Plant Molecular Biology **31:** 585–593, 1996.

Thi1, a thiamine biosynthetic gene in *Arabidopsis thaliana*, complements bacterial defects in DNA repair

C.R. Machado[1], R.L. Costa de Oliveira[1], S. Boiteux[2], U.M. Praekelt[1], P.A. Meacock[3] and C.F.M. Menck[1,4,*]

[1] *Depto. de Biologia, Instituto de Biociências, Universidade de São Paulo, CP 11461, São Paulo, 05422-970, SP, Brazil (*author for correspondence);* [2] *Institut Gustave Roussy, Villejuif, France;* [3] *Dept of Genetics, Univ. of Leicester, Univ. Road, Leicester LEI 7RH, UK;* [4] *Depto. de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo.*

Received 30 August 1995; accepted in revised form 17 March 1996

Key words: Arabidopsis, DNA repair, tolerance, DNA damage, thiamine biosynthesis

Abstract

An *Arabidopsis thaliana* cDNA was isolated by complementation of the *Escherichia coli* mutant strain BW535 (*xth, nfo, nth*), which is defective in DNA base excision repair pathways. This cDNA partially complements the methyl methane sulfonate (MMS) sensitive phenotype of BW535. It also partially corrects the UV-sensitive phenotype of *E. coli* AB1886 (*uvrA*) and restores its ability to reactivate UV-irradiated λ phage. It has an insert of ca. 1.3 kb with an open reading frame of 1047 bp (predicting a protein with a molecular mass of 36 kDa). This cDNA presents a high homology to a stress related gene from two species of *Fusarium* (*sti35*) and to genes whose products participate in the thiamine biosynthesis pathway, *THI4*, from *Saccharomyces cerevisiae* and *nmt2* from *Schizosaccharomyces pombe*. The *Arabidopsis* predicted polypeptide has homology to several protein motifs: amino-terminal chloroplast transit peptide, dinucleotide binding site, DNA binding and bacterial DNA polymerases. The auxotrophy for thiamine in the yeast *thi4::URA3* disruption strain is complemented by the *Arabidopsis* gene. Thus, the cloned gene, named *thi1*, is likely to function in the biosynthesis of thiamine in plants. The data presented in this work indicate that *thi1* may also be involved in DNA damage tolerance in plant cells.

Introduction

Evolution has produced several systems to protect the genetic material from damage by environmental agents or from by-products of normal cell metabolism. These systems, which are DNA repair and lesion tolerance mechanisms, exist in all living organisms and the recent characterization of several related genes has revealed an important degree of homology and analogy even in phylogenetically unrelated species [18, 19]. In plants, studies of these mechanisms are far behind the present knowledge of bacterial and mammalian systems [41], although these organisms are directly exposed to the environment, particularly ultraviolet (UV) light from the solar spectrum and soil contaminants, and also accumulate highly reactive oxygen species as by-products of normal metabolism.

There is physiological evidence for the existence of DNA excision repair mechanisms in plants in response to chemical agents, UV light and ionizing radiation [45, 18, 19]. An efficient system for photoreactivation of UV damage was found in *Arabidopsis thaliana* [30]. At least one UV-sensitive mutant of *Arabidopsis* was isolated, which is deficient in dark DNA repair of pyrimidine-pyrimidinone (6-4) lesions [5]. At the molecular level, several plant genes related to DNA repair have been isolated. Homologues of the yeast *RAD6* gene have been rescued from *A. thaliana* and wheat by their ability to conjugate ubiquitin to other cellular proteins [42, 43]. On the basis of the high conservation of DNA repair mechanisms, some genes were

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number U17589.

isolated via homologies at the nucleotide and amino acid levels. This is the case for the *Arabidopsis* topoisomerase I [22]; a chloroplast homologue of the *E. coli recA* gene from *Arabidopsis* and *Pisium sativum* [6]; and an apurinic/apyrimidinic (AP) endonuclease, that also has redox activity, and which displays homologies to animal and bacterial enzymes [2].

Several other plant genes were identified by complementation assays in bacteria deficient in DNA repair. They were isolated from *E. coli* mutant cells transformed with a plant cDNA expression library by virtue of their ability to restore, at least partially, the resistance of the mutants to treatment with DNA damaging agents. Pang and colleagues [31, 28, 29] have employed this technique to obtain clones which complement *E. coli* strains either *recA*$^-$, *uvr*$^-$, *phr*$^-$ (genes related to DNA repair) or *ruvC*$^-$; *recG*$^-$ (whose products act on DNA recombination processes). One of the isolated cDNAs partially complements recombination related (RecA$^-$) phenotypes. Another 5 cDNAs also afforded partial correction of the *E. coli* mutant phenotypes, but their molecular roles were not clearly identified. Interestingly, most of these cDNAs are apparently targeted to chloroplasts. More recently, Santerre and Britt [38], using a similar strategy, reported the cloning of an *Arabidopsis thaliana* cDNA that complements the MMS-sensitive phenotype of the *E. coli tagA*, *alkA*, a double mutant deficient in the two 3-methyladenine glycosylases. The product of this cDNA has a 3-methyladenine glycolylase activity and its predicted amino acid sequence is homologous to other eukaryotic enzymes with similar activity.

