The TCP domain: a motif found in proteins regulating plant growth and development

Pilar Cubas1,†, Nick Lauter2, John Doebley2 and Enrico Coen1,*
1John Innes Centre, Colney Lane, Norwich NR4 7UH, UK, and
2Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108, USA

Summary
The cycloidea (cyc) and teosinte branched 1 (tb1) genes code for structurally related proteins implicated in the evolution of key morphological traits. However, the biochemical function of CYC and TB1 proteins remains to be demonstrated. To address this problem, we have analysed the predicted secondary structure of regions conserved between CYC and TB1, and looked for related proteins of known function. One of the conserved regions is predicted to form a non-canonical basic-Helix-Loop-Helix (bHLH) structure. This domain is also found in two rice DNA-binding proteins, PCF1 and PCF2, where it has been shown to be involved in DNA-binding and dimerization. This indicates that the conserved domain most probably defines a new family of transcription factors, which we have termed the TCP family after its first characterised members (TB1, CYC and PCFs). Other plant proteins of unknown function also belong to this family. We have studied two of these in Arabidopsis and have shown that they are expressed in rapidly growing floral primordia. This, together with the proposed involvement of cyc and tb1 in influencing meristem growth, suggests that many members of the TCP family may affect cell division. Some of these genes may have been recruited during plant evolution to generate new morphological traits.

Introduction
The cycloidea (cyc) and teosinte branched 1 (tb1) genes have been implicated in the evolution of key morphological traits. The cyc gene is involved in the control of floral symmetry, a character that has changed many times during plant evolution (Carpenter and Coen, 1990; Luo et al., 1996; Stebbins, 1974). The tb1 gene controls developmental switches that contributed to the evolution of maize from its wild ancestor teosinte (Doebley et al., 1995; Doebley et al., 1997). Although both cyc and tb1 have been isolated, the biochemical function of their encoded proteins is unclear (Doebley et al., 1997; Luo et al., 1996). To address this problem, we have analysed the predicted secondary structure of these and related proteins, and compared some of the gene expression patterns.

The cyc gene is required, together with a related gene, dichotoma (dich), to establish dorsoventral asymmetry of the Antirrhinum flower. In flowers mutant for both cyc and dich, differences between dorsal, lateral and ventral organs are eliminated, rendering the flower radially symmetrical. The cyc gene is expressed in the dorsal region of wild-type floral meristems, from very early through to later stages of development. The initial activity reduces the growth rate in the dorsal region of the wild-type meristem and controls primordium initiation. Late expression prevents the full development of the dorsal stamen and affects the asymmetry, size and cell types of the dorsal and lateral petals (Luo et al., 1996).

The tb1 gene affects the fate of maize axillary meristems; at lower nodes it prevents the outgrowth of buds and at upper nodes it promotes the development of female inflorescences (ears). In tb1 mutants of maize, axillary buds of lower nodes grow out to give basal branches (tillers), and the buds of upper nodes give branches tipped with male inflorescences (tassels), a phenotype reminiscent of the ancestor of maize, teosinte (Doebley et al., 1997).

Although the processes controlled by cyc and tb1 appear to be unrelated, there are some common themes. First, both genes are involved in the development of axillary structures, either flowers or branches. Second, both genes affect petals and stamens, organs whose development is regulated by the B class of floral organ identity genes (Coen and Meyerowitz, 1991). Third, both genes have been proposed to function, at least in part, as modifiers of organ growth.

To understand the biological role of cyc and tb1, we have investigated the possible functions of their proteins. In particular, we have studied regions conserved between cyc and tb1 that may act as functional domains. The predicted secondary structure of one of the regions is a basic-Helix-Loop-Helix (bHLH). This region is unrelated to the bHLH structure found in canonical bHLH transcription factors (Murre et al., 1989), but is closely related to a bHLH domain.
found in two rice DNA binding proteins, PCF1 and PCF2 (Kosugi and Ohashi, 1997). Based on this homology we define a new class of proteins, the TCP family, that most likely act as transcription factors. We have also characterised two further members of the TCP family from Arabidopsis and have shown that they are expressed in floral organs undergoing rapid growth. Taken together with the phenotypic effects of cyc and tb1, this suggests that members of the TCP family may influence cell division and growth. Recruitment of some of these genes to play new roles during plant development may underlie some key morphological changes during angiosperm evolution.

