Isolation, characterization, and chromosomal location of a gene encoding the Δ^1-pyrroline-5-carboxylate synthetase in *Arabidopsis thaliana*

Arnould Savouré, Samir Jaoua, Xue-Jun Hua, Wilson Ardiles, Marc Van Montagu*, Nathalie Verbruggen

Laboratorium voor Genetica, Universiteit Gent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

Received 31 July 1995

Abstract **A full-length cDNA and the corresponding *At-P5S* gene encoding the first enzyme of the proline biosynthetic pathway, the Δ^1-pyrroline-5-carboxylate (P5C) synthetase, were isolated in *Arabidopsis thaliana*. The *At-P5S* cDNA encodes a protein of 717 amino acids showing high identity with the P5C synthetase of *Vigna aconitifolia*. Strong homology is also found at the N-terminus to bacterial and yeast γ-glutamyl kinase and at the C-terminus to bacterial γ-glutamyl phosphate reductase. Putative ATP- and NAD(P)H-binding sites are suggested in the At-P5S protein. The transcribed region of the *At-P5S* gene is 4.8 kb long and contains 20 exons. Southern analysis suggests the presence of only one *At-P5S* gene in the *A. thaliana* genome mapped at the bottom of the chromosome two. Expression analysis of *At-P5S* in different organs reveals abundant *At-P5S* transcripts in mature flowering plant. Rapid induction of the *At-P5S* gene followed by accumulation of proline was observed in NaCl-treated seedlings suggesting that *At-P5S* is osmoregulated.**

Key words: *Arabidopsis thaliana*; Pyrroline-5-carboxylate synthetase; Proline; Salt stress

1. Introduction

Hyperosmotic stresses caused by drought and salinity are the most important factors limiting plant growth and crop productivity [1]. Proline accumulation in response to various environmental stresses has been described in many phylogenetically diverse organisms such as bacteria [2], marine invertebrates [3], algae [4] and plants [5]. The higher halotolerance of the proline-overproducing *proB74 Escherichia coli* mutant [6] and the increase in the regeneration of drought-stressed calli to intact plants by exogenous supply of proline [7] suggest that proline plays an adaptive role in osmotic stress [5].

In eukaryotes and specifically in higher plants, proline accumulation is mainly due to a de novo synthesis from glutamate [5]. Despite the importance of this pathway, little is known about the mechanisms governing its regulation.

Proline is synthesized from glutamate via two successive reductions. In *E. coli*, γ-glutamyl kinase first phosphorylates glutamate into γ-glutamyl phosphate which is reduced into glutamate semialdehyde by γ-glutamyl phosphate reductase. Glutamate semialdehyde cyclises spontaneously into pyrroline-5-carboxylate (P5C). P5C is then reduced to proline by P5C reductase. In plants, the proline biosynthetic pathway from glutamate occurs as in microorganisms. However, reduction of glutamate to its semialdehyde was demonstrated to be performed by one bifunctional enzyme, P5C synthetase, since one cDNA has been cloned in *Vigna aconitifolia* that could complement an *E. coli* mutant lacking the two first enzymes of the proline biosynthetic pathway [8]. Recently, a cDNA showing high identity to the *P5CS* cDNA from *V. aconitifolia* was isolated from *Arabidopsis thaliana* [9].

As part of our efforts toward understanding the proline biosynthetic pathway in *Arabidopsis* and to unravel the role of proline in response to osmotic stress, we have previously cloned and characterized a cDNA and the corresponding gene encoding the second enzyme of the pathway, the P5C reductase [10].

In this paper, we report on the isolation, the structure, and the chromosomal location of the *At-P5S* gene in *A. thaliana* as well as on features of the encoded polypeptide. We also present expression analysis of *At-P5S* in different plant organs and during NaCl treatment.

2. Materials and methods

2.1. Plant material and growth conditions

Arabidopsis thaliana (L.) Heynh., ecotypes Columbia and Landsberg erecta (provided by M. Anderson, Nottingham *Arabidopsis* Stock Centre, UK), were grown in soil under 16 h light conditions at 22°C. Seedlings were grown for 10 days on modified Murashige and Skoog medium (Kl medium) [11] and treated either with Kl medium for control or with 170 mM NaCl as described previously [10]. Axenic cultures of *A. thaliana* were set up in liquid medium according to [12] and the roots collected after 3 weeks of growth.

2.2. Cloning of a P5C synthetase cDNA

To identify *A. thaliana* P5C synthetase cDNA(s), degenerate oligonucleotide primers in conserved sequences between *P5CS* cDNA from *V. aconitifolia* [8], *proB* and *proA* sequences from *E. coli* [13] and *Serratia marcescens* [14] (primer 1 = 5′-ATWGCYGATGCCCTKGAAGC-AAA-3′; primer 2 = 5′-TCWSWGAATCTKGTGCTKGCRTT-3′) were used in polymerase chain reactions (PCR) to amplify a DNA fragment from genomic DNA of *A. thaliana* ecotype Columbia. A 2-kb PCR fragment showing sequence similarity to the *V. aconitifolia P5CS* cDNA was cloned into pBluescript (Stratagene) and used as a probe to screen a λZAP-cDNA library from drought-stressed roots [15]. Of 150,000 colonies screened, 10 positive clones were isolated and the clones containing the longest insert of 2.6 kb was converted to pBluescript SK(−) phagemid clone by in vivo excision according to the Statagene protocol. This cDNA was named *At-P5S* according to the *Arabidopsis* nomenclature [16].

*Corresponding author. Fax: (32) (9) 264-5349.
Bitnet: mamon@gengenp.rug. ac.be

The nucleotide sequences reported in this paper have been deposited in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number X87330 for the cDNA and X89914 for the corresponding gene.

Abbreviations: P5C, pyrroline-5-carboxylate; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate.

SSDI 0014-5793(95)00935-3

2.3. Isolation of the At-P5S *gene*

An *A. thaliana* ecotype Columbia genomic DNA library constructed in the λGEM11 (obtained from J.T. Mulligan and R.W. Davis via the EEC-BRIDGE *Arabidopsis* DNA Stock Center, Köln, Germany) was screened by plaque hybridization with the *At-P5S* cDNA probe as described [17]. Of 300,000 plaques, 15 positive clones were isolated and one further analyzed. Two *Xho*I fragments of 3.4 and 3.8 kb in length were isolated and subcloned in pBluescript KS (Stratagene).

