Branched-chain-amino-acid biosynthesis in plants: molecular cloning and characterization of the gene encoding acetohydroxy acid isomeroreductase (ketol-acid reductoisomerase) from *Arabidopsis thaliana* (thale cress)

Renaud DUMAS,*† Gilles CURIEN,* Richard T. DeROSE† and Roland DOUCE*

*Unité Mixte CNRS/Rhône-Poulenc (Unité associée au Centre National de la Recherche Scientifique, U. M. 41)

and †Service de Biologie Moléculaire et Biochimie Cellulaire, Rhône-Poulenc Agrochimie, 14-20 rue Pierre Baizet, 69263 Lyon Cédex 09, France

Towards the goal of gaining a better understanding of the molecular mechanisms controlling branched-chain-amino-acid biosynthesis in plants, we have isolated, sequenced and characterized a gene encoding acetohydroxy acid isomeroreductase (ketol-acid reductoisomerase) from *Arabidopsis thaliana* (thale cress). Comparison between the acetohydroxy acid isomeroreductase cDNA and the genomic sequence has

allowed us to determine the exon structure of the coding region. The isolated acetohydroxy acid isomeroreductase gene is distributed over approx. 4.5 kbp and contains nine introns (79–347 bp). The transcriptional start site was found to be 52 bp upstream of the translational initiation site. Southern-blot analysis of *A. thaliana* genomic DNA shows that the acetohydroxy acid isomeroreductase is encoded by a single-copy gene.

INTRODUCTION

Acetohydroxy acid isomeroreductase (ketol-acid reductoisomerase, EC 1.1.1.86), the second enzyme in the parallel biosynthetic pathway of isoleucine/valine, catalyses a two-step reaction in which the substrate, either 2-acetolactate or 2-aceto-2-hydroxybutyrate, is converted via an alkyl migration and a NADPH-dependent reduction into 2,3-dihydroxy-3-methylbutyrate or 2,3-dihydroxy-3-methylvalerate respectively. The demonstration that selective inhibitors of acetohydroxy acid isomeroreductase (Schulz et al., 1988; Aulabaugh and Schloss, 1990) give rise to herbicidal effects has led to a renewed interest in the study of this enzyme. Most of our biochemical knowledge concerning acetohydroxy acid isomeroreductase come from studies conducted with purified prokaryotic enzymes (Arfin and Umbarger, 1969; Shematek et al., 1973; Hoffer et al., 1975; Chunduru et al., 1989; Aulabaugh and Schloss, 1990). Detailed molecular analyses have described the isolation of genes encoding this enzyme from several prokaryotes (Blazey and Burns, 1984; Wek and Hatfield, 1986; Aguilar and Grasso, 1991; Godon et al., 1992; Rieble and Beale, 1992) and only from a few eukaryotes, such as *Saccharomyces cerevisiae* (baker's yeast) (Petersen and Holmberg, 1986) and the fungus *Neurospora crassa* (Sista and Bowman, 1992). Such studies have led to greater insights into the structure–function relationship of this enzyme.

Bacterial acetohydroxy acid isomeroreductase has been used as a model enzyme for the design of new herbicide molecules (Schulz et al., 1988; Aulabaugh and Schloss, 1990). In order to characterize better the true herbicide target, that is, the acetohydroxy acid isomeroreductase from a higher plant, we have purified this enzyme from the stroma of spinach (*Spinacia oleracea*) chloroplasts (Dumas et al., 1989). We have also isolated cDNAs encoding this protein from both spinach (Dumas et al., 1991) and thale cress (*Arabidopsis thaliana*) (Curien et al., 1993) and have shown that the spinach acetohydroxy acid isomero-

reductase gene is present as a single copy per haploid genome (Dumas et al., 1991). To investigate kinetic properties of the plant higher-acetohydroxy acid isomeroreductase, the spinach cDNA was overexpressed in *Escherichia coli* (Dumas et al., 1992b). These studies showed that the structural and kinetic properties of the plant enzyme (Dumas et al., 1992a) are very different from those determined for its bacterial counterpart. Furthermore, we must consider the possibility that herbicides may act upon the target enzyme at the level of gene expression. For example, in *E. coli*, it is known that expression of the acetohydroxy acid isomeroreductase gene is induced by the binding of either 2-acetolactate or 2-aceto-2-hydroxybutyrate to a positive activator encoded by the gene *ilvY* (Wek and Hatfield, 1986, 1988). Here, as a preliminary approach to understanding the molecular regulation of acetohydroxy acid isomeroreductase in higher plants, we report the isolation and characterization of the gene encoding this enzyme from *A. thaliana*.

