

# Transcriptional Regulation of NADP-Dependent Malate Dehydrogenase: Comparative Genetics and Identification of DNA-Binding Proteins

Steffen Hameister · Beril Becker · Simone Holtgreffe · Inga Strodtkötter · Vera Linke · Jan E. Backhausen · Renate Scheibe

Received: 8 January 2007 / Accepted: 2 July 2007 / Published online: 10 October 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** The transcriptional regulation of NADP-malate dehydrogenase (NADP-MDH) was analyzed in Arabidopsis ecotypes and other Brassicaceae. The amount of transcript increased twofold after transfer into low temperature (12°C) or high light (750 μE) in all species. Analysis of the genomic DNA reveals that the NADP-MDH gene (At5g58330 in *A. thaliana*) in Brassicaceae is located between two other genes (At5g58320 and At5g58340 in Arabidopsis), both encoded on the opposite DNA strand. No promoter elements were identified in 5' direction of the NADP-MDH gene, and the expression of NADP-MDH was not affected in knock-out plants carrying a DNA insert in the 5' region. A yeast-one hybrid approach yielded only three DNA-binding proteins for the 500-bp fragment located upstream of the ATG sequence, but 34 proteins for its coding region. However, in *Chlamydomonas* and in some Poaceae, which do not possess any genes within the 1200 bp upstream region, typical

promoter elements were identified. Alignments of genomic DNA reveal that, in contrast to Poaceae, the introns are highly conserved within Brassicaceae. We conclude that in Brassicaceae the majority of regulatory elements are located within the coding region. The NADP-MDH gene of both families evolved from a common precursor, similar to the gene in *Chlamydomonas*. Changes in the selection pressure allowed the insertion of At5g58340 into the promoter region of a common ancestor. When the demand for transcriptional regulation increased, At5g58340 disappeared in Poaceae, and a promoter developed in the 5' region. In contrast, Brassicaceae maintained At5g58340 and shifted all regulatory elements into the coding region of NADP-MDH.

**Keywords** Arabidopsis · Brassicaceae · Evolution · Gene organization · Genomic DNA · Knock-out · Poaceae · Yeast-one hybrid

S. Hameister and B. Becker contributed equally to this paper.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-007-9025-9) contains supplementary material, which is available to authorized users.

S. Hameister · B. Becker · S. Holtgreffe · I. Strodtkötter · V. Linke · J. E. Backhausen · R. Scheibe  
Pflanzenphysiologie, Fachbereich Biologie/Chemie,  
Universität Osnabrück, D-49069 Osnabrück, Germany

S. Hameister  
Spezielle Botanik, Fachbereich Biologie/Chemie,  
Universität Osnabrück, D-49069 Osnabrück, Germany

J. E. Backhausen (✉)  
Pflanzenphysiologie, Fachbereich Biologie/Chemie,  
Universität Osnabrück, Barbarastr. 11, D-49069 Osnabrück,  
Germany  
e-mail: backhausen@biologie.uni-osnabrueck.de

## Introduction

Plants respond to changes in their environment on various levels. Short-term changes in light intensity are compensated for by using already existing proteins, while persisting changes lead to altered gene expression (e.g., Murchie et al. 2005; Scheibe et al. 2005). During high-light acclimation, the amount of thylakoid components, especially of light-harvesting complexes (LHC), is decreased, and the amount of stromal enzymes, especially of Calvin-cycle enzymes, is increased (Anderson et al. 1995). To prevent oxidative stress, the amounts of glutathione-synthesizing enzymes and of chloroplast NADP-MDH also increase (Becker et al. 2006). The operation of the malate valve is important to stabilize the chloroplast NADPH/ATP ratio, especially under rapidly changing conditions

(Scheibe 2004). With respect to the nuclear-encoded chloroplast NADP-MDH, the key enzyme of the malate valve, it has long been known that it is subject to light/dark modulation, mediated by the ferredoxin–thioredoxin system (Buchanan 1980; Scheibe 2004). The activation state of NADP-MDH is influenced by the NADPH/NADP ratio (Scheibe and Jacquot 1983), thus NADP-MDH activation, and consequently the export rate of malate from the chloroplast, decreases when little NADPH is available.

In addition to these fast changes in the activation state, it has also been shown that the total amount of NADP-MDH is regulated by several endogenous and exogenous factors. In tobacco plants, the amount of NADP-MDH, defined as capacity after full activation with reducing agents, showed strong age dependence. The capacity was highest in young leaves, and decreased with leaf and plant age (Faske et al. 1997). When tobacco plants were grown under various CO<sub>2</sub> concentrations, the expression of NADP-MDH was reduced when elevated CO<sub>2</sub> was present (Backhausen and Scheibe 1999). On a chlorophyll (Chl) basis, the NADP-MDH capacity was nearly tenfold higher in cold-grown winter wheat, compared to control plants grown at 20°C (Savitch et al. 2000). In cold-acclimated Arabidopsis plants, the NADP-MDH capacity was also higher than in plants grown at 20°C (Savitch et al. 2001). Monitoring the expression changes in NADP-MDH after a transfer into conditions of persisting over-reduction, i.e., a transfer of Arabidopsis plants into high light levels and moderately decreased temperature, it turned out that both NADP-MDH transcript and protein amount are upregulated within a few hours (Becker et al. 2006). From an array analysis conducted in parallel, it was concluded that the expression change of NADP-MDH is not a single event, but a part of the overall process of high-light acclimation (Becker et al. 2006). There is some evidence, at least in the case of fully developed Arabidopsis leaves, that an alteration in the chloroplast redox state is responsible for the release of redox signals. An excitation imbalance between photosystem (PS) I and PS II has the potential to generate such redox signals, and also regulates the transcription of several other nuclear-encoded genes (Walters et al. 1999; Fey et al. 2005; Holtgreffe et al. 2007).

However, in Arabidopsis, the expression pattern of NADP-MDH not only depends on the environmental conditions, but is modulated by the endogenous state of the plant. The transcriptional upregulation of NADP-MDH was evident in plants grown under short-day conditions, while in plants grown under long-day conditions no such increase was observed. The duration of the light period resulted in different responses towards an identical stimulus, i.e., high light and decreased temperature. It has been concluded that the endogenous systems that measure day length interact with redox regulation. Redox-mediated acclimation signals are redirected during short days in a way that allows light

to be used more efficiently, and prevents oxidative damage under long-day conditions (Becker et al. 2006).

In summary, a complex pattern for the transcriptional regulation of NADP-MDH emerges. It is evident that, for the regulation of transcription, either the altered binding of transcription factors or the binding of different transcription factors to regulatory elements of the target gene is required. In the majority of cases described to date in plants, transcription factors do not interact with the coding region, but all *cis*-regulatory elements are localized in the 5' upstream region of a gene. A typical promoter of the RNA-polymerase II-type consists of the core promoter, located directly in front of the beginning of the section to be transcribed, and the proximal and distal promoter regions. The core promoter contains the transcription start site (TSS), located in the initiation region (INR), together with some binding sites for general transcription factors. A well-known transcription-factor binding site is the TATA box. It is usually located 25–40 bp upstream of the TSS, and it is the recognition sequence for the TATA-binding protein that forms part of the transcription complex (Werner 1999).

Upstream of the core promoter is the proximal promoter, which usually has a length of 200–300 bp and contains binding sites for additional transcription factors. Transcription factors that bind to the proximal promoter are often involved in the regulation of transcription. A well-studied example is the CCAAT box, typically located 80–110 bp in front of the TSS. In contrast to the TATA box, which only regulates the start of transcription, the CCAAT box controls the intensity of gene expression. The distal promoter, which is located upstream of the proximal promoter, contains additional regulatory elements and can extend up to 1500 bp from the transcription start. The distal promoter is characterized by its highly variable structure, and by containing additional binding sites for transcription factors (Werner 1999).

The occurrence of different transcription-factor binding sites in the proximal and the distal promoter are thought to be responsible for the differential expression patterns of a gene. Transcription factors act as independent modules. Apart from constitutive promoters, development-dependent, tissue-specific promoters, and inducible promoters, which respond to specific environmental signals, are known (Tyagi 2001). In some cases, additional elements involved in positive or negative control of transcription can be located further upstream or downstream in the 3'-untranslated region of a gene (Timko et al. 1985). They differ from the distal promoter in acting in a manner that is independent of position and orientation (Rippe et al. 1995).

However, several modifications from the classical promoter type have been described. The modifications can be grouped in four types, all of which lead to correct transcription of the target genes: (1) The TATA box-core

promoter contains neither INR nor downstream elements; (2) The INR-core promoter does not contain the TATA box and downstream elements; (3) The composite-core promoter contains only the TATA box and INR; and (4) the null core promoter contains only up- and downstream elements. Only 30–50% of all known promoters contain a TATA box. Many housekeeping genes, and most photosynthetic genes, are characterized by the lack of the TATA box (Shahmuradov et al. 2005). There are also some known examples of transcription factors that can bind to the coding region. The leaf ferredoxin gene, for example, contains important regulatory elements in the coding region (Dickey et al. 1992).

In this work we use different strategies, namely comparative genetics, analysis of knock-out plants, and a yeast-one hybrid system (YOH) screen, to put forward the hypothesis that, due to the presence of coding DNA segments and overlapping UTRs on the opposite DNA strand, in both the 3' and 5' directions, the regulatory elements that control the complex transcription pattern of NADP-MDH in Arabidopsis were shifted into the coding region of the gene during evolution.

## Materials and Methods

### Plant Material and Growth Conditions

Seeds of *Capsella rubella* Reut., *Capsella bursa-pastoris* (L.) Med., *Cardaminopsis petraea* (L.) Hiit., *Cochlearia officinalis* L., *Lepidium densiflorum* Schrad. and *Lepidium latifolium* L. were obtained from the collection of the Botanical Garden of the Universität Osnabrück. Seeds of *Arabidopsis thaliana* L. (Heynh.) from the accessions Columbia, Landsberg erecta, and Wassilewskija were obtained from the SALK institute (<http://www.signal.salk.edu>).

Inbred seeds were placed on soil and kept in darkness for 48 hours at 4°C in order to synchronize germination. Cultivation conditions were defined as a light intensity of 120  $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ , a relative humidity of 65%, and 8 hours of light, with 22°C to 18°C day/night temperatures. Six weeks after germination, the plants were transferred into a growth chamber with the respective conditions (120  $\mu\text{E}$  at 12°C, or 750  $\mu\text{E}$  at 22°C). SON-T AGRO 400 lamps (Philips, Eindhoven, The Netherlands) were used as light sources.

Genomic DNA from *Arabis turrita* L. and *Arabidopsis wallichii* (Hook. f. et Thoms.) Busch was a kind gift from M. Koch (Universität Heidelberg, Germany).

Leaf samples were taken by quickly removing all rosette leaves, and transferring them into liquid nitrogen within less than 10 seconds. Leaves from two different

plants were combined to form each sample. The leaves were homogenized under liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until use.

### Identification and Analysis of Homozygous At5g58340 Plants

The Arabidopsis knock-out mutant lines *At5g58340::tDNA-1* (SALK\_053119) and *At5g58340::tDNA-2* (SALK\_018118) were received from the Arabidopsis Biological Resource Stock Center (<http://www.arabidopsis.org/abrc>). Homozygous knock-out plants were identified by PCR for T-DNA insertion within the gene region of *At5g58340*. Genomic DNA was isolated from plant tissues by standard methods. The sequence information for the gene- and T-DNA specific primers was obtained from the SALK-Institute (<http://www.signal.salk.edu>). The insertion positions of the T-DNA products were confirmed by sequencing the PCR products.

### NADP-MDH Capacity Measurements in Leaf Samples

The samples used for the estimation of the NADP-MDH capacity were obtained from individual leaves. The leaves were cut from the plant and immediately transferred into liquid nitrogen. The leaves were homogenized under liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until use. The activity of NAD-MDH, and total activity (capacity) of NADP-MDH were determined after exhaustive reduction with reduced dithiothreitol (DTT) and corrected for non-specific NADP-dependent activity due to NAD-MDH, as described by Scheibe and Stitt (1988). For each experiment, duplicate measurements from three different plants were made.