2.4. DNA and protein sequence analysis

Double-stranded plasmid DNA was sequenced on both strands by the dideoxy chain termination method on an automated DNA sequencer using dye primers (Applied Biosystem). The sequences of the cDNA and genomic *At-P5S* were established by using directed nucleotides. Analysis of nucleotide and amino acid sequences were carried out using programs in the GCG package of the University of Wisconsin through a UNIX system. Exon/intron boundaries within the *At-P5S* gene were identified by sequence comparison with the *At-P5S* cDNA. Sequence comparison with the databases were performed using BLAST through the NCBI E-mail server [18]. Protein secondary structure analysis was determined by using the PredictProtein program (EMBL Heidelberg, Germany) [19,20] with multiple sequence alignments of the seven proteins showing highest similarity, P5CS from *V. aconitifolia* [8], γ-glutamyl kinase and γ-glutamyl phosphate reductase from *E. coli* [13], *S. marcescens* [14] and *Bacillus subtilis* [21]. Inclusion of protein homologs increases the prediction accuracy of the secondary structure to 70% [19].

2.5. Southern blot analysis and chromosomal location

Genomic DNA was isolated from young leaves as described [12]. For Southern blot analysis, genomic DNA (1 μg) was digested, separated on 0.8% agarose gel in 1 × TBE buffer and transferred onto Hybond-N membrane (Amersham). After UV cross-linking, the membrane was hybridized with the ^{32}P-labelled full-length *At-P5S* cDNA clone at 65°C according to [22]. Membranes were washed at the final stringency of 0.2 × SSC, 0.1% SDS at 65°C for 20 min according to the manufacturer's protocol.

Polymorphism generated by *Eco*RI restriction was used for mapping of *At-P5S* in the *A. thaliana* genome. Data from 61 recombinant inbred (RI) lines provided by C. Lister and C. Dean (John Innes Institute, Norwich, UK; [23]) were scored and the map distances were calculated by C. Lister (Norwich, UK). Mapping data are available through AAtDB.

2.6. At-P5S *expression*

Total RNA was isolated from 10-day-old seedlings as well as from roots grown in liquid medium, from leaves, flowers, and siliques of mature plants according to [24]. Total RNA (10 or 20 μg) was electrophoresed on a 1.2% agarose-formaldehyde gel, transferred, and cross-linked to nylon membranes. The filters were subsequently treated as described for Southern blotting.

2.7. Proline determination

Free proline content was measured according to the method of Bates [25] using L-proline as standard.

3. Results

3.1. Isolation and characterization of A. thaliana At-P5S *cDNA*

A 2.6-kb full-length cDNA was isolated by screening a cDNA library prepared from roots of drought-stressed seedlings [15] with a 2-kb PCR fragment. The deduced amino acid sequence of the open reading frame encodes a putative protein of 717 amino acids with a molecular mass of 77.7 kDa and a calculated isoelectric point of 6.2. Analysis of hydropathy and sequence revealed no signal peptide or transmembrane domain.

We confirmed the identity of *At-P5S* cDNA on the basis of its deduced amino acid sequence showing high identity with the P5C synthetase from *V. aconitifolia* (74%) [8]. A cDNA encoding the P5C synthetase from *A. thaliana* ecotype Columbia has recently been reported [9]. It differs from the *At-P5S* cDNA described here in that the two phenylalanine at positions 42 and 467 are replaced respectively by cysteine and leucine. Protein sequence comparison using protein databases revealed similarity with the N-terminus of γ-glutamyl kinase from *E. coli* (35%), *S. marcescens* (33%), *Thermus aquaticus* (29%), and yeast (26%) and with the C-terminus of γ-glutamyl phosphate reductase from *E. coli* (39%), *S. marcescens* (39%) and *T. aquaticus* (35%).

Alignment of At-P5S with the P5C synthetase from *V. aconitifolia*, with γ-glutamyl kinase and γ-glutamyl phosphate reductase from *E. coli* and *S. marcescens* is shown in Fig. 1. Interestingly, the kinase and the reductase bacterial proteins overlapped at their C- and N-terminal ends, respectively (42% similarity; [8]). The predicted secondary structure of At-P5S was determined with the PHD Email server using multiple sequence alignments with the seven proteins showing highest similarity (Fig. 1). The predicted secondary structure composition of At-P5S is a mixed class of α-helix, β-sheet, and loop structures. At the N-terminus at position 61 just after the first β-sheet, a conserved sequence GXXXXGR between bacteria and plant proteins was observed. This motif may contribute to the phosphate-binding loop structure although it does not correspond exactly to the conserved GXXXXGKT sequence fingerprint [26]. In the C-terminal domain corresponding to the γ-glutamyl phosphate reductase domain, two β-sheets separated by an α-helix are predicted where glycine residues may allow a tight turn providing a favorable interaction for NAD(P)H [27]. The presence of two putative leucine zipper motifs in the N- and C-terminal domains was postulated [8]. However, these motifs do not perfectly match the general consensus of four contiguous heptad repeats and are not, as expected, in a continuous α-helix structure.

An aspartate residue is present at position 107 in the γ-glutamyl kinase bacterial proteins. This aspartate residue was shown to be implicated in the feedback inhibition of the γ-glutamyl kinase by the end product proline [28]. This aspartate residue is also present in *V. aconitifolia* at position 128 but not in *A. thaliana*. However, a conserved aspartate residue is present in both the *V. aconitifolia* and *A. thaliana* proteins at positions 126 and 125, respectively.

3.2. Genomic structure and organization of At-P5S *gene in* A. thaliana

A genomic library was screened with *At-P5S* cDNA clone and one of 15 positive phages was purified. The genomic clone yielded four DNA fragments when the phage DNA was digested by *Xho*I. The 3.8- and the 3.4-kb *Xho*I fragments hybridized to the full-length *At-P5S* cDNA (data not shown). The *At-P5S* gene is spanning 4.8 kb in length and is organized into 20 exons that are interrupted by 19 introns (Fig. 2). All the intron/exon junctions of *At-P5S* gene contained GT/AG splice donor and acceptor dinucleotides and conform well with the consensus splice sequences [29]. The sizes of introns ranged from 72 to 288 nucleotides. Two putative polyadenylation signal sequences AAAATA and ATACTA were identified starting at positions 189 and 308 downstream from the termination codon. The coding sequence of the *At-P5S* gene corresponds exactly to the sequence of the isolated cDNA.