EXPERIMENTAL

Gene isolation

The acetohydroxy acid isomeroreductase gene was cloned from an *A. thaliana* (ssp. *columbia*) genomic library (λEMBL3) purchased from Clontech (Palo Alto, CA, U.S.A.). Approx. 50000 recombinant phage were plated on the host strain *E. coli* NM539, grown overnight at 37 °C and screened using an *A. thaliana* acetohydroxy acid isomeroreductase cDNA (Curien et al., 1993) as the hybridization probe. Plaques were transferred to nitrocellulose filters (Schleicher and Schuell), and the DNA was denatured for 1 min in 1.5 M NaOH/1.5 M NaCl, neutralized 10 min in 0.5 M Tris/HCl (pH 7.2)/1.5 M NaCl/1 mM EDTA and washed 5 min in 3 × SSC (1 × SSC is 0.15 M NaCl/15 mM sodium citrate buffer, pH 7.0). Filters were u.v.-cross-linked and then prehybridized for 4 h at 65 °C in 6 × SSC/0.5% (w/v) SDS/0.125% (w/v) dry milk. The cDNA probe was ³²P-labelled

† To whom correspondence should be sent.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X69880.

Figure 1 Restriction map, exon-intron pattern and sequencing strategy of the *A. thaliana* acetohydroxy acid isomeroreductase gene

Shown above are the restriction sites used during subcloning into the vector pBluescript II. Open boxes indicate introns, shaded boxes show open reading frames and hatched boxes represent 5'- and 3'-non-coding transcribed regions. Arrows indicate the sequence strategy on both strands. Numbers indicate intron designation.

with a random-primed-DNA labelling kit (Boehringer) and hybridization was allowed to proceed overnight at 65 °C. Membranes were washed consecutively at 65 °C for 30 min in (a) 2 × SSC containing 0.2 % (w/v) SDS and (b) 0.5 × SSC containing 0.2 % (w/v) SDS. Filters were autoradiographed 24 h at -80 °C. Two positive plaques were re-plated and rescreened until all plaques gave a positive signal. The acetohydroxy acid isomeroreductase-coding regions of these two positive clones were further identified by Southern hybridization, subcloned into plasmid vector pBluescript II (Stratagene) and transformed by electroporation into *E. coli* strain DH10B for amplification (Dower et al., 1988). Plasmid DNA was prepared by the method of Birnboim and Doly (1979).

DNA sequencing

DNA sequence analysis was carried out on both strands by the dideoxy chain-termination method (Sanger et al., 1977) with T7 DNA polymerase (Pharmacia) using [α -³⁵S]dATP as the radiolabelled nucleotide (Amersham). Restriction fragments were subcloned into plasmid vector pBluescript II (Stratagene) and sequenced by using 18-mer oligonucleotide primers synthesized with a Cyclone Plus DNA synthesizer (Millipore). Sequencing products were then analysed on denaturing 6%-(w/v)-polyacrylamide gels. Sequence data were analysed using Gene Works (IntelliGenetics, Mountain View, CA, U.S.A.) and DNA Strider (CEA, Saclay, France) programs on a Macintosh Quadra 950 microcomputer.

Southern-blot analysis

Total DNA isolated from *A. thaliana* (ssp. *columbia*) leaves (10 µg/digestion) was digested for 4 h with several restriction enzymes (New England Biolabs) and concentrated by ethanol precipitation. DNA was fractionated by agarose (0.7%)-gel electrophoresis, blotted and hybridized to a radiolabelled [α -³²P]dCTP probe, as described by Maniatis et al. (1982). Hybridization conditions were identical with those used for the library screening.

Primer extension

A specific 27-mer oligonucleotide (5' CATTT TCTAA GTTGT AGCAA ATTGA CT 3') complementary to the 5'-end of the cDNA was end-labelled with [γ -³²P]ATP using T4-polynucleotide kinase (New England Biolabs) and purified by electro-

phoresis through a 20%-polyacrylamide/urea sequencing gel. After co-precipitation of mRNA (5 µg) and oligonucleotide (20000 c.p.m.), the pellet was resuspended in 30 µl of annealing buffer composed of 20 mM Tris/HCl (pH 8.0)/200 mM NaCl/1 mM EDTA. Annealing was carried out for 1 h at 65 °C, followed by slowly cooling to 37 °C for 1 h. Reverse transcription was carried out for 1 h at either 37 or 42 °C by adding 70 µl of a buffer containing 7.5 mM MgCl₂/7.5 mM dithiothreitol/0.7 mM-dNTP/10 units M-MuLV reverse transcriptase (New England Biolabs)/human placental ribonuclease inhibitor (200 units; Pharmacia). The elongation products were extracted with phenol/chloroform, and precipitated with ethanol. The pellet was resuspended in 4 µl of 80 mM NaOH and 4 µl of 0.05 % (w/v) Bromophenol Blue/0.05 % (w/v) xylene cyanol/95 % (v/v) formamide. The elongation products were analysed on a denaturing 6%-(w/v)-polyacrylamide gel.

