Functional analysis of the promoter of the glycine decarboxylase PA gene (*GLDPA*) of *Flaveria trinervia* (C_4)

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Christian Wiludda

aus Düsseldorf

Düsseldorf, November 2011

aus dem Institut für Entwicklungs- und Molekularbiologie der Pflanzen
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. P. Westhoff
Koreferent: Prof. Dr. R. Simon

Tag der mündlichen Prüfung: 19.01.2012

Ich habe die vorliegende Dissertation eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation habe ich in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, 17.11.2011

(Christian Wiludda)

Contents

Abbreviations

A	Adenine
A. thaliana	*Arabidopsis thaliana*
bp	Base pairs
BREd	Downstream transcription factor IIB recognition element
BREu	Upstream transcription factor IIB recognition element
°C	Degree Celsius
C_1, C_3, C_4	One-, three-, four-carbon molecule
^{13}C	Carbon-13, stable isotope of carbon
C	Cytosine
CA	Carbonic anhydrase
CO_2	Carbon dioxide
DNA	Deoxyribonucleic acid
DPE	Downstream promoter element
F.	*Flaveria*
Ft	*Flaveria trinervia*
G	Guanine
GDC	Glycine decarboxylase complex
GLDPA	Glycine decarboxylase PA gene of *Flaveria trinervia*
GUS	β-glucuronidase
h	Hour(s)
HCO_3^-	Bicarbonate (hydrogen carbonate)
Inr	Initiator
kb	Kilobases
N_2	Molecular nitrogen
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NH_3	Ammonia
MDH	Malate dehydrogenase
ME	Malic enzyme

MEM1	Mesophyll expression module 1
mRNA	Messenger ribonucleic acid
NMD	Nonsense-mediated mRNA decay
nt	Nucleotides
O_2	Molecular oxygen
OAA	Oxaloacetate
ORF	Open reading frame
PEP	Phospho*enol*pyruvate
PEPC	Phospho*enol*pyruvate carboxylase
PEPCK	Phospho*enol*pyruvate carboxykinase
3PGA	3-Phosphoglycerate
2PG	2-Phosphoglycolate
PPDK	Pyruvate, orthophosphate dikinase
P_{R2}	Distal promoter (defined by region 2 of the *GLDPA* promoter)
P_{R7}	Proximal promoter (defined by region 7 of the *GLDPA* promoter)
R	Purine (adenine or guanine)
5' RACE	Rapid amplification of 5' complementary DNA ends
RNA	Ribonucleic acid
Rubisco	Ribulose 1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
SHMT	Hydroxymethyltransferase
T	Thymine
TBP	TATA box-binding protein
T-DNA	Transfer DNA
TFII	Transcription factor for RNA polymerase II
TF	Transcription factor
TSS	Transcription start site
uORF	Upstream open reading frame
5' UTR	5' Untranslated region
Y	Pyrimidine (cytosine or thymine)

I. Introduction

1. The biochemistry of C_4 photosynthesis

1.1 The ribulose 1,5-bisphosphate carboxylase/oxygenase – a bifunctional enzyme

Terrestrial plants can convert atmospheric carbon dioxide (CO_2) into organic compounds with the energy of the sun by three different pathways. The most common one is represented by C_3 photosynthesis from which C_4 photosynthesis and the crassulacean acid metabolism (CAM) are derived (West-Eberhard et al., 2011; Wang et al., 2011).

The C_3 pathway represents the single largest flux of organic carbon in the majority of photosynthetic organisms leading to the assimilation of about 100 billion tons of carbon annually, which corresponds to 15% of the atmospheric carbon (Raines, 2011). About 85% of all plant species assimilate CO_2 by C_3 photosynthesis, including agronomically relevant crops such as wheat and rice (Ehleringer et al., 1991; Kutschera and Niklas, 2007; Bauwe et al., 2010). In C_3 plants, CO_2 is fixed by ribulose 1,5-bisphosphate (RuBP) carboxylase/oxygenase (Rubisco). Rubisco catalyzes the carboxylation of RuBP, resulting in the generation of two molecules of 3-phosphoglycerate (3PGA) as the first stable products of this cycle. As 3PGA contains three carbon atoms, plants using Rubisco as initial enzyme for CO_2 fixation are referred to as C_3 plants and the photosynthetic pathway they utilize is termed C_3 photosynthesis (Hibberd and Covshoff, 2010; Raines, 2011). 3PGA then enters the Calvin cycle resulting in the production of triose phosphates while RuBP is regenerated to serve as substrate for Rubisco again (Ogren, 1984). For the plant all carbon compounds formed in this cycle such as starch or sucrose are essential for development and growth (Raines and Paul, 2006; Smith and Stitt, 2007).

However, Rubisco is a bifunctional enzyme that can also bind oxygen (O_2) and fix it into RuBP. This initiates a process which is termed photorespiration and generates – apart from 3PGA – 2-phosphoglycolate (2PG), a toxic compound for plants (Bowes et al., 1971; Tolbert, 1971; Leegood et al., 1995). The energy-consuming recycling of 2PG to 3PGA during photorespiratory processes leads to the loss of 25%–30% of CO_2 already fixed. This reduces C_3 plant net-photosynthesis by about 20% under moderate conditions and can even have a stronger impact under certain conditions such as high temperature (Cegelski and Schaefer, 2006; Bauwe et al., 2010; Raines, 2011). As the ratio of soluble CO_2 and O_2 decreases with increasing temperature, Rubisco's oxygenase activity is favored at leaf temperatures over 20–25 °C (Ehleringer and Björkman, 1977; Ehleringer and Pearcy, 1983).

The wasteful photorespiratory effects can be counteracted by increasing CO_2 or reducing O_2, each resulting in both raised maximum photosynthetic rate and photosynthetic efficiency at limiting light (Hatch, 1987). Even though Rubisco has a higher specificity for CO_2, the present atmospheric conditions (0.035% CO_2, 21% O_2 and 78% N_2) lead to approximately 1000-fold higher O_2 concentrations compared to CO_2 in the chloroplasts of C_3 plants. This favors the fixation of O_2 by Rubisco and thereby photorespiratory processes particularly at elevated temperature (Ehleringer and Monson, 1993; Andersson, 2008; Foyer et al., 2009). Rubisco's oxygenase activity which is referred to as "Rubisco penalty" by Edwards et al. (2001b) and its low turnover rate for CO_2 fixation result in the production of large amounts of this enzyme in the plant to compensate its inefficiency. Thus, Rubisco represents the single most abundant soluble protein on earth (Edwards et al., 2004) and accounts for about 25% of nitrogen and 50% of soluble protein in C_3 leaves (Ellis, 1979; Portis and Parry, 2007).

1.2 The CO_2-concentrating mechanism of C_4 plants suppresses photorespiration

In contrast to C_3 species, C_4 plants have succeeded in overcoming the Rubisco penalty. C_4 photosynthesis represents a mechanism to concentrate CO_2 at the site of Rubisco, resulting in an increase in photosynthetic efficiency by suppressing photorespiration (Sage, 2004). The establishment of such a biochemical CO_2 pump is based on the division of labor between two distinct photosynthetically active leaf tissues, the mesophyll and the bundle-sheath (Figure 1). The bundle-sheath cells form a ring around the vascular bundles and are surrounded by the mesophyll cells. This C_4-characteristic leaf anatomy is therefore referred to as Kranz-type anatomy first described by Haberlandt (1881) (Hattersley, 1984; Dengler and Nelson, 1999). However, many C_3 plants – even *Arabidopsis thaliana* – also have bundle-sheath cells, but they are not well characterized yet and only little is known about their function (Kinsman and Pyke, 1998; Leegood, 2008).

All relevant enzymes involved in the C_4 cycle are compartmentalized into mesophyll and bundle-sheath cells. While in C_3 plants the C_3 pathway occurs in all photosynthetic cells, it is restricted to the bundle-sheath cells in C_4 plants (Hatch, 1987; Ehleringer and Monson, 1993). Within the mesophyll cells CO_2, in the form of HCO_3^-, is initially fixed by phospho*enol*pyruvate carboxylase (PEPC) into the substrate phospho*enol*pyruvate (PEP) which leads to the formation of the four-carbon acid oxaloacetate (OAA). Therefore, this pathway is referred to as C_4 photosynthesis and plants performing the C_4 cycle are termed C_4 plants. OAA is then converted into either malate or aspartate which diffuses to the bundle-sheath cells. There CO_2 is released during the decarboxylation of the transport metabolites by

one of the three enzymes, NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME) or PEP carboxykinase (PEPCK). The remaining pyruvate (NADP-ME type) or alanine (NAD-ME type) returns to the mesophyll cells where PEP is regenerated to maintain the C_4 cycle (Figure 1). In bundle-sheath cells of the PEPCK type aspartate is converted to OAA which is then decarboxylated so that PEP is directly regenerated and diffuses to the mesophyll cells in order to be carboxylated by PEPC again (Hatch, 1987; Leegood and Walker 1999). Although C_4 plants are assigned to one of these three decarboxylation types, recent findings indicate that there exists certain flexibility between the decarboxylating pathways. For instance, several C_4 plants of the NADP-ME type such as maize are able to additionally decarboxylate aspartate by PEPCK apart from the general decarboxylation of malate (Furbank, 2011).

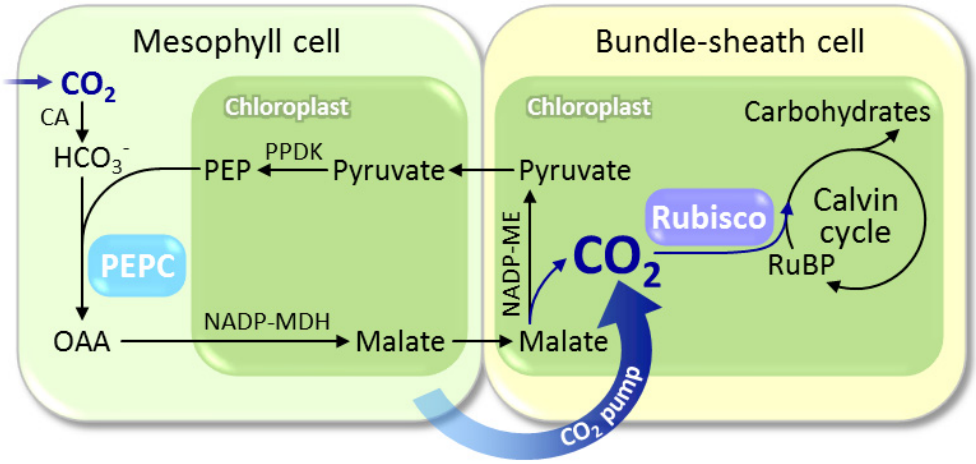


Figure 1. The reactions of the C_4 photosynthetic pathway in C_4 plants of the NADP-malic enzyme type.

C_4 photosynthesis represents a biochemical pump to concentrate CO_2 at the site of the ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) to repress its oxygenase activity and thereby photorespiration. CA, Carbonic anhydrase; NADP-MDH, NADP-malate dehydrogenase; NADP-ME, NADP-malic enzyme; OAA, Oxaloacetate; PEP, Phospho*enol*pyruvate; PEPC, Phospho*enol*pyruvate carboxylase; PPDK, Pyruvate, orthophosphate dikinase; RuBP, Ribulose 1,5-bisphosphate.

The prefixing of inorganic carbon in the mesophyll and the decarboxylation specifically in the bundle-sheath represents a biochemical pump that causes elevated CO_2 concentrations within bundle-sheath cells (Figure 1). This almost saturates the active site of Rubisco and suppresses its oxygenase activity, resulting in the efficient repression of photorespiration (von Caemmerer and Furbank, 2003; Sage, 2004; Furbank, 2011). Compared to C_3 species, in most of the C_4 plants photorespiration is almost undetectable. Nevertheless, photorespiratory processes are not completely abolished but occur even in C_4 species (Yoshimura et al., 2004). It was shown that a photorespiratory maize (C_4) mutant deficient in activity of glycolate oxidase survived only in the presence of increased CO_2 concentrations that suppress

photorespiration. This provides evidence that photorespiration is needed in C_4 plants to avoid accumulation of toxic glycolate (Zelitch et al., 2009).

1.3 Specific adaptations and characteristics of C_4 plants

PEPC exhibits no oxygenase activity and has a higher affinity for HCO_3^- than Rubisco for CO_2. Therefore, this carboxylase can bind adequate amounts of inorganic carbon even in the presence of low CO_2 concentrations in the leaf that occur as the result of stomatal closure due to heat or drought (Hatch, 1987; Kanai and Edwards, 1999). The maintenance of efficient photosynthetic rates in the C_4 leaf, even when stomata are closed, reduces transpiration effectively which is reflected in a better water-use efficiency compared to C_3 species (Long, 1999). Thus, many C_4 plants even grow in arid and hot habitats with high light intensities (Ehleringer et al., 1997).

Suberized lamellae are present in the outer cell walls of the bundle-sheath cells of many C_4 plants. They represent a diffusion barrier that diminishes the CO_2 efflux in order to maintain high CO_2 concentrations within the bundle-sheath. However, all NAD-ME species lack the suberin layer but appear to compensate this by tending to concentrate chloroplasts at the inner or centripedal side of the bundle-sheath cell around mitochondria. Presumably, this decreases the loss of CO_2 released by decarboxylation in mitochondria and retards its leakage because the large vacuole has to be passed (Hattersley and Browning, 1981; Dengler and Nelson, 1999; von Caemmerer and Furbank, 2003; Sage, 2004; Yoshimura et al., 2004).

The chloroplasts of C_4 plants in the two distinct photosynthetically active tissues are different with respect to their enzymes and the accumulation of starch. They also have a dimorphic ultrastructure: The bundles-sheath chloroplasts from C_4 plants of the NADP-ME type differ from those of the mesophyll cells in the reduction of grana where O_2 is produced by the water-splitting complex of photosystem II. The decrease of grana in chloroplasts of bundle-sheath cells prevents the enrichment of O_2 through photosynthetic processes and thereby suppresses Rubisco's oxygenase activity (Meierhoff and Westhoff, 1993; Edwards et al., 2004; Yoshimura et al., 2004).

As Rubisco is restricted to the bundle-sheath cells and not present in mesophyll cells, C_4 species do not waste that much nitrogen in forming this enzyme. This results in three to six times less amounts of Rubisco in comparison to C_3 plants (Ku et al., 1979; Sage et al., 1987). Generally, the nitrogen content of the leaf is reduced in C_4 plants, which enables them to achieve higher nitrogen-use efficiencies than C_3 species (Brown, 1978; Hatch, 1987; Sage and Pearcy, 1987).

1.4 The phenomenon of single-cell C_4 photosynthesis

Apart from the well-known two-cell or dual-cell C_4 photosynthesis of Kranz-type C_4 plants, single-cell C_4 photosynthesis exists. In each photosynthetic cell of the two aquatic species *Hydrilla verticillata* and *Egeria densa*, CO_2 is fixed by PEPC in the cytoplasm and concentrated in chloroplasts where Rubisco and NADP-ME are localized. Despite the absence of a diffusion barrier to reduce the leakage of CO_2 out of the chloroplast, the intracellular C_4 cycle in *Hydrilla* leads to increased carbon yield at low CO_2 concentrations that predominantly occur in ponds at elevated temperature (Reiskind et al., 1997; Casati et al., 2000; Bowes et al., 2002). The two terrestrial plants *Bienertia cycloptera* and *Borszczowia aralocaspica* also perform single-cell C_4 photosynthesis. Similar to the aquatic species, the PEPC of *Bienertia* accumulates in the cytoplasm within distinct pockets, whereas Rubisco and the decarboxylating enzymes are present in chloroplasts which are assembled as a central core in each cell. In contrast, the photosynthetic cells of *Borszczowia* are bipolar. The carboxylation and regeneration of PEP occur at the end in contact with the intercellular air space, whereas decarboxylation by NADP-ME and CO_2 fixation by Rubisco are performed exclusively at the cell's end in proximity to the vascular bundles (Edwards et al., 2004).

1.5 C_4 plants are highly productive in warm habitats

The anatomical and biochemical adaptations of C_4 plants to efficiently concentrate CO_2 at the site of Rubisco lead to the suppression of photorespiration. This increases the photosynthetic rates compared to those of C_3 plants at higher temperatures. Warmth promotes the insolubility of CO_2 which favors the binding of O_2 by Rubisco and thereby photorespiratory processes predominantly in C_3 species that lack a CO_2-concentrating mechanism to suppress Rubisco's oxygenase activity (Sage, 2004). Thus, in warmer habitats the efficiency of C_4 photosynthesis and the dominance of C_4 species among the grasses of the tropical and subtropical climates are responsible for the productivity of C_4 plants. C_4 species contribute approximately 20–30% to the terrestrial biomass production despite the fact that they only constitute about 3% of all vascular plant species (Lloyd and Farquhar, 1994; Gillion and Yakir, 2001; Edwards et al., 2010). For this reason, C_4 species such as maize, sorghum and sugar cane belong to the most productive agricultural crops (Brown, 1999).

2. Evolution of the C_4 syndrome

2.1 The polyphyletic origin of C_4 photosynthesis

C_4 plants have evolved several times independently from C_3 ancestors among the advanced members of the Dicotyledonae (dicots) and Monocotyledonae (monocots) (Ehleringer et al., 1997). The evolution of the C_4 photosynthetic pathway occurred at least 62 times in 19 different families of the angiosperms including approximately 7500 species. Thirty-six of the 62 lineages are found in the eudicots, while 26 lineages occur in the monocots with 18 and six lineages belonging to the grasses and sedges respectively (Sage et al., 2011). Although the majority of evolutionary C_4 lineages are distributed among the approximately 165000 dicots, only about 1600 (~ 1%) of them represent C_4 species. In contrast, one-third of the Poales (monocots) with about 18000 C_3 and C_4 species in total are C_4 plants with approximately 4600 (~ 25%) C_4 grasses and 1300 (~ 7%) C_4 sedges (Ehleringer et al., 1997; Sage et al., 1999a; Bruhl and Wilson, 2007; Roalson, 2011; Sage et al., 2011). Thus, C_4 photosynthesis is a prime example of convergent evolution (Muhaidat et al., 2007).

All available evidence indicates that C_4 photosynthesis is of rather recent evolutionary origin (Ehleringer et al., 1997). Presumably all lineages of C_4 plants evolved during the last 30 million years. It was shown that C_4 origins in dicots were contemporaneous with those in monocots (Christin et al., 2011). C_4 dicots should therefore not be considered evolutionary younger than C_4 monocots as it was previously proposed (Ehleringer et al., 1997; Sage, 2004; Christin et al., 2011). Based on $\delta^{13}C$ evidence the drastic increase of C_4 plant biomass is dated back six to eight million years ago (Edwards et al., 2001b).

C_4 dicots evolved in arid regions of low latitude, indicating that environmental factors such as high temperatures, drought and salinity might have forwarded C_4 evolution. Forty-seven of the 62 lineages originated geographically in areas of southwestern North America, south-central South America, northeastern and southern Africa, central Asia and inland Australia (Sage, 2004; Sage et al., 2011). Nowadays, almost all tropic and subtropic grasslands and those of the warm temperate zones are dominated by C_4 grasses and sedges which account for the majority of the plant species of arid regions from the tropics to the temperate zones (Archibold, 1995; Sage et al., 1999b)

2.2 C_4 photosynthesis as an evolutionary adaptation to counteract photorespiration

The decrease of CO_2 in the atmosphere in the late Oligocene increased photorespiration and might be the major selective force for C_4 photosynthesis to evolve (Ehleringer et al., 1991; Sage et al., 2011). The occurrence of C_4 monocots is tightly correlated with temperature, whereas the distribution of C_4 dicot plants is rather influenced by other environmental factors such as aridity, indicating that these parameters promote C_4 evolution (Ehleringer et al., 1997). Thus, the evolution of C_4 photosynthesis has to be considered as an adaptation to reduce photorespiration. Environmental factors such as low CO_2 concentration, drought, increased salinity, low humidity and high temperatures enhance photorespiratory processes and CO_2 deficiency, which presumably led to the selection for C_4-characteristic traits and finally to the establishment of C_4 photosynthesis (Ehleringer and Monson, 1993; Sage, 2004).

Another evolutionary adaptation to counteract photorespiration even at decreasing CO_2 levels might have resulted in a Rubisco variant that is free of oxygenase activity. Even though different Rubisco forms with varying specificities for CO_2 relative to O_2 have evolved, it seems as if limits in optimizing Rubisco have been reached. The active site biochemistry of Rubisco is restricted by similarities in the carboxylase and oxygenase reactions which apparently allow no further improvement of this enzyme (Andrews and Lorimer, 1987; Roy and Andrews, 2000; Tcherkez et al., 2006; Kapralov et al., 2011). Hence, the CO_2-concentrating mechanism underlying C_4 photosynthesis represents the only effective evolutionary adaptation to efficiently suppress photorespiration promoted predominantly by heat and drought.

2.3 All enzymes involved in C_4 photosynthesis are already present in C_3 plants

The polyphyletic origin of C_4 photosynthesis indicates that the conversion from C_3 towards C_4 plants did not require drastic changes but might have occurred rather easily (Westhoff and Gowik, 2004; Brown et al., 2011; Gowik and Westhoff, 2011). It is assumed that the reorganization of metabolic processes which already exist in C_3 species have led to the evolution of C_4 photosynthesis (West-Eberhard et al., 2011). Generally, during complex trait evolution the existing biochemistry is exploited rather than creating novel enzymes. Thus, changes in the kinetics, the regulation and tissue-specificities of the C_3 enzymes were necessary to establish a functional CO_2-concentrating mechanism (Doebley and Lukens, 1998; Sage, 2004).

Actually, all enzymes involved in C_4 photosynthesis do not represent evolutionary reinventions but are already present in C_3 plants where they are embedded in the carbohydrate and nitrogen metabolism (Sage, 2004; Aubry et al., 2011; Brown et al., 2011). The C_3 PEPC, for example, is involved in anaplerotic reactions, amino acid synthesis, the regulation of turgor-dependent movements such as stomatal opening, the fiber elongation of *Gossypium hirsutum* (cotton) and ammonium assimilation in *Oryza* (rice) (Cockburn, 1983; Miyao and Fukayama, 2003; Cousins et al., 2007; Li et al., 2010; Masumoto et al., 2010). The three C_4 acid decarboxylases PEPCK, NADP-ME and NAD-ME have various metabolic roles in C_3 plants which differ depending on tissue type and the developmental stage (Aubry et al., 2011). PEPCK can be found in stomatal guard cells, fruits, roots and vascular tissues where it may contribute to anaplerotic reactions in the phloem (Walker et al., 1999; Leegood and Walker, 2003; Brown et al., 2010). In germinating seeds of C_3 species sugars are mobilized from lipids by PEPCK via gluconeogenesis (Rylott et al., 2003). NADP-ME and NAD-ME have various housekeeping functions in C_3 plants (Wedding, 1989; Edwards and Andreo, 1992; Drincovich et al., 2001). NADP-ME contributes to the oxidative pentose phosphate pathway, lignin and lipid biosynthesis (Gerrard-Wheeler et al., 2005, 2008). NAD-ME was shown to be important for coordinating carbon and nitrogen metabolism (Tronconi et al., 2008).

2.4 The stepwise transition from C_3 to C_4 photosynthesis during C_4 evolution

Sage (2004) proposes a model that includes the main phases of C_4 evolution as stepwise transition from C_3 to C_4 photosynthesis (Figure 2). Respective steps and certain developmental stages might have occurred earlier or later in singles evolutionary lineages. The basis for the evolution of the C_4 photosynthetic pathway was the duplication of relevant genes so that multiple copies of a single gene were present. As general preconditioning, one or more copies could then undergo modifications which might have resulted in neo- or nonfunctionalization, while the original gene maintained its function (Marshall et al., 1996; Lynch and Conery, 2000; Monson, 2003). Afterwards the establishment of anatomical preconditions occurred, such as the reduction of the distance between mesophyll and bundle-sheath cells by decreasing the interveinal space and/or enlarging the bundle-sheath. The activation of bundle-sheath cells included the increase of the number of chloroplasts and mitochondria due to the progressive importance of these cells in photosynthesis (Sage, 2004).

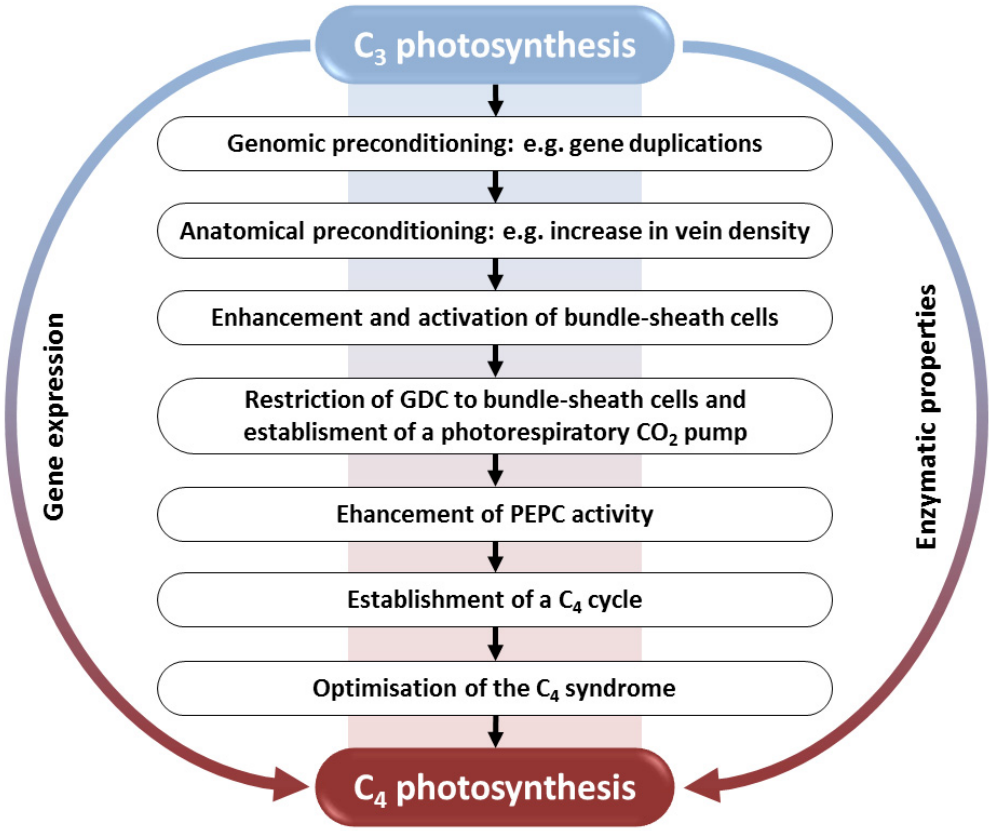


Figure 2. The stepwise evolution of C_4 photosynthesis.

This model represents the main phases during the evolution of the C_4 syndrome. GDC, Glycine decarboxylase complex; PEPC, Phospho*enol*pyruvate carboxylase. Adapted from Sage (2004) and Gowik and Westhoff (2011).

A very crucial step towards C_4 photosynthesis was the restriction of the glycine decarboxylase complex (GDC) to the bundle-sheath cells. GDC converts glycine to serine while releasing CO_2 during photorespiration. The absence of functional GDC in the mesophyll reduces the loss of photorespiratory CO_2 in this tissue because the decarboxylation of glycine is prevented. In the bundle-sheath cells glycine can be decarboxylated by GDC. This imbalance causes a constant flow of glycine from the mesophyll into bundle-sheath cells where the enrichment of photorespiratory CO_2 by GDC favors Rubisco's carboxylation reaction. Such a functional photorespiratory CO_2 pump (Figure 3) represents a mechanism to concentrate CO_2 in bundle-sheath cells to suppress the oxygenase activity of Rubisco and thereby photorespiration in this tissue (Rawsthorne, 1992; Bauwe and Kolukisaoglu, 2003; Sage, 2004; Gowik et al., 2011; Gowik and Westhoff, 2011).

In C_3 species, photorespiration occurs in all photosynthetically active cells because Rubisco and GDC are present in these tissues (Rawsthorne et al., 1988; Yoshimura et al., 2004). In contrast to C_3, in C_4 plants Rubisco and GDC and hence photorespiratory processes are restricted to the bundle-sheath cells (Ohnishi and Kanai, 1983; Hatch, 1987; Morgan et al.,

1993; Koprivova et al., 2001; Yoshimura et al., 2004; Sudderth et al., 2007). C_3-C_4 intermediate species which represent an evolutionary link in the transition from C_3 to C_4 plants lack functional GDC in mesophyll cells, which is consistent with the theory of relocating GDC to the bundle-sheath during C_4 evolution (Rawsthorne et al., 1988; Rawsthorne, 1992; Morgan et al., 1993). In the C_3-C_4 intermediate *Moricandia arvensis* the P-subunit of GDC is restricted to the bundle-sheath cells, whereas the other three GDC subunits are still present within the mesophyll cells. This prevents the assembly of a functional GDC in the mesophyll (Hylton et al., 1988; Morgan et al. 1993).

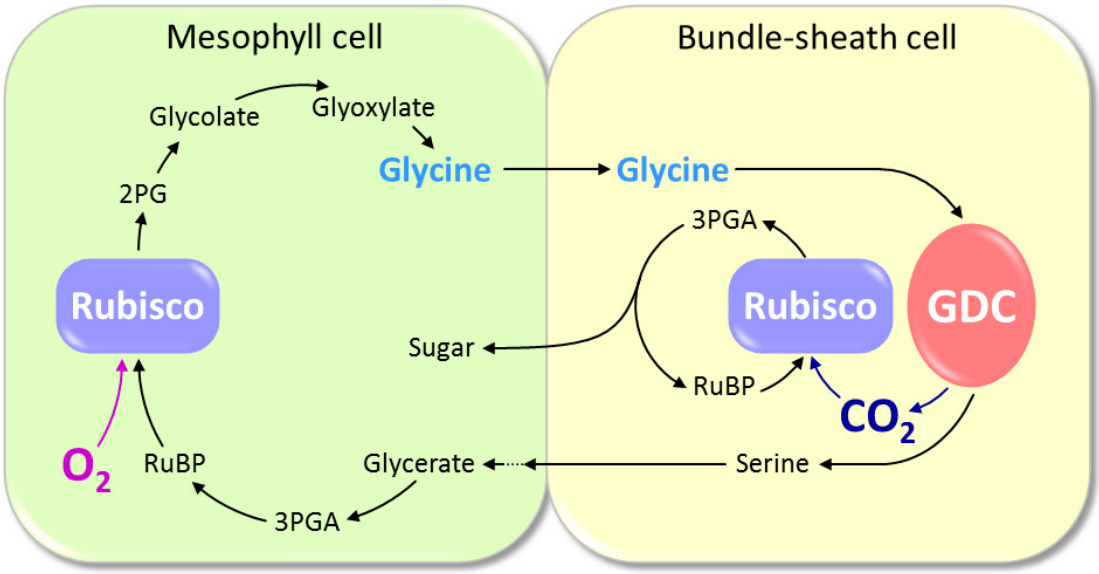


Figure 3. The photorespiratory CO_2 pump.

The constant flux of glycine into the bundle-sheath cells and its decarboxylation by the glycine decarboxylase complex (GDC) exclusively in these cells leads to the enrichment of CO_2 and the suppression of the oxygenase activity of the ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco). 2PG, 2-Phosphoglycolate; 3PGA, 3-Phosphoglycerate; RuBP, Ribulose 1,5-bisphosphate. Adapted from Sage (2004).

For establishment of a true C_4-type CO_2 pump PEPC activity had to be enhanced in the mesophyll cells for efficient fixation of CO_2. Then, the produced C_4 acids could be directed to the bundle-sheath cells where they were decarboxylated, resulting in the enrichment of CO_2 (Monson, 1999). In intermediates of the genus *Flaveria*, PEPC activity increases approximately 40-fold from C_3 to C_4 species. In *F. linearis* and *F. ramosissima* (both C_3-C_4 intermediates) and *F. brownii* (C_4-like intermediate) the activity values are five, seven and 20 times greater than the C_3 values respectively (Monson and Moore, 1989; Monson and Rawsthorne, 2000; Svensson et al., 2003).

In order to operate a functional C_4 pathway the expression pattern of most of the enzymes involved in C_4 photosynthesis had to be reorganized to allow the strict compartmentalization

into either mesophyll or bundle-sheath cells, which is achieved by differential gene expression (Wyrich et al., 1998; Sheen, 1999; Hibberd and Covshoff, 2010; Gowik et al., 2011; West-Eberhard et al., 2011). Compared to their C_3 homologs, the C_4-specific genes are highly expressed and exhibit organ- and cell-specific expression patterns. Mesophyll gene expression is predominantly regulated at the transcriptional level, while bundle-sheath-specific gene expression is controlled transcriptionally and posttranscriptionally (Sheen, 1999; Edwards et al., 2001a; Gowik et al., 2004; Patel et al., 2006). Even little variations in one single *cis*-regulatory module can alter the gene expression pattern to mesophyll-specificity in C_4 plants (Akyildiz et al., 2007). This indicates that the establishment of C_4-characteristic gene expression did not require drastic alterations but could have been implemented rather easily in genetic terms (Westhoff and Gowik, 2004; Gowik and Westhoff, 2011). Doebley and Lukens (1998) conclude that changes in the *cis*-regulatory elements of transcriptional regulators are the predominant driving force for the generation of novel phenotypes. However, *cis*-regulatory elements of genes from C_3 species can already be recognized in a way that ensures their recruitment into C_4 photosynthesis without alterations to sequence, suggesting that *trans*-acting factors had to be modified to confer cell-specific expression (Brown et al., 2011). There are even cell-specific promoters of C_4 genes that lead to mesophyll- or bundle-sheath-specific gene expression in closely or even widely related C_3 plants, which indicates that the genetic regulatory mechanisms are generally controlled in a very similar way in C_3 and C_4 plants (Matsuoka et al., 1994; Hibberd and Covshoff, 2010; Westhoff and Gowik, 2010). In C_3-C_4 intermediate species, many of the enzymes required for the performance of a functional C_4 photosynthesis are already restricted to one of the two distinct photosynthetic tissues, while Rubisco is present in both, mesophyll and bundle-sheath cells (Hylton et al., 1988; Rawsthorne, 1992; Drincovich et al., 1998; Monson and Rawsthorne, 2000).

The last evolutionary step towards C_4 photosynthesis is the optimisation of photosynthetic efficiency. This includes the alteration of kinetic properties and regulatory characteristics of various enzymes to adjust them to a functional C_4-metabolic environment (Leegood and Walker, 1999; Westhoff and Gowik, 2004). PEPC is a good example to study kinetic adaptation that occurred during C_4 evolution. The C_3 variant of this enzyme is normally inhibited by malate. As malate is present in high concentrations in the mesophyll cells of C_4 plants, the sensitivity to malate of the C_4 PEPC is reduced, while the activation through glucose-6-phosphate is enhanced (Svensson et al., 1997; Bläsing et al., 2000; Svensson et al., 2003; Westhoff and Gowik, 2004).

2.5 The genus Flaveria as model system to study C_4 evolution

C_4 photosynthesis is a convergent and complex evolutionary phenomenon and represents an excellent system to study the mechanisms of evolutionary adaptation in response to environmental change (Monson, 2003; Sage et al., 2011). The genus *Flaveria* is especially suitable to analyze the evolutionary development of C_4 photosynthesis because it contains species that perform C_3 (e.g. *F. pringlei*) or C_4 photosynthesis (e.g. *F. trinervia* and *F. bidentis*) as well as C_3-C_4 intermediate (e.g. *F. pubescens* and *F. ramosissima*) and C_4-like (e.g. *F. brownii*) species (Powell, 1978; McKown et al., 2005). *Flaveria* therefore represents a distinguished model system to study molecular mechanisms underlying the evolutionary transition from C_3 to C_4 photosynthesis (Westhoff and Gowik, 2004; Brown et al., 2005; Akyildiz et al., 2007; Engelmann et al., 2008).

3. The transcriptional control region of eukaryotic protein-coding genes

3.1 Structure of the eukaryotic RNA polymerase II-dependent promoter

In eukaryotes, the RNA polymerase II-dependent promoter comprises *cis*-regulatory elements located at the 5' end of a gene near the transcription start site (TSS). It consists of a core promoter for basal transcription initiation and a regulatory promoter region, often referred to as proximal promoter, directly upstream of the core promoter (Figure 4). The regulatory promoter contains binding sites for transcription factors that increase the efficiency of transcription initiation. Further *cis*-regulatory sequences, so-called enhancers, interact with the promoter and control the spatial and temporal gene expression pattern (Levine and Tjian, 2003; Vedel and Scotti, 2011).

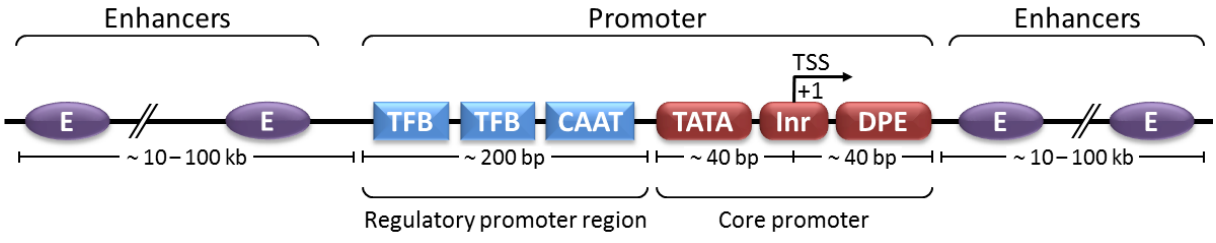


Figure 4. Schematic structure of the transcriptional control region of eukaryotic protein-coding genes.

Typical *cis*-regulatory elements of the core promoter are the TATA-Box (TATA), the Initiator sequence (Inr) and the downstream promoter element (DPE) which all initiate transcription. Proximal transcription factor binding sites (TFB) such as the CAAT-Box (CAAT) of the regulatory promoter region affect the efficiency of the core promoter. Enhancers (E) are located up- and/or downstream of the promoter and control the gene expression pattern. The approximate distances are indicated in base pairs (bp) or kilobases (kb). The arrow at position +1 represents the transcription start site (TSS). Adapted from Vedel and Scotti (2011).

3.2 Cis-regulatory elements of the core promoter

The core promoter commonly ranges from -40 bp to +40 bp with regard to the TSS at position +1 and represents the DNA region that recruits the RNA polymerase II machinery to initiate transcription precisely (Juven-Gershon and Kadonaga, 2010). Typical core promoter elements are the TATA box, the upstream/downstream transcription factor IIB (TFIIB) recognition element (BREu/BREd), the Initiator motif (Inr), the motif ten element (MTE) and the downstream core promoter element (DPE). These elements are normally present in focused core promoters that are characterized by a single TSS or a distinct cluster of TSSs within a short nucleotide region. However, these elements do not appear in all core promoters, indicating that they are not universal. Even core promoters exist that contain none of these elements at all, which suggests that the structure and function of core promoters is widely diverse (Juven-Gershon et al., 2008; Vedel and Scotti, 2011). Various general transcription factors, including transcription factor for RNA polymerase II (TFII) A, B, D, E, F and H, bind to *cis*-regulatory elements of the core promoter to recruit RNA polymerase II for RNA synthesis (Juven-Gershon and Kadonaga, 2010). The location of core promoter elements to which proteins of the transcription complex bind regulates promoter function by determining the position of the TSS and the direction of transcription (Tsai and Sigler, 2000; Gershenzon et al., 2006).

For example, the Inr which encompasses the TSS is assumed to interact with TFIID, a multi subunit complex composed of the TATA box-binding protein (TBP) and various TBP-associated factors (TAFs) (Smale and Baltimore, 1989; Purnell et al., 1994; Smale and Kadonaga, 2003). Inr motifs were shown to be present also in plants where they are even able to compensate the lack of the TATA box by initiating transcription on its own (Nakamura et al., 2002). In rice and *A. thaliana* the YR (Y = C or T, R = A or G) consensus sequence (YR rule) at position -1/+1 (+1 represents the TSS) is considered to represent a less stringent form of Inr with the CA and TA sequence occurring most frequently (Yamamoto et al., 2007) (Figure 5).

The TATA box, detected by Goldberg (1979), is normally located at -31 bp or -30 bp with regard to the TSS (Carninci et al., 2006; Ponjavic et al., 2006) and is bound by TBP associated with the remaining subunits of the TFIID complex (Patikoglou et al., 1999). BREu and BREd flank the TATA box and interact with TFIIB. They were shown to interfere with the TATA box, resulting in an increase or decrease of the basal transcription levels (Lagrange et al., 1998; Deng and Roberts, 2005, 2007). As in other eukaryotes, the TATA box is a common core promoter motif in plants where it is located usually 25–40 bp upstream of the

TSS with a space of 32 bp in between occurring most frequently (Figure 5). However, only 20–30% of all promoters of *A. thaliana* contain a TATA box or a TATA variant (Joshi, 1987; Molina and Grotewold, 2005; Yamamoto et al., 2009; Bernard et al., 2010; Zuo and Li, 2011). These findings are consistent with the observation that less than 20% of genes in human and yeast contain a TATA box (Basehoar et al., 2004; Shi and Zhou, 2006). Zuo and Li (2011) were able to show that in plants about 35% of the examined TATA-less promoters harbored TC-elements preferentially in a range of -50 bp to +50 bp with regard to the TSS. These TC-elements – TC-rich DNA sequences with a length of 6 bp (Figure 5) – might represent a novel class of regulatory elements controlling transcription in plants (Bernard et al., 2010). In the absence of a TATA box, TBP-related factors (TRFs) are involved in transcription by RNA polymerase II (Juven-Gershon and Kadonaga, 2010).

Based on transcriptome analysis of *A. thaliana* it was shown that genes which contain a TATA box exhibit tissue-specific expression profiles that can be regulated in response to a variety of stimuli. In contrast, genes with coreless promoters that have no recognizable core element are expressed constitutively (Schug et al., 2005; Yamamoto et al., 2009, 2011). Promoters that harbor a TC-element are expressed in specific conditions (Bernard et al., 2010), indicating that TC-elements are important for gene regulation as response to environmental or developmental stimuli. Thus, the architecture of the core promoter is very important for the regulation of gene expression. For example, the light-regulated promoter of the *psaDb* gene encoding the ferredoxin-binding subunit of PSI from *Nicotiana sylvestris* contains an Inr motif but lacks a TATA box. By interchanging the Inr with a TATA box it was shown that the presence of the Inr is essential for the transcriptional regulation of *psaDb* in response to light (Nakamura et al., 2002).

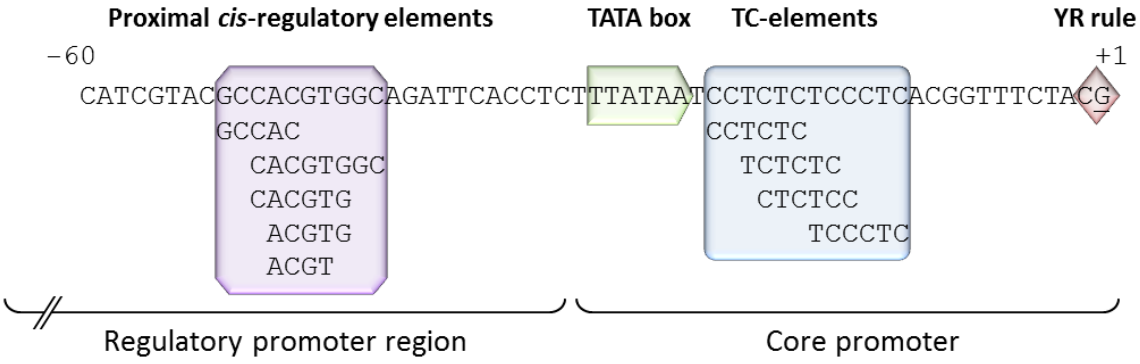


Figure 5. Molecular structure of a typical plant promoter.

The *cis*-regulatory elements of the promoter of the Arabidopsis *At1g10960* gene encoding ferredoxin 1 are shown. The YR rule (-1/+1; Y = C or T, R = A or G), TC-elements and the TATA box of the core promoter and several proximal *cis*-regulatory elements of the regulatory promoter region such as ABRELATERD1 (ACGTG), ACGTATERD1 (ACGT), CACGTGMOTIF (CACGTG), EMBP1TAEM (CACGTGGC) and SORLIP1AT (GCCAC) are highlighted. The transcription start site (+1) is underlined. Adapted from Yamamoto et al. (2007).

3.3 Enhancers, silencers and insulators influence gene expression

The core promoter is important for basal transcription but can be influenced by further *cis*-acting regulatory sequences. Among these are enhancers that are usually located upstream of the TSS but can also be present downstream of it, for example in introns of the corresponding genes (Figures 4 and 6). The distance of an enhancer with regard to the TSS is not defined so that they can be found up to several kilobases up- and/or downstream of the respective gene. Enhancers are usually responsible for the expression of a gene in a tissue- or cell-specific pattern. *Cis*-regulatory elements represent conserved short motifs of a length of five to 20 nucleotides that are specifically bound by *trans*-acting regulatory proteins, referred to as transcription factors (TFs) which then influence transcription. Each of the at least 1500 TFs identified in plants regulates the expression of hundreds of target genes (Levine and Tjian, 2003; Rombauts et al., 2003; Vedel and Scotti, 2011). By binding appropriate TFs, enhancers can activate or stimulate transcription regardless of their location and orientation with respect to the TSS. A typical enhancer has a length of about 500 bp and is composed of multiple binding sites for different sequence-specific transcription factors (Banerji et al., 1981; Levine and Tjian, 2003; Arnosti and Kulkarni, 2005; Bulger and Groudine, 2011).

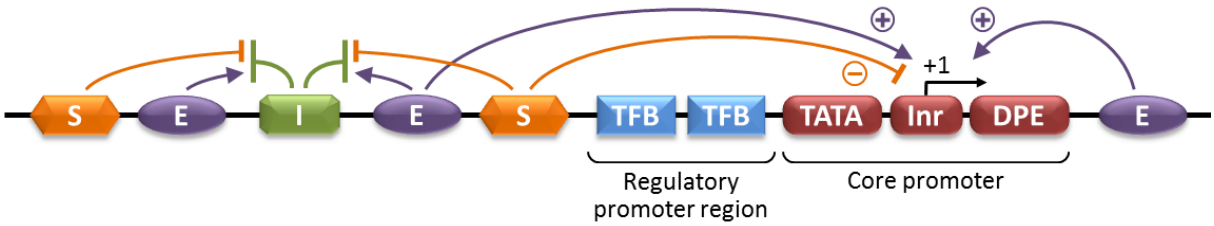


Figure 6. The complex interplay of *cis*-regulatory sequences in the transcriptional control region of multicellular eukaryotes.

The activity of the promoter can be increased by enhancers (E) or repressed by silencers (S). Insulators (I) can block the enhancing or silencing effect when they are located between the promoter and the enhancer or silencer. The position +1 represents the transcription start site. DPE, Downstream promoter element; Inr, Initiator; TATA, TATA-Box; TFB, Transcription factor binding site. Adapted from Levine and Tjian (2003).

Apart from enhancers there are further *cis*-regulatory sequences, silencers and insulators, that influence gene expression (Figure 6). Silencers are specifically recognized by TFs which inhibit transcription by heterochromatin formation. Insulators are elements which insulate genes by restricting the activity of long-range enhancers and silencers in a position-dependent manner. Insulators which are located between the enhancer and its interacting promoter are termed enhancer-blockers, whereas those insulators that are present between a silencer and a promoter and thereby shield the promoter from being silenced are referred to as barriers (Valenzuela and Kamakaka, 2006; Raab and Kamakaka, 2010). These different *cis*-regulatory

sequences, enhancers, silencers and insulators, can be distributed over distances of 100 kb in mammals and 10 kb in Drosophila (Levine and Tjian, 2003). The combination of all *cis*-regulatory elements and the integration of signals of various TFs result in a precise and specific temporal and spatial expression pattern (Priest et al., 2009; Vedel and Scotti, 2011).

3.4 The mesophyll expression module 1 for C_4-specific gene expression

Little is known about the molecular structure of *cis*-elements and *trans*-regulatory factors required for the C_4-specific expression of genes. Only one *cis*-regulatory module for specific expression in mesophyll cells could have been described at the nucleotide level so far (Figure 7). The 41-bp mesophyll expression module 1 (MEM1) is located within the distal promoter region of the *ppcA1* gene of the C_4 plant *Flaveria trinervia* which encodes the C_4-isoform of PEPC. This C_4-MEM1 confers mesophyll-specificity in transgenic *Flaveria bidentis* (C_4) by enhancing mesophyll expression and repressing expression in the bundle-sheath cells and the vascular bundles. It contains the tetranucleotide CACT and a G to A substitution in comparison to the C_3-MEM1 of the orthologous *ppcA1* gene from the C_3 plant *Flaveria pringlei*. The C_3-MEM1 does not direct mesophyll-specific gene expression in transgenic *Flaveria bidentis*. These small changes in the molecular composition of MEM1 during C_4 evolution emphasize that the transition from C_3- to C_4-characteristic gene expression might have occurred in easy steps (Gowik et al., 2004; Akyildiz et al., 2007; Westhoff and Gowik, 2010). However, no bundle-sheath-specific *cis*-regulatory element has been characterized on the nucleotide level so far. Thus, it would be interesting to identify such elements to unravel the molecular basis underlying bundle-sheath-specific gene expression in the context of C_4 evolution.

