The T-box family

Citation for published version:

Digital Object Identifier (DOI):
10.1186/gb-2002-3-6-reviews3008

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Genome Biology

Publisher Rights Statement:
RoMEO green

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Protein family review

The T-box family

Val Wilson* and Frank L Conlon†

Addresses: *Centre for Genome Research, King's Buildings, West Mains Rd, Edinburgh EH9 3JQ, UK. †Department of Genetics, 220 Fordham Hall, University of North Carolina, Chapel Hill, NC 27599-3280, USA.

Correspondence: Frank L Conlon. E-mail: frank_conlon@med.unc.edu

Published: 31 May 2002

Genome Biology 2002, 3(6):reviews3008.1–3008.7

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/6/reviews/3008

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Summary

Transcription factors of the T-box family are required both for early cell-fate decisions, such as those necessary for formation of the basic vertebrate body plan, and for differentiation and organogenesis. When mutated, T-box genes give dramatic phenotypes in mouse and zebrafish, and they have been implicated both in fundamentals of limb patterning and in a number of human congenital malformations such as Holt-Oram, ulnar-mammary and DiGeorge syndromes, as well as being amplified in a subset of cancers. Genes encoding members of the T-box family have recently been shown to comprise approximately 0.1% of genomes as diverse as those of nematodes and humans and have been identified in a wide variety of animals from ctenophores (comb jellies) to mammals; they are, however, completely absent from genomes from other organisms (such as the model plant Arabidopsis thaliana).

Gene organization and evolutionary history

Positional cloning and sequencing of the genes defective in the mouse gastrulation mutant Brachyury, also known as T, and of the Drosophila behavior mutant optomotor-blind (omb), show extensive sequence similarity between the amino-terminal regions of the two proteins [1,2]; the region of similarity contains a unique sequence-specific DNA-binding domain. Since these initial observations, over 50 proteins have been identified with sequence similarity to the DNA-binding domain of Brachyury and Omb. This domain is now referred to as the T-box and the genes are collectively referred to as the T-box gene family. Members of this family are expressed in, and are required for, the development of multiple cell types in diverse organisms, as demonstrated by genetic studies in flies, worms, fish, mice, dogs, and humans [3-7]. For many of these genes, such as Brachyury, there are clear orthologs (direct homologs), with a high degree of sequence similarity, expression pattern, and function between a variety of vertebrates, including fish, frogs, dogs, and mice [1-7]. Other T-box genes appear to be unique to a particular species; for instance, VegT, a T-box gene thought to be required for endoderm formation in Xenopus, has no apparent ortholog in mice or humans. The family has 18 members in mammals; representatives have been identified from a wide range of animals, including various chordates, Drosophila melanogaster (11 members), Caenorhabditis elegans (14 members), annelids, and cnidarians.

Analysis of T-box genes shows most of their loci to be dispersed randomly throughout chordate genomes (see Table 1 for their locations in the human genome), although several examples of clustering have been reported. One instance occurs in C. elegans, for which genomic sequencing has shown a tight linkage between Tbx8 and Tbx9 [8]; a second is in mouse, where Tbx2 and Tbx4 are tightly linked on chromosome 11 and Tbx3 and Tbx5 are linked on chromosome 5. The association between these latter T-box genes appears to be conserved in other mammals, as human Tbx2 and Tbx4, and Tbx3 and Tbx5, have a similar arrangement on chromosomes 17 and 12, respectively [9,10]. Phylogenetic analysis of these cognate pairs suggests they arose through initial duplication of an ancestral gene by an unequal crossover event.
between two alleles; in the case of Tbx2 and Tbx4, and Tbx3 and Tbx5, these events occurred at least 600 million years ago [8,11]. Although possible mechanisms for the duplication have been put forward, such as duplication of entire chromosomes or genomes, the functional consequences of the pairs of linked Tbx genes remains to be established.

T-box genes contain multiple exons, and the T-box is generally encoded by at least five exons dispersed over a relatively large distance. For example, the human Tbx5 gene contains eight exons distributed over 53 kilobases (kb) of chromosome 12 [10]. As has been found for other gene families, the intron-exon boundaries of T-box homologs are conserved throughout evolution, but the lengths of the introns vary between species [10,12]. Most T-box family members encode a single transcript and there are few direct demonstrations of alternative exon splicing. One exception is Xenopus VegT/Antipodean protein, which is found in two different isoforms resulting from alternative splicing of the 3’ end of the VegT gene. Interestingly, the isoforms appear to be tissue-specific: one is present maternally in the endodermal layer of the embryo and the other is expressed zygotically in the mesodermal layer [13].