In this work, we intended to isolate DNA repair related genes from *Arabidopsis thaliana* by using complementation screening in bacteria defective in enzymes that participate in the initial steps of base excision repair. The *E. coli* strain employed was the triple mutant BW535 (*xth, nfo, nth*), defective for the DNA repair endonucleases: exonuclease III, endonuclease IV and endonuclease III. These enzymes participate in the first steps of base excision repair of DNA lesions normally mediated by oxidative and alkylation products [12]. The mutant strain presents a hypersensitive phenotype to agents that produce oxygen radicals, such as hydrogen peroxide, and to alkylating agents, such as MMS [10]. The screening was performed by transfecting a cDNA library from *Arabidopsis thaliana* into BW535 and selecting clones that had increased resistance to MMS. One clone was isolated and characterized. Its predicted polypeptide sequence has strong homology with that of a thiamine biosynthetic gene from *S. cerevisiae* and *S. pombe*, as well as with a stress-inducible gene from two *Fusarium* species. The function of the *Arabidopsis* gene was tested by complementation of a *S. cerevisiae* strain disrupted for the homologous gene. The plant cDNA restores the yeast mutant to thiamine prototrophy, and it is most likely a thiamine biosynthetic gene also in *Arabidopsis*. Therefore, this gene was named *thi1*. The data presented, however, suggest that *thi1* may also have a function in DNA repair in plants, and that it may therefore encode a bi-functional protein.

Materials and methods

Bacterial strains, bacteriophage and media

Most of the strains employed were DNA repair mutant derivatives of the *E. coli* K12 strain AB1157 (F$^-$ *thi*1 *his*4 Δ(*gpt-proA*) *argE3 thr*1 *leuB*6 *ara*14 *lacY*1 *galK*2 *xyl*5 *mtl*1 *tsc*33 *supE*44 *rpsL*31), which is wild type for DNA repair functions [3]. The isogenic mutant strains were BW535, *nfo::kan nth*1*::kan* Δ*(xth-pncA)*90 [10]; BW9109, Δ*(xth-pncA)*90 [47] and AB1886, *uvrA*6 [3]. BNN132 strain, employed to convert the λ YES cDNA library to a plasmid library, was a λKC (*kan, cre*) lysogen of JM107 [13]. Wild-type λ phage were employed for host reactivation essays. The cells were routinely grown in LB broth liquid medium or plates (containing 1.5% agar). When necessary for plasmid selection, ampicillin (Ap, 50 μg/ml) was added to the medium [36].

cDNA library in plasmid expression vector

The *Arabidopsis thaliana* cDNA expression library was contained in a λ expression vector system, λYES [13], kindly given by Dr R.W. Davis, Stanford University, CA. The phage bears the 7.8 kb yeast-*E. coli* shuttle plasmid pSE936 between two P1 phage *lox* sites, which allows conversion from phage λ to plasmid clones by using the λKC lysogen BNN132. In the expression library the cDNAs were cloned into *Xho*I restriction site flanked by two *Eco*RI sites. The bacteria were infected with the λYES library and plated onto LB plus ampicillin. Plasmid DNA was purified from the pooled colonies (a total of ca. 7.2×10^5).

Screening clones for MMS resistance

The cDNA plasmid expression library was transfected into the *E. coli* BW535 by the method of Hanahan [15]. Five independent transfections were performed (300 ng DNA per transfection) yielding a total of ca. 1.2×10^5 colonies. These were plated directly onto Ap plates containing MMS (5 mM) and incubated for 18 h at 37 °C. Fifty clones (ten per transfection, from a total of ca. 10^3) were chosen among the largest colonies (it was assumed that these showed better growth in selective conditions) and restreaked on selective medium (5 and 10 mM MMS). Only seventeen clones survived. From these, the six largest colonies were assayed for resistance to 1 mM MMS. Two of these isolates showed increased survival in comparison to the parental strain. Plasmid characterization by restriction mapping and 5′ DNA sequence indicated that both contained identical cDNAs. One of them, designated pYXN3, was chosen for further studies.

Survival measurement

MMS in agar plates
Mid log-phase cells (OD = 0.4 at 580 nm) were diluted and plated in LB-Ap plates supplemented with 1 mM of MMS. After incubation for 18 h at 37 °C, the colonies were counted to measure cell survival.

MMS in liquid medium and UV irradiation
The cells were grown to OD = 0.4 at 580 nm, pelleted and resuspended in the original volume of 10 mM $MgSO_4$. Cell suspensions were irradiated with UV light (254 nm) or treated with MMS, as indicated, spread on duplicate LB plates, and incubated overnight at 37 °C.

Phage reactivation
Suspensions of λ phage (in 10 mM $MgSO_4$) were UV-irradiated a different doses, diluted, infected into overnight grown bacterial cultures and plated on agar plates. After 18 h of incubation at 37 °C, the number of plaque-forming units (pfu) was scored [46].

Determination and analysis of DNA sequence

The nucleotide sequence was determined by using the dideoxynucleotide-mediated chain termination method [37]. Primers complementary to sequences flanking cDNA inserts in the pSE939 were firstly employed [13]. The insert from the pYXN3 was subcloned into (*Eco*RI site) pBluescript KS^- ($pBSKS^-$, Stratagene), producing the plasmid pBXN3, with the *Arabidopsis* cDNA in the correct orientation concerning the *lacZ′* promoter. The insert was progressively resected with exonuclease III, from both extremities, using the technique described by Henikoff [16]. These exonuclease III products were sequenced using T3 and T7 primers. The whole original insert was entirely sequenced on both strands. All ambiguities were resolved. The GenBank [1] database was searched for homologous genes and SBASE [32] was searched for homologous motifs.

Expression of the cloned cDNA in a yeast thi4 *mutant strain.*

The cloned cDNA was excised with *Xho*I from pYXN3 and subcloned into the *Sal*I site of the yeast expression vector pG1 in both orientations. A similar vector, pG3, without the *thi1* cDNA, was employed in control experiments [39]. The recombinant vector (pG1-thi1) was transfected into the *S. cerevisiae thi4::URA3* disruption strain [33]. The genotype of this strain is *MATα, Leu2-3, ade2-1, trp1-1, his3-11, ura3, thi4::URA3.* Minimal medium was prepared as described [48] with the omission of thiamine and tryptophane unless indicated otherwise. Transformants were recovered at 30 °C on medium lacking tryptophan, but with thiamine, for vector selection.