RESULTS

Gene isolation

By using an *A. thaliana* acetohydroxy acid isomeroreductase cDNA clone (Curien et al., 1993) as a probe, two genomic clones were isolated from a λ EMBL3 *A. thaliana* (ssp. *columbia*) genomic library (Clontech). The isolated genomic clones were characterized by restriction mapping and Southern-blot analyses. One of these clones contained the entire coding region of acetohydroxy acid isomeroreductase sequence of the gene, whereas the other contained only 1200 bp of the 5'-end. Five restriction fragments corresponding to the entire acetohydroxy acid isomeroreductase sequence [1.26 kbp (*EcoRV*-*Bam*HI), 1.67 kbp (*Bam*HI-*EcoRV*), 0.68 kbp (*EcoRV*-*EcoRV*), 0.46 kbp (*EcoRV*-*EcoRV*) and 1.4 kbp (*EcoRI*-*EcoRI*)] were subcloned into the plasmid vector pBluescript II and sequenced (Figure 1). Synthetic oligonucleotides corresponding to internal regions of the acetohydroxy acid isomeroreductase gene were used as primers during sequencing reactions. The completed nucleotide sequence of the isolated gene was compared with the acetohydroxy acid isomeroreductase cDNA nucleotide sequence.

Gene structure

Comparison of the genomic and cDNA sequences of *A. thaliana* acetohydroxy acid isomeroreductase split the known transcribed region into ten exons, as shown in Figures 1 and 2. All the introns have consensus dinucleotide splice junctions at their 5' (GT) and

3' (AG) borders (Table 1). Interestingly, an adenine was always found 2 bp upstream of the 5'(GT) border. As with most other eukaryotic genes, the exons were found to have a lower A + T percentage (48–58 %, with an average of 53 %) than the introns (64–71 %, with an average of 67 %).

Primer-extension analysis was performed to map the initiation site of transcription. A 27-mer synthetic oligonucleotide complementary to the beginning of the mRNA was used during reverse transcriptase primer-extension experiments. The major product was 54 nucleotides in length, suggesting that the start of transcription occurs 52 bp upstream of the translational initiation site (Figure 3). The position +1 in the gene nucleotide sequence was attributed to the nucleotide corresponding to the elongation product found in the primer-extension experiment (–52 from the initiation codon). A potential TATA box with the sequence TATTAA was found at position –27, and a CAAT box was located at nucleotide –170. A possible alternative TATA box with the sequence TATAAA was found at position –107, and a ten-nucleotide perfect palindromic sequence GGCCT/AGGCC can be found at position –115. The nucleotide sequence around the initiation codon (GAAAATGG) is slightly different from the corresponding sequence of spinach (ATCAATGG) and to the proposed consensus sequence (AAAAATGG) for translation initiation in dicotyledonous plants (Cavener and Ray, 1991). The last exon contains a putative polyadenylation signal AATAAA located 161 bp downstream from the translational stop signal. Interestingly, the polyadenylation signal is followed by a putative stem-loop secondary structure (Figure 4). Comparison of the 3'-end of the mRNA (Curien et al., 1993) and the gene corresponding sequence shows that the cleavage and polyadenylation site of the mRNA was located in the middle of the perfect palindromic sequence (CGAAATTTTCG) found in the putative stem-loop secondary structure (Figure 4). Except for a single base-pair difference located at position +3092 of the gene (gene, C/cDNA, T), changing an alanine (gene) into a valine residue (cDNA), the coding regions of the gene and cDNA are identical.

Southern-blot analysis

Southern-blot analysis was used to examine the number of genes encoding acetohydroxy acid isomeroreductase in *A. thaliana*. Total DNA was digested with enzymes that do not cut (*PvuII*) or do cut (*EcoRV* and *EcoRI*) within the genomic sequence and probed with ³²P-labelled genomic restriction fragments corresponding to the centre of the gene [*BamHI*–*EcoRV* (1.67 kbp)] or to the 3'-end of the gene [*EcoRI*–*EcoRI* (1.4 kbp)] (see Figure 1). As Figure 5 shows, these probes hybridized with a single band of 9 kbp when total DNA was digested with *PvuII*. After digestion by *EcoRV*, the *BamHI*–*EcoRV* probe hybridized with a single band of 2.9 kbp whose size agrees with the genomic restriction fragment determined by sequencing of the genomic clone (Figures 1 and 5). In the same fashion, when DNA was digested with *EcoRI*, the *EcoRI*–*EcoRI* probe hybridized to a single band whose size (1.4 kbp) is identical with that of the probe. We conclude, therefore, as for the spinach acetohydroxy acid isomeroreductase gene (Dumas et al., 1991), that, in *A. thaliana*, this gene is present as a single copy.