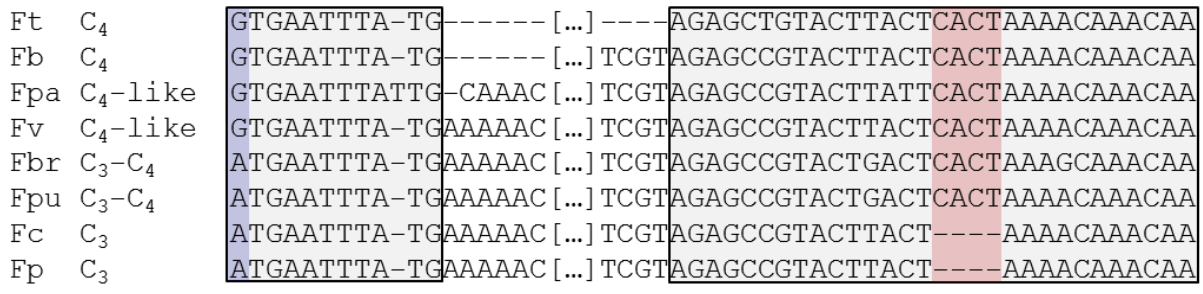


Figure 7. Molecular structure of the mesophyll expression module 1 of the *ppcA1* gene in various *Flaveria* species.

The structure of the 41-bp mesophyll expression module 1 (MEM1), emphasized with grey boxes, is shown on the nucleotide level as it can be found in different C_4, C_4-like, C_3-C_4 intermediate and C_3 *Flaveria* species. The tetranucleotide CACT and the G to A substitution are highlighted in red and blue respectively. *Fb*, *F. bidentis*; *Fbr*, *F. brownii*; *Fc*, *F. cronquistii*; *Fp*, *F. pringlei*; *Fpa*, *F. palmerii*; *Fpu*, *F. pubescens*; *Ft*, *F. trinervia*; *Fv*, *F. vaginata*. Adapted from Akyildiz et al. (2007).

3.5 The phenomenon of multiple transcription start sites in plants

Although representing a rare phenomenon in plants, the presence of more than one transcription start site (TSS) in the promoter region of the same gene can occur. These multiple TSSs serve as control mechanism to initiate gene expression in different organs of the plant (Bassett et al., 2004), to respond to environmental stimuli differently (Lee et al., 1994) or to provide different protein isoforms targeted either to the chloroplast or the cytoplasm (Matsuoka et al., 1988; Sheen, 1991; Rosche and Westhoff, 1995; Luo et al., 1997; Parsley and Hibberd, 2006).

The promoter of the *inrpk1* gene of *Ipomoea nil* (morning glory) encoding a leucine-rich receptor protein kinase contains three different TATA boxes. Two of them initiate transcription predominantly in leaves and cotyledons, whereas the third TATA box contributes only to root-specific transcription. These findings indicate that the organ-specific expression of the *inrpk1* gene is regulated by the presence of multiple TSSs within its promoter (Bassett et al., 2004).

In the promoter of the *PAL5* gene of tomato encoding the phenylalanine ammonia-lyase, two TSSs lead to the synthesis of a short and a long transcript variant. Under normal conditions basal transcription from both TSSs occurs. Environmental factors, such as wounding, light and pathogen infection, preferentially stimulate the accumulation of the short RNA variant (Lee et al., 1994). This demonstrates certain flexibility in controlling gene expression in response to stress due to the presence of two different TSSs.

The pyruvate, orthophosphate dikinase (*PPDK*) gene of the C_4 plant maize is regulated by two TSSs, leading to the generation of a long and a short transcript version. The long transcript gives rise to a protein which is located in chloroplasts because the first exon encodes an appropriate transit peptide. The short transcript lacks the first exon so that no chloroplastic targeting sequence can be translated, resulting in the formation of a cytosolic protein variant (Matsuoka et al., 1988; Sheen, 1991). The same transcriptional regulation of the *PPDK* gene can be found in *Flaveria trinervia* (C_4) and *Arabidopsis thaliana* (C_3) (Rosche and Westhoff, 1995; Parsley and Hibberd, 2006). Two transcripts generated at different TSSs within the promoter of the carrot gene encoding dihydrofolate reductase-thymidylate synthase also lead to the translation of two protein isoforms. One variant is targeted to the chloroplast, whereas the other isoform is retained in the cytoplasm (Luo et al., 1997). These findings demonstrate that two different TSSs of the same gene can cause a different subcellular distribution of the encoded protein variants, namely to the chloroplast or the cytoplasm.

4. The glycine decarboxylase complex

4.1 Composition and reaction mechanism of the glycine decarboxylase complex

The mitochondrial glycine decarboxylase multienzyme complex (GDC) is present in plants, animals and bacteria. It catalyzes the conversion of two molecules of glycine to one molecule of serine, CO_2 and NH_3 in cooperation with the serine hydroxymethyltransferase (Oliver, 1994) (Figure 8). Phylogenetic analyses have shown that the GDC subunits in mitochondria of plants and algae originated from α-proteobacteria (Kern et al., 2011). GDC accounts for approximately one-third of the soluble proteins in the mitochondrial matrix of pea leaves. This leads to concentrations up to 0.13 g/ml that can even change the density of mitochondria (Oliver et al., 1990a, 1990b; Vauclare et al., 1996).

GDC is composed of four different subunits, the P-, H-, T- and L-subunit which are assembled with a predicted stoichiometry of $1L_2$:$2P_2$:27H:9T with the L- and P-subunit acting as homodimers respectively (Oliver et al., 1990b). The P-subunit is a pyridoxal phosphate-containing protein that catalyzes the release of CO_2 by decarboxylation, the T-subunit or aminomethyl-transferase is required for the tetrahydrofolate-dependent reaction, the lipoamide-containing H-subunit functions as carrier protein and the L-protein represents a dihydrolipoamide dehydrogenase (Kikuchi et al., 2008) (Figure 8).

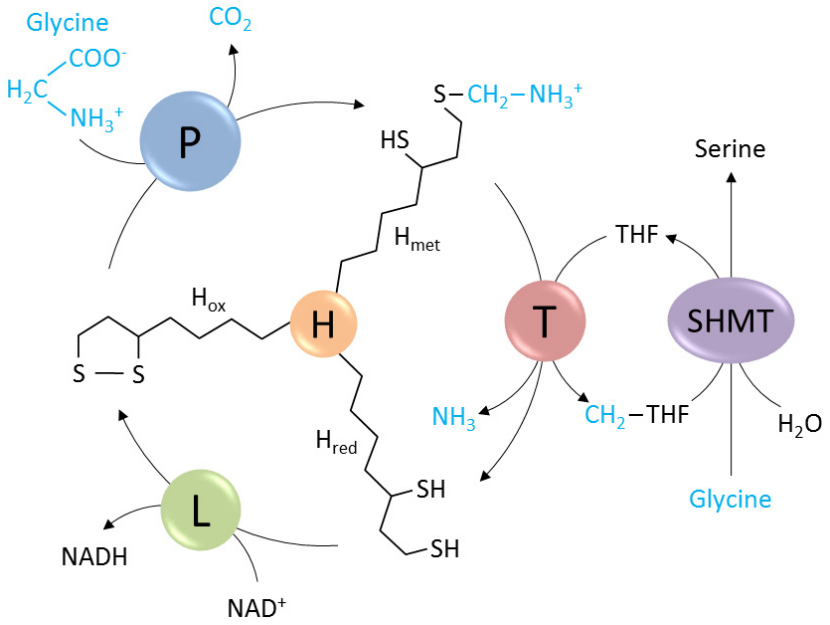


Figure 8. Reaction of the glycine decarboxylase complex.

The glycine decarboxylase complex (GDC) consists of the P-, H-, L- and T-protein. The H-protein undergoes a cycle of reductive methylamination, methylamine transfer and electron transfer, which is catalyzed by the P-, T- and L-protein respectively. The cleavage of glycine results in CO_2, NH_3 and a tetrahydrofolate-bound C_1 residue (CH_2–THF). A second molecule of glycine together with CH_2–THF is converted by serine hydroxymethyltransferase (SHMT) to serine. H_{met}, H_{red} and H_{ox}: methylaminated, reduced and oxidized forms of the H-protein. Adapted from Douce et al. (2001).

4.2 Function of the glycine decarboxylase complex in plants

In all organisms, GDC plays an important role in interconnecting the metabolism of one-, two- and three-carbon compounds (Oliver, 1994). In plants, GDC is involved in photorespiration occurring in all photosynthetically active cells and in the one-carbon (C_1) metabolism in all biosynthetic tissues (Engel et al., 2007; Douce et al., 2001; Hanson and Roje 2001). Photorespiration is assumed to be one of the most wasteful processes on the planet because energy and previously fixed CO_2 is lost, which decreases the rate of photosynthesis (Foyer et al., 2009). Transgenic *A. thaliana* plants which perform a shortened photorespiratory cycle by integrating the *Escherichia coli* glycolate catabolic pathway produce higher shoot and root biomass (Kebeish et al., 2007). Nonetheless, the photorespiratory metabolism represents a mechanism to regenerate 3PGA from the toxic compound 2PG which accumulates as the result of Rubisco's oxygenase activity, and to impede poisoning by 2PG itself, glyoxylate or glycine (Bauwe et al., 2010). Apart from that, a further positive function of photorespiration seems to be the prevention of photoinactivation of the photosynthetic apparatus under sunlight when CO_2 concentrations are decreased as a result of stomatal closure (Heber et al., 1996; Douce and Neuburger, 1999).

An Arabidopsis T-DNA insertion double mutant lacking functional GDC, due to the knockout of both genes encoding the P-subunit, was lethal even under conditions suppressing photorespiration. This led to the conclusion that the GDC reaction cannot be bypassed in higher plants but is essential for other metabolic processes such as the glycine serine cycling of the C_1 metabolism (Engel et al., 2007). The C_1 metabolism presumably occurs in all plant tissues. GDC is present even in non-photosynthetic cells, although only little amounts were detectable (Bourguignon et al., 1993; Mouillon et al., 1999). C_1-metabolic reactions are important for supplying C_1 molecules that are used for the synthesis of proteins, pantothenates, nucleic acids and many methylated compounds such as lignin, betaines and alkaloids (Cossins and Chen, 1997; Hanson and Roje, 2001). Probably, the C_1 demand varies in the different plant tissues, which might be achieved by regulating the expression of genes involved in the C_1 metabolism (Hanson and Roje, 2001).

Consistent with the involvement of GDC in photorespiration, the expression of all four GDC subunits is light-regulated. Although the transcripts encoding the H-, P- and T-protein were present at low levels in etiolated pea tissues, their abundance increased when the plants were illuminated. In contrast, the mRNA of the L-protein is enriched even in the dark and its concentration only changes little after the exposure to light. This can be explained by the fact that the L-protein is a subunit of different multienzyme complexes. The light-dependent

accumulation of transcripts encoding the H- and P-protein is regulated transcriptionally (Kim et al., 1991; Srinivasan et al., 1992, Oliver, 1994; Srinivasan and Oliver, 1995; Douce et al., 2001).

4.3 The GLDPA gene encodes the P-protein of GDC in the C_4 plant Flaveria trinervia

The genome of the C_4 plant *Flaveria trinervia* contains two genes encoding the P-subunit of GDC, *GLDPA* and *GLDPB* (Cossu and Bauwe, 1998). Based on phylogenetic analyses the *GLDPB* gene was renamed to *GLDPE* (personal communication with Stefanie Schulze). *GLDPE* seems to represent a pseudogene. It encodes no functional P-protein due the absence of corresponding mRNA most likely as the result of an insertion in the first intron that shows similarity with retrotransposons. However, the *GLDPA* gene is transcribed, leading to the expression of a functional P-protein that consists of 1034 amino acids. After targeting to mitochondria and cleavage of the 63 amino acid long presequence the mature *GLDPA* protein comprises 971 amino acids (Cossu and Bauwe, 1998). The -1571 to -1 5' upstream region (with regard to the translational start at +1) of the *GLDPA* gene of *F. trinervia* is referred to as the *GLDPA* promoter. This promoter was shown to activate expression of the β-glucuronidase (*GUS*) reporter gene specifically in the bundle-sheath cells and the vascular bundles in the closely related C_4 plant *F. bidentis* (Burscheidt, 1998).

In contrast to C_3 plants where GDC is present in all photosynthetically active cells, in C_4 species GDC is restricted to the bundle-sheath cells (Rawsthorne et al., 1988; Morgan et al., 1993; Yoshimura et al., 2004). The transcriptional regulation through the *GLDPA* promoter seems to be essential for the bundle-sheath-specific expression of the *GLDPA* gene. Thus, the *GLDPA* promoter can be used to identify *cis*-regulatory elements which are required for gene expression specifically in the bundle-sheath.

II. Scientific aims

The mitochondrial glycine decarboxylase complex (GDC) is involved in photorespiration and the C_1 metabolism. Photorespiratory processes occur in all photosynthetically active tissues in C_3 species, but are restricted to the bundle-sheath in C_4 plants. The *GLDPA* gene of *Flaveria trinervia* (C_4) encodes the P-subunit of GDC. The present work shall give an insight into the molecular mechanisms underlying the bundle-sheath- and vasculature-specific activity of the *GLDPA* promoter of *F. trinervia* in leaves of the closely related C_4 plant *Flaveria bidentis* and the phylogenetically distant C_3 plant *Arabidopsis thaliana*. This might help to reveal adaptive changes in the regulation of gene expression during C_4 evolution.

1) The *GLDPA* promoter directs expression of the β-glucuronidase (*GUS*) reporter gene in bundle-sheath cells and the vasculature in leaves of both transgenic *F. bidentis* (C_4) and *A. thaliana* (C_3). To exclude diffusion effects caused during the histochemical GUS staining procedure, a variant of the green fluorescent protein (GFP) targeted to the endoplasmic reticulum, and a yellow fluorescent protein (YFP) fused to histone 2B (H2B) which is retained in the nucleus were expressed under the control of the *GLDPA* promoter in *A. thaliana*. A detailed functional promoter analysis was performed in *A. thaliana* and *F. bidentis* by fusing several *GLDPA* promoter deletion and recombination constructs to the *GUS* reporter gene. The aim was to identify potential *cis*-regulatory determinants for bundle-sheath-specific gene expression and to verify whether these findings can be applied to the C_4 and C_3 context (Manuscript 1: Engelmann et al., 2008).

2) Based on the previous findings (Manuscript 1: Engelmann et al., 2008), further *GLDPA* promoter deletion and recombination constructs fused to the *GUS* reporter gene were analyzed in both *F. bidentis* (C_4) and *A. thaliana* (C_3) to elucidate the regulation of this promoter more precisely. The transcription start sites of the *GLDPA* gene were determined by analyzing mRNA 5' ends using rapid amplification of 5' complementary DNA ends (5' RACE). By means of 454 pyrosequencing data, the abundance of *GLDPA* transcripts as well as their splicing pattern were detected. Two different RNA variants transcribed from the *GLDPA* promoter give rise to proteins with either a full-length or a truncated presequence for mitochondrial targeting. Both presequence versions were fused with the *GFP* reporter gene and transiently expressed in leaf protoplasts of *Nicotiana benthamiana* under the control of the constitutive *cauliflower mosaic virus* (*CaMV*) *35S* promoter. The influence of both presequence variants on the subcellular localization of GFP was analyzed by confocal microscopy (Manuscript 2: Wiludda et al., 2011).

III. Theses

The *GLDPA* promoter of the C_4 plant *Flaveria trinervia* is composed of two sub-promoters that regulate gene expression in tandem. Together these promoters ensure strong expression in the bundle-sheath cells and the vasculature, but apparently additional weak expression in the mesophyll to precisely adjust the accumulation of GLDPA protein to the metabolic needs of the different leaf tissues.

1) The *GLDPA* promoter of *F. trinervia* (C_4) is specifically active in bundle-sheath cells and the vasculature of both the closely related C_4 species *Flaveria bidentis* and the phylogenetically distant C_3 species *Arabidopsis thaliana*. Two components of the *GLDPA* promoter are essential for its regulation: the distal region enhances promoter activity, while an intermediate segment confers bundle-sheath and vasculature specificity by repressing mesophyll expression. Both components act in the same or at least a very similar way in the C_4 and the C_3 plant. This indicates that bundle-sheath-specific *cis*-regulatory determinants within the *GLDPA* promoter are recognized already in the C_3 background in a C_4-characteristic manner (Manuscript 1: Engelmann et al., 2008).

2) The expression of the *GLDPA* gene of *F. trinervia* (C_4) is regulated by an intricate interplay of transcriptional and post-transcriptional mechanisms. Two sub-promoters within the full-length *GLDPA* promoter act in tandem. The proximal promoter is active in bundle-sheath cells and the vasculature, while the distal promoter exhibits additional activity in the mesophyll in both *F. bidentis* (C_4) and *A. thaliana* (C_3). To confer bundle-sheath and vasculature specificity, the proximal promoter is enhanced transcriptionally, and the output of the distal promoter is repressed apparently by transcript destabilization due to inefficient splicing of an intron within the 5' untranslated region which might elicit RNA decay. In *F. bidentis*, the proximal promoter suffices to suppress the output of the distal promoter, whereas in *A. thaliana* the intermediate segment which represses mesophyll expression is additionally needed for stable suppression. Despite their low abundance, completely spliced RNAs transcribed from the distal promoter occur, and appear to be stable. Although these transcripts give rise to a protein variant with a truncated mitochondrial targeting sequence, this has no effect on protein localization to mitochondria. In this way, C_4 plants are able to highly express GLDPA in the bundle-sheath needed for photorespiration, while only low amounts of GLDPA protein seem to be required in the mesophyll for the C_1 metabolism (Manuscript 2: Wiludda et al., 2011).

IV.A Summary

The successful process of C_4 photosynthesis is based on the strict compartmentalization of the enzymes involved in this pathway into mesophyll and bundle-sheath cells of the leaf. Differential gene expression ensures the restriction to one of these distinct tissues. The *GLDPA* gene of *Flaveria trinervia* (C_4) encodes the P-subunit of the mitochondrial glycine decarboxylase complex (GDC). GDC is involved in photorespiration which occurs in all photosynthetically active tissues in C_3 species, but is restricted to the bundle-sheath in C_4 plants. In all biosynthetically active cells, GDC is also essential for the C_1 metabolism.

In this study, promoter-reporter gene fusion analyses showed that the 1571 base pairs full-length promoter of the *GLDPA* gene is specifically active in bundle-sheath cells and the vasculature in leaves of transgenic plants of both *Flaveria bidentis* (C_4) and *Arabidopsis thaliana* (C_3). This indicates that the genetic control mechanisms of C_3 plants do not differ substantially from those of C_4 species. Detailed studies of promoter deletion and recombination constructs fused to the β-glucuronidase (*GUS*) reporter gene in transgenic *F. bidentis* and *A. thaliana* plants, and analyses of RNA 5' ends by rapid amplification of 5' complementary DNA ends (5' RACE) revealed that the *GLDPA* promoter is composed of two sub-promoters which coordinate gene expression in tandem. In both species, the proximal promoter already suffices to cause the same expression pattern as the full-length promoter, while the distal promoter alone exhibits uniform activity in all inner leaf tissues including the mesophyll. The proximal promoter can silence the distal promoter in *F. bidentis*, but requires an additional *GLDPA* promoter segment in Arabidopsis for stable suppression.

In the context of the full-length *GLDPA* promoter, the output of the distal promoter appears to be repressed post-transcriptionally by transcript destabilization. Based on 454 pyrosequencing data, splicing of an intron in the 5' untranslated region of transcripts derived from the distal promoter only inefficiently occurs, but seems to be essential for the accumulation of stable RNAs. These completely spliced transcripts are rare, and encode a protein with a truncated mitochondrial targeting sequence, which was shown to have no effect on localization to mitochondria. Further promoter segments were identified that enhance the transcriptional activity of the proximal promoter to ensure strong bundle-sheath expression.

In this way, C_4 plants do not only succeed in expressing GDC in different tissues, but are also able to adjust its amount to the needs of the respective metabolic pathway. GDC is apparently needed in large quantities for photorespiration in bundle-sheath cells, while its amount is reduced to an adequate level in mesophyll cells to serve the C_1 metabolism.

IV.B Zusammenfassung

Der erfolgreiche Ablauf der C_4-Photosynthese beruht auf der strikten Kompartimentierung der an dieser Reaktion beteiligten Enzyme in den Mesophyll- und Bündelscheidenzellen des Blattes. Die differentielle Genexpression gewährleistet hierbei die Begrenzung auf eines dieser speziellen Gewebe. Das *GLDPA*-Gen von *Flaveria trinervia* (C_4) kodiert die P-Untereinheit des mitochondrialen Glycin-Decarboxylase-Komplexes. Dieser ist an der Photorespiration beteiligt, die in C_3-Spezies in allen photosynthetisch aktiven Geweben abläuft, jedoch auf die Bündelscheide in C_4-Pflanzen beschränkt ist. Außerdem ist der Glycin-Decarboxylase-Komplex für den C_1-Metabolismus in allen biosynthetischen Zellen essentiell.

Im Rahmen dieser Arbeit konnte mittels Promoter-Reportergen-Fusionen gezeigt werden, dass der 1571 Basenpaare große Volllängenpromotor des *GLDPA*-Gens spezifisch in den Bündelscheidenzellen und dem Leitgewebe in Blättern transgener Pflanzen von *Flaveria bidentis* (C_4) und *Arabidopsis thaliana* (C_3) aktiv ist. Dies deutet darauf hin, dass sich die genetischen Kontrollmechanismen der C_3-Pflanzen nicht wesentlich von denen der C_4-Pflanzen unterscheiden. Detaillierte Studien an transgenen Pflanzen von *F. bidentis* und *A. thaliana*, die verschiedene Promotordeletions- und Promotorrekombinationskonstrukte enthielten, welche mit dem β-Glucuronidase-Reportergen (*GUS*) fusioniert waren, sowie die Bestimmung von RNA-5'-Enden durch die schnelle Amplifizierung von 5'-komplementären DNA-Enden (5' RACE) haben aufgedeckt, dass der *GLDPA*-Promotor ein Tandempromotor ist, der sich aus zwei Teilpromotoren zusammensetzt, die gemeinsam die Genexpression koordinieren. Der proximale Promoter reicht in beiden Spezies aus, um das gleiche Expressionsmuster wie der Volllängenpromotor zu bewirken, wohingegen der distale Promotor alleine eine gleichmäßige Aktivität in allen inneren Blattgeweben einschließlich des Mesophylls aufweist. In *F. bidentis* kann der proximale Promotor den distalen Promotor stilllegen, benötigt in Arabidopsis jedoch für die stabile Unterdrückung ein zusätzliches *GLDPA*-Promotorsegment.

Hierbei scheint die Ausgabe des distalen Promotors im Kontext des *GLDPA*-Volllängenpromotors posttranskriptionell durch destabilisierte Transkripte verringert zu werden. Mit Hilfe von 454-Sequenzierungsdaten konnte gezeigt werden, dass das Spleißen eines Introns innerhalb des 5' untranslatierten Bereichs der Transkripte, die vom distalen Promotor stammen, nur ineffizient abläuft, jedoch essentiell für die Anreicherung stabiler RNAs zu sein scheint. Diese vollständig gespleißten Transkripte sind selten und kodieren ein Protein mit einer verkürzten mitochondrialen Zielsequenz, was jedoch keine Auswirkung auf

die Lokalisierung in den Mitochondrien hat. Es konnten weitere Promotorsegmente identifiziert werden, die die transkriptionelle Aktivität des proximalen Promotors verstärken, um eine starke bündelscheidenspezifische Expression zu gewährleisten.

Auf diese Weise gelingt es C_4-Pflanzen nicht nur, den Glycin-Decarboxylase-Komplex in verschiedenen Geweben zu exprimieren, sondern sie schaffen es auch, dessen Menge präzise den Bedürfnissen des jeweiligen Stoffwechselweges anzupassen. Der Glycin-Decarboxylase-Komplex wird scheinbar in großen Mengen in den Bündelscheidenzellen für die Photorespiration, aber nur in geringen Mengen im Mesophyll für den C_1-Metabolismus benötigt.

V. Literature

Akyildiz M, Gowik U, Engelmann S, Koczor M, Streubel M, Westhoff P (2007) Evolution and function of a *cis*-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. Plant Cell **19:** 3391–3402

Andersson I (2008) Catalysis and regulation in Rubisco. J Exp Bot **59:** 1555–1568

Andrews TJ, Lorimer GH (1987) Rubisco: structure, mechanisms, and prospects for improvement. **In:** Hatch MD, Boardman NK (eds) The biochemistry of plants, Vol. 10. Academic Press, New York, NY, USA, pp 131–218

Archibold OW (1995) Ecology of world vegetation. London, UK: Chapman & Hall

Arnosti DN, Kulkarni MM (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? J Cell Biochem **94:** 890–898

Aubry S, Brown NJ, Hibberd JM (2011) The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. J Exp Bot **62:** 3049–3059

Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell **27:** 299–308

Basehoar AD, Zanton SJ, Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. Cell **116:** 699–709

Bassett CL, Nickerson ML, Farrell RE Jr, Harrison M (2004) Multiple transcripts of a gene for a leucine-rich repeat receptor kinase from morning glory (*Ipomoea nil*) originate from different TATA boxes in a tissue-specific manner. Mol Genet Genomics **271:** 752–760

Bauwe H, Kolukisaoglu U (2003) Genetic manipulation of glycine decarboxylation. J Exp Bot **54:** 1523–1535

Bauwe H, Hagemann M, Fernie AR (2010) Photorespiration: players, partners and origin. Trends Plant Sci **15:** 330–336

Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. BMC Genomics **11:** 166

Bläsing OE, Westhoff P, Svensson P (2000) Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. J Biol Chem **275:** 27917–27923

Bourguignon J, Vauclare P, Merand V, Forest E, Neuburger M, Douce R (1993) Glycine decarboxylase complex from higher plants. Molecular cloning, tissue distribution and mass spectrometry analyses of the T-protein. Eur J Biochem **217:** 377–386

Bowes G, Ogren WL, Hageman RH (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. Biochem Biophys Res Commun **45:** 716–722

Bowes G, Rao SK, Estavillo GM, Reiskind JB (2002) C_4 mechanisms in aquatic angiosperms: comparisons with terrestrial C_4 systems. Funct Plant Biol **29:** 379–392

Brown RH (1978) A difference in N use efficiency in C_3 and C_4 plants and its implications in adaptation and evolution. Crop Sci **18:** 93–98

Brown RH (1999) Agronomic implications of C_4 photosynthesis. **In:** Sage RF, Monson RK (eds) C_4 plant biology. Academic Press, San Diego, pp 473–507

Brown NJ, Parsley K, Hibberd JM (2005) The future of C_4 research – maize, *Flaveria* or *Cleome*? Trends Plant Sci **10:** 215–221

Brown NJ, Palmer BG, Stanley S, Hajaji H, Janacek SH, Astley HM, Parsley K, Kajala K, Quick WP, Trenkamp S, Fernie AR, Maurino VG, Hibberd JM (2010) C_4 acid decarboxylases required for C_4 photosynthesis are active in the mid-vein of the C_3 species *Arabidopsis thaliana*, and are important in sugar and amino acid metabolism. Plant J **61:** 122–133

Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM (2011) Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. Science **331:** 1436–1439

Bruhl JJ, Wilson KL (2007) Towards a comprehensive survey of C_3 and C_4 photosynthetic pathways in Cyperaceae. Aliso **23:** 99–148

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell **144:** 327–339

Burscheidt J (1998) *Cis*-regulatorische Determinanten für mesophyll- und bündelscheidenspezifische Genexpression in C_4-Spezies der Gattung *Flaveria* – Die Promotoren der Phospho*enol*pyruvat-Carboxylase- und der Glycin-Decarboxylasegene. Ph.D. thesis. Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

von Caemmerer S, Furbank RT (2003) The C_4 pathway: an efficient CO_2 pump. Photosynth Res **77:** 191–207

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J et al. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet **38:** 626–635

Casati P, Lara MV, Andreo CS (2000) Induction of a C_4-like mechanism of CO_2 fixation in *Egeria densa*, a submersed aquatic species. Plant Physiol **123:** 1611–1621

Cegelski L, Schaefer J (2006) NMR determination of photorespiration in intact leaves using *in vivo* $^{13}CO_2$ labeling. J Magn Reson **178:** 1–10

Christin PA, Osborne CP, Sage RF, Arakaki M, Edwards EJ (2011) C_4 eudicots are not younger than C_4 monocots. J Exp Bot **62:** 3171–3181

Cockburn W (1983) Stomatal mechanism as the basis of the evolution of CAM and C_4 photosynthesis. Plant Cell Environ **6:** 275–279

Cossins EA, Chen L (1997) Folates and one-carbon metabolism in plants and fungi. Phytochemistry **45:** 437–52

Cossu R, Bauwe H (1998) Two genes of the *GDCSP* gene family from the C_4 plant *Flaveria trinervia*: *GDCSPA* encoding P-protein and *GDCSPB*, a pseudo-gene. Plant Physiol **116:** 445–446

Cousins AB, Baroli I, Badger MR, Ivakov A, Lea PJ, Leegood RC, von Caemmerer S (2007) The role of phosphoenolpyruvate carboxylase during C_4 photosynthetic isotope exchange and stomatal conductance. Plant Physiol **145:** 1006–1017

Deng W, Roberts SG (2005) A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev **19:** 2418–2423

Deng W, Roberts SG (2007) TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma **116:** 417–429

Dengler NG, Nelson T (1999) Leaf structure and development in C_4 plants. **In:** Sage RF, Monson RK (eds) C_4 plant biology. Academic Press, San Diego, pp 133–172

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell **10:** 1075–1082

Douce R, Neuburger M (1999) Biochemical dissection of photorespiration. Curr Opin Plant Bio **2:** 214–222

Douce R, Bourguignon J, Neuburger M, Rebeille F (2001) The glycine decarboxylase system: a fascinating complex. Trends Plant Sci **6:** 167–176

Drincovich MF, Casati P, Andreo CS, Chessin SJ, Franceschi VR, Edwards GE, Ku MSB (1998) Evolution of C_4 photosynthesis in *Flaveria* species. Isoforms of NADP-malic enzyme. Plant Physiol **117:** 733–744

Drincovich MF, Casati C, Andreo CS (2001) NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. FEBS Lett **490:** 1–6

Edwards GE, Andreo CS (1992) NADP-Malic enzyme from plants. Phytochemistry **31:** 1845–1857

Edwards GE, Franceschi VR, Ku MSB, Voznesenskaya EV, Pyankov VI, Andreo CS (2001a) Compartmentation of photosynthesis in cells and tissues of C_4 plants. J Exp Bot **52:** 577–590

Edwards GE, Furbank RT, Hatch MD, Osmond CB (2001b) What does it take to be C_4? Lessons from evolution of C_4 photosynthesis. Plant Physiol **125:** 46–49

Edwards GE, Franceschi VR, Voznesenskaya EV (2004) Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. Annu Rev Plant Biol **55:** 173–196

Edwards EJ, Osborne CP, Strömberg CAE, Smith SA, C_4 Grasses Consortium (2010) The origins of C_4 grasslands: integrating evolutionary and ecosystem science. Science **328:** 587–591

Ehleringer J, Björkman O (1977) Quantum yields for CO_2 uptake in C_3 and C_4 plants-dependence on temperature, CO_2, and O_2 concentration. Plant Physiol **59:** 86–90

Ehleringer J, Pearcy RW (1983) Variation in quantum yield for CO_2 uptake among C_3 and C_4 plants. Plant Physiol **73:** 555–559

Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW (1991) Climate change and the evolution of C_4 photosynthesis. Trends Ecol Evol **6:** 95–99

Ehleringer JR, Monson RK (1993) Evolutionary and ecological aspects of photosynthetic pathway variation. Annu Rev Ecol Syst **24:** 411–439

Ehleringer JR, Cerling TE, Helliker BR (1997) C_4 photosynthesis, atmospheric CO_2, and climate. Oecologia **112:** 285–299

Ellis RJ (1979) Most abundant protein in the world. Trends Biochem Sci **4:** 241–244

Engel N, Van Den Daele K, Kolukisaoglu Ü, Morgenthal K, Weckwerth W, Pärnik T, Keerberg O, Bauwe H (2007) Deletion of glycine decarboxylase in Arabidopsis is lethal under nonphotorespiratory conditions. Plant Physiol **144:** 1328–1335

Engelmann S, Zogel C, Koczor M, Schlue U, Streubel M, Westhoff P (2008) Evolution of the C_4 phospho*enol*pyruvate carboxylase promoter of the C_4 species *Flaveria trinervia*: the role of the proximal promoter region. BMC Plant Biol **8:** 4

Foyer CH, Bloom AJ, Queval G, Noctor G (2009) Photorespiratory metabolism: Genes, mutants, energetics, and redox signaling. Annu. Rev Plant Biol **60:** 455–484

Furbank RT (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? J Exp Bot **62:** 3103–3108

Gerrard-Wheeler MC, Tronconi MA, Drincovich MF, Andreo CS, Flugge UI, Maurino VG (2005) A comprehensive analysis of the NADP-malic enzyme gene family of Arabidopsis. Plant Physiol **139:** 39–51

Gerrard-Wheeler MC, Arias CL, Tronconi MA, Maurino VG, Andreo CS, Drincovitch MF (2008) *Arabidopsis thaliana* NADP-malic enzyme isoforms: high degree of identity but clearly distinct properties. Plant Mol Biol **67:** 231–242

Gershenzon NI, Trifonov EN, Ioshikhes IP (2006) The features of *Drosophila* core promoters revealed by statistical analysis. BMC Genomics **7:** 161

Gillion J, Yakir D (2001) Influence of carbonic anhydrase activity in terrestrial vegetation on the ^{18}O content of atmospheric CO_2. Science **291:** 2584–2587

Goldberg, ML (1979) Ph.D. thesis. Stanford University, Stanford, CA, U.S.A.

Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, Westhoff P (2004) *Cis*-regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. Plant Cell **16:** 1077–1090

Gowik U, Bräutigam A, Weber KL, Weber APM, Westhoff P (2011) Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? Plant Cell **23:** 2087–2105

Gowik U, Westhoff P (2011) The path from C_3 to C_4 photosynthesis. Plant Physiol **155:** 56–63

Haberlandt G (1881) Vergleichende Anatomie des assimilatorischen Gewebesytems bei Pflanzen. Jahrbuch der Wissenschaftlichen Botanik **13:** 74–188

Hanson AD, Roje S (2001) One-carbon metabolism in higher plants. Annu Rev Plant Physiol Plant Mol Biol **52:** 119–137

Hatch MD (1987) C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim Biophys Acta **895:** 81–106

Hattersley PW, Browning AJ (1981) Occurrence of the suberized lamella in leaves of grasses of different photosynthetic type. I. In parenchymatous bundle sheath and PCR ('Kranz') sheaths. Protoplasma **109:** 371–401

Hattersley PW (1984) Characterization of C_4 type leaf anatomy in grasses (Poaceae). Mesophyll:Bundle sheath area ratios. Ann Bot **53:** 163–179

Heber U, Bligny R, Streb P, Douce R (1996) Photorespiration is essential for the protection of the photosynthetic apparatus of C_3 plants against photoinactivation under sunlight. Bot Acta **109:** 307–315

Hibberd JM, Covshoff S (2010) The regulation of gene expression required for C_4 photosynthesis. Annu Rev Plant Biol **61:** 181–207

Hylton CM, Rawsthorne S, Smith AM, Jones DA (1988) Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. Planta **175:** 452–459

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res **15:** 6643–6651

Juven-Gershon T, Hsu J-Y, Theisen JW, Kadonaga JT (2008) The RNA polymerase II core promoter – the gateway to transcription. Curr Opin Cell Biol **20:** 253–259

Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol **339:** 225–229

Kanai R, Edwards GE (1999) The biochemistry of C_4 photosynthesis. **In:** Sage RF, Monson RK (eds), C_4 plant biology. Academic Press, San Diego, pp 49–87

Kapralov MV, Kubien DS, Andersson I, Filatov DA (2011) Changes in Rubisco kinetics during the evolution of C_4 photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. Mol Biol Evol **28:** 1491–1503

Kebeish R, Niessen M, Thiruveedhi K, Bari R, Hirsch HJ, Rosenkranz R, Stabler N, Schonfeld B, Kreuzaler F, Peterhansel C (2007) Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. Nat Biotechnol **25:** 593–599

Kern R, Bauwe H, Hagemann M (2011) Evolution of enzymes involved in the photorespiratory 2-phosphoglycolate cycle from cyanobacteria via algae toward plants. Photosynth Res **doi:** 10.1007/s11120-010-9615-z

Kikuchi G, Motokawa Y, Yoshida T, Hiraga K (2008) Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. Proc Jpn Acad Ser B Phys Biol Sci **84:** 246–63

Kim Y, Shah K, Oliver DJ (1991) Cloning and light-dependent expression of the gene coding for the P-protein of the glycine decarboxylase complex from peas. Physiol Plant **81:** 501–506

Kinsman EA, Pyke KA (1998) Bundle sheath cells and cell-specific plastid development in Arabidopsis leaves. Development **125:** 1815–1822

Koprivova A, Melzer M, Von Ballmoos P, Mandel T, Brunold C, Kopriva S (2001) Assimilatory sulfate reduction in C_3, C_3-C_4, and C_4 species of *Flaveria*. Plant Physiol **127:** 543–550

Ku MSB, Schmitt MR, Edwards GE (1979) Quantitative determination of RuBP carboxylase-oxygenase protein in leaves of several C_3 and C_4 plants. J Exp Bot **114:** 89–98

Kutschera U, Niklas KJ (2007) Photosynthesis research on yellowtops: Macroevolution in progress. Theory Biosci **125:** 81–92

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev **12:** 34–44

Lee SW, Heinz R, Robb J, Nazar RN (1994) Differential utilization of alternate initiation sites in a plant defense gene responding to environmental stimuli. Eur J Biochem **226:** 109–114

Leegood RC, Lea PJ, Adcock MD, Häusler RE (1995) The regulation and control of photorespiration. J Exp Bot **46:** 1397–1414

Leegood RC, Walker RP (1999) Regulation of the C_4 pathway. **In:** Sage RF, Monson RK (eds) C_4 plant biology. Academic Press, San Diego, pp 89–132

Leegood RC, Walker RP (2003) Regulation and roles of phosphoenolpyruvate carboxykinase in plants. Arch Biochem Biophys **414:** 204–210

Leegood R (2008) Roles of the bundle sheath cells in leaves of C_3 plants. J Exp Bot **59:** 1663–1673

Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature **424:** 147–151

Li XR, Wang L, Ruan YL (2010) Developmental and molecular physiological evidence for the role of phosphoenolpyruvate carboxylase in rapid cotton fibre elongation. J Exp Bot **61:** 287–295

Lloyd J, Farquhar GD (1994) $\delta^{13}C$ discrimination during CO_2 assimilation by the terrestrial biosphere. Oecologia **99:** 201–215

Long SP (1999) Environmental responses. **In:** RF Sage, RK Monson (eds) C_4 plant biology. Academic Press, San Diego, pp 215–249

Luo M, Orsi R, Patrucco E, Pancaldi S, Cella R (1997) Multiple transcription start sites of the carrot dihydrofolate reductasethymidylate synthase gene, and sub-cellular localization of the bifunctional protein. Plant Mol Biol **33:** 709–722

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155

Marshall JS, Stubbs JD, Taylor WC (1996) Two genes encode highly similar chloroplastic NADP-malic enzymes in *Flaveria*. Plant Physiol **111:** 1251–1261

Masumoto C, Miyazawa SI, Ohkawa H, Fukuda T, Taniguchi Y, Murayama S, Kusano M, Saito K, Fukayama H, Miyao M (2010) Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. Proc Natl Acad Sci U S A **107:** 5226–5231

Matsuoka M, Ozeki Y, Yamamoto N, Hirano H, Kano-Murakami Y, Tanaka Y (1988) Primary structure of maize pyruvate orthophosphate dikinase as deduced from cDNA sequence. J Biol Chem **263:** 11080–83

Matsuoka M, Kyozuka J, Shimamoto K, Kano-Murakami Y (1994) The promoters of two carboxylases in a C_4 plant (maize) direct cell-specific, light-regulated expression in a C_3 plant (rice). Plant J **6:** 311–319

McKown AD, Moncalvo JM, Dengler NG (2005) Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. Am J Bot **92:** 1911–1928

Meierhoff K, Westhoff P (1993) Differential biogenesis of photosystem II in mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C_4 plants: the nonstoichiometric abundance of the subunits of photosystem II in the bundle-sheath chloroplasts and the translational activity of the plastome-encoded genes. Planta **191:** 23-33

Miyao M, Fukayama H (2003) Metabolic consequences of overproduction of phosphoenolpyruvate carboxylase in C_3 plants. Arch Biochem Biophys **414:** 197–203

Molina C, Grotewold E (2005) Genome wide analysis of Arabidopsis core promoters. BMC Genomics **6:** 25

Monson RK, Moore BD (1989) On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. Plant Cell Environ **12:** 689–699

Monson RK (1999) The origins of C_4 genes and evolutionary pattern in the C_4 metabolic phenotype. **In:** Sage RF, Monson RK (eds) C_4 plant biology. Academic Press, San Diego, pp 377–410

Monson RK, Rawsthorne S (2000) CO_2 assimilation in C_3-C_4 intermediate plants. **In:** Leegood RC, Sharkey TD, von Caemmerer SC (eds) Photosynthesis: Physiology and Metabolism. Kluwer Academic, Dordrecht, the Netherlands, pp 533–550

Monson RK (2003) Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. Int J Plant Sci **164:** S43–S54

Morgan CL, Turner SR, Rawsthorne S (1993) Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and serine hydroxymethyltransferase in leaves of C_3-C_4 intermediate species from different genera. Planta **190:** 468–473

Mouillon JM, Aubert S, Bourguignon J, Gout E, Douce R, Rébeillé F (1999) Glycine and serine catabolism in non-photosynthetic higher plant cells: their role in C_1 metabolism. Plant J **20:** 197–205

Muhaidat R, Sage RF, Dengler NG (2007) Diversity of kranz anatomy and biochemistry in C_4 eudicots. Am J Bot **94:** 362–381

Nakamura M, Tsunoda T, Obokata J (2002) Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. Plant J **29:** 1–10

Ogren, WL (1984) Photorespiration: pathways, regulation, and modification. Ann Rev Plant Physiol **35:** 415–442

Ohnishi J, Kanai R (1983) Differentiation of photorespiratory activity between mesophyll and bundle sheath cells of C_4 plants. I. Glycine oxidation by mitochondria. Plant Cell Physiol **24:** 1411–1420

Oliver DJ, Neuburger M, Bourguignon J, Douce R (1990a) Glycine metabolism by plant mitochondria. Physiol Plant **80:** 487–491

Oliver DJ, Neuburger M, Bourguignon J, Douce R (1990b) Interaction between the component enzymes of the glycine decarboxylase multienzyme complex. Plant Physiol **94:** 833–839

Oliver DJ (1994) The glycine decarboxylase complex from plant mitochondria. Annu Rev Plant Physiol Plant Mol Biol **45:** 323–327

Parsley K, Hibberd JM (2006) The Arabidopsis *PPDK* gene is transcribed from two promoters to produce differentially expressed transcripts responsible for cytosolic and plastidic proteins. Plant Mol Biol **62:** 339–49

Patel M, Siegel AJ, Berry JO (2006) Untranslated regions of *FbRbcS1* mRNA mediate bundle sheath cell-specific gene expression in leaves of a C_4 plant. J Biol Chem **281:** 25485–25491

Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev **13:** 3217–3230

Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, Sandelin A (2006) Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome Biol **7:** R78

Portis AR, Parry MAJ (2007) Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. Photosynth Res **94:** 121–143

Powell AM (1978) Systematics of *Flaveria* (Flaveriinae-Asteraceae). Ann Mo Bot Gard **65:** 590–636

Priest HD, Filichkin SA, Mockler TC (2009) *Cis*-regulatory elements in plant cell signaling. Curr Opin Plant Biol **12:** 643–649

Purnell BA, Emanuel PA, Gilmour DS (1994) TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. Genes Dev **8:** 830–842

Raab JR, Kamakaka RT (2010) Insulators and promoters: closer than we think. Nat Rev Genet **11:** 439–446

Raines CA, Paul MJ (2006) Products of leaf primary carbon metabolism modulate the developmental programme determining plant morphology. J Exp Bot **57:** 1857–1862

Raines CA (2011) Increasing photosynthetic carbon assimilation in C_3 plants to improve crop yield: current and future strategies. Plant Physiol **155:** 36–42

Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW (1988) Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C_3 and C_3-C_4 intermediate species of *Moricandia*. Planta **173:** 298–308

Rawsthorne S (1992) C_3-C_4 intermediate photosynthesis: linking physiology to gene expression. Plant J **2:** 267–274

Reiskind JB, Madsen TV, van Ginkel LC, Bowes G (1997) Evidence that inducible C_4-type photosynthesis is a chloroplastic CO_2-concentrating mechanism in *Hydrilla*, a submersed monocot. Plant Cell Environ **20:** 211–220

Roalson EH (2011) Origins and transitions in photosynthetic pathway types in monocots: a review and reanalysis. **In:** Raghavendra AS, Sage RF (eds) C_4 photosynthesis and related CO_2 concentrating mechanisms. Springer, Dordrecht, The Netherlands, pp 319–338

Rombauts S, Florquin K, Lescot M, Marchal K, Rouze P, van de Peer Y (2003) Computational approaches to identify promoters and *cis*-regulatory elements in plant genomes. Plant Physiol **132:** 1162–1176

Rosche E, Westhoff P (1995) Genomic structure and expression of the pyruvate, orthophosphate dikinase gene of the dicotyledonous C_4 plant *Flaveria trinervia* (Asteraceae). Plant Mol Biol **29:** 663–678

Roy H, Andrews TJ (2000) Rubisco: assembly and mechanism. **In:** Leegood RC, Sharkey TD, von Caemmerer S (eds) Photosynthesis: physiology and metabolism. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp 53–83

Rylott EL, Gilday AD, Graham IA (2003) The gluconeogenic enzyme phosphoenolpyruvate carboxykinase in Arabidopsis is essential for seedling establishment. Plant Physiol **131:** 1834–1842

Sage RF, Pearcy RW (1987) The nitrogen use efficiency of C_3 and C_4 plants. I. Leaf nitrogen, growth, and biomass partitioning in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). Plant Physiol **84:** 954–58

Sage RF, Pearcy RW, Seemann JR (1987) The nitrogen use efficiency of C_3 and C_4 plants. III. Leaf nitrogen effects on the activity of carboxylating enzymes in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). Plant Physiol **85:** 355–59

Sage RF, Li M, Monson RK (1999a) The taxonomic distribution of C_4 photosynthesis. **In:** Sage RF, Monson RK (eds) C_4 plant biology. Academic Press, San Diego, pp 551–584

Sage RF, Wedin DA, Li MR (1999b) The biogeography of C_4 photosynthesis: patterns and controlling factors. **In:** Sage RF, Monson RK (eds) C_4 plant biology. Academic Press, San Diego, pp 313–373

Sage RF (2004) The evolution of C_4 photosynthesis. New Phytol **161:** 341–370

Sage RF, Christin PA, Edwards EJ (2011) The C_4 plant lineages of planet Earth. J Exp Bot **62:** 3155–3169

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ Jr (2005) Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol **6:** R33

Sheen J (1991) Molecular mechanisms underlying the differential expression of maize *PPDK* genes. Plant Cell **3:** 225–45

Sheen J (1999) C_4 gene expression. Annu Rev Plant Physiol Plant Mol Biol **50:** 187–217

Shi W, Zhou W (2006) Frequency distribution of TATA Box and extension sequences on human promoters. BMC Bioinformatics **7(Suppl 4):** S2

Smale ST, Baltimore D (1989) The "initiator" as a transcription control element. Cell **57:** 103–113

Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem **72:** 449–479

Smith AM, Stitt M (2007) Coordination of carbon supply and plant growth. Plant Cell Environ **30:** 1126–1149

Srinivasan R, Kraus C, Oliver OJ (1992) Developmental expression of the glycine decarboxylase multienzyme complex in greening pea leaves. **In:** Lambers H, van der Plas LHW (eds) Molecular, biochemical, and physiological aspects of plant respiration. SPB Academic, the Hague, pp 323–334

Srinivasan R, Oliver DJ (1995) Light-dependent and tissue-specific expression of the H-protein of the glycine decarboxylase complex. Plant Physiol **109:** 161–168

Sudderth EA, Muhaidat RM, McKown AD, Kocacinar F, Sage RF (2007) Leaf anatomy, gas exchange and photosynthetic enzyme activity in *Flaveria kochiana*. Funct Plant Biol **34:** 118–129

Svensson P, Bläsing OE, Westhoff P (1997) Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase – A comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3). Eur J Biochem **246:** 452–460

Svensson P, Bläsing OE, Westhoff P (2003) Evolution of C_4 phosphoenolpyruvate carboxylase. Arch Biochem Biophys **414:** 180–188

Tcherkez GG, Farquhar GD, Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proc Natl Acad Sci U S A **103:** 7246–7251

Tolbert NE (1971) Microbodies: peroxisomes and glyoxysomes. Annu Rev Plant Physiol **22:** 45–74

Tronconi MA, Fahnenstich H, Gerrard Weehler MC, Andreo CS, Flugge UI, Drincovich MF, Maurino VG (2008) Arabidopsis NAD-malic enzyme functions as a homodimer and heterodimer and has a major impact on nocturnal metabolism. Plant Physiol **146:** 1540–1552

Tsai FT, Sigler PB (2000) Structural basis of preinitiation complex assembly on human pol II promoters. Embo J **19:** 25–36

Valenzuela L, Kamakaka RT (2006) Chromatin insulators. Annu Rev Genet **40:** 107–138

Vauclare P, Diallo N, Bourguignon J, Macherel D, Douce R (1996) Regulation of the expression of the glycine decarboxylase during pea leaf development. Plant Physiol **112:** 1523–1530

Vedel V, Scotti I (2011) Promoting the promoter. Plant Science **180:** 182–189

Walker RP, Chen ZH, Tecsi LI, Famiani F, Lea PJ, Leegood RC (1999) Phosphoenolpyruvate carboxykinase plays a role in interactions of carbon and nitrogen metabolism during grape seed development. Planta **210:** 9–18

Wang L, Peterson RB, Brutnell TP (2011) Regulatory mechanisms underlying C_4 photosynthesis. New Phytologist **190:** 9–20

Wedding R (1989) Malic enzymes of higher plants: characteristics, regulation, and physiological function. Plant Physiol **90:** 367–371

West-Eberhard MJ, Smith JAC, Winter K (2011) Photosynthesis, reorganized. Science **332:** 311–312

Westhoff P, Gowik U (2004) Evolution of C_4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus *Flaveria*. Ann Bot **93:** 13–23

Westhoff P, Gowik U (2010) Evolution of C_4 photosynthesis – Looking for the master switch. Plant Physiol **154:** 598–601

Wyrich R, Dressen U, Brockmann S, Streubel M, Chang C, Qiang D, Paterson AH, Westhoff P (1998) The molecular basis of C_4 photosynthesis in sorghum: isolation, characterization and RFLP mapping of mesophyll- and bundle-sheath-specific cDNAs obtained by differential screening. Plant Mol Biol **37:** 319–335

Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics **8:** 67

Yamamoto YY, Yoshitsugu T, Sakurai T, Seki M, Shinozaki K, Obokata J (2009) Heterogeneity of Arabidopsis core promoters revealed by high density TSS analysis. Plant J **60:** 350–362

Yamamoto YY, Yoshioka Y, Hyakumachi M, Obokata J (2011) Characteristics of core promoter types with respect to gene structure and expression in *Arabidopsis thaliana*. DNA Res: 1–10

Yoshimura Y, Kubota F, Ueno O (2004) Structural and biochemical bases of photorespiration in C_4 plants: quantification of organelles and glycine decarboxylase. Planta **220:** 307–317

Zelitch I, Schultes NP, Peterson RB, Brown P, Brutnell TP (2009) High glycolate oxidase activity is required for survival of maize in normal air. Plant Physiol **149:** 195–204

Zuo YC, Li QZ (2011) Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. Genomics **97:** 112–120

VI. Manuscripts

1. Sascha Engelmann, Christian Wiludda, Janet Burscheidt, Udo Gowik, Ute Schlue, Maria Koczor, Monika Streubel, Roberto Cossu, Hermann Bauwe, Peter Westhoff (2008). **The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: Analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and Arabidopsis (C_3).** *Plant Physiology* **146:** 1773–1785

2. Christian Wiludda, Stefanie Schulze, Udo Gowik, Sascha Engelmann, Maria Koczor, Monika Streubel, Hermann Bauwe, Peter Westhoff (2011). **Regulation of the photorespiratory *GLDPA* gene in C_4 *Flaveria* – an intricate interplay of transcriptional and post-transcriptional processes.** Submitted for publication to *The Plant Cell*.