Hybridization analysis

DNA and RNA were isolated from whole *Arabidopsis thaliana* (ecotype Columbia) plants as described [14]. DNA was digested by restriction nucleases and analyzed by Southern blot following standard procedures [36]. Northern blotting was performed as described [44]. Hybridization was performed using the pBXN3 insert as probe, ^{32}P-labelled by the random primer DNA labelling kit (Gibco/BRL), according to the manufacturer's instructions.

Results

Screening the Arabidopsis *cDNA library*

We have employed the *Arabidopsis thaliana* cDNA library prepared by Davis and co-workers [13] in the λYES phage. This multifunctional phage library was

converted into a plasmid (pSE936) library by using the *cre-lox* site-specific recombination system. The pSE936 is an expression vector in *E. coli*, so this cDNA *Arabidopsis* library was transformed into the *E. coli* strain BW535, a triple mutant defective for the genes *xth, nfo, nth*, which encode exonuclease III, endonuclease IV and endonuclease III, respectively. The transformed bacteria were plated on LB agar with antibiotic and the selective agent MMS. The surviving clones were restreaked in MMS-containing medium; these two selection steps yielded 6 clones. These were individually retested for MMS resistance, which was only confirmed for two clones. Analysis of these two clones revealed that they contain plasmids that have inserts with similar sizes and identical 5′ and 3′ sequence. One of the plasmids was named pYXN3 and used in subsequent investigations. The pYXN3 was retransfected into BW535 to avoid possible revertant phenotype due to bacterial modifications and not really carried by the plasmid. The transformed bacteria were plated in medium without MMS selection and independent clones were tested. The data confirmed that the plasmid carries a heritable MMS-resistant factor.

Table 1. Complementation of the MMS-sensitive phenotype of *E. coli* mutants by the *Arabidopsis thi1* gene.

E. coli strain	Plasmid	Survival (%)*
AB1157	none	51.0
BW535	none	0.6 ± 0.4
BW535	pYXN3	19.0 ± 16
BW9109	pSE936	0.26 ± 0.28
BW9109	pYXN3	20.5 ± 13.2
BW9109	pBSKS$^-$	0.72 ± 0.23
BW9109	pBXN3	8.2 ± 1.8

*Treatment was performed in LB agar plates supplemented with 1 mM MMS during 18 h, at 37 °C. The numbers indicate the average of, at least, 3 different experiments.

Complementation of sensitive phenotypes

The degree of complementation of the *E. coli* MMS-sensitive phenotype by the plant gene was tested in several independent experiments (summarized in Table 1). The colony forming ability was evaluated in MMS containing media, and compared to wild-type and untransformed mutant cells. It can be seen that the presence of pYXN3, containing the *Arabidopsis* cDNA, increases at least 30-fold the MMS resistance of the triple mutant, although it was not able to restore completely to the wild-type phenotype. This plasmid (pYXN3) also conferred a partial MMS resistant phenotype on the *E. coli* strain BW9109, single mutant for the *xth* gene (exonuclease III), in comparison to bacteria carrying the plasmid without any insert (pSE936). In order to confirm these data, the cDNA carried by pYXN3 was subcloned into a commercial vector (pBSKS$^-$), resulting in a plasmid named pBXN3, in which the *Arabidopsis* cDNA is under the control of the *lacZ* promoter. The new plasmid also partially corrected the MMS-sensitive phenotype of the BW9109 mutant (Table 1) and it was able to restore MMS resistance of the BW535 mutant when treatment was performed in liquid medium (Fig. 1A). The results confirm that the cloned *Arabidopsis* cDNA helps *E. coli* mutant strains to repair and/or tolerate the action of MMS, a DNA damaging agent. The cDNA was also tested for survival complementation of the AB1886 mutant strain (*uvrA6*), which is UV-sensitive, as it is unable to accomplish nucleotide excision repair. Significant complementation was observed in the *uvrA*-deficient strain (Fig. 1B). The presence of pBXN3 did not alter the UV survival levels of wild-type (AB1157) bacterial cells (Fig. 1B). Higher UV doses (yielding ca. 10% survival) were also employed for AB1157 cells with similar results (not shown).

To determine whether the *Arabidopsis* gene product acts by repairing DNA lesions rather than preventing them, we analysed its ability in recovering lesions in exogenous DNA. Infection of a *uvrA*-deficient strain with UV-irradiated DNA phage yielded a lower number of pfu relative to infection of wild-type bacteria. The mutant strain bearing a plasmid with the *Arabidopsis* cDNA recovered the ability to reactivate the damaged phage almost to wild-type levels (Fig. 2). These data confirm that the product of this cDNA is involved in repair mechanisms of DNA damage.

DNA sequence analysis

The sequence of the cDNA was determined after progressive digestion with exonuclease III and S1 nuclease [16] of pBXN3. Both strands were sequenced and all possible ambiguities were eliminated. DNA sequence for this cDNA is available via GENBANK, access number U17589. Analysis of the sequence reveals one single open reading frame (ORF) of 1047 bp starting at an ATG 22 bases from the 5′ end of the cDNA.