Deduced amino acid sequence and sequence-similarity studies

Coding sequence comparison between *S. cerevisiae*, *N. crassa* and *S. oleracea* previously described this enzyme as having two conserved regions around (a) the putative NADPH-binding site (Dumas et al., 1991; Curien et al., 1993) and (b) a putative Mg²⁺-

binding site (Sista and Bowman, 1992). If we compare the *A. thaliana* amino acid sequence with the previously reported eukaryotic amino acid sequences, we note that the *S. oleracea* (595 amino acids) and the *A. thaliana* (591 amino acids) proteins are 30 % longer than the *S. cerevisiae* (395 amino acids) and the *N. crassa* (400 amino acids) proteins. As Figure 6 shows, this difference corresponds essentially to (a) a difference of 40 amino acids between the plant chloroplastic transit peptide (*S. oleracea* 72 amino acids, *A. thaliana* 69 amino acids) and the fungus mitochondrial transit peptide (*N. crassa*, *S. cerevisiae*, approx. 30 amino acids) and (b) to an insertion of a region containing 141 amino acids (325–465 amino acids, *A. thaliana*; 331–471 amino acids, *S. oleracea*). These comparisons also revealed that the plant proteins are divergent in their transit peptides and around their putative processing site (Curien et al., 1993). However, they are highly conserved in the mature protein region, including the segment corresponding to the 141-amino-acid insertion. To assign a function to these 141 amino acids, we have compared the inserted peptide sequence with the sequences within the EMBL, GenBank and SwissProt libraries, without revealing any significant similarities with these known sequences.

DISCUSSION

As a first step towards understanding the molecular regulation of branched-chain amino acids in plants, we have isolated and characterized a gene encoding acetohydroxy acid isomeroreductase from *A. thaliana*. Of the three eukaryotic acetohydroxy acid isomeroreductase genes sequenced, namely those of *S. cerevisiae* (Petersen and Holmberg, 1986), *N. crassa* (Sista and Bowman, 1992) and *A. thaliana* (the present study), only the last two contain introns. The *N. crassa* gene was found to contain fewer introns (four) than the *A. thaliana* gene (nine). Also, the location of these introns is very different between these two genes. For instance, the binding site for the pyrophosphate moiety of NADPH is in exon 1 in *N. crassa*, whereas the same site is split between exons 2 and 3 in *A. thaliana*.

In higher eukaryotes, a TATA box is usually located 26–34 bp upstream of the transcriptional start site (Breathnach and Chambon, 1981). A TATA-like sequence (TATTAA) is present at position –27, as determined from primer-extension experiments. A CAAT box can be found also at position –170. Because the cDNA sequence previously described contains only 24 bp upstream of the initiation codon, the possibility that an intron exists in the 5' untranslated region cannot be dismissed. Interestingly, a second putative TATA box (TATAAA) is located at position –107, and sequences showing similarity to consensus intron splice junctions are found at nucleotide –48 (5'-border, AG GTAAA; consensus AG GTAAAG) and +30 (3'-border, CGCAG T; consensus TGCAG G), leading to a possible intron of 78 bp. Assuming this second TATA box is used and the existence of an intron within the 5' untranslated region is correct, the transcriptional start site, located 52 bp upstream of the initiation codon, would then be displaced to a thymidine residue located 29 bp downstream of the second TATAAA box. The upstream CAAT box would then be located at a more normal 92 bp upstream of this thymidine residue. Experiments are currently underway to determine if an intron exists within the 5' untranslated region of this gene. Unlike the spinach cDNA, which does not contain a consensus AATAAA polyadenylation signal, the *A. thaliana* gene contains a consensus polyadenylation signal located 161 bp downstream from the translational stop signal. Interestingly, the cleavage and polyadenylation site of the mRNA was within a palindromic sequence located in a putative stem-loop secondary structure.