Manuscript 1

The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: Analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and Arabidopsis (C_3)

The Gene for the P-Subunit of Glycine Decarboxylase from the C_4 Species *Flaveria trinervia*: Analysis of Transcriptional Control in Transgenic *Flaveria bidentis* (C_4) and Arabidopsis (C_3)[1][W][OA]

Sascha Engelmann, Christian Wiludda, Janet Burscheidt, Udo Gowik, Ute Schlue, Maria Koczor, Monika Streubel, Roberto Cossu, Hermann Bauwe, and Peter Westhoff*

Institut für Entwicklungs- und Molekularbiologie der Pflanzen, Heinrich-Heine-Universität, 40225 Duesseldorf, Germany (S.E., C.W., J.B., U.G., U.S., M.K., M.S., P.W.); Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany (R.C.); and Abteilung Pflanzenphysiologie der Universität Rostock, 18051 Rostock, Germany (H.B.)

Glycine decarboxylase (GDC) plays an important role in the photorespiratory metabolism of plants. GDC is composed of four subunits (P, H, L, and T) with the P-subunit (GLDP) serving as the actual decarboxylating unit. In C_3 plants, GDC can be found in all photosynthetic cells, whereas in leaves of C_3-C_4 intermediate and C_4 species its occurrence is restricted to bundle-sheath cells. The specific expression of GLDP in bundle-sheath cells might have constituted a biochemical starting point for the evolution of C_4 photosynthesis. To understand the molecular mechanisms responsible for restricting *GLDP* expression to bundle-sheath cells, we performed a functional analysis of the *GLDPA* promoter from the C_4 species *Flaveria trinervia*. Expression of a promoter-reporter gene fusion in transgenic plants of the transformable C_4 species *Flaveria bidentis* (C_4) showed that 1,571 bp of the *GLDPA* 5′ flanking region contain all the necessary information for the specific expression in bundle-sheath cells and vascular bundles. Interestingly, we found that the *GLDPA* promoter of *F. trinervia* exhibits a C_4-like spatial activity also in the C_3 plant Arabidopsis (*Arabidopsis thaliana*), indicating that a mechanism for bundle-sheath-specific expression is also present in this C_3 species. Using transgenic Arabidopsis, promoter deletion studies identified two regions in the *GLDPA* promoter, one conferring repression of gene expression in mesophyll cells and one functioning as a general transcriptional enhancer. Subsequent analyses in transgenic *F. bidentis* confirmed that these two segments fulfill the same function also in the C_4 context.

Net photosynthetic CO_2 assimilation rates in C_3 plants are reduced by photorespiration, a process that results from the oxygenase activity of Rubisco. C_4 plants usually show no apparent photorespiration, and this is achieved by splitting the photosynthetic reactions between two morphologically and biochemically distinct cell types, the mesophyll and the bundle-sheath cells. Initial CO_2 fixation in C_4 plants occurs exclusively in the mesophyll cells and is performed by the enzyme phospho*enol*pyruvate carboxylase (PEPC) to form a C_4 acid, oxaloacetate. Depending on the C_4 subtype, oxaloacetate is converted to either malate or Asp, which subsequently move to the bundle-sheath and become decarboxylated, resulting in significant elevation of the CO_2 concentration in these cells. Final refixation of CO_2 is achieved by Rubisco, which in C_4 plants is only present in bundle-sheath cells. The enrichment of CO_2 in the vicinity of Rubisco effectively inhibits the enzyme's oxygenase activity (Hatch, 1987).

In C_4 plants, the CO_2 assimilatory enzymes are compartmentalized into either mesophyll or bundle-sheath cells, and this is governed by differential gene expression. It has been shown that mesophyll-specific expression of C_4 cycle genes is mainly regulated at the transcriptional level (Schaffner and Sheen, 1992; Stockhaus et al., 1994; Rosche et al., 1998; Sheen, 1999), whereas bundle-sheath-specific expression is controlled at both transcriptional and posttranscriptional levels (Long and Berry, 1996; Marshall et al., 1997; Sheen, 1999; Patel et al., 2006).

There are indications that photorespiration also exists in C_4 plants, albeit at a much lower level than in C_3 plants (Osmond and Harris, 1971; Furbank and Badger, 1983; de Veau and Burris, 1989). This is likely due to the fact that photorespiration in C_4 species is strictly confined to the bundle-sheath cells in the leaves, while in

[1] This work was supported by the Deutsche Forschungsgemeinschaft within the SFB 590. Initial studies on the *Flaveria trinervia GLDP* gene were funded by the Bundesministerium für Bildung und Forschung (grant to H.B.).

* Corresponding author; e-mail west@uni-duesseldorf.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Peter Westhoff (west@uni-duesseldorf.de).

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.107.114462

C_3 plants photorespiration occurs in all photosynthetically active mesophyll cells (Ohnishi and Kanai, 1983).

The mitochondrial multienzyme complex Gly decarboxylase (GDC) plays a key role in the photorespiratory pathway. GDC is composed of four different subunits (P, H, T, and L) and catalyzes, in cooperation with Ser hydroxymethyltransferase, the oxidative decarboxylation of Gly that originates from the breakdown of photorespiratory phosphoglycolate. In the course of these reactions, two molecules of Gly are converted to one molecule each of Ser, NH_3, and CO_2 (Neuburger et al., 1986). Consistent with the compartmentation of the photorespiratory cycle, GDC is present in all photosynthetic cells of C_3 plant leaves, but strictly confined to the bundle-sheath cells in C_4 species (Ohnishi and Kanai, 1983). None of the GDC subunits has been detected in the mesophyll cells of C_4 plants (Morgan et al., 1993).

Plant species possessing a C_3-C_4 intermediate type of photosynthesis are of special interest for studying the evolution of C_4-characteristic traits. Some C_3-C_4 plants are to some extent able to fix CO_2 into malate and Asp (Monson et al., 1986), but most of these species do not have a C_4 metabolism. They can be characterized as "intermediate," for example, by their CO_2 compensation points, which are lower than those of C_3 plants but higher than those of C_4 plants (Edwards and Ku, 1987; Rawsthorne, 1992). As in C_4 plants, functional GDC occurs only in the bundle-sheath cells of the leaves of C_3-C_4 intermediate plants, as it was first demonstrated for *Moricandia arvensis* (Rawsthorne et al., 1988a, 1988b). Later on, it was discovered that the loss of GDC activity in the mesophyll is due to a lack of the P-subunit (GLDP). Because of the absence of GDC activity in mesophyll cells of *M. arvensis* leaves, photorespiratory Gly moves to the bundle-sheath cells to be processed by GDC. The bundle-sheath cells of *M. arvensis* contain a large number of mitochondria that are arranged at the centripetal cell wall adjacent to the vascular tissue, whereas the chloroplasts are located at the cell periphery. This special distribution of organelles and the restriction of Gly oxidation to the bundle-sheath compartment result in an efficient recapture of released photorespiratory CO_2, thereby lowering the CO_2 compensation point when compared to a typical C_3 plant (Rawsthorne et al., 1998). C_3-C_4 intermediate species are thought to represent a stage in the evolutionary transition from C_3 to C_4 photosynthesis (Edwards and Ku, 1987). It was therefore tempting to speculate that the confinement of GDC to the bundle-sheath cells has been one of the biochemical starting points for the evolution of C_4 plants (Morgan et al., 1993; Bauwe and Kolukisaoglu, 2003; Sage, 2004). However, the possible effects of this relocation for C_4 evolution are discussed controversially (Edwards et al., 2001). The loss of the P-subunit seems to be the initial step to inactivate GDC in the mesophyll, and the absence of all GDC subunits in the leaf mesophyll of other C_3-C_4 intermediate species suggests that they have developed further toward C_4 photosynthesis than *M. arvensis* (Morgan et al., 1993).

A well-established experimental system for investigating the evolution of C_4-characteristic traits is the genus *Flaveria* of the Asteraceae (Powell, 1978). This small genus comprising 23 known species includes both C_3 and C_4 species, but also a large number of C_3-C_4 intermediate species (Edwards and Ku, 1987). In this study we examined the promoter of the gene encoding GLDP from the C_4 species *Flaveria trinervia* to gain insight into the molecular basis of bundle-sheath-specific gene expression. Two genes encoding GLDP have been identified in *F. trinervia*, *GLDPA* and the pseudogene *GLDPB* (Cossu, 1997). The *GLDPA* promoter was fused to a GUS reporter gene and promoter activity was analyzed in transgenic *Flaveria bidentis* (C_4) and Arabidopsis (*Arabidopsis thaliana*; C_3). Similar expression patterns were observed in these two species, which allowed the use of Arabidopsis as a heterologous system for testing a series of promoter deletions to identify C_4-characteristic regulatory elements within the *GLDPA* promoter. These analyses resulted in the identification of regions within the *GLDPA* promoter that contribute mainly to the regulation of expression quantity or to the spatial expression pattern of the *GLDPA* gene, respectively.

RESULTS

In Situ RNA Hybridization

Immunolabeling studies have shown that, in C_4 plants, GLDP accumulates exclusively in bundle-sheath cells of leaves (Hylton et al., 1988; Morgan et al., 1993; Yoshimura et al., 2004). To examine whether this C_4-characteristic localization of the P-protein is due to specific accumulation of *GLDPA* mRNA in this compartment, we analyzed the expression pattern of the *GLDPA* gene in leaves of the C_4 species *F. trinervia* by in situ hybridization. As a probe we used a 2.4-kb fragment of the *GLDPA* cDNA from *F. trinervia*, and control hybridizations were performed with the corresponding sense probe.

In leaves of *F. trinervia*, transcripts of the *GLDPA* gene could only be detected in bundle-sheath and not in mesophyll cells (Fig. 1A). The *GLDPA* mRNA accumulated near the centripetal cell walls of the bundle-sheath cells due to the concentration of cytoplasm in this region. The confinement of the P-protein to the bundle-sheath cells therefore is controlled by the specific accumulation of *GLDPA* mRNA in this compartment. The same result was obtained by in situ hybridization of the *GLDPA* probe to leaf cross sections of the C_4 species *F. bidentis* (Fig. 1C).

Expression of a GUS Reporter Gene under the Control of the *GLDPA* Promoter from *F. trinervia* in Transgenic *F. bidentis*

The in situ RNA hybridization analysis showed that the occurrence of *GLDPA* transcripts is restricted to the

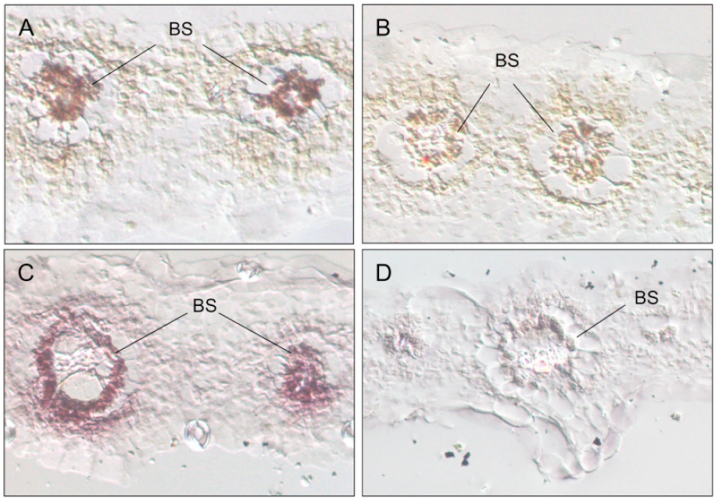


Figure 1. In situ localization of *GLDPA* RNA accumulation in leaves of *F. trinervia* (A) and *F. bidentis* (C). Hybridizations with a *GLDPA* sense probe to leaf cross sections of *F. trinervia* (B) and *F. bidentis* (D) served as negative controls. BS, Bundle-sheath cell.

bundle-sheath cells in *F. trinervia* and *F. bidentis* (Fig. 1). To test whether the available 1,571 bp of the 5′ flanking region of the *GLDPA* gene (including the 5′ untranslated region upstream of the AUG start codon) harbor all the necessary information for this bundle-sheath-specific expression pattern, we fused this region to GUS reporter gene (construct *GLDPA*-Ft; Fig. 2A) and examined its expression behavior in transgenic *F. bidentis* plants. The C_4 species *F. bidentis* is a close relative to *F. trinervia*, but unlike *F. trinervia* it is suitable for transformation by *Agrobacterium tumefaciens*-mediated gene transfer (Chitty et al., 1994).

Histochemical analysis of the expression of the *GLDPA*-Ft promoter-GUS construct revealed an intense blue staining in the bundle-sheath cells but not in the mesophyll cells (Fig. 2B). GUS activity could also be observed in most vascular bundles, with the degree of GUS expression varying with the size of the veins. The small minor veins usually exhibited a strong blue staining, while higher-order vascular bundles showed only moderate GUS activity. Additional weak *GLDPA* promoter activity was also detected in the guard cells of the stomatal complexes (Fig. 2C).

The Expression Pattern of the *GLDPA*-GUS Construct in Arabidopsis Is Similar to That in *F. bidentis*

Bundle-sheath cells are not a unique feature of C_4 plants. They are also present in many C_3 plants, but compared to the situation in C_4 species, these cells exhibit fewer chloroplasts and mitochondria (Kinsman and Pyke, 1998; Leegood, 2002). To examine whether the *GLDPA* promoter of *F. trinervia* shows a cell-specific activity in a C_3 background, we introduced the *GLDPA*-GUS construct into Arabidopsis.

The histochemical analysis revealed GUS expression in the vascular tissue and in the surrounding bundle-sheath cells (Fig. 3, C and D). Notably, very similar to the expression pattern in *F. bidentis*, no GUS activity could be detected in the mesophyll cells of transgenic Arabidopsis plants. The quantification of GUS levels showed that the median activity of the reporter protein was comparable in Arabidopsis and *F. bidentis* leaves (Figs. 2D and 3B).

To verify the results obtained from the histochemical GUS analysis, the *GLDPA* promoter was also fused to the GFP reporter gene *mgfp5-ER* (Siemering et al., 1996; Haseloff et al., 1997) and a histone 2B/ yellow fluorescent protein (*H2B:YFP*) fusion gene (Boisnard-Lorig et al., 2001), which are targeted to the endoplasmic reticulum and the nucleus, respectively. These reporter proteins allow a nondestructive analysis by fluorescence or confocal laser microscopy, thereby avoiding any potential diffusion of the reporter protein that might occur during a histochemical staining procedure. In both cases, the reporter proteins could be detected in bundle-sheath cells and vascular bundles, but not in mesophyll cells (Fig. 3, E–L; Supplemental Videos S1 and S2).

Deletion Analysis of the *GLDPA* Promoter

The "C_4-like" expression pattern of *GLDPA*-GUS in transgenic Arabidopsis provided the opportunity to functionally dissect the *GLDPA* promoter by using this C_3 model organism as an experimental system. To identify cis-regulatory determinants that are responsible for the activity of the *GLDPA* promoter in bundle-sheath cells and the vascular bundle, we produced a set of 5′ deletions and analyzed their expression specificity and level in transgenic Arabidopsis. The *GLDPA* promoter was subdivided into seven fragments that are referred to as region 1 to region 7 in the following (Fig. 4A).

The removal of a 182-bp segment (region 1) from the 5′ end of the *GLDPA* promoter, resulting in construct *GLDPA*-Ft-Δ1 (Fig. 4A), did not alter the spatial expression pattern of the GUS reporter gene when com-

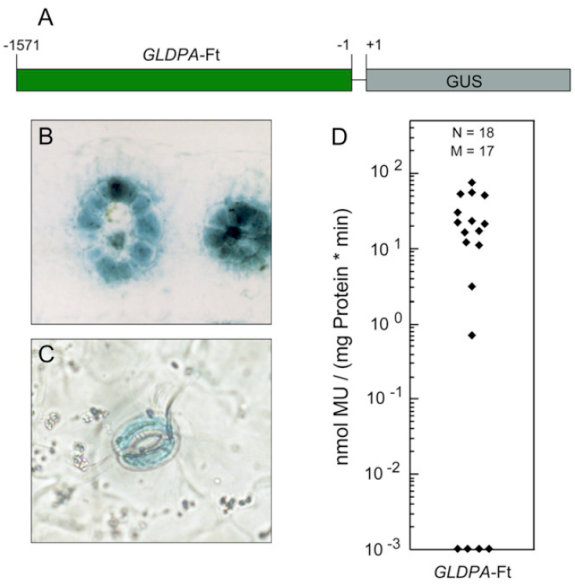


Figure 2. Analysis of *GLDPA*-Ft promoter activity in transgenic *F. bidentis*. A, Structure of the chimeric *GLDPA*-Ft::*GUS* gene. B and C, Histochemical localization of GUS activity (blue staining) in leaf sections of transgenic *F. bidentis* transformed with *GLDPA*-Ft. The photograph (C) was taken from the leaf surface displaying GUS activity in the stomatal guard cells of *F. bidentis* leaves. Incubation times were 1 h. D, GUS activities in leaves of transgenic *F. bidentis* plants transformed with *GLDPA*-Ft. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per milligram of protein per minute. The numbers of independent transgenic plants investigated (N) and the median values of GUS activity (M) are indicated at the top of each column.

pared to the original full-length promoter, i.e. this promoter variant was still capable of directing GUS expression specifically in the vascular bundles and bundle-sheath cells of Arabidopsis (Fig. 4B). This indicates the absence of cis-regulatory elements conferring cell specificity in this most distal part of the *GLDPA* promoter. However, quantitative GUS assays revealed an approximately 20-fold lower GUS activity in leaves of transgenic plants expressing *GLDPA*-Ft-Δ1 relative to leaves of *GLDPA*-Ft plants (Fig. 4D). The transcriptional activity of the promoter further decreased when deleting another 251 bp (region 2) from the 5′ end of *GLDPA*-Ft-Δ1 (construct *GLDPA*-Ft-Δ2; Fig. 4A). GUS activity in leaves of Arabidopsis plants harboring the *GLDPA*-Ft-Δ2 transgene was about 60 times lower than that of *GLDPA*-Ft (Fig. 4D), but the spatial expression pattern in the Arabidopsis leaf was still identical to that of the full-length promoter construct (Fig. 4C). Region 3 of the *GLDPA* promoter included the sequences between −1,138 and −927. Deletion of this promoter fragment in construct *GLDPA*-Ft-Δ3 decreased the GUS activity below the sensitivity limit of the histochemical GUS assay. Hence, no GUS expression could be detected in leaf cross sections of plants carrying this promoter construct. Similar low GUS activities (using the quantitative GUS assay) were also observed for the constructs *GLDPA*-Ft-Δ4 and *GLDPA*-Ft-Δ5 (Fig. 4, A and D), while practically no GUS activity was observed for *GLDPA*-Ft-Δ6 (Fig. 4, A and D).

This deletion analysis clearly demonstrated the pronounced importance of regions 1, 2, and 3 for the transcriptional activity of the *GLDPA* promoter in the leaves of transgenic Arabidopsis plants. While truncation of regions 1 and 2 causes a dramatic decrease of transcriptional activity without affecting the spatial expression pattern, the additional deletion of region 3 results in a further reduction of promoter activity, which impeded further analysis of cell type-specific expression within the leaf.

Regions 1 and 2 of the *GLDPA* Promoter Together Function as a General Enhancer of Transcription in the Arabidopsis Leaf

To investigate whether the *GLDPA* promoter fragment reaching from −1,571 to −1,139 (regions 1 and 2) was able to act as a transcriptional enhancer, we combined this segment of the promoter with region 7 of the *GLDPA* promoter (−298 to −1). Region 7 harbors a putative TATA box and the starting point of transcription, but as reported above this part of the promoter alone is not sufficient to drive GUS expression in the Arabidopsis leaf (Fig. 4D).

The transformation of construct *GLDPA*-Ft-1-2-7 (Fig. 5A) into Arabidopsis caused a substantial level of GUS expression in mesophyll and bundle-sheath cells as well as in the vascular strands of the leaves (Fig. 5, B and C) and confirms that regions 1 and 2 contain transcriptional enhancers with no apparent cell type specificity within the leaf.

To test whether regions 1 and 2 of the *GLDPA* promoter function also in a heterologous promoter context, we fused this segment in front of the proximal promoter region of the *ppcA1* gene of *F. trinervia* (Fig. 5A). The *ppcA1* gene encodes the C_4 isoform of PEPC (Hermans and Westhoff, 1992), and its complete 2,188-bp promoter directs high and mesophyll-specific GUS expression in transgenic *F. bidentis* (Stockhaus et al., 1997). In contrast, the activity of the 570-bp-long proximal *ppcA* promoter part (*ppcA*-PR_{Ft}) is extremely low and can hardly be visualized in histochemical GUS assays (Gowik et al., 2004). In plants in which the low activity permits a histological analysis, the *ppcA*-PR_{Ft} promoter fragment directs a uniform expression in all cells of the leaves of *F. bidentis*, including the vascular bundles (Akyildiz et al., 2007).

The fusion of regions 1 and 2 of the *GLDPA* promoter with the *ppcA*-PR_{Ft} promoter fragment resulted in strong GUS expression in leaves of Arabidopsis (Fig. 5B). The GUS reporter gene was active in both mesophyll and bundle-sheath cells as well as in the vascular bundles (Fig. 5D), and the expression profile of this chimeric promoter is thus indistinguishable

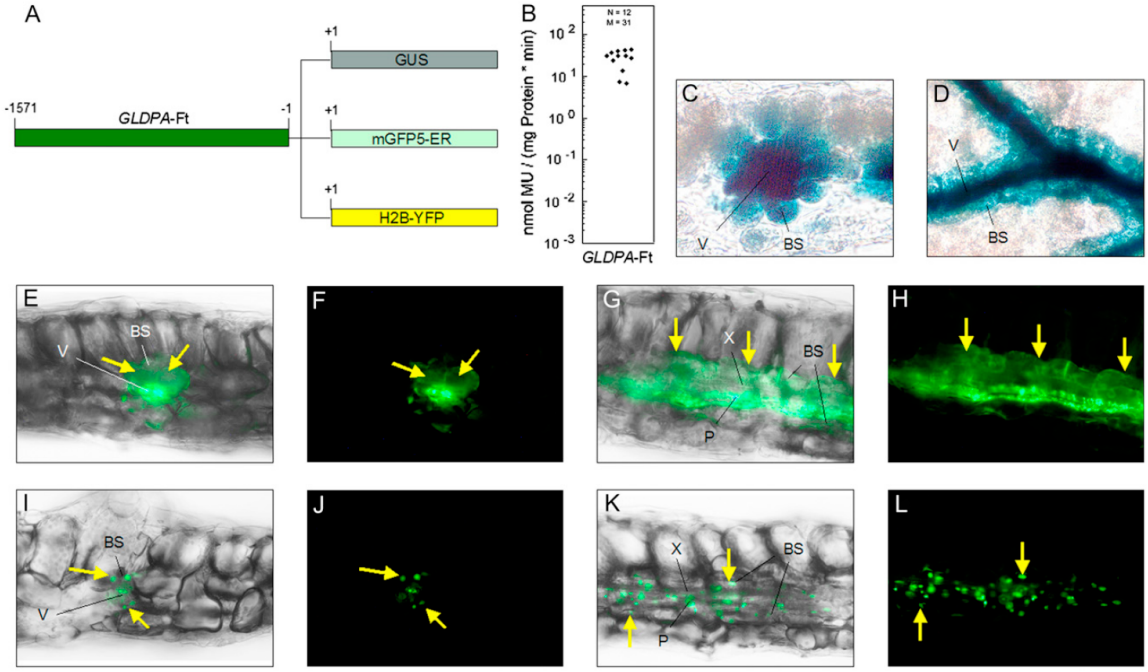


Figure 3. Analysis of *GLDPA*-Ft promoter activity in transgenic Arabidopsis. A, Schematic structure of the chimeric *GLDPA*-Ft::*GUS*, *GLDPA*-Ft::*mGFP5-ER*, and *GLDPA*-Ft::*H2B:YFP* genes. B, GUS activities in leaves of transgenic Arabidopsis plants transformed with *GLDPA*-Ft::*GUS*. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per milligram of protein per minute. The numbers of independent transgenic plants investigated (N) and the median values of GUS activity (M) are indicated at the top of each column. C and D, Histochemical localization of GUS activity (blue staining) in leaf sections of transgenic Arabidopsis transformed with *GLDPA*-Ft::*GUS*. Picture (D) was taken from the leaf surface displaying GUS activity in the vascular strands and the adjacent layer of bundle-sheath cells in Arabidopsis. Incubation times were 1 h. E to L, Fluorescence microscopy images taken from leaf cross sections of Arabidopsis plants transformed with the construct *GLDPA*-Ft::*mGFP5-ER* (E–H) or *GLDPA*-Ft::*H2B:YFP* (I–L). For E, G, I, and K, images taken with a GFP imaging filter system were merged with corresponding bright-field images. Photographs in F, H, J, and L were captured using the GFP imaging filter system alone. Selected bundle-sheath cells that express the reporter gene are highlighted by yellow arrows. BS, Bundle-sheath cell; V, vascular tissue; X, xylem; P, phloem.

from that of *GLDPA*-Ft-1-2-7. We conclude from these experiments that regions 1 and 2 of the *GLDPA* promoter constitute a general transcriptional enhancer module that, in combination with a basal promoter, stimulates the expression of a linked reporter gene in all types of interior leaf cells of Arabidopsis.

Region 3 of the *GLDPA* Promoter Acts as a Mesophyll-Specific Repressor of Gene Expression

The role of region 3 (−1,138 to −927) in regulating *GLDPA* promoter activity was investigated by introducing the relevant promoter fragment into construct *GLDPA*-Ft-1-2-7, resulting in the production of the chimeric promoter *GLDPA*-Ft-1-2-3-7 (Fig. 6A). The addition of region 3 to *GLDPA*-Ft-1-2-7 caused a significant change in the spatial expression pattern of the GUS reporter gene. While *GLDPA*-Ft-1-2-7 plants expressed the GUS reporter gene in mesophyll and bundle-sheath cells as well as in the vascular tissue (Fig. 5C), GUS activity of *GLDPA*-Ft-1-2-3-7 plants was strictly confined to the bundle-sheath cells and the vascular compartment (Fig. 6B). These observations indicate that region 3 of the *GLDPA* promoter from *F. trinervia* functions as a mesophyll-specific repressor of gene expression in the Arabidopsis leaf.

Analysis of *GLDPA* Promoter Regions 4, 5, and 6

We have shown that the *GLDPA* promoter fragment comprising base pairs −1,571 to −927 (regions 1–3) in combination with the most proximal promoter part (region 7) is sufficient to direct GUS expression in the bundle-sheath cells and vascular bundles of transgenic Arabidopsis plants. Nevertheless, additional cis-regulatory determinants that could be involved in the spatial regulation of transcriptional activity might also be present in promoter regions 4, 5, and 6. To investigate the occurrence of cis-regulatory elements within these promoter regions, it was necessary to raise the GUS expression levels of constructs *GLDPA*-Ft-Δ3 to *GLDPA*-Ft-Δ6 to a level that allowed a histochemical

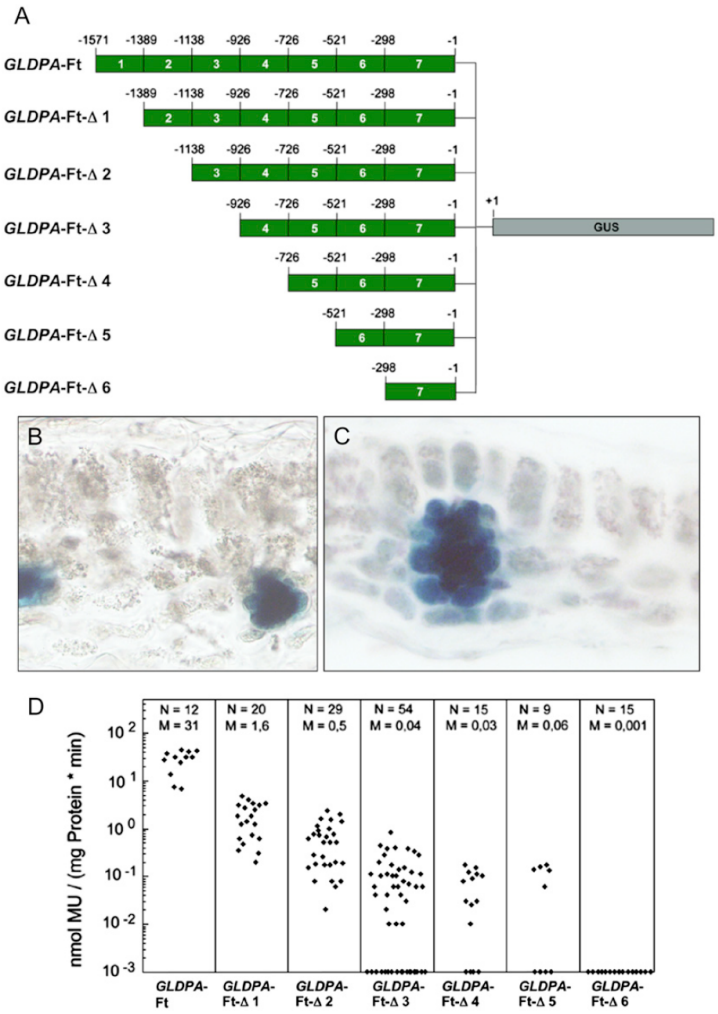


Figure 4. Deletion analysis of the *GLDPA* promoter of *F. trinervia* in the C_3 plant Arabidopsis. A, Overview of *GLDPA* promoter deletion constructs analyzed in transgenic Arabidopsis. Due to the cloning strategy, the *GLDPA* promoter of *F. trinervia* was subdivided into seven regions. B and C, Histochemical localization of GUS activity in leaf sections of transgenic Arabidopsis transformed with deletion construct *GLDPA*-Ft-Δ1 (B) or *GLDPA*-Ft-Δ2 (C). Incubation times were 12 h (A) and 20 h (B). D, GUS activities in leaves of transgenic Arabidopsis plants transformed with different *GLDPA* deletion constructs. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per milligram of protein per minute. The numbers of independent transgenic plants investigated (N) and the median values of GUS activity (M) are indicated at the top of each column.

analysis. Since regions 1 and 2 of the *GLDPA* promoter contain a general transcriptional enhancer with no apparent leaf cell specificity, we attached this *GLDPA* transcriptional enhancer module to the 5′ borders of the truncated promoters (Fig. 7A).

As expected, the transcriptional activity of these constructs was dramatically higher than that of their "enhancerless" counterparts and was therefore suitable for performing GUS stainings in situ (Figs. 4D and 7B). In construct *GLDPA*-Ft-1-2-4-5-6-7, only region 3 was removed from the original full-length *GLDPA* promoter. While about 50% of the transgenic lines displayed a uniform GUS expression in mesophyll and bundle-sheath cells and the vascular bundles (Fig. 7F), GUS expression in the other half of the plant lines was still restricted to the bundle-sheath cells and the vascular tissue (Fig. 7C). The same distribution of transgenic plants displaying either a uniform or restricted expression of the reporter gene was also observed for the promoter constructs *GLDPA*-Ft-1-2-5-6-7 and *GLDPA*-Ft-1-2-6-7 in which regions 4 and 5 were further deleted (Fig. 7, D, E, G, and H). In contrast, as already reported above, the additional deletion of region 6 in construct *GLDPA*-Ft-1-2-7 resulted in a uniform expression pattern in the leaf (Fig. 5C). These findings suggest that additional cis-regulatory elements conferring repression of gene expression in the mesophyll are located in region 6 of the *GLDPA* promoter. However, when compared to the highly effective repressor elements located in region 3, these additional elements in region 6 do not provide robust repression.

Analysis of Promoter Construct *GLDPA*-Ft-1-2-3-7 in Transgenic *F. bidentis*

A truncated promoter containing the transcription-enhancing regions 1 and 2, the mesophyll repressor region 3, and the basal expression segment 7 generated

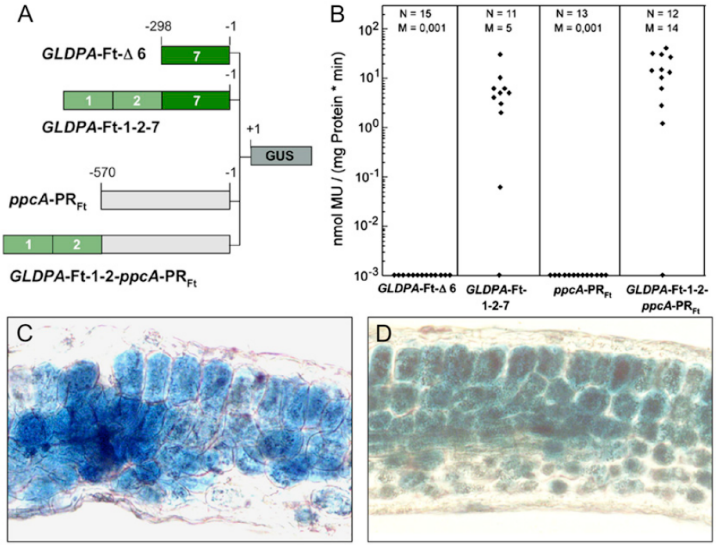


Figure 5. Functional analysis of *GLDPA* promoter regions 1 and 2 in transgenic Arabidopsis. A, Chimeric promoter constructs used for plant transformation. B, GUS activities in leaves of transgenic Arabidopsis plants transformed with different promoter constructs. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per milligram of protein per minute. The numbers of independent transgenic plants investigated (N) and the median values of GUS activity (M) are indicated at the top of each column. C and D, Histochemical localization of GUS activity in leaf sections of transgenic Arabidopsis plants transformed with *GLDPA*-Ft-1-2-7 (C) and *GLDPA*-Ft-1-2-*ppcA*-PR_{Ft} (D). Incubation times were 12 h.

the same spatial expression profile in the leaf of the C_3 plant Arabidopsis as the complete *GLDPA* promoter, i.e. the promoter regions 4, 5, and 6 (−926 to −299) were not essential for creating the C_4-characteristic spatial expression pattern of a reporter gene. We now wished to examine whether this chimeric *GLDPA*-Ft-1-2-3-7 promoter is capable of providing this C_4 expression profile also in the C_4 background of *F. bidentis*. This chimeric promoter construct was therefore transformed into *F. bidentis*, and its expression was examined in the leaves of the transgenic plants (Fig. 8).

No differences between the spatial expression patterns of *GLDPA*-Ft-1-2-3-7 and the full-length promoter construct *GLDPA*-Ft were observed (compare Figs. 2 and 8). In both cases, GUS expression was found exclusively in the bundle-sheath cells and—with variable intensities—in the vascular strands. While GUS staining was strong in some minor veins, it was absent from other minor and all major vascular strands (Fig. 8, B and C). These results indicate that regions 1 to 3 in combination with the basal TATA box-containing segment 7 of the *GLDPA* promoter are sufficient to direct reporter gene expression in bundle-sheath cells and the vascular bundles of both the homologous C_4 species *F. bidentis* and the heterologous C_3 plant Arabidopsis.

DISCUSSION

The correct functioning of the C_4 photosynthetic cycle requires strict compartmentalization of C_4 enzymes in either mesophyll or bundle-sheath cells of the leaf. This cell type-specific accumulation of proteins is governed by differential gene expression (Sheen, 1999). To broaden our knowledge on the molecular basis of bundle-sheath-specific gene expression in C_4 plants, we have performed a functional analysis of the 5′ flanking sequences of the *GLDPA* gene from *F. trinervia* (C_4). The *GLDPA* gene encodes GLDP, which is specifically located in the bundle-sheath cells of C_4 species (Morgan et al., 1993). To determine whether the bundle-sheath-specific accumulation of the GLDP protein in the C_4 leaf is paralleled by the accumulation of the corresponding mRNA, we studied the occurrence of *GLDPA* transcripts within the leaves of *F. trinervia* and *F. bidentis* by in situ hybridizations. *GLDPA* RNA was exclusively found in the bundle-sheath cells of both C_4 species, indicating that the presence of this protein in bundle-sheath cells and its absence in mesophyll cells are caused by differential *GLDPA* mRNA accumulation.

We then investigated whether the available 1,571 bp of the 5′ flanking region of the *GLDPA* coding sequence harbor all the necessary information for this bundle-sheath-specific expression. Fusion of these sequences—including the 5′ untranslated segment of the *GLDPA* gene—to the GUS reporter gene resulted in reporter gene activity in the bundle-sheath but not in the mesophyll cells of transgenic *F. bidentis* plants. In addition, GUS activity could be detected in the vascular bundles. Here, GUS activity was clearly visible in minor veins but very low in major veins.

The expression of the reporter gene in the bundle-sheath cells and the absence of GUS activity in the mesophyll are consistent with the accumulation pattern of the *GLDPA* RNA. The additional activity of the GUS reporter gene in the vascular tissue, however, is in contrast to the lack of detectable *GLDPA* RNA in this tissue. There are two possible explanations that could account for this discrepancy in the patterns of RNA accumulation and reporter gene activity. First, there could be additional cis-regulatory sequences further upstream within the introns or even downstream of the *GLDPA* gene that might control *GLDPA* transcription in

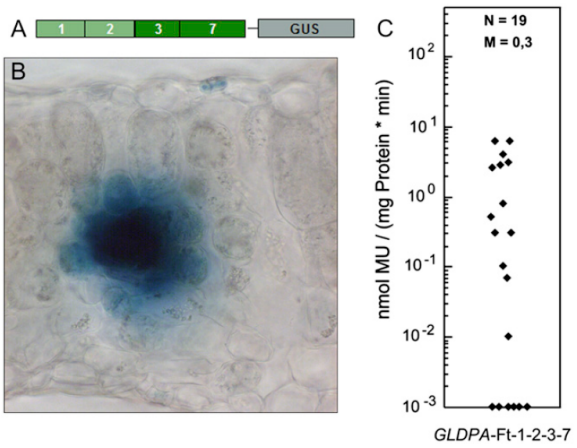


Figure 6. Functional analysis of *GLDPA* promoter region 3 in transgenic Arabidopsis. A, Schematic structure of construct *GLDPA*-Ft-1-2-3-7. B, Histochemical localization of GUS activity in the leaf sections of a transgenic Arabidopsis plant transformed with *GLDPA*-Ft-1-2-3-7. Incubation time was 20 h. C, GUS activities in leaves of transgenic Arabidopsis plants transformed with promoter construct *GLDPA*-Ft-1-2-3-7. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per milligram of protein per minute. The numbers of independent transgenic plants investigated (N) and the median values of GUS activity (M) are indicated at the top of each column.

the natural genome context, but are absent in the *GLDPA*-Ft promoter-GUS construct. These elements would be required for repressing reporter gene expression in the vascular tissue. Alternatively, the absence of *GLDPA* mRNA in the vascular tissue might be caused by the low *GLDPA* RNA stability in this tissue (Parker and Song, 2004; Moore, 2005). While the 5′ untranslated region of the *GLDPA* mRNA is included in the *GLDPA* promoter construct, the 3′ untranslated segment is not. If this latter RNA segment or the coding region contains stability determinants leading to degradation of the *GLDPA* mRNA in the vascular tissue, the GUS mRNA lacking these segments would accumulate.

Interestingly, expression of the *GLDPA* promoter-GUS construct in transgenic Arabidopsis showed a spatial activity of the *GLDPA* promoter that was very similar to that observed in transgenic *F. bidentis* (Fig. 3). As found for the C_4 species, the *GLDPA* promoter of *F. trinervia* was inactive in the leaf mesophyll of the C_3 plant Arabidopsis, but active in the bundle-sheath cells and the vascular bundle. The level of reporter gene expression was similar in both species. We therefore conclude that the C_4-characteristic cis-regulatory transcriptional determinants are recognized in the same spatial manner also in the C_3 context. This indicates that the transcription factors necessary for the correct interpretation of these cis-regulatory sequences are already present in this C_3 species and are operating in the same spatial pattern. Moreover, the similar spatial expression profiles in C_4 and C_3 leaves allow us to conclude that the gene regulatory networks operating in mesophyll and bundle-sheath cells of dicot C_3 and C_4 species share common elements, as it was previously proposed by Matsuoka et al. (1993).

The exact physiological and biochemical functions of bundle-sheath cells in C_3 species are poorly understood. They are involved in phloem loading and unloading (van Bel, 1993), and for tobacco (*Nicotiana tabacum*) it was shown that the bundle-sheath cells of stems and petioles exhibit high activities of enzymes characteristic of C_4 photosynthesis, thus allowing the decarboxylation of four-carbon organic acids derived from the xylem and phloem (Hibberd and Quick, 2002). Additionally, a class of Arabidopsis mutants termed *dov* (differential development of vascular associated cells) demonstrates that differential chloroplast development occurs between bundle-sheath and mesophyll cells in the Arabidopsis leaf (Kinsman and Pyke, 1998).

These observations from tobacco and Arabidopsis provide some evidence that bundle-sheath cells in C_3 plants are somehow predetermined to evolve C_4-characteristic features. The special physiology of bundle-sheath cells in Arabidopsis and the fact that preexisting transcription factors in this C_3 species are able to recognize heterologous C_4-characteristic cis-regulatory elements in the correct fashion provide further evidence for the view that the evolution of C_4 plants must have been relatively simple in genetic terms (Westhoff and Gowik, 2004). C_4-like spatial activities of C_4 promoters in transgenic C_3 plants have also been reported for the C_4 isoform of PEPC of maize (*Zea mays*; Matsuoka et al., 1994; Nomura et al., 2000), the pyruvate orthophosphate dikinase gene of maize (Matsuoka et al., 1993), and the phospho*enol*pyruvate carboxykinase of *Zoysia japonica* (Nomura et al., 2005). On the other hand, the C_4-PEPC promoter of *F. trinervia* loses mesophyll specificity when it is introduced in Arabidopsis (Akyildiz et al., 2007). Similarly, the NADP-dependent malic enzyme promoter from maize loses its bundle-sheath specificity in rice (*Oryza sativa*; Nomura et al., 2005). This shows that the functionality of C_4-specific regulatory cis-elements in C_3 plants cannot be generalized.

The C_4-like expression pattern of the *GLDPA*-Ft promoter in Arabidopsis provided us with the possibility to dissect the functional organization of this promoter in Arabidopsis. A series of *GLDPA* promoter deletion and recombination constructs were analyzed, and two major functional modules were identified and localized, a non-cell-type-specific transcriptional enhancer and a segment that represses gene expression in mesophyll cells.

The transcriptional enhancer is located within the outermost distal regions 1 and 2 of the *GLDPA* promoter comprising base pairs −1,571 to −1,139. The enhancer functioned in all interior leaf tissues of Arabidopsis, i.e. in mesophyll and bundle-sheath cells as well as in the vascular bundle. The transcription-enhancing activity of these regions was not restricted to the context of the *GLDPA* promoter but was also

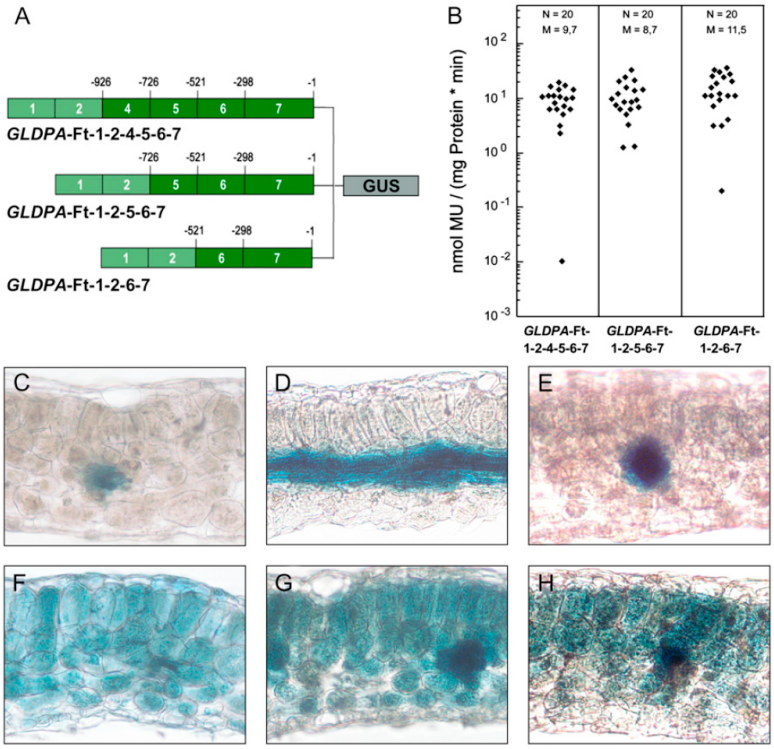


Figure 7. Analysis of promoter constructs *GLDPA*-Ft-1-2-4-5-6-7, *GLDPA*-Ft-1-2-5-6-7, and *GLDPA*-Ft-1-2-6-7 in transgenic Arabidopsis. A, Schematic structure of constructs *GLDPA*-Ft-1-2-4-5-6-7, *GLDPA*-Ft-1-2-5-6-7, and *GLDPA*-Ft-1-2-6-7. B, GUS activities in leaves of transgenic Arabidopsis plants transformed with different promoter constructs. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per milligram of protein per minute. The numbers of independent transgenic plants investigated (N) and the median values of GUS activity (M) are indicated at the top of each column. C to H, Histochemical localization of GUS activity in leaf sections of transgenic Arabidopsis plants transformed with *GLDPA*-Ft-1-2-4-5-6-7 (C and F), *GLDPA*-Ft-1-2-5-6-7 (D and G), or *GLDPA*-Ft-1-2-6-7 (E and H). Incubation times were 4 h (C, D, F, and G) and 5 h (E and H).

functional when combined with the proximal part of the *ppcA1* promoter of *F. trinervia*. The enhancer is thus not *GLDPA* gene specific but functions as a general enhancer module.

The quantitative analysis of promoter activities (Fig. 4D) indicates that region 1 has a higher potential for transcriptional enhancement than region 2. A search for known cis-regulatory elements (Prestridge, 1991; Higo et al., 1999) revealed the presence of a motif with similarity to the simian virus 40 enhancer core (GTGGWWHG) at positions −1,455 to −1,448 in region 1. This motif is also present in a region of the *GLDPA* promoter of *Flaveria pringlei* that is associated with an increase in expression quantity (Bauwe et al., 1995).

Region 3 (−1,138 to −927) harbors cis-regulatory elements that confer cell specificity to the *GLDPA* promoter by repressing its activity in the mesophyll cells of the Arabidopsis leaf. A chimeric promoter consisting of the transcription-enhancing regions 1 and 2, region 3, and the proximal basal expression region 7 is also not active in the mesophyll cells of the C_4 species *F. bidentis*. This indicates that region 3 can repress expression in mesophyll cells also in the C_4 context, i.e. the mesophyll-repressing function of region 3 is conserved between the C_3 and the C_4 species. The lack of *GLDPA* expression in the mesophyll is thus caused by transcriptional regulation and not by posttranscriptional regulation as it was reported for the *FbRbc*S1 gene of *F. bidentis*. *FbRbc*S1 encodes the small subunit of Rubisco, and its bundle-sheath-specific expression is entirely established by selective *rbc*S transcript stabilization in the bundle-sheath cells (Patel et al., 2006).