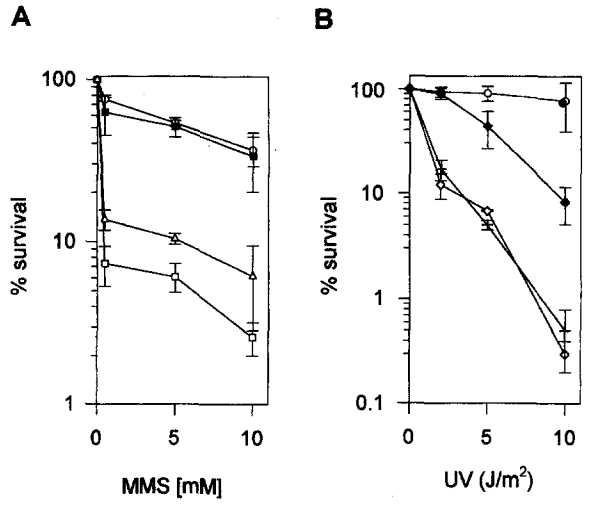

Figure 1. Survival after MMS and UV treatments of bacteria containing *thi1*. A. Cells were treated with MMS (10 minutes in liquid suspension, with the indicated doses): *E. coli* wild-type AB1157 (○) and BW535 (*xth, nfo, nth*) (□); BW535 with pBSKS⁻ (△); BW535 with pBXN3, containing the *thi1* cDNA (■). B. Cells were UV-irradiated: *E. coli* wild-type AB1157 (○); AB1157 with pBXN3 (●); AB1886 (*uvrA*) (◇); AB1886 with pBSKS⁻ (+); AB1886 with pBXN3 (*thi1*) (◆). Each value correspond to the average of at least 4 different experiments.

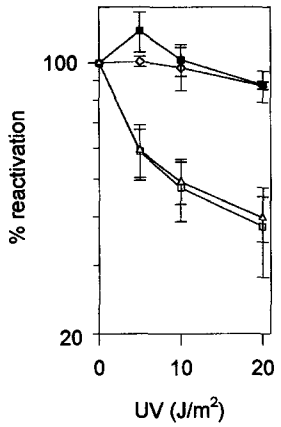

Figure 2. Reactivation of UV irradiated λ phage when infected into *E. coli* wild-type and *uvrA* mutant. λ phage was UV-irradiated before infection into AB1157 (wt) (◇), AB1886 (*uvrA*) (□), AB1886 with pBSKS⁻ (△) and AB1886 with pBXN3 (■). Each value correponds to the average of 8 different experiments.

A stop codon is found 178 bases 5′ from a small (9 bases) poly(A) tail. The bases bordering the first ATG codon of this ORF (AAAAAUGGC) are good match to the plant translational initiation consensus sequence [25]. The predicted protein is 349 amino acids in length (Fig. 3), with a molecular mass of ca. 36 kDa and positive charge. Screening the database for related sequences revealed a high homology between the plant gene and three fungal proteins. Two of these are stress-inducible proteins from two species of *Fusarium* (57.8% identity and 71.2% similarity, excluding the 55 amino acids at the N-terminus of *thi1*, which have no homology to the other gene products, see below) [7]. The other two are genes required for thiamine biosynthesis: the THI4 protein from *S. cerevisiae* (51.7% identity and 65.4% similarity) [33, 34], and the *nmt2* gene product from *S. pombe* (55.1% identity and 67.8% similarity) [26] both of which have been shown to be required for the biosynthesis of the thiazole precursor of thiamine. The alignment of the four polypeptides is shown in Fig. 3. On the basis of this strong homolo-

```
                                                I
P       MAAIASTLSLSSTKPQRLFDSSFHGSAISAAPISIGLKPRSFSVRATTAGYDLNAF
F       ............Ma-paAvsppsRsaeLatstklPVmskninTrtVremLgqwdd-
Sp      .......maPaTaVVTpqtafktdLpVekt-hnTVVksemgalsK-yptysldes-
Sc      ..................msaTstATst--sqLhLnst-vThCLsdivkke-wSd-

                                                    II          III
P   57 TFDPIKESIVSREMTRRYMTDMITYAETDVVVVGAGSAGLSAAYEISKN.PNVQVA
F   45 k-a--R--q---a-----fq-Ldn---S-I-II----c------iLg-kr-DLKI-
Sp  50 S-a--R--t---a-----fS-Ldk---S-I-I--------T---y-gtrr-DLKI-
Sc  39 k-a--R--t---a--s--fk-LdkF-vS--II-----S------v-A--r-DLK-C

                                                IVa
P  112 IIEQSVSPGGGAWLGGQLFSAMIVRKPAHLFLDEIGVAYDEQDTYVVVKHAALFTS
F  101 ---a-------------------M----da--r-V--p-EDEgn------------
Sp 107 ---a--A---------------V-----d---N----p-EDEgd------------
Sc  95 ---S--A----S----------VM--------q-LeIp-EDEgd----------i-

P  168 TIMSKLLARPNVKLFNAVAAEDLIV....K....GNRVGGVVTNWALVAQNHHTQS
F  157 -----V-qm--I-----tcv----trpseE....-v-Ia------t--SmH-dd--
Sp 163 -V-ARt--l--------t-v-------...egkdgqeg-Ia------t--Sl--gl--
Sc 151 -VL--V-ql--------tcv---VtrpptEk.g.evt-A------t--t-a-g---c

P  216 CMDPNVMEAK.............IVVSSCGHDGPFGATGVKRLKSIGMIDHVPG
F  209 -----tIn-p.............LII-Tt-----m--fc----v-M-R-EkLg-
Sp 216 -----tIn-h.............L---At--------fc----a-aqLVsNLhd
Sc 205 ------I-lagykndgtrdlsqkhgVIL-Tt--------fca--Ivd-dqnqkLg-

                                                IVb
P  257 MKALDMNTAEDAIV..RLTREVVPGMIVTGMEVAEIDGAPRMGPTFGAMMISGQKA
F  250 -Rg----l-------..Kg---I---L--g---LS-V---n---------aL--l--
Sp 257 -Rp----r---l--..Kg----f-----g---LS-f---n-----SVV.......
Sc 261 --g----h--hdV-ihsgayag-dn-yfa------L--ln---------aL--vh-

P  311 GQLALKALGLPNAIDGTLVGNLSPELVLAAADSAETVDA
F  304 aEe---ifdtrkkqndl.....................
Sp     .....................................
Sc 317 aEqi--hfaa.....................................
```