10	20	30	40	50	60	70	80	90	
1234567890123456789012345678901234567890123456789012345678901234567890									
GATATCACTTAGCATGATATTGGTTAAATTAATGTACTCTTTATCACAAGCCTTGAATCTATCATACGAAATCTCCTTGAAAAGATT									-941
TTCGCTGATAGTTACATGTGCAATCGATCACCATTGCAAATCCAAACAAAATTTTCGAGTGATCAAGACTCAACGTCAAGTTTTCACATA									-851
TGCAAATTAAGACAACCTTCTCAATACCATTGCAAGAAATCTATAACACCAGATTGAATTTACGTTATAACTAATTATCAGATCCAAAAC									-761
ATCTCTTAAATCACCAGTGAAGTTGCAAAGAAAGGATGTAAGTTGAGCATAAAAGACACAAAACATCAAAAATTATAAAATCACAT									-671
CTCATGATAAGCCAAGAAAAGTTACCAAATAAACATAGCATTTCAGGCCTTGGAAACAGAGCAATCTCTATAGCTTCTCTCTCCTCGTTAT									-581
CATCGACATTACAGGATCAGTGAAGACTTCTTGCTCTCTCATAGTCGAAGAAGACGAAGCTTGATCCATCTCAGATACAAAGAATTT									-491
AACTTGAGAGAAAACCCAGATGAGACTAATCGATCAAACAAACCCGTAAATCTTAAATGTTAAGTCTTTCCCGAACTGTGATTACAACAA									-401
TCTCACGAGTCGCGAAGACGAAGAACGTAGAGGTTTCGCGTTAGGTGTGTAGTGTATACCATTACTTCAATCTGAGCTACTTCAAAA									-311
CCATCAGATCTAAGCTACCTCTGTTAAATCTGAACCGTCTATTACGTTGACTTTGAAACCGTGATATAGACTAGACACCTATGTATCCTC									-221
CACCGCTGGGAAGTCAACATTGAAGCAGTTGATGGACCAACACCAACCAATAGAAAGAAGCTCCACATTTACTTTAATGGGCCTTTTAAA									-131
CTAATAGGCCTAGGCCATTATATATACATGGATTGTTTTATTTATTTTCTCATATCTGTCTCTTGGTGGTGATGAGTAGGTAAACTG									-41
TGTGCGCTTATATATCTCTCTCACACTGACTCTGCGACCCTAATTTGTGAGCTCTCTTCTCTTTGCGCAGTCAATTGTCTACAACCTAGA									50
AAATGGCGGCGGCTACTTCATCCATCGCTCTTCTCTTTTCATGCCCATCTCTTCTTCTTCATCCAAAACCCCTTTGGTCTTCCAAAGCCA									140
MetAlaAlaAlaThrSerSerIleAlaProSerLeuSerCysProSerProSerSerSerSerLysThrLeuTrpSerSerLysAlaA									
GAACCTTGGCTCTACCCAATATCGGTTTCTCTCGTCTTCTTCCAAGTCTCTGAGGTCGCTTACTGCCACCGTCGCTGGAAATGGCGCCA									230
rgThrLeuAlaLeuProAsnIleGlyPheLeuSerSerSerSerLysSerLeuArgSerLeuThrAlaThrValAlaGlyAsnGlyAlaT									
CTGGATCCTCCCTTGGCGCTCGCATGGTTTCTTCGCTGCGGTCAAAGCCCTGTTTCTCTCGATTTTGAGACATCTGTCTTCAAAAAGG									320
hrGlySerSerSerLeuAlaAlaArgMetValSerSerSerAlaValLysAlaProValSerLeuAspPheGluThrSerValPheLysLysG									
AGAAAGTTTCTCTTGCTGGTTACGAAGAGgtttatctcatttaccctcttattcatttctgttctaatgtgtcttattctataatcatt									410
luLysValSerLeuAlaGlyTyrGluGlu									
ttgatgtgtattgattttgttagttgttgctaatgtttggtttctgctgattcttgttttgaTACATTGTGAGAGGAGGAAGAGACTTGT									500
TyrIleValArgGlyGlyArgAspLeuP									
TCAAGCATCTCCAGATGCTTTCAAGGGGATTAAGCAGATTGGTGTGATTGGCTGGGGATCTCAGgtacttaccttttttgggtctcttta									590
heLysHisLeuProAspAlaPheLysGlyIleLysGlnIleGlyValIleGlyTrpGlySerGln									
tctgtagctactgtgcttctgttttgattgcttttttaattgcttacggttgattgcgaataaggaatcttggttctgtgttgatagc									680
aaattggtcagaaaaaggatatgaataaagtgttttcattttcatacttttgatgcatttgagttggtggaatgaacaagattaatat									770
tgtttcatattctttttattcagcatctcttctcttgtaataaacgctttatctatatatgcagGGACCTGCCAGGCTCAGAATTTAAG									860
GlyProAlaGlnAlaGlnAsnLeuAr									
GGATTCACTTGTGGAGGCAAAGTCTGACATTGTTGTCAAGgtatttcaactgcagagttggcttcatacttgatttggttgcgttctgt									950
gAspSerLeuValGluAlaLysSerAspIleValValLys									
tgttacgtgaaggctctgttctgactttgttttcgttttaaataatactattacgtgaATTGGTCTCAGAAAAGGGTCTCGCTCATTGAGG									1040
IleGlyLeuArgLysGlySerArgSerPheGluG									
AGGCACGTGCTGCTGGCTTCACTGAAGAAAGTGGTACTTTGGGTGATATATGGGAAACTATCGCTGCCAGTGATCTGTATTGCTTTTGA									1130
luAlaArgAlaAlaGlyPheThrGluGluSerGlyThrLeuGlyAspIleTrpGluThrIleAlaGlySerAspLeuValLeuLeuLeuI									
TCTCTGATGCTGCTCAAgttaagtgcacttttgttatatatttccactactacttgttgaggattagtcctaagggtgtatacatcgattt									1220
leSerAspAlaAlaGln									