3.3. At-P5S *is a single-copy gene*

Southern blots with genomic DNA isolated from Columbia and Landsberg ecotypes were carried out to investigate the

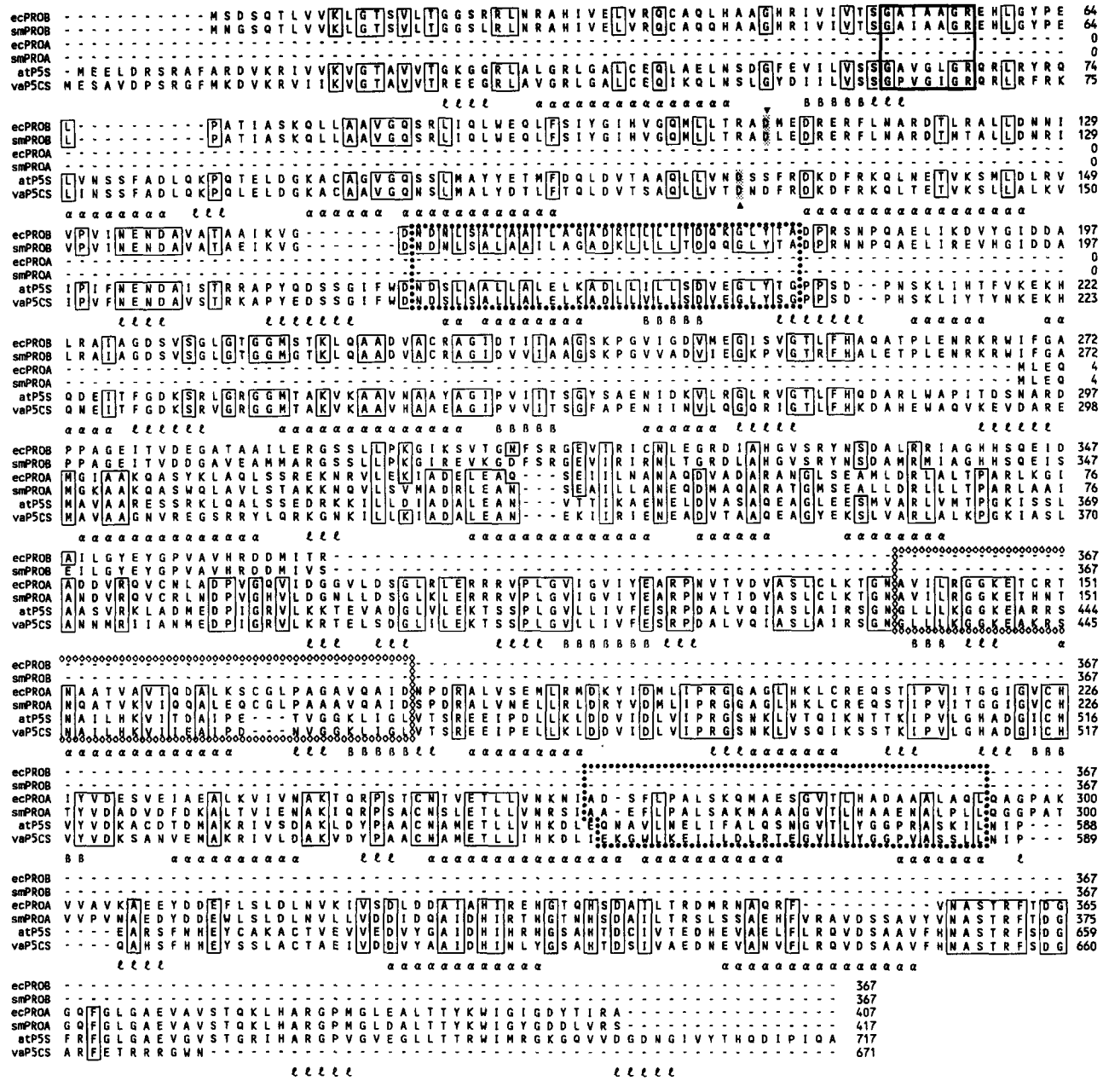
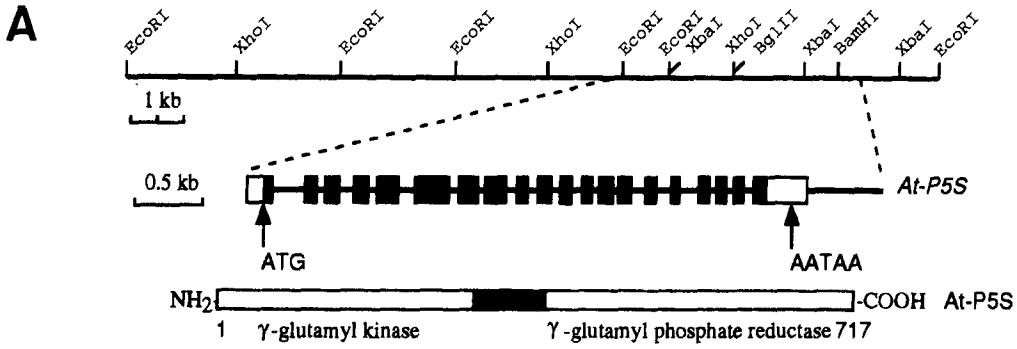


Fig. 1. Sequence alignment of the predicted At-P5S amino acid sequence of *Arabidopsis thaliana* with plant and prokaryotic protein sequences. The comparison includes sequences from *Vigna aconitifolia* (vaP5CS), γ-glutamyl-kinase (ecPROB) and γ-glutamyl-phosphate reductase (ecPROA) from *Escherichia coli*, γ-glutamyl-kinase (smPROB) and γ-glutamyl-phosphate reductase (smPROA) from *Serratia marcescens*. Identical amino acids are boxed. Secondary structure of At-P5S was calculated with the program PredictProtein (Heidelberg, Germany) and the result is shown in the seventh line: α-helices are indicated by α, β-strands by β, and loops by ℓ. Boxes indicate a putative ATP-binding site with thick lines, putative leucine domains with filled circles, and a putative NAD(P)H-binding domain with open diamonds. Triangles point out shaded aspartate residue implicated in proline feedback inhibition (▼) and their putative homologues in plants (▲). Gaps indicated by hyphens were introduced to maximize the sequence alignment.