### Northern Blot Analysis

For Northern blot analysis, total RNA was isolated from frozen leaf material using the Purescript RNA extraction kit (Gentra Systems, Minneapolis, Minn., USA). For RNA gel-blot hybridization analysis, 10  $\mu\text{g}$  of total RNA was denatured and separated on a 1.2% (w/v) agarose 2.5% (v/v) formaldehyde gel. Homidium bromide was included in the loading buffer to ensure equal sample loading. RNA was blotted onto a nylon membrane (Hybond-N, Amersham Biosciences) using downstream capillary transfer. RNA was cross-linked to the membrane by ultraviolet (UV) irradiation. Prehybridization and hybridization were performed at 65°C in Church-buffer medium [0.25 M sodium phosphate buffer, pH 7.2, 1 mM ethylenediamine

tetraacetic acid (EDTA), 7% (w/v) sodium dodecyl sulfate (SDS), and 1% bovine serum albumin (BSA)]. Hybridization was performed by using an alpha-[<sup>32</sup>P]-dCTP-labeled full-length NADP-MDH (*A. thaliana*)-cDNA-specific probe (Ready-To-Go DNA labeling beads, Amersham Biosciences). Membranes were washed twice for 15 minutes at 65°C in washing buffer [40 mM sodium phosphate buffer, pH 7.2, 1 mM EDTA, 0.5% (w/v) SDS and 0.5% (w/v) BSA], then for 10 minutes at room temperature in washing buffer containing 1% (w/v) SDS. Finally, membranes were exposed to Kodak MS X-ray film at -80°C. The densitometric analysis of the blots was performed using the Gelix-1 gel scan software.

#### Isolation of Genomic DNA, Amplification and Sequencing of Gene Fragments

Genomic DNA was isolated using a modified protocol based on the cetyltrimethyl ammonium bromide (CTAB) method (Mummenhoff and Koch 1994). Homologous primers were designed for the amplification of NADP-MDH promoter fragments, based on the genomic NADP-MDH sequence from *Arabidopsis thaliana* var. Columbia (AB019228.1; clone MCK7). As forward primers, NADP-MDH-for (5'-tctagacgcctcgttgccgatc-3') and NADP-MDH-for2 (5'-ttgagctactgattcaagagagga-3') were used, whereas NADP-MDH-rev (5'-ccatggccattatgcacagacac-3') was used as the reverse primer. The polymerase-chain reaction (PCR) conditions were 0.4–1.4 µg template DNA, 0.5 U *Taq*-Polymerase (Fermentas, St. Leon-Rot, Germany) and 0.4 mM of each primer. PCR was carried out using a gradient thermocycler (Eppendorf, Hamburg, Germany). The PCR cycle conditions were five minutes initial denaturation at 95°C, 45 cycles of one minute at 95°C, one minute at 55–60°C, one minute at 72°C for final extending, then 10°C cooling. The PCR products were purified with the QIAquick<sup>®</sup> PCR-purification kit (Qiagen, Hilden, Germany). Cycle-sequencing amplification was carried out for all DNA fragments in forward and reverse orientation by using the Big Dye<sup>™</sup>-terminator cycle-sequencing ready-reaction kit (Applied Biosystems, PE Biosystems, Foster City, USA).

#### Western Blot and Determination of Protein Content

Equal amounts of soluble protein (30 µg) were loaded onto two 15% discontinuous SDS-polyacrylamide gels using a vertical minigel system (Mini-Protean II, BioRad, Germany). After separation by electrophoresis, one of the gels was stained with Coomassie. The other gel was blotted onto nitrocellulose membrane. Immunodetection

was performed essentially as described in Graeve et al. (1994). For the detection of NADP-MDH, an antibody against NADP-MDH from pea leaves (1:3000) was used. For the detection of the second antibody (1:15000), Luminol (Amersham, Freiburg, Germany) was used as a substrate. The protein content of leaf extracts was determined according to Bradford (1976), with BSA as a standard.

#### Yeast-One Hybrid

For construction of the NADP-MDH-HIS3 fusion genes, parts of the 5' UTR, the 3' UTR, and the coding sequence of genomic DNA fragments of NADP-MDH (clone MCK7), were amplified by PCR using specific primers (see the primer list in the supplementary information). The potential promoter fragments were cloned in pGEM-T Easy, digested with *Eco*RI, and cloned into pHisi and pHisi-1 in order to position the fragments directly in front of the His3-reporter gene. These constructs were digested with *Xho*I and transferred into the genome of the yeast strain YM4271 by homologous recombination, according to the manufacturer's instructions (Matchmaker One-Hybrid System; Clontech, Heidelberg, Germany). A small-scale transformation, with 15 µg linearized pHisi-1 carrying the promoter fragments, was performed. For selection, the strains were grown on SD-His medium, so that only colonies that had a functional His3-reporter gene were able to grow. The background activity of His3 was competitively inhibited by 3-AT, an inhibitor of the His3 gene. Only strains with a His3-background activity below 45 mM 3-AT were used for one-hybrid screens.

An *Arabidopsis*-cDNA library (Nemeth et al. 1998), which had a GAL4-activation domain (AD) and a leucine gene at the vector pACT2, was used to express proteins fused with GAL4-AD, which bind to the selected DNA fragments of NADP-MDH. The undigested cDNA plasmids were transformed into the promoter-reporter strains using the manufacturer's protocol for a large-scale transformation (Matchmaker One-Hybrid System, Clontech, Heidelberg, Germany). The binding of a protein, expressed by the cDNA library fused to the GAL4-AD, to a DNA fragment resulted in expression of the reporter gene His3 and allowed growth of the cells on media without histidine and leucine. From these colonies, the cDNA-carrying plasmids were isolated and transformed to electrocompetent *Escherichia coli* cells. For sequencing, the plasmids isolated from *E. coli* were used. All proteins identified with the YOH screen were denoted as "binds at NADP-MDH gene" (BANG) and numbered according to the fragment they are binding to.

## Bioinformatical Analysis

The DNA sequences of NADP-MDH were obtained from the National Center for Biotechnology Information (NCBI) gene bank (<http://www.ncbi.nlm.nih.gov/>), and from the eukaryote projects of the Institute for Genome Research, (<http://www.tigr.org/tdb/euk/>). The genomic DNA sequence of *A. thaliana* was used to generate homologous primers for promoter and coding region. The comparison of the obtained sequences was performed with the BLAST algorithm (Altschul et al. 1997), using the BLAST portal of NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>). To identify conserved promoter regions, multiple sequence alignments were performed with ClustalW (Version 1.8.2), using standard parameters. To allow further analysis of the promoter sequences and the identification of potential transcription-factor binding sites, the MatInspector tool of the Genomatix homepage (<http://www.genomatix.de/cgi-bin/matinspector>) was used. The expression patterns of At5g58340 and At5g58330 were compiled from the At-GenExpress tool (<http://www.jsp.weigelworld.org/expviz/expviz.jsp>).

The identification of the cDNA sequences, for the results obtained with the YOH screen, was performed with BLAST on NCBI and TAIR (<http://www.arabidopsis.org/Blast/>). For unknown interaction partners, a function search was performed with <http://www.ncbi.nlm.nih.gov>, using the “Short, nearly exact matches in Arabidopsis” and the “Short, nearly exact matches in all organisms, search by domain architecture” options. Sequence alignments and the analysis of unknown proteins were done with Lalign ([http://www.ch.embnet.org/software/LALIGN\\_form.html](http://www.ch.embnet.org/software/LALIGN_form.html)), Predict protein (<http://www.expasy.ch>), Psort (<http://www.psort.nibb.ac.jp/form.html>), and Meta-search ([http://www.cubic.bioc.columbia.edu/pp/submit\\_meta.html](http://www.cubic.bioc.columbia.edu/pp/submit_meta.html)).

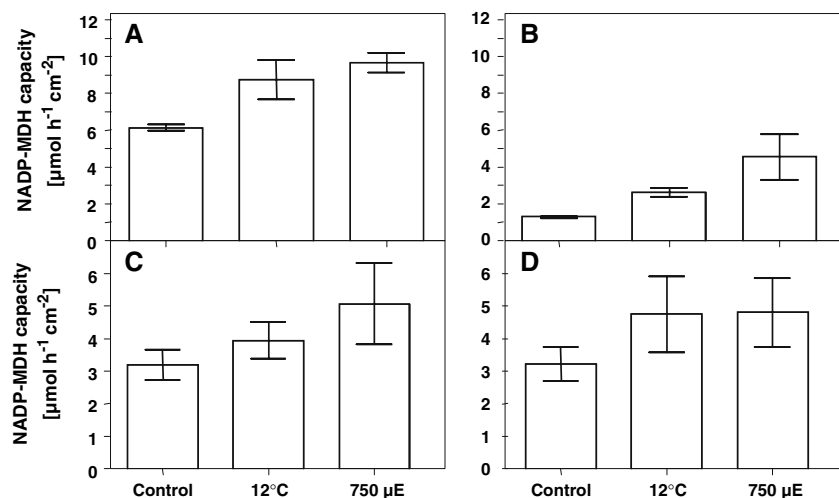
## Results

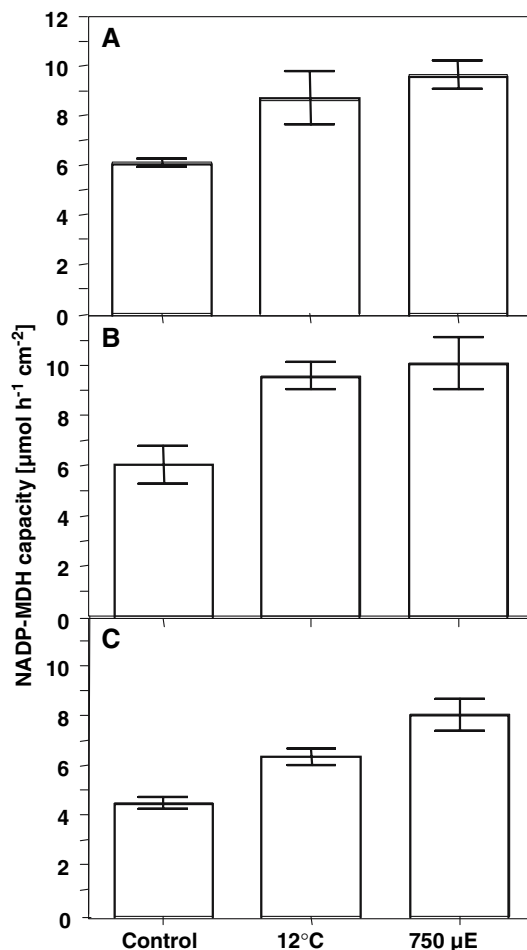
## Capacity Changes of NADP-MDH in Various Brassicaceae Species and Varieties

It has recently been shown that in *A. thaliana* var. Columbia the capacity of NADP-MDH is upregulated at the transcriptional level under a combination of high light and low temperature (Becker et al. 2006). We analyzed how the two parameters act, separately, to alter transcript level and enzyme capacity in different Arabidopsis ecotypes and in selected other Brassicaceae species. Low-light-acclimated plants (120  $\mu\text{E}$  at 22°C) were transferred into conditions of either decreased temperature (120  $\mu\text{E}$  at 12°C) or high light (750  $\mu\text{E}$  at 22°C), and samples were taken after four days.