Results

CYC and TB1 contain a basic-Helix-Loop-Helix domain

To investigate the biochemical action of the CYC and TB1 proteins, we analysed their predicted secondary structures. In particular, we studied the regions showing sequence conservation, as these might constitute functional domains.

The first conserved region is predicted to form a basic-Helix-Loop-Helix (bHLH, Figures 1a and 2a). This bHLH domain is defined by structural criteria and is unrelated in sequence to the bHLH domain found in MyoD, E12 and related proteins (Murre et al., 1989). The basic region of the bHLH domain of both CYC and TB1 is 21 residues long and includes a putative bipartite nuclear localisation signal (NLS) (Dingwall and Laskey, 1991; Doebley et al., 1997; Luo et al., 1996). The helical regions are amphipathic and comprise alternating conserved hydrophobic residues and partially conserved hydrophilic residues (Figure 1b). The second helix contains a LXXLL-motif (indicated in Figure 2a), which has been shown to mediate binding of transcriptional co-activators to liganded nuclear receptors in animals (Heery et al., 1997). Helix II also contains three potential sites of phosphorylation (serine or threonine), two of them in conserved positions. The region linking the two helices has conserved glycine, aspartate and serine residues found with high frequency in loops (Leszczynski and Rose, 1986), as well as proline in the case of CYC. In addition to the bHLH domain, CYC and TB1 have a second conserved region termed the R-domain (Figure 2b), which is rich in polar residues (arginine, lysine and glutamic acid). The R-domain is predicted to form a hydrophilic α-helix (not shown). Both the bHLH and R domains are also predicted to form coiled coils, similar to those formed by leucine zippers (Lupas et al., 1991; Lupas, 1996).

The TCP domain

When CYC and TB1 were first described, no sequence similarity to proteins of known biochemical function was found. However, more recent database searches revealed two additional proteins with similarity to CYC and TB1 in the bHLH region: PCF1 and PCF2. These proteins were isolated on the basis of their ability to bind specifically to promoter elements of the rice gene for the proliferating cell nuclear antigen, PCNA, a protein involved in DNA replication and cell cycle control (for a review see Zophonias and Hubscher, 1997). The PCF proteins may bind to DNA as homo- or heterodimers. A region containing the bHLH domain has been shown to be sufficient for DNA binding and necessary for dimerization (Kosugi and Ohashi, 1997). This suggests that CYC and TB1 may also function as DNA-binding proteins and may interact with other proteins through their bHLH domain.

In addition to CYC, TB1 and the PCFs, several sequences of unknown function from Arabidopsis and maize show
Figure 2. The TCP gene family.
(a) Alignment of the predicted amino acid sequence from members of the TCP family. Only the TCP domain is shown and founder members of the TCP family are indicated in bold. Identical amino acids are in black boxes, while amino acids with similar charge or hydrophobicity are in grey. Circles along the top indicate conserved basic residues; underlined circles indicate residues forming part of the putative bipartite NLS; asterisks indicate conserved hydrophobic residues in the helices; underlined asterisks indicate the LXXLL motif; black arrowheads point to residues (glycine or proline) that disrupt α-helix formation. The variable predicted length of helix II is indicated by a dotted line.
(b) Alignment of the R-domain in the CYC/TB1 subfamily. Note that TCP1 has several glycines in this region that are likely to prevent the formation of a helix.
(c) Consensus tree showing relationship between various TCP proteins. The percentage of bootstrap samples in which particular clades were monophyletic is indicated for those clades with support of 70% or more.

Amino acids are expressed in the standard single letter code; Ψ stands for any hydrophobic residue. In the PCF subfamily the N-terminus of the bipartite domain lies outside the conserved region (Kosugi and Ohashi, 1997). The predicted length of the helices varies slightly for each member of the TCP family. Note that the LXXLL motif is not fully conserved, however, the VXXLL sequence found in TB1 has also shown to tolerate interactions with liganded nuclear receptors (Heery et al., 1997). TCP1 (AC002130) corresponds to gene FIN21.14, BAC FIN21, chromosome 1. TCP2 (AF072691) corresponds to cDNA clone 1614T7 (EST R30409). TCP3 (AF072134) corresponds to cDNA clone 75H5T7 (EST T45419). TCP4 corresponds to the maize clone php20581. TCP5 corresponds to an ORF found in the genomic clone MSL3 (AB008269), chromosome 5. TCP6 corresponds to an ORF found in the genomic clone MEE6 (AB010072). TCP7 (N97289) is an EST from cDNA clone 241F10T7. TCP8 (H36511) is an EST from cDNA clone 174B16T7. TCP9 (T88453) is an EST from cDNA clone 156J1T7. Only non-redundant ESTs or BACs have been used for the alignment.

homology to the bHLH domain (Figure 2a; Doebley et al., 1997; Kosugi and Ohashi, 1997; Luo et al., 1996). Alignment of these protein sequences shows a high conservation of key residues in the bHLH domain: two short stretches of residues in the basic region, hydrophobic residues along the apolar face of both α-helices, a tryptophan in helix II and a helix-breaking glycine in the loop between the helices (Figure 2a). However, the residues in the loop and the hydrophilic residues of the helices are not as well conserved.