Additional mesophyll-repressing cis-regulatory sequences are located in region 6 (−521 to −299). They can partially compensate for the lack of the mesophyll-repressing cis-regulatory sequences in region 3, when this segment is not present in the promoter construct. However, these cis-regulatory elements are not able to establish a robust repression of reporter gene activity in the mesophyll cells of Arabidopsis and appear to be of minor importance. This is documented by the cell type-specific expression of construct *GLDPA*-Ft-1-2-3-7 that consists of the transcription-enhancing regions 1 and 2, region 3, and the proximal basal promoter region 7. *GLDPA*-Ft-1-2-3-7 directs a C_4-characteristic expression profile in *F. bidentis*. This demonstrates that the cis-regulatory motives present in region 6 are not necessary for the repression of *GLDPA* expression in mesophyll cells of this C_4 species. Moreover, we can infer from the expression profile of this truncated promoter that regions 4 and 5 are also not necessary to achieve a C_4-typical GUS expression pattern in both Arabidopsis and *F. bidentis*. Regarding the mechanism of mesophyll repression, no predictions can be made at this moment, since searching for known cis-regulatory elements in region 3 did not identify any robust candidate motifs.

The cis-regulatory determinants for the mesophyll-specific repression of *GLDPA* expression in the leaf have

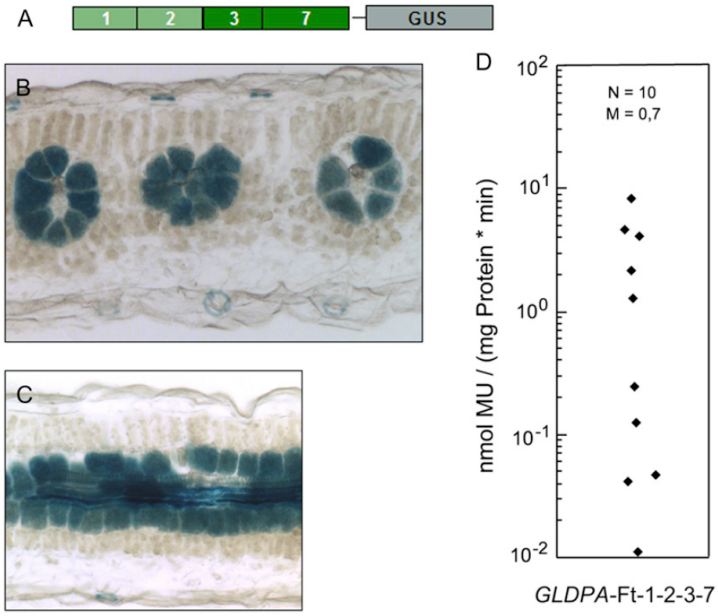


Figure 8. Analysis of the promoter construct *GLDPA*-Ft-1-2-3-7 in transgenic *F. bidentis*. A, Schematic structure of construct *GLDPA*-Ft-1-2-3-7. B and C, Histochemical localization of GUS activity in leaf sections of transgenic *F. bidentis* plants transformed with *GLDPA*-Ft-1-2-3-7. Incubation times were 28 h. D, GUS activities in leaves of transgenic *F. bidentis* plants transformed with *GLDPA*-Ft-1-2-3-7. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per milligram of protein per minute. The numbers of independent transgenic plants investigated (N) and the median values of GUS activity (M) are indicated at the top of each column.

not been determined yet, and no other gene system for bundle-sheath-specific expression has been investigated in such detail that its cis- and trans-regulatory elements are known. However, cis-regulatory elements for mesophyll-specific gene expression have recently been identified at the nucleotide level (Gowik et al., 2004; Akyildiz et al., 2007) for the C_4-PEPC gene *ppcA1* from *F. trinervia*. It was found that variation at two positions in a 41-bp element, a G to A transition and the presence or absence of the tetranucleotide CACT, determines whether the promoter is active in all interior tissues of the leaf of transgenic *F. bidentis* or only in the mesophyll cells (Akyildiz et al., 2007). Hence, for both genes, *GLDPA* and *ppcA1*, cell type-specific gene expression is achieved by repressing the activity of the gene in the respective other cell type. Further dissection of repressor regions 3 and 6 within the *GLDPA* promoter of *F. trinervia* should allow precise mapping of the involved cis-regulatory elements. Such identification of a bundle-sheath-specific expression module consisting of a cis-regulatory sequence and the corresponding transcription factor would mark an important step in our understanding of the evolution of C_4 photosynthesis. Moreover, this information would also be of great value for attempts to convert a C_3 species into a C_4 species, as it has been proposed for rice (Mitchell and Sheehy, 2006).

MATERIALS AND METHODS

Construction of Chimeric Promoters

DNA manipulations and cloning were carried out according to Sambrook and Russell (2001). Construct *GLDPA*-Ft (Cossu, 1997) served as the basis for the series of *GLDPA* promoter deletions. In *GLDPA*-Ft, the 5′ upstream region of the *GLDPA* gene of *Flaveria trinervia* from −1,571 to −1 (+1 describes the first base of the translational start codon) was fused to the GUS cDNA in the binary plant transformation vector pBI121 (CLONTECH Laboratories). Different 5′-deleted fragments of the *GLDPA* promoter were generated by PCR amplification (Tables I and II). The primers added an *Xba*I restriction site at the 5′ border of the DNA fragments and an *Xma*I site at the 3′ end. Therefore, the deleted promoters could be inserted into *Xba*I/*Xma*I-cut pBI121, resulting in the formation of the constructs *GLDPA*-Ft-Δ1, *GLDPA*-Ft-Δ3, *GLDPA*-Ft-Δ4, *GLDPA*-Ft-Δ5, and *GLDPA*-Ft-Δ6. For the production of *GLDPA*-Ft-Δ2, plasmid *GLDPA*-Ft was digested with *Hin*dIII and *Eco*72I. The remaining plasmid was purified by gel electrophoresis and blunt ends were generated by treatment with the Klenow fragment of *Escherichia coli* DNA polymerase I. The vector was religated to form *GLDPA*-Ft-Δ2. The DNA fragment comprising regions 1 and 2 of the *GLDPA*-Ft promoter was generated by PCR amplification with primers *GLDPA*-5′-Xba and Eco72-R. *Xba*I sites were introduced at both ends of the PCR product, which allowed the insertion of the DNA fragment into *Xba*I-cut *GLDPA*-Ft-Δ3, *GLDPA*-Ft-Δ4, *GLDPA*-Ft-Δ5, *GLDPA*-Ft-Δ6, and *ppcA*-S-Ft (Stockhaus et al., 1994). The resulting plasmid constructs were named *GLDPA*-Ft-1-2-4-5-6-7, *GLDPA*-Ft-1-2-5-6-7, *GLDPA*-Ft-1-2-6-7, *GLDPA*-Ft-1-2-7, and *GLDPA*-Ft-1-2-*ppcA*-PR_{Ft}. The cloning of construct *GLDPA*-Ft-1-2-3-7 involved PCR amplification of the promoter region between −1,571 and −927 using primers *GLDPA*-5′-Xba and gdcs3-R. The PCR product was digested with *Xba*I and ligated with *Xba*I-cut *GLDPA*-Ft-Δ6. The correct orientation of the inserted DNA fragment in *GLDPA*-Ft-1-2-3-7 was shown by sequencing.

To produce construct *GLDPA*-Ft::*H2B:YFP*, the *H2B:YFP* gene fusion was excised from plasmid pBI121-*35S*::*H2B:YFP* (Boisnard-Lorig et al., 2001) with *Bam*HI and *Sac*I. The *GLDPA* promoter of *F. trinervia* was amplified from plasmid *GLDPA*-Ft (Cossu, 1997) with primers *GLDPA*5′-*Hin*dIII and *GLDPA*3′-*Bam*HI. A pBI121 backbone was generated by removing the 35S::*H2B:YFP* insert from plasmid pBI121-*35S*::*H2B:YFP* via incubation with *Hin*dIII and *Sac*I, and a triple ligation between pBI121 (*Hin*dIII/*Sac*I), the *GLDPA* promoter (*Hin*dIII/*Bam*HI), and the *H2B:YFP* fragment (*Bam*HI/*Sac*I) finally resulted in the formation of construct *GLDPA*-Ft::*H2B:YFP*.

The cloning of construct *GLDPA*-Ft::*mGFP5-ER* was achieved by PCR amplification of the *mgfp5-ER* gene (Haseloff et al., 1997) from genomic DNA of the Arabidopsis (*Arabidopsis thaliana*) enhancer trap line E2443 (generated by Scott Poethig, http://www.arabidopsis.org/abrc/poethig.jsp) with primers 5′-mGFP5ER-*Bam*HI and 3′-mGFP5ER-*Sac*I. The PCR product was then cloned into *Bam*HI/*Sac*I-cut *GLDPA*-Ft::*H2B:YFP* to yield *GLDPA*-Ft::*mGFP5-ER*.

Table I. *Oligonucleotide primers used for the construction of promoter fragments and for the synthesis of probes for in situ and northern hybridizations*

Primer Name	Sequence (5′–3′)
GLDPA-5′-Xba	TGCTCTAGAAGCTTTACTCCTCTC
GLDPA-Xma	AAATCCCGGGTAGTGTAAGATGGG
GLDPA-Xba-2	TGCTCTAGATGAAACAGGATGAGC
GLDPA-Xba-3	TGCTCTAGATGAAACAGGATGAGC
GLDPA-Xba-4	TGCTCTAGACAAGTTTGATACTAG
GLDPA-Xba-5	TGCTCTAGACAAGTTTGATACTAG
GLDPA-Xba-6	TGCTCTAGACATTTGATCTATAACG
Eco72-R	TGCTCTAGAGTGGAGATGATAGTTG
gdcs3-R	TGCTCTAGACAAAAGTTCAAAACTTG
gdcsF-4	GAAAGGATCCTCGAATGCTTCATTGC
gdcsR-4	TTCACTCGAGGAATTTAGCAGCACG
*GLDPA*5′-*Hin*dIII	AAGCTTTACTCCTCTCAACT
*GLDPA*3′-*Bam*HI	TTAGGATCCAGTGTAAGATGGGGTCTA
5′-mGFP5ER-*Bam*HI	TTAGGATCCAAGGAGATATAACAATGA
3′-mGFP5ER-*Sac*I	AATGAGCTCTTAAAGCTCATCATGTTT

Plant Transformation

In all transformation experiments, the *Agrobacterium tumefaciens* strain AGL1 was used (Lazo et al., 1991). The promoter-GUS constructs were introduced into AGL1 by electroporation. Arabidopsis plants were transformed via the floral dip method according to Clough and Bent (1998). The transformation of *Flaveria bidentis* was performed as described by Chitty et al. (1994). The integration of the transgenes into the genome of regenerated *F. bidentis* or T_1 Arabidopsis plants was proved by PCR analyses.

Measurement of GUS Activity and Histochemical Analysis of Reporter Gene Activity

F. bidentis T_0 plants used for GUS analysis were 40 to 50 cm tall and before flower initiation; the Arabidopsis T_1 plants were examined around 3 weeks after germination. Fluorometrical quantification of GUS activity was performed according to Jefferson et al. (1987) and Kosugi et al. (1990). For histochemical analysis of GUS activity, leaves were cut manually with a razorblade and the sections were transferred to incubation buffer (100 mM Na_2HPO_4, pH 7.5, 10 mM EDTA, 50 mM $K_4[Fe(CN)_6]$, 50 mM $K_3[Fe(CN)_6]$, 0.1% (v/v) Triton X-100, 2 mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronid acid). After brief vacuum infiltration, the sections were incubated at 37°C for 1 to 20 h. After incubation, chlorophyll was removed from the tissue by treatment with 70% ethanol. Fluorescence microscopy was performed using a Zeiss Axiophot (Carl Zeiss AG) equipped with an Olympus DP50 camera (Olympus Optical) and a Zeiss GFP imaging filter system (BP 450–490, FT 510, BP 515–565). Bright-field and fluorescence images were overlaid with Adobe Photoshop 7.0 (Adobe Systems). For confocal laser microscopy, a Zeiss LSM 510 with a Plan-Neofluar 25× objective was used. YFP fluorescence was monitored with a 505- to 550-nm band pass emission filter (488-nm excitation line). Chlorophyll autofluorescence was visualized with a long pass 560-nm emission filter.

Table II. *Oligonucleotide primer combinations used for the construction of GLDPA promoter fragments*

Construct	Primer 1	Primer 2
GLDPA-Ft-Δ1	*GLDPA*-Xba-2	*GLDPA*-Xma
GLDPA-Ft-Δ3	*GLDPA*-Xba-3	*GLDPA*-Xma
GLDPA-Ft-Δ4	*GLDPA*-Xba-4	*GLDPA*-Xma
GLDPA-Ft-Δ5	*GLDPA*-Xba-5	*GLDPA*-Xma
GLDPA-Ft-Δ6	*GLDPA*-Xba-6	*GLDPA*-Xma

In Situ RNA Hybridization

Nonradioactive in situ hybridization experiments were performed according to the protocol described by Simon (2002). Embedded leaves of *F. bidentis* and *F. trinervia* were cut into cross sections of 20 μm thickness using a standard microtome. For the generation of the *GLDPA* probe, a 2.4-kb cDNA fragment of the *GLDPA* gene of *F. trinervia* was amplified by PCR (primers gdcsF-4 and gdcsR-4; see Table I). The PCR product was digested with *Bam*HI and *Xho*I and cloned into *Bam*HI/*Xho*I-cut pBluescript II KS+ (Stratagene). The use of two different RNA polymerases (T3 and T7) then allowed the production of both antisense and sense probes for in situ hybridization.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Video S1. Histochemical localization of *H2B:YFP* in Arabidopsis plants transformed with the construct *GLDPA*-Ft::*H2B:YFP* by confocal laser microscopy.

Supplemental Video S2. Histochemical localization of *H2B:YFP* in Arabidopsis plants transformed with the construct *GLDPA*-Ft::*H2B:YFP* by confocal laser microscopy.

ACKNOWLEDGMENTS

We thank Smita Kurup for the generous gift of plasmid pBI121-*35S::H2B:YFP* and Scott Poethig for the donation of the Arabidopsis enhancer trap line E2443. We are indebted to Ute Hoecker for carefully reading the manuscript.

Received December 2, 2007; accepted February 17, 2008; published February 27, 2008.

LITERATURE CITED

Akyildiz M, Gowik U, Engelmann S, Koczor M, Streubel M, Westhoff P (2007) Evolution and function of a cis-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. Plant Cell **19:** 3391–3402

Bauwe H, Chu C, Kopriva S, Nan Q (1995) Structure and expression analysis of the *gdcsPA* and *gdcsPB* genes encoding two P-isoproteins of the glycine-cleavage system from *Flaveria pringlei*. Eur J Biochem **234:** 116–124

Bauwe H, Kolukisaoglu U (2003) Genetic manipulation of glycine decarboxylation. J Exp Bot **54:** 1523–1535

Boisnard-Lorig C, Colon-Carmona A, Bauch M, Hodge S, Doerner P, Bancharel E, Dumas C, Haseloff J, Berger F (2001) Dynamic analyses of the expression of the HISTONE:YFP fusion protein in *Arabidopsis* show that syncytial endosperm is divided in mitotic domains. Plant Cell **13:** 495–509

Chitty JA, Furbank RT, Marshall JS, Chen Z, Taylor WC (1994) Genetic transformation of the C4 plant, Flaveria bidentis. Plant J **6:** 949–956

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16:** 735–743

Cossu R (1997) Charakterisierung der Glycindecarboxylase-Gene von *Flaveria trinervia* (C_4) und ihre Expression in transgenen *Nicotiana tabacum*, *Flaveria pubescens* und *Solanum tuberosum*. PhD thesis. Universität Hannover, Hannover, Germany

de Veau EJ, Burris JE (1989) Photorespiratory rates in wheat and maize as determined by O-labeling. Plant Physiol **90:** 500–511

Edwards GE, Furbank RT, Hatch MD, Osmond CB (2001) What does it take to be C_4? Lessons from the evolution of C_4 photosynthesis. Plant Physiol **125:** 46–49

Edwards GE, Ku MSB (1987) Biochemistry of C_3-C_4 intermediates. *In* MD Hatch, NK Boardman, eds, The Biochemistry of Plants, Vol 10. Academic Press, New York, pp 275–325

Furbank RT, Badger MR (1983) Photorespiratory characteristics of isolated bundle sheath strands of C_4 monocotyledons. Aust J Plant Physiol **10:** 451–458

Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, Westhoff P (2004) cis-Regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. Plant Cell **16:** 1077–1090

Haseloff J, Siemering KR, Prasher DC, Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. Proc Natl Acad Sci USA **94:** 2122–2127

Hatch MD (1987) C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim Biophys Acta **895:** 81–106

Hermans J, Westhoff P (1992) Homologous genes for the C4 isoform of phosphoenolpyruvate carboxylase in a C3 and a C4 Flaveria species. Mol Gen Genet **234:** 275–284

Hibberd JM, Quick WP (2002) Characteristics of C4 photosynthesis in stems and petioles of C3 flowering plants. Nature **415:** 451–454

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res **27:** 297–300

Hylton CM, Rawsthorne S, Smith AM, Jones DA, Woolhouse HW (1988) Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. Planta **175:** 452–459

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J **6:** 3901–3907

Kinsman EA, Pyke KA (1998) Bundle sheath cells and cell-specific plastid development in *Arabidopsis* leaves. Development **125:** 1815–1822

Kosugi S, Ohashi Y, Nakajima K, Arai Y (1990) An improved assay for beta-glucuronidase in transformed cells: Methanol almost completely suppresses a putative endogenous beta-glucuronidase activity. Plant Sci **70:** 133–140

Lazo GR, Stein PA, Ludwig RA (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. Biotechnology (N Y) **9:** 963–967

Leegood RC (2002) C(4) photosynthesis: principles of CO(2) concentration and prospects for its introduction into C(3) plants. J Exp Bot **53:** 581–590

Long JJ, Berry JO (1996) Tissue-specific and light-mediated expression of the C_4 photosynthetic NAD-dependent malic enzyme of amaranth mitochondria. Plant Physiol **112:** 473–482

Marshall JS, Stubbs JD, Chitty JA, Surin B, Taylor WC (1997) Expression of the C_4 *Me1* gene from *Flaveria bidentis* requires an interaction between 5′ and 3′ sequences. Plant Cell **9:** 1515–1525

Matsuoka M, Kyozuka J, Shimamoto K, Kano-Murakami Y (1994) The promoters of two carboxylases in a C4 plant (maize) direct cell-specific, light-regulated expression in a C3 plant (rice). Plant J **6:** 311–319

Matsuoka M, Tada Y, Fujimura T, Kano-Murakami Y (1993) Tissue-specific light-regulated expression directed by the promoter of a C4 gene, maize pyruvate,orthophosphate dikinase, in a C3 plant, rice. Proc Natl Acad Sci USA **90:** 9586–9590

Mitchell PL, Sheehy JE (2006) Supercharging rice photosynthesis to increase yield. New Phytol **171:** 688–693

Monson RK, Moore BD, Ku MSB, Edwards GE (1986) Co-function of C_3- and C_4-photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. Planta **168:** 493–502

Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. Science **309:** 1514–1518

Morgan CL, Turner SR, Rawsthorne S (1993) Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and serine hydroxymethyltransferase in leaves of C3-C4 intermediate species from different genera. Planta **190:** 468–473

Neuburger M, Bourguignon J, Douce R (1986) Isolation of a large complex from the matrix of pea leaf mitochondria involved in the rapid transformation of glycine into serine. FEBS Lett **207:** 18–22

Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, Tajima S (2005) Differential expression pattern of C4 bundle sheath expression genes in rice, a C3 plant. Plant Cell Physiol **46:** 754–761

Nomura M, Sentoku N, Nishimura A, Lin JH, Honda C, Taniguchi M, Ishida Y, Ohta S, Komari T, Miyao-Tokutomi M, et al (2000) The evolution of C4 plants: acquisition of cis-regulatory sequences in the promoter of C4-type pyruvate, orthophosphate dikinase gene. Plant J **22:** 211–221

Ohnishi J, Kanai R (1983) Differentiation of photorespiratory activity between mesophyll and bundle sheath cells of C_4 plants. I: Glycine oxidation by mitochondria. Plant Cell Physiol **24:** 1411–1420

Osmond CB, Harris B (1971) Photorespiration during C 4 photosynthesis. Biochim Biophys Acta **234:** 270–282

Parker R, Song H (2004) The enzymes and control of eukaryotic mRNA turnover. Nat Struct Mol Biol **11:** 121–127

Patel M, Siegel AJ, Berry JO (2006) Untranslated regions of FbRbcS1 mRNA mediate bundle sheath cell-specific gene expression in leaves of a C4 plant. J Biol Chem **281:** 25485–25491

Powell AM (1978) Systematics of *Flaveria* (Flaveriinae-Asteraceae). Ann Mo Bot Gard **65:** 590–636

Prestridge DS (1991) SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. Comput Appl Biosci **7:** 203–206

Rawsthorne S (1992) C_3-C_4 photosynthesis. Linking physiology to gene expression. Plant J **2:** 267–274

Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW (1988a) Distribution of photorespiratory enzymes between bundle sheath and mesophyll cells in leaves of the C_3-C_4 intermediate species *Moricandia arvensis* (L.) DC. Planta **176:** 527–532

Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW (1988b) Photorespiratory metabolism and immunogold localisation of photorespiratory enzymes in leaves of C_3 and C_3-C_4 intermediate species of *Moricandia*. Planta **173:** 298–308

Rawsthorne S, Morgan CL, O'Neill CM, Hylton CM, Jones DA, Frean ML (1998) Cellular expression pattern of the glycine decarboxylase P protein in leaves of an intergenic hybrid between the C_3-C_4 intermediate species *Moricandia nitens* and the C_3 species *Brassica napus*. Theor Appl Genet **96:** 922–927

Rosche E, Chitty J, Westhoff P, Taylor WC (1998) Analysis of promoter activity for the gene encoding pyruvate orthophosphate dikinase in stably transformed C_4 *Flaveria* species. Plant Physiol **117:** 821–829

Sage RF (2004) The evolution of C_4 photosynthesis. New Phytol **161:** 341–370

Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schaffner AR, Sheen J (1992) Maize C4 photosynthesis involves differential regulation of phosphoenolpyruvate carboxylase genes. Plant J **2:** 221–232

Sheen J (1999) C4 gene expression. Annu Rev Plant Physiol Plant Mol Biol **50:** 187–217

Siemering KR, Golbik R, Sever R, Haseloff J (1996) Mutations that suppress the thermosensitivity of green fluorescent protein. Curr Biol **6:** 1653–1663

Simon R (2002) Molecular and biochemical analysis of Arabidopsis. *In* DIG Application Manual for Nonradioactive In Situ Hybridization, Ed 3. Roche Applied Science, Mannheim, Germany, pp 197–207

Stockhaus J, Poetsch W, Steinmuller K, Westhoff P (1994) Evolution of the C4 phosphoenolpyruvate carboxylase promoter of the C4 dicot Flaveria trinervia: an expression analysis in the C3 plant tobacco. Mol Gen Genet **245:** 286–293

Stockhaus J, Schlue U, Koczor M, Chitty JA, Taylor WC, Westhoff P (1997) The promoter of the gene encoding the C_4 form of phosphoenolpyruvate carboxylase directs mesophyll-specific expression in transgenic C_4 *Flaveria* spp. Plant Cell **9:** 479–489

van Bel AJE (1993) Strategies of phloem loading. Annu Rev Plant Physiol Plant Mol Biol **44:** 253–281

Westhoff P, Gowik U (2004) Evolution of c4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus Flaveria. Ann Bot (Lond) **93:** 13–23

Yoshimura Y, Kubota F, Ueno O (2004) Structural and biochemical bases of photorespiration in C4 plants: quantification of organelles and glycine decarboxylase. Planta **220:** 307–317

The authors' contributions

SE wrote this manuscript and performed all experiments except those listed below.

CW constructed *GLDPA*-Ft::*H2B:YFP* and *GLDPA*-Ft::*mGFP5-ER*, generated transgenic *Arabidopsis thaliana* plants carrying *GLDPA*-Ft::*H2B:YFP* or *GLDPA*-Ft::*mGFP5-ER*, performed the fluorescence microscopic and confocal microscopic analyses of transgenic *GLDPA*-Ft::*H2B:YFP* and *GLDPA*-Ft::*mGFP5-ER* plants of *A. thaliana* (Figure 3, E–L; Supplemental Videos S1 and S2), and carried out the histochemical GUS staining as well as the fluorometrical quantification of GUS activity in leaves of transgenic *Flaveria bidentis* plants harboring the *GLDPA*-Ft-1-2-3-7 construct (Figure 8).

JB analyzed the histochemical GUS staining in leaf cross sections of transgenic *F. bidentis* plants harboring the *GLDPA*-Ft construct (Figure 2B).

UG analyzed the histochemical GUS staining in guard cells of transgenic *GLDPA*-Ft plants of *F. bidentis* (Figure 2C).

US, **MK** and **MS** performed the transformation of *F. bidentis* plants.

RC and **HB** provided the *GLDPA*-Ft construct.

PW and **UG** participated in drafting of the manuscript.

Manuscript 1 was published in *Plant Physiology* (Impact Factor: 7.016).

Manuscript 2

Regulation of the photorespiratory *GLDPA* gene in C_4 *Flaveria* – an intricate interplay of transcriptional and post-transcriptional processes

Regulation of the photorespiratory *GLDPA* gene in C_4 *Flaveria* – an intricate interplay of transcriptional and post-transcriptional processes

Running title:

GLDPA gene regulation

Authors:

Christian Wiludda[a], Stefanie Schulze[a], Udo Gowik[a], Sascha Engelmann[a], Maria Koczor[a], Monika Streubel[a], Hermann Bauwe[b], and Peter Westhoff[a,1]

[a] Heinrich-Heine-Universität Düsseldorf, Institut für Entwicklungs- und Molekularbiologie der Pflanzen, Universitätsstraße 1, 40225 Düsseldorf, Germany

[b] Universität Rostock, Abteilung Pflanzenphysiologie, Albert-Einstein-Straße 3, 18059 Rostock, Germany

[1] To whom the correspondence should be addressed: Peter Westhoff, [a]Institute of Plant Molecular and Developmental Biology, Universitaetsstrasse 1, Heinrich-Heine-University, 40225 Duesseldorf, Germany. Tel: 49 (0) 211 8112338. Fax: 49 (0) 211 8114871. E-mail: west@uni-duesseldorf.de

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Peter Westhoff (west@uni-duesseldorf.de).

Estimated length of the article: 14.6 pages

ABSTRACT

The mitochondrial glycine decarboxylase complex (GDC) is a key component of the photorespiratory pathway that occurs in all photosynthetically active tissues of C_3 plants, but is restricted to bundle-sheath cells in C_4 species. GDC is also required for the general cellular C_1 metabolism. In the Asteracean C_4 species *Flaveria trinervia* a single functional *GLDP* gene encoding the P-subunit of GDC, *GLDPA*, has been identified. *GLDPA* promoter reporter gene fusion studies revealed that this promoter is active in bundle-sheath cells and the vasculature of transgenic *F. bidentis* (C_4) and the Brassicacean C_3 species *Arabidopsis thaliana* suggesting the existence of an evolutionary conserved gene regulatory system in the bundle-sheath. Here, we demonstrate that *GLDPA* gene regulation is controlled by an intricate interplay of transcriptional and post-transcriptional mechanisms. The *GLDPA* promoter is composed of two tandem promoters, P_{R2} and P_{R7}, that together ensure a strong bundle-sheath expression. While the proximal promoter (P_{R7}) is active in the bundle-sheath and vasculature like the full-length *GLDPA* promoter, the distal promoter (P_{R2}) drives uniform expression in all leaf chlorenchyma cells and the vasculature. Post-transcriptional regulation is based on the inefficient splicing of an intron in the 5' untranslated leader of P_{R2}-derived transcripts appearing to elicit RNA decay.

INTRODUCTION

C_4 photosynthesis is based on the division of labor between two distinct photosynthetically active cell types, mesophyll and bundle-sheath cells. CO_2 is initially prefixed in mesophyll cells by phospho*enol*pyruvate carboxylase and then transported into bundle-sheath cells in the form of malate or aspartate. There CO_2 is released, refixed by ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco), and finally enters the Calvin-Benson cycle as it occurs in C_3 plants. As a bifunctional enzyme Rubisco is able to catalyze the carboxylation as well as the oxygenation of its substrate ribulose 1,5-bisphosphate. When fixing O_2 photorespiration is initiated, a process leading to the loss of previously bound CO_2 and thus decreasing the efficiency of photosynthesis. However, the concentration of CO_2 around Rubisco in the bundle-sheath cells by the C_4 cycle suppresses photorespiration effectively (Ogren, 1984; Hatch, 1987; Leegood et al., 1995; Foyer et al., 2009).

C_4 plants have evolved several times independently from C_3 ancestors, indicating that the conversion from C_3 towards C_4 photosynthesis did not require drastic alterations but could have been implemented rather easily in genetic terms (Sage, 2004; Brown et al., 2011; Gowik

and Westhoff, 2011). It is assumed that a crucial step towards C_4 photosynthesis was the establishment of a functional photorespiratory CO_2 pump, which required the restriction of the glycine decarboxylase complex (GDC) to the bundle-sheath cells (Bauwe and Kolukisaoglu, 2003; Sage, 2004).

GDC is located in the mitochondria and consists of four subunits, the L-, H-, P- and T-protein. Together the four proteins cleave glycine resulting in the release of CO_2, NH_3 and a tetrahydrofolate-bound C_1 residue (Oliver, 1994; Douce et al., 2001). Aside from its involvement in the photorespiratory pathway, GDC also contributes to the C_1 metabolism in all biosynthetic tissues which is essential for the synthesis of proteins, nucleic acids, pantothenates and methylated molecules (Mouillon et al., 1999; Hanson and Roje, 2001).

In C_3 plants, GDC accumulates in all photosynthetically active cells. In contrast, in C_4 plants GDC occurs only in bundle-sheath but not in mesophyll cells, and consequently photorespiratory activity of C_4 plants is restricted to the bundle-sheath. In C_3-C_4 intermediate species which are considered an evolutionary link in the transition from C_3 to C_4 plants, GDC activity has been reported to be already restricted to the bundle-sheath cells (Ohnishi and Kanai, 1983; Rawsthorne et al., 1988; Morgan et al., 1993; Yoshimura et al., 2004). This finding is consistent with the evolutionary scenario that predicts such a relocation of GDC during C_4 evolution (Sage, 2004). The genus *Flaveria* includes C_4, C_3 and several C_3-C_4 intermediate species (Powell, 1978; McKown et al., 2005) and therefore represents a distinguished model system to study molecular mechanisms of the evolutionary transition from C_3 to C_4 photosynthesis (Westhoff and Gowik, 2004; Brown et al., 2005).

The *GLDPA* gene encodes the P-subunit of GDC in the C_4 plant *F. trinervia* (Cossu and Bauwe, 1998). *In situ* hybridization studies showed that *GLDPA* transcripts accumulate only in bundle-sheath cells suggesting that the *GLDPA* gene is specifically transcribed in this tissue (Engelmann et al., 2008). In agreement with this finding, the *GLDPA* 5'-flanking region from -1 to -1571 (with regard to the translational start site at +1; here referred to as the *GLDPA* promoter) directs expression of reporter genes in the bundle-sheath but not in the mesophyll cells. In *F. bidentis* (C_4), the promoter is also active, although to a varying degree, in the vasculature of leaves and roots, stomata and in the pericycle cells of roots (Figure 1). Surprisingly, the *GLDPA* promoter exhibited a similar activity in *A. thaliana* (C_3) (Engelmann et al., 2008; Figure 1) suggesting that the regulatory networks controlling bundle-sheath gene expression are similar in the Brassicacean C_3 species *A. thaliana* and the Asteracean C_4 species *F. bidentis*. Promoter deletion and recombination experiments identified a distal

region that enhanced promoter activity and an intermediate segment that appeared to contain mesophyll-repressing sequences (Engelmann et al., 2008).

In this study we demonstrate that the functional architecture of the *GLDPA* promoter is much more complex than previously thought. We show that the *GLDPA* promoter is in fact composed of two sub-promoters acting in tandem. The proximal promoter (P_{R7}) equivalent to region 7 (-1 to -298; Engelmann et al., 2008) directs reporter gene expression specifically in the bundle-sheath and the vasculature. The distal promoter (P_{R2}) corresponding to region 2 (-1139 to -1389; Engelmann et al., 2008) is active in the vasculature and in all chlorenchyma tissues of the leaf including the mesophyll. When combined, P_{R7} efficiently suppresses the activity of P_{R2}, but only in *F. bidentis*, while in Arabidopsis the full and robust suppression of P_{R2} requires in addition the action of region 3 (-927 to -1138; Engelmann et al., 2008). By rapid amplification of 5' complementary DNA ends (5' RACE) and mining of RNA-seq data (Gowik et al., 2011) we show that the suppression of the mesophyll activity in the *GLDPA* promoter is not complete, allowing also very small amounts of *GLDPA* transcripts starting from P_{R2} to accumulate. We show that the removal of an intron in the 5' untranslated leader of P_{R2}-derived transcripts is essential for generating GLDPA protein whose mitochondrial targeting sequence is truncated but that can be imported into mitochondria nevertheless. We discuss why the bundle-sheath-exclusive expression of a single leaf-specific *GLDP* gene, as predicted and requested by the functional model of C_4 photosynthesis, must be somewhat leaky, but not too much.

RESULTS

Analysis of 5' ends of transcripts of the *GLDPA* gene of *Flaveria trinervia* revealed two independent transcription start sites

Since the transcription start site (TSS) of the *GLDPA* gene of *F. trinervia* had not been determined experimentally yet, rapid amplification of 5' complementary DNA ends (5' RACE; Frohman et al., 1988) was used for mapping 5' ends of *GLDPA* transcripts as present in total leaf extracts. 5' RACE analysis revealed two RNA 5' end classes with one starting in the most proximal region 7 predominantly at nucleotide (nt) -100 upstream of the predicted translational start codon at +1 (ATG_{+1}), while the other one started in the distal region 2 between nucleotides -1185 and -1174 (Figure 2A). Half of the analyzed transcripts starting from region 7 contained a 5' untranslated region (5' UTR_{R7}) of 100 nt. The remaining 5' UTR_{R7}s were slightly shorter with a length between 66 nt and 99 nt. The UTRs of the two

detected 5' ends of RNAs transcribed from region 2 (5' UTR_{R2}s) included parts of region 2 and 3 but lacked regions 4, 5, 6 and 7 as well as the first 17 nt of the predicted *GLDPA* open reading frame (Figure 2A). Fourteen individual and randomly selected 5' RACE products were sequenced. Twelve of them started in region 7 and only two in region 2 (Figure 2A) indicating that the dominant TSS is that one located in region 7.

The comparison of the 5' UTR_{R2}s with the DNA sequence of the *GLDPA* promoter identified the signatures of a spliceosomal intron with two putative GT splice donor sites at -1103 and -1037 respectively within region 3 and a shared AG splice acceptor site at +16 within the open reading frame. If splicing occurs, regardless of which donor site is used, the next available putative start codon at position +25 (ATG_{+25}; Figure 2A) could be used resulting in the shortening of the mitochondrial GLDPA presequence by eight amino acids.

The analysis of the leaf transcriptome of *F. trinervia* by 454 pyrosequencing confirmed the 5' RACE data (Figure 2B; Gowik et al., 2011). The most distal reads detected for the *GLDPA* gene started exactly at position -1185 within region 2. Additionally, the 91-nt 5' UTR_{R2} splicing variant starting at -1185 (Figure 2A) was also found twice by 454 sequencing. In contrast to the low abundance of transcripts in the range from -1185 to -100, the frequency of mRNAs increased at or downstream of position -100 within region 7. This suggests that region 7 is transcriptionally more active than region 2, which is consistent with the results obtained by 5' RACE. We conclude that the *GLDPA* promoter contains two putative TSSs with the major and proximal TSS located in region 7 predominantly at position -100 (TSS_{R7}) and the distal and minor TSS in region 2 around position -1185 (TSS_{R2}).

The proximal and distal transcription start sites of the *GLDPA* gene of *F. trinervia* are functional in transgenic *F. bidentis* and *A. thaliana*

To test whether the two putative TSSs are used in a transgenic promoter-reporter gene context, 5' RACE experiments were performed with transgenic *F. bidentis* and *A. thaliana* both containing the *GLDPA*-Ft:*GUS* chimeric gene (see Supplemental Figure 1 online; Engelmann et al., 2008). In both *F. bidentis* and *A. thaliana*, all RNA 5' ends started between position -90 and -100, i.e. in region 7. Transcripts that originated from region 2 at position -1185 were detected, despite their very low abundance. The 5' UTRs of these mRNAs were not spliced. This is to be expected, because the splice acceptor site that occurs in the *GLDPA* reading frame is not available due to the substitution of the *GLDPA* reading frame by the *GUS* sequence. Taken together, these findings suggest that region 2 and 7 of the *GLDPA* promoter function as separate promoters in transgenic plants of both *A. thaliana* and *F. bidentis*.

Both GLDPA transit peptide variants ensure mitochondrial import

The P-subunit of glycine decarboxylase is located in the mitochondria and hence the GLDPA precursor protein should contain a mitochondrial targeting sequence (presequence) at its very amino terminus (Tanudji et al., 1999; Huang et al., 2009). Analysis of the *GLDPA* coding sequence by UniProtKB predicts a transit peptide of 63 amino acids which is equivalent to 189 nt of the nucleotide sequence (Cossu and Bauwe, 1998). The use of the distal promoter and the removal of the intron shift the putative translational start site to position +25 nt. This would result in a presequence truncated by eight amino terminal amino acid residues. We investigated therefore whether the full size transit peptide of 63 amino acids can target the green fluorescent protein (GFP) to mitochondria and, if so, whether a deletion of the eight amino terminal residues would interfere with a mitochondrial targeting. The two different *GLDPA* presequence variants were fused with the *GFP* reporter gene and the various constructs were transiently expressed in leaf protoplasts of *Nicotiana benthamiana* under the control of the constitutive *cauliflower mosaic virus* (*CaMV*) *35S* promoter (Odell et al., 1985). The distribution of GFP throughout the cell was analyzed by confocal microscopy (Figure 3).

When construct *35S:GLDPA*$_{mt}$-Ft-*mgfp6* containing the full-length *GLDPA* presequence (*GLDPA*$_{mt}$-Ft) fused to *GFP* was analyzed, GFP fluorescence was exclusively detected in the mitochondrial network. As expected, in the absence of any mitochondrial targeting peptide (construct *35S:mgfp6*) GFP was evenly distributed throughout the cytoplasm with no visible association to any cellular organelle. When the truncated transit peptide was investigated (*35S:GLDPA*$_{mt\Delta 24}$-Ft-*mgfp6*), the cellular pattern of GFP fluorescence was indistinguishable from that obtained with the *35S:GLDPA*$_{mt}$-Ft-*mgfp6* construct. The absence of the first eight amino acids from the GLDPA presequence, therefore, did not affect the mitochondrial targeting of the passenger protein. We conclude that both transit peptide variants are capable of targeting the GLDPA protein into mitochondria.

Region 7 of the *GLDPA* promoter directs bundle-sheath- and vasculature-specific gene expression in both *F. bidentis* and *A. thaliana*

The presence of the putative TSS identified in region 7 predominantly at position -100 (TSS_{R7}) and in region 2 at position -1185 (TSS_{R2}) respectively, raise the question whether regions 2 and 7 function as promoters that initiate transcription at these positions. To test the promoter function of region 7, the corresponding segment was fused to *GUS* (construct *GLDPA*-Ft-7; Figure 4A), and the expression pattern of the chimeric gene was analyzed in leaves of both transgenic *F. bidentis* and *A. thaliana* (Figure 4).

Transgenic *GLDPA*-Ft-7 plants of *F. bidentis* exhibited GUS activity only in bundle-sheath cells and vascular bundles (Figures 4B to 4D). This expression pattern was indistinguishable from that of the full-length *GLDPA* promoter (Figure 1; Engelmann et al., 2008). However, the promoter activity of *GLDPA*-Ft-7 was much lower than that of the full-length promoter (Figure 4H; Engelmann et al., 2008).

An almost identical expression pattern was also observed in leaves of transgenic *GLDPA*-Ft-7 plants of *A. thaliana* (Figures 4E to 4G). The two species differed only in the extent of GUS staining within the vasculature which was less in the C_4 plant compared to the C_3 plant. Similarly as in *F. bidentis*, the *GLDPA*-Ft-7 promoter activity in Arabidopsis was very low (Figure 4H).

These findings demonstrate that region 7 is a functional promoter (P_{R7}) that can initiate transcription on its own. Furthermore, P_{R7} directs gene expression specifically in bundle-sheath cells and the vascular bundles like the full-length *GLDPA* promoter.

The 5' UTR in P_{R7}-derived transcripts does not contribute to gene expression specificity

RNAs transcribed from P_{R7} at TSS_{R7} contain a 100-nt long 5' untranslated region (5' $UTR_{R7}100$) which is part of the sub-promoter P_{R7} as defined above. It has been reported that the 5' UTRs of transcripts from C_4 genes may be responsible for the bundle-sheath-specific accumulation of the corresponding RNAs (Patel et al., 2006). To analyze whether the 5' $UTR_{R7}100$ of the *GLDPA* gene contributes to or may be even responsible for the observed bundle-sheath specificity of *GLDPA* expression, the 5' $UTR_{R7}100$ was fused to the *GUS* coding sequence and the transgene was stably expressed in Arabidopsis driven by the *CaMV 35S* promoter (*35S:GLDPA*-Ft-5'$UTR_{R7}100$-*GUS*). Transgenic Arabidopsis plants containing a *35S:GUS* gene served as controls (Figure 5A).

Independently of whether the 5' $UTR_{R7}100$ was inserted between the *35S* promoter and the *GUS* gene or not, the *GUS* gene was expressed in all inner tissues of mature rosette leaves as well as in cotyledons, roots and partially in hypocotyls of young seedlings (Figures 5B to 5I). Thus, the 5' $UTR_{R7}100$ did not alter the expression pattern of the reporter gene. However, transgenic Arabidopsis plants showed a sevenfold higher GUS activity in total leaf extracts, when the 5' $UTR_{R7}100$ was present compared to those plants expressing the 5' $UTR_{R7}100$-less *GUS* variant (Figure 5J). In addition, young seedlings harboring *35S:GLDPA*-Ft-5'$UTR_{R7}100$-*GUS* exhibited much stronger GUS staining within the whole primary root than *35S:GUS* seedlings (Figures 5B and 5F).

These findings indicate that the 5' $UTR_{R7}100$ is not involved in the bundle-sheath specificity of transcript accumulation. It neither destabilizes transcript accumulation in the mesophyll cells, nor does it enhance transcript accumulation in the bundle-sheath cells and the vasculature. We conclude that the bundle-sheath- and vasculature-specific expression of genes driven by P_{R7} is regulated transcriptionally.

Region 2 activates gene expression in the leaf chlorenchyma and vascular tissues in both *F. bidentis* and *A. thaliana*

To investigate the promoter activity contained in region 2 of the *GLDPA* promoter, the sequence of region 2 was fused to the *GUS* reporter gene and the expression pattern and strength of the resulting *GLDPA*-Ft-2*:GUS* gene (construct *GLDPA*-Ft-2; Figure 6A) was analyzed in transgenic *F. bidentis* and *A. thaliana* (Figure 6). Transgenic *GLDPA*-Ft-2 plants of both *F. bidentis* and *A. thaliana* showed an indistinguishable *GUS* expression pattern. Uniform GUS staining was detectable in all inner leaf tissues, namely the chlorenchyma (mesophyll and bundle-sheath cells) and the vascular bundles. No tissue was stained preferentially (Figures 6B to 6E). Region 2 of the *GLDPA* promoter is, therefore, essentially a general leaf promoter (P_{R2}) that functions in both *F. bidentis* and *A. thaliana* with no obvious cell or tissue preference. The promoter activity of region 2 was much stronger than that of region 7 in both *F. bidentis* (~ 400-fold) and *A. thaliana* (~ 1000-fold) (Figures 4H and 6F) reaching almost the promoter activity of the full-length *GLDPA* promoter at least in Arabidopsis (Engelmann et al., 2008).

Region 1 enhances the promoter activities of regions 2 and 7 of the *GLDPA* promoter

Region 1 and 2 together were previously suggested to act as a general transcriptional enhancing module of the *GLDPA* promoter (Engelmann et al., 2008). However, the present findings have revealed that region 2 alone is a strong autonomous promoter. To investigate whether region 1 alone enhances transcriptional activity, it was combined with either the proximal promoter, region 7 (P_{R7}), or the distal promoter, region 2 (P_{R2}), fused to *GUS* and analyzed in transgenic *F. bidentis* and Arabidopsis plants. The activity of P_{R7} or P_{R2} in the presence of region 1 (constructs *GLDPA*-Ft-1-7 and *GLDPA*-Ft-1-2) was then compared with that of the constructs *GLDPA*-Ft-7 and *GLDPA*-Ft-2, that lack region 1 (see Supplemental Figure 2 online).

In both *F. bidentis* and *A. thaliana*, the addition of region 1 to the P_{R2} and P_{R7} promoter segments caused similar effects. In combination with P_{R7}, region 1 enhanced promoter

activity 15- to 18-fold, while the enhancing effect of region 1 on P_{R2} was small to moderate (~ twofold). In both cases, the addition of region 1 did not alter the spatial expression patterns of the attached promoters. Region 1 therefore exhibits only a quantitative enhancing effect, but contains no cell or tissue specificity component.

Region 7 represses the promoter activity of region 2 of the *GLDPA* promoter stably in *F. bidentis*, but only partially in Arabidopsis

The presented data showed that the full-length *GLDPA* promoter functions essentially as a bundle-sheath- and vasculature-specific promoter with minute amounts of transcripts derived from the non-specific distal sub-promoter P_{R2} (region 2). The question therefore arose how this expression pattern could be achieved in view of the fact that the non-specific sub-promoter P_{R2} is about two to three magnitudes stronger than the specific proximal sub-promoter P_{R7} (region 7). Previous experiments had shown that a recombined promoter consisting of regions 1, 2, 3 and 7 in the order given (*GLDPA*-Ft-1-2-3-7) directed an expression pattern that was indistinguishable from that of the full-length promoter (Engelmann et al., 2008). A plausible hypothesis is that the activity of P_{R2} within the *GLDPA* promoter is repressed by the proximal promoter P_{R7} and/or region 3. To identify the component in the *GLDPA* promoter that suppresses its activity in the mesophyll tissue, various combinations of regions 1, 2, 3 and 7 were analyzed with regard to their expression specificities in both *F. bidentis* and *A. thaliana*.

Transgenic *F. bidentis* plants expressing *GLDPA*-Ft-1-2-7:*GUS* (*GLDPA*-Ft-1-2-7; Figure 7A) retained the expression specificity in bundle-sheath cells and the vasculature (Figure 7B) suggesting that, in *F. bidentis*, region 3 is not required for promoter specificity. As expected, omission of region 1 (*GLDPA*-Ft-2-7; Figure 7A) did not affect the spatial GUS staining pattern (Figure 7C). In *F. bidentis*, therefore, the presence of P_{R7} alone suffices to repress the activity of P_{R2}.

In contrast, in *A. thaliana* the absence of region 3 in the promoter constructs caused a loss of bundle-sheath and vasculature specificity. All transgenic *A. thaliana* plants harboring *GLDPA*-Ft-1-2-7 showed *GUS* expression in all inner leaf tissues (Engelmann et al., 2008). Thus, the expression pattern resembles that of P_{R2} alone (Figure 6). In contrast to the stable *GUS* expression pattern of transgenic *GLDPA*-Ft-1-2-7 plants, Arabidopsis lines containing the *GLDPA*-Ft-2-7 construct varied in their *GUS* expression patterns between two extremes, bundle-sheath-/vasculature-specific GUS staining to an expression in all inner leaf tissues (Figure 8). In the presence of region 3 (*GLDPA*-Ft-2-3-7; Figure 9A) all transgenic *A.*

thaliana plants exhibited *GUS* expression in the bundle-sheath and the vasculature (Figures 9C and 9G). In Arabidopsis, therefore, P_{R7} is not sufficient to suppress P_{R2}, but needs region 3 for stable repression. With regard to *GLDPA*-Ft-1-2-7 Arabidopsis plants the presence of region 1 enhances the activity of P_{R2}, and the partially repressive function of P_{R7} is overcome.

In Arabidopsis, region 3 cannot maintain bundle-sheath specificity on its own, but needs the presence of P_{R7}

Since region 3 of the *GLDPA* promoter is absolutely required for the suppression of P_{R2} activity in Arabidopsis, the question arose whether also in Arabidopsis P_{R7} is required for region 3 to be functional. In the natural context of the *GLDPA* promoter region 3 is located 3' to the sub-promoter P_{R2} followed further downstream by P_{R7}. We wanted to know therefore whether the position of region 3 with respect to P_{R2} is important for its suppressing activity and whether region 3 can also influence the activity of P_{R7}.

To investigate whether region 3 alone could repress P_{R2} activity, it was fused downstream (*GLDPA*-Ft-2-3) as well as upstream (*GLDPA*-Ft-3-2) of P_{R2} (Figure 9A). In both cases all transgenic *A. thaliana* plants exhibited the same uniform *GUS* expression pattern in the chlorenchyma and vasculature as detected for P_{R2} alone (Figures 9D, 9E, 9H and 9I). In contrast, the combination of P_{R2}, region 3 and P_{R7} (*GLDPA*-Ft-2-3-7; Figure 9A) caused specific *GUS* expression in the bundle-sheath and the vasculature (Figures 9C and 9G). Therefore, region 3 alone is not sufficient to suppress P_{R2}, independent of its location down- or upstream of P_{R2}, but the presence of P_{R7}, in addition, is necessary.

To examine whether region 3 could affect also P_{R7} activity, the corresponding transgene (*GLDPA*-Ft-3-7; Figure 9A) was constructed and analyzed in transgenic Arabidopsis. As expected the combination of region 3 and P_{R7} led to the same bundle-sheath and vasculature-specific expression pattern as P_{R7} alone (Figures 9B and 9F), indicating that region 3 has no influence on the spatial activity of P_{R7}.