copy number of *At-P5S*. Two strongly hybridizing bands were observed with *Bgl*II; *Hin*dIII, and *Xba*I although *Eco*RI generated one intense and one weak band (Fig. 3). This pattern compared with the restriction map of the genomic sequence suggests the presence of only one *At-P5S* gene in *A. thaliana* genome. Polymorphism was detected in the DNA of the two parental lines Columbia and Landsberg restricted with *Eco*RI and *Bgl*II (Fig. 3). Sixty-one recombinant inbred lines were used to map *At-P5S* using *Eco*RI polymorphism. *At-P5S* mapped at the bottom of the chromosome two on two RFLP markers, the g4514 cosmid and the m429 phage clones. This result confirms the single-copy nature of *At-P5S* in *A. thaliana* genome.

3.4. *Expression of* At-P5S *in plant tissues and in NaCl-treated seedlings*

To study expression of *At-P5S* in planta, roots, leaves, flowers, and siliques were analyzed for transcript level by Northern



B

```
tttattttcttaccttaaatacgacggtgcttcactgagtccgactcagttaactcgttcctctctctgtgtgtggttttggtagacgacgacgacgataATGGAGGAGCTAGATCGTTCACGTGCTTTT 130
                                                                                                    M  E  E  L  D  R  S  R  A  F  10
GCCAGAGACGTCAAACGTATCGTCGTTAAGgttcgttgagatacgttcgcatttttcagattttgttgttgatgattagattcttaatttgtgataatgtggaaatgaatattatgtaatttaagtgcatc 260
A  R  D  V  K  R  I  V  V  K                                                                       20
taaactctttgtttattgaattcgtgaatctgaatatatttttctaatcccagaaactaaaacttctcgtatgaatcttaatttgcatgtcattagagacgaatgaataatcagaatattcgagggatttt 390
ttttctgtttggtgattaaaatttggatttttgtttatattatgtaaaaaaaaaaagGTTGGGACAGCAGTTGTTACTGGAAAAGGTGGAAGATTGGCTCTTGGTCGTTTAGGAGCACTGTGTGAACAG 520
                                                                V  G  T  A  V  V  T  G  K  G  G  R  L  A  L  G  R  L  G  A  L  C  E  Q  44
gtaattgtcaaattttaataatctccttttttgtattgtgtttataaaaaagtgtaaaggtttcatttttttttcacgaaagacatgtgaaattattcatgcgtagtggcaacttttaatttgtaaaaaaata 650
tatatatataatgtcagCTTGCGGAATTAAACTCGGATGGATTTGAGGTGATATTGGTGTCATCTGGTGCGGTTGGTCTTGGCAGGCAAAGGCTTCGTTATCGACAATTAGTCAATAGCAGgttaaagct 780
                  L  A  E  L  N  S  D  G  F  E  V  I  L  V  S  S  G  A  V  G  L  G  R  Q  R  L  R  Y  R  Q  L  V  N  S  S  79
taatggctacacttcattattaatccctttcccttataacaacatttggaaacaaaaaaaaagggtgatgatggatggaccattttggcttatgttttattgctcaataacagtgacatgtgtttatgt 910
gtgttatgatttaaaagttttgttttttttgtgatggatttgttttttttctttttttttgttaatggcttttgcagCTTTGCGGATCTTCAGAAGCCTCAGACTGAACTTGATGGGAAGGCTTGTGC 1040
                                                                F  A  D  L  Q  K  P  Q  T  E  L  D  G  K  A  C  A  96
TGGTGTTGGACAAAGCAGTCTTATGGCTTACTATGAGACTATGTTTGACCAGgtgatttttcctttgttatcgaatttctagattattgtgtaagacatccaaatattgatgctgttgttttttctttggtt 1170
 G  V  G  Q  S  S  L  M  A  Y  Y  E  T  M  F  D  Q                                               113
agcTTGATGTGACGGCAGCTCAACTTCTGGTGAATGACAGTAGTTTTAGAGACAAGGATTTCAGGAAGCAACTTAATGAAACTGTCAAGTCTATGCTTGATTTGAGGGTTATTCCAATTTTCAATGAGAA 1300
  L  D  V  T  A  A  Q  L  L  V  N  D  S  S  F  R  D  K  D  F  R  K  Q  L  N  E  T  V  K  S  M  L  D  L  R  V  I  P  I  F  N  E  N  156
TGATGCTATTAGCACCCGAAGAGCCCCATATCAGgtttgtccctttgacatgaacttttctacacactctgagatgtgagggattctttgaatctcgtagtctaatgttcagcttcactggatcttgat 1430
 D  A  I  S  T  R  R  A  P  Y  Q                                                                  167
atatgcagGATTCTTCTGGTATTTTCTGGGATAACGATAGCTTAGCTGCTCTACTGGCGTTGGAACTGAAAGCTGATCTTCTGATTCTTCTGAGCGATGTTGAAGGTCTTTACACAGGCCCTCCAAGTGA 1560
       D  S  S  G  I  F  W  D  N  D  S  L  A  A  L  L  A  L  E  L  K  A  D  L  L  I  L  L  S  D  V  E  G  L  Y  T  G  P  P  S  D  208
TCCTAACTCAAAGTTGATCCACACTTTTGTTAAAGAAAAACATCAAGATGAGATTACATTCGGCGACAAATCAAGATTAGGGAGAGGGGGTATGACTGCAAAAGTCAAAGCTGCAGTCAATGCAGCTTAT 1690
 P  N  S  K  L  I  H  T  F  V  K  E  K  H  Q  D  E  I  T  F  G  D  K  S  R  L  G  R  G  G  M  T  A  K  V  K  A  A  V  N  A  A  Y  251
GCTGGGATTCCTGTCATCATAACCAGgtgaggaacctttctaagctcaccatgcataatgataggggtgatatgcttgttcaaatttggttagatggtatattgatatctttcttgcttctgaagTGGGTAT 1820
A  G  I  P  V  I  I  T  S                                                                    G  Y  262
TCAGCTGAGAACATAGATAAAGTCCTCAGAGGACTACGTGTTGGAACCTTGTTTCATCAAGATGCTCGTTTATGGGCTCCGATCACAGATTCTAATGCTCGTGACATGGCAGTTGCTGCGAGGGAAAGTT 1950
S  A  E  N  I  D  K  V  L  R  G  L  R  V  G  T  L  F  H  Q  D  A  R  L  W  A  P  I  T  D  S  N  A  R  D  M  A  V  A  A  R  E  S  S  306
CCAGAAAGCTTCAGgtaattgtgacttatgcatggctttctttcatgttcgtaacgtcaaaaaccattcttgctcggcatagagttacttaacttttttttacatttgctatagGCCTTATCTTCGGAA 2080
  R  K  L  Q                                                                                    A  L  S  S  E  315
GACAGGAAAAAAAATTCTGCTTGATATTGCCGATGCCCTTGAAGCAAATGTTACTACAATCAAAGCTGAGAATGAGTTAGATGTAGCTTCTGCACAAGAGGCTGGGTTGGAAGAGTCAATGGTGGCTCGCT 2210
D  R  K  K  I  L  L  D  I  A  D  A  L  E  A  N  V  T  T  I  K  A  E  N  E  L  D  V  A  S  A  Q  E  A  G  L  E  E  S  M  V  A  R  L  359
TAGTTATGACACCTGGAAAGgtaagaaagtattcatggccatagatagttgcttttttgttgctatggcttgggcaaacatattgtgccaatgtaacctctcctattatgtttcttatttttgtgcttgat 2340