The alignment shows that the proteins form two subfamilies: one closely related to CYC and TB1, and another more related to the PCFs. The basic region of the CYC/TB1 subfamily contains a bipartite NLS (Dingwall and Laskey, 1991; Doebley et al., 1997; Luo et al., 1996) while the basic region of the PCF subfamily contains only a portion of a bipartite NLS. Although largely conserved within each subfamily, the residue composition of the loop is different between the two subfamilies. Helix II of the CYC/TB1 subfamily is longer than that of the PCF subfamily, which is interrupted after nine residues by a proline (Figure 2a). Each subfamily shares specific regions outside the bHLH domain: members of the PCF subfamily share regions adjacent to the bHLH domain (Kosugi and Ohashi, 1997), while some members of the CYC/TB1 subfamily share an R-domain located at various distances from the TCP domain (Figure 2b and D. Luo, M. Chadwick and P. Cubas, unpublished results). The tree obtained from this alignment confirms that the proteins fall into two clusters (Figure 2c).

Thus, CYC, TB1 and the PCFs belong to a large family of...
proteins sharing a common motif that we propose to call the 'TCP domain', based on the initial letters of the founding members (TB1, CYC, PCF). According to the evidence presented for the PCFs (Kosugi and Ohashi, 1997), the TCP domain is probably involved in DNA binding and protein–protein interactions.

Role of TCP proteins in meristem proliferation

A common feature of the TCP proteins characterised thus far is a role in meristem growth: CYC controls the growth of the floral meristems and primordia; TB1 affects axillary meristem growth; PCF1 and PCF2 bind to the promoter of a gene (PCNA) involved in meristematic cell division. To investigate whether this correlation with growth held for other members of the TCP family, we analysed two Arabidopsis ESTs, here termed TCP2 and TCP3, that are closely related to CYC and TB1. If these were involved in cell growth and proliferation, they might be expected to be expressed in dividing tissue. The TCP2 and TCP3 cDNAs were sequenced (Figure 3a,b; GenBank accession numbers AF072691 and AF072134, respectively), genetically mapped (Figure 3c) and their expression patterns characterised by in situ hybridisation.

Sections of plant tissue were collected and embedded 4, 6, 8, 10, 12, 16 and 22 days after sowing, and hybridised with digoxigenin-labelled probes from TCP2 and TCP3 templates. The transcription patterns of the two genes were qualitatively very similar. At 4–8 days, when the apical meristem was producing vegetative leaf primordia, there was no detectable mRNA from TCP2 and TCP3. After 8 days, the apical meristem started producing flower primordia. At this time, weak expression of both genes was detected in vegetative primordia (not shown).

Transcripts were also detectable from the earliest stages of flower development (Smyth et al., 1990). At stages 1–2, the floral meristem was dome-shaped and the signal was diffuse throughout (Figure 4a,e). From stage 3, when the floral organ primordia started to form, expression patterns were monitored separately for each of the four whorls. (i) In sepals, mRNA accumulation was highest at stages 3–4, during which sepal primordia grew to enclose the bud (Figure 4a,e). After this, the signal progressively decayed. (ii) In petals, weak signal was detected when the petal primordia first became visible (stages 5–6, not shown), and peaked during stages 8–12 when they were growing rapidly (Figure 4c,d,h). During this time, most signal was observed at the tips and the margins of the petal primordia (Figure 4h). Dorsal (adaxial) and ventral (abaxial) petals showed similar levels of expression. (3) In the stamens, signal increased during stages 7–8 when the stamen primordia were growing rapidly and anthers were starting to become distinguishable (Figure 4b). By stages 8–9 signal was restricted to the developing anthers, being mainly...
detectable in the pollen sacs where microspore mother cells were undergoing meiosis (Figure 4c,g). Transcript levels were similar in dorsal (adaxial), lateral and ventral (abaxial) stamens. By stage 10, when pollen grains were mature, transcripts were no longer detectable in stamens (Figure 4c). In carpels, signal was detected in the placental tissue during stage 9 when ovule primordia were forming (Figure 4g). The expression pattern in roots was not analysed.