The capacity of NADP-MDH in low-light-acclimated plants ranged between 1.29  $\mu\text{mol h}^{-1} \text{cm}^{-2}$  leaf area in *Cochlearia officinalis* (Fig. 1B), and 6.15  $\mu\text{mol h}^{-1} \text{cm}^{-2}$  leaf area in *Arabidopsis thaliana* var. Columbia (Fig. 1A). After the transfer, the capacity increased under both conditions. In *Arabidopsis thaliana* var. Columbia, values of 8.75  $\mu\text{mol h}^{-1} \text{cm}^{-2}$  and 9.68  $\mu\text{mol h}^{-1} \text{cm}^{-2}$  were reached, while in *Cochlearia officinalis* 2.63  $\mu\text{mol h}^{-1} \text{cm}^{-2}$  and 4.54  $\mu\text{mol h}^{-1} \text{cm}^{-2}$  were reached. In general, the different Arabidopsis ecotypes and the other Brassicaceae showed a similar behavior. The largest increase was always obtained after transfer into high light. Decreased temperatures also resulted in higher NADP-MDH capacities, but the final values were, on average, 25% below the values obtained in high light (Figs. 1 and 2). Northern blot analysis of NADP-MDH mRNA in the different Arabidopsis ecotypes indicates a specific increase in gene expression in all cases (Fig. 3), which correlates with the activity measurements. Thus, the increase in NADP-MDH capacity after the

**Fig. 1** NADP-malate dehydrogenase (MDH) capacity changes in selected Brassicaceae species. Plants from the indicated species were analyzed four days after a transfer into low temperature (12°C), into high light (750  $\mu\text{E}$ ) or under control conditions (Control). Full activation of NADP-MDH was achieved by excessive reduction with DTT<sub>red</sub> (A) *A. thaliana* var. Columbia (B) *Cochlearia officinalis* (C) *Capsella rubella* (D) *Lepidium latifolium*. The error bars represent the standard error





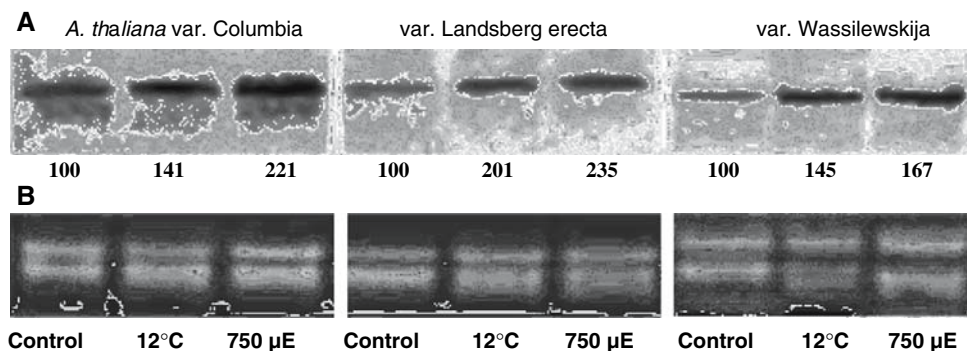
**Fig. 2** NADP-malate dehydrogenase (MDH) capacity changes in Arabidopsis ecotypes. Plants from three Arabidopsis ecotypes were analyzed four days after transfer into low temperature (12°C), into high light (750  $\mu\text{E}$ ) or under unaltered control conditions (Control). The mean values of two independent samples for each condition are shown. Full activity of NADP-MDH was measured after reduction with  $\text{DTT}_{\text{red}}$  (A) *A. thaliana* var. Columbia, (B) *A. thaliana* var. Landsberg erecta, (C) *A. thaliana* var. Wassilewskija

transfer into high light or decreased temperature is most likely the consequence of a transcriptional upregulation.

#### Structure of the NADP-MDH Gene and the Upstream and Downstream DNA Regions

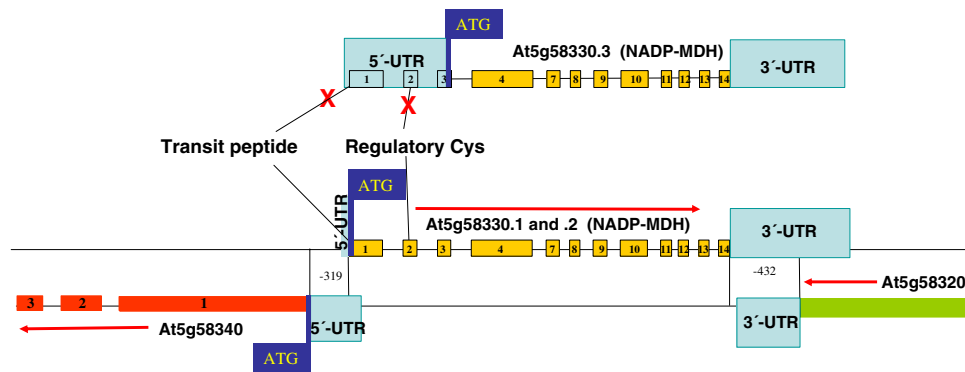
In Arabidopsis, the gene for NADP-MDH is located on chromosome five and consists of between 12 and 14 exons, and between 11 and 13 introns (Fig. 4). For this gene, three different splicing variants are available in the Arabidopsis database (<http://www.arabidopsis.org>). The first splicing variant (At5g58330.1) has a total length of 3180 bp, encoding for a protein of 444 amino acids. Its 5' UTR is predicted to be 94 bp long, and its 3' UTR is 808 bp long. Splicing variant At5g58330.2 is 57 bp shorter, and encodes a protein of 443 amino acids. Both differ in their amino acid sequence between position 16 and 48, located in the transit peptide of the protein. For At5g58330.2, the 5' UTR is predicted to have a length of only 37 bp, while the 3' UTR is also 808 bp long. The third splicing variant (At5g58330.3) is 3094 bp long, but the predicted protein should have a size of only 335 amino acids, because the 5' UTR is predicted to cover the first 544 bp of the coding region as well. Therefore, the first 108 amino acids of the resulting protein, carrying the transit peptide and the N-terminal extension with the regulatory cysteine residues, would be missing. Apart from the lack of the ability to be transferred into chloroplasts and to be regulated via the N-terminal disulfide bridge, this would result in a much smaller NADP-MDH protein than has ever been observed in vivo. In addition, the corresponding transcript could not be found in the mRNA-data collection of NCBI (<http://www.ncbi.nlm.nih.gov/>).

Assuming that only the first two splicing variants (At5g58330.1 and At5g58330.2) occur in vivo, the distance



**Fig. 3** Northern blot analysis of the NADP-malate dehydrogenase (MDH) transcript in Arabidopsis ecotypes. Plants from three Arabidopsis ecotypes were analyzed four days after a transfer into low temperature (12°C), into high light (750  $\mu\text{E}$ ) or under unaltered control conditions. (A) Autoradiograph of the Northern blot, hybrid-

ized with a [ $^{32}\text{P}$ ]-NADP-MDH (*A. thaliana*)-specific probe. The numbers below the image indicate the band intensity. (B) As a loading control for the gels, the rRNA content is shown. Left side, *A. thaliana* var. Columbia; middle, *A. thaliana* var. Landsberg erecta; right side, *A. thaliana* var. Wassilewskija



**Fig. 4** Schematic representation of At5g58340, the splicing variants of NADP-MDH (At5g58330) and At5g58320 in *A. thaliana* var. Columbia. This figure gives an overview of the localizations of At5g58340, At5g58330 and At5g58320. The exons of At5g58340 are shown in red.

The exons of the three splicing variants of At5g58330 (NADP-MDH) are shown in yellow, and At5g58320 is shown in green. All untranslated regions (UTRs) are shown in light blue. Exons 5 and 6 are not assigned to the Arabidopsis NADP-MDH gene (cf. Fig. 6)

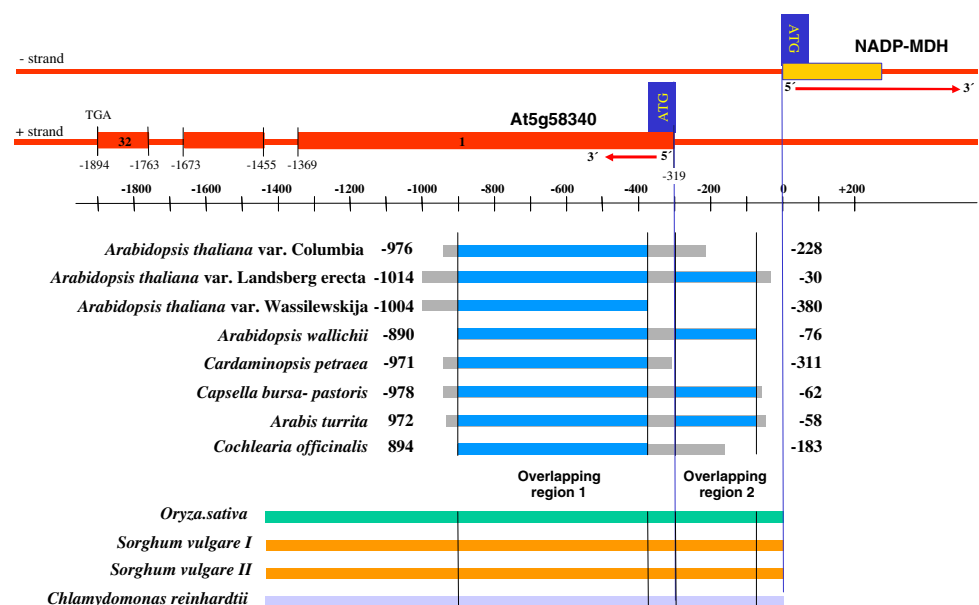
to the next genes on chromosome 5 is very small (summarized in Fig. 4). In the 3' direction, the coding region of At5g58320 starts 410 bp after the end of the last exon of NADP-MDH, but the UTRs are predicted to overlap by 829 bp. At5g58320 is transcribed in the opposite direction. The function of that gene is not known, and it must be questioned whether it is transcribed at all, since no expression data are available in the internet databases.

In 5' direction, the transcription start of the gene At5g58340, also encoded on the opposite DNA strand, is only 319 bp away from the ATG of NADP-MDH. The coding region is 1575 bp long and contains two introns close to the end of the gene. The first exon, located nearest to the NADP-MDH transcription start, is 1050 bp long and thus covers the region which should carry the distal and the proximal promoter of NADP-MDH. The UTRs of both genes also overlap (Fig. 4). At5g58340 is denoted as an

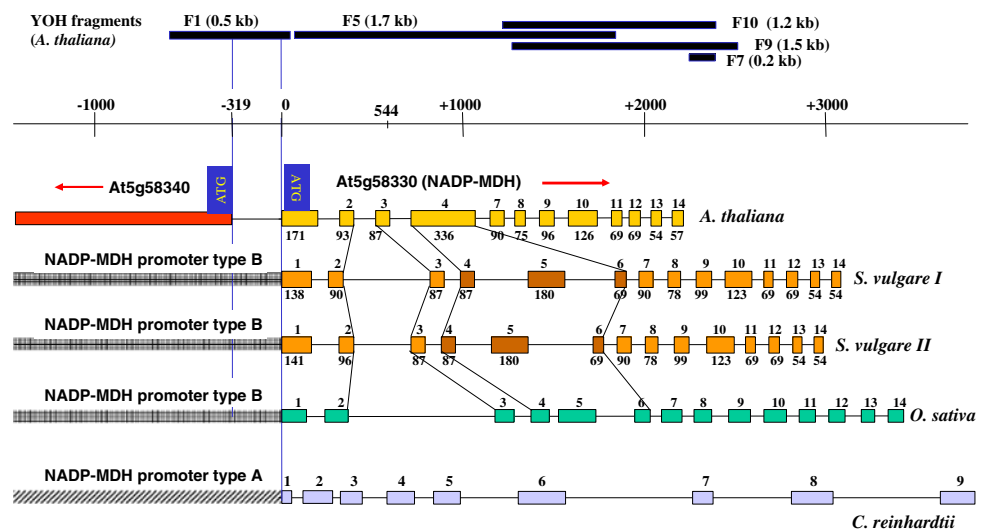
unknown gene, and a BLAST search for similar genes indicates that it may be a myb-type transcription factor, because it probably contains a myb-DNA binding domain. In order to assess whether At5g58340 is really transcribed and detectable on gene arrays, the expression data on Affymetrix chips available in the internet were used. According to AtGenExpress (<http://www.jsp.weigelworld.org/expviz/expviz.jsp>), At5g58340 is expressed in all parts of the plant. The lowest expression values are reported for leaves, and the highest expression is reported for pollen and seeds in the late developmental states. However, the intensity of the signals is less than 10% of the intensity found for NADP-MDH. It should also be noted that the expression pattern of NADP-MDH does not show any similarity with that of At5g58340.