 V  M  T  P  G  K                                                                               365
agATCTCGAGCCTTGCAGCTTCAGTTCGTAAGCTAGCTGATATGGAAGATCCAATCGGCCGTGTTTTAAAGAAAACAGAGgtgatcagaggacaattgttaccatatagttaatttacatactcttgagt 2470
  I  S  S  L  A  A  S  V  R  K  L  A  D  M  E  D  P  I  G  R  V  L  K  K  T  E                   390
taaataagggatatgactatcctcctagttgacatacaatagttgtttatgctatttgttctttgtggcaattccttttacagGTGGCAGATGGTCTTGTCTTAGAGAAGACCTCATCACCATTAGGCGT 2600
                                                                V  A  D  G  L  V  L  E  K  T  S  S  P  L  G  V  406
ACTTCTGATTGTTTTTGAATCCCGACCTGATGCACTTGTACAGgtatgttaatagtcaaaattcatttcccttcttaatatgtgaatttcctaaagctgtgctttatccacaaaccaaacagATAGCTTC 2730
 L  L  I  V  F  E  S  R  P  D  A  L  V  Q                                                     I  A  S  423
ACTTGCCATCCGTAGTGGAAATGGTCTTCTGCTGAAGGGTGGAAAGGAGGCCCGGCGATCAAATGCTATCTTACACAAGgtaccattgcctcagatttcatatcattatttgcctcaaaatttatcacta 2860
 L  A  I  R  S  G  N  G  L  L  L  K  G  G  K  E  A  R  R  S  N  A  I  L  H  K                   452
cagctcttttaagttcatggtaaatttctagGTGATCACTGATGCAATTCCAGAGACTGTTGGGGGTAAACTCATTGGACTTGTGACTTCAAGAGAAGAGATTCCTGATTTGCTTAAGgtaagaacagat 2990
                  V  I  T  D  A  I  P  E  T  V  G  G  K  L  I  G  L  V  T  S  R  E  E  I  P  D  L  L  K  480
ttacaagctaggagctgcaacagttcttttgtatcttttgtaaactggaacccaccatttgcatttgtgttacagCTTGATGACGTTATCGATCTTGTGATCCCAAGAGGAAGCAACAAGCTTGTTACTC 3120
                                                                L  D  D  V  I  D  L  V  I  P  R  G  S  N  K  L  V  T  Q  499
AGATAAAAAAATACTACAAAAATCCCTGTGCTAGGTCATGCTGgtatggttgcaagtttgtttttccagaagattctttacttggattgtgctagagtgtgacgatggcttaattgtgtacttgcagATG 3250
  I  K  N  T  T  K  I  P  V  L  G  H  A  D                                                     G  514
GAATCTGTCATGTATATGTCGACAAGGCTTGTGATACGGATATGGCAAAGCGCATAGTTTCTGATGCAAAGTTGGACTATCCAGCAGCCTGTAATGCGATGgtaagagaacttttttaccttccatcgaga 3380
  I  C  H  V  Y  V  D  K  A  C  D  T  D  M  A  K  R  I  V  S  D  A  K  L  D  Y  P  A  A  C  N  A  M                   547
tttaattaatacagtgggagattctaaagttcaactgactcatttcatcttctctcgtctctttcagGAAACCCTTCTTGTGCATAAGGATCTAGAGCAGAATGCTGTGCTTAATGAGCTTATTTTTGCT 3510
                                                                E  T  L  L  V  H  K  D  L  E  Q  N  A  V  L  N  E  L  I  F  A  568
CTGCAGAGCAATGgtacgtcataaatggcccaatcatttgttggtctatcttaaccatttatttgacctcttgttaccttccatctggatgtctcatagatatacatgtagcctgtttgattataaatat 3640
L  Q  S  N  G                                                                                  572
tgaatggtcatctcatgaaaacatttctagagtggcataactcatgagatatattaaactacagGAGTCACTTTGTATGGTGGACCAAGGGCAAGTAAGATACTGAACATACCAGAAGCACGGTCATTCA 3770
                                                                V  T  L  Y  G  G  P  R  A  S  K  I  L  N  I  P  E  A  R  S  F  N  594
ACCATGAGTACTGTGCCAAGGCTTGCACTGTTGAAGTTGTAGAAGACGTTTATGGTGCTATAGATCACATTCACCGACATGGGAGgtagaaactcgacatatacaggcattgactttagaaattctttgca 3900
  H  E  Y  C  A  K  A  C  T  V  E  V  V  E  D  V  Y  G  A  I  D  H  I  H  R  H  G  S                622
tatgtagtggaaatgttcactcgttatcttgtcttgtatgttgttacgagcagTGCACACACAGACTGCATTGTGACAGAGGATCACGAAGTTGCAGAGCTATTCCTTCGCCAAGTGGATAGGtaaagta 4030
                                                  A  H  T  D  C  I  V  T  E  D  H  E  V  A  E  L  F  L  R  Q  V  D  S                645
ctgaatctttaacttgcttattatctgtctttgatttttcttggaaactgactgtaagatgttgcgaccttgaacagCGCTGCTGTGTTCCACAACGCCAGCACAAGATTCTCAGATGGTTTCCGATTTG 4160
                                                                A  A  V  F  H  N  A  S  T  R  F  S  D  G  F  R  F  G  663
GACTTGGTGCAGAGgtaagtcagagacatacacataagtctatagattaaaaacaaatgaaaaagaggaagagtgagtgatgaaaaaagtattggttgtggtatatagGTGGGGGTAAGCACGGGCAGGATC 4290
  L  G  A  E                                                                                    V  G  V  S  T  G  R  I  675
CATGCTCGTGGTCCAGTCGGGGTCGAAGGATTACTTACAACGAGATGgtacaattttagttactcaaagcaccattgttatgtcaatgaaagacccacaataagccttttttcctatgcttcttttaatttt 4420
H  A  R  G  P  V  G  V  E  G  L  L  T  T  R  W                                                691
tcatggtgaaatggttgcagGATAATGAGAGGAAAAGGACAAGTTGTCGACGGAGACAATGGAATTGTTTACACCCATCAGGACATTCCCATCCAAGCTTAAacaagacttccggagtgtgtgtttgtgta 4550
                  I  M  R  G  K  G  Q  V  V  D  G  D  N  G  I  V  Y  T  H  Q  D  I  P  I  Q  A  *  717
tttggttgagacttgaggagagacacagaggaggatgggctttttgtttcctctctgcttagtactcatatcctatcattattattactactcattattgaaaccctcgcttatgtagtggtt 4680
ttgatttagggttaggattgcaccaaaaataaagatccacttttaccacttagtcttgctcataagtacgatgaagaacatttaattagcttctcttcttgtcattgtaagctacctacacatttctgatct 4810
ttatcaaggtactactactttttcatttcgcttatctataaatatatttcgatttgcattggaaatcacaagttgaatcagaactggaaactcttaaccataaattctcaaagattgtgctacatttgaaa 4940
gctaacaatgaacacagaaaagaacctcacaatctcattcttaggtacttagctatcaacctaacgtttcactgttccatgaactccttaacattcttcaaccataaaaagtactctcggttcactttc 5070
ctcgtcatgtactcgtgcaggaaattcgtcgtgctagctttcactcctctgatctccaaaaacttgtagaacgtcttcttcagatttctcgtccaagtctctgcaaaaagcgatgaaacacaaaccaacat 5200
```