However, the *GLDPA*-Ft-3-7 construct was 20-fold more active than *GLDPA*-Ft-7 (Figures 4H and 9J). Interestingly, region 3 was also able to enhance P_{R2} activity (compare constructs *GLDPA*-Ft-3-2 [Figure 9J] and *GLDPA*-Ft-2 [Figure 6F]), although to a much less degree than observed for P_{R7} (twofold vs. 20-fold). Thus, region 3 can enhance transcription of P_{R7} and P_{R2} with a comparable strength as detected for region 1.

DISCUSSION

The commonly believed evolutionary scenario of C_4 photosynthesis predicts that relatively early along the path towards C_4 photosynthesis a photorespiratory CO_2 pump was established by compartmentalization of glycine decarboxylase activity in the bundle-sheath (Sage, 2004; Bauwe, 2011). All available experimental data confirm the final outcome of this evolutionary process, namely that in present C_4 species glycine decarboxylase accumulates exclusively in the bundle-sheath (Majeran et al., 2010; Li et al., 2010). Along these lines the activity of the *GLDPA* promoter of the Asteracean C_4 species *F. trinervia* was found to be restricted to the bundle-sheath cells and vasculature in leaves of both transgenic *F. bidentis* (C_4) and *A. thaliana* (C_3) (Engelmann et al., 2008). This suggested that the bundle-sheath-specific accumulation of *GLDPA* mRNAs should be essentially controlled by transcription (Engelmann et al., 2008). Data presented in this paper indicate that *GLDPA* gene regulation is much more complex than previously thought. We discuss and provide evidence that the mRNA output of this gene is controlled by a combination of transcriptional and post-transcriptional means.

The *GLDPA* promoter consists of two tandem promoters

The *GLDPA* promoter is composed of two tandem promoters, P_{R2} and P_{R7}, that together ensure a strong bundle-sheath expression. The two sub-promoters are not easily recognized by inspection of their corresponding nucleotide sequences. No reliable candidates for TATA boxes can be detected in the predicted distance of 25 to 40 bp upstream of the two transcriptional initiation sites (Joshi, 1987; Bernard et al., 2010; Zuo and Li, 2011). This may not surprise since, for instance in *A. thaliana*, only 20 to 30% of all promoters contain a TATA box/variant (Molina and Grotewold, 2005; Yamamoto et al., 2009; Bernard et al., 2010). Recently, TC-elements have been proposed as a novel class of regulatory elements that control transcription in plants (Bernard et al., 2010). Indeed, several TC-elements are located around TSS_{R2} but predominantly around TSS_{R7} within the predicted range of 50 bp up- and/or downstream of the corresponding TSS (see Supplemental Figure 3 online; Zuo and Li, 2011). The possible importance of TC-elements for transcriptional regulation of P_{R7} is supported by the fact that the motifs CCCTTT, CCTTCT and TCTTCT are unique to region 7 within the *GLDPA* promoter and that TCTTCT even belongs to the three TC-elements most frequently observed (Bernard et al., 2010). TSS_{R2} is flanked by a sequence repeat that is very similar to the predicted Initiator (Inr) motif shown to be essential for the light-dependent activity of the *psaDb* promoter from *Nicotiana sylvestris* (Nakamura et al., 2002; see Supplemental Figure 3

online). According to the YR rule (YR, Y = C or T, R = A or G, TSS underlined) most of the Arabidopsis promoters contain the CA dimer sequence around their TSSs (Yamamoto et al., 2007), which is also true for TSS_{R2}. Yamamoto et al. (2007) consider this YR rule to represent a less stringent form of Inr. This indicates that Inr elements might be crucial for transcriptional activity of P_{R2}.

The *GLDPA* sub-promoters diverge in their specificities

When analyzed separately, the two sub-promoters diverge in their spatial expression profiles. The proximal promoter (P_{R7}), defined by region 7, is relatively weak and specifically active in the bundle-sheath and the vasculature like the full-length *GLDPA* promoter (Figure 4). The distal promoter (P_{R2}), defined by region 2, is strong and drives expression in all inner leaf cells including the mesophyll (Figure 6). In contrast, in the context of the full-length *GLDPA* promoter the final RNA output from both promoters, as measured by 5' RACE and RNA sequencing experiments, is just the reverse. RNAs transcribed from the proximal promoter dominate the *GLDPA* transcript population, and RNAs derived from P_{R2} are in the minority (Figure 2). When the two sub-promoters, i.e. regions 2 and 7, are combined and fused to the *GUS* reporter gene, the read-out of this chimeric gene in *F. bidentis* is indistinguishable from that of the complete *GLDPA* promoter (cf. Figures 1 and 7). The finding suggests that the proximal bundle-sheath-specific promoter P_{R7} turns off the activity of the unspecific distal promoter P_{R2}. How could a downstream promoter interfere with the activity of an upstream promoter and even disable it?

That a strong upstream promoter can shut-off the activity of a downstream promoter is well documented (Mazo et al., 2007). This phenomenon is called transcriptional interference and may be defined as “the in *cis* suppression of one transcriptional process by another”. Transcriptional interference has been documented as a general regulatory process affecting the transcription from adjacent convergent or tandem promoters (Palmer et al., 2011). Different ways are considered of how promoters could impede or even block one another. One possibility is the transcription from a strong regulatory promoter that might impair the recruitment of the transcription initiation complex or the transcriptional elongation of a neighboring target gene (Mazo et al., 2007; Palmer et al., 2011).

The regulatory interactions between the yeast *SRG1* gene that encodes a non-coding RNA and the downstream target gene *SER3* encoding an enzyme of the serine biosynthesis pathway may serve as the best investigated example of transcriptional interference. Both genes are arranged in tandem. Transcription of the non-coding *SRG1* RNA suppresses the adjacent

SER3 gene by inhibiting the binding of transcriptional activators to the *SER3* promoter (Martens et al., 2004). The binding of these activators is prevented because *SRG1* transcription leads to random positioning of nucleosomes that occlude the *SER3* promoter and thereby block *SER3* transcription (Hainer et al., 2011; Thebault et al., 2011). Transcriptional interference has also been reported for plants. The strong *35S* promoter of a T-DNA inserted upstream of the *AtRibA1* gene results in large transcripts that run over the *AtRibA1* promoter and thereby inhibit *AtRibA1* transcription (Hedtke and Grimm, 2009).

In all known cases of transcriptional interference occurring with tandem promoters, the upstream promoter blocks the activity of the downstream promoter. However, with respect to the *GLDPA* promoter just the opposite is true: the downstream promoter P_{R7} inhibits the output from the upstream promoter P_{R2}. Could a roadblock mechanism explain the transcriptional interference between the *GLDPA* sub-promoters, i.e. a pausing RNA polymerase II (Levine, 2011) or a scaffold of general transcription factors (Yudkovsky et al., 2000) residing at the downstream promoter inhibit the progress of RNA polymerase II from the upstream promoter? There is increasing evidence that RNA polymerase II pauses quite often after having initiated transcription and having produced a nascent transcript of about 30 to 50 nucleotides (Levine, 2011). In human lung fibroblasts 30% of all protein-coding genes carry a paused RNA polymerase (Core et al., 2008), and in the early *Drosophila* embryo the stalling of RNA polymerase II in promoter-proximal regions occurs in hundreds of genes that are regulated by environmental or developmental stimuli (Muse et al., 2007; Zeitlinger et al., 2007). It is conceivable therefore that a RNA polymerase II pausing at P_{R7} might represent a roadblock that impedes the elongation of transcripts originating from P_{R2} and leads to the suppression of the P_{R2} output.

The *GLDPA* 5' flanking region – a player in post-transcriptional control

One hallmark of the *GLDPA* gene of *F. trinervia* is an intron located within the 5' UTR of P_{R2}-derived transcripts. Depending on the splice donor site used the intron commences 84 or 139 nucleotides behind the respective transcription start site of the distal promoter P_{R2} within region 3 and ends 17 nucleotides behind the first nucleotide of the *GLDPA* reading frame (Figure 2). RNA sequencing experiments using the 454 technology (Gowik et al., 2011) revealed that this 5' intron is present in *GLDPA* transcripts, and the sequence reads cover the intron region uniformly (see Supplemental Figure 4 online). In contrast, sequence reads from the gene-internal introns are not detectable. The accumulation of unspliced P_{R2}-derived

transcripts with respect to their 5' UTR suggests that the splicing efficiency of the 5' intron is drastically lower than that of the gene-internal ones.

To prevent the accumulation of aberrant mRNAs, i.e. mRNAs that are erroneously or not completely spliced, eukaryotes have developed various quality control systems (Egecioglu and Chanfreau, 2011). Spliceosomal DExD/H box ATPases provide the first layer of defense. They act as kinetic proofreading systems and limit the escape of unspliced or erroneously spliced RNAs from the spliceosome (Egecioglu and Chanfreau, 2011). Despite the accuracy of these proofreading activities unspliced mRNAs may escape detection, and therefore external quality control systems have been built up. The nuclear exosome takes part in the degradation of unspliced RNAs (Houseley et al., 2006; Fasken and Corbett, 2009), although the molecular mechanisms by which unspliced RNAs are recognized are not yet clear. Exon-junction complexes that are deposited on spliced RNAs might be involved in the recognition mechanism (Egecioglu and Chanfreau, 2011). These protein complexes play also a prominent role in nonsense-mediated mRNA decay. Nonsense-mediated mRNA decay is a eukaryotic mRNA surveillance mechanism that detects and degrades mRNAs containing premature termination codons (Chang et al., 2007; Brogna and Wen, 2009). In plants, long 3' UTRs, introns that are located in 3' UTRs and upstream open reading frames (uORFs) within the 5' UTR can trigger nonsense-mediated mRNA decay (Kertész et al., 2006; Hori and Watanabe, 2007; Nyikó et al., 2009).

The 5' intron sequence that is present in unspliced transcripts derived from P_{R2} contains several uORFs. The one that starts directly upstream of TSS_{R7} encodes more than 35 amino acids (see Supplemental Figure 5 online). uORFs are considered to have the potential to elicit the nonsense-mediated mRNA decay response when they give rise to proteins that are larger than the critical threshold of 35 amino acids (Nyikó et al., 2009). When P_{R2} is combined with P_{R7} (*GLDPA*-Ft-2-7), a 135 nucleotides long uORF commencing directly upstream of TSS_{R7} might encode a protein of 45 amino acids that should promote the nonsense-mediated mRNA decay (see Supplemental Figure 6 online). This 135 nucleotides long uORF is only present in transcripts originating from P_{R2}. 5' RACE analyses of transgenic *GLDPA*-Ft-2-7 plants of *A. thaliana* and *F. bidentis* showed that transcription started from either TSS_{R2} (P_{R2}) or TSS_{R7} (P_{R7}). While P_{R7}-derived RNAs exhibited more or less the same length, P_{R2}-derived transcripts appeared to be destabilized, because many RNAs started randomly between TSS_{R2} and TSS_{R7} indicating RNA degradation (see Supplemental Figure 6 online). Taken together, the presented data strongly suggest that the mRNA output of the *GLDPA* gene of *F. trinervia* is controlled by an intricate interplay of transcriptional and post-transcriptional mechanisms.

Evolution of the *GLDPA* promoter – the necessity of being bundle-sheath-specific, but not completely

The presence of the 5' intron in P_{R2}-derived transcripts and of an alternative ATG codon 25 nucleotides behind the major translational start site result in a GLDPA protein variant whose mitochondrial targeting peptide is truncated by eight amino acids. Our import experiments showed that nevertheless the truncated transit peptide is capable of directing an attached passenger protein to the mitochondria (Figure 3). This indicates that P_{R2}- as well as P_{R7}-derived mRNAs yield a GLDPA protein that accumulates in the mitochondria. One wonders why the *GLDPA* promoter contains one sub-promoter with the desired specificity in the bundle-sheath and a second sub-promoter that is active in all internal leaf cells including the mesophyll. Moreover, the second, non-specific promoter is not allowed to express its full potential, but is almost, even though not completely, switched off by a combination of transcriptional and post-transcriptional means.

Mutational analysis with *Arabidopsis thaliana* revealed that a *GLDP* double mutant in which both of the two *GLDP* genes were knocked out is lethal, even under nonphotorespiratory conditions (Engel et al., 2007). This indicates that the activity of GDC is indispensable and that all biosynthetically active cells need glycine decarboxylase activity for one-carbon metabolism (Hanson and Roje, 2001). According to 454 pyrosequencing data, in C_4 *Flaveria* species the *GLDPA* gene is the only active leaf *GLDP* gene (Gowik et al., 2011). We hypothesize therefore that the *GLDPA* gene of *F. trinervia* (C_4) must fulfill two purposes: firstly, it has to serve the requirements of the photorespiratory pathway and its activity should therefore be restricted to the bundle-sheath cells; secondly, however, small amounts of GDC activity are needed in all biosynthetically active cells – also in the mesophyll cells, and therefore the regulatory system of the *GLDPA* gene has to be somewhat leaky (Figure 10). We have provided conclusive evidence that an intricate combination of transcriptional and post-transcriptional control ensures small amounts of *GLDPA* mRNAs to accumulate in the mesophyll cells of C_4 *Flaveria* species. It remains to be investigated, how this pattern of gene expression control evolved in the genus *Flaveria*. These studies are underway and should elucidate the adaptive changes in gene expression that are a central component of C_4 evolution.

METHODS

Generation of chimeric promoters and cloning of promoter-reporter gene constructs

The amplification and cloning procedure of DNA was accomplished according to Sambrook and Russell (2001). The dissection of the *GLDPA* promoter from *Flaveria trinervia* (GenBank accession number Z99767) into seven regions and the cloning of the constructs *GLDPA*-Ft-7 (previously referred to as *GLDPA*-Ft-Δ6), *GLDPA*-Ft-1-2-7 and *GLDPA*-Ft:*H2B-YFP* (previously referred to as *GLDPA*-Ft::*H2B:YFP*) have been described in Engelmann et al. (2008). All *GLDPA* promoter regions were amplified by PCR by means of the Phusion High-Fidelity DNA Polymerase (New England Biolabs) or the *Pfu* DNA Polymerase (Stratagene) using the *GLDPA*-Ft construct (Engelmann et al., 2008; Cossu, 1997) as template and the corresponding oligonucleotides containing respective restriction sites (Tables 1 and 2). For *GLDPA*-Ft-2-3-7 the regions 2-3 as *Xba*I/*Bcu*I and 7 as *Bcu*I/*Xma*I fragment were cloned together into the *Xba*I/*Xma*I-digested binary plant transformation vector pBI121 (Clontech Laboratories; Chen et al., 2003; Jefferson et al., 1987) lacking the *CaMV 35S* promoter. *GLDPA*-Ft-2-7 and *GLDPA*-Ft-3-7 were constructed by exchanging region 2-3 of *GLDPA*-Ft-2-3-7 with region 2 and 3 as *Xba*I/*Bcu*I fragments respectively. For cloning of *GLDPA*-Ft-2 and *GLDPA*-Ft-2-3 the regions 2 and 2-3 were ligated respectively as *Xba*I/*Xma*I fragment into the *Xba*I/*Xma*I-restricted pBI121 vector lacking the *35S* promoter. Region 3 was cloned as *Xba*I/*Xba*I fragment into *GLDPA*-Ft-2 prior cut with *Xba*I to generate *GLDPA*-Ft-3-2. The correct orientation of region 3 was proven by DNA sequencing. *35S:GUS* was constructed by inserting the *35S* promoter amplified from pBI121-*35S::H2B:YFP* (Boisnard-Lorig et al., 2001) as *Hind*III/*Xba*I-*Xma*I fragment into the *Hind*III/*Xma*I-restricted pBI121 vector, whereas for *35S:GLDPA*-Ft-5'UTR$_{R7}$100-*GUS* the *35S* promoter was directly excised from pBI121-*35S::H2B:YFP* as *Hind*III/*Xba*I fragment to ligate it together with the *Xba*I/*Xma*I-digested 100-bp 5' UTR of region 7 (5' UTR$_{R7}$100) into the *Hind*III/*Xma*I-restricted pBI121 vector. For *35S:mgfp6*, *35S:GLDPA*$_{mt}$-Ft-*mgfp6* and *35S:GLDPA*$_{mt\Delta 24}$-Ft-*mgfp6* the Gateway Technology (Invitrogen) was applied starting with the generation of Gateway compatible recombination fragments by PCR: Regarding *35S:mgfp6* the primers ATG-5'-*att*B1 and ATG-3'-*att*B2 (Tab. I) were used for simple primer dimer formation, elongation and amplification generating a start codon (ATG) with flanking *att*B sites, whereas for *35S:GLDPA*$_{mt}$-Ft-*mgfp6* and *35S:GLDPA*$_{mt\Delta 24}$-Ft-*mgfp6 F. trinervia* cDNA was used for the amplification of the full-length N-terminal presequence of the *GLDPA* gene (*GLDPA*$_{mt}$, 189 bp, primers: *GLDPA*$_{mt}$-5'-*att*B1/*GLDPA*$_{mt}$-3'-*att*B2 and *att*B1/*att*B2 adapter) as annotated by GenBank (accession number Z99767) or a shorter version lacking the first 24 bp (*GLDPA*$_{mt\Delta 24}$, 165 bp, primers:

GLDPA$_{mt\Delta 24}$-5'-*att*B1/*GLDPA*$_{mt}$-3'-*att*B2 and *att*B1/*att*B2 adapter). The *att*B-flanked PCR products were recombined into pDONR^{TM}221 (Invitrogen) and afterwards into pMDC83 (Curtis and Grossniklaus, 2003) respectively. All generated constructs were verified by DNA sequencing.

Transformation of *Arabidopsis thaliana* and *Flaveria bidentis*

All chimeric promoter-reporter gene constructs were transformed into either the *Agrobacterium tumefaciens* strain AGL1 (Lazo et al., 1991) or GV3101 (pMP90) (Koncz and Schell, 1986) via electroporation. Transgenic *Flaveria bidentis* were generated according to Chitty et al. (1994) by means of *A. tumefaciens* AGL1 containing the respective construct. *Arabidopsis thaliana* (Ecotype Columbia) was transformed by the floral dip method (Clough and Bent, 1998) modified according to Logemann et al. (2006) by using the *A. tumefaciens* strain GV3101 harboring the appropriate construct. The presence of the respective transgene within the genome of each single independent *F. bidentis* T_0 and *A. thaliana* T_1 line was verified by PCR after DNA isolation as described by Edwards et al. (1991).

***In situ* analysis of the β-glucuronidase and detection of its activity**

The fifth leaf from the top of 40–50 cm tall transgenic *F. bidentis* T_0 plants or 3 mature rosette leaves of 3–4 weeks old transgenic *A. thaliana* T_1 plants prior to flowering were used respectively for the fluorometrical quantification of the β-glucuronidase (GUS) activity according to Jefferson et al. (1987) and Kosugi et al. (1990). The statistical significance of the difference between two data sets was analyzed by means of the Mann-Whitney U test (*http://elegans.swmed.edu/~leon/stats/utest.html*). Histochemical GUS analyses were performed as described by Engelmann et al. (2008). The fifth leaf from the top of transgenic *F. bidentis* T_0 plants (40–50 cm) and single rosette leaves as whole blades or manually cut cross sections as well as young seedlings of transgenic T_1 *A. thaliana* plants were used for the histochemical GUS analysis *in situ* respectively.

Transient gene expression in leaves of *Nicotiana benthamiana* and isolation of protoplasts

The *Agrobacterium*-mediated infiltration of leaves of *Nicotiana benthamiana* was performed according to Waadt and Kudla (2008) by means of *A. tumefaciens* GV3101 (pMP90) containing the respective construct and the *A. tumefaciens* strain p19 for suppression of gene

silencing (Voinnet at al., 2003). After four days two infiltrated leaves per plant were harvested respectively for protoplast isolation. Four leaf pieces (approximately 0.7 x 0.7 cm each) per blade were cut out with a razor blade, transferred into 5 ml enzyme solution (Yoo et al., 2007), vacuum-infiltrated three times for 30 sec and then incubated at room temperature for 2 h. After light shaking to release protoplasts, remaining leaf pieces were removed and MitoTracker Orange CMTMRos (Invitrogen) was added to a final concentration of 150 nM to the suspension of protoplasts for labeling mitochondria. After incubation for 15 min at 37 °C protoplasts were centrifuged at 500*g* for 1 min, the supernatant was removed, and the sedimented leaf cells were resuspended in 100 µl W5 solution (Yoo et al., 2007) for analysis by confocal microscopy.

Confocal laser scanning microscopy and fluorescence microscopy

The analysis of protoplasts by confocal laser scanning microscopy was performed with the LSM 510 (Carl Zeiss AG). Protoplasts were excited at 488 nm (for detection of GFP and chlorophyll fluorescence) and at 561 nm (for detection of MitoTracker fluorescence), respectively. To visualize specifically GFP fluorescence, a 505–550-nm band pass emission filter was used. The fluorescence of MitoTracker Orange CMTMRos-labeled mitochondria was observed by utilizing a 575–615-nm band pass filter, and the autofluorescence of chlorophyll was recorded by means of a 650-nm long pass filter.

Leaf cross sections, complete leaf blades and roots of transgenic *F. bidentis* as well as roots of transgenic *A. thaliana* carrying the *GLDPA*-Ft:*H2B-YFP* chimeric gene were analyzed with the aid of an Axiophot fluorescence microscope (Carl Zeiss AG) that was equipped with an integrated HBO-UV lamp (Carl Zeiss AG) and a DP50-CU camera (Olympus Optical Co.) by using the filter set F41-028 (excitation: HQ 500/20, beam splitter: Q 515 LP, emission: HQ 535/30; AHF Analysentechnik). Bright-field and fluorescent pictures were merged with Adobe Photoshop 7.0 (Adobe Systems). Prior to fluorescence microscopy leaf blades were extracted with 95% ethanol according to Zhou et al. (2005). Leaf blades were harvested from 30 cm tall *F. bidentis* grown in the greenhouse. Roots of young seedlings of Arabidopsis and *F. bidentis* were taken from plants cultivated on agar medium in a climate chamber.

Analysis of mRNA 5' ends by the rapid amplification of 5' complementary DNA ends (5' RACE)

Total RNA from leaves of *F. trinervia* was isolated according to Westhoff et al. (1991). After enrichment by the Oligotex mRNA Midi Kit (Qiagen) 0.5 µg of poly A^+ mRNA was used for cDNA first-strand synthesis performed with the SMART™ RACE cDNA Amplification Kit (Clontech Laboratories) and the PowerScript™ Reverse Transcriptase (Clontech Laboratories) according to the manufacturers' protocols. For PCR amplification of 5' UTRs with the Phusion High-Fidelity DNA Polymerase (New England Biolabs) the gene-specific 3' oligonucleotide *GLDPA*-RACE4 (5'-GAGATCTTGGACTTGTACTGTC-3') and the SMART-II-A-Primer (5'-AAGCAGTGGTATCAACGCAGAGT-3') were used. The PCR fragment was cloned subsequently using the CloneJET™ PCR Cloning Kit (Fermentas). 60 independent clones were analyzed by colony PCR using the SMART-II-A-Primer and the gene-specific *GLDPA*-RACE6 oligonucleotide (5'-ACACCGTACATAGCAGCCATG-3'). These PCR products were verified by restriction endonuclease analyses leading to the identification of 51 potentially correct clones. Plasmid DNA was isolated from 14 of them for DNA sequencing.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Analysis of mRNA 5' ends in leaves of transgenic *A. thaliana* and *F. bidentis* harboring the *GLDPA*-Ft:*GUS* transgene.

Supplemental Figure 2. Functional analysis of region 1 of the *GLDPA* promoter in transgenic *A. thaliana* and *F. bidentis*.

Supplemental Figure 3. Distribution of TC-rich elements and Initiator-like motifs within region 2 and region 7 of the *GLDPA* promoter.

Supplemental Figure 4. Splicing pattern of the *GLDPA* transcript analyzed by 454 sequencing.

Supplemental Figure 5. Distribution of upstream open reading frames in P_{R2}-derived transcripts.

Supplemental Figure 6. Analysis of mRNA 5' ends in leaves of transgenic *A. thaliana* and *F. bidentis* harboring *GLDPA*-Ft-2-7.

Supplemental Methods

ACKNOWLEDGEMENTS

This work was supported by the "Sonderforschungsbereich 590 – Inhärente und adaptive Differenzierungsprozesse" and the "Forschergruppe 1186" funded by the Deutsche Forschungsgemeinschaft.

AUTHOR CONTRIBUTIONS

C.W., S.S., U.G., H.B., and P.W. designed the research. C.W., S.S., U.G., S.E., M.K., and M.S. performed the research. C.W., S.S., U.G., and P.W. analyzed data. C.W., U.G., and P.W. wrote the paper.

FIGURE LEGENDS

Figure 1. Fluorescence microscopic analysis of *GLDPA* promoter activity in transgenic *F. bidentis* and *A. thaliana*.

(A) Schematic presentation of the *GLDPA*-Ft:*H2B*-*YFP* construct which was transformed into *F. bidentis* or *A. thaliana* to express the yellow fluorescent protein (YFP) fused to histone 2B (H2B) under the control of the *GLDPA* promoter.

(B) to **(H)** The localization of YFP was examined by fluorescence microscopy in longitudinal **(B)** and cross sections **(C)** of leaves, whole leaf blades in top view **(D)**, guard cells of both the upper **(E)** and lower epidermis **(F)** and in roots (**[G]** and **[H]**). The fluorescence image is displayed underneath and the corresponding merge of the fluorescent signal and the bright-field picture above respectively. In the root single endodermis (EN) and pericycle (PE) cells are depicted by an arrowhead.

Figure 2. Analysis of transcript 5' ends of the endogenous *GLDPA* gene of *F. trinervia* by rapid amplification of 5' complementary DNA ends (5' RACE) (A) and 454 pyrosequencing (B).

(A) The *GLDPA* promoter and its transcriptional output. The dissection of the *GLDPA* promoter into 7 regions has been described in Engelmann et al. (2008). The schematic structure of the 5' untranslated regions (5' UTRs) of the two types of RNAs originating from region 2 (5' UTR_{R2}s) or region 7 (5' UTR_{R7}s) and their corresponding cDNA sequences are depicted below the schematic drawing of the *GLDPA* promoter. The transcription start sites (TSSs) within region 2 (TSS_{R2}) and 7 (TSS_{R7}) are indicated as well as the start codons used

when transcription starts from region 2 (ATG_{+25}) or region 7 (ATG_{+1}) and the number (No.) of the 5' UTRs detected for each 5' UTR variant whose length is stated in nucleotides (nt).
(B) Diagram showing the read coverage of the *GLDPA* contig derived from 454 sequencing reads (Gowik et al., 2011). The coverage upstream of the translational start site (ATG_{+1}) up to 100 nt downstream was analyzed in 50 nt windows. A contig corresponding to the 91-nt spliced variant starting from TSS_{R2} that was detected by 5' RACE **(A)** was represented by only two 454 reads. The transcription start sites (TSS_{R2} and TSS_{R7}) and the start codons (ATG_{+1} and ATG_{+25}) are marked by arrowheads. The different *GLDPA* promoter regions 2 to 7 shown as columns are allocated to their respective positions.

Figure 3. Localization study of the two different transit peptide variants of the GLDPA protein.
The structures of the three constructs used for transient expression in leaves of *Nicotiana benthamiana* are diagrammed on top of the figure. *35S:GLDPA*$_{mt}$-Ft-*mgfp6* contains the full-length *GLDPA* sequence encoding the predicted presequence for mitochondrial targeting (*GLDPA*$_{mt}$-Ft). In *35S:GLDPA*$_{mt\Delta 24}$-Ft-*mgfp6* the transit peptide lacks the eight amino terminal residues. *35S:mgfp6* is devoid of any transit peptide sequence and served as a control. For visualizing mitochondria MitoTracker staining was carried out (+MT) or omitted as negative control (-MT). Three different channels were utilized to separate the fluorescence signals of MitoTracker-labeled mitochondria (magenta color), GFP (green color) and chlorophyll of chloroplasts (blue color) from each other. When merging MitoTracker and GFP fluorescences (M + GFP), white color indicates overlapping of both signals. All three different fluorescence signals are merged in the last column (M + GFP + CP). C, chloroplasts; GFP, green fluorescent protein; M, mitochondria; MT, MitoTracker; WT, wild type.

Figure 4. Functional analysis of region 7 of the *GLDPA* promoter in leaves of transgenic *F. bidentis* and *A. thaliana*.
(A) Schematic presentation of the *GLDPA*-Ft-7 construct. TSS, transcription start site; 5' UTR_{R7}100, 100-bp 5' untranslated region of *GLDPA* region 7.
(B) to **(G)** Histochemical localization of GUS activity in cross sections (**[B]**, **[C]**, **[E]** and **[F]**) and leaf blades in top view (**[D]** and **[G]**) of leaves of transgenic *F. bidentis* and *A. thaliana*. Single bundle-sheath cells are emphasized by arrowheads. Incubation times for the GUS staining were 17 h (**[E]** and **[F]**), 29 h **(G)**, 43 h **(B)**, 66 h **(C)** and 70.5 h **(D)**.

(H) Fluorometrical quantification of GUS activities of transgenic *F. bidentis* and *A. thaliana* plants transformed with the *GLDPA*-Ft-7 construct. Each single dot represents one independent transgenic line. The number of lines examined (n) is indicated above as well as the median of all values ($\tilde{x}$), also displayed as black line in the diagram. MU, 4-methylumbelliferone.

Figure 5. Analysis of the gene-regulatory properties of the 100-bp 5' untranslated region of *GLDPA* region 7 (5' UTR$_{R7}$100) in transgenic *A. thaliana*.

(A) Schematic structure of the two constructs used for transformation. *35S:GUS* consists of the *CaMV 35S* promoter and the *GUS* reporter gene, while *35S:GLDPA*-Ft-5'UTR$_{R7}$100-*GUS* additionally contains the 100-bp long 5' UTR$_{R7}$. TSS, transcription start site.

(B) to **(I)** Histochemical GUS staining of *A. thaliana* transformed with *35S:GUS* or *35S:GLDPA*-Ft-5'UTR$_{R7}$100-*GUS* in seedlings (**[B]** and **[F]**), young leaf blades (**[C]** and **[G]**) and cross sections of mature rosette leaves (**[D]**, **[E]**, **[H]** and **[I]**). Staining was for 1 h (**[H]** and **[I]**), 3 h (**[C]** and **[G]**), 4 h (**[B]** and **[F]**) or 16 h (**[D]** and **[E]**).

(J) Quantitative measurements of expression strength by analyzing GUS activities in leaf extracts of transgenic Arabidopsis plants transformed with the *35S:GUS* or *35S:GLDPA*-Ft-5'UTR$_{R7}$100-*GUS* construct. Each single dot represents one independent transgenic *A. thaliana* line. The number of lines examined (n) is indicated above as well as the median of all values ($\tilde{x}$) which is additionally charted as black line in the diagram (*** $p < 0.001$). MU, 4-methylumbelliferone.

Figure 6. Functional analysis of region 2 of the *GLDPA* promoter in leaves of transgenic *F. bidentis* and *A. thaliana*.

(A) Schematic structure of the *GLDPA*-Ft-2 construct. TSS, transcription start site; 5' UTR$_{R2}$, 5' untranslated region of *GLDPA* region 2.

(B) to **(E)** Histochemical GUS staining in cross sections of leaves of transgenic *F. bidentis* or *A. thaliana* harboring the *GLDPA*-Ft-2 construct. Incubation times for the GUS staining procedure were 1.5h **(E)** and 2 h (**[B]** to **[D]**).

(F) Fluorometrical quantification of GUS activities of transgenic *F. bidentis* and *A. thaliana* plants transformed with the *GLDPA*-Ft-2 construct. Each single dot represents one independent transgenic line. The number of lines examined (n) is indicated above as well as the median of all values ($\tilde{x}$), also displayed as black line in the diagram. MU, 4-methylumbelliferone.

Figure 7. Functional analysis of the interactions of regions 2 and 7 of the *GLDPA* promoter in transgenic *F. bidentis*.

(A) Schematic structure of the constructs *GLDPA*-Ft-1-2-7 and *GLDPA*-Ft-2-7. TSS, transcription start site.

(B) and **(C)** Histochemical GUS staining in leaf cross sections of transgenic *F. bidentis* transformed with either *GLDPA*-Ft-1-2-7 or *GLDPA*-Ft-2-7. Incubation times for the GUS staining procedure were 2 h **(B)** and 6 h **(C)**.

Figure 8. Functional analysis of the interactions of regions 2 and 7 of the *GLDPA* promoter in transgenic *A. thaliana*.

(A) Schematic structure of the promoter-reporter gene construct *GLDPA*-Ft-2-7. TSS, transcription start site.

(B) to **(K)** Analysis of GUS staining patterns in leaf cross sections of five independent transgenic *A. thaliana* lines (**[B]**/**[G]**, **[C]**/**[H]**, **[D]**/**[I]**, **[E]**/**[J]** and **[F]**/**[K]**) carrying the *GLDPA*-Ft-2-7 construct. Incubation times for the GUS staining procedure were 3.5 h (**[C]**, **[E]**, **[H]** and **[J]**), 4 h (**[B]**, **[D]**, **[F]**, **[I]** and **[K]**) and 5 h **(G)**.

(L) GUS staining in leaf blades of four different *GLDPA*-Ft-2-7 Arabidopsis lines showing exemplarily the smooth transition of the various expression patterns detected. This transition is schematically depicted as bluish bars representing the varying intensity of GUS staining of the mesophyll (MC) and the bundle-sheath cells including the vascular bundles (BS + VB). Incubation times for the GUS staining procedure were 6 h, 3h, 3.5h and 6h (from left to right).

Figure 9. Functional analysis of region 3 of the *GLDPA* promoter in transgenic *A. thaliana*.

(A) Schematic structure of the constructs *GLDPA*-Ft-3-7, *GLDPA*-Ft-2-3-7, *GLDPA*-Ft-2-3 and *GLDPA*-Ft-3-2. TSS, transcription start site.

(B) to **(I)** Histochemical GUS staining in cross sections of leaves of transgenic *A. thaliana* transformed with *GLDPA*-Ft-3-7, *GLDPA*-Ft-2-3-7, *GLDPA*-Ft-2-3 or *GLDPA*-Ft-3-2. Incubation times for GUS staining were 0.5 h (**[D]** and **[H]**), 1 h **(E)**, 2.5 h **(I)**, 3.5 h (**[C]** and **[G]**), 5 h **(B)** or 6 h **(F)**.

(J) Fluorometrical measurement of GUS activities in transgenic *A. thaliana* transformed with *GLDPA*-Ft-3-7, *GLDPA*-Ft-2-3-7, *GLDPA*-Ft-2-3 or *GLDPA*-Ft-3-2. Each single dot represents one independent transgenic line. The number of plants analyzed (n) is indicated at

the top of each diagram as well as the median values ($\tilde{x}$), also added as black lines in the diagrams respectively (*** $p < 0.001$; ** $p < 0.01$). MU, 4-methylumbelliferone.

Figure 10. The expression of the *GLDPA* gene of *Flaveria trinervia* is regulated by an intricate interplay of transcriptional and post-transcriptional mechanisms.

The proximal promoter P_{R7} is sufficient to confer expression specifically in bundle-sheath cells and the vascular bundles of leaves but can be effectively enhanced by regions 1 and 3. Transcripts generated at TSS_{R7} are presumably stabilized by their 5' untranslated region (5' UTR_{R7}), finally resulting in the accumulation of GLDPA protein in the distinct cell types to contribute to photorespiration. The activity of the distal promoter P_{R2} in all inner leaf tissues is also enhanced by regions 1 and 3. Transcripts from TSS_{R2} are supposed to be destabilized when they contain the sequence of region 7, which impedes RNA accumulation. The problem of RNA instability can be overcome by splicing out impairing elements assuring at least little amounts of stable *GLDPA* transcript and thus GLDPA protein additionally in the mesophyll cells in order to serve the C_1 metabolic pathway.

TABLES

Table 1. Oligonucleotides used for the amplification of the different *GLDPA*-Ft promoter regions or the *GLDPA*-Ft presequence

Oligonucleotide name	Sequence (5'–3') with restriction sites underlined
*GLDPA*2-5'-*Xba*I	TTATCTAGATGAAACAGGATGAGCCAC
*GLDPA*2-3'-*Bcu*I	TTAACTAGTGTGGAGATGATAGTTGTTG
*GLDPA*2-3'-*Xma*I	TTACCCGGGGTGGAGATGATAGTTGTT
*GLDPA*3-5'-*Xba*I	TTATCTAGAGTGGTTCCGTGCCGC
*GLDPA*3-3'-*Xba*I	ATCTCTAGAAAAAGTTCAAAACTTG
*GLDPA*3-3'-*Xma*I	TTACCCGGGAAAAGTTCAAAACTTGAT
*GLDPA*3-3'-*Bcu*I	TTAACTAGTAAAAGTTCAAAACTTGAT
*GLDPA*7-5'-*Bcu*I	TTAACTAGTCATTTGATCTATAACGAT
GLDPA-3'-*Xma*I	AAATCCCGGGAGTGTAAGATGGG
*GLDPA*5'UTR_{R7}100-5'-*Xba*I	TTATCTAGAAACCGATCAGAAAAAG
*GLDPA*5'UTR_{R7}100-3'-*Xma*I	TTACCCGGGAGTGTAAGATGGG
35S-5'-*Hind*III	GCCAAGCTTGCATGCCTGC
35S-3'-*Xba*I-*Xma*I	AATCCCGGGTCTAGAGTCCCCCGTG
ATG-5'-*att*B1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGACCCA
ATG-3'-*att*B2	GGGGACCACTTTGTACAAGAAAGCTGGGTCCATAGCCT
$GLDPA_{mt}$-5'-*att*B1	AAAAAGCAGGCTATGGAGCGTGCACGCAGG
$GLDPA_{mt\Delta 24}$-5'-*att*B1	AAAAAGCAGGCTATGTTGGGGCGCCTTGTG
$GLDPA_{mt}$-3'-*att*B2	AGAAAGCTGGGTCGGTTCTAACTTGTGAACC
*att*B1 adapter	GGGGACAAGTTTGTACAAAAAAGCAGGCT
*att*B2 adapter	GGGGACCACTTTGTACAAGAAAGCTGGGT

Table 2. Oligonucleotide combinations for the amplification of the appropriate promoter parts or Gateway compatible fragments

Construct	PCR product	5'-Oligonucleotide	3'-Oligonucleotide
GLDPA-Ft-2-3-7	*GLDPA*-2-3	*GLDPA*2-5'-*Xba*I	*GLDPA*3-3'-*Bcu*I
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-2-7	*GLDPA*-2	*GLDPA*2-5'-*Xba*I	*GLDPA*2-3'-*Bcu*I
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-3-7	*GLDPA*-3	*GLDPA*3-5'-*Xba*I	*GLDPA*3-3'-*Bcu*I
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-2-3	*GLDPA*-2-3	*GLDPA*2-5'-*Xba*I	*GLDPA*3-3'-*Xma*I
GLDPA-Ft-2	*GLDPA*-2	*GLDPA*2-5'-*Xba*I	*GLDPA*2-3'-*Xma*I
GLDPA-Ft-3-2	*GLDPA*-2	*GLDPA*2-5'-*Xba*I	*GLDPA*2-3'-*Xma*I
	GLDPA-3	*GLDPA*3-5'-*Xba*I	*GLDPA*3-3'-*Xba*I
35S:GUS	*CaMV 35S*	*35S*-5'-*Hind*III	*35S*-3'-*Xba*I-*Xma*I
35S:GLDPA-Ft-5'UTR$_{R7}$100-*GUS*	*GLDPA*-Ft-5'UTR$_{R7}$100	*GLDPA*5'UTR$_{R7}$100-5'-*Xba*I	*GLDPA*5'UTR$_{R7}$100-3'-*Xma*I
35S:mgfp6	ATG	ATG-5'-*att*B1	ATG-3'-*att*B2
35S:GLDPA$_{mt}$-Ft-*mgfp6*	*GLDPA*$_{mt}$	*GLDPA*$_{mt}$-5'-*att*B1	*GLDPA*$_{mt}$-3'-*att*B2
	*att*B1-*GLDPA*$_{mt}$-*att*B2	*att*B1 adapter	*att*B2 adapter
35S:GLDPA$_{mt\Delta 24}$-Ft-*mgfp6*	*GLDPA*$_{mt\Delta 24}$	*GLDPA*$_{mt\Delta 24}$-5'-*att*B1	*GLDPA*$_{mt}$-3'-*att*B2
	*att*B1-*GLDPA*$_{mt\Delta 24}$-*att*B2	*att*B1 adapter	*att*B2 adapter

REFERENCES

Bauwe, H., and Kolukisaoglu, U. (2003). Genetic manipulation of glycine decarboxylation. J. Exp. Bot. **54:** 1523–1535.

Bauwe, H. (2011). Photorespiration: The bridge to C_4 photosynthesis. In C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms, Advances in Photosynthesis and Respiration Series Vol. 32, Raghavendra, A.S., and Sage R.F., eds (Springer, Dordrecht, The Netherlands), pp. 81–108.

Bernard, V., Brunaud, V., and Lecharny, A. (2010). TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. BMC Genomics **11:** 166.

Boisnard-Lorig, C., Colon-Carmona, A., Bauch, M., Hodge, S., Doerner, P., Bancharel, E., Dumas, C., Haseloff, J., and Berger, F. (2001). Dynamic analyses of the expression of the HISTONE:YFP fusion protein in Arabidopsis show that syncytial endosperm is divided in mitotic domains. Plant Cell **13:** 495–509.

Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. Nat. Struct. Mol. Biol. **16:** 107–113.

Brown, N.J., Parsley, K., and Hibberd, J.M. (2005). The future of C_4 research – maize, *Flaveria* or *Cleome*? Trends Plant Sci. **10:** 215–221.

Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K., and Hibberd, J.M. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. Science **331:** 1436–1439.

Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. Annu. Rev. Biochem. **76:** 51–74.

Chen, P.Y., Wang, C.K., Soong, S.C., and To, K.Y. (2003). Complete sequence of the binary vector pBI121 and its application in cloning T-DNA insertion from transgenic plants. Mol. Breed. **11:** 287–293.

Chitty, J.A., Furbank, R.T., Marshall, J.S., Chen, Z., and Taylor, W.C. (1994). Genetic transformation of the C_4 plant, *Flaveria bidentis*. Plant J. **6:** 949–956.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735–743.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science **322:** 1845–1848.

Cossu, R. (1997). Charakterisierung der Glycindecarboxylase-Gene von *Flaveria trinervia* (C_4) und ihre Expression in transgenen *Nicotiana tabacum*, *Flaveria pubescens* und *Solanum tuberosum*. PhD thesis. Universität Hannover, Hannover, Germany.

Cossu, R., and Bauwe, H. (1998). Two genes of the *GDCSP* gene family from the C_4 plant *Flaveria trinervia*: *GDCSPA* encoding P-protein and *GDCSPB*, a pseudo-gene. Plant Physiol. **116:** 445–446.

Curtis, M.D., and Grossniklaus, U. (2003). A Gateway cloning vector set for high-throughput functional analysis of genes in plants. Plant Physiol. **133:** 462–469.

Douce, R., Bourguignon, J., Neuburger, M., and Rebeille, F. (2001). The glycine decarboxylase system: a fascinating complex. Trends Plant Sci. **6:** 167–176.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. **19:** 1349.

Egecioglu, D.E., and Chanfreau, G. (2011). Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. RNA **17:** 383–389.

Engel, N., Van Den Daele, K., Kolukisaoglu, Ü., Morgenthal, K., Weckwerth, W., Pärnik, T., Keerberg, O., and Bauwe, H. (2007). Deletion of glycine decarboxylase in Arabidopsis is lethal under nonphotorespiratory conditions. Plant Physiol. **144:** 1328–1335.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., Streubel, M., Cossu, R., Bauwe, H., and Westhoff, P. (2008). The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: Analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and Arabidopsis (C_3). Plant Physiol. **146:** 1773–1785.

Fasken, M.B., and Corbett, A.H. (2009). Mechanisms of nuclear mRNA quality control. RNA Biol. **6:** 237–241.

Foyer, C.H., Bloom, A.J., Queval, G., Noctor, G. (2009). Photorespiratory metabolism: Genes, mutants, energetics, and redox signaling. Annu. Rev. Plant Biol. **60:** 455–84.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. U.S.A. 85**:** 8998–9002.

Gowik, U., and Westhoff, P. (2011). The path from C_3 to C_4 photosynthesis. Plant Physiol. **155:** 56–63.

Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? Plant Cell **23:** 2087–2105.

Hainer, S.J., Pruneski, J.A., Monteverde, R.M., and Martens, J.A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev. **25:** 29–40.

Hanson, A.D., and Roje, S. (2001). One-carbon metabolism in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **52:** 119–137.

Hatch, M.D. (1987). C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim. Biophys. Acta **895:** 81–106.

Hedtke, B., and Grimm, B. (2009). Silencing of a plant gene by transcriptional interference. Nucleic Acids Res. **37:** 3739–3746.

Hori, K., and Watanabe, Y. (2007). Context analysis of termination codons in mRNA that are recognized by plant NMD. Plant Cell Physiol. **48:** 1072–1078.

Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. Nat. Rev. Mol. Cell Biol. **7:** 529–539.

Huang, S., Taylor, N.L., Whelan, J., and Millar, A.H. (2009). Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. Plant Physiol. **150:** 1272–1285.

Jefferson, R.A., Kavangh, T.A., and Bevan, M.W. (1987). GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6:** 3901–3907.

Joshi, C.P. (1987). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res. **15:** 6643–6651.

Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfy, T., Barta, E., and Silhavy D. (2006). Both introns and long 3'-UTRs operate as *cis*-acting elements to trigger nonsense-mediated decay in plants. Nucleic Acids Res. **34:** 6147–6157.

Koncz, C., and Schell, J. (1986). The promoter of the T_L-DNA gene *5* controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol. Gen. Genet. **204:** 383–396.

Kosugi, S., Ohashi, Y., Nakajima, K., and Arai, Y. (1990). An improved assay for β-glucuronidase in transformed cells: Methanol almost completely suppresses a putative endogenous β-glucuronidase activity. Plant Sci. **70:** 133–140.

Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1991). A DNA transformation-competent Arabidopsis genomic library in *Agrobacterium*. Nat. Biotechnol. **9:** 963–967.

Leegood, R.C., Lea, P.J., Adcock, M.D., and Häusler, R.E. (1995). The regulation and control of photorespiration. J. Exp. Bot. **46:** 1397–1414.

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. Cell **145:** 502–511.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers, C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., and Brutnell, T.P. (2010). The developmental dynamics of the maize leaf transcriptome. Nat. Genet. **42:** 1060–1067.

Logemann, E., Birkenbihl, R.P., Ülker, B., and Somssich, I.E. (2006). An improved method for preparing *Agrobacterium* cells that simplifies the Arabidopsis transformation protocol. Plant Methods **2:**16.

Majeran, W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., Zhang, C., Asakura, Y., Bhuiyan, N.H., Sun, Q., Turgeon, R., and van Wijk, K.J. (2010). Structural and metabolic transitions of C_4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. Plant Cell **22:** 3509–3542.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae SER3* gene. Nature **429:** 571–574.

Mazo, A., Hodgson, J.W., Petruk, S., Sedkov, Y., and Brock, H.W. (2007). Transcriptional interference: an unexpected layer of complexity in gene regulation. J. Cell. Sci. **120:** 2755–2761.

McKown, A.D., Moncalvo, J.M., and Dengler, N.G. (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. Am. J. Bot. **92:** 1911–1928.

Molina, C., and Grotewold, E. (2005). Genome wide analysis of Arabidopsis core promoters. BMC Genomics **6:** 25.

Morgan, C.L., Turner, S.R., and Rawsthorne, S. (1993). Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and serine hydroxymethyltransferase in leaves of C_3-C_4 intermediate species from different genera. Planta **190:** 468–473.

Mouillon, J.M., Aubert, S., Bourguignon, J., Gout, E., Douce, R., and Rébeillé, F. (1999). Glycine and serine catabolism in non-photosynthetic higher plant cells: their role in C_1 metabolism. Plant J. **20:** 197–205.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. **39:** 1507–1511.

Nakamura, M., Tsunoda, T., and Obokata, J. (2002). Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. Plant J. **29:** 1–10.

Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., and Silhavy, D. (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant Mol. Biol. **71:** 367–378.

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

Ogren, W.L. (1984). Photorespiration: pathways, regulation, and modification. Annu. Rev. Plant Physiol. **35:** 415–442.

Ohnishi, J., and Kanai, R. (1983). Differentiation of photorespiratory activity between mesophyll and bundle sheath cells of C_4 plants. I. Glycine oxidation by mitochondria. Plant Cell Physiol. **24:** 1411–1420.

Oliver, D.J. (1994). The glycine decarboxylase complex from plant mitochondria. Annu. Rev. Plant. Physiol. Plant Mol. Biol. **45:** 323–327.

Palmer, A.C., Egan, J.B., and Shearwin, K.E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. Transcription **2:** 9–14.

Patel, M., Siegel, A.J., and Berry, J.O. (2006). Untranslated regions of *FbRbcS1* mRNA mediate bundle sheath cell-specific gene expression in leaves of a C_4 plant. J. Biol. Chem. **281:** 25485–25491.

Powell, A.M. (1978). Systematics of *Flaveria* (Flaveriinae-Asteraceae). Ann. Mo. Bot. Gard. **65:** 590–636.

Rawsthorne, S., Hylton, C.M., Smith, A.M., and Woolhouse, H.W. (1988). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C_3 and C_3-C_4 intermediate species of *Moricandia*. Planta **173:** 298–308.

Sage, R.F. (2004). The evolution of C_4 photosynthesis. New Phytol. **161:** 341–370.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Tanudji, M., Sjöling, S., Glaser, E., and Whelan, J. (1999). Signals required for the import and processing of the alternative oxidase into mitochondria. J. Biol. Chem. **274:** 1286–1293.

Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J., and Nourani, A. (2011). Transcription regulation by the noncoding RNA *SRG1* requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. Mol. Cell. Biol. **31:** 1288–1300.

Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J. **33:** 949–956.

Waadt, R., and Kudla, J. (2008). *In Planta* Visualization of Protein Interactions Using Bimolecular Fluorescence Complementation (BiFC). CSH Protoc., **doi:** 10.1101/pdb.prot4995.

Westhoff, P., Offermann-Steinhard, K., Höfer, M., Eskins, K., Oswald, A., and Streubel, M. (1991). Differential accumulation of plastid transcripts encoding photosystem II components in the mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C_4 plants. Planta **184:** 377–388.

Westhoff, P., and Gowik, U. (2004). Evolution of C4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus *Flaveria*. Ann. Bot. **93:** 13–23.

Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K., and Abe, T. (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics **8:** 67.

Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., and Obokata, J. (2009). Heterogeneity of Arabidopsis core promoters revealed by high density TSS analysis. Plant J. **60:** 350–362.

Yoo, S.D., Cho, Y.H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. **2:** 1565–1572.

Yoshimura, Y., Kubota, F., and Ueno, O. (2004). Structural and biochemical bases of photorespiration in C_4 plants: quantification of organelles and glycine decarboxylase. Planta **220:** 307–317.

Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature **408:** 225–229.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. Nat. Genet. **39:** 1512–1516.

Zhou, X., Carranco, R., Vitha, S., and Hall, T.C. (2005). The dark side of green fluorescent protein. New Phytol. **168:** 313–322.

Zuo, Y.C., and Li, Q.Z. (2011). Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. Genomics **97:** 112-120.

FIGURES

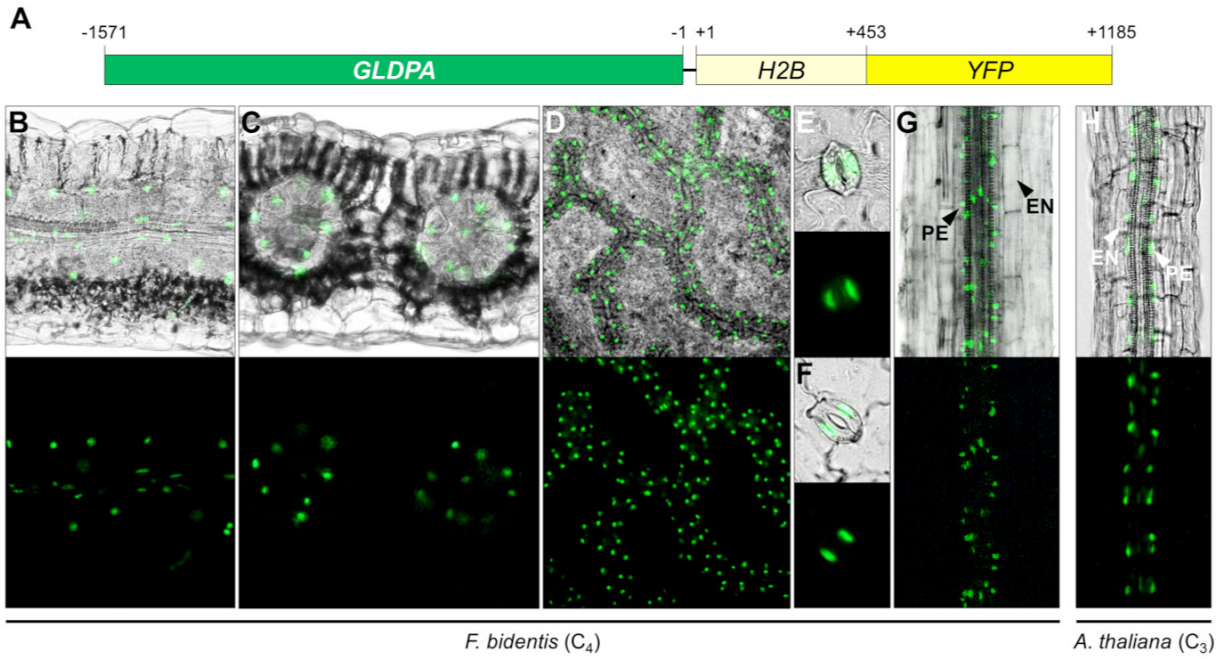


Figure 1. Fluorescence microscopic analysis of *GLDPA* promoter activity in transgenic *F. bidentis* and *A. thaliana*.

(A) Schematic presentation of the *GLDPA*-Ft:*H2B-YFP* construct which was transformed into *F. bidentis* or *A. thaliana* to express the yellow fluorescent protein (YFP) fused to histone 2B (H2B) under the control of the *GLDPA* promoter.

(B) to **(H)** The localization of YFP was examined by fluorescence microscopy in longitudinal **(B)** and cross sections **(C)** of leaves, whole leaf blades in top view **(D)**, guard cells of both the upper **(E)** and lower epidermis **(F)** and in roots (**[G]** and **[H]**). The fluorescence image is displayed underneath and the corresponding merge of the fluorescent signal and the bright-field picture above respectively. In the root single endodermis (EN) and pericycle (PE) cells are depicted by an arrowhead.

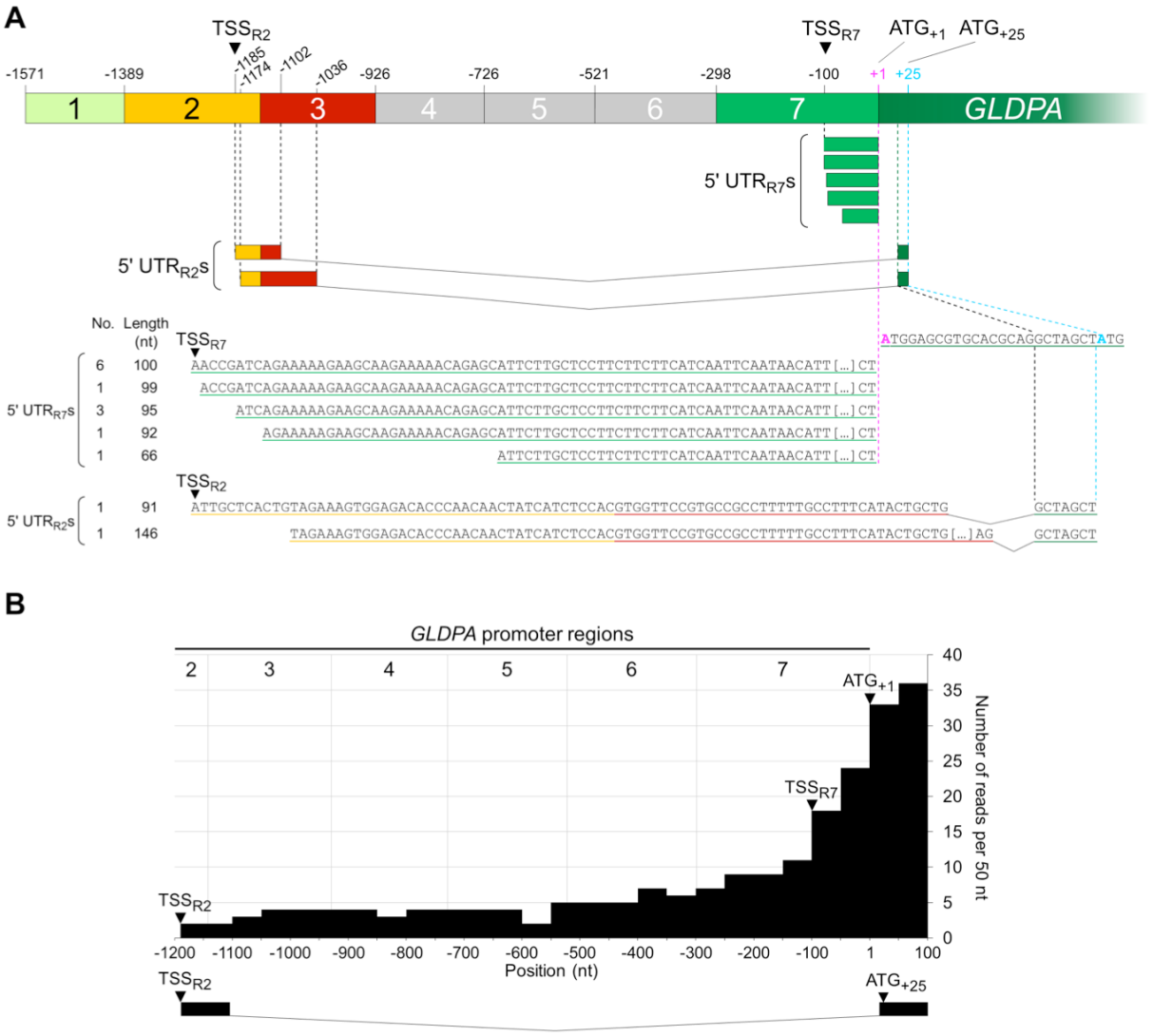


Figure 2. Analysis of transcript 5' ends of the endogenous *GLDPA* gene of *F. trinervia* by rapid amplification of 5' complementary DNA ends (5' RACE) (A) and 454 pyrosequencing (B).

(A) The *GLDPA* promoter and its transcriptional output. The dissection of the *GLDPA* promoter into 7 regions has been described in Engelmann et al. (2008). The schematic structure of the 5' untranslated regions (5' UTRs) of the two types of RNAs originating from region 2 (5' UTR_{R2}s) or region 7 (5' UTR_{R7}s) and their corresponding cDNA sequences are depicted below the schematic drawing of the *GLDPA* promoter. The transcription start sites (TSSs) within region 2 (TSS_{R2}) and 7 (TSS_{R7}) are indicated as well as the start codons used when transcription starts from region 2 (ATG_{+25}) or region 7 (ATG_{+1}) and the number (No.) of the 5' UTRs detected for each 5' UTR variant whose length is stated in nucleotides (nt).

(B) Diagram showing the read coverage of the *GLDPA* contig derived from 454 sequencing reads (Gowik et al., 2011). The coverage upstream of the translational start site (ATG_{+1}) up to 100 nt downstream was analyzed in 50 nt windows. A contig corresponding to the 91-nt spliced variant starting from TSS_{R2} that was detected by 5' RACE **(A)** was represented by only two 454 reads. The transcription start sites (TSS_{R2} and TSS_{R7}) and the start codons (ATG_{+1} and ATG_{+25}) are marked by arrowheads. The different *GLDPA* promoter regions 2 to 7 shown as columns are allocated to their respective positions.

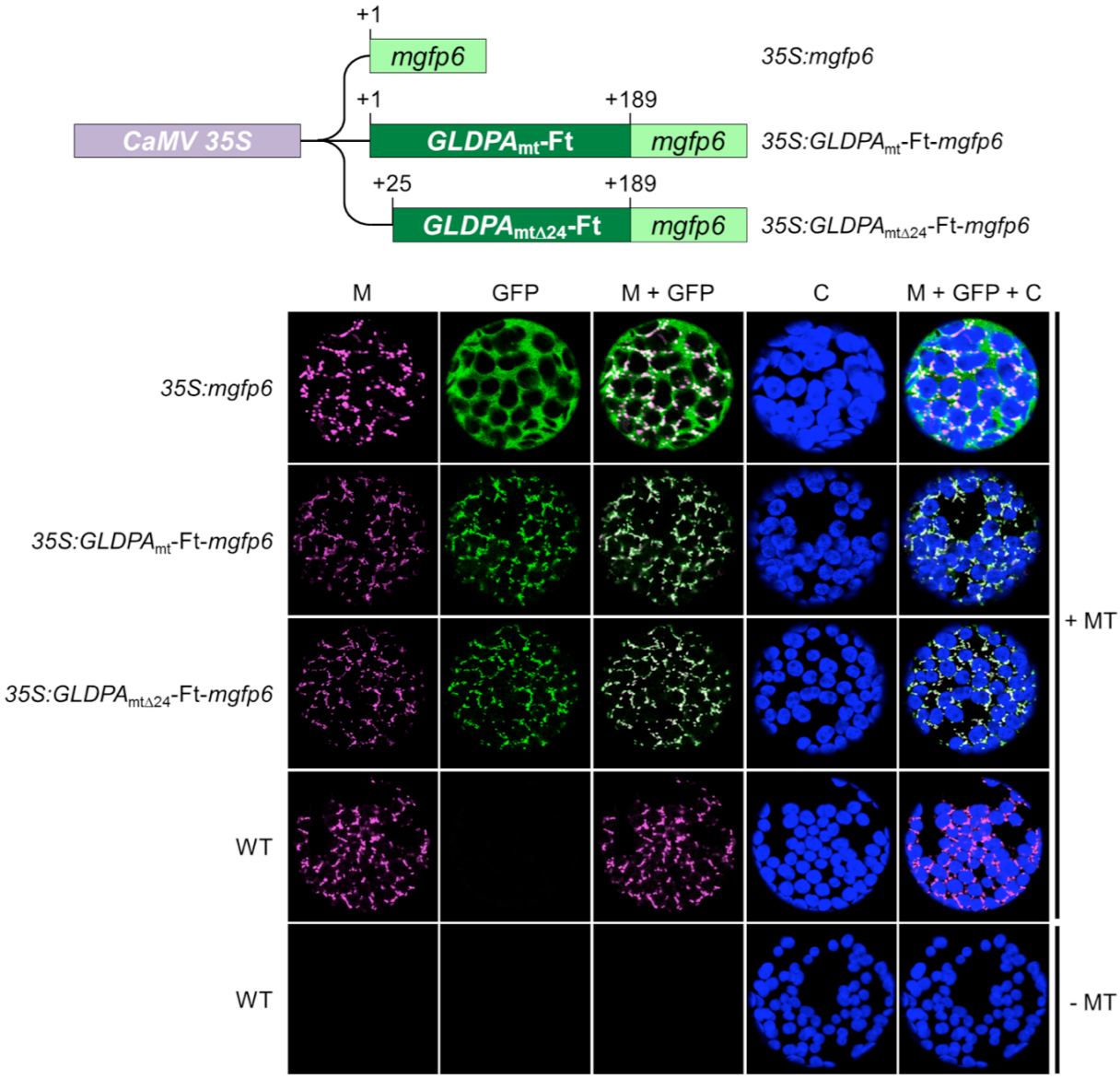


Figure 3. Localization study of the two different transit peptide variants of the GLDPA protein.

The structures of the three constructs used for transient expression in leaves of *Nicotiana benthamiana* are diagrammed on top of the figure. *35S:GLDPA*$_{mt}$-Ft-*mgfp6* contains the full-length *GLDPA* sequence encoding the predicted presequence for mitochondrial targeting (*GLDPA*$_{mt}$-Ft). In *35S:GLDPA*$_{mtΔ24}$-Ft-*mgfp6* the transit peptide lacks the eight amino terminal residues. *35S:mgfp6* is devoid of any transit peptide sequence and served as a control. For visualizing mitochondria MitoTracker staining was carried out (+MT) or omitted as negative control (-MT). Three different channels were utilized to separate the fluorescence signals of MitoTracker-labeled mitochondria (magenta color), GFP (green color) and chlorophyll of chloroplasts (blue color) from each other. When merging MitoTracker and GFP fluorescences (M + GFP), white color indicates overlapping of both signals. All three different fluorescence signals are merged in the last column (M + GFP + CP). C, chloroplasts; GFP, green fluorescent protein; M, mitochondria; MT, MitoTracker; WT, wild type.

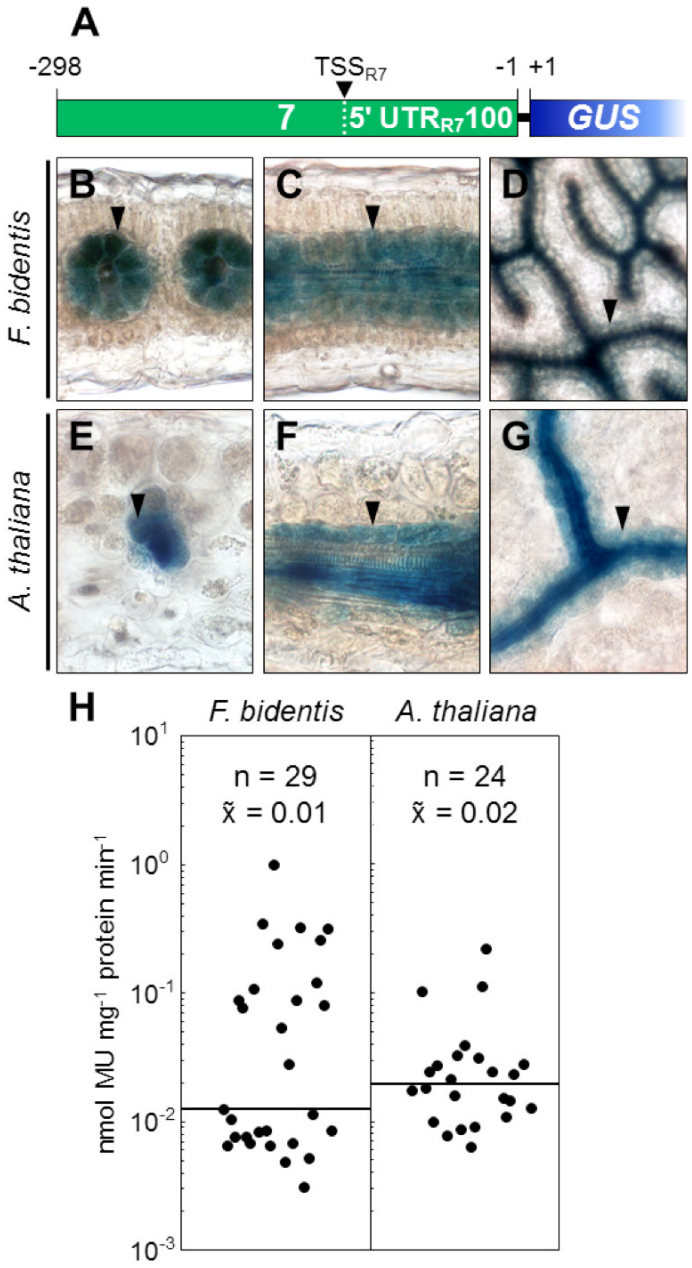


Figure 4. Functional analysis of region 7 of the *GLDPA* promoter in leaves of transgenic *F. bidentis* and *A. thaliana*.

(A) Schematic presentation of the *GLDPA*-Ft-7 construct. TSS, transcription start site; 5' UTR$_{R7}$100, 100-bp 5' untranslated region of *GLDPA* region 7.

(B) to **(G)** Histochemical localization of GUS activity in cross sections (**[B]**, **[C]**, **[E]** and **[F]**) and leaf blades in top view (**[D]** and **[G]**) of leaves of transgenic *F. bidentis* and *A. thaliana*. Single bundle-sheath cells are emphasized by arrowheads. Incubation times for the GUS staining were 17 h (**[E]** and **[F]**), 29 h **(G)**, 43 h **(B)**, 66 h **(C)** and 70.5 h **(D)**.

(H) Fluorometrical quantification of GUS activities of transgenic *F. bidentis* and *A. thaliana* plants transformed with the *GLDPA*-Ft-7 construct. Each single dot represents one independent transgenic line. The number of lines examined (n) is indicated above as well as the median of all values ($\tilde{x}$), also displayed as black line in the diagram. MU, 4-methylumbelliferone.

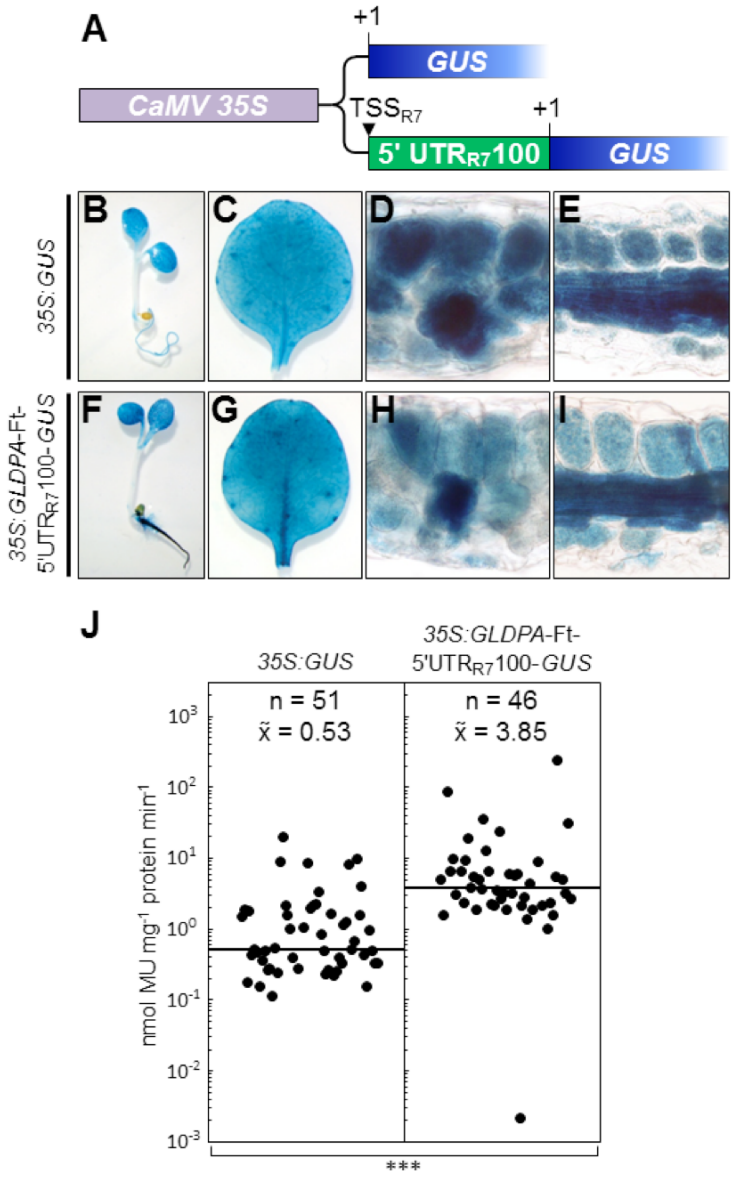


Figure 5. Analysis of the gene-regulatory properties of the 100-bp 5' untranslated region of *GLDPA* region 7 (5' UTR$_{R7}$100) in transgenic *A. thaliana*.

(A) Schematic structure of the two constructs used for transformation. *35S:GUS* consists of the *CaMV 35S* promoter and the *GUS* reporter gene, while *35S:GLDPA*-Ft-5'UTR$_{R7}$100-*GUS* additionally contains the 100-bp long 5' UTR$_{R7}$. TSS, transcription start site.

(B) to **(I)** Histochemical GUS staining of *A. thaliana* transformed with *35S:GUS* or *35S:GLDPA*-Ft-5'UTR$_{R7}$100-*GUS* in seedlings (**[B]** and **[F]**), young leaf blades (**[C]** and **[G]**) and cross sections of mature rosette leaves (**[D]**, **[E]**, **[H]** and **[I]**). Staining was for 1 h (**[H]** and **[I]**), 3 h (**[C]** and **[G]**), 4 h (**[B]** and **[F]**) or 16 h (**[D]** and **[E]**).

(J) Quantitative measurements of expression strength by analyzing GUS activities in leaf extracts of transgenic Arabidopsis plants transformed with the *35S:GUS* or *35S:GLDPA*-Ft-5'UTR$_{R7}$100-*GUS* construct. Each single dot represents one independent transgenic *A. thaliana* line. The number of lines examined (n) is indicated above as well as the median of all values ($\tilde{x}$) which is additionally charted as black line in the diagram (*** $p < 0.001$). MU, 4-methylumbelliferone.

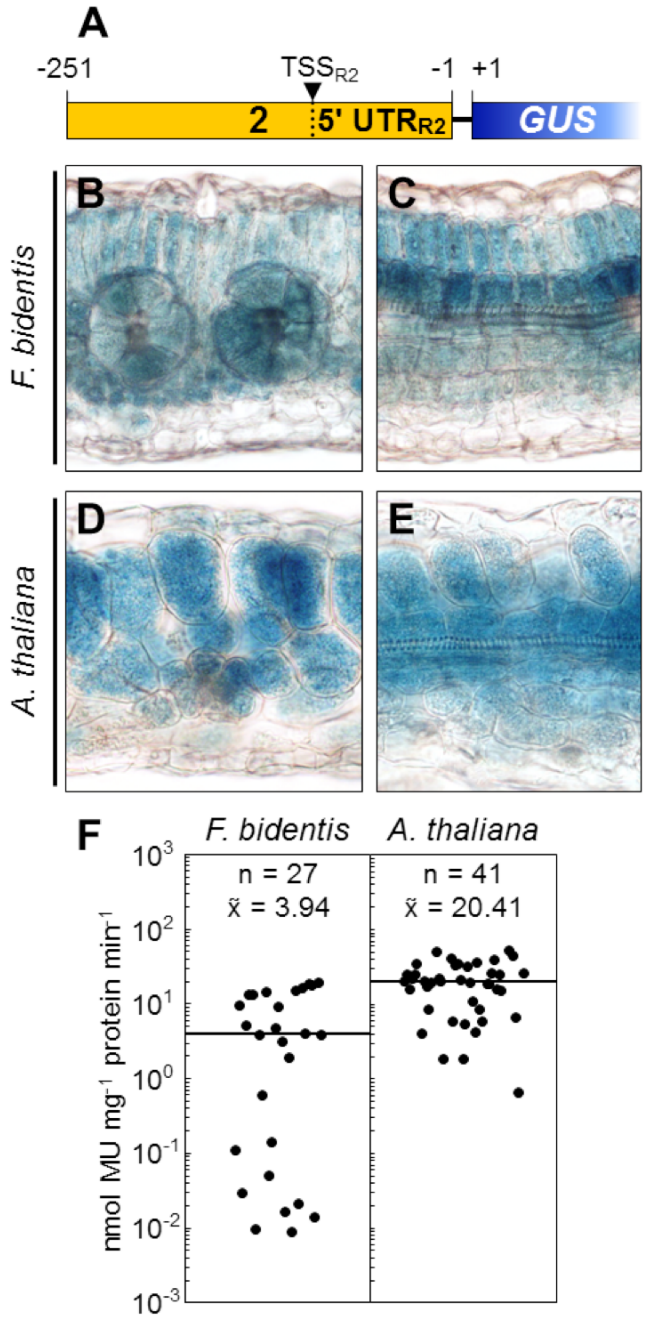


Figure 6. Functional analysis of region 2 of the *GLDPA* promoter in leaves of transgenic *F. bidentis* and *A. thaliana*.

(A) Schematic structure of the *GLDPA*-Ft-2 construct. TSS, transcription start site; 5' UTR$_{R2}$, 5' untranslated region of *GLDPA* region 2.

(B) to **(E)** Histochemical GUS staining in cross sections of leaves of transgenic *F. bidentis* or *A. thaliana* harboring the *GLDPA*-Ft-2 construct. Incubation times for the GUS staining procedure were 1.5h **(E)** and 2 h (**[B]** to **[D]**).

(F) Fluorometrical quantification of GUS activities of transgenic *F. bidentis* and *A. thaliana* plants transformed with the *GLDPA*-Ft-2 construct. Each single dot represents one independent transgenic line. The number of lines examined (n) is indicated above as well as the median of all values ($\tilde{x}$), also displayed as black line in the diagram. MU, 4-methylumbelliferone.

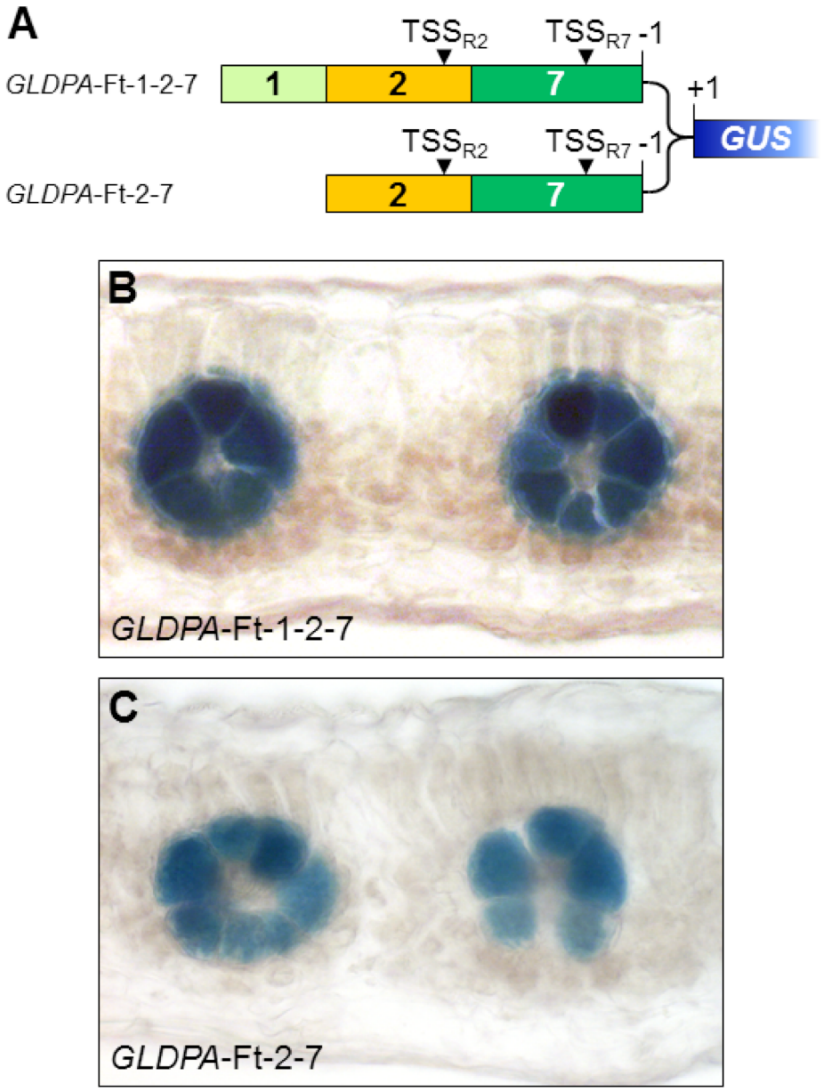


Figure 7. Functional analysis of the interactions of regions 2 and 7 of the *GLDPA* promoter in transgenic *F. bidentis*.

(A) Schematic structure of the constructs *GLDPA*-Ft-1-2-7 and *GLDPA*-Ft-2-7. TSS, transcription start site. **(B)** and **(C)** Histochemical GUS staining in leaf cross sections of transgenic *F. bidentis* transformed with either *GLDPA*-Ft-1-2-7 or *GLDPA*-Ft-2-7. Incubation times for the GUS staining procedure were 2 h **(B)** and 6 h **(C)**.

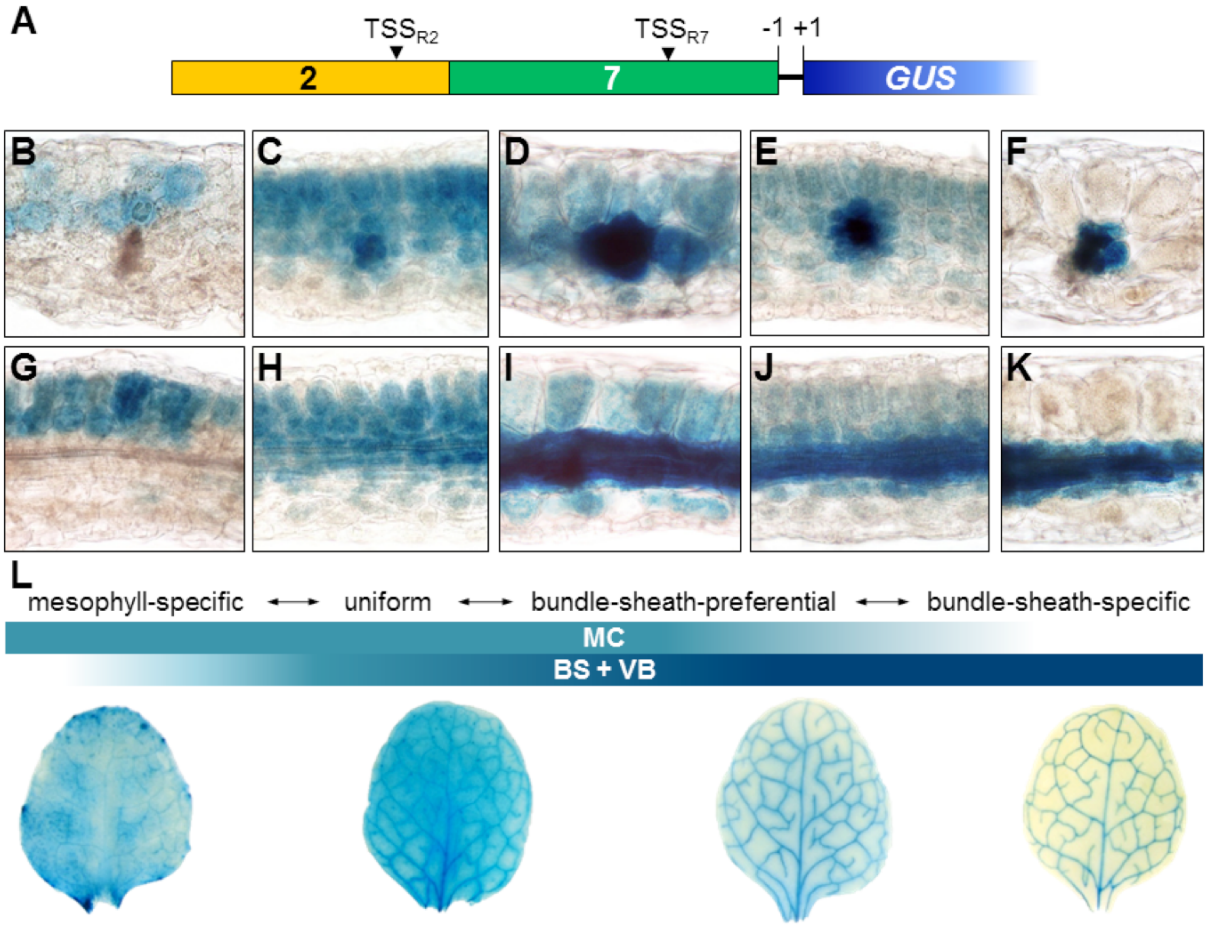


Figure 8. Functional analysis of the interactions of regions 2 and 7 of the *GLDPA* promoter in transgenic *A. thaliana*.

(A) Schematic structure of the promoter-reporter gene construct *GLDPA*-Ft-2-7. TSS, transcription start site.

(B) to **(K)** Analysis of GUS staining patterns in leaf cross sections of five independent transgenic *A. thaliana* lines (**[B]**/**[G]**, **[C]**/**[H]**, **[D]**/**[I]**, **[E]**/**[J]** and **[F]**/**[K]**) carrying the *GLDPA*-Ft-2-7 construct. Incubation times for the GUS staining procedure were 3.5 h (**[C]**, **[E]**, **[H]** and **[J]**), 4 h (**[B]**, **[D]**, **[F]**, **[I]** and **[K]**) and 5 h **(G)**.

(L) GUS staining in leaf blades of four different *GLDPA*-Ft-2-7 Arabidopsis lines showing exemplarily the smooth transition of the various expression patterns detected. This transition is schematically depicted as bluish bars representing the varying intensity of GUS staining of the mesophyll (MC) and the bundle-sheath cells including the vascular bundles (BS + VB). Incubation times for the GUS staining procedure were 6 h, 3h, 3.5h and 6h (from left to right).

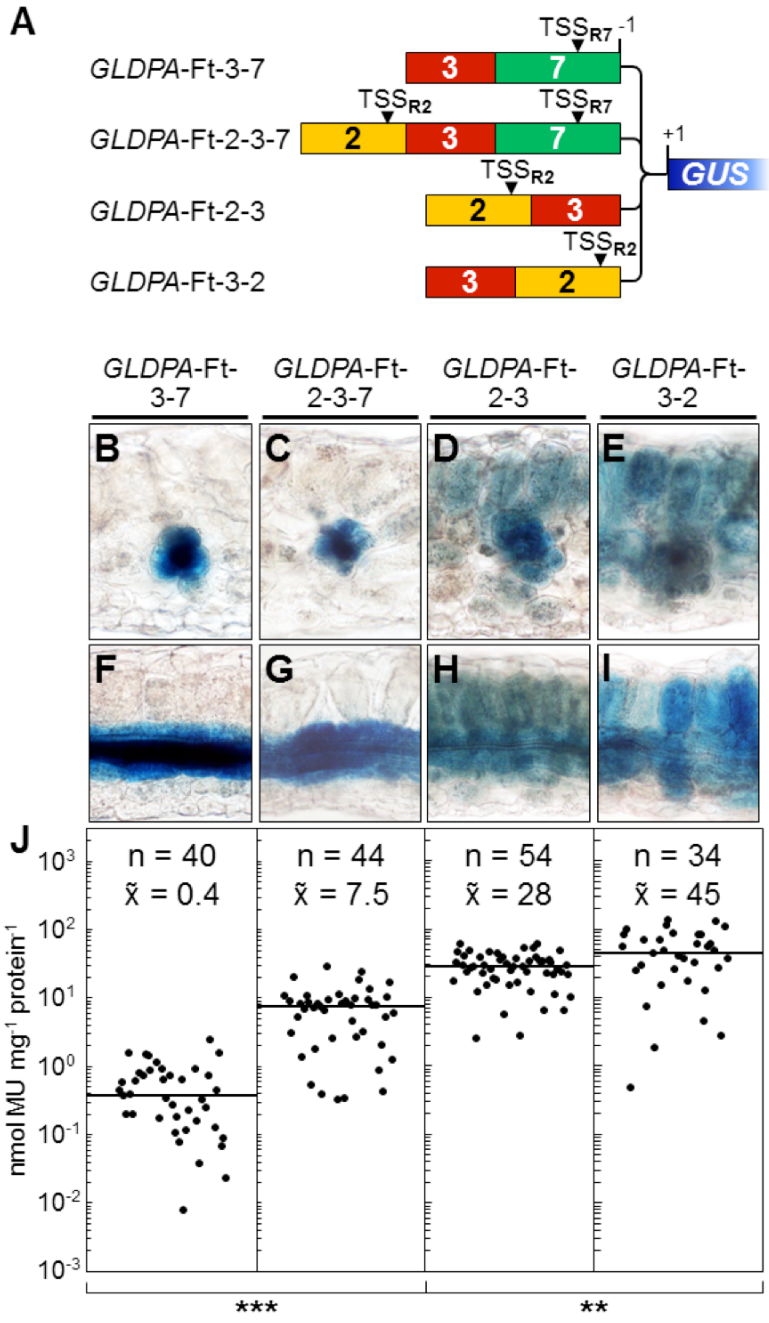


Figure 9. Functional analysis of region 3 of the *GLDPA* promoter in transgenic *A. thaliana*.

(A) Schematic structure of the constructs *GLDPA*-Ft-3-7, *GLDPA*-Ft-2-3-7, *GLDPA*-Ft-2-3 and *GLDPA*-Ft-3-2. TSS, transcription start site.

(B) to **(I)** Histochemical GUS staining in cross sections of leaves of transgenic *A. thaliana* transformed with *GLDPA*-Ft-3-7, *GLDPA*-Ft-2-3-7, *GLDPA*-Ft-2-3 or *GLDPA*-Ft-3-2. Incubation times for GUS staining were 0.5 h (**[D]** and **[H]**), 1 h **(E)**, 2.5 h **(I)**, 3.5 h (**[C]** and **[G]**), 5 h **(B)** or 6 h **(F)**.

(J) Fluorometrical measurement of GUS activities in transgenic *A. thaliana* transformed with *GLDPA*-Ft-3-7, *GLDPA*-Ft-2-3-7, *GLDPA*-Ft-2-3 or *GLDPA*-Ft-3-2. Each single dot represents one independent transgenic line. The number of plants analyzed (n) is indicated at the top of each diagram as well as the median values ($\tilde{x}$), also added as black lines in the diagrams respectively (*** $p < 0.001$; ** $p < 0.01$). MU, 4-methylumbelliferone.

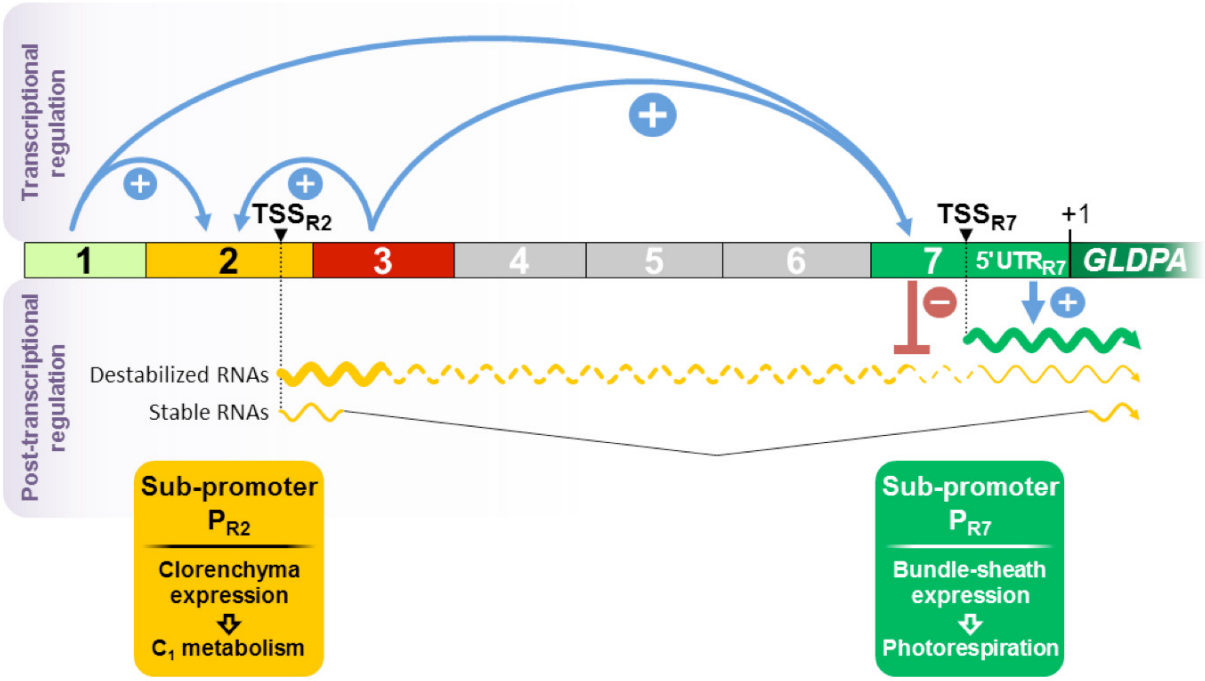
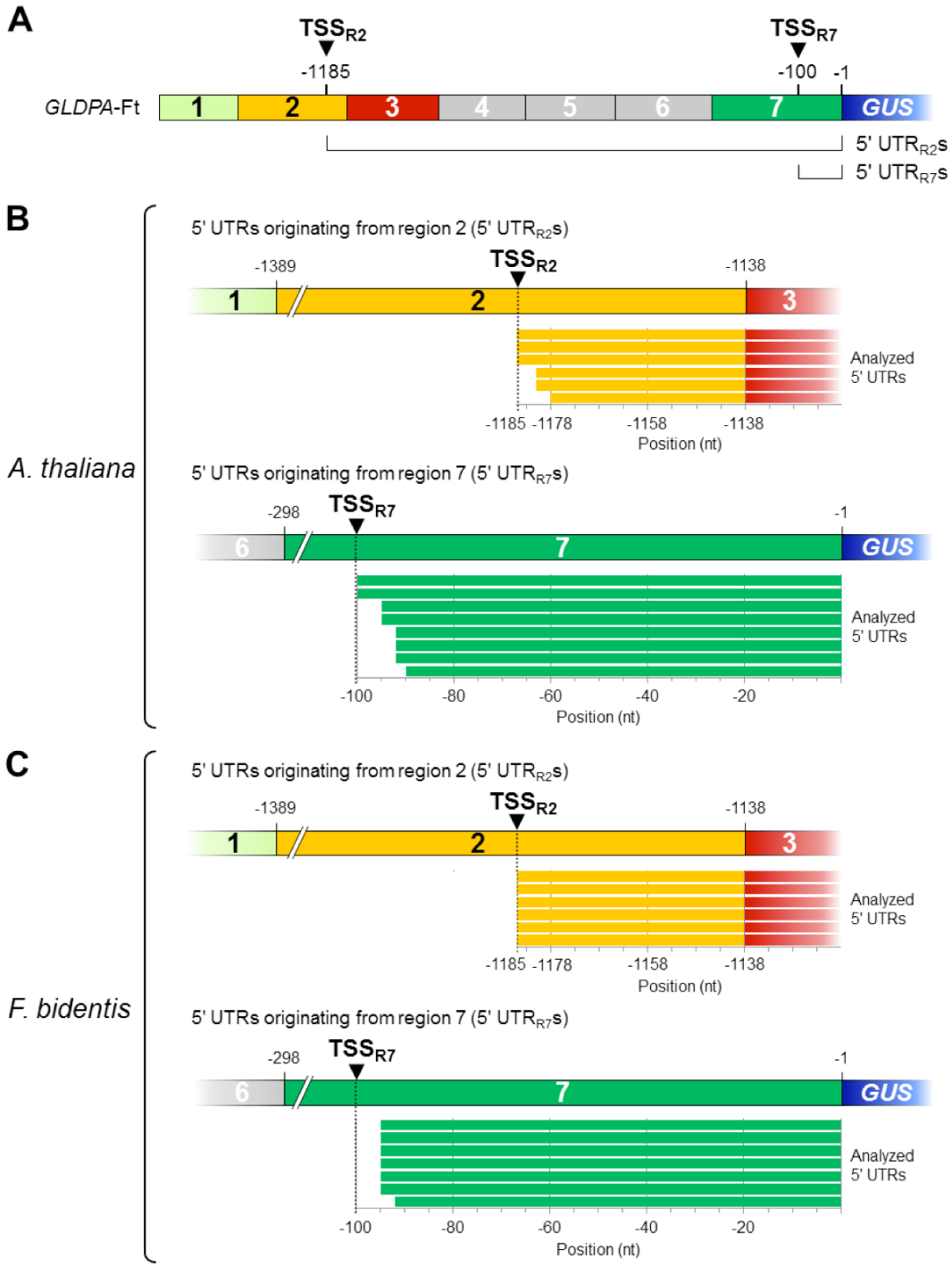


Figure 10. The expression of the *GLDPA* gene of *Flaveria trinervia* is regulated by an intricate interplay of transcriptional and post-transcriptional mechanisms.

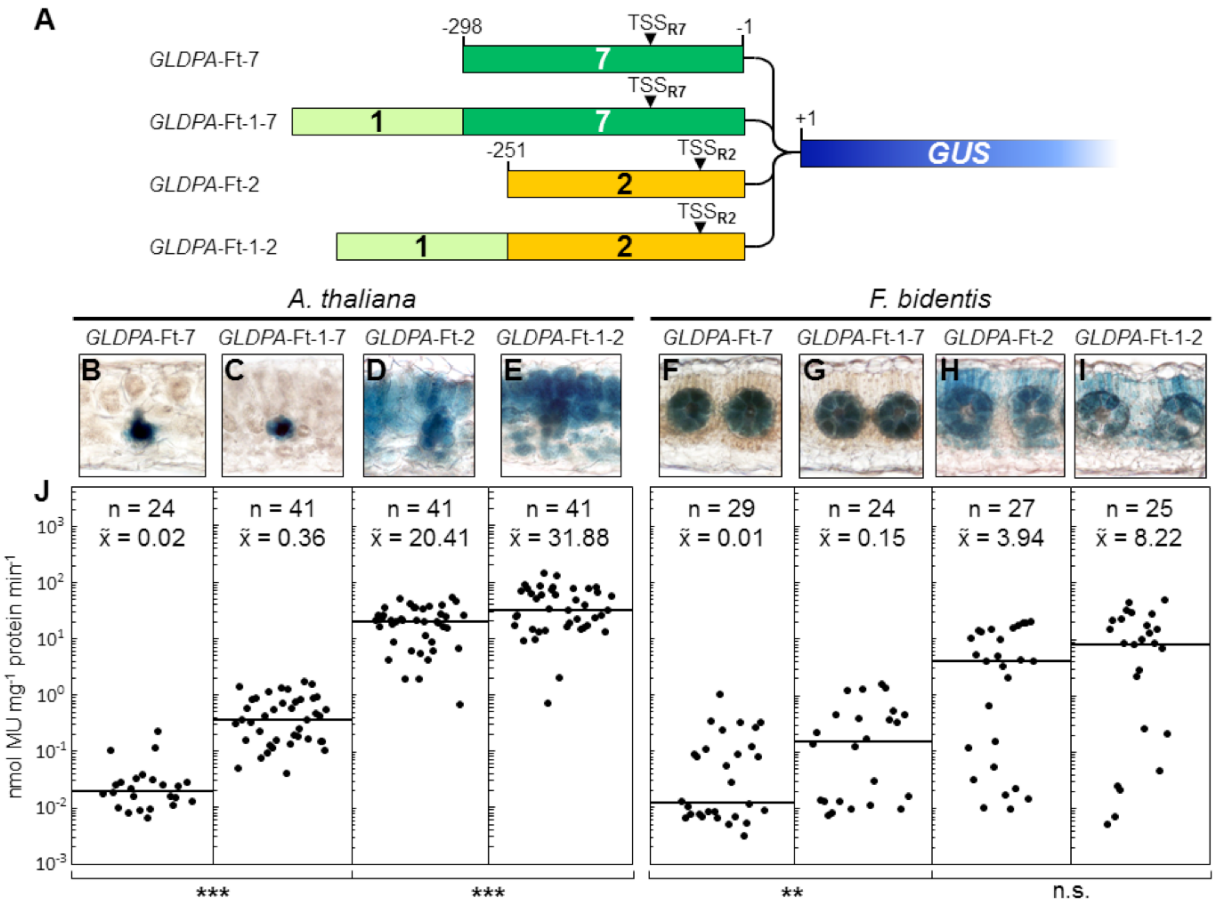
The proximal promoter P_{R7} is sufficient to confer expression specifically in bundle-sheath cells and the vascular bundles of leaves but can be effectively enhanced by regions 1 and 3. Transcripts generated at TSS_{R7} are presumably stabilized by their 5' untranslated region (5' UTR_{R7}), finally resulting in the accumulation of GLDPA protein in the distinct cell types to contribute to photorespiration. The activity of the distal promoter P_{R2} in all inner leaf tissues is also enhanced by regions 1 and 3. Transcripts from TSS_{R2} are supposed to be destabilized when they contain the sequence of region 7, which impedes RNA accumulation. The problem of RNA instability can be overcome by splicing out impairing elements assuring at least little amounts of stable *GLDPA* transcript and thus GLDPA protein additionally in the mesophyll cells in order to serve the C_1 metabolic pathway.