Figure 3. Amino acid sequence comparison of the predicted product of the *thi1* gene from *Arabidopsis* (P), *sti35* from *Fusarium oxysporum* (F), *thi4* from *S. cerevisiae* (Sc) and *nmt2* from *S. pombe* (Sp). The alignment was obtained by using the FASTA program. Trace marks (–) denote identical residues, aminoacids in upper caps denote functionally similar residues, points (.) denote deletions and lower caps denote no similarity, always in relation to the plant gene. The indicated regions correspond to amino acid sequences with significant homology with (I) chloroplast transite peptides, (II) dinucleotide-binding site consensus (III). DNA-binding motif consensus and (IVa and IVb) bacterial DNA polymerases.

gy to the yeast sequence, the *Arabidopsis* cDNA was designated *thi1*.

The N-terminal portion of the predicted polypeptide is rich in the hydroxylated amino acids serine and threonine and the hydrophobic amino acids alanine and valine. It has a net positive charge. These features are common to transit peptides targeted to chloroplasts [21], suggesting that the gene product may be located in the chloroplast. Moreover, comparison of the THI1 predicted protein to motifs in the data base indicates the presence of at least three other interesting regions, which match a dinucleotide-binding site [27], a DNA-binding motif [17] and two sites homologous to bacterial DNA polymerases [23].

Arabidopsis thi1 *complements a yeast strain disrupted for the homologous gene*

Because of the high homology of the plant cDNA to yeast genes involved in the synthesis of thiamine, the *thi1* cDNA was tested for complementation of yeast mutant strain. It was subcloned into a *S. cerevisiae* expression vector, under the control of the glyceraldehyde-3-phosphate dehydrogenase promoter, and then transformed into a *thi4::URA3* deletion strain. This strain is unable to synthesize the thiazole precursor of thiamine and is, therefore, thiamine auxotroph [33]. Transformants, selected for tryptophane prototrophy, were tested for growth in medium lacking thiamine. The results of this complementation test are shown in Fig. 4. Cells containing vectors with the plant cDNA inserted in the sense or antisense orientation grow well in thiamine-supplemented medium. In the absence of thiamine only cells bearing the *thi1* gene in the correct orientation for expression could grow. These results confirm that the cloned cDNA encodes a functional homologue of the yeast *THI4* gene.

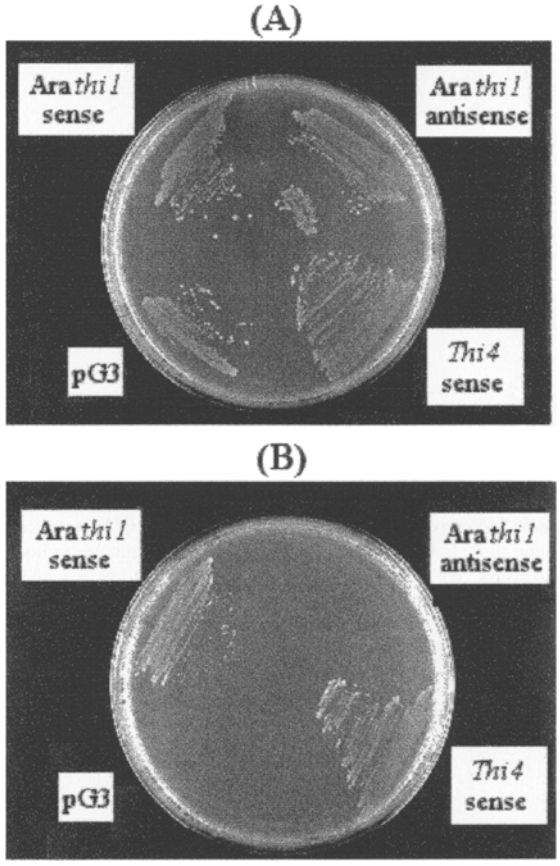

Figure 4. Functional complementation of an *S. cerevisiae thi4* disruption strain by the *Arabidopsis thi1* cDNA. The yeast wild-type and *thi4::URA3* disruption strain were transformed with yeast episomal expression vectors: pG3, pG1-thi1, containing the *thi1* gene (Ara *thi1*) in the sense or antisense orientation, and pG1-Thi4, containing the yeast gene. The resulting clones were streaked on agar plates supplemented with (A) or lacking thiamine (B). The plates were incubated at 30 °C for 3 days.

Arabidopsis genome and RNA analysis of the thi1 *gene.*

Figure 5 shows a Southern blot of the *Arabidopsis thaliana* genome digested with several restriction enzymes and hybridized with the entire ^{32}P-labelled *thi1* cDNA. Only one band is observed in lanes 1 and 2, for enzymes that do not cleave the cDNA. The *Bgl*II digest (lane 2) yields a small fragment, indicating that the whole coding sequence may be contained in a small region of the genome (at most 4.3 kb). However, the *Hpa*II digestion pattern (lane 3), which cleaves the cDNA twice with an internal fragment of 152 bp, show three different bands, the smaller corresponding to ca. 0.7 kb. This result reveals the presence of at least one intron. The band intensities, relative to self hybridized DNA (data not shown), indicate that the *thi1* gene is present in a single copy per *Arabidopsis* genome. This observation implies that *thi1* is a nuclear gene, since organellar genomes are normally observed in several hundred copies per cell [35].