Fig. 2. Structure and restriction map of the *At-P5S* gene. (A) The gene map is represented to scale as indicated. Exons within the coding region are marked with black boxes and introns by solid lines. White boxes indicate 5′ and 3′ non-translated regions. The shaded box in the protein scheme indicates the conserved region between bacterial γ-glutamyl kinase and γ-glutamyl phosphate reductase. (B) Nucleotide sequence and deduced amino acid sequence of the *At-P5S* gene. Protein-deduced sequences are shown in uppercase and non-coding sequences in lowercase. The stop codon is denoted by an asterisk. At the 3′ end, putative polyadenylation sites are underlined, and the adenine residue in which polyadenylation is found is indicated in bold type. Arrows represent the position of the primers used for PCR.

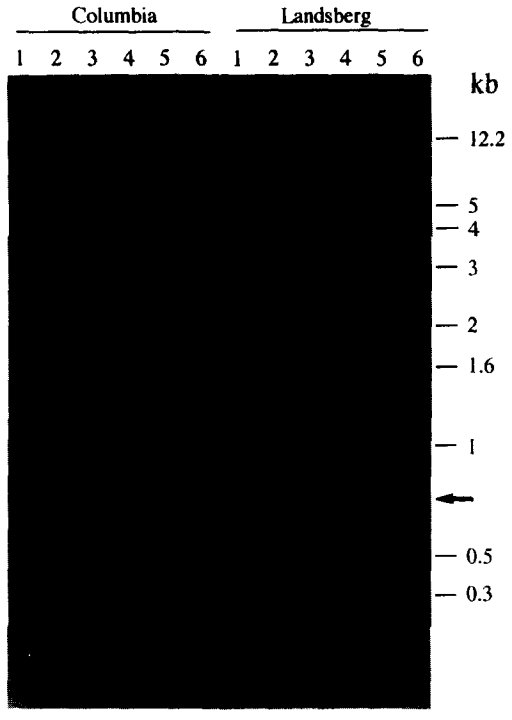


Fig. 3. Southern hybridization of *At-P5S* to genomic DNA of *Arabidopsis thaliana* cv. Columbia and Landsberg. DNA was digested with *Bgl*II (lane 1), *Dde*I (lane 2), *Dra*I (lane 3), *Eco*RI (lane 4), *Hin*dIII (lane 5) and *Xba*I (lane 6). Hybridization was carried out with the full-length *At-P5S* cDNA. Molecular sizes are given in kb.

blot analysis (Fig. 4A). *At-P5S* transcript is more abundant in flowers than in roots and leaves. *At-P5S* expression was relatively low in siliques. The highest concentration of proline was also found in flowers and, to a lesser extent, in siliques and in roots (Fig. 5A). In contrast, mature leaves contained a relatively lower level of proline.

When 10-day-old seedlings were treated with 170 mM NaCl, *At-P5S* transcript accumulates after 4 h to reach a maximum level at 8 h (Fig. 4B). The level of proline measured after 24 h of treatment was 7-fold higher than in non-treated seedlings (Fig. 5B).

4. Discussion

We report here the isolation and characterization of a cDNA and the corresponding *At-P5S* gene encoding the Δ^1-pyrroline-5-carboxylate synthetase from *A. thaliana* as well as features of the encoded polypeptide.