SUPPLEMENTAL DATA



Supplemental Figure 1. Analysis of mRNA 5' ends in leaves of transgenic *A. thaliana* and *F. bidentis* harboring the *GLDPA*-Ft:*GUS* transgene.

(A) Schematic structure of the *GLDPA*-Ft construct. TSS, transcription start site; UTR, untranslated region. **(B)** and **(C)** Identification of 5' UTRs by the rapid amplification of 5' complementary DNA ends (5' RACE) in leaves of transgenic *A. thaliana* **(B)** and *F. bidentis* **(C)** plants carrying the *GLDPA*-Ft construct. Six (**[B]** and **[C]**), seven **(C)** or eight **(B)** cloned 5' RACE products were randomly selected and analyzed by DNA sequencing. 5' UTR$_{R2}$s reached to position -1. nt, nucleotide.

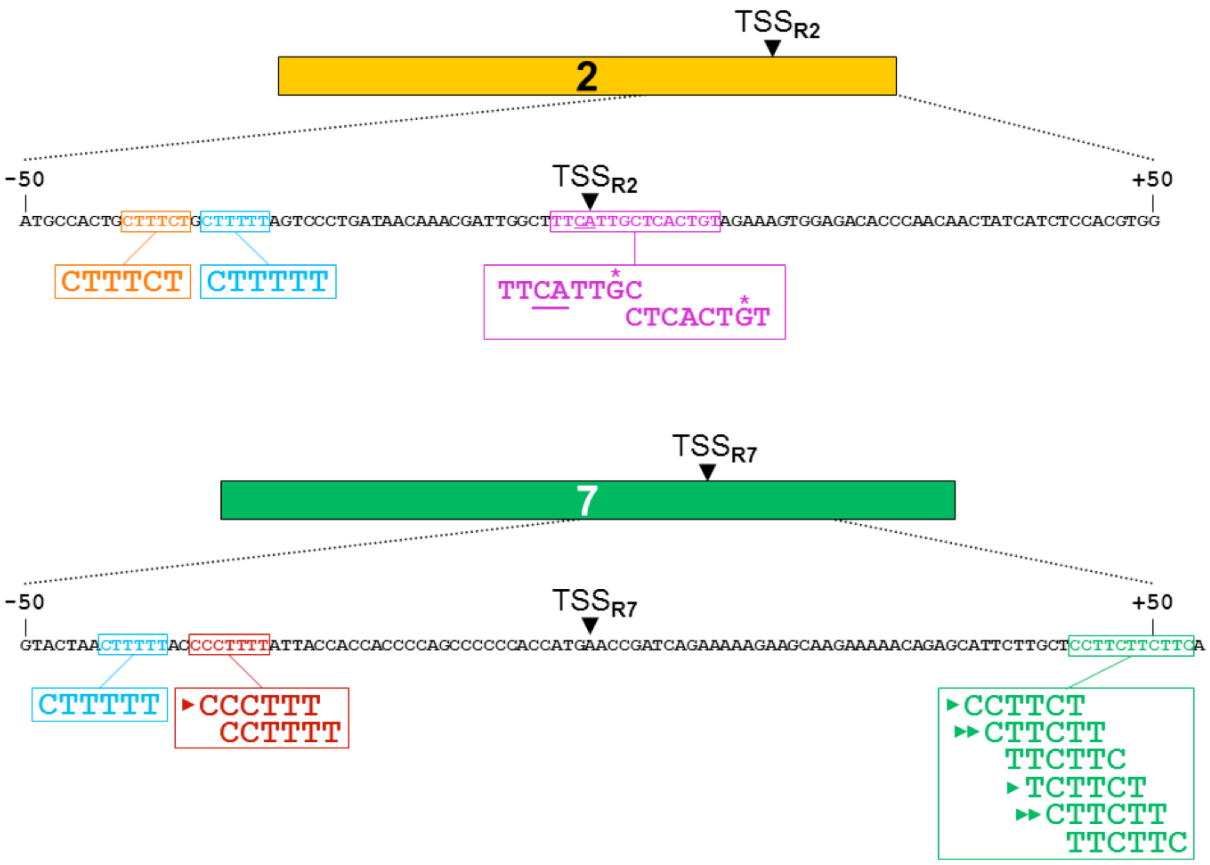


Supplemental Figure 2. Functional analysis of region 1 of the *GLDPA* promoter in transgenic *A. thaliana* and *F. bidentis*.

(A) Schematic presentation of the transformed constructs *GLDPA*-Ft-7, *GLDPA*-Ft-1-7, *GLDPA*-Ft-2 and *GLDPA*-Ft-1-2. TSS, transcription start site.

(B) to **(I)** Histochemical GUS staining in cross sections of leaves of transgenic *A. thaliana* and *F. bidentis* plants carrying *GLDPA*-Ft-7, *GLDPA*-Ft-1-7, *GLDPA*-Ft-2 or *GLDPA*-Ft-1-2. GUS staining incubation times were 1.5 h **(D)**, 2 h (**[E]** and **[I]**), 3 h **(H)**, 16 h **(C)**, 17 h **(B)**, 43 h **(F)** or 46 h **(G)**.

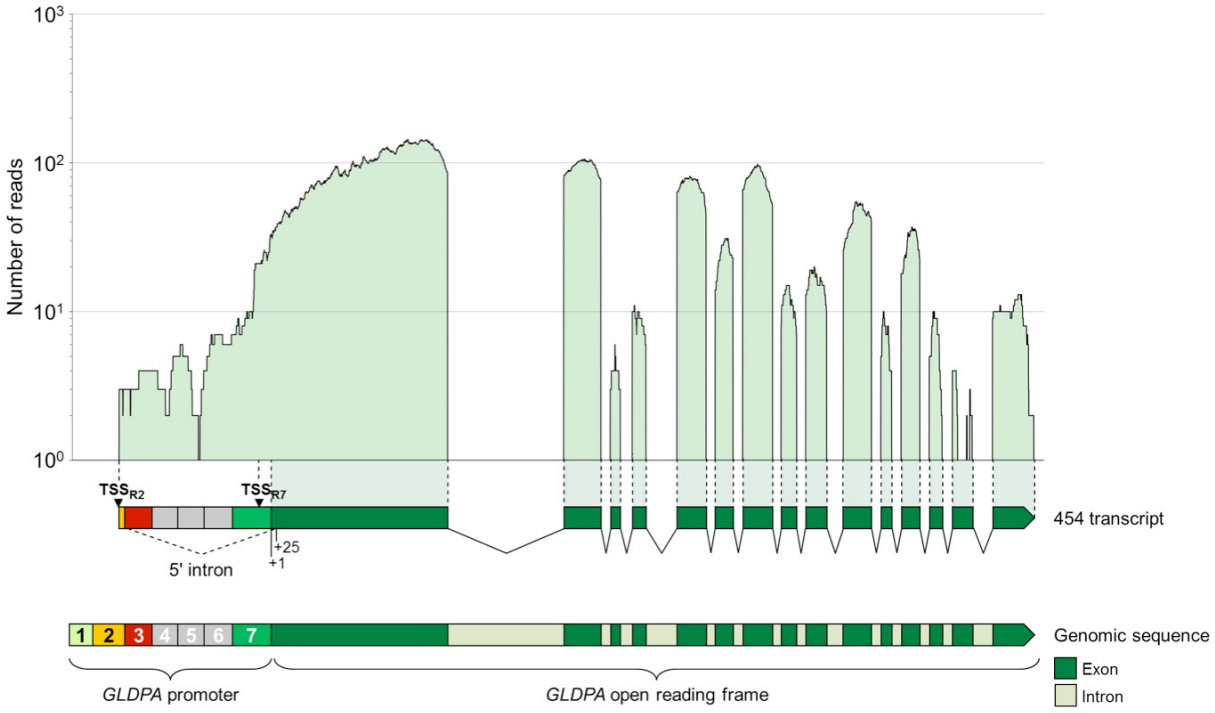
(J) Quantitative GUS measurements in leaf extracts of *A. thaliana* and *F. bidentis* carrying *GLDPA*-Ft-7, *GLDPA*-Ft-1-7, *GLDPA*-Ft-2 or *GLDPA*-Ft-1-2. Each single dot represents one independent transgenic plant line. The number of lines examined (n) is indicated at the top of each diagram as well as the median values ($\tilde{x}$), also added as black lines in the diagrams (*** $p < 0.001$; ** $p < 0.01$; n.s., not significant, $p > 0.05$). MU, 4-methylumbelliferone.



Supplemental Figure 3. Distribution of TC-rich elements and Initiator-like motifs within region 2 and region 7 of the *GLDPA* promoter.

The localization of TC-rich elements found within a range of -50 nt to +50 nt around TSS_{R2} and TSS_{R7} is shown. The TC-elements that are unique to region 7 within the *GLDPA* promoter are highlighted with a single arrowhead. The double arrowhead in front of CTTCTT emphasizes that this TC-element occurs twice but is unique to region 7. The two Initiator-like elements TTCATTGC and CTCACTGT encompassing TSS_{R2} are colored in purple with the asterisk indicating the single nucleotide that differs from the consensus sequence YTCANTYY (N = A, C, G or T; Y = C or T; Nakamura et al., 2002) respectively. The CA dimer of the Initiator motif is underlined with A representing TSS_{R2}.

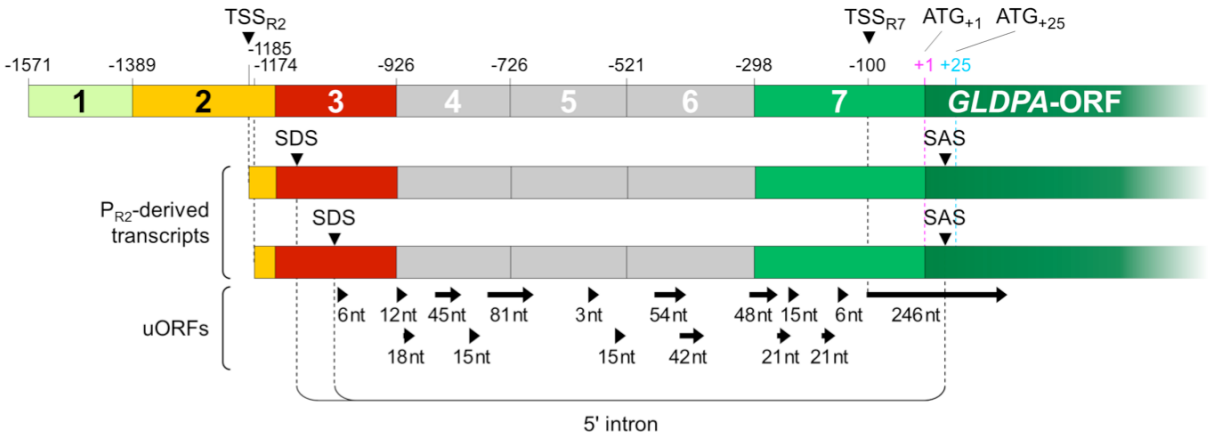
Nakamura, M., Tsunoda, T., and Obokata, J. (2002). Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. Plant J. **29:** 1–10.



Supplemental Figure 4. Splicing pattern of the *GLDPA* transcript analyzed by 454 sequencing.

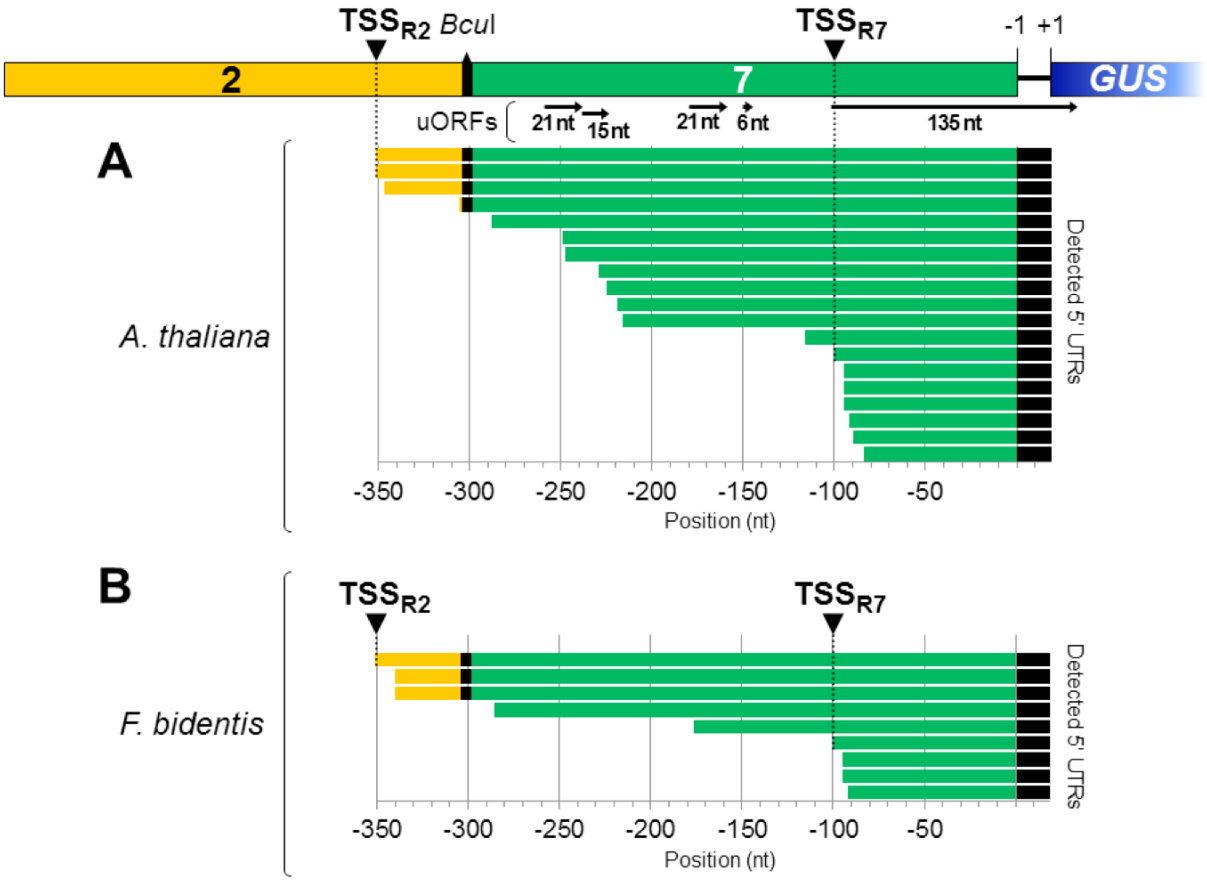
The coverage of 454 reads (Gowik et al., 2011) along the genomic sequence of the *GLDPA* gene of *Flaveria trinervia* and the resulting consensus sequence of these reads (454 transcript) are shown. The uniform distribution of reads throughout the *GLDPA* promoter sequence indicates that the 5' intron, ranging from region 3 of the *GLDPA* promoter to +17 of the *GLDPA* open reading frame, is only inefficiently spliced out. TSS, transcription start site in region 2 (TSS_{R2}) and 7 (TSS_{R7}).

Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? Plant Cell **23:** 2087–2105.



Supplemental Figure 5. Distribution of upstream open reading frames in P_{R2}-derived transcripts.

The *GLDPA* promoter and the 5' end of the open reading frame of the *GLDPA* gene (*GLDPA*-ORF) are schematically presented above. Underneath the two transcript variants originating from P_{R2} (region 2) at position -1185 (TSS_{R2}) or -1174 are shown that are incompletely spliced and still contain the 5' intron. Within this 5' intron several upstream open reading frames (uORFs; depicted by arrows) are present with their length being indicated in nucleotides (nt). The 246-nt long uORF starts directly upstream of TSS_{R7} so that the corresponding start codon is not embedded within transcripts originating from TSS_{R7}.The transcription start site in region 2 (TSS_{R2}) and 7 (TSS_{R7}) as well as the respective splice donor site (SDS) and the splice acceptor site (SAS) are highlighted by arrowheads. The major ATG codon at position +1 (ATG_{+1}) and the first ATG codon (ATG_{+25}) that follows when the 5' intron is spliced out are indicated.



Supplemental Figure 6. Analysis of mRNA 5' ends in leaves of transgenic *A. thaliana* and *F. bidentis* harboring *GLDPA*-Ft-2-7.

(A) and **(B)** The *GLDPA*-Ft-2-7 construct is depicted above with the two different transcription start sites (TSSs), TSS_{R2} and TSS_{R7}, being indicated. Underneath a random selection of 5' untranslated regions (5' UTRs) analyzed by DNA sequencing is shown. These 5' UTRs were identified by rapid amplification of 5' complementary DNA ends (5' RACE) in leaves of a transgenic *GLDPA*-Ft-2-7 *A. thaliana* plant **(A)** that exhibited *GUS* expression in the mesophyll but preferentially in bundle-sheath cells and the vasculature (see Figures 8D and 8I) or in leaves of a transgenic *GLDPA*-Ft-2-7 *F. bidentis* plant **(B)** (see Figure 7C). The arrows beneath region 7 of the *GLDPA* promoter represent upstream open reading frames (uORFs) including their respective length in nucleotides (nt) that are located at the indicated position within transcripts originating from TSS_{R2}. The 135-nt long uORF does not start at TSS_{R7} but directly upstream of it with the start codon thus not being embedded within RNAs transcribed from TSS_{R7}.

SUPPLEMENTAL METHODS

Supplemental Figures 1 and 6:

For 5' RACE total RNA was isolated from mature leaves of single *A. thaliana* and *F. bidentis* lines harboring *GLDPA*-Ft or *GLDPA*-Ft-2-7 by means of the RNeasy Plant Mini Kit (Qiagen) combined with the RNase-Free DNase Set (Qiagen). 0.5 µg (*F. bidentis*) or 1 µg (*A. thaliana*) of total RNA was directly used for cDNA first-strand synthesis by utilizing the SMARTer^{TM} RACE cDNA Amplification Kit (Clontech Laboratories) and the SMARTScribe^{TM} Reverse Transcriptase (Clontech Laboratories) according to the manufacturers' instructions. For PCR amplification of 5' UTRs with the Advantage 2 Polymerase Mix (Clontech Laboratories) or the Phusion High-Fidelity DNA Polymerase (New England Biolabs) the Universal Primer A Mix (UPM; Clontech Laboratories) and the *GUS* gene-specific 3' oligonucleotide 5RACE-3GUS-3 (5'-GTTTTCGTCGGTAATCACCATTCCC-3') or pBI121GUS3'-2 (5'-ACTGCCTGGCACAGCAATTGC-3') were used. After cloning of the obtained PCR fragments with the CloneJET^{TM} PCR Cloning Kit (Fermentas) several independent clones were analyzed by colony PCR with the oligonucleotides pJetFW (5'-ATCAACTGCTTTAACACTTGTGCC-3') and pJetRW (5'-CGGTTCCTGATGAGGTGGTTAG-3'). From single, potentially correct clones plasmid DNA was isolated for DNA sequencing.

Supplemental Figure 2:

GLDPA-Ft-1-7 was constructed by exchanging region 2-3 of *GLDPA*-Ft-2-3-7 against region 1 as *Xba*I/*Bcu*I fragment amplified by PCR with the 5' primer *GLDPA*-5'-*Xba*I (5'-TGCTCTAGAAGCTTTACTCCTCTC-3') and the 3' primer *GLDPA*1-3'-*Bcu*I (5'-TTAACTAGTCACTTTCACATTCGCCTT-3'). For cloning of *GLDPA*-Ft-1-2 region 1 and 2 together were amplified by PCR with the 5' primer *GLDPA*-5'-*Xba*I (5'-TGCTCTAGAAGCTTTACTCCTCTC-3') and the 3' primer *GLDPA*2-3'-*Xma*I (5'-TTACCCGGGGTGGAGATGATAGTTGTT-3'). This 1-2 module was ligated as *Xba*I/*Xma*I fragment into the *Xba*I/*Xma*I-restricted pBI121 vector lacking the *35S* promoter. Histochemical GUS analysis and the fluorometrical measurement of GUS activities was performed as described in methods.

Supplemental Figure 4:

The 454 sequencing reads (Gowik et al., 2011) were bioinformatically analyzed with the program CLC Genomics Workbench 4 (version 4.8, CLC bio). For this analysis the reads were mapped against the sequence of the *GLDPA* gene (GenBank accession number Z99767) with the high throughput process “map reads to reference”.

Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? Plant Cell **23:** 2087–2105.

The authors' contributions

CW wrote this manuscript and performed all experiments except those listed below.

SS performed the fluorescence microscopic analysis of roots of transgenic *Flaveria bidentis* plants harboring the *GLDPA*-Ft:*H2B-YFP* construct (Figure 1G), analyzed the histochemical GUS staining in leaves of transgenic *F. bidentis* plants harboring the *GLDPA*-Ft-2-7 construct (Figure 7C), carried out the bioinformatic analysis of *GLDPA* transcripts obtained by 454 pyrosequencing (Supplemental Figure 4), and detected mRNA 5' ends by 5' RACE in transgenic *GLDPA*-Ft plants of *F. bidentis* and *Arabidopsis thaliana* (Supplemental Figure 1 [the 5' UTR_{R7}s sequences in [B] were detected by **CW**]) as well as in transgenic *GLDPA*-Ft-2-7 plants of *F. bidentis* (Supplemental Figure 6B).

UG performed the 454 pyrosequencing of transcripts of *Flaveria trinervia* (Figure 2B and Supplemental Figure 4) and provided total RNA of leaves of *F. trinervia*.

SE generated the *GLDPA*-Ft-1-2-7 construct and provided transgenic seeds of *A. thaliana* plants which had been transformed with the construct *GLDPA*-Ft-7 by him.

MK and **MS** performed the transformation of *F. bidentis* plants.

HB provided the *GLDPA*-Ft construct.

PW and **UG** participated in drafting of the manuscript.

Manuscript 2 was submitted for publication to *The Plant Cell* (Impact Factor: 10.648).

VII. Addendum

1. The influence of the 50-bp flanking sequences of P_{R2} on gene expression

The *GLDPA* promoter of *Flaveria trinervia* (C_4) consists of two sub-promoters. The proximal promoter P_{R7}, defined by region 7, is bundle-sheath- and vasculature-specific, whereas the distal promoter P_{R2}, defined by region 2, directs uniform expression in all inner leaf tissues in both transgenic *Flaveria bidentis* (C_4) and *Arabidopsis thaliana* (C_3) plants (Manuscript 2). When combined (construct *GLDPA*-Ft-2-7), P_{R7} is able to suppress the activity of P_{R2} stably in *F. bidentis*, but only partially in *A. thaliana* (Manuscript 2; Figure Ad1). The majority of the *GLDPA*-Ft-2-7 Arabidopsis lines (60%) expressed the *GUS* reporter gene preferentially in the bundle-sheath and vasculature but additionally in the mesophyll (Figure Ad1, C and D). This expression pattern indicates the activity of both P_{R2} and P_{R7}. However, few *GLDPA*-Ft-2-7 lines (15%) exhibited specific *GUS* expression in the bundle-sheath cells and the vascular tissues (Figure Ad1D), which suggests the predominant or even sole activity of P_{R7}. In contrast, some *GLDPA*-Ft-2-7 plants (25%) showed uniform GUS staining in all inner leaf tissues (Figure Ad1D), which reflects the activity of P_{R2} alone. However, when adding the transcriptional enhancing region 1 upstream of the P_{R2}-P_{R7} module (*GLDPA*-Ft-1-2-7) all transgenic Arabidopsis plants exhibited uniform *GUS* expression in all inner leaf tissues. Thus, P_{R2} appears to overcome the partially repressive function of P_{R7} in the presence of region 1. In contrast, specific expression in bundle-sheath cells and vascular bundles could be stably maintained by inserting region 3 in between P_{R2} and P_{R7} (*GLDPA*-Ft-2-3-7). Hence, P_{R2} is assumed to be entirely suppressed in the presence of region 3 and P_{R7} (Manuscripts 1 and 2).

These findings raise the question whether the flanking sequences of P_{R2} are important to influence – or even sufficient to alter – the gene expression pattern in combination with P_{R7} rather than the presence of the whole region 1 or 3. Thus, the removal of region 1 and 3 up- and downstream of P_{R2} respectively, as it is the case in the *GLDPA*-Ft-2-7 construct, might lead to the disruption of functional *cis*-regulatory elements at the left border of P_{R2} and region 1 (1-2 border) and at the right border of P_{R2} and region 3 (2-3 border). Addressing this question, P_{R2} was elongated 50 bp upstream into region 1 (1_{50}-2) or 50 bp downstream into region 3 ($2\text{-}3_{50}$), combined with P_{R7} and fused to the *GUS* gene resulting in *GLDPA*-Ft-1_{50}-2-7 and *GLDPA*-Ft-$2\text{-}3_{50}$-7 respectively (Figure Ad1A). The activities of these both constructs in transgenic *A. thaliana* lines were compared with those of plants harboring *GLDPA*-Ft-2-7 (Figure Ad1, B–D).

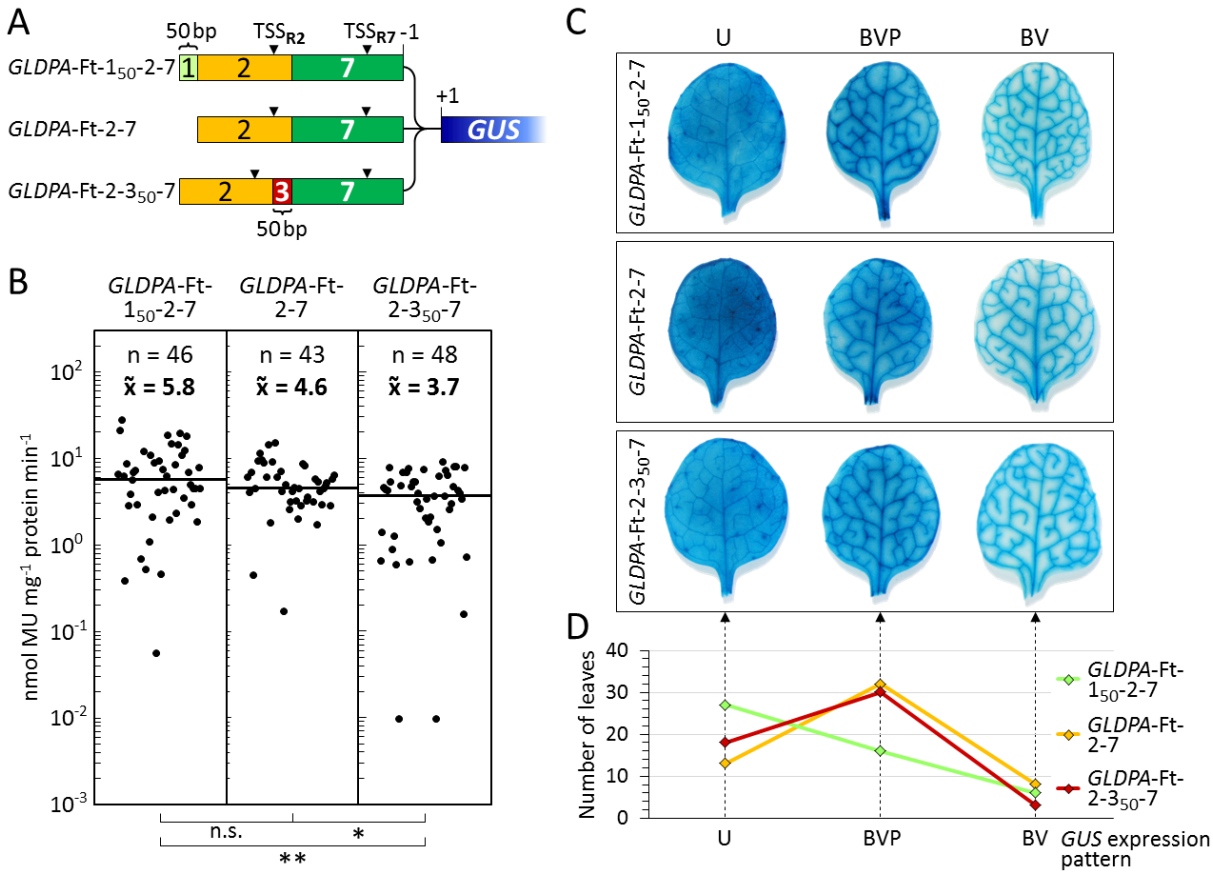


Figure Ad1. Functional analysis of the 50-bp flanking sequences of P_{R2} (region 2) in transgenic *A. thaliana*.

(A) Schematic structure of the constructs *GLDPA*-Ft-1$_{50}$-2-7, *GLDPA*-Ft-2-7 and *GLDPA*-Ft-2-3$_{50}$-7. TSS, Transcription start site.
(B) Fluorometrical measurement of GUS activities in transgenic *A. thaliana* harboring *GLDPA*-Ft-1$_{50}$-2-7, *GLDPA*-Ft-2-7 or *GLDPA*-Ft-2-3$_{50}$-7. Each single dot represents one independent transgenic line. The number of plants analyzed (n) and the median value (x̃), also added as black line in the diagram, are indicated above (n.s., not significant, $p > 0.05$; * $p < 0.05$; ** $p < 0.01$). MU, 4-Methylumbelliferone.
(C) and **(D)** Histochemical GUS staining of whole leaf blades in top view of independent transgenic Arabidopsis plants carrying *GLDPA*-Ft-1$_{50}$-2-7, *GLDPA*-Ft-2-7 or *GLDPA*-Ft-2-3$_{50}$-7. According to the observed GUS staining pattern, one single leaf blade of the independent transgenic *A. thaliana* lines was assigned to one of the three categories: uniform staining in all inner leaf tissues (U), bundles-sheath- and vasculature-preferential (BVP) or bundle-sheath- and vasculature-specific staining (BV). Incubation times for GUS staining were 20 h respectively.

Although the quantitative activities of the three constructs were very similar (Figure Ad1B), the distribution of their spatial activities differed (Figure Ad1, C and D). Most of the lines harboring *GLDPA*-Ft-2-7 (60%) exhibited GUS staining preferentially in the bundle-sheath cells and the vascular bundles, which was true for plants carrying *GLDPA*-Ft-2-3$_{50}$-7 (59%). In contrast, the majority of Arabidopsis plants with the *GLDPA*-Ft-1$_{50}$-2-7 construct (55%) showed uniform expression in all inner leaf tissues (Figure Ad1D). This indicates the predominant activity of P_{R2}. Hence, *GLDPA*-Ft-1$_{50}$-2-7 plants tend to exhibit an expression pattern like *GLDPA*-Ft-1-2-7 plants suggesting that the 50 bp upstream of P_{R2} or the 1-2 border sequence itself is already able to influence the expression pattern of the P_{R2}-P_{R7} module but cannot replace the function of the entire region 1.

Region 1 (182 bp in total) was shown to enhance the transcriptional activities of both P_{R2} and P_{R7} (Manuscript 2). Thus, the 50 bp of region 1 adjacent to P_{R2} already seem to be capable of slightly enhancing P_{R2} transcriptionally in the context of *GLDPA*-Ft-1_{50}-2-7, reflected by the predominant activity of P_{R2}. However, the 50-bp flanking sequence upstream of P_{R2} and the intact 1-2 border sequence are not sufficient for the stable maintenance of uniform expression in all inner leaf tissues in combination with P_{R2} and P_{R7} as observed for *GLDPA*-Ft-1-2-7 plants. These findings indicate that further *cis*-regulatory elements embedded within the remaining 132 bp of region 1 are essential for strong transcriptional enhancement of P_{R2} in order to maintain uniform expression in all inner leaf tissues stably in the presence of P_{R7}.

With regard to *GLDPA*-Ft-2-3_{50}-7, most of the *A. thaliana* plants (59%) exhibited GUS staining preferentially in the bundle-sheath and vasculature like 60% of *GLDPA*-Ft-2-7 plants (Figure Ad1D). Hence, the presence or absence of the 50-bp flanking sequence downstream of P_{R2} (representing the 5' end of region 3) and the 2-3 border sequence itself do not appear to have any effect on the expression pattern of the P_{R2}-P_{R7} module. However, the entire region 3 (212 bp in total) is able to maintain specific expression in the bundle-sheath cells and the vascular bundles when being inserted in between P_{R2} and P_{R7} (*GLDPA*-Ft-2-3-7). This leads to the conclusion that the remaining 162 bp, but not the 50 bp adjacent to P_{R2} including the 2-3 border sequence, are essential for cell-specific expression.

Region 3 is assumed to act as a specificity component – but only in the presence of P_{R7} – by a yet unknown mechanism (Manuscripts 1 and 2). In contrast to RNAs originating from P_{R7}, P_{R2}-derived transcripts contain sequences of region 3 (50 nt in the case of *GLDPA*-Ft-2-3_{50}-7 or 212 nt with regard to *GLDPA*-Ft-2-3-7). Region 3 contains two splice donor sites 37 nt and 103 nt away from its 5' end (Manuscript 2). Together with an appropriate splice acceptor site located in the *GUS* reporter gene, splicing might result in aberrant transcripts. According to this theory, in the presence of a splice acceptor site, the single splice donor site within the 50 nt of region 3 adjacent to P_{R2} does not suffice to initiate splicing efficiently, but both splice donor sites are needed for the repressive function of region 3. In the remaining 162 nt of region 3 an upstream open reading frame (uORF) can be found that might destabilize P_{R2}-derived transcripts and/or have an effect on translational efficiency (see chapter 5). This might explain the stable suppression of P_{R2} only in the presence of the complete region 3, together with region 7 (P_{R7}) which contains further uORFs (see chapter 5).

2. Fine mapping of the transcriptional enhancing regions 1 and 3 of the *GLDPA* promoter

Despite its weak activity, P_{R7} (region 7) alone is sufficient for specific gene expression in bundle-sheath cells and the vascular bundles. Region 1, and also region 3 – apart from its ability to stably repress P_{R2} in combination with P_{R7} – were shown to substantially enhance the transcriptional activity of P_{R7} respectively (Manuscript 2). Transgenic *A. thaliana* plants carrying *GLDPA*-Ft-1-3-7, the combination of region 1, 3 and P_{R7}, exhibited 20-fold higher GUS activities than *GLDPA*-7 plants expressing *GUS* under the control of P_{R7} alone (Figure Ad2; Manuscript 2). To identify more precisely where relevant *cis*-regulatory determinants which enhance expression strength may be located within region 1 and 3, these regions were dissected into four parts of approximately the same size with each single part overlapping 25 bp with the adjacent part(s) in order to prevent putative *cis*-regulatory elements at the borders from becoming disrupted (Figures Ad2A and Ad3A). Each part of region 1 (1.1–1.4) was then fused in front of region 3 and P_{R7} and linked to the *GUS* gene (*GLDPA*-Ft-1.1-3-7 to *GLDPA*-Ft-1.4-3-7; Figure Ad2A). The four partial fragments of region 3 (3.1–3.4) were inserted in between region 1 and P_{R7} respectively and fused to *GUS* (*GLDPA*-Ft-1-3.1-7 to *GLDPA*-Ft-1-3.4-7; Figure Ad3A). The presence of region 3 in the case of *GLDPA*-Ft-1.1-3-7 to *GLDPA*-Ft-1.4-3-7 and the presence of region 1 with respect to *GLDPA*-Ft-1-3.1-7 to *GLDPA*-Ft-1-3.4-7 should ensure a sufficient promoter activity above threshold for a successful detection of the quantitative and spatial GUS activities. As expected, all of these constructs caused a specific *GUS* expression in the bundle-sheath cells and the vascular bundles due to the presence of P_{R7} which alone suffices for gene expression in these tissues (Figures Ad2B and Ad3B).

In the presence of the parts 1.2 (*GLDPA*-Ft-1.2-3-7) and 1.3 (*GLDPA*-Ft-1.3-3-7) very similar median GUS activities could be detected compared to that of *GLDPA*-Ft-1-3-7 plants, whereas the *GLDPA*-Ft-1.1-3-7 construct was about fourfold less and *GLDPA*-Ft-1.4-3-7 approximately threefold more active (Figure Ad2C). These findings indicate that the partial fragments 1.2–1.4 are very important for enhancing P_{R7} transcriptionally with fragment 1.4 having the strongest effect. In contrast, the promoter part 1.1 does not appear to contribute to the enhancement of P_{R7} but rather represses its activity slightly, as the absence of fragment 1.1 (*GLDPA*-Ft-3-7) led to an increase in activity of about fourfold (Figure Ad2C; Manuscript 2). Nonetheless, the whole region 1 presumably contains several *cis*-regulatory elements for transcriptional enhancement of P_{R7}, but also such/those one(s) that can have a repressive effect on transcription.

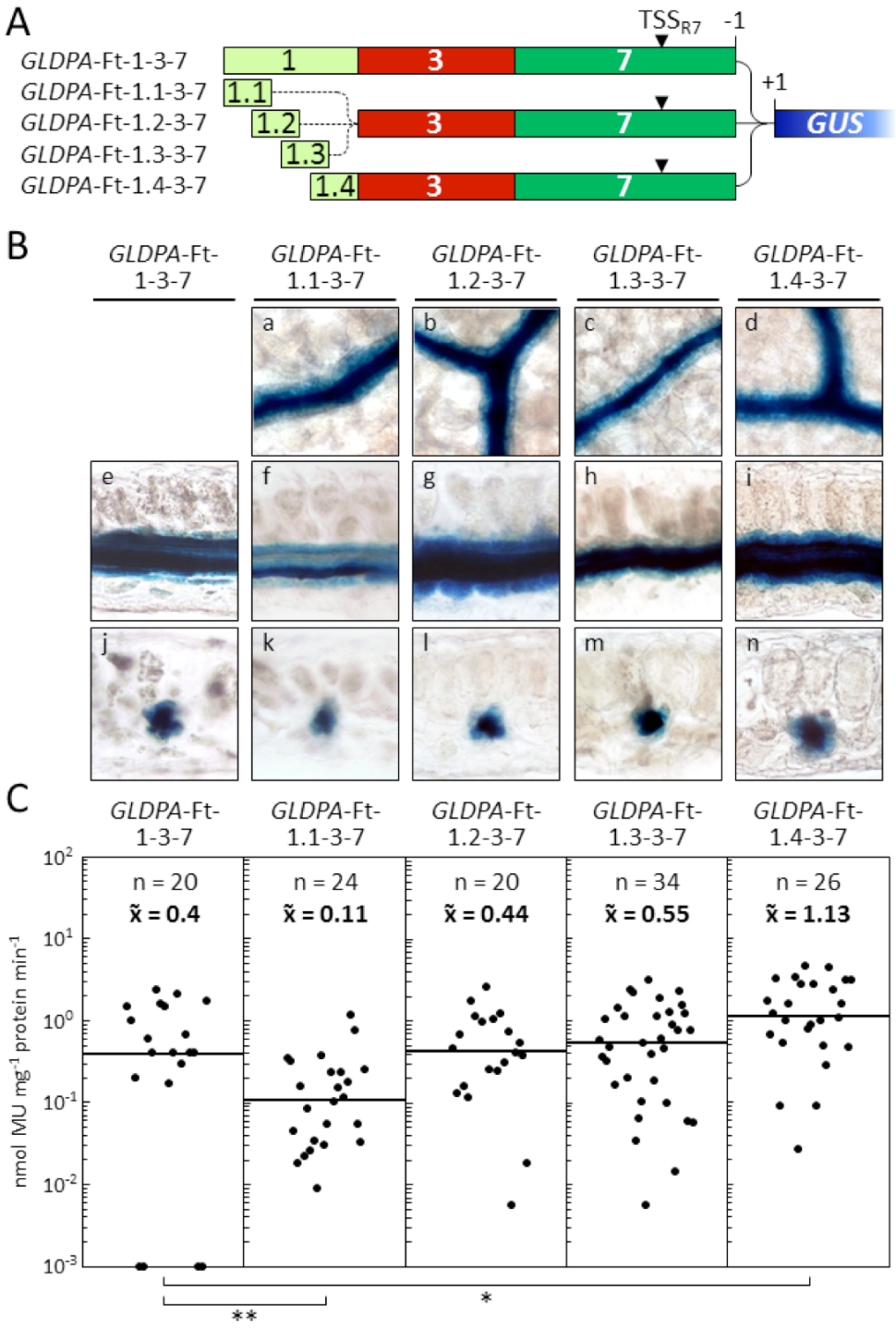


Figure Ad2. Fine mapping of the transcriptional enhancing region 1 in transgenic *A. thaliana* plants.

(A) Schematic structure of the constructs *GLDPA*-Ft-1-3-7 and *GLDPA*-Ft-1.1-3-7 to *GLDPA*-Ft-1.4-3-7. The partial fragments 1.1–1.3 and 1.4 have a size of 64 base pairs (bp) and 65 bp respectively with each part overlapping 25 bp with the adjacent part(s). TSS, Transcription start site.

(B) Histochemical localization of GUS activity in whole leaf blades in top view (a–d) and cross sections (e–n) of leaves from transgenic *A. thaliana* harboring *GLDPA*-Ft-1-3-7 (e and j), *GLDPA*-Ft-1.1-3-7 (a, f and k), *GLDPA*-Ft-1.2-3-7 (b, g and l), *GLDPA*-Ft-1.3-3-7 (c, h and m) or *GLDPA*-Ft-1.4-3-7 (d, i and n). Incubation times for GUS staining were 1.5 h (c), 3 h (b), 4 h (d, i and n), 6 h (m), 16 h (e and j), 19.5 h (a), 21 h (g and l), 23 h (h) or 46 h (f and k).

(C) Fluorometrical measurement of GUS activities in transgenic *A. thaliana* carrying the respective construct as stated above. Each single dot represents one independent transgenic line. The number of plants analyzed (n) and the median value (x̃), also added as black line in the diagram, are indicated above respectively (* $p < 0.05$; ** $p < 0.01$). MU, 4-Methylumbelliferone.

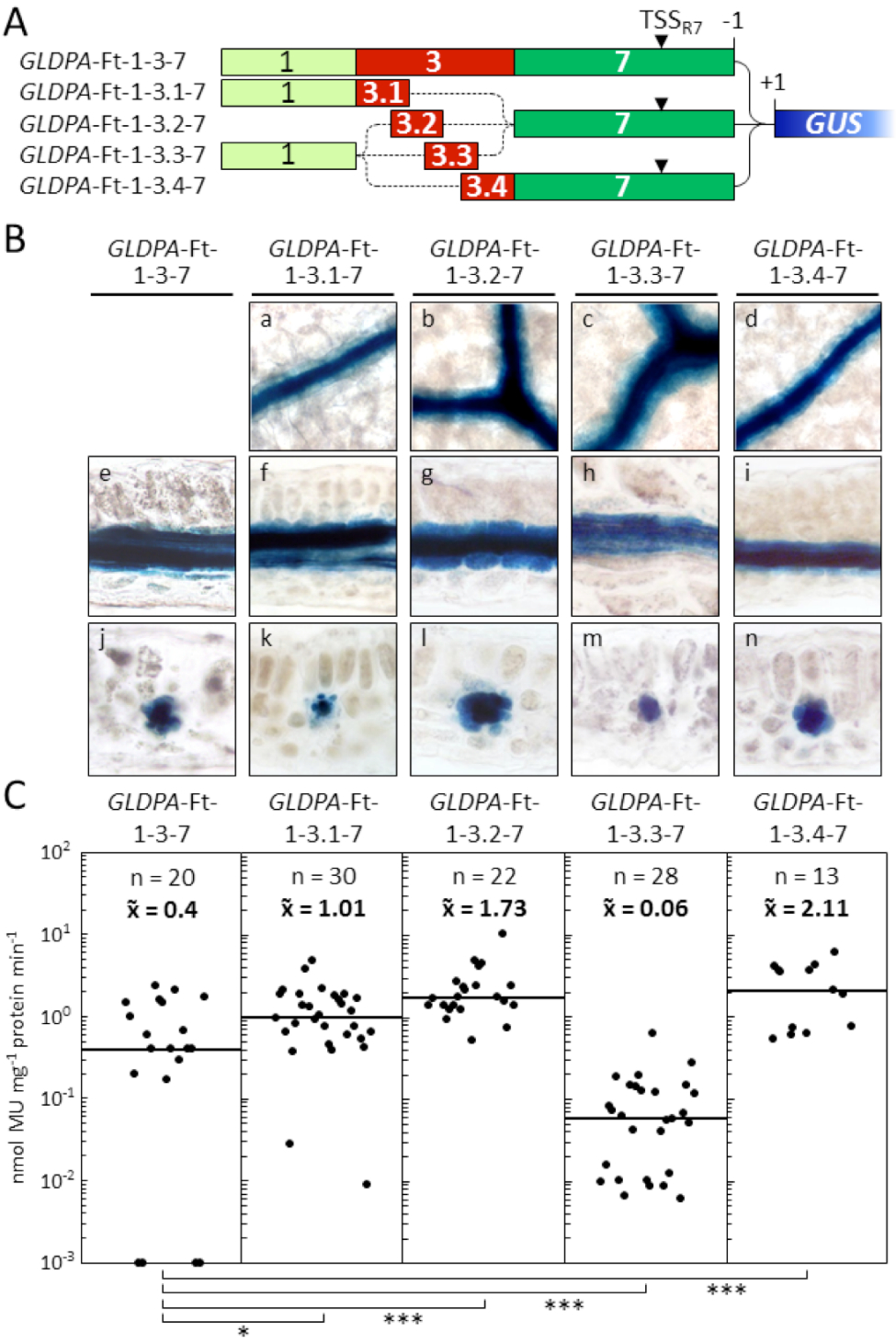


Figure Ad3. Fine mapping of the transcriptional enhancing region 3 in transgenic *A. thaliana* plants.

(A) Schematic structure of the constructs *GLDPA*-Ft-1-3-7 and *GLDPA*-Ft-1-3.1-7 to *GLDPA*-Ft-1-3.4-7. The pieces 3.1–3.3 and 3.4 have a size of 72 base pairs (bp) and 71 bp respectively with each single promoter fragment overlapping 25 bp with the adjacent fragment(s). TSS, Transcription start site.
(B) Histochemical localization of GUS activity in whole leaf blades in top view (a–d) and cross sections (e–n) of leaves from transgenic *A. thaliana* harboring *GLDPA*-Ft-1-3-7 (e and j), *GLDPA*-Ft-1-3.1-7 (a, f and k), *GLDPA*-Ft-1-3.2-7 (b, g and l), *GLDPA*-Ft-1-3.3-7 (c, h and m) or *GLDPA*-Ft-1-3.4-7 (d, i and n). Incubation times for GUS staining were 2 h (b and d), 4 h (a and n), 6 h (g, i, k and l), 16 h (e, f, h and j), 21 h (m) or 23 h (c).
(C) Fluorometrical measurement of GUS activities in transgenic *A. thaliana* carrying the respective construct as stated above. Each single dot represents one independent transgenic line. The number of plants analyzed (n) and the median value (x̃), also added as black line in the diagram, are indicated above respectively (* $p < 0.05$; *** $p < 0.001$). MU, 4-Methylumbelliferone.

A very similar situation could be observed for the effect of the single fragments 3.1–3.4 on the transcriptional activity of P_{R7}. While the pieces 3.1 (*GLDPA*-Ft-1-3.1-7), 3.2 (*GLDPA*-Ft-1-3.2-7) and 3.4 (*GLDPA*-Ft-1-3.4-7) caused a two and a half-fold, about four- and fivefold increase in activity respectively compared to the complete region 3 (*GLDPA*-Ft-1-3-7), the presence of fragment 3.3 (*GLDPA*-Ft-1-3.3-7) led to a reduction in activity of about six and a half-fold (Figure Ad3C). Furthermore, the GUS activities of *GLDPA*-Ft-1-3.3-7 Arabidopsis plants were decreased sixfold compared to that of plants carrying the *GLDPA*-Ft-1-7 construct which completely lacks region 3 (Figure Ad3C; Manuscript 2). Thus, the partial promoter fragment 3.3 seems to have a strong repressive effect on the activity of P_{R7}. These findings lead to the conclusion that region 3 also contains several *cis*-regulatory elements for the enhancement of gene expression, but additionally at least one element that represses the transcriptional activity of P_{R7}.

3. Region 2 of the *GLDPA* promoter can enhance transcription of P_{R7}

The proximal promoter P_{R7} (region 7) directs *GUS* reporter gene expression specifically in bundle-sheath cells and the vasculature in leaves of transgenic *A. thaliana* plants carrying the *GLDPA*-Ft-7 construct (Figure Ad4, B and F; Manuscript 2). This expression pattern is indistinguishable from that of the full-length *GLDPA* promoter (*GLDPA*-Ft construct) (Manuscript 1). In comparison, transgenic Arabidopsis plants harboring the *GLDPA*-Ft construct exhibited about 1500-fold higher GUS activities than transgenic *GLDPA*-Ft-7 plants (Manuscripts 1 and 2). These findings demonstrate that the activity of P_{R7} is drastically reduced in the absence of the *GLDPA* promoter regions 1 to 6.