RNA extracted from whole plants was also analysed by northern blot and the data are shown in Fig. 6. A single RNA band with an approximate size of 1.3 kb hybridizes with the *thi1* probe. It should be noted that overloaded lanes are provided to show that only one RNA message is observed in the plant cells.

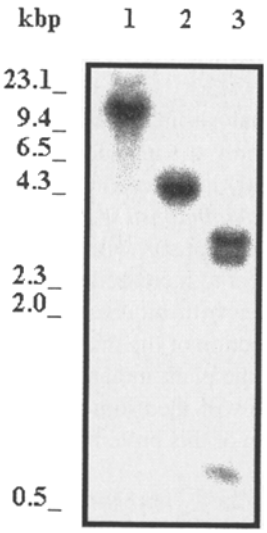

Figure 5. Genome analysis of the *thi1* gene. Southern blot using the *thi1* cDNA as a ^{32}P-labelled probe. The different lanes represent ca. 3 μg of *Arabidopsis thaliana* DNA digested with (1) *Bam*HI, (2) *Bgl*II and (3) *Hpa*II. Numbers on the left indicate DNA sizes (kb).

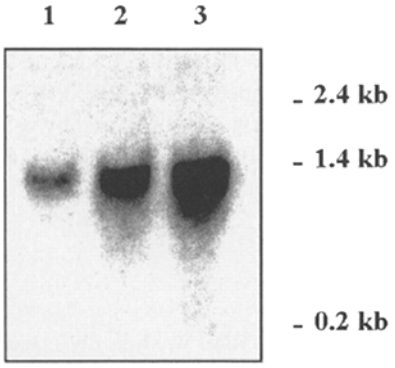

Figure 6. thi1 RNA expression in whole plants. Northern blot of *Arabidopsis thaliana* RNA: 1.5 μg; 2.15 μg; 3.30 μg; probed with ^{32}P-labelled *thi1* cDNA. Numbers on the right indicate RNA sizes (kb).

Discussion

We isolated one *Arabidopsis* cDNA which partially corrects the *E. coli* strain BW535 MMS-sensitive phenotype. This strain lacks three endonucleases (exonuclease III, endonuclease IV and endonuclease III) that play important roles in the excision of modified bases in DNA [11]. This *Arabidopsis* cDNA also confers MMS resistance to *E. coli* strain BW9109, which is defective only in the product of the *xth* gene (exonuclease III) and it is also able to enhance the survival after UV irradiation of the UV sensitive *uvrA E. coli* strain AB1886, defective in nucleotide excision repair. The presence of a plasmid expressing this cDNA also restores the ability of the *uvrA* bacterial mutant to reactivate UV damaged λ phage. Surprisingly, this cDNA was found to be homologue of yeast thiamine biosynthesic genes, *THI4* from *S. cerevisiae* and *nmt2* from *S. pombe*, that are involved in the formation of the thiazole precursor of thiamine. The ability of the *Arabidopsis* gene, *thi1*, to functionally complement the yeast *thi4* disruption strain suggests that it has a similar role in thiamine biosynthesis of plants.

The mechanism by which *Arabidopsis thi1* is able to enhance the survival of *E. coli* mutants is not clear. The data indicate that it is not a trivial effect of intracellular thiamine acting as scavenger of DNA damaging agents: the phage reactivation experiments clearly show that damaged DNA is recovered efficiently in mutants carrying the *thi1* gene. Also, wild-type strain survival was not affected by the expression of this gene implying that *thi1* is not simply inducing a general enhancement of bacterial resistance to these damaging agents. The *E. coli* strains employed in this work require thiamine for their growth, but this requirement is not complemented by expression of the *thi1* gene (data not shown). In addition, the whole selection strategy and survival experiments were performed in the rich LB broth media and the correction of the sensitive phenotypes was also observed in experiments carried out with excess thiamine (5 mM). These observations, together with the phage reactivation data, make very unlikely the possibility that any effect on direct thiamine synthesis *per se* by *thi1* would affect cell resistance. Therefore, dual independent functions for the *thi1* gene product, that is, thiamine biosynthesis and DNA damage tolerance, may occur within the cells.

The sequence analysis of the *thi1* cDNA reveals that it has no significant homology to the AP endonuclease recently isolated from *A. thaliana* that displays partial similarity to several known endonucleases, including the *E. coli* exonuclease III [2]. However, the different consensus motifs for DNA binding [17] and bacterial DNA polymerases [23] found in the *thi1* polypeptide sequence are consistent with a role of this protein in the metabolism of damaged DNA. Additionally, the homologue of *Arabidopsis thi1* in two species of the phytopathogenic fungus *Fusarium*, *sti35*, is inducible by various stresses including heat shock, ethanol and phytoalexins [7]. Its predicted polypeptide sequence suggests that it is likely to function also in thiamine biosynthesis, however, inducibility by various stresses, which is commonly observed for DNA repair genes,

would support a dual function of this group of proteins in thiamine biosynthesis and DNA repair. It will be interesting to determine whether *Arabidopsis thi1* is induced by environmental stress, particularly UV radiation and heat shock.

The biosynthesis of the thiazole precursor of thiamine has previously been reported to occur in the chloroplasts of plants [20] and the presence of an N-terminal sequence with characteristics of a chloroplast targeting signal predicts that the thi1 protein of *Arabidopsis* is translocated to the chloroplasts. In the case of yeast, the structure of the N-terminal coding region of THI4 predicts that this protein is located in the mitochondria, suggesting that complementation of the *thi4* disruption strain with the plant genes was dependent on targeting to this organelle. Although most chloroplast targeting signals are specific for chloroplasts, dual targeting to both chloroplasts and mitochondria was recently reported for pea glutathione reductase, an enzyme with important function in the protection of chloroplasts and mitochondria from oxidative damage and encoded by a single gene [9]. We are currently investigating the subcellular location of the yeast THI4 protein and of the *Arabidopsis thi1* gene product in the yeast *thi4* disruption strain.