In summary, the Arabidopsis TCP2 and TCP3 genes are most strongly expressed during flower development; expression being highest in petal and stamen primordia, but also being detectable in sepals and carpels. Expression coincides with the stages when primordia are growing rapidly, consistent with a role for these genes in primordial growth, but by no means conclusive as many other functions may also be compatible with such expression patterns. Unlike cyc, there is no distinction in expression levels between dorsal and ventral primordia.

Discussion

We have shown that CYC and TB1 belong to a family of proteins sharing a common region, the TCP domain, that is predicted to form a basic-Helix-Loop-Helix (bHLH) structure. This region is unrelated in sequence to the canonical bHLH domain found in transcription factors such as MyoD (Murre et al., 1989). However, it is similar to the bHLH domain found in PCF1 and PCF2, plant DNA-binding proteins that most probably act as transcription factors (Kosugi and Ohashi, 1997). The main conserved features of the TCP domain are: two short stretches of residues in the basic region, hydrophobic residues along the apolar face of both α-helices, a tryptophan in helix II, and a helix-breaking glycine in the loop between the helices. So far, members of the TCP family have only been found in plants.

What is the biochemical function of the TCP domain? Important clues have been obtained from the analysis of PCF1 and PCF2. In these proteins, the basic region of the TCP domain is necessary for specific binding to promoter elements of the PCNA gene (Kosugi and Ohashi, 1997). Basic regions have also been shown to be involved in DNA binding in the case of bHLH, bHLHZ and bZIP proteins (Hurst, 1994; Littlewood and Evans, 1994). In these transcription factors, the basic domain adopts an α-helical structure that interacts with the major groove of the DNA. In contrast, the basic region of the TCP domain contains residues that prevent helix formation, suggesting that in this case the basic region binds to DNA through a different mechanism (Kosugi and Ohashi, 1997). It is possible that other TCP proteins bind DNA through their basic domain.
in a similar way. In addition, the basic region of the TCP domain may target these proteins into the nucleus as it contains a complete (CYC/TB1 subfamily) or partial (PCF subfamily) bipartite NLS (Dingwall and Laskey, 1991; Doebley et al., 1997; Luo et al., 1996). This would fit the observation that NLSs often overlap or flank nucleic acid-binding domains (LaCasse and Lefebvre, 1995; Littlewood and Evans, 1994).

The role of the HLH region of the PCFs and other TCP proteins is less clear. One possibility is that the amphipathic helices mediate protein–protein interactions through their hydrophobic surfaces. This would be similar to the proposed role of the amphipathic helix (K domain) of MADS box genes (Davies and Schwarz-Sommer, 1994; Shore and Sharrocks, 1995). Amphipathic helices also mediate homomeric and heterodimerization in bZIP and bHLH proteins of the MyoD type (Landschulz et al., 1988; Murre et al., 1989). In the PCFs, a region containing the TCP domain is essential for homo- and heterodimerization, although the role of the helices has not been tested (Kosugi and Ohashi, 1997). Amphipathic helices might also be involved in interactions with non-TCP proteins. For instance, helix II of the TCP domain contains a conserved sequence that resembles the LXXLL motif shown to be involved in protein interactions with liganded nuclear receptors (Heery et al., 1997). It is possible that TCP proteins interact with as yet unidentified plant nuclear receptors or other proteins through this sequence.

Most members of the TCP family contain up to three potential phosphorylation sites in serine and threonine in the basic domain and helix II. Phosphorylation has been shown to affect nuclear localization, DNA binding and transcriptional activation of regulatory proteins (Hunter and Karin, 1992), raising the possibility that the activity of the TCP proteins might be regulated by a similar mechanism.

The TCP proteins fall into two subfamilies (one including CYC and TB1 and the other including the PCFs) based on features both within and outside the TCP domain. Within the TCP domain, each subfamily has a different linker for the bipartite NLS, a distinct residue composition in the loop and hydrophilic faces of the helices, and a different length for helix II. Outside the TCP domain, most members of the CYC/TB1 subfamily have an R-domain, predicted to form a coiled coil that may mediate protein–protein interactions (Lupas et al., 1991), and all members of the PCF subfamily share short regions flanking the TCP domain. These differences between subfamilies may reflect differences in the DNA-binding specificities and/or protein–protein interactions.