In contrast to P_{R7}, the distal promoter P_{R2} (region 2) alone is strong and shows activity in all inner leaf tissues in transgenic *A. thaliana* plants. The combination of P_{R2} and P_{R7} (*GLDPA*-Ft-2-7) led to a variety of different *GUS* expression patterns throughout the independent transgenic Arabidopsis lines (Figure Ad4, C–E and G–I; Manuscript 2). This suggests that both promoters were active, but that the output of each promoter varied from line to line. Among the independent *GLDPA*-Ft-2-7 lines, there were also plants that exhibited specific *GUS* expression in the bundle-sheath and vasculature (Figure Ad4, E and I). Although these lines expressed *GUS* in the same pattern as plants that carried the *GLDPA*-Ft-7 construct (Figure Ad4, B and F), the GUS activity was increased 180-fold in the presence of region 2 (Figure Ad4J). This leads to the conclusion that region 2 (P_{R2}) can enhance the transcriptional activity of P_{R7} efficiently.

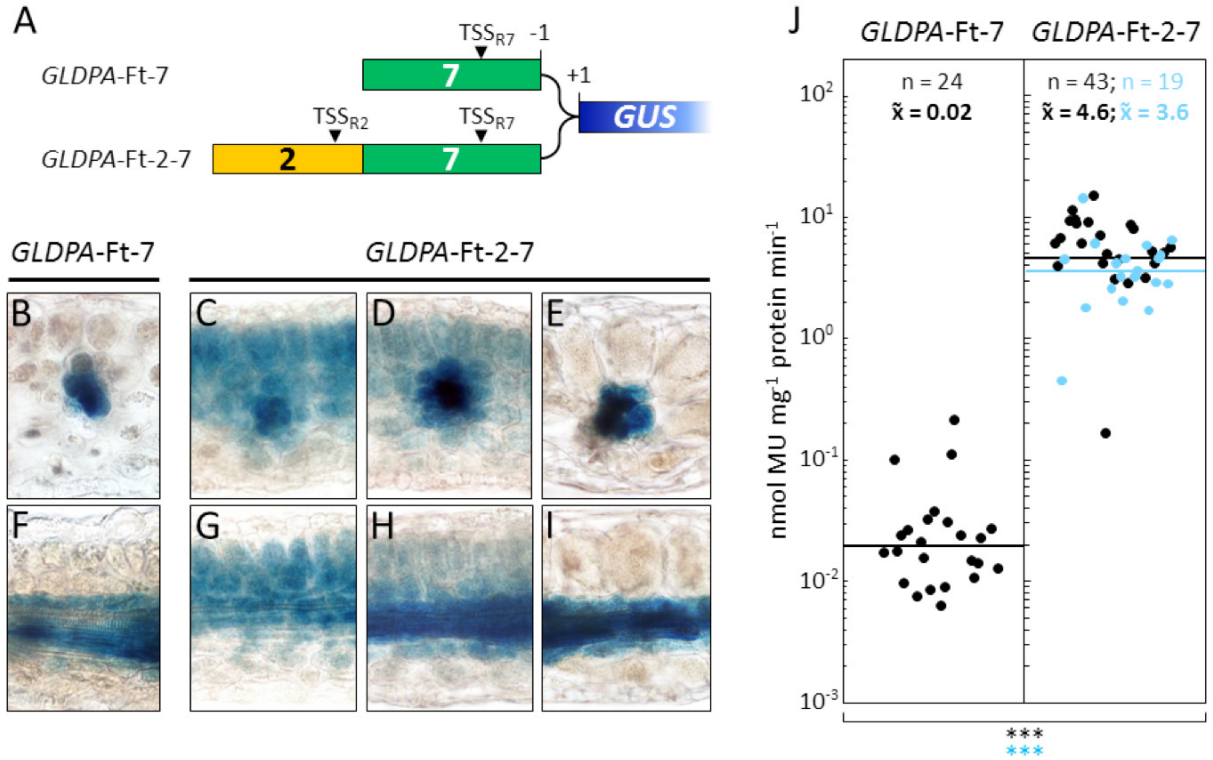


Figure Ad4. Region 2 of the *GLDPA* promoter enhances the activity of P_{R7} in transgenic *A. thaliana* plants.

(A) Schematic structure of the constructs *GLDPA*-Ft-7 and *GLDPA*-Ft-2-7. TSS, Transcription start site.
(B) to **(I)** Histochemical GUS staining in cross sections (B–I) of leaves of four independent transgenic *A. thaliana* lines (B/F, C/G, D/H and E/I) harboring *GLDPA*-Ft-7 (B and F) or *GLDPA*-Ft-2-7 (C–E and G–I). Incubation times for the GUS staining procedure were 3.5 h (C, D, G and H), 4 h (E and I) and 17 h (B and F).
(J) Fluorometrical quantification of GUS activities of Arabidopsis plants transformed with *GLDPA*-Ft-7 or *GLDPA*-Ft-2-7. Each single dot represents one independent transgenic line. With regard to *GLDPA*-Ft-2-7, blue dots represent plants exhibiting specific expression in bundle-sheath cells and vascular bundles as displayed in E/I, whereas the remaining black dots can be allocated to one of the other expression patterns (C/G and D/H). The number of all lines examined (n) and the median of all values (x̃) are indicated at the top of the diagram in black. The number of those plants that exhibited specific expression (blue dots) and their median value is highlighted in blue. Additionally, the median value of all plants analyzed and that one of the plants with specific staining in the bundle-sheath and vasculature (blue dots) are indicated as black or blue line in the diagram respectively. The statistical evaluation included all values (black asterisks) or only the values (blue dots) of plants with specific GUS expression (blue asterisks) (*** $p < 0.001$). MU, 4-Methylumbelliferone.

Transgenic *F. bidentis* plants that harbored the *GLDPA*-Ft-7 construct exhibited weak GUS activity specifically in the bundle-sheath and vasculature (Figure Ad5, A, B, D and F; Manuscript 2). In contrast to Arabidopsis, in *F. bidentis* the addition of region 2 (P_{R2}) upstream of region 7 (P_{R7}) (construct *GLDPA*-Ft-2-7) did not change the *GUS* expression pattern, but increased the GUS activity about 390-fold (Figure Ad5, A, C, E and F). These findings suggest that region 2 has the ability to enhance the transcriptional activity of P_{R7} in both *A. thaliana* and *F. bidentis*.

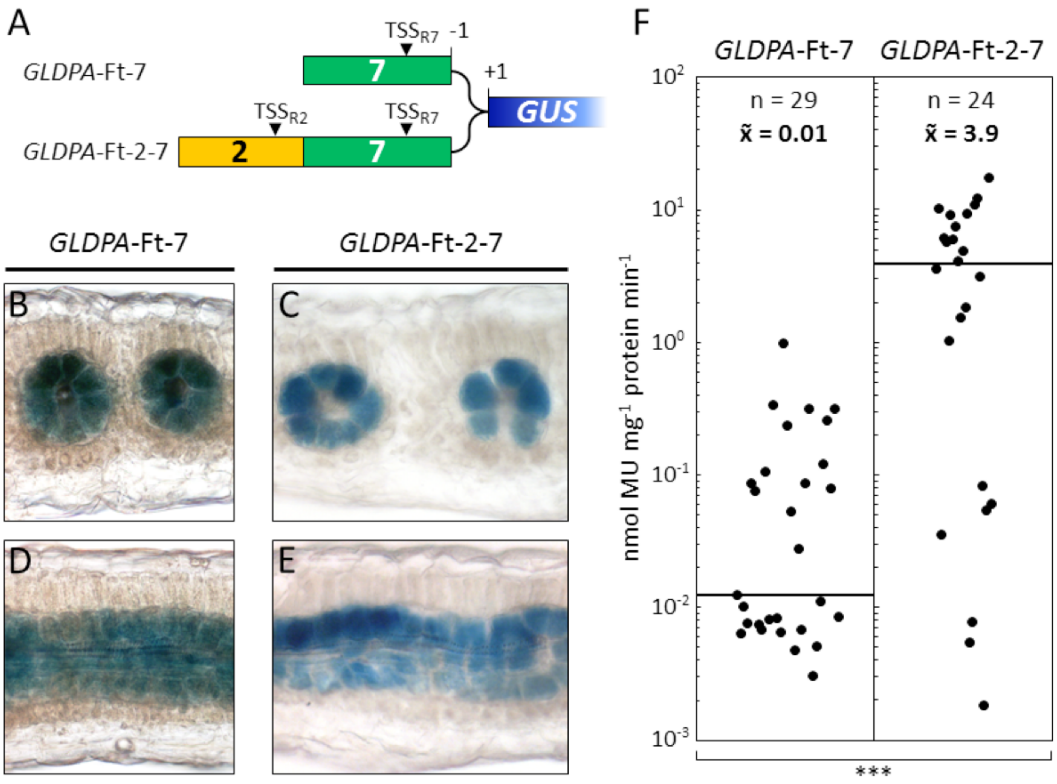


Figure Ad5. Region 2 of the *GLDPA* promoter enhances the activity of P_{R7} in transgenic *F. bidentis* plants.

(A) Schematic structure of the constructs *GLDPA*-Ft-7 and *GLDPA*-Ft-2-7. TSS, Transcription start site.
(B) to **(E)** Histochemical GUS staining in cross sections (B–E) of leaves of transgenic *F. bidentis* plants harboring *GLDPA*-Ft-7 (B and D) or *GLDPA*-Ft-2-7 (C and E). Incubation times for the GUS staining procedure were 6 h (C and E), 43 h (B) and 66 h (D).
(F) Fluorometrical quantification of GUS activities in leaves of *F. bidentis* plants transformed with *GLDPA*-Ft-7 or *GLDPA*-Ft-2-7. Each single dot represents one independent transgenic line. The number of all lines examined (n) and the median of all values (x̃), also added as black line in the diagram, are indicated above the single values respectively (*** $p < 0.001$). MU, 4-Methylumbelliferone.

4. The position of region 3 influences the output of P_{R2}

Region 3 of the *GLDPA* promoter is able to enhance the transcriptional activity of P_{R2} without affecting its spatial activity in all inner leaf tissues (Manuscript 2; Figure Ad6). This enhancing effect appears to be stronger when region 3 is located upstream of P_{R2} (constructs *GLDPA*-Ft-3-2 and *GLDPA*-Ft-3-1-2) rather than being downstream of it (constructs *GLDPA*-Ft-2-3 and *GLDPA*-Ft-1-2-3). Transgenic Arabidopsis plants carrying *GLDPA*-Ft-3-2 exhibited 1.6-fold higher GUS activities than *GLDPA*-Ft-2-3 lines (Manuscript 2). The addition of the transcriptional enhancing region 1 (constructs *GLDPA*-Ft-1-2-3 and *GLDPA*-Ft-3-1-2; Figure Ad6A) raised the GUS activities generally, but *A. thaliana* plants with the *GLDPA*-Ft-3-1-2::*GUS* transgene were 2.7-fold more active than plants harboring the *GLDPA*-Ft-1-2-3::*GUS* transgene (Figure Ad6F). These findings suggest that the position of

region 3 influences its ability to enhance the activity of P_{R2}. The enhancing effect of region 3 appears to be weaker when it is located downstream of P_{R2}.

In *A. thaliana*, region 3 of the *GLDPA* promoter might be involved in the post-transcriptional regulation of gene expression (see chapter 5), apart from its function to enhance the transcriptional activities of P_{R2} and P_{R7}. When region 3 is located downstream of P_{R2} (constructs *GLDPA*-Ft-2-3 and *GLDPA*-Ft-1-2-3), all RNAs that are transcribed from P_{R2} include the sequence of region 3 and might be less stable or less efficiently translated than transcripts that start from P_{R2} but lack region 3 (constructs *GLDPA*-Ft-3-2 and *GLDPA*-Ft-3-1-2). This indicates that the observed positional effect of region 3 on the output of P_{R2} may be caused post-transcriptionally rather than transcriptionally.

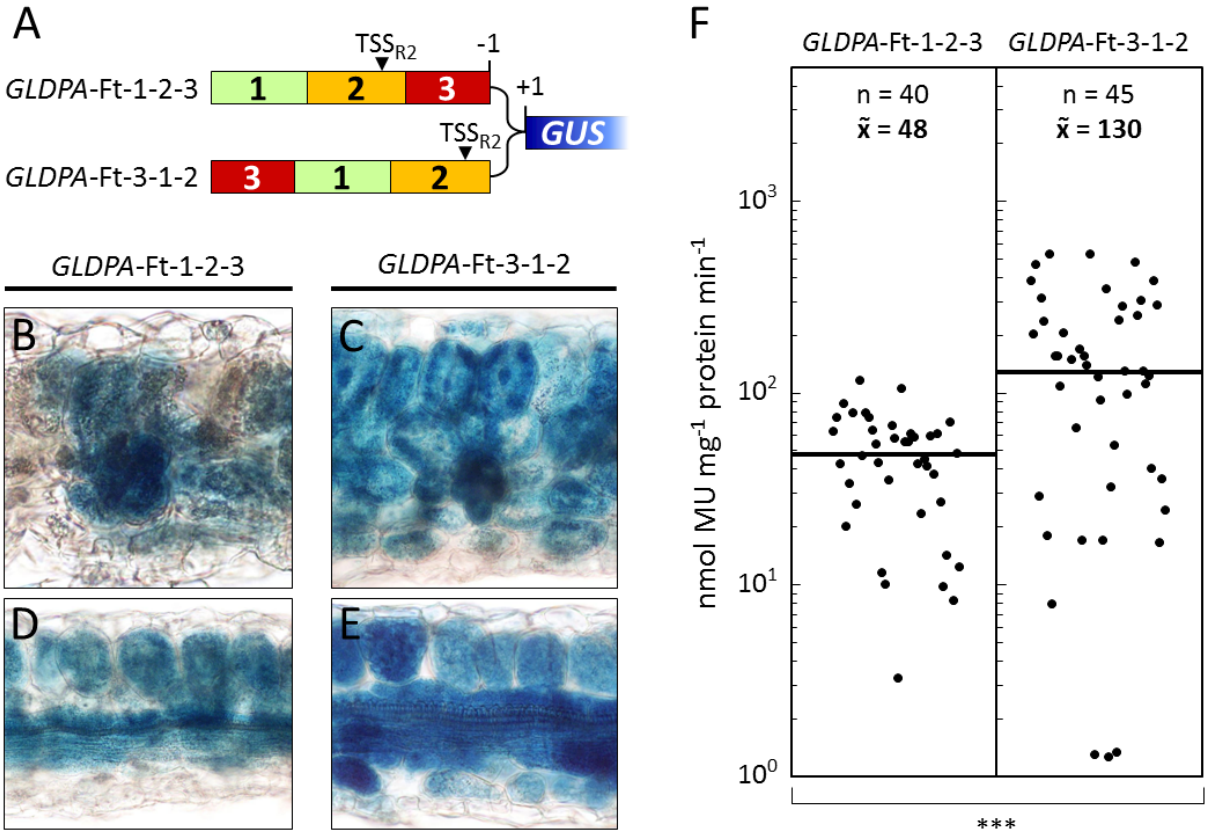


Figure Ad6. The positional effect of region 3 on the output of region 2 (P_{R2}) in transgenic *A. thaliana* plants.

(A) Schematic structure of the constructs *GLDPA*-Ft-1-2-3 and *GLDPA*-Ft-3-1-2. TSS, Transcription start site.
(B) to **(E)** Histochemical GUS staining in cross sections (B–E) of leaves of transgenic *A. thaliana* lines carrying *GLDPA*-Ft-1-2-3 (B and D) or *GLDPA*-Ft-3-1-2 (C and E). Incubation times for the GUS staining procedure were 0.5 h (D), 1 h (B and C) and 2 h (E).
(F) Fluorometrical quantification of GUS activities of Arabidopsis plants transformed with *GLDPA*-Ft-1-2-3 or *GLDPA*-Ft-3-1-2. Each single dot represents one independent transgenic line. The number of all lines examined (n) and the median of all values ($\tilde{x}$), also added as black line in the diagram, are indicated at the top of the diagram respectively (*** $p < 0.001$). MU, 4-Methylumbelliferone.

5. P_{R2}-derived RNAs are destabilized in the presence of P_{R7}

P_{R7} alone is weak and specific, whereas P_{R2} alone shows strong and uniform activity in all inner leaf tissues. In combination, the proximal promoter P_{R7} is able to suppress the distal promoter P_{R2}. The output of P_{R2} appears to be efficiently reduced post-transcriptionally by nonsense-mediated mRNA decay (NMD) (Manuscript 2). NMD is a eukaryotic control mechanism that detects and degrades mRNAs which contain premature termination codons (Chang et al., 2007; Brogna and Wen, 2009). In plants, upstream open reading frames (uORFs) of the 5' untranslated region (5' UTR) can trigger NMD when the encoded protein exceeds the critical size of 35 amino acids (Nyikó et al., 2009).

When P_{R2} is combined with P_{R7} (construct *GLDPA*-Ft-2-7), RNAs transcribed from P_{R2} contain several uORFs and one of these has a length of 135 nucleotides that might give rise to a protein of 45 amino acids (Figure Ad7; Manuscript 2). To find out whether these RNAs might become degraded, possibly by NMD, 5' RACE analyses were performed with various transgenic *A. thaliana* and *F. bidentis* plants carrying the *GLDPA*-Ft-2-7 construct (Figures Ad7 and Ad9). In a transgenic *GLDPA*-Ft-2-7 *A. thaliana* line, RNA 5' ends could be detected starting from either TSS_{R2} (P_{R2}) or TSS_{R7} (P_{R7}) (Figures Ad7A and Ad9B). However, many transcripts with 5' UTRs of different lengths were found that apparently started randomly between TSS_{R2} and TSS_{R7}. These transcripts presumably had all been transcribed from TSS_{R2} but might have become degraded due to RNA instability. The effect of possibly degraded transcripts originating from TSS_{R2} in the presence of P_{R7} was also observed for a transgenic *F. bidentis* plant that harbored the *GLDPA*-Ft-2-7 construct (Figures Ad7B and Ad9B). In Arabidopsis, 5' RACE analyses did not reveal a general correlation between the observed activities of P_{R2} and P_{R7}, based on the detected GUS staining pattern, and their transcriptional activities. With respect to the RNA output, P_{R7} always appeared to be the dominant or even the only active promoter regardless of the detected *GUS* expression pattern of the respective *GLDPA*-Ft-2-7 line (Figure Ad9B). These findings indicate that transcripts originating from P_{R2} are favoured for degradation in the presence of P_{R7}. At least one uORF beyond the critical size in P_{R2}-derived transcripts might elicit NMD and thereby lead to RNA degradation.

Transgenic Arabidopsis plants expressing *GUS* under the control of P_{R2} or P_{R7} alone showed RNA 5' ends of more or less the same length (Figures Ad8 and Ad9B) suggesting that transcripts from P_{R2} and P_{R7} seem to be stable. In both cases no uORF is present in P_{R2}- or P_{R7}-derived transcripts, so that NMD is not expected to occur.

Apart from their destabilizing effect on transcripts, uORFs affect translational efficiency. Translation of the major ORF is even strongly reduced when it overlaps with a uORF (Kozak,

2002; Kozak, 2005; Sachs and Geballe, 2006). This is true for the 135-nt long uORF that starts upstream of TSS_{R7} and extends into the ORF of the *GUS* gene (Figure Ad7).

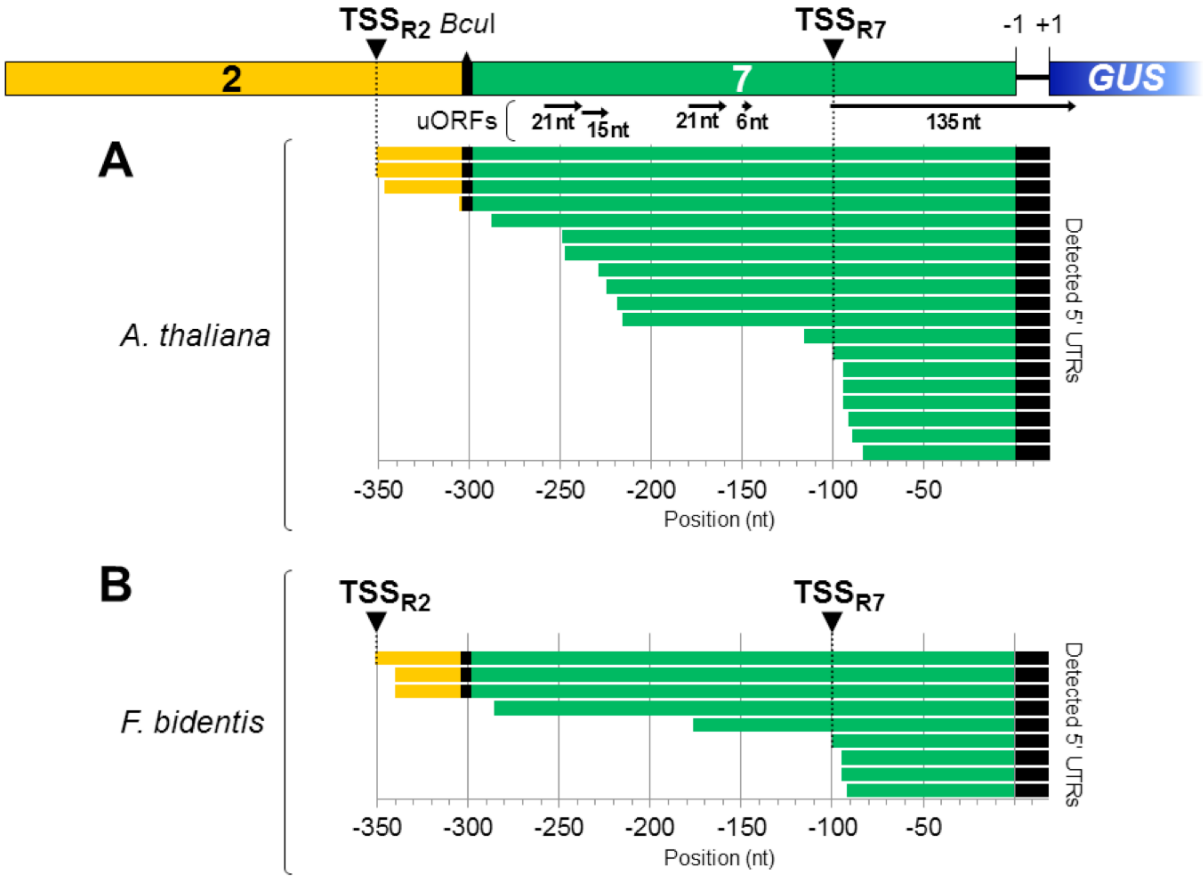


Figure Ad7. Analysis of mRNA 5' ends in leaves of transgenic *A. thaliana* and *F. bidentis* harboring *GLDPA*-Ft-2-7.

(A) and **(B)** The *GLDPA*-Ft-2-7 construct is depicted above with the two different transcription start sites (TSSs), TSS_{R2} and TSS_{R7}, indicated by an arrowhead. Underneath a random selection of 5' untranslated regions (5' UTRs) analyzed by DNA sequencing is shown. These 5' UTRs were identified by rapid amplification of 5' complementary DNA ends (5' RACE) in leaves of a transgenic *GLDPA*-Ft-2-7 *A. thaliana* plant (A) that exhibited *GUS* expression in the mesophyll but preferentially in bundle-sheath cells and the vasculature (see Figure Ad9B, lanes 6 and 7) or in leaves of a transgenic *GLDPA*-Ft-2-7 *F. bidentis* plant (B) (see Figure Ad9B, lane 1). The arrows beneath region 7 of the *GLDPA* promoter represent upstream open reading frames (uORFs) including their respective length in nucleotides (nt) which are located at the indicated position within transcripts originating from TSS_{R2}. The 135-nt long uORF does not start at TSS_{R7} but directly upstream of it with the start codon thus not being embedded within RNAs transcribed from TSS_{R7}.

In *F. bidentis*, P_{R2} seems to be stably repressed by P_{R7} because all transgenic *GLDPA*-Ft-2-7 plants exhibited *GUS* expression exclusively in the bundle-sheath and the vasculature. In contrast, the combination of P_{R2} and P_{R7} led to various *GUS* expression patterns in transgenic *A. thaliana* plants (Figures Ad4 and Ad5; Manuscript 2). Bundle-sheath- and vasculature-specific *GUS* expression could be stably maintained by adding region 3 in between P_{R2} and P_{R7} (construct *GLDPA*-Ft-2-3-7) (Manuscript 2). The fusion of region 3 up- or downstream to

P_{R2} in the absence of P_{R7} did not alter the expression pattern from unspecific to specific (Figure Ad6), which led to the conclusion that region 3 is only functional in the presence of P_{R7} and supports the partially repressive activity of P_{R7} on P_{R2} (Manuscript 2). In transgenic Arabidopsis plants harboring *GLDPA*-Ft-2-3-7, no RNA 5' ends transcribed from P_{R2} could be detected by 5' RACE, but only those starting from P_{R7} (Figure Ad9B). Although region 3 contains a uORF, this one has a length of only six nucleotides (Manuscript 2) and presumably does not promote NMD efficiently. Region 3 was shown to enhance the transcriptional activity of both P_{R2} and P_{R7} (Figure Ad6; Manuscript 2). The strong enhancement of P_{R7} in the presence of region 3 combined with the decay of RNAs transcribed from P_{R2} might explain the fact that all *GLDPA*-Ft-2-3-7 *A. thaliana* plants exhibited bundle-sheath- and vasculature-specific *GUS* expression. Apart from the transcriptional and possibly post-transcriptional regulation, translational efficiency might be further reduced by the additional small uORF in region 3 in *GLDPA*-Ft-2-3-7 plants compared to *GLDPA*-Ft-2-7 lines. The presence of the two splice donor sites within region 3 (Manuscript 2) might induce splicing in the presence of an appropriate splice acceptor site within the *GUS* reading frame and result in aberrant and/or destabilized transcripts, which also affects translational efficiency.

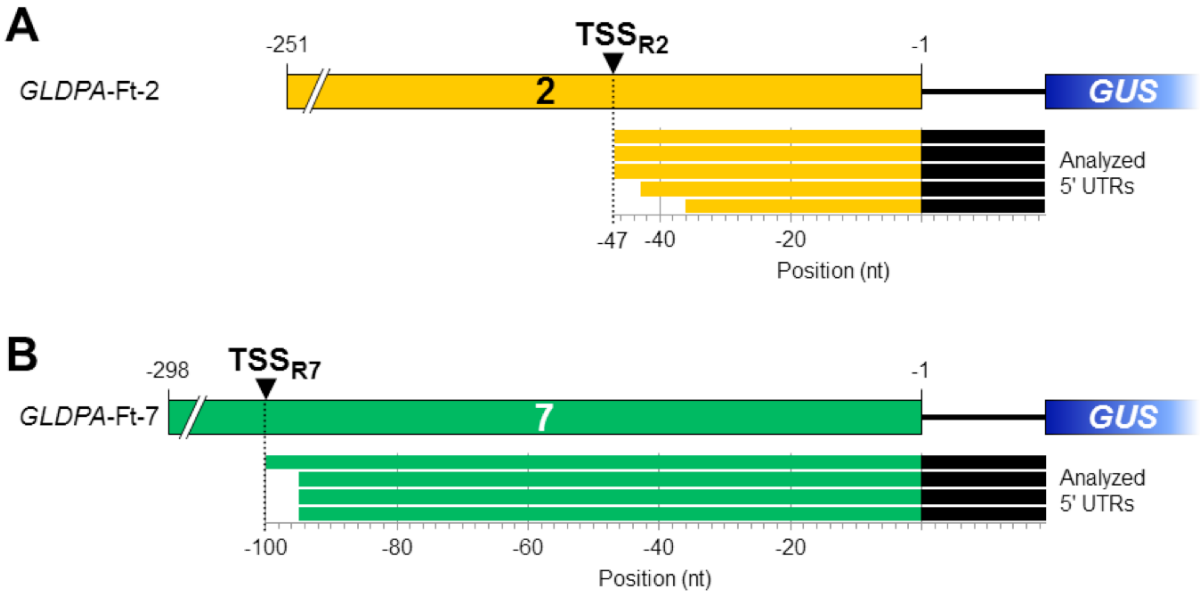


Figure Ad8. Analysis of mRNA 5' ends in leaves of transgenic *A. thaliana* harboring *GLDPA*-Ft-2 or *GLDPA*-Ft-7.

(A) and **(B)** Identification of 5' untranslated regions (5' UTRs) by the rapid amplification of 5' complementary DNA ends (5' RACE) in leaves of transgenic *GLDPA*-Ft-2 (A) or *GLDPA*-Ft-7 (B) *A. thaliana* plants (see Figure Ad9B, lanes 10–14). Five (A) and four (B) cloned 5' RACE products were randomly selected and analyzed by DNA sequencing. The respective construct is depicted schematically above the detected 5' UTR sequences with the corresponding transcription start site (TSS) within region 2 (TSS_{R2}) or region 7 (TSS_{R7}) indicated by an arrowhead. nt, nucleotide.

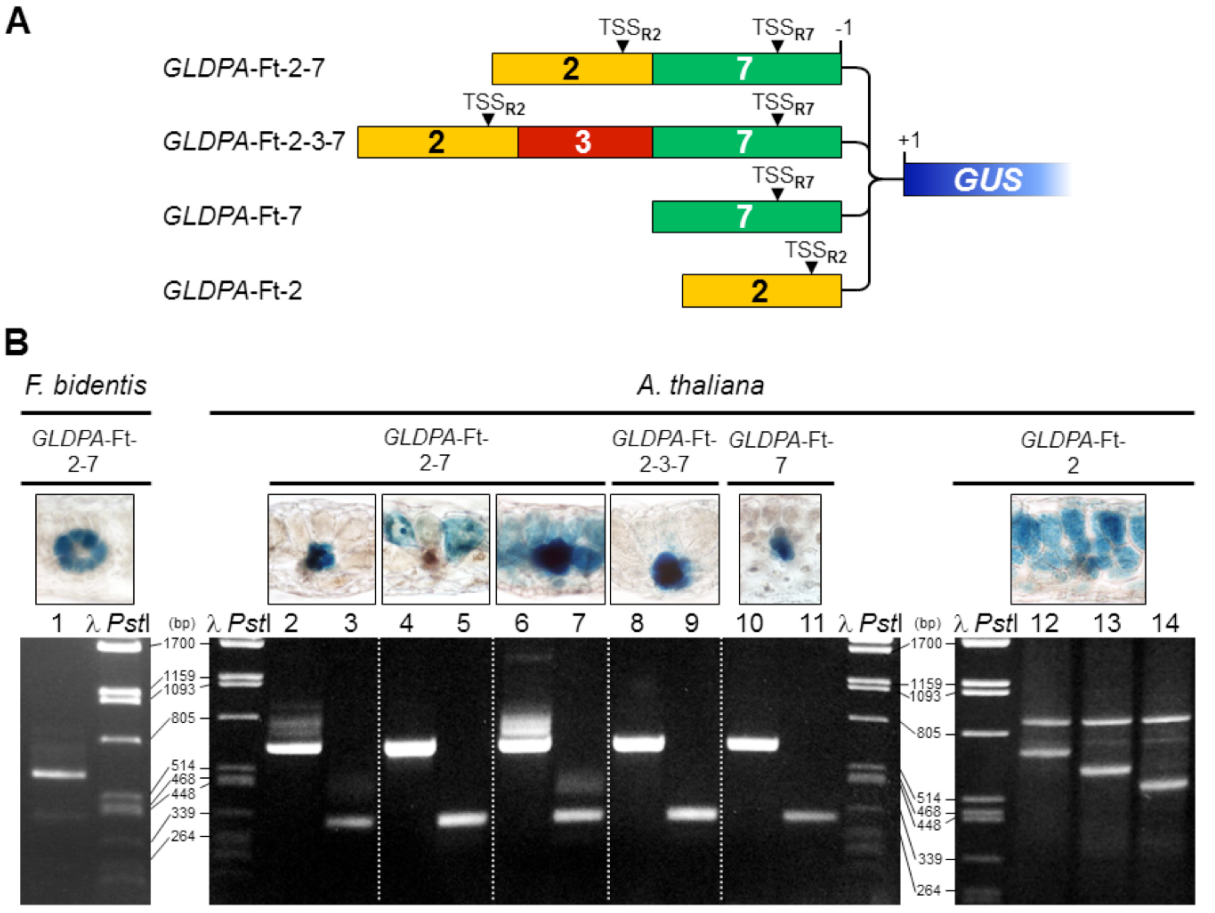


Figure Ad9. Analysis of mRNA 5' ends in leaves of various transgenic *A. thaliana* and *F. bidentis* lines.

(A) Schematic structure of the constructs *GLDPA*-Ft-2-7, *GLDPA*-Ft-2-3-7, *GLDPA*-Ft-7 and *GLDPA*-Ft-2 with indication of the transcription start sites (TSSs) within region 2 (TSS_{R2}) or region 7 (TSS_{R7}).
(B) Electrophoretic separation of PCR products in agarose gels amplified during rapid amplification of 5' complementary DNA ends (5' RACE) from cDNA of a transgenic *F. bidentis* line carrying *GLDPA*-Ft-2-7 (lane 1) and of transgenic *A. thaliana* plants harboring *GLDPA*-Ft-2-7 (lanes 2–7), *GLDPA*-Ft-2-3-7 (lanes 8 and 9), *GLDPA*-Ft-7 (lanes 10 and 11) or *GLDPA*-Ft-2 (lanes 12–14). GUS staining within a leaf cross section of the respective line is displayed above. The distinct band (lanes 1–11) represents RNA 5' ends originating from TSS_{R7} with the differences in size (lanes 2, 4, 6, 8 and 10 vs. lanes 3, 5, 7, 9 and 11) caused by using different primer combinations. The smear in lanes 2, 3, 6 and 7 represents RNA 5' ends presumably generated at TSS_{R2}. The distinct lower band (lanes 12–14) represents 5' ends of transcripts originating from TSS_{R2} with the differences in size caused by using different primer combinations, whereas the other detected fragments are unspecific PCR products. Selected 5' RACE products (lanes 1, 6, 10 and 13) were verified by DNA sequencing (see Figures Ad7 and Ad8). 5 µl PCR product was loaded on the gels respectively. bp, base pairs; λ *Pst*I, *Pst*I-restricted Lambda DNA.

6. Material and Methods

Generation of promoter-reporter gene constructs

The amplification, electrophoretic separation in agarose gels, and the cloning of DNA was accomplished according to Sambrook and Russell (2001). The generation and the cloning of the various promoter-reporter gene constructs were carried out as described in Manuscript 2.

All *GLDPA* promoter regions were amplified by PCR by means of the corresponding oligonucleotides containing respective restriction sites (Tables Ad1 and Ad2). The cloning of the constructs *GLDPA*-Ft-2, *GLDPA*-Ft-7, *GLDPA*-Ft-2-7 and *GLDPA*-Ft-2-3-7 has been previously described (Manuscript 2). *GLDPA*-Ft-1_{50}-2-7 and *GLDPA*-Ft-2-3_{50}-7 were constructed by exchanging the 2-3 module of *GLDPA*-Ft-2-3-7 against region 2 (P_{R2}) elongated 50 base pairs upstream into region 1 (1_{50}-2) or downstream into region 3 (2-3_{50}) as *Xba*I/*Bcu*I fragment respectively. *GLDPA*-Ft-1-3-7 was generated by cloning region 1 as *Xba*I/*Bam*HI and region 3 as *Bam*HI/*Bcu*I fragment together into pBluescript II KS(+) (Stratagene) so that the 1-3 module could be excised as *Xba*I/*Bcu*I fragment in order to be ligated with region 7 (P_{R7}) as *Bcu*I/*Xma*I fragment into the *Xba*I/*Xma*I-restricted pBI121 vector (Clontech Laboratories) lacking the *CaMV 35S* promoter. *GLDPA*-Ft-1.1-3-7, *GLDPA*-Ft-1.2-3-7, *GLDPA*-Ft-1.3-3-7 and *GLDPA*-Ft-1.4-3-7 were constructed by replacing region 1 of *GLDPA*-Ft-1-3-7 by its partial regions 1.1 (64 bp), 1.2 (64 bp), 1.3 (64 bp) or 1.4 (65 bp), each one overlapping 25 bp with the adjacent part(s), as *Xba*I/*Bam*HI fragment respectively. *GLDPA*-Ft-1-3.1-7, *GLDPA*-Ft-1-3.2-7, *GLDPA*-Ft-1-3.3-7 and *GLDPA*-Ft-1-3.4-7 were generated by exchanging region 3 of *GLDPA*-Ft-1-3-7 against its partial regions 3.1 (72 bp), 3.2 (72 bp), 3.3 (72 bp) or 3.4 (71 bp), each one overlapping 25 bp with the adjacent part(s), as *Bam*HI/*Bcu*I fragment respectively. The *GLDPA*-Ft-1-3-7 construct (Figure Ad2, A and C; Figure Ad3, A and C) was composed of region 1 as *Xba*I/*Xho*I, region 3 as *Xho*I/*Hind*III and region 7 (P_{R7}) as *Hind*III/*Xma*I fragment, and was cloned by Sascha Engelmann. To generate the *GLDPA*-Ft-1-2-3 construct the regions 1, 2 and 3 were amplified as 1-2-3 module and ligated as *Xba*I/*Xma*I fragment into the *Xba*I/*Xma*I-restricted pBI121 vector lacking the *35S* promoter. Region 3 was cloned as *Xba*I/*Xba*I fragment into *GLDPA*-Ft-1-2 (Manuscript 2) previously restricted with *Xba*I to generate *GLDPA*-Ft-3-1-2. The correct orientation of region 3 was verified by DNA sequencing.

Transformation of *Arabidopsis thaliana* and *Flaveria bidentis*

Transgenic plants of *A. thaliana* and *F. bidentis* were generated according to Manuscript 2.

***In situ* analysis of the β-glucuronidase and fluorometrical detection of its activity**

The *in situ* analysis of the β-glucuronidase (GUS) and the fluorometrical measurement of the GUS activity were performed as described in Manuscript 2. The generation of transgenic Arabidopsis plants harboring *GLDPA*-Ft-1-3-7 and the histochemical and fluorometrical detection of GUS activities in leaves of these plants was carried out by Sascha Engelmann.

The histochemical GUS staining in leaves of transgenic *F. bidentis* plants carrying the *GLDPA*-Ft-2-7 construct, and the fluorometrical measurement of GUS activities in leaves of these lines were performed by Stefanie Schulze.

Analysis of mRNA 5' ends by rapid amplification of 5' complementary DNA ends

For rapid amplification of 5' complementary DNA ends (5' RACE), total RNA was isolated from mature rosette leaves of single independent *A. thaliana* lines carrying *GLDPA*-Ft-7, *GLDPA*-Ft-2-7 or *GLDPA*-Ft-2-3-7 or from mature leaves of a transgenic *F. bidentis* plant harboring *GLDPA*-Ft-2-7 by means of the RNeasy Plant Mini Kit (Qiagen) combined with the RNase-Free DNase Set (Qiagen). 0.5 µg (*F. bidentis*) or 1 µg (*A. thaliana*) of total RNA was directly used for cDNA first-strand synthesis by utilizing the SMARTer^{TM} RACE cDNA Amplification Kit (Clontech Laboratories) and the SMARTScribe^{TM} Reverse Transcriptase (Clontech Laboratories) according to the manufacturers' instructions. For PCR amplification of 5' untranslated regions (5' UTRs) with the Advantage 2 Polymerase Mix (Clontech Laboratories) or the Phusion High-Fidelity DNA Polymerase (New England Biolabs) the Universal Primer A Mix (UPM; Clontech Laboratories) and the *GUS* gene-specific 3' oligonucleotide 5RACE-3GUS-3 (5'-GTTTTCGTCGGTAATCACCATTCCC-3') or pBI121GUS3'-2 (5'-ACTGCCTGGCACAGCAATTGC-3') were used. After cloning of the obtained PCR fragments with the CloneJET^{TM} PCR Cloning Kit (Fermentas), several independent clones were analyzed by colony PCR with the oligonucleotides pJetFW (5'-ATCAACTGCTTTAACACTTGTGCC-3') and pJetRW (5'-CGGTTCCTGATGAGGTGGTTAG-3'). From single, potentially correct clones plasmid DNA was isolated for DNA sequencing. The isolation of RNA from leaves of a single transgenic *GLDPA*-Ft-2-7 plant of *F. bidentis* and the analysis of these transcripts by 5' RACE was performed by Stefanie Schulze.

5' RACE of a single independent *A. thaliana* line carrying *GLDPA*-Ft-2 was performed as described before with the exception that the SMART^{TM} RACE cDNA Amplification Kit (Clontech Laboratories) was used. The SMART-II-A-Primer (5'-AAGCAGTGGTATCAACGCAGAGT-3') in combination with either the *GUS* gene-specific 3' primer 5RACE-3GUS-1 (5'-AACAGACGCGTGGTTACAGTCTTGC-3'), 5RACE-3GUS-2 (5'-TACGCTGCGATGGATTCCGGCATAG-3') or 5RACE-3GUS-3 (5'-GTTTTCGTCGGTAATCACCATTCCC-3') was used respectively for PCR amplification of 5' UTRs with the Advantage 2 Polymerase Mix (Clontech Laboratories). Cloning and verifying of single clones was carried out as described before.

Table Ad1. Oligonucleotide sequences

Oligonucleotide name	Sequence (5'–3'), restriction sites underlined
GLDPA-5'-*Xba*I	TGCTCTAGAAGCTTTACTCCTCTC
*GLDPA*1-3'-*Bam*HI	TTAGGATCCCACTTTCACATTCGCCTT
GLDPA1_{50}-5'-*Xba*I	TTATCTAGAATCATGTTTTGATGG
*GLDPA*1.1-3'-*Bam*HI	TTAGGATCCTTGGGAAGAACACAAACC
*GLDPA*1.2-5'-*Xba*I	TTATCTAGAGTTCGTAGGTTTGTGTTC
*GLDPA*1.2-3'-*Bam*HI	TTAGGATCCTAATCATGAATTAACCGA
*GLDPA*1.3-5'-*Xba*I	TTATCTAGATTTTTCATCGGTTAATTC
*GLDPA*1.3-3'-*Bam*HI	TTAGGATCCAAAACATGATGTGAAATA
*GLDPA*1.4-5'-*Xba*I	TTATCTAGAATTCCACTATTTCACATC
*GLDPA*2-5'-*Xba*I	TTATCTAGATGAAACAGGATGAGCCAC
*GLDPA*2-3'-*Bcu*I	TTAACTAGTGTGGAGATGATAGTTGTTG
*GLDPA*2-3'-*Xma*I	TTACCCGGGGTGGAGATGATAGTTGTT
*GLDPA*3-5'-*Xba*I	TTATCTAGAGTGGTTCCGTGCCGC
*GLDPA*3-3'-*Xba*I	ATCTCTAGAAAAAGTTCAAAACTTG
*GLDPA*3-3'-*Xma*I	TTACCCGGGAAAAGTTCAAAACTTGAT
GLDPA3_{50}-3'-*Bcu*I	TTAACTAGTACAAATGACGCACC
*GLDPA*3.1-5'-*Bam*HI	TTAGGATCCGTGGTTCCGTGCCGCCTT
*GLDPA*3.1-3'-*Bcu*I	TTAACTAGTGTGGGTGAGTGCTTTGAT
*GLDPA*3.2-5'-*Bam*HI	TTAGGATCCTGTCGTCATCAAAGCACT
*GLDPA*3.2-3'-*Bcu*I	TTAACTAGTATCAAAGCATAAAAACCT
*GLDPA*3.3-5'-*Bam*HI	TTAGGATCCAATCTTTAGGTTTTTATG
*GLDPA*3.3-3'-*Bcu*I	TTAACTAGTCACAGATAACAAAAAGTC
*GLDPA*3.4-5'-*Bam*HI	TTAGGATCCCGTATCTGACTTTTTGTT
*GLDPA*3.4-3'-*Bcu*I	TTAACTAGTAAAAGTTCAAAACTTGAT
*GLDPA*7-5'-*Bcu*I	TTAACTAGTCATTTGATCTATAACGAT
GLDPA-3'-*Xma*I	AAATCCCGGGAGTGTAAGATGGG

Table Ad2. Oligonucleotide combinations for the amplification of the appropriate promoter parts

Construct	Promoter part	5'-Oligonucleotide	3'-Oligonucleotide
GLDPA-Ft-1_{50}-2-7	*GLDPA*-1_{50}-2	*GLDPA*1_{50}-5'-*Xba*I	*GLDPA*2-3'-*Bcu*I
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-2-3_{50}-7	*GLDPA*-2-3_{50}	*GLDPA*2-5'-*Xba*I	*GLDPA*3_{50}-3'-*Bcu*I
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-1-3-7	*GLDPA*-1	*GLDPA*-5'-*Xba*I	*GLDPA*1-3'-*Bam*HI
	GLDPA-3	*GLDPA*3.1-5'-*Bam*HI	*GLDPA*3.4-3'-*Bcu*I
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-1.1-3-7,	*GLDPA*-1.1	*GLDPA*-5'-*Xba*I	*GLDPA*1.1-3'-*Bam*HI
GLDPA-Ft-1.2-3-7,	*GLDPA*-1.2	*GLDPA*1.2-5'-*Xba*I	*GLDPA*1.2-3'-*Bam*HI
GLDPA-Ft-1.3-3-7,	*GLDPA*-1.3	*GLDPA*1.3-5'-*Xba*I	*GLDPA*1.3-3'-*Bam*HI
GLDPA-Ft-1.4-3-7	*GLDPA*-1.4	*GLDPA*1.4-5'-*Xba*I	*GLDPA*1-3'-*Bam*HI
	GLDPA-3	*GLDPA*3.1-5'-*Bam*HI	*GLDPA*3.4-3'-*Bcu*I
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-1-3.1-7,	*GLDPA*-3.1	*GLDPA*3.1-5'-*Bam*HI	*GLDPA*3.1-3'-*Bcu*I
GLDPA-Ft-1-3.2-7,	*GLDPA*-3.2	*GLDPA*3.2-5'-*Bam*HI	*GLDPA*3.2-3'-*Bcu*I
GLDPA-Ft-1-3.3-7,	*GLDPA*-3.3	*GLDPA*3.3-5'-*Bam*HI	*GLDPA*3.3-3'-*Bcu*I
GLDPA-Ft-1-3.4-7	*GLDPA*-3.4	*GLDPA*3.4-5'-*Bam*HI	*GLDPA*3.4-3'-*Bcu*I
	GLDPA-1	*GLDPA*-5'-*Xba*I	*GLDPA*1-3'-*Bam*HI
	GLDPA-7	*GLDPA*7-5'-*Bcu*I	*GLDPA*-3'-*Xma*I
GLDPA-Ft-1-2-3	*GLDPA*-1-2-3	*GLDPA*-5'-*Xba*I	*GLDPA*3-3'-*Xma*I
GLDPA-Ft-3-1-2	*GLDPA*-3	*GLDPA*3-5'-*Xba*I	*GLDPA*3-3'-*Xba*I
	GLDPA-1-2	*GLDPA*-5'-*Xba*I	*GLDPA*2-3'-*Xma*I

7. Literature

Brogna S, Wen J (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol **16:** 107–113

Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem **76:** 51–74

Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. Gene **299:** 1–34

Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene **361:** 13–37

Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, Silhavy D (2009) Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant Mol Biol **71:** 367–378

Sachs MS, Geballe AP (2006) Downstream control of upstream open reading frames. Genes Dev **20:** 915–921

Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Danksagungen

Ich danke...

...**Prof. Dr. Peter Westhoff** für die Möglichkeit, meine Doktorarbeit an seinem Institut durchführen zu können und für die hilfreiche Betreuung während dieser Zeit.

...**Prof. Dr. Rüdiger Simon** für die Übernahme des Koreferates.

...meinen Freunden **Chris**, **Julchen** und **Steffi** für eine unvergessliche Zeit, die ständige Unterstützung und Aufmunterung, unsere schönen und unterhaltsamen Schoki-Pausen und den gemeinsamen Spaß mit euch während der gesamten Zeit, insbesondere…

…dir, **Chris**, für deine lehrreiche Pflanzennachhilfe, unsere witzigen (Nacht)Stunden mit den Wurzeln, die unvergessliche Woche im Maisfeld, die schönen gemeinsamen Mittagspausen im Botanischen Garten, deine lieben Mitbringsel und für noch so vieles mehr…

…dir, **Julchen**, für deine Redseligkeit und Fröhlichkeit, du kleiner Sonnenschein, die gleiche Wellenlänge, dein großes Herz und dein offenes Ohr.

…dir, **Steffi**, für unsere witzigen Spielchen zwischendurch als Fast-Labornachbarn, deine Hilfe während der krisenhaften Paper-Zeit, deinen Zauberstab und dein Schwarzes Herz.

Ich danke euch für alles und werde euch nie vergessen!

...**Nino** für ihr sonniges Gemüt, ihr herzliches Lachen und den Spaß, den wir immer hatten.

...**Maria** und **Monika** für die besten Flaverias aller Zeiten und die liebe Pflege der Pflänzchen, insbesondere dir, **Maria**, für die lustigen Mittagspausen und die witzigen Anekdoten aus deiner Heimat.

...**Susanne** für das Dulden meines langjährigen „Reserve"-Kühlfachs in ihrem Labor.

...**Udo** für die nette Betreuung.

...dem Rest **der Botanik IV-Truppe** für die schöne Zeit, das angenehme Arbeitsklima und die nette Zusammenarbeit.

...**den Gärtnern** für die Betreuung meiner Pflanzen.

...**Meli** für das Anhören meiner Uni-Probleme und das ständige Aufmuntern.

...**meiner Familie**, die immer für mich da war und mich die ganze Zeit tatkräftig unterstützt hat. Vielen Dank für alles! Ich habe euch sehr lieb!