The deduced amino acid sequence of *At-P5S* showed highest similarity with P5C synthethase from *V. aconitifolia*. Alignment of *At-P5S* with all presently available similar proteins revealed two enzymatic domains corresponding to bacterial γ-glutamyl kinase and to γ-glutamyl phosphate reductase. At-P5C synthetase is thus a bifunctional enzyme as it was proposed for P5C synthetase from *V. aconitifolia* [8].

In eubacteria, γ-glutamyl kinase and γ-glutamyl phosphate reductase which are encoded by the *proBA* operon form a multimolecular complex allowing a direct and protected transfer of the very labile γ-glutamyl phosphate intermediate [13,14,30]. Interestingly in yeast, γ-glutamyl kinase [31] and γ-glutamyl phosphate reductase ([32]; M.C. Brandriss, personal communication) functions are carried by two different proteins. P5C synthetase activity was also reported from mammalian cells [33]. However, it is not known whether this P5C synthetase activity is mediated by one or two proteins. A conserved region is present in the bacterial kinase and reductase proteins. This sequence similarity may have facilitated gene fusion during plant evolution through homologous recombination. This conserved region could also correspond to two similar functional domains important to channel the unstable intermediate, γ-glutamyl phosphate. One of the two domains may have been eliminated during plant evolution leading to the formation of a bifunctional enzyme. This phenomenon may be linked to a better regulation of proline synthesis for improved adaptation of the plant cell to osmotic stress. From an evolutionary perspective, it will be interesting to identify gene(s) involved in the synthesis of P5C in plants that do not accumulate proline for osmotic adjustment and in this way to determine whether this putative gene fusion presents an advantage for plants through evolution.

The determination of the secondary structure of *At-P5S* allows a first identification of important motifs which could be implicated in the multiple functions of At-P5S. Analysis of the primary and secondary structure of At-P5S suggests the presence of a potential ATP-binding domain at the N-terminus downstream from a β-sheet with the GXXXXGR signature. The presence of a βαβ secondary structure at the C-terminus may be part of the non-covalent NAD(P)H-binding domain. Based on sequence analysis, a leucine zipper motif was identified in each of the two enzymatic domain of P5CS of *V. aconitifolia* [8]. These motifs are not present in an α-helix and do not match the highly conserved consensus of heptads repeats with leucine at every seventh position forming a ridge on one side of an α-helix that interacts with a complementary helix [34]. However, these leucine motifs may still participate, with a reduced affinity, in protein–protein interaction to maintain inter- or intra-molecular interaction of the P5C synthetase.

In the *proB74 E. coli* mutant, a mutation in γ-glutamyl kinase protein of the aspartate at position 107 to an asparagine re-

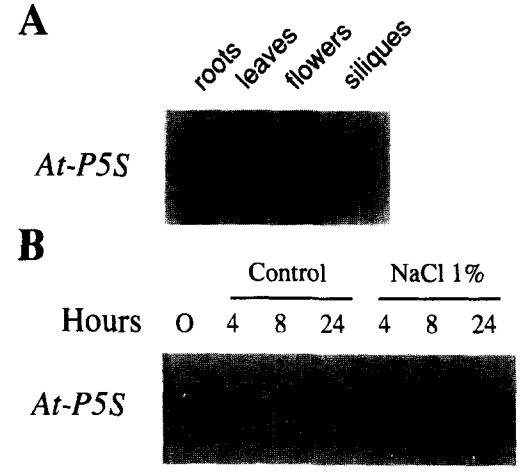


Fig. 4. *At-P5S* transcript analysis in *Arabidopsis thaliana*. (A) RNA gel blot analysis of total RNA from roots, leaves, flowers, and siliques (10 μg/lane). (B) *At-P5S* mRNA accumulation in *Arabidopsis* during NaCl treatment. Total RNA was isolated from 10-day-old seedlings after 4, 8, and 24 h treatment and analyzed by Northern blotting (20 μg/lane). Blots were hybridized with the full-length *At-P5S* cDNA.

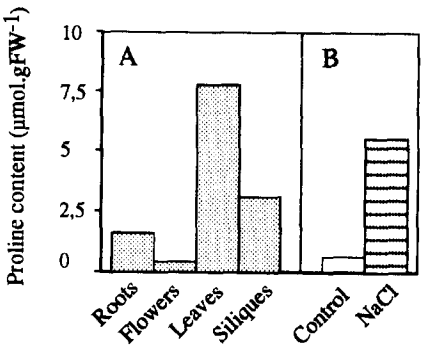


Fig. 5. Proline level in different organs of mature flowering plants (A) and in 10-day-old seedlings treated with 170 mM NaCl for 24 h (B). Standard deviation was less than 5%.

sulted in an overproduction of proline and enhanced tolerance to osmotic stress [28]. The aspartate 107 is implicated in the feedback inhibition of the protein by the end product proline [28]. Interestingly, this aspartate is conserved in P5C synthetase in *V. aconitifolia* at position 128 but not in *A. thaliana*. However, two aspartate residues in *V. aconitifolia* and in *A. thaliana* at position 126 and 125 respectively are conserved. Site-directed mutagenesis may indicate whether they are implicated in the feedback inhibition by proline.

Nucleotide sequence analysis reveals identical sequences in the coding region between *At-P5S* cDNA and the corresponding gene. The *At-P5S* sequence presented in this paper shows two different amino acids with the recently published P5C synthetase from the same *Arabidopsis* ecotype [9], a cysteine and a leucine residues at positions 42 and 467 in *At-P5S* are replaced by two phenylalanines in P5C synthetase [9]. This discrepancy may be due to the utilization of sub-ecotypes of Columbia. The *At-P5S* gene is separated into 20 exons by 19 introns. The total length of the introns represents almost 50% of the transcribed sequence. The restriction map of the genomic region containing *At-P5S* is consistent with Southern blot analysis of the *Arabidopsis* genomic DNA suggesting a single-gene copy of *At-P5S* in *A. thaliana* genome, corroborating the results of [9]. *At-P5S* mapped at the bottom of the chromosome two, its map position not matching any mapped mutants.