Considering the role of the Arabidopsis *thi1* gene in DNA damage tolerance, it is interesting that many plant genes cloned by functional complementation of *E. coli* DNA repair mutants encode plastid proteins [31, 28, 29]. The preferential selection in bacteria of plant genes whose products function in the chloroplast may be accounted for by the likely prokaryotic origin of chloroplasts. However, it also points to very active DNA damage tolerance mechanisms present in plant chloroplasts, which have to cope with constant exposure to oxidizing agents produced by the photosynthetic electron transport chain.

Recently, two other plant genes homologous to *thi1* were cloned: from maize [4] and from *Alnus glutinosa* (Ribeiro and Pawlowski, personal communication). In maize, two different cDNA copies were cloned suggesting its duplication, and it was shown that one of them also complements the yeast *thi4* mutant. By immunoblot analysis, this gene product was found to be located in the plastid cell fraction confirming that it is targeted to cell organelles. Expression of the maize gene was highest in developing embryos and in green leaves. The *Alnus* gene was found to be expressed at high levels in nodules and shoot tips. Both patterns of expression are consistent with a role of these homologous genes in thiamine synthesis, but highly replicating tissues would also need protection against DNA lesions, particularly those related to oxidative metabolism.

Mutation analysis in *Arabidopsis* and yeast has so far defined only a single locus affecting thiazole biosynthesis [24], and experiments are underway to test whether the *Arabidopsis tz* locus and *thi1* are allelic. If so, these *tz* mutants will be very interesting to study with respect to increased sensitivity to UV irradiation. Together with biochemical investigations of the catalytic function of the thiazole biosynthetic protein, studies of the plant mutant and of the yeast *thi4* disruption strain will shed some light on the nature of the dual function of this protein.

Acknowledgements

We thank Dr Ronald W. Davis (Stanford University, CA) that kindly provided us with the *Arabidopsis* cDNA library. We thank also Dr M.A. Van Sluys (University of São Paulo) for providing plants (*Arabidopsis thaliana*) for hybridization analysis. R.C.O. and U.M.P. had fellowships from La Ligue Nationale contre le Cancer (France) and FAPESP (São Paulo), respectively. This work was supported by grants from FAPESP (São Paulo, Brazil), CNPq (Brasília, Brazil), and ARC (Villejuif, France).

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman D: Basic local alignment search tool. J Mol Biol 215: 403–410 (1990).
2. Babiychuk E, Kushnir S, Montagu MV, Inz D: The *Arabidopsis thaliana* apurinic endonuclease Arp reduces human transcription factors Fos and Jun. Proc Natl Acad Sci USA 91: 3299–3303 (1994).
3. Bachmann B: Pedigrees of some mutant strains of *Escherichia coli* K-12. Bact Rev 36: 525–557 (1972).
4. Belanger FC, Leustek T, Chu B, Kriz AL: Evidence for the thiamine biosynthetic pathway in higher-plant plastids and its developmental regulation. Plant Mol Biol 29: 809–821 (1995).
5. Britt AB, Chen J-J, Mitchell D: A UV-sensitive mutant of Arabidopsis defective in the repair of pyrimidine-pyrimidinone (6–4) dimers. Science 261: 1571–1574 (1993).
6. Cerutti H, Osman N, Grandoni P, Jagendorf AT: A homologue of *Escherichia coli* RecA proteins in plastids of higher plants. Proc Natl Acad Sci USA 89: 8068–8072 (1992).
7. Choi GH, Marek ET, Schardl LC, Richey MG, Chang S, Smith DA: *sti35*, a stress-responsive gene in *Fusarium* spp. J Bact 172: 4522–4528 (1990).
8. Cieminis KGK, Rancelien VM, Prijalgauskien AJ, Tiunaitien NV, Rudzianskait AM, Jancys ZJ: Chromosome and DNA damage and their repair in higher plants irradiated with short-wave ultraviolet light. Mutat Res 181: 9–16 (1987).