Unlike the homeo-domain containing genes of metazoans, many of which occur in tandem clusters in the genome (Holland et al., 1994), the Arabidopsis TCP genes are dispersed throughout the genome. In this respect they resemble the MADS box family of higher plants, which are also dispersed (Hauge et al., 1993; Rounsley et al., 1995). All the members of the TCP family investigated thus far function in processes related to cell proliferation. CYC retards growth of the dorsal region of young floral meristems, affecting the number of primordia initiated. Later, it arrests dorsal stamen development and promotes dorsal petal growth (Luo et al., 1996). TB1 is involved in arresting growth of some axillary buds, repressing internode growth in branches, and arresting petal (lodicule) and stamen development in the female flowers (Doebley et al., 1995; Doebley et al., 1997). PCF1 and PCF2 most probably control the transcription of PCNA, a gene expressed only in meristematic tissue where it is involved in DNA replication and cell cycle control. Here we show that the expression patterns of two Arabidopsis members of the CYC/TB1 subfamily, TCP2 and TCP3, also correlate with actively dividing regions of the floral meristem, suggesting a possible involvement of these genes in regulating cell growth and division. However, such expression patterns may also be compatible with other biological roles. Further experiments, such as inactivation of the TCP2 and TCP3 genes, will be needed to determine whether these genes are indeed involved in regulating growth.

The TCP2 and TCP3 genes are most probably not orthologues of CYC or TB1, as CYC and TB1 are more closely related to each other than to either of these genes (Figure 2c). However, they do share with CYC and TB1 some features in their expression patterns: the TCP2 and TCP3 genes are upregulated in petal and stamen primordia, similar to cycc; the tb1 gene is also likely to be expressed in these organ primordia as tb1 mutants affect the development of petals and stamens. One distinctive feature, however, is that cycc is expressed only in dorsal primordia whereas TCP2 and TCP3 are expressed uniformly along the dorsoventral axis.

It is possible that many members of the TCP family function in proliferating tissues where they may act in combination with other proteins to influence cell division and growth, and perhaps recruitment of some of these regulatory genes for new developmental functions has been involved in generating key changes in plant morphology during angiosperm evolution.

**Experimental procedures**

**Secondary structure analysis**

The secondary structure prediction for CYC was obtained by probabilistic Discrete State-Space Models analysis, DSMs (Stultz et al., 1993; White et al., 1994), submitting the protein sequences to the Protein Sequence Analysis (PSA) server (http://bmerc-www.bu.edu/psa/coment3.htm). Helical wheels were plotted with PROTEAN (DNASTAR for Windows 3.10a) and refined by hand. The prediction of coiled-coil regions (Lupas et al., 1991) was
obtained by submitting the protein sequences to the Network Protein Sequence analysis server (http://pbil.ibcp.fr).

Multiple sequence comparison

The TCP and R-domain alignments were constructed with MEGALIGN (DNASTAR for Windows 3.10a), using CLUSTALW (Thompson et al., 1994) and PAM250 weighted distances and were manually refined. Phylogenetic analysis was performed with the PHYLIP package (Felsenstein, 1985). Two hundred bootstrap re-samplings of the original data were generated with the SEQBOOT program. Distance matrices were made for each bootstrap dataset using the PRODIST program with the Dayhoff PAM distance method, and 200 trees constructed from these by the Neighbor-joining method.

In situ hybridisations

Plant material was grown under long days and collected for in situ hybridisation as described by Ratcliffe et al. (1998). Digoxigenin labelling of RNA probes, tissue preparation and in situ hybridisation were done as described by Coen et al. (1990).

Sequencing

Plasmid clones of the R30409 and T45419 ESTs (corresponding to TCP2 and TCP3, respectively) were obtained from the ABRC at Ohio University. These clones were completely sequenced at the Advanced Genetic Analysis Center (University of Minnesota) using gene specific internal oligonucleotide primers.

Genetic mapping

Four TCP genes were placed on the Arabidopsis genetic map using the R30409 (TCP2) and T45419 (TCP3) EST clones as probes on DNA blots of a set of 99 recombinant inbred lines (ABRC Stock Number CS1899). For R30409, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP3 and maps to chromosome 1 between markers m213 and nga128. T45419 also hybridised to two other loci more faintly: one between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that correspo

References


GenBank accession numbers AF072691 and AF072134.