10	20	30	40	50	60	70	80	90	
1234567890123456789012345678901234567890123456789012345678901234567890									
tattctatcctggtttttgcagGCTGATAACTATGAGAAAATATTCTCTCATGAAGCCAAACAGCATTCTTGGTTTATCACACGGGTTT									1310
AlaAspAsnTyrGluLysIlePheSerHisMetLysProAsnSerIleLeuGlyLeuSerHisGlyPhe									
CTACTAGGGCATTACAGTCTCGGACTCGATTTCCTCAAGAACATCAGTGTGGTCGCTGTATGCCCTAAGGGAATGGGTCCTTCTGTG									1400
LeuLeuGlyHisLeuGlnSerSerGlyLeuAspPheProLysAsnIleSerValValAlaValCysProLysGlyMetGlyProSerVal									
AGGAGGCTTTACGTCCAAGGAAAAGAAATTAACGGTGCTGGAATCAACGCCAGTTTTCAGTCCACCAGgttttctatctctttatctct									1490
ArgArgLeuTyrValGlnGlyLysGluIleAsnGlyAlaGlyIleAsnAlaSerPheAlaValHisGln									
actgcaaaagctatctttactgaaaagctcttatacctcatgtgttcattttttcaccactatttctgctattctgttggcagGATGTTG									1580
AspValA									
ACGGTAGAGCCGCCGATGTTGCATTGGGATGGTCAGTAGCACTTGGTTCTCCGTTTACTTTTGTACTACTCTTGAACAGGAGTACAGGA									1670
spGlyArgAlaAlaAspValAlaLeuGlyTrpSerValAlaLeuGlySerProPheThrPheAlaThrThrLeuGluGlnGluTyrArgS									
GTGACATCTTTGGAGAAAGAGgtgagtttgatcccccttttagacctgtctttatatcactcaatcccatacttacatatcttctctttt									1760
erAspIlePheGlyGluArg									
ttgttaaaagGAATTTGCTTGGTGTGTTTACGGAATCGTGGAGTCTCTGTTTAGAAGATACACCGAAAATGGGATGAGTGAAGACTTG									1850
GlyIleLeuLeuGlyAlaValHisGlyIleValGluSerLeuPheArgArgTyrThrGluAsnGlyMetSerGluAspLeu									
GCTTACAAGAACACAGTAGAATGCATCACAGGAACAATTTCAAGGACTATCTTACCCAGgttagtaggaagatatctgtagagaacaga									1940
AlaTyrLysAsnThrValGluCysIleThrGlyThrIleSerArgThrIleSerThrGln									
ctaattcggttttagccaaatataaaaattttaaacctgtttctaatagaattgtgaacatcagGGAATGTTGGCTGTGTACAACTCCTTG									2030
GlyMetLeuAlaValTyrAsnSerLeu									
TCTGAAGAAGGTAAAAAGATTTTGAGACTGCATACAGCGCATCCTTCTATCCTTGTATGGAGATTCTCTATGAATGTTACGAGGATGTA									2120
SerGluGluGlyLysLysAspPheGluThrAlaTyrSerAlaSerPheTyrProCysMetGluIleLeuTyrGluCysTyrGluAspVal									
CAAAGTGGCAGCGAAATCCGAAGTGTGTCTTAGCCGGTCGTCGTTCTATgtaatgactcttccgaatcttccctttgtttacaattac									2210
GlnSerGlySerGluIleArgSerValValLeuAlaGlyArgArgPheTyr									
ccaaacacagtttgtgcaggagttttgatggattctcttacttataacagGAAAAGGAAGGCTTGCCAGCATTCCCTATGGGAAATATTG									2300
GluLysGluGlyLeuProAlaPheProMetGlyAsnIleA									
ACCAGACAAGAATGTGAAGGTGGGTGAACGCGTCAGGAAGTCCAGACCTGCTGGTGAAGTGGGTCCATTGTATCCCTTACCAGCTGGAG									2390
spGlnThrArgMetTrpLysValGlyGluArgValArgLysSerArgProAlaGlyAspLeuGlyProLeuTyrProPheThrAlaGlyV									
TTTATGTAGCACTTATGATGGCTCAGgtatgtctatgtttatatctggtgtgccaccttttgctttggatttttcggtttatgcaatgta									2480
alTyrValAlaLeuMetMetAlaGln									
gagtcggtttcaatttccggttcagatttttagaaaatatacctctaattggaatattttgtcattgcaataggtttgtatagtctaattg									2570
gattcttcttacttccattttgtttcagagaaattgtgatatcgatctgtttgtattcaattttgaattctttttgtatggttttagt									2660
tcgaattagccgcatggacaaccctagaatttaatatgctatcaattttcagtgatttcttagaatctaaatcatcggtggtttctatg									2750
tggatttgtgtagATTGAGATCTTGAGGAAGAAAGGTCACCTTACTCAGAAATCATCAACGAGAGTGTGATTGAATCCGTGGACTCTCT									2840
IleGluIleLeuArgLysLysGlyHisSerTyrSerGluIleIleAsnGluSerValIleGluSerValAspSerLe									
AAACCCATTATGCACGCCAGGGAGTGTCTTCATGGTTGACAACTGCTCAACACAGCAAGATTGGGTTCCAGGAAATGGCGCCACG									2930
uAsnProPheMetHisAlaArgGlyValSerPheMetValAspAsnCysSerThrThrAlaArgLeuGlySerArgLysTrpAlaProAr									
GTTTGACTACATCTGACCAACAAGCTCTTGTGGCTGTGGACAGTGGTGCAGCAATCAACAGAGACCTAATCAGTAACCTCTCTCTGA									3020
gPheAspTyrIleLeuThrGlnGlnAlaLeuValAlaValAspSerGlyAlaAlaIleAsnArgAspLeuIleSerAsnPhePheSerAs									
TCCAGTCCATGGTGCCATTGAGGTCTGCGCACAGCTCAGGCCTACCGTTGATATCTCTGTGCCTGCTGATGCAGACTTTGTTCCAGCTGA									3110
pProValHisGlyAlaIleGluValCysAlaGlnLeuArgProThrValAspIleSerValProAlaAspAlaAspPheValArgProGl									
GTTGCGTCAATCTAGCAACTGAGTGAAGGTTGAAAGTTTGTCACTCTCTTATTTGTAATCGAGATTAAAGTCGAGAGTTTGTGATGTT									3200
uLeuArgGlnSerSerAsn									
TTCTTAGGCGTGACTGTTTGTGTTTGAAGGATTATGTCTCTTTGCTTTGGTCTTAAATCTACTTAAATCATAAATATAGTTTAACG									3290
AAATTTCGCTTTAACTATTTTCGTAACAACATTGCAAACTTTCAACTTAATTTGTCTAGAATTGTGTATTGTCA									3366