9. Creissen G, Reynolds H, Xue Y, Mullineaux P: Simultaneous targeting of pea glutathione reductase and a bacterial fusion protein to chloroplasts and mitochondria in transgenic tobacco. Plant J 8: 167–175 (1995).
10. Cunningham RP, Saporito SM, Spitzer SG, Weiss B: Endonuclease IV (nfo) mutant of *Escherichia coli*. J Bact 168: 1120–1127 (1986).
11. Demple B, Harrison L: Repair of oxidative damage to DNA: enzymology and biology. Annu Rev Biochem 63: 915-948 (1994).
12. Doetsch PW, Cunningham RP: The enzymology of apurinic/apyrimidinic endonucleases. Mutat Res 336: 173–201 (1990).
13. Elledge SJ, Mulligan JT, Ramler SW, Spottswood M, Davis RW: λYes: a multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations. Proc Natl Acad Sci USA 88: 1731–1735 (1991).
14. Fedoroff NV: Biochemical and molecular techniques in maize research. In: Setlow J, Hollaender A (eds) Genetic Engineering: Principles and Methods, pp. 115–134. Plenum Publishing Corporation, New York (1985).
15. Hanahan D: Studies on transformation of *E. coli* with plasmids. J Mol Biol 166: 557–580 (1983).
16. Henikoff S: Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359 (1984).
17. Henkin TM, Grundy FJ, Nicholson WL, Chambliss GH: Catabolite repression of alpha-amylase gene expression in *Bacillus subtilis* involves a trans-acting gene product homologous to the *Escherichia coli* lacI and galR repressors. Mol Microbiol 5: 575–584 (1991).
18. Hoeijmakers JHJ: Nucleotide excision repair I: from *E. coli* to yeast. Trends Genet 9 (5): 173–177 (1993).
19. Hoeijmakers JHJ: Nucleotide excision repair II: from yeast to mammals. Trends Genet 9 (6): 211–217 (1993).
20. Julliard J-H, Douce R: Biosynthesis of the thiazole moiety of thiamin (vitamin B1) in higher plant chloroplasts. Proc Natl Acad Sci USA 88: 2042–2045 (1991).
21. Keegstra K, Olsen LJ, Theg SM: Chloroplastic precursors and their transport across the envelope membranes. Annu Rev Plant Physiol Plant Mol Biol 40: 471–501 (1989).
22. Kieber JJ, Tissier AF, Signer ER: Cloning and characterization of an *Arabidopsis thaliana* Topoisomerase I gene. Plant Physiol 99: 1493–1501 (1992).
23. Lawyer FC, Stoffel S, Saiki RK, Myambo K, Drummond R, Gelfand DH: Isolation, characterization, and expression in *Escherichia coli* of the DNA Polymerase gene from *Thermus aquaticus*. J Biol Chem 264: 6427–6437 (1989).
24. Li SL, Redei GP: Thiamine mutants of the crucifer *Arabidopsis*. Biochem Genet 3: 163–170 (1969).
25. Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA: Selection of AUG inititation codons differs in plants and animals. EMBO J 6: 43–48 (1987).
26. Manetti AGO, Rosetto M, Maundrell, KG: *nmt2* of fission yeast: a second thiamine-repressible gene co-ordinately regulated with *nmt1*. Yeast 10: 1075–1082 (1994).
27. Moeller W, Amons R: Phosphate-binding sequences in nucleotide-binding proteins. FEBS Lett 186: 1–7 (1985).
28. Pang Q, Hays JB, Rajagopal I, Schaefer TS: Selection of *Arabidopsis* cDNAs that partially correct phenotypes of *Escherichia coli* DNA-damage-sensitive mutants and analysis of two plant cDNAs that appear to express UV-specific dark repair activities. Plant Mol Biol 22: 411–426 (1993).
29. Pang Q, Hays JB, Rajagopal I: Two cDNAs from the plant *Arabidopsis thaliana* that partially restore recombination proficiency and DNA-damage resistance to *E. coli* mutants lacking recombination-intermediate-resolution activities. Nucl Acids Res 21: 1647–1653 (1993).
30. Pang Q, Hays JB: UV-B-inducible and temperature sensitive photoreactivation of cyclobutane pyrimidine dimers in *Arabidopsis thaliana*. Plant Physiol 95: 536–543 (1991).
31. Pang Q, Hays JB Rajagopal I: A plant cDNA that partially complements *Escherichia coli recA* mutations predicts a polypeptide not strongly homologous to RecA proteins. Proc Natl Acad Sci USA 89: 8073–8077 (1992).
32. Pongor S, Skerl V, Cserzo M, Hatsagi Z, Simon G, Bevilacqua V: The SBASE domain library release 2.0 & A collection of annotated protein sequence segment. Nucl Acids Res 21: 3111–3115 (1992).
33. Praekelt UM, Byrne KL, Meacock PA: Regulation of THI4 (MOL1), a thiamine-biosynthetic gene of *Saccharomyces cerevisiae*. Yeast 10: 481–490 (1994).
34. Praekelt UM, Meacock PA: MOL1, a *Saccharomyces cerevisiae* gene that is highly expressed in early stationary phase during growth on molasses. Yeast 8: 699–710 (1992).
35. Pruitt RE, Meyerowitz EM: Characterization of the genome of *Arabidopsis thaliana*. J Mol Biol 187: 169–183 (1986).
36. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, NY (1989).
37. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
38. Santerre A, Britt AB: Cloning of a 3-methyladenine-DNA glycosylase from *Arabidopsis thaliana*. Proc Natl Acad Sci USA 91: 2240–2244 (1994).
39. Schena M, Picard D, Yamamoto KR: Vectors for constitutive and inducible gene expression in yeast. Meth Enzymol 194: 389–398 (1991).
40. Soyfer VN: Influence of physiological conditions on DNA repair and mutagenesis in higher plants. Physiol Plant 58: 373–380 (1983).
41. Stapleton AE: Ultraviolet radiation and plants: burning questions. Plant Cell 4: 1353–1358 (1992).
42. Sullivan ML, Vierstra RD: An ubiquitin carrier protein from wheat germ is structurally and functionally similar to the yeast DNA repair enzyme encoded by RAD6. Proc Natl Acad Sci USA 86: 9861–9865 (1989).
43. Sullivan ML, Vierstra RD: Cloning af a 16 KDa ubiquitin carrier protein from wheat and *Arabidopsis thaliana*. J Biol Chem 266: 23878–23885 (1991).
44. Thomas PS: Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc Natl Acad Sci USA 77: 5201–5205 (1980).
45. Veleminsky J, Gichner T: DNA repair in mutagen-injured higher plants. Mutat Res 55: 71–84 (1978).
46. Weigle JJ: Induction of mutations in a bacterial virus. Proc Natl Acad Sci USA 39: 628 (1953).
47. White BJ, Hochauser SJ, Cintron NM, Weiss B: Genetic mapping of *xthA*, the structural gene for exonuclease III in *Escherichia coli* K-12. J Bact 126: 1082–1088 (1976).
48. Wickerham LJ: Taxonomy of yeast. US Dept Agric Tech Bull 1029: 11–56 (1951).