**Characteristic structural features**

T-box proteins generally range in size from 50 kDa to 78 kDa. *Brachyury*, the founding member of the T-box gene family, has been shown to encode a sequence-specific DNA-binding protein that functions as a transcriptional activator [1,14-17]. Although crystallographic analysis of T-box proteins has been achieved only for a truncated version of the *Xenopus* homolog of Brachyury, Xbra, and a truncated version of Tbx3, the results clearly demonstrate that the T-box is unlike any other DNA-binding domain [18] (Figure 1). Studies with a number of T-box proteins have shown that they comprise at least two structural and functional domains: a sequence-specific DNA-binding domain (the T-box) and a transcriptional activator or repressor domain [3,4]. The relative position of the domains varies between different members of the family, but the order is conserved for any one member of the T-box family and its orthologs [10,12].

**The T-box**

The T-box is defined as the minimal region within the T-box protein that is both necessary and sufficient for sequence-specific DNA binding [3-5,14,17]. Despite the sequence variations within the T-box between family members, examination of downstream targets and binding-site selection experiments for a number of T-box proteins show that all members of the family so far examined bind to the DNA consensus sequence TCACACCT. In several binding-site selection studies, members of the T-box family preferentially bound sequences that contain two or more core motifs arranged in various orientations. The ability to bind the sequence is protein-specific; for example, Xbra can bind to two core motifs arranged head-to-head, whereas VegT cannot; conversely, VegT can bind to two core motifs arranged tail-to-tail whereas Xbra cannot. The biological relevance of these findings remains unknown, as no downstream target of any T-box gene has been found to contain a double site [17].

The T-box is a relatively large DNA-binding domain, generally comprising about a third of the entire protein (17-26 kDa), and individual T-box gene family members show varying degrees of homology across the domain. Specific residues within the T-box are 100% conserved in all members of the family, however. This observation has provided the basis for subdivision of the family (see Figure 2) [19]. It has recently been demonstrated that the specificity of several T-box proteins for their target sites lies mainly within the T-box. But specificity does not appear to reflect binding affinity [17], suggesting that other functions may lie in the T-box, such as regions required for protein-protein interactions. Consistent with this proposal, our mutational analysis has identified a single amino-acid residue within the T-box of *Xenopus* Xbra, Eomesodermin, and VegT (Lys149, Asn155 or Asn353, respectively) that is required for the correct target specificity of the respective proteins [17]. In addition, one T-box protein, Mga, contains both a T-box and a basic helix-loop-helix leucine zipper (bHLH-zip) domain [20]. When heterodimerized with the bHLH protein Max, Mga is converted to a transcriptional activator with apparent dual specificity, regulating genes containing either a Max-binding or a T-box-domain-binding site [20]. Conversely, human Tbx22 has been found to contain a truncated T-box lacking residues found in the amino-terminal portion of all other family members and would be predicted not to bind DNA [21]. Tbx22 may therefore represent a case in which the T-box has functions other than DNA binding.

**Transcriptional regulatory domains**

T-box proteins have been demonstrated to function both as transcriptional activators and as repressors. In all cases studied, the transcriptional regulation activity has been shown to require sequences located in the carboxy-terminal portion of the protein. Only in the case of Brachyury and its frog and zebrafish orthologs, however, has the region both necessary and sufficient for transcriptional regulation been accurately mapped [16,22]. Interestingly, there are only a few small blocks of conservation between the Brachyury orthologs in this region, and the overall level of similarity is low.

**Localization and function**

The T-box genes share two characteristics of interest to researchers studying cell specification and differentiation: they tend to be expressed in specific organs or cell types, especially during development, and they are generally required for the development of those tissues (Table 1). In the
few cases for which intracellular localization has been analyzed, T-box proteins have shown to be localized exclusively in the nucleus. These considerations, together with their DNA-binding and transcriptional activation/repression capacity, mean that T-box proteins are well placed to fulfill a wide array of important regulatory roles in development. This is supported by the observation that mutant alleles commonly give a phenotype even in heterozygotes (that is, they show haploinsufficiency), indicating that the level of a T-box protein is important for determining its function. In addition, mutational studies have demonstrated that T-box genes are required cell-autonomously (active in the cell in which they are expressed). For example, Brachyury is expressed in posterior mesoderm and in the developing notochord, and it is required for the formation of these cells in mice [1,23-28].