The structure of the NADP-MDH gene, as described above, is not unique for Arabidopsis. All other

**Fig. 5** Schematic representation of the 5' upstream region of the NADP-malate dehydrogenase (MDH) gene in various plants. The ATG of NADP-MDH is set to zero. Only the first exon of NADP-MDH is shown (in yellow at the right end). The classical promoter region of NADP-MDH, which contains At5g58340 on the opposite DNA strand, is labeled with negative numbers. The 5' upstream regions that were sequenced and aligned for the species listed in Table 4 are indicated. All sequenced DNA fragments are shown in grey and the sequence parts used for the alignments in Table 4 are shown in blue



**Fig. 6** Schematic representation of the NADP-MDH gene and of the fragments used for the YOH screen. The transcription start of NADP-MDH (ATG) is set at zero. The classical promoter region of NADP-MDH is labeled with negative numbers. Intron and exon structure, and numbering of introns and exons, as used in Tables 2 and 3, is indicated. At the top of the figure, the position of the fragments used in the YOH screen is indicated



Brassicaceae species sequenced in this study (see below) display the same gene structure, with 11 introns and only a small distance of approximately 320 bp between the coding regions of NADP-MDH and the unknown gene on the opposite DNA strand. The genomic sequence of the NADP-MDH gene and a large part of the 5' UTR is also available for some Poaceae. *Sorghum vulgare* is unique in having two genes for NADP-MDH, and the sequence from *Oryza sativa* is also available from databases. Three striking differences in the gene structures form Brassicaceae and Poaceae are evident (Figs. 5 and 6). At first, the NADP-MDH genes from the Poaceae possess 13 introns, but DNA alignments indicate that the large exon 4 of *A. thaliana* var. Columbia is disrupted by two additional introns which may be as long as 350 bp (Fig. 6). Thus, exons 5 and 6 are absent in Arabidopsis in the figure. The second major difference is that intron 2 is much larger, up to 450 bp in *Oryza sativa*, compared to only approximately 50 bp in the Brassicaceae. The most striking difference with respect to transcriptional regulation is, however, that no other genes are located in a 1200-bp region upstream of the transcriptional start. In rice, a pseudogene may be located at -897, but its DNA sequence does not display any similarity with At5g58340.

#### Sequence Comparison of the NADP-MDH Gene and the 5'-DNA Region

Among Brassicaceae, genes, especially exons, are often highly conserved (Rossberg et al. 2000). However, the NADP-MDH protein is strongly conserved even between other organisms. The identity of NADP-MDH from *A. thaliana* var. Columbia and other Brassicaceae compared to that of Poaceae is 76–78%, and even the protein

from *Chlamydomonas reinhardtii* is 60 or 62% identical with that of Poaceae or Brassicaceae, respectively (Table 1). In all cases, most differences in the translated sequence occur within the first 100 amino acids, carrying the transit peptide, and the region containing the N-terminal regulatory cysteines. The exception is NADP-MDH from *Chlamydomonas reinhardtii* which already carries the N-terminal sequence extension with several conserved amino acids but does not possess the two N-terminal cysteines. Also, the 40 amino acids of the sequence with the C-terminal regulatory cysteine residues are not as highly conserved as the other amino acids (Table 1).

However, at the DNA level, some differences are apparent. In the coding region, the exons of all Brassicaceae are highly conserved (Table 2). Identities vary between 87 and 89% (Exon 14) and 96 and 100% (exons 7 to 9), as was already reported for other Brassicaceae exons (Acaran et al. 2000; Boivin et al. 2004). This is slightly higher than those observed within the Poaceae (Table 3) with average identities between 90 and 100%. The identity of Poaceae with Arabidopsis is only 10% for the exon 1 (encoding the transit peptide), and otherwise between 34% and 88%. A comparison of the introns shows large differences between Poaceae and Brassicaceae. In Brassicaceae, the introns of the NADP-MDH gene are nearly as highly conserved as the exons. In Poaceae, however, identities of 80% or above are rare, and only occur between the two *Sorghum* sequences. The identities between both species are around 25%, and only intron 10 is 63% identical between rice and *Sorghum*.

The same tendency continues in the 5' UTR of the NADP-MDH gene. Here, the two regions indicated in Fig. 5 were analyzed in more detail. Overlapping region 1 has a length of approximately 520 bp and lies between -380 and -890 bp, upstream of the ATG of the Arabidopsis



**Table 1** Alignment of the NADP-MDH amino-acid sequences

Sorghum	MGLSTAYSP-VGSHLAPAPLGHRRSAQLHRP---RRALLATVRCVSDAA-KQVQDG---	51
Oryza	MAAIDLSSP-ARS--SPAPLSPRRGSLHLLL---RRPRRPTLRCSLDAAPKQAQAQGP	54
Arabidopsis	MAMAELESTPKTTSFPLNSSSLRLSSKHLHLSNHFRHLLPLPHTTTPNSKISCSVSQ	60
Chlamydomonas	MALNMKQQQ-----AGLS-RKAARVSS-----RAPVVVRAVAAP-----	34
	*. . . : : * .: . . : : . . . :	
Sorghum	-VATAEA--PATRKDQFVVFCTTYDLKAEDK--TKSWKKLVNIAVSGAAGMISNHLLFKL	106
Oryza	AVAAEEA--PTARKECYGVFCTTYDLRADEK--TKSWKSLVNAVSGAAGMISNHLLFKL	110
Arabidopsis	PVAVQENGLVKTKKECYGVFCLTYDLKAEET--TRSWKKLINIAVSGAAGMISNHLLFKL	118
Chlamydomonas	-VAPAAE---AEAKKAYGVFRLSYDTQNEASLTRSWKTKVAVTGAAGNIAHLLFML	90
	.** . . . *..*** :** ::: *:*.. :::*:**:* *:** ** :*** **	
Sorghum	ASGEVFGQDQPIALKLLGSEFSFQALEGVAMELEDSLYPLLRVSVIGIDPYEVFEDVDWA	166
Oryza	ASGEVFGPDQPIALKLLGSEFSIQALEGVAMELEDSLYPLLRVSVIGIDPVPVFEDEAWA	170
Arabidopsis	ASGEVFGPDQPIALKLLGSEFSIQALEGVAMELEDSLFPLLRVSDIGIDPNEVFQDVEWA	178
Chlamydomonas	ASGEVYKQDQPIALQLLGSERSKEALEGVAMELEDSLYPLLRVSVIGIDPYEVFGDADWA	150
	****:* *****:***** :*****:*****.* ** * .:*	
Sorghum	LLIGAKPRGPGMERAALLDINGQIFADQKALNAVASKNVKVLVGVNPCNTNALICLKNA	226
Oryza	LLIGAKPRGPGMERSALLDINGQIFAEQKALNSVASRNKVIIVGVNPCNTNALICLKNA	230
Arabidopsis	ILIGAKPRGPGMERADLLDINGQIFAEQKALNKAASPNVKVLVGVNPCNTNALICLKNA	238
Chlamydomonas	LMTGAKPRGPGMERADLLQNGEIFQVQGRALNESASRNCKVLVGVNPCNTNALIAMENA	210
	::*****: * : **:* **.* ** ** * .*****: ::**	
Sorghum	PDIPAKNFHALTRLDENRAKQQLALKAGVFYDKVSNVTIWGNHSTTQVPDFLNAKIDGRP	286
Oryza	PNIPAKNFHALTRLDENRAKQQLALKAGVFYDKVSNMTIWGNHSTTQVPDFLNAKINGRP	290
Arabidopsis	PNIPAKNFHALTRLDENRAKQQLALKAGVFYDKVSNMTIWGNHSTTQVPDFLNARINGLP	298
Chlamydomonas	PNIPRKNFHALTRLDENRAKQQLALKSGKPYTSVSRMAIWGNHSTTQVPDFVNARIGGLP	270
	*:* *****:* ** .: *****:*. * . *	
Sorghum	VKEVIKDTKWLEEFITTVQKRGGALIQKWGRSSAASTAVSIADAIKSLVTPPEGDWFS	346
Oryza	VKEVIKDTKWLEDFITTVQKRGGALIQKWGRSSAASTAVSIVDAIRSLVNTPPEGDWFS	350
Arabidopsis	VKEVITDHWLEEGFTESVQKRGGLLIQKWGRSSAASTAVSIVDAIKSLVTPPEGDWFS	358
Chlamydomonas	APDVIRDMKWFREEFTPKVALRGGALIKKWGRSSAASTAVSVADAIRALVVPTAPGDCFS	330
	. :* * **.: * . * ** * :*****:*****:* ** . * **	
Sorghum	TGVYTTGNPYGIAEDIVFSMPCRSKGDDYELATDVSMDDFLWERIKKSEAELLAEKRCV	406
Oryza	TGVYTTGNPYGIAEDIVFSMPCRSKGDDYELVKDVAMDDFLWERIKKSEAELLAEKRCV	410
Arabidopsis	TGVYTDGNPYGI EELVFSMPCRSKGDDYELVKDVETD DYLQRIAKSEAELLAEKRCV	418
Chlamydomonas	TGVIDSDGNPYGREGLIFSPCRSKGDDYEICDNFIVDEWLRKIRASEDELQKEKECV	390
	** : *****: * .: **:* *****: . : .:* * : * ** ** * .**	
Sorghum	AHLTGEENAYCDV--PEDTMLPGEV	429
Oryza	AHLTGEENAFCDL--PGDTMLPGEV	433
Arabidopsis	AHLTGEIAYCDL--PVDTMLPGEV	442
Chlamydomonas	SHLIGMMGGSCALRGAEEDTTVPGEN	415
	::** * . * : . ** :***	

The amino-acid sequences were generated from the DNA sequences. The overall identities are: Arabidopsis NADP-MDH with Oryza 78%, with Sorghum 76% and with Chlamydomonas 62%. NADP-MDH from Sorghum is 82% identical with Oryza and 62% with Chlamydomonas. The identity between Oryza and Chlamydomonas is 60%. The regulatory cysteine residues are colored yellow

NADP-MDH. In the case of Brassicaceae, this is within the coding region of At5g58340. In overlapping region 1, the promoter sequences of eight Brassicaceae species, namely *A. thaliana* var. Columbia, *A. thaliana* var. Landsberg *erecta*, *A. thaliana* var. Wassilewskija, *Cardaminopsis petraea*, *Capsella bursa-pastoris*, *Arabis turrta*, *Arabidopsis wallichii*, and *Cochlearia officinalis*, are compared with *Oryza sativa* and the two Sorghum sequences. The summary (Table 4A) indicates high homologies, which are similar to the introns or exons of NADP-MDH within this region, while (with the exception of Sorghum I and II) this region is not conserved in Poaceae.

Overlapping region 2 has a length of 242 bp and lies between -76 and -318 bp, which is between the start of the NADP-MDH and At5g58340, and covers a large part of the 5' UTR of At5g58340. In overlapping region 2, the sequences of four Brassicaceae (*A. thaliana* var. Landsberg

*erecta*, *Capsella bursa-pastoris*, *Arabis turrta*, and *Arabidopsis wallichii*) are compared with NADP-MDH from *Oryza sativa* and the two Sorghum sequences. The degree of identity is not as high as in overlapping region 1, but with an average of 80%, it is much higher than in Poaceae, where only values below 10% were obtained (Table 4B). The gene from Chlamydomonas was not included in this comparison, because it differs in all aspects (promoter, intron structure, and length) from the higher plant genes.