The expression of *At-P5S* is rapidly triggered by salt stress in seedlings and is followed by proline accumulation, as already observed [9]. It indicates that *At-P5S* is osmoregulated and plays a key role in proline biosynthesis. In addition, higher basal level of *At-P5S* transcripts were found in roots, leaves, and flowers of mature plant compared to that in seedlings. However, leaves contain a relatively low level of proline. This result suggests that proline produced in leaves may be transported towards the flowers. Proline may have a role in the differentiation of flowers as well as in the desiccation processes of seeds [10,35–37].

At-P5S is present as a single copy in the *A. thaliana* genome and its expression appears to be regulated during development and under osmotic stress. It will be of interest to identify the *cis* elements important for activation of the promoter.

Accumulation of proline is proposed to play an important role in the osmo-adaptation of the plant cell; however, no direct experiment has been performed that directly addresses the role of proline in osmotic adaptation of plants. The availability of the *At-P5S* cDNA allows the generation of transgenic plants with *At-P5S* antisense constructs under the control of an inducible promoter to specifically block the pathway during osmotic stress. In addition, analysis of transgenic plants that overexpress *At-P5S* may open the route for engineering crops more tolerant to environmental stress.

Acknowledgements: The authors thank J. Giraudat and F. Gosti for providing the *Arabidopsis* cDNA library, C. Dean and C. Lister for the recombinant inbred lines and for the calculation of linkage, R. Villarroel for help in the sequencing work, B. Van de Cotte for skilful technical assistance, P. Rouzé for fruitful discussion, M. May and P. Frendo for critical reading of the manuscript, and Martine De Cock for final preparation of the manuscript. This work was supported by grants from the Belgian Programme on Interuniversity Poles of Attraction (Prime Minister's Office, Science Policy Programming, #38) and the Vlaams Actieprogramma Biotechnologie (ETC 002). A.S. was supported by an EMBO (ALTF106-1994) long-term postdoctoral fellowship, and S.J. by an International Scientific Cooperation Initiative 'Marie Curie', EC fellowship (ERBCISTGT940005). N.V. is a Postdoctoral Fellow of the National Fund for Scientific Research (Belgium).

References

[1] Boyer, J.S. (1982) Science 218, 443–448.
[2] Csonka, L.N. and Hanson, A.D. (1991) Annu. Rev. Microbiol. 45, 569–606.
[3] Poulin, R., Larochelle, J. and Hellebust, J.A. (1987) J. Exp. Zool. 243, 365–378.
[4] Brown, L.M. and Hellebust, J.A. (1978) Can. J. Bot. 56, 676–679.
[5] Delauney, A.J. and Verma, D.P.S. (1993) Plant J. 4, 215–223.
[6] Csonka, L.N. (1989) Microbiol. Rev. 53, 121–147.
[7] Itai, C. and Paleg, L.G. (1982) Plant Sci. Lett. 25, 329–335.
[8] Hu, C.-A.A., Delauney, A.J. and Verma, D.P.S. (1992) Proc. Natl. Acad. Sci. USA 89, 9354–9358.
[9] Yoshiba, Y., Kiyosue, T., Katagiri, T., Ueda, H., Mizoguchi, T., Yamaguchi-Shinozaki, K., Wada, K., Harada, Y. and Shinozaki, K. (1995) Plant J. 7, 751–760.
[10] Verbruggen, N., Villarroel, R. and Van Montagu, M. (1993) Plant Physiol. 103, 771–781.
[11] Murashige, T. and Skoog, F. (1962) Physiol. Plant. 15, 473–497.
[12] Pruitt, R.E. and Meyerowitz, E.M. (1986) J. Mol. Biol. 187, 169–183.
[13] Deutch, A.H., Rushlow, K.E. and Smith, C.J. (1984) Nucleic Acids Res. 12, 6337–6355.
[14] Omori, K., Suzuki, S., Imai, Y. and Komatsubara, S. (1991) J. Gen. Microbiol. 137, 509–517.
[15] Gosti, F., Bertauche, N., Vartanian, N. and Giraudat, J. (1995) Mol. Gen. Genet. 246, 10–18.
[16] Koornneef, M. and Stam, P. (1992) in: Methods in *Arabidopsis* Research (Koncz, C., Chua, N.-H. and Schell, J. eds.) pp. 83–99, World Scientific, Singapore.
[17] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, a Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
[18] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[19] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[20] Rost, B. and Sander, C. (1993) Proc. Natl. Acad. Sci. USA 90, 7558–7562.
[21] Ogura, M., Kawata-Mukai, M. and Itaya, M. (1994) J. Bacteriol. 176, 5673–5680.
[22] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.
[23] Lister, C. and Dean, C. (1993) Plant J. 4, 745–750.
[24] Logemann, J., Schell, J. and Willmitzer, L. (1987) Anal. Biochem. 163, 16–20.
[25] Bates, L.S. (1973) Plant and Soil 39, 205–207.
[26] Swindells, M.B. (1993) Protein Sci. 2, 2146–2153.
[27] Wierenga, R.K., Terpstra, P. and Hol, W.G.L. (1986) J. Mol. Biol. 187, 101–107.

[28] Csonka, L.N., Gelvin, S.B., Goodner, B.W., Orser, C.S., Siemieniak, D. and Slightom, J.L. (1988) Gene 64, 199–205.
[29] Luehrsen, K.R., Taha, S. and Walbot, V. (1994) Prog. Nucleic Acids Res. Mol. Biol. 47, 149–193.
[30] Baich, A. (1969) Biochim. Biophys. Acta 192, 462–467.
[31] Li, W. and Brandriss, M.C. (1992) J. Bacteriol. 174, 4148–4156.
[32] Tomenchock, D.M. and Brandriss, M.C. (1987) J. Bacteriol, 169, 5364–5372.
[33] Smith, R.J., Downing, S.J., Phang, J.M., Lodato, R.F. and Aoki, T.T. (1980) Proc. Natl. Acad. Sci. USA 77, 5221–5225.
[34] O'Shea, E., Klemm, J.D., Kim, P.S. and Alber, T. (1991) Science 254, 539–544.
[35] Mutters, R.G., Ferreira, L.G.R. and Hall, A.E. (1989) Crop Sci. 29, 1497–1500.
[36] Walton, E.F., Clark, C.J. and Boldingh, H.L. (1991) Plant Physiol. 97, 1256–1259.
[37] Venekamp, J.H. and Koot, J.T.M. (1984) J. Plant Physiol. 116, 343–349.