
Mermod,N., O'Neil,E.A., Kelley,T.J. and Tjian,R. (1989) The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. *Cell*, **58**, 741–753.
Mitchell,P.J. and Tjian,R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371–378.
Nagpal,S., Friant,S., Nakshatri,H. and Chambon,P. (1993) RARs and RXRs: evidence for two autonomous transactivation functions (AF-1 and AF-2) and heterodimerization *in vivo*. *EMBO J.*, **12**, 2349–2360.
Nye,J., Petersen,J.M., Gunther,C.V., Jonson,M.D. and Graves,B.J. (1992) Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA binding motif. *Genes Dev.*, **6**, 975–990.
Pani,L., Overdier,D.G., Porcella,A, Qian,X., Lai,E. and Costa,R.H. (1992) Hepatocyte nuclear factor 3b contains two transcriptional activation domains, one of which is novel and conserved with the *Drosophila* fork head protein. *Mol. Cell. Biol.*, **12**, 3723–3732.
Pflugfelder,G.O., Roth,H. and Poeck,B. (1992) A homology domain shared between *Drosophila optomotor-blind* and mouse *Brachyury* is involved in DNA binding. *Biochem. Biophys. Res. Commun.*, **196**, 918–925.
Qian,Y.Q., Billeter,M., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1989) The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580.
Rao,Y. (1994) Conversion of a mesodermalizing molecule, the *Xenopus Brachyury* gene, into a neuralizing factor. *Genes Dev.*, **8**, 939–947.
Schulte-Merker,S., Ho,R.K., Herrmann,B.G. and Nüsslein-Volhard,C. (1992) The protein product of the zebrafish homologue of the mouse *T* gene is expressed in nuclei of the germ ring and the notochord of the early embryo. *Development*, **116**, 1021–1032.
Smith,J.C., Price,B.M.J., Green,J.B.A., Weigel,D. and Herrmann,B.G. (1991) Expression of a *Xenopus* homolog of *Brachyury* (*T*) is an immediate-early response to mesoderm induction. *Cell*, **67**, 79–87.
Stunnenberg,H.G. (1993) Mechanisms of transactivation by retinoic acid receptors. *BioEssays*, **15**, 309–315.
Treisman,R. and Ammerer,G. (1992) The SRF and MCM1 transcription factors. *Curr. Opin. Genet. Dev.*, **2**, 221–226.
Treisman,R., Harris,E. and Desplan,C. (1991) The paired box encodes a second DNA-binding domain in the paired homeodomain protein. *Genes Dev.*, **5**, 594–604.
Wasylyk,C., Kerckaert,J.P. and Wasylyk,B. (1992) A novel modulator domain of Ets transcription factors. *Genes Dev.*, **6**, 965–974.
Webster,N.J.G., Green,S., Jin,J.R. and Chambon,P. (1988) The hormone-binding domains of the estrogen and glucocorticoid receptors contain an inducible transcription activation function. *Cell*, **54**, 199–207.
Weigel,D. and Jäckle,H. (1990) fork head: a new eukaryotic DNA binding motif? *Cell*, **63**, 455–456.
Wilkinson,D.G., Bhatt,S. and Herrmann,B.G. (1990) Expression pattern of the mouse *T* gene and its role in mesoderm formation. *Nature*, **343**, 657–659
Xiao,J.H., Davidson,I., Matthes,H., Wintzerith,M., Schatz,C., Wildemann,A. and Chambon,P. (1991) Cloning, expression, and transcriptional activation properties of human enhancer factor TEF-1. *Cell*, **65**, 551–568.
Xu,W., Rould,M.A., Jun,S. Desplan,C. and Pabo,C.O. (1995) Crystal structure of a paired domain–DNA complex at 2.5 Å resolution reveals structural basis for *Pax* developmental mutations. *Cell*, **80**, 639–650.
Yasuo,H. and Satoh,N. (1993) Function of vertebrate *T* gene. *Nature*, **364**, 582–583.
Yasuo,H. and Satoh,N. (1994) An ascidian homolog of the mouse *Brachyury* (*T*) gene is expressed exclusively in notochord cells at the fate restricted stage. *Dev. Growth Differ.*, **36**, 9–18.

Received on March 20, 1995; revised on July 13, 1995