#### Analysis of Knock-Out Plants for At5g58340

In order to assess whether the emerging picture, namely the lack of regulatory elements in 5' position of the NADP-MDH gene, but regulatory elements located within the coding region, can be confirmed by a more direct method,

**Table 2** Nucleotide identities of the NADP-MDH genes of selected Brassicaceae

Exon	Species	<i>A. th.</i>	<i>C. b.-p.</i>	<i>C. rub.</i>	Intron	Species	<i>A. th.</i>	<i>C. b.-p.</i>	<i>C. rub.</i>
1	<i>A. th.</i>	1.00	.	.	1	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.96	1.00	.		<i>C. b.-p.</i>	0.82	1.00	.
	<i>C. rub.</i>	0.95	0.98	1.00		<i>C. rub.</i>	0.82	0.97	1.00
2	<i>A. th.</i>	1.00	.	.	2	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.94	1.00	.		<i>C. b.-p.</i>	0.75	1.00	.
	<i>C. rub.</i>	0.94	1.00	1.00		<i>C. rub.</i>	0.74	0.94	1.00
3	<i>A. th.</i>	1.00	.	.	3	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.91	1.00	.		<i>C. b.-p.</i>	0.86	1.00	.
	<i>C. rub.</i>	0.88	0.96	1.00		<i>C. rub.</i>	0.87	0.97	1.00
4	<i>A. th.</i>	1.00	.	.	6	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.92	1.00	.		<i>C. b.-p.</i>	0.73	1.00	.
	<i>C. rub.</i>	0.92	0.97	1.00		<i>C. rub.</i>	0.72	1.00	1.00
7	<i>A. th.</i>	1.00	.	.	7	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.96	1.00	.		<i>C. b.-p.</i>	0.82	1.00	.
	<i>C. rub.</i>	0.96	1.00	1.00		<i>C. rub.</i>	0.82	1.00	1.00
8	<i>A. th.</i>	1.00	.	.	8	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.96	1.00	.		<i>C. b.-p.</i>	0.82	1.00	.
	<i>C. rub.</i>	0.96	1.00	1.00		<i>C. rub.</i>	0.82	1.00	1.00
9	<i>A. th.</i>	1.00	.	.	9	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.96	1.00	.		<i>C. b.-p.</i>	0.83	1.00	.
	<i>C. rub.</i>	0.96	1.00	1.00		<i>C. rub.</i>	0.84	1.00	1.00
10	<i>A. th.</i>	1.00	.	.	10	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.95	1.00	.		<i>C. b.-p.</i>	0.86	1.00	.
	<i>C. rub.</i>	0.96	0.92	1.00		<i>C. rub.</i>	0.89	0.86	1.00
11	<i>A. th.</i>	1.00	.	.	11	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.89	1.00	.		<i>C. b.-p.</i>	0.85	1.00	.
	<i>C. rub.</i>	0.87	0.93	1.00		<i>C. rub.</i>	0.76	0.90	1.00
12	<i>A. th.</i>	1.00	.	.	12	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.87	1.00	.		<i>C. b.-p.</i>	0.80	1.00	.
	<i>C. rub.</i>	0.96	0.90	1.00		<i>C. rub.</i>	0.89	0.79	1.00
13	<i>A. th.</i>	1.00	.	.	13	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.90	1.00	.		<i>C. b.-p.</i>	0.67	1.00	.
	<i>C. rub.</i>	0.92	0.90	1.00		<i>C. rub.</i>	0.85	0.73	1.00
14	<i>A. th.</i>	1.00	.	.	14	<i>A. th.</i>	1.00	.	.
	<i>C. b.-p.</i>	0.91	1.00	.		<i>C. b.-p.</i>	0.91	1.00	.
	<i>C. rub.</i>	0.96	0.90	1.00		<i>C. rub.</i>	0.96	0.90	1.00

Genomic DNA sequences were obtained by DNA sequencing. *A. th.*, *Arabidopsis thaliana* var. Columbia; *C. b.-p.*, *Capsella bursa-pastoris*; *C. rub.*, *Capsella rubella*. The values were calculated on the basis of the genomic sequence of *Arabidopsis thaliana* var. Columbia (clone MCK7; TAIR, <http://www.arabidopsis.org>), using the ClustalW 1.82 software. The numbering of the introns and exons is as indicated in Fig. 6

knock-out plants that carry large T-DNA inserts in the proximal and distal promoter region were analyzed. Homozygous lines were selected as described in the Materials and Methods section, and the positions of the inserts were confirmed by DNA sequencing. Unfortunately, the expression level of At5g58340 is extremely low, and it was neither possible to analyze its expression in wild-type (WT) plants, nor to confirm the lack of expression in the homozygous knock-out plants.

In line ko1, the insert is located 420 bp in front of the ATG of NADP-MDH. This is outside the 5' UTR of NADP-MDH, but inside the first exon of At5g58340. In

line ko2, the insert position is 1524 bp in front of the ATG, which is within the second exon of At5g58340 (Fig. 7A). The position of the inserts is shown in Fig. 7A. In both knock-out lines, the expression of functional At5g58340 should be abolished. However, the homozygous knock-out plants did not display any visible differences to the WT plants. Also the expression of NADP-MDH was unaltered (Figs. 7B and 7C). This confirms our conclusion that no regulatory elements are located in the classical proximal and distal promoter region. However, it cannot be ruled out that regulatory elements are located in the 420 bp between the ATG and the position of the insert in ko1.

**Table 3** Nucleotide identities of the NADP-malate dehydrogenase (MDH) gene from *A. thaliana* var. Columbia and selected Poaceae

Exon	Species	<i>A. th.</i>	<i>S. v. I</i>	<i>S. v. II</i>	<i>O. s.</i>	Intron	Species	<i>A. th.</i>	<i>S. v. I</i>	<i>S. v. II</i>	<i>O. s.</i>
1	<i>A. th.</i>	<b>1.00</b>	.	.	.	1	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.09	<b>1.00</b>	.	.		<i>S. v. I</i>	0.10	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.07	0.83	<b>1.00</b>	.		<i>S. v. II</i>	0.09	0.46	<b>1.00</b>	.
	<i>O. s.</i>	0.14	0.17	0.16	<b>1.00</b>		<i>O. s.</i>	0.23	0.12	0.10	<b>1.00</b>
2	<i>A. th.</i>	<b>1.00</b>	.	.	.	2	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.40	<b>1.00</b>	.	.		<i>S. v. I</i>	0.10	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.34	0.90	<b>1.00</b>	.		<i>S. v. II</i>	0.10	0.19	<b>1.00</b>	.
	<i>O. s.</i>	0.35	0.68	0.79	<b>1.00</b>		<i>O. s.</i>	0.10	0.09	0.05	<b>1.00</b>
3	<i>A. th.</i>	<b>1.00</b>	.	.	.	3	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.75	<b>1.00</b>	.	.		<i>S. v. I</i>	0.14	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.74	0.95	<b>1.00</b>	.		<i>S. v. II</i>	0.13	0.82	<b>1.00</b>	.
	<i>O. s.</i>	0.75	0.80	0.80	<b>1.00</b>		<i>O. s.</i>	0.25	0.21	0.20	<b>1.00</b>
4	<i>A. th.</i>	<b>1.00</b>	.	.	.	4	<i>A. th.</i>	-	.	.	.
	<i>S. v. I</i>	0.76	<b>1.00</b>	.	.		<i>S. v. I</i>	-	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.74	0.91	<b>1.00</b>	.		<i>S. v. II</i>	-	0.40	<b>1.00</b>	.
	<i>O. s.</i>	0.75	0.82	0.85	<b>1.00</b>		<i>O. s.</i>	-	0.47	0.54	<b>1.00</b>
5	<i>A. th.</i>	-	.	.	.	5	<i>A. th.</i>	-	.	.	.
	<i>S. v. I</i>	-	<b>1.00</b>	.	.		<i>S. v. I</i>	-	<b>1.00</b>	.	.
	<i>S. v. II</i>	-	0.92	<b>1.00</b>	.		<i>S. v. II</i>	-	0.64	<b>1.00</b>	.
	<i>O. s.</i>	-	0.88	0.83	<b>1.00</b>		<i>O. s.</i>	-	0.41	0.38	<b>1.00</b>
6	<i>A. th.</i>	-	.	.	.	6	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	-	1.00	.	.		<i>S. v. I</i>	0.09	<b>1.00</b>	.	.
	<i>S. v. II</i>	-	0.97	<b>1.00</b>	.		<i>S. v. II</i>	0.13	0.76	<b>1.00</b>	.
	<i>O. s.</i>	-	0.82	0.87	<b>1.00</b>		<i>O. s.</i>	0.14	0.49	0.25	<b>1.00</b>
7	<i>A. th.</i>	<b>1.00</b>	.	.	.	7	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.77	<b>1.00</b>	.	.		<i>S. v. I</i>	0.21	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.80	0.94	<b>1.00</b>	.		<i>S. v. II</i>	0.19	0.94	<b>1.00</b>	.
	<i>O. s.</i>	0.81	0.90	0.91	<b>1.00</b>		<i>O. s.</i>	0.50	0.21	0.28	<b>1.00</b>
8	<i>A. th.</i>	<b>1.00</b>	.	.	.	8	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.77	<b>1.00</b>	.	.		<i>S. v. I</i>	0.11	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.76	0.98	<b>1.00</b>	.		<i>S. v. II</i>	0.14	0.75	<b>1.00</b>	.
	<i>O. s.</i>	0.88	0.90	0.89	<b>1.00</b>		<i>O. s.</i>	0.29	0.18	0.40	<b>1.00</b>
9	<i>A. th.</i>	<b>1.00</b>	.	.	.	9	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.70	<b>1.00</b>	.	.		<i>S. v. I</i>	0.19	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.69	0.95	<b>1.00</b>	.		<i>S. v. II</i>	0.27	0.82	<b>1.00</b>	.
	<i>O. s.</i>	0.72	0.87	0.85	<b>1.00</b>		<i>O. s.</i>	0.25	0.25	0.26	<b>1.00</b>
10	<i>A. th.</i>	<b>1.00</b>	.	.	.	10	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.73	<b>1.00</b>	.	.		<i>S. v. I</i>	0.20	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.74	0.92	<b>1.00</b>	.		<i>S. v. II</i>	0.34	0.90	<b>1.00</b>	.
	<i>O. s.</i>	0.77	0.86	0.89	<b>1.00</b>		<i>O. s.</i>	0.32	0.15	0.63	<b>1.00</b>
11	<i>A. th.</i>	<b>1.00</b>	.	.	.	11	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.78	<b>1.00</b>	.	.		<i>S. v. I</i>	0.27	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.78	<b>1.00</b>	<b>1.00</b>	.		<i>S. v. II</i>	0.19	0.89	<b>1.00</b>	.
	<i>O. s.</i>	0.81	0.94	0.94	<b>1.00</b>		<i>O. s.</i>	0.22	0.35	0.12	<b>1.00</b>
12	<i>A. th.</i>	<b>1.00</b>	.	.	.	12	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.59	<b>1.00</b>	.	.		<i>S. v. I</i>	0.13	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.59	0.98	<b>1.00</b>	.		<i>S. v. II</i>	0.17	0.96	<b>1.00</b>	.
	<i>O. s.</i>	0.63	0.89	0.88	<b>1.00</b>		<i>O. s.</i>	0.19	0.29	0.30	<b>1.00</b>

**Table 3** continued

Exon	Species	<i>A. th.</i>	<i>S. v. I</i>	<i>S. v. II</i>	<i>O. s.</i>	Intron	Species	<i>A. th.</i>	<i>S. v. I</i>	<i>S. v. II</i>	<i>O. s.</i>
13	<i>A. th.</i>	<b>1.00</b>	.	.	.	13	<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.77	<b>1.00</b>	.	.		<i>S. v. I</i>	0.17	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.77	<b>1.00</b>	<b>1.00</b>	.		<i>S. v. II</i>	0.17	0.86	<b>1.00</b>	.
	<i>O. s.</i>	0.75	0.92	0.92	<b>1.00</b>		<i>O. s.</i>	0.16	0.29	0.11	<b>1.00</b>
14	<i>A. th.</i>	<b>1.00</b>	.	.	.		<i>A. th.</i>	<b>1.00</b>	.	.	.
	<i>S. v. I</i>	0.74	<b>1.00</b>	.	.		<i>S. v. I</i>	0.17	<b>1.00</b>	.	.
	<i>S. v. II</i>	0.70	0.79	<b>1.00</b>	.		<i>S. v. II</i>	0.17	0.86	<b>1.00</b>	.
	<i>O. s.</i>	0.74	0.81	0.83	<b>1.00</b>		<i>O. s.</i>	0.16	0.29	0.11	<b>1.00</b>

Genomic DNA sequences for Sorghum and rice were obtained from database sources. *A. th.*, *Arabidopsis thaliana* var. Columbia; *S. v. I*, *Sorghum vulgare* I; *S. v. II*, *Sorghum vulgare* II; *O. s.*, *Oryza sativa*. The values were calculated on the basis of the genomic sequence of *Arabidopsis thaliana* var. Columbia (clone MCK7). The numbering of introns and exons is as indicated in Fig. 6