T-box genes are also required in specific tissue types at later stages of development: for example, Tbx2, Tbx3, Tbx4 and

---

**Figure 1**
Ribbon diagram of crystal structures of (a,b) Xenopus Xbra and (c) human TBX3 bound to DNA. Beta strands are depicted in red and alpha helices in (a,b) orange or (c) turquoise. Reproduced with permission from [18,61].

---

**Figure 2**
Conservation of selected T-box residues and the presence of diagnostic residues for different members of the family. Position 149 is always a lysine in Xbra proteins from different species (blue) but not in other T-box proteins (red). A diagram of Xenopus Xbra is above, showing the relative positions of the DNA-binding domain, the nuclear localization signal, and the transcriptional activation domain.
Tbx5 in the developing limb (reviewed in [29]). Tbx2 and Tbx3 are expressed in the anterior and posterior margins of both forelimb and hindlimb buds [30]. The posterior expression of Tbx3 is crucial for the development of the more distal limb elements, as shown in human patients lacking TBX3, who have ulnar-mammary syndrome and lack the ulna (a forearm bone) and digits [31,32].

In contrast to the overlap in expression of Tbx2 and Tbx3, their close homologs, Tbx4 and Tbx5, are expressed exclusively in the hindlimb bud and the forelimb bud, respectively [30]. Although the signals that direct expression of these genes to the forelimb or hindlimb are unknown, it is likely that at least part of this involves an interpretation of the ‘Hox code’, the set of Hox genes expressed in the mesoderm that eventually produces the mesenchyme cells that migrate into the limb buds [33,34]. The expression of Tbx4 lies downstream of Ptx1, a homeobox-containing gene expressed in posterior lateral plate mesoderm. Expression is, however, independent of the signals that direct limb-bud outgrowth. Significantly, these genes not only delineate forelimb and hindlimb territories but also specify forelimb or hindlimb type [33,34]. Retroviral overexpression of Tbx4 inappropriately in forelimb mesenchyme of chick embryos, or of Tbx5 in the hindlimb, can transform the tissue into hindlimb or forelimb type, respectively [33]. The transformation includes both the mesodermally derived skeletal elements and the overlying ectoderm, which develops feathers or scales depending on which gene is expressed [33]. Tbx4 and Tbx5 are thus thought to act as ‘selector’ genes for the limb bud, defining what type of limb develops; this is corroborated by the correlation of Tbx4 or Tbx5 expression with hindlimb or

<table>
<thead>
<tr>
<th>Subfamily</th>
<th>Gene</th>
<th>Human Chromosome</th>
<th>Expression</th>
<th>Heterozygous phenotype in human</th>
<th>Reference (mutations only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brachyury</td>
<td>BRACHYURY</td>
<td>6</td>
<td>Primitive streak, tail bud, and notochord</td>
<td>Spinal cord defects (anteroposterior axis defects)</td>
<td>[22-26]</td>
</tr>
<tr>
<td>Tbx2/3</td>
<td>TBX19 (TPIT)</td>
<td>1</td>
<td>Pituitary</td>
<td></td>
<td>[49,50]</td>
</tr>
<tr>
<td>T-box l</td>
<td>T-BRAIN1</td>
<td>2</td>
<td>Cerebral cortex</td>
<td></td>
<td>[51]</td>
</tr>
<tr>
<td>EOMESODERMIN/ (T-BRAIN2)</td>
<td></td>
<td>3</td>
<td>Trophoblast, early primitive streak, and cerebral cortex</td>
<td>(Early postimplantation failure)</td>
<td>[44,45]</td>
</tr>
<tr>
<td>Tbx1</td>
<td>TBX1</td>
<td>22</td>
<td>Heart and pharyngeal arches</td>
<td>DiGeorge syndrome</td>
<td>[38-40]</td>
</tr>
<tr>
<td>Tbx2</td>
<td>TBX2</td>
<td>17</td>
<td>X-linked cleft palate</td>
<td></td>
<td>[59]</td>
</tr>
<tr>
<td>Tbx3</td>
<td>TBX3 (including an alternative splice form)</td>
<td>12</td>
<td>Limbs and heart</td>
<td>Ulnar-mammary syndrome</td>
<td>[31,32]</td>
</tr>
<tr>
<td>Tbx4</td>
<td>TBX4</td>
<td>17</td>
<td>Allantois, hindlimb</td>
<td></td>
<td>[30,60]</td>
</tr>
<tr>
<td>Tbx5</td>
<td>TBX5 (including an alternative splice form)</td>
<td>12</td>
<td>Forelimb</td>
<td>Holt-Oram syndrome (failure of heart development)</td>
<td>[30,31,35-37]</td>
</tr>
<tr>
<td>Tbx6</td>
<td>TBX6</td>
<td>16</td>
<td>Primitive streak and tail bud</td>
<td>(Respecification of posterior paraxial mesoderm as neurorteoderm)</td>
<td>[60]</td>
</tr>
</tbody>
</table>