Figure 2 Nucleotide sequence and deduced amino acid sequences of the *A. thaliana* acetohydroxy acid isomeroreductase gene

The nine introns in the gene are shown in lower-case letters. Putative TATA boxes, CAAT box and the polyadenylation signal are designated by sequences within boxes. The major transcriptional start site, as determined by primer extension experiments (see Figure 3), is denoted by a residue with a triangle above it. A region of dyad symmetry and the 5'- and 3'-borders of introns are underlined.

Table 1 Splice sites, length and percentage A+T of introns in the *A. thaliana* acetohydroxy acid isomeroreductase gene

Intron	Length (bp)	5'-Splice site	3'-Splice site	A + T (%)
1	124	AG GTTTA	TTTAG T	71
2	269	AG GTACT	TGCAG G	69
3	106	AG GTATT	CGTAG A	64
4	94	AA GTAAG	TGCAG G	67
5	104	AG GTTTT	GGCAG G	65
6	79	AG GTGAG	AAAAG G	67
7	93	AG GTTAG	ATCAG G	71
8	89	AT GTAAT	AACAG G	64
9	347	AG GTACT	TGTAG A	68
Plant consensus sequence...		AG GTAAG	TGCAG G

Figure 3 Identification of the transcriptional start site of the *A. thaliana* acetohydroxy acid isomeroreductase gene by primer-extension analysis

A 27-mer oligonucleotide complementary to the cDNA sequence was used as a primer for reverse transcription at 37 °C of mRNA (5 µg) and for sequencing the corresponding DNA clone. Lanes G, A, T and C correspond to the sequence of the sense strand. Numbers indicate base-pairs from the adenine within the initiation codon. Lane PE, primer extension product.

Southern-blot analysis discloses that, in *A. thaliana*, the acetohydroxy acid isomeroreductase is encoded by a single copy gene, as found for *S. oleracea* (Dumas et al., 1991), *S. cerevisiae* (Kakar and Wagner, 1964), *N. crassa* (Wagner et al., 1964) and previously reported prokaryotes. In prokaryotes, operons encoding enzymes involved in the biosynthesis of branched-chain amino acids are clustered in distinct regions of the genome. The number and organization of these operons vary among the different bacterial species studied (Somers et al., 1973; Pattee, 1976; Mackey et al., 1982; Squires et al., 1983; Umbarger, 1983; Wek et al., 1985; Wek and Hatfield, 1986; Lawther et al., 1987; Godon et al., 1992). In eukaryotes such as *S. cerevisiae*, genes involved in this pathway are not clustered, but are located on separate chromosomes (Mortimer et al., 1991). In higher plants, nothing is known about the chromosomal location for these genes.

In *E. coli*, expression of the acetohydroxy acid isomeroreductase gene is induced by binding of either 2-acetolactate or

Figure 4 Potential secondary structure around the site of cleavage and polyadenylation of the acetohydroxy acid isomeroreductase mRNA

The polyadenylation signal is designated by the sequence within the box.

Figure 5 Southern-blot analysis of *A. thaliana* acetohydroxy acid isomeroreductase gene

Total leaf DNA (10 µg) was digested with *PvuII*, *EcoRV* and *EcoRI*, as indicated, and probed with ³²P-labelled genomic restriction fragments corresponding to (a) the centre [*Bam*H1–*EcoRV* (1.67 kbp)] and (b) to the 3'-end [*EcoRI*–*EcoRI* (1.4 kbp)] of the gene.