### DNA Binding Proteins Identified by YOH

A YOH screen for DNA binding proteins was performed with *A. thaliana* var. Columbia genomic DNA. The YOH screens were performed with one fragment (F1) that covers the 5' UTR of the NADP-MDH gene and the first exon (−411 to 106) and four fragments located inside the coding region (Fig. 6). Fragment 5 (F5) starts in the first exon (106

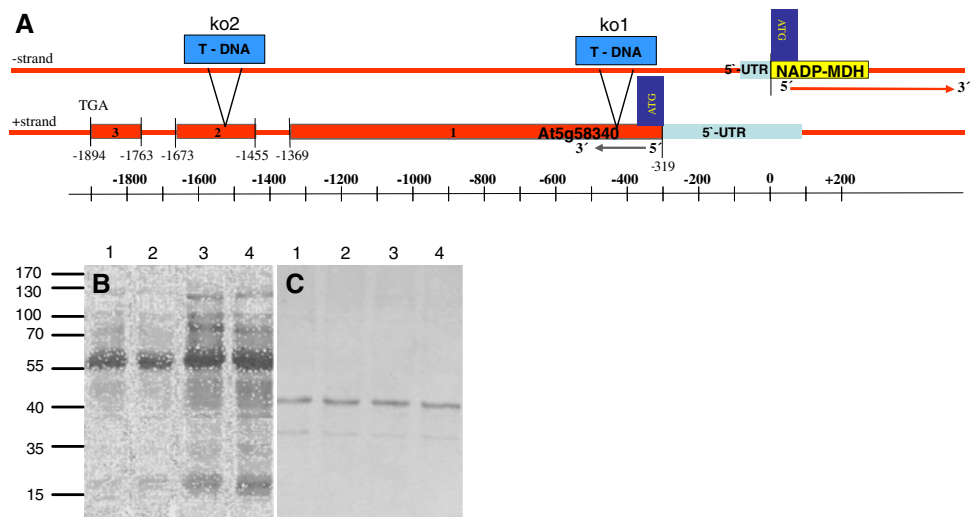
bp behind the ATG) and has a total size of 1740 bp. Fragment 9 (F9) spans the second half of the coding region and part of the 3' UTR. It starts at position 1211 and ends at position 2788. Fragment 7 (F7) covers a part of the 3' UTR and reaches from 2372 to 2530, and fragment 10 (F10) reaches from 1211 to 2437. The position of the fragments is indicated in Fig. 6. For the classical promoter region (F1), only three different proteins could be identified (Table 5).

**Table 4** Nucleotide identities within the 5' region of the NADP-MDH gene

Genomic DNA sequences were either obtained by DNA sequencing or from databank sources (*S. vulgare* I and II, *O. sativa*, *C. reinhardtii*). *A. th. Col.*, *Arabidopsis thaliana* var. Columbia; *A. th. Ler.*, *Arabidopsis thaliana* var. Landsberg erecta; *A. th. Was.*, *Arabidopsis thaliana* var. Wassilevskija; *A. tu.*, *Arabidopsis turruta*; *A. wa.*, *Arabidopsis wallichii*; *C. b.-p.*, *Capsella bursa-pastoris*; *C. p.*, *Cardaminopsis petraia*; *C. rub.*, *Capsella rubella*; *C. of.*, *Cochlearia officinalis*; *O. s.*, *Oryza sativa*; *S. v. I*, *Sorghum vulgare* I; *S. v. II*, *Sorghum vulgare* II. The values were calculated on the basis of the genomic sequence of *A. thaliana* var. Columbia (clone MCK7). The size and position of the overlapping regions 1 and 2 are indicated in Fig. 5

A: overlapping region 1											
	<i>A. th. Col</i>	<i>A. th. Ler</i>	<i>A. th. Was</i>	<i>C. p</i>	<i>C. b.-p.</i>	<i>A. tu.</i>	<i>A.wa.</i>	<i>C. of.</i>	<i>O. s.</i>	<i>S. v. I</i>	<i>S. v. II</i>
<i>A. th. Col</i>	<b>1.00</b>	.	.	.	.	.	.	.	.	.	.
<i>A. th. Ler</i>	0.97	<b>1.00</b>	.	.	.	.	.	.	.	.	.
<i>A. th. Was</i>	0.90	0.87	<b>1.00</b>	.	.	.	.	.	.	.	.
<i>C. p.</i>	0.87	0.85	0.92	<b>1.00</b>	.	.	.	.	.	.	.
<i>C. b.-p.</i>	0.88	0.86	0.88	0.85	<b>1.00</b>	.	.	.	.	.	.
<i>A. tu.</i>	0.90	0.87	0.87	0.85	0.88	<b>1.00</b>	.	.	.	.	.
<i>A. wa.</i>	0.88	0.86	0.84	0.83	0.89	0.88	<b>1.00</b>	.	.	.	.
<i>C. of.</i>	0.79	0.77	0.74	0.75	0.77	0.81	0.80	<b>1.00</b>	.	.	.
<i>O. s.</i>	0.04	0.09	0.04	0.05	0.02	0.02	0.04	0.05	<b>1.00</b>	.	.
<i>S. v. I</i>	0.02	0.02	0.04	0.03	0.02	0.03	0.02	0.06	0.05	<b>1.00</b>	.
<i>S. v. II</i>	0.06	0.05	0.07	0.06	0.06	0.06	0.06	0.05	0.02	0.63	<b>1.00</b>
B: overlapping region 2											
	<i>A. th. Col</i>	<i>A. th. Ler</i>	<i>A. wa.</i>	<i>A. tu.</i>	<i>C. b.-p.</i>	<i>O. s.</i>	<i>S. v. I</i>	<i>S. v. II</i>			
<i>A. th. Col</i>	<b>1.00</b>	.	.	.	.	.	.	.			
<i>A. th. Ler</i>	0.99	<b>1.00</b>	.	.	.	.	.	.			
<i>A. wa.</i>	0.84	0.84	<b>1.00</b>	.	.	.	.	.			
<i>A. tu.</i>	0.72	0.72	0.77	<b>1.00</b>	.	.	.	.			
<i>C. b.-p.</i>	0.80	0.81	0.84	0.73	<b>1.00</b>	.	.	.			
<i>O. s.</i>	0.06	0.06	0.10	0.09	0.06	<b>1.00</b>	.	.			
<i>S. v. I</i>	0.05	0.05	0.04	0.06	0.12	0.06	<b>1.00</b>	.			
<i>S. v. II</i>	0.04	0.04	0.05	0.10	0.05	0.08	0.69	<b>1.00</b>			

**Fig. 7** NADP-MDH expression in knock-out plants for At5g58340. **(A)** The position of the inserts in lines ko1 and ko2, as confirmed by DNA sequencing, is indicated; **(B)** sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE); **(C)** Western blot using an antibody raised against NADP-MDH from pea leaves. For both gels, a total amount of 30  $\mu$ g protein was used. Lanes 1 and 3, WT; lane 2, homozygous At5g58340-ko1; lane 4, homozygous At5g58340-ko2



Two of them have an unknown function, and there is little evidence that they are involved in transcriptional regulation by a known mechanism.

BANG 1-1 (At4g08240), the first protein that was found to bind to F1, has a calculated molecular weight of 14.5 kDa, and is predicted to possess between three and five  $\alpha$ -helices. The three-dimensional structure could not be predicted and no known domains could be identified. It is characterized by a very high content of leucine residues, and the serine and threonine contents are also unusually high. A similar protein (55% identity) was only found in the rice genome (P0020D05.17, classified as ‘unknown protein’), but not in any other organism. The second protein, BANG 1-2 codes for At5g11680. The deduced amino acid sequence consists of 207 amino acids and has a calculated molecular weight of 23.1 kDa. No three-dimensional structure could be predicted and no known domains could be recognized. Again, only a protein identified within the rice genome has a high identity (75%) with BANG 1–2. Psort (<http://www.expasy.ch>) predicts that both proteins are localized in the cytosol. However, for such small proteins a nuclear localization signal (NLS) would not be required for nuclear translocation (Yoneda 1997). The third protein that binds to F1 was identified as EIN2 (BANG 1–3), an ethylene receptor (Chang and Shockey 1999). However, since this protein is probably membrane-associated (Alonso et al. 1999), it is not very likely that it can bind to the NADP-MDH gene, although it has been reported that clathrin-coated vesicles may provide such possibility (Benmerah et al. 2003).

In additional YOH screens, fragments that cover most of the coding region of the NADP-MDH gene were used. Fragment 5 covers introns and exons 1 to 10. Fragment 10 slightly overlaps with F5, and covers introns 8 to 13 and exons 8 to 14. A total of 34 proteins were identified as binding to the coding region. The Arabidopsis gene

numbers and the description are summarized in Table 5. The properties of the proteins are summarized in Table 6. For 10 of the proteins, Psort predicts a nuclear localization, and for 19 additional proteins, a nuclear localization is possible, even without NLS, due to their small size. In eight cases, literature data indicate nuclear localization or the ability to bind DNA.

## Discussion

### Regulation of NADP-MDH Expression in Brassicaceae and Poaceae

Under in vivo conditions, a complex pattern for the transcriptional regulation of NADP-MDH is found. Changes in the expression of NADP-MDH are caused by exogenous factors (e.g., CO<sub>2</sub> availability, light intensity, light period, or temperature) and endogenous factors (plant and leaf age, flowering). This transcriptional regulation is not unique in Arabidopsis and other Brassicaceae; it was also observed in other species such as wheat (Savitch et al. 2000) and tobacco (Faske et al. 1997). Unfortunately, no data are available as to whether changes in NADP-MDH expression also occur in Chlamydomonas or related algae, under changing environmental conditions.

Although the amino acid sequence of NADP-MDH is similar in Chlamydomonas, Poaceae and Brassicaceae (Table 1), the gene structure displays striking differences (Fig. 6). In Poaceae, typical promoter elements could be identified. The Chlamydomonas NADP-MDH gene also possesses a classical promoter, but it does not share conserved motifs with the Poaceae promoter. In addition, the gene structure differs between the two. Although the Brassicaceae gene is organized in a similar way as in Poaceae, another gene is located close to the transcription