*These sequences have been reported in mouse but not human; the human genes are hypothetical.
forelimb identity in the limb buds induced by ectopic expression of fibroblast growth factors in the flank [34]. Finally, mutations in human TBX5 affect forelimb growth and heart development [35,36]. Interestingly, missense mutations within the T-box of human TBX5 that contact the minor groove of DNA (such as Arg237Gln) result primarily in limb abnormalities, whereas the other aspect of Holt-Oram syndrome, aberrant heart development, is predominantly seen as a result of a missense mutation that alters a residue that contacts the major groove (Gly80Arg) [37]. This suggests that tissue-specific target genes are affected by mutations in different residues within the T-box of TBX5.

Mutations in T-box proteins have also been implicated in DiGeorge syndrome, a complex disease that includes abnormalities of the heart’s outflow tract [38-40], and Tbx2 is amplified in some types of breast cancer [41].

Frontiers

Despite the essential role for individual members of the T-box gene family in a wide variety of developmental processes, relatively little is known about the genetic and biochemical pathways in which T-box genes act [42,43]. Thus, one of the critical areas for future research is to identify the factors that act directly upstream and downstream of individual T-box genes.

At the cellular level, genetic mutations have provided clues about the requirement for T-box genes in a variety of developmental processes, but the exact function of T-box genes, including questions of genetic redundancy, still needs to be established. For example, in ulnar-mammary syndrome patients it is not clear why there is no corresponding defect in the legs, a region that normally expresses Tbx3 [30,31]. This may perhaps be due to a redundant function with other T-box genes expressed in the hindlimb, such as Tbx2 or Tbx4.

In order to dissect the function of individual T-box genes, it will also be necessary to generate allelic series (as has been useful for the study of Holt-Oram syndrome) and conditional mutations. A case for the latter has already been demonstrated for Eomesodermin, a gene expressed in all vertebrates just prior to gastrulation in the prospective mesoderm and, in the mouse, in the trophoderm, an extraembryonic tissue that is required for placenta formation and is thus unique to mammals [44]. Mouse lacking Eomesodermin fail at, or shortly after, gastrulation, because of a defect in the trophoderm. This phenotype can be rescued by wild-type trophoderm, even if the embryo itself is mutant, but when embryonic tissues lack Eomesodermin, mesoderm differentiation and migration fails completely [45].

Finally, relatively little is known about post-translational processing, protein turnover, or protein stability for any T-box protein. In addition, only in the case of Tbx5 [46,47], Tbr1 [48], and Mga [20] has a protein-protein interaction domain been reported and, with the exception of Mga, the biological significance of these interactions has not yet been determined.

References

   The seminal piece of work identifying Brachyury as the gene disrupted in a classical mouse mutation.
   First clues that Brachyury and Omb may be members of a protein family.
   A comprehensive review of the T-box gene family.
   A review of the molecular and cellular aspects of the T-box gene family.
   A review of T-box family members.
   This paper demonstrates the high degree of conservation in the function of Brachyury and explains why Corgis have short tails.
   A review of T-box family members and their role in early development.
   Shows the conservation of the T-box family from worms to mice.
   Work on the evolutionary history of the T-box family.
10. Entrez Nucleotide
   One of the main sequence databases.
    A proposal for how an ancestral locus gave rise to two T-box gene clusters.
    A proposal for classifying the T-box family based on genomic structure of the DNA-binding domain.
    The first direct demonstration of alternative splicing of a T-box gene.
    Brachyury can bind to DNA in a sequence-specific fashion.
    Further evidence that the sequence and expression of Brachyury is evolutionarily conserved.


24. Gluecksohn-Schoenheimer S: Cloning of a T-box gene family member with a truncated version of the DNA-binding domain.


47. Gluecksohn-Schoenheimer S: Brachury regulates cell migration.


53. Gluecksohn-Schoenheimer S: Brachury regulates cell migration.


59. Gluecksohn-Schoenheimer S: Brachury regulates cell migration.