**Table 5** List of proteins identified by the YOH with fragments of the NADP-MDH gene

Internal name	Sequence alignment with TAIR database (MAtdB)	Size of the CDS (kb)	Common name	Predicted protein size (kDa)
<b>BANG1-1</b>	At4g08240	0.4	Expressed protein	14.5
<b>BANG1-2</b>	At5g11680	0.6	Unknown protein	23.1
<b>BANG1-3</b>	At5g03280	4.0	Ethylene-insensitive 2 (EIN2)	140.9
<b>BANG 5-1</b>	At2g41230	0.27	Expressed protein	9.6
<b>BANG 5-2</b>	At2g32280	0.5	Expressed protein	17.5
<b>BANG 5-3</b>	At5g41080	1.3	Glycerophosphoryl-diester phosphodiesterase	42
<b>BANG 5-4</b>	At3g16640	0.5	Translationally controlled tumor-family protein	18.9
<b>BANG 5-5</b>	At1g13440	1.0	Cytosolic GAPDH C1	37
<b>BANG 5-6</b>	At2g40700	1.8	DEAD/DEAH helicase	68.2
<b>BANG 5-7</b>	At3g17390 / At3g17380	0.9	S-adenosyl-methionine synthetase/ meprin and TRAF-homology domain containing protein	42.7
<b>BANG 5-8</b>	At1g07230	1.6	Phosphoesterase-family protein	60
<b>BANG 5-9</b>	At5g14170	1.6	SWIB complex	59
<b>BANG 5-10</b>	At1g04210	3.3	Leucine-rich repeat family protein	125
<b>BANG 5-11</b>	At5g59890	0.4	Actin-depolymerizing factor 4	16
<b>BANG 5-12</b>	At4g32720	1.3	RNA-recognition motif (RRM)	48
<b>BANG 7-1</b>	At5g60390	1.35	Elongation factor 1 alpha	49.5
<b>BANG 7-2</b>	At1g25350	2.4	Glutamine-tRNA ligase	91.2
<b>BANG 7-3</b>	At4g09320	0.5	Nucleoside-diphosphate kinase (NDK1)	18.8
<b>BANG 7-4</b>	At1g56070	2.5	Elongation factor 2	94
<b>BANG 7-5</b>	At3g03100	0.5	NADH:ubiquinone oxidoreductase family	18.3
<b>BANG 9-1</b>	At3g60450	0.8	Expressed protein	30.8
<b>BANG 9-2</b>	At3g41950	2.74	Not found (TAIR)	100
<b>BANG 9-3</b>	At4g24210	0.45	F-protein “sleepy”	17.4
<b>BANG 10-1</b>	At4g17790	0.8	Expressed protein	29
<b>BANG 10-2</b>	At4g10280	0.4	Expressed protein	15.7
<b>BANG 10-3</b>	At3g26890	2.0	Expressed protein	70.5
<b>BANG 10-4</b>	At1g65290	0.6	Acyl-carrier family	14
<b>BANG 10-5</b>	At1g33980	1.44	Smg-4/UPF3 family protein	54
<b>BANG 10-6</b>	At5g59550	1.2	Zinc finger (C3HC4-type RING finger)	44.3
<b>BANG 10-7</b>	At2g36910	3.9	Multidrug-resistance P-glycoprotein (PGP1)	141
<b>BANG 10-8</b>	At2g40510	0.4	40S ribosomal protein S26	14.8
<b>BANG 10-9</b>	At4g09320	0.5	Nucleoside-diphosphate kinase 1 (NDK1)	18.8
<b>BANG 10-10</b>	At3g52930 / At2g36460	1.0	Fructose-bisphosphate aldolase	38
<b>BANG 10-11</b>	At2g19730	0.4	60S ribosomal protein L28	15.8
<b>BANG 10-12</b>	At5g13070	0.55	MSF1-like family protein	21
<b>BANG 10-13</b>	At5g10360	0.75	40S ribosomal protein	28.1
<b>BANG 10-14</b>	At1g13750	1.8	Calcineurin-like phosphoesterase family	68
<b>BANG 10-15</b>	At1g43170	1.1	60S ribosomal protein	44.5
<b>BANG 10-16</b>	At1g55030	1.25	F-box family protein	48.1

The proteins listed below have been identified using the YOH system, as described in the Materials and Methods section. The localization of the fragments is indicated in Fig. 6

start of NADP-MDH in Brassicaceae, and promoter elements are absent from the 5' region. The results of the YOH screen, sequence alignments, and the NADP-MDH expression in Arabidopsis knock-out plants for At5g58340 provide evidence that in Brassicaceae most regulatory

elements, which are responsible for the many-folded environmental effects on NADP-MDH expression, must be located inside the coding region of the gene.

In the knock-out lines for At5g58340, a large insert disrupts the DNA region where the proximal and the distal

**Table 6** Properties of putative NADP-MDH interaction partners

Interaction partner	Fully unknown	cDNA-restriction fragment fits to the size of the cDNA	Full-length sequence	Nuclear localization predicted	Evidence from the literature	Smaller than 45 kDa
BANG 1-1	+	+	+			+
BANG 1-2	+	+	+			+
BANG 1-3						
BANG 5-1	+	+			+	+
BANG 5-2	+	+	+			+
BANG 5-3			+			+
BANG 5-4		+	+			+
BANG 5-5		+	+		+	+
BANG 5-6		+		+		
BANG 5-7						+
BANG 5-8			+		+	
BANG 5-9		+		+		
BANG 5-10				+		
BANG 5-11		+	+			+
BANG 5-12				+		
BANG 7-1						
BANG 7-2			+			
BANG 7-3		+	+		+	+
BANG 7-4				+		
BANG 7-5			+	+		+
BANG 9-1	+	+				+
BANG 9-2						+
BANG 9-3		+	+		+	+
BANG 10-1	+	+			+	+
BANG 10-2	+					+
BANG 10-3	+					
BANG 10-4		+				+
BANG 10-5						
BANG 10-6		+	+	+		+
BANG 10-7						
BANG 10-8		+	+	+		+
BANG 10-9		+	+		+	+
BANG 10-10		+	+		+	+
BANG 10-11		+	+	+		+
BANG 10-12		+	+			+
BANG 10-13		+	+	+		+
BANG 10-14						
BANG 10-15		+	+	+		+
BANG 10-16					+	+

The interaction partners of the NADP-MDH gene were tested by a restriction analysis of the cDNA-carrying vectors from yeast. The possibility that the proteins are able to move into the nucleus was assessed by Psort, literature data, and by the molecular weight. Proteins smaller than approximately 45 kDa should be able to enter the nucleus without nuclear localization signals. Positive evidence is marked by +; empty field indicates no evidence

promoters of typical genes should be located. However, both knock-out lines are indistinguishable from WT plants in all aspects. In particular, the expression of NADP-MDH is unaffected. It was expected that interruption of the promoter by large DNA inserts would affect gene

expression, because regulatory elements are removed from their original positions. This supports the results obtained from motif prediction programs, which indicate that no regulatory elements are located in 5' position of the NADP-MDH gene of Brassicaceae. With the YOH screen, only

three proteins that bind to the 5' region were identified, while 34 proteins appear to be able to bind to the coding region. However, most of the proteins are as yet unknown. Bioinformatical analysis confirmed that some of the proteins (BANG 5-6, BANG 5-9, BANG 5-10, BANG 5-12, BANG 7-4, BANG 7-5, BANG 10-6, BANG 10-8, BANG 10-11, BANG 10-13, and BANG 10-15) contain an NLS sequence, which would facilitate nuclear localization. Since more possibilities exist for nuclear import, the absence of an NLS does not automatically mean that nuclear localization is impossible. Future work will focus on the identification of DNA-binding proteins *in vitro*, and identification of their target motifs.

Secondly, the comparison of sequence identities in the coding and non-coding parts of the NADP-MDH gene indicates large differences between the Brassicaceae and the other organisms. The position of NADP-MDH in Poaceae is covered in Brassicaceae by another gene on the opposite DNA strand. The first exon of At5g58340 starts at -319 bp, before the transcription start of NADP-MDH. It is 1050 bp long and covers a large part of the proximal and distal promoter region. As a consequence, this region is highly conserved within the Brassicaceae. The evolution rate is as low as is expected for coding regions. One single nucleotide exchange in every 500–1000 bp is the average value. For non-coding regions, a higher average rate of one exchange in every 200–500 bp would be typical. (Brumfield et al. 2003). In fact, within the Poaceae, there was no similarity between the sequences in this DNA segment. The same difference between Brassicaceae and Poaceae holds true for the DNA segment between NADP-MDH and At5g58340. In Brassicaceae it is even more conserved than are the exons of At5g58340.

Additional characterization of the 5' region of the NADP-MDH gene was performed by searching for *cis*-acting elements in transcription factor databases (Genomatix, MatInspector). For the genomic sequence of NADP-MDH from *A. thaliana* var. Columbia (At5g58330; T-DNA-Express), no potential binding sites for transcription factors could be identified, and a TATA box is also absent. For genes lacking the TATA box, a direct interaction between regulatory transcription factors, the TATA-box binding protein associated factor (TAF), and the preinitiation complex is possible and would also result in the binding of RNA-polymerase II (Wieczorek et al. 1998). Genes without a TATA box may be not so rare. The TATA box is absent from approximately one third of all Arabidopsis genes (Shahmuradov et al. 2005).

An unusually high degree of conservation was found not only for the 5' region of the NADP-MDH gene, but also for the introns in the Brassicaceae NADP-MDH gene. Intron size and position are highly conserved in Brassicaceae, but not in Poaceae. This indicates that some kind of selection

pressure prevents any exchange of nucleotides at a higher rate. Even the two genes of Sorghum NADP-MDH, which arise from a recent gene duplication (Luchetta et al. 1991; Rondeau et al. 2005), are more different than the introns of the Brassicaceae species sequenced in this work. Thus, three different types of promoters are found for the NADP-MDH gene in the different species. For *Chlamydomonas* and for rice, internet resources such as Gene2Promoter (<http://www.genomatix.de/>) predict a promoter, and a CCAAT box is present, while in Arabidopsis, the prediction programs do not find any promoter-like structures. This supports our conclusion that in the case of Brassicaceae, regulatory transcription-factor binding sites lie within the coding region of NADP-MDH, especially inside the introns. It seems that regulatory elements for NADP-MDH expression could not be inserted into the exon of another gene (At5g58340), even when this is encoded on the opposite DNA strand. One may expect that the loss or modification of At5g58340 may have serious consequences for fundamental processes in Arabidopsis, but the lack of any visible differences between the WT plants and both knock-out lines indicate that this is not the case.

Apart from the transcription-factor binding motifs, which are usually very small, promoters are not highly conserved. In particular, the distal promoter region is very variable, and there are some examples of distal promoter elements several kb away from the gene they regulate (Kirchhamer et al. 1996). There is evidence that downstream-activating sequences also play a role in transcriptional regulation. Dickey et al. (1992) demonstrated for the *fed-1* gene from *Pisum sativum*, that regulatory elements can be located inside the coding region, even within exons. *Fed-1* encodes the photosynthetic protein ferredoxin I. A light-responsive element was identified in the 5' UTR and in the first third of the coding region. With respect to the *Acl 1.4* gene, which encodes ACP4, an acyl-carrier protein, Bonaventure and Ohlrogge (2002) demonstrated that it also contains elements for transcriptional regulation within the coding region.

#### Evolution of NADP-MDH Structure and Function

It is interesting to ask how such a difference in gene structure could evolve. NADP-MDH is originally derived from NAD-malate dehydrogenase (NAD-MDH) and still displays several structural similarities to NAD-dependent malate dehydrogenases, especially around the active site. Apart from the altered coenzyme usage, the major difference between NAD- and NADP-MDH is the presence of extra amino acids in the C- and N-terminal extensions (Scheibe 1990). Both possess regulatory cysteine residues that act in two different ways to inactivate the enzyme upon oxidation



(in darkness) in the chloroplast (Miginiac-Maslow and Lancelin 2002). The NADP-MDH gene of *Chlamydomonas* possibly resembles a common ancestor of the recent Brassicaceae and Poaceae gene. It differs from both in possessing only the C-terminal disulfide bridge. At the N-terminus of the *Chlamydomonas* protein the sequence extension is already present, but it does not contain any cysteine residues (Ocheretina et al. 2000). The absence of the second regulatory disulfide bridge alters the activation/inactivation properties of NADP-MDH (Lemaire et al. 2005). Although the protein sequence of *Chlamydomonas* NADP-MDH is quite similar to higher plant NADP-MDH (Table 1), the gene structure differs remarkably. As summarized in Fig. 6, the *Chlamydomonas* gene possesses only eight introns, some of which are quite long, and their position is different from those in the Brassicaceae and Poaceae. In addition, it possesses a classical promoter in 5' direction.

The early evolution of NADP-MDH must have occurred in an atmosphere that consisted of nearly 1% CO<sub>2</sub> with only traces of oxygen. Later, when the oxygen concentration increased, eukaryotic algae developed the CO<sub>2</sub>-concentrating mechanism (CCM) to maintain a high CO<sub>2</sub>/O<sub>2</sub> ratio inside the chloroplasts. An enzyme with limited regulatory properties, as found for NADP-MDH of the recent *Chlamydomonas*, may be sufficient for a malate-valve function under such conditions. The ability to regulate the malate valve at the level of gene expression may have been no longer necessary. The insertion of another gene into the promoter (the progenitor of At5g58340 in this case) could occur without any negative consequences. During the evolution of multicellular organisms, the CCM of the algal type was lost, and after plants had entered land, the selection pressure changed again. Light intensities on land are not only much higher than in the water, but its intensity can also change more dramatically. In addition, the O<sub>2</sub>-partial pressure increased further, paralleled by a decrease of the atmospheric CO<sub>2</sub> content. The regulatory disulfide bridge on the N-terminal sequence of NADP-MDH possibly evolved first, allowing for more efficient post-translational regulation of the enzyme, but the introduction of additional regulatory motifs into the DNA was also required. Our data do not clearly indicate the point from which Poaceae and Brassicaceae went in different directions, but the last common ancestors probably existed about 65 million years ago. Since both have conserved cysteine residues on N- and C-terminus, their introduction possibly happened in a common ancestor. The time point of At5g58340 insertion is more difficult to estimate. One possibility is that At5g58340 was inserted into a common ancestor of both, which already had both regulatory disulfide bridges, and was later lost in Poaceae. The 5' region of *Arabidopsis* NADP-MDH can be aligned with the rice promoter to achieve 41% identity, but it must be questioned how far the algorithms used for such

an alignment are reliable, especially when gaps need to be introduced into noncoding DNA. In this case, the progenitor of At5g58340 was removed, and the elements required for the transcriptional regulation of NADP-MDH, now forming a new promoter, were introduced instead. However, it is also possible that At5g58340 was inserted later in an ancestor of the Brassicaceae, but not into the genome of the Poaceae progenitors. Our data only indicate that, at the position of At5g58340, Poaceae do not possess a coding region. Genes very similar to At5g58340 are only found in other Brassicaceae, but not in rice, Sorghum or *Chlamydomonas*.

With the exception of some Poaceae, most recent plants still possess only one copy of NADP-MDH. The last step in evolution of the NADP-MDH gene in Poaceae seems to be a gene duplication in some C<sub>4</sub> species such as Sorghum (Luchetta et al. 1991; Rondeau et al. 2005). This may be due to the special function of NADP-MDH in C<sub>4</sub> metabolism, requiring a much higher expression level in the mesophyll cells only. However, maize successfully performs the C<sub>4</sub> pathway with only one copy of the NADP-MDH gene.

In conclusion, our data support the hypothesis that NADP-MDH expression in Brassicaceae is regulated by a new type of promoter. We conclude that during evolution no regulatory elements could be introduced into an exon of another gene encoded on the opposite DNA strand without influencing its function. The NADP-MDH promoter can be described as a null–null promoter, because the NADP-MDH gene does not contain any regulatory elements within the 5' region. All regulatory elements must then be located either inside the coding region, or in the 3' UTR. Although all of the elements involved in transcriptional regulation may be located within the coding region, the expression of NADP-MDH follows the same complex pattern as in other organisms which use a null–core promoter, as is the case in Poaceae.

**Acknowledgments** The authors would like to thank Agnieszka Mickiewicz and Mareike Panz for their help with the YOH screens, Sabine Heuer for her help with DNA sequencing, and Heike Wolf-Wibbelmann and Sabine Steinbach for growing the plants. Further thanks are due to the Csaba Konecz, MPI für Züchtungsforschung, Köln, Germany for providing the *Arabidopsis* cDNA library for the YOH experiments, and to Herbert Hurka and the staff at the Botanischer Garten, Universität Osnabrück, for providing the Brassicaceae seeds, and Marcus Koch (Universität Heidelberg, Germany) for the genomic DNA. This work was financially supported by grants from the Deutsche Forschungsgemeinschaft (FOR 387, TP1 and EM 166/2-1).

## References

- Acarkan A, Rossberg M, Koch M, Schmidt R (2000) Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant J* 23:55–62

- Alonso JM, Hirayama T, Roman G, Nourizadeh S, Ecker JR (1999) EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science* 284:2148–2152
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Anderson JM, Chow WS, Park YI (1995) The grand design of photosynthesis: Acclimation of the photosynthetic apparatus to environmental cues. *Photosynth Res* 46:129–139
- Backhausen JE, Scheibe R (1999) Adaptation of tobacco plants to elevated CO<sub>2</sub>: Influence of leaf age on changes in physiology, redox state and NADP-malate dehydrogenase activity. *J Exp Bot* 50:665–675
- Becker B, Holtgreffe S, Jung S, Wunrau C, Kandlbinder A, Baier M, Dietz K-J, Backhausen JE, Scheibe R (2006) Influence of the photoperiod on redox regulation and stress responses in *Arabidopsis thaliana* L. (Heynh.) plants under long- and short-day conditions. *Planta* 224:380–393
- Benmerah A, Scott M, Poupon V, Marullo S (2003) Nuclear functions for plasma membrane-associated proteins? *Traffic* 4:503–511
- Boivin K, Acarkan A, Mbulu RS, Clarenz O, Schmidt R (2004) The *Arabidopsis* genome sequence as a tool for genome analysis in *Brassicaceae*. A comparison of the *Arabidopsis* and *Capsella rubella* genomes. *Plant Physiol* 135:735–744
- Bonaventure G, Ohlrogge JB (2002) Differential regulation of mRNA levels of acyl carrier protein isoforms in *Arabidopsis*. *Plant Physiol* 128:223–235
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254
- Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* 18:249–256
- Buchanan BB (1980) Role of light in the regulation of chloroplast enzymes. *Annu Rev Plant Physiol Plant Mol Biol* 31:341–374
- Chang C, Shockey JA (1999) The ethylene-response pathway: Signal perception to gene regulation. *Curr Opin Plant Biol* 2:352–358
- Dickey LF, Gallo-Meagher M, Thompson WF (1992) Light regulatory sequences are located within the 5' portion of the *Fed-1* message sequence. *EMBO J* 11:2311–2317
- Faske M, Backhausen JE, Sendker M, Singer-Bayrle M, Scheibe R, von Schaewen A (1997) Transgenic tobacco plants expressing a heterologous pea chloroplast NADP-MDH cDNA in sense and antisense orientation: Effect on NADP-MDH level, stability of transformants and plant growth. *Plant Physiol* 115:705–715
- Fey V, Wagner R, Brautigam K, Wirtzt M, Hell R, Dietzmann A, Leister D, Oelmüller R, Pfanschmidt T (2005) Retrograde plastid redox signals in the expression of nuclear genes for chloroplast proteins of *Arabidopsis thaliana*. *J Biol Chem* 280:5318–5328
- Graeve K, von Schaewen A, Scheibe R (1994) Purification, characterization, and cDNA sequence of glucose-6-phosphate dehydrogenase from potato (*Solanum tuberosum* L.). *Plant J* 5:353–361
- Holtgreffe S, Voss I, Jung S, Khorobrykh S, Scheibe R, Backhausen JE (2007) Altered high light acclimation in transgenic potato (*Solanum tuberosum* L. var. Désirée) plants with decreased leaf-ferredoxin content. *Plant Physiol*, submitted
- Kirchhamer CV, Yuh CH, Davidson EH (1996) Modular cis-regulatory organization of developmentally expressed genes: Two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proc Natl Acad Sci USA* 93:9322–9328
- Lemaire SD, Quesada A, Merchan F, Corral JM, Igeno MI, Keryer E, Issakidis-Bourguet E, Hirasawa M, Knaff D, Miginiac-Maslow M (2005) NADP-malate dehydrogenase from unicellular green alga *Chlamydomonas reinhardtii*. A first step toward redox regulation? *Plant Physiol* 137:514–521
- Lottspeich F, Zorbas H (1998) Bioanalytik. Spektrum Akademischer Verlag, Heidelberg
- Luchetta P, Cretin C, Gadal P (1991) Organization and expression of the two homologous genes encoding the NADP-malate dehydrogenase in *Sorghum vulgare* leaves. *Mol Gen Genet* 228:473–481
- Miginiac-Maslow M, Lancelin JM (2002) Intracellular inhibition in redox signalling: light activation of NADP-malate dehydrogenase. *Photosynth Res* 72:1–12
- Mummenhoff K, Koch M (1994) Chloroplast DNA restriction site variation and phylogenetic relationships in the genus *Thlaspi sensu lato* (Brassicaceae). *Syst Bot* 19:73–88
- Murchie EH, Hubbart S, Peng S, Horton P (2005) Acclimation of photosynthesis to high irradiance in rice: Gene expression and interactions with leaf development. *J Exp Bot* 56:449–460
- Nemeth K, Salchert K, Putnoky P, Bhalerao R, Koncz-Kalman Z, Stankovic-Stangeland B, Bako L, Mathur J, Ökresz L, Stabel S, Geigenberger P, Stitt M, Redei GP, Schell J, Koncz C (1998) Pleiotropic control of glucose and hormone response by PRL1, a nuclear WD protein, in *Arabidopsis*. *Genes Dev* 12:3059–3073
- Ocheretina O, Haferkamp I, Tellioglu H, Scheibe R (2000) Light-modulated NADP-malate dehydrogenase from moss fern and green algae: Insights into evolution of the enzymes regulation. *Gene* 258:147–154
- Rippe K, Vonhippel PH, Langowski J (1995) Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem Sci* 20:500–506
- Rondeau P, Rouch C, Besnard G (2005) NADP-malate dehydrogenase gene evolution in *Andropogoneae* (Poaceae): Gene duplication followed by sub-functionalization. *Ann Bot* 96:1307–1314
- Rossberg M, Theres K, Acarkan A, Herrero R, Schmitt T, Schumacher K, Schmitz G, Schmidt R (2001) Comparative sequence analysis reveals extensivemicrocolinearity in the lateral suppressor regions of the tomato, *Arabidopsis* and *Capsella* genomes. *Plant Cell* 13:979–988
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Savitch LV, Massacci A, Gray GR, Huner NPA (2000) Acclimation to low temperature or high light mitigates sensitivity to photoinhibition: Roles of the Calvin cycle and the Mehler reaction. *Aust J Plant Physiol* 27:253–264
- Savitch LV, Barker-Astrom J, Ivanov AG, Hurry V, Öquist G, Huner NPA, Gardeström P (2001) Cold acclimation of *Arabidopsis thaliana* results in incomplete recovery of photosynthetic capacity, associated with an increased reduction of the chloroplast stroma. *Planta* 214:295–303
- Scheibe R (1990) Light-dark modulation: Regulation of chloroplast metabolism in a new light. *Bot Acta* 103:327–334
- Scheibe R (1991) Redox-modulation of chloroplast enzymes. *Plant Physiol* 96:1–3
- Scheibe R (2004) Malate valves to balance cellular energy supply. *Physiol Plant* 120:21–26
- Scheibe R, Jacquot JP (1983) NADP regulates the light activation of NADP-dependent malate dehydrogenase. *Planta* 157:548–553
- Scheibe R, Stitt M (1988) Comparison of NADP-malate dehydrogenase activation, Q<sub>A</sub> reduction and O<sub>2</sub> evolution in spinach leaves. *Plant Physiol Biochem* 26:473–481
- Scheibe R, Backhausen JE, Emmerlich V, Holtgreffe S (2005) Strategies to maintain redox homeostasis during photosynthesis under changing conditions. *J Exp Bot* 56:1481–1489

- Shahmuradov IA, Solovyev VV, Gammerrmann AJ (2005) Plant promoter prediction with confidence estimation. *Nucleic Acids Res* 33:1069–1076
- Timko MP, Kausch AP, Castresana C, Fassler J, Herrera-Estrella L, Van den Broeck G, Van Montagu M, Schell J, Cashmore AR (1985) Light regulation of plant gene expression by an upstream enhancer-like element. *Nature* 318:579–582
- Tyagi AK (2001) Plant genes and their expression. *Curr Sci* 80:161–169
- Von Schaewen A, Langenkämper G, Graeve K, Wenderoth I, Scheibe R (1995) Molecular characterization of the plastidic glucose-6-phosphate dehydrogenase from potato in comparison to its cytosolic counterpart. *Plant Physiol* 109:1327–1335
- Walters RG, Rogers JJM, Shephard F, Horton P (1999) Acclimation of *Arabidopsis thaliana* to the light environment: The role of photoreceptors. *Planta* 209:517–527
- Wang XT, Gaasterland T, Chua NH (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol* 6:R30.1–R30.11
- Werner T (1999) Models for prediction and recognition of eukaryotic promoters. *Mamm Genome* 10:168–175
- Wieczorek E, Brand M, Jacq X, Tora L (1998) Function of TAFII-containing complex without TBP in transcription by RNA polymerase II. *Nature* 393:187–191
- Yoneda Y (1997) How proteins are transported from cytoplasm to the nucleus. *J Biochem (Tokyo)* 121:811–817