2-aceto-2-hydroxybutyrate to a positive activator encoded by the gene *ilvY* (Wek and Hatfield, 1986, 1988). In fungi and plants, the regulation of acetohydroxy acid isomeroreductase may be more complex, because this enzyme is nuclear-encoded and exported either to chloroplasts (plants) or mitochondria (fungi). In *S. cerevisiae*, the biosynthesis of the branched-chain amino acids is transcriptionally regulated, in part, by *leu3*. *leu3* is a transcriptional activator of *leu1* (encoding isopropylmalate iso-

<i>A. thaliana</i>	MAAATS SIAPSLSCPS PSSSSKTLWS SKARTLALPN IGFSSSSKS	46
<i>S. oleracea</i>	MAATAATIFS LSSSSSTSA AASKALKSPK PSALNLGFLG SSSTIKACRS	50
<i>S. cerevisiae</i>	-----MLR TQARLICMS	21
<i>N. crassa</i>	MAARNCTK ALRFLARQLA	26
<i>A. thaliana</i>	LRSLTATVAG NGATGSSSLAA RMVSSSAVKA P--VSLDFET SVFKKAVL	94
<i>S. oleracea</i>	LKAARVLPSG ANGGGSALSA QMVSAFINT PSATITDFDS SVFKKAVL	100
<i>S. cerevisiae</i>	RVITAKRTFA LATRAAAYSR PAARFVKPM- ITTRGLKQIN FGGTVEIV	60
<i>N. crassa</i>	TPAVQRRTFV AAASAVRASV AVKAVAAPAR QQVRGVKTM FAGHNEIV	66
NADPH Binding site		
<i>A. thaliana</i>	AGYEYIVRG GRDLFKHLD MFKIKQIGV IGGSSQPAQ ANILDELVE	144
<i>S. oleracea</i>	SGHDEYIVRG GRNLFPILED MFKIKQIGV IGGSSQPAQ ANILDELVE	150
<i>S. cerevisiae</i>	--YE--FA DWPKEKILL MFKN-DTFAL IGGSSQYGO GINILDD	98
<i>N. crassa</i>	--HE--FA DWPKEKILL MFKN-DTIAL IGGSSQHGQ GINILDD	104
<i>A. thaliana</i>	AKSDIVKIG MKNSSSFEE ARAAGTEES GTLEDIWEIT AGSDIVLIL	194
<i>S. oleracea</i>	AKSDIVKIG MKNSSSFEE ARAAGTEES GTLEDIWEIT AGSDIVLIL	200
<i>S. cerevisiae</i>	--NGLAVIG MKNSSSFEE ARAAGTEES GTLEDIWEIT AGSDIVLIL	145
<i>N. crassa</i>	--NGLAVIG MKNSSSFEE ARAAGTEES GTLEDIWEIT AGSDIVLIL	151
<i>A. thaliana</i>	SDAQADNYE KIFSHKPI --NSILDLGH GFLGHLOSS GLDFQNIIV	240
<i>S. oleracea</i>	SDAQADNYE KIFSHKPI --NSILDLGH GFLGHLOSS GLDFQNIIV	246
<i>S. cerevisiae</i>	SDAQADNYE KIFSHKPI --NSILDLGH GFLGHLOSS GLDFQNIIV	191
<i>N. crassa</i>	SDAQADNYE KIFSHKPI --NSILDLGH GFLGHLOSS GLDFQNIIV	197
<i>A. thaliana</i>	VAVPKMGCP SVRLVYVQK EINGAGINS FAVHCVVGR ADVDRLGWSV	290
<i>S. oleracea</i>	VAVPKMGCP SVRLVYVQK EINGAGINS FAVHCVVGR ADVDRLGWSV	296
<i>S. cerevisiae</i>	IIVPKMGCP SVRLVYVQK EINGAGINS FAVHCVVGR ADVDRLGWSV	236
<i>N. crassa</i>	IIVPKMGCP SVRLVYVQK EINGAGINS FAVHCVVGR ADVDRLGWSV	242
Mg ²⁺ Binding site		
<i>A. thaliana</i>	ANGSPFITAT TEFEVNSDI FGERGLHGA VHGVECLFR RYTENGMSD	340
<i>S. oleracea</i>	ANGSPFITAT TEFEVNSDI FGERGLHGA VHGVECLFR RYTENGMSD	346
<i>S. cerevisiae</i>	ANGSPFITAT TEFEVNSDI FGERGLHGA VHGVECLFR RYTENGMSD	270
<i>N. crassa</i>	ANGSPFITAT TEFEVNSDI FGERGLHGA VHGVECLFR RYTENGMSD	276
<i>A. thaliana</i>	LAYKNTVECI TGTISRTIST QGMLAVYNSL SEEGKKDFET AYSASFYPCM	390
<i>S. oleracea</i>	LAYKNTVECI TGTISRTIST QGMLAVYNSL SEEGKKDFET AYSASFYPCM	396
<i>S. cerevisiae</i>	LAYKNTVECI TGTISRTIST QGMLAVYNSL SEEGKKDFET AYSASFYPCM	270
<i>N. crassa</i>	LAYKNTVECI TGTISRTIST QGMLAVYNSL SEEGKKDFET AYSASFYPCM	276
<i>A. thaliana</i>	EILYECYEDV QSGSEIRSVV LAGRRFYKE GLPAFFMGNI DQTRMWKVE	440
<i>S. oleracea</i>	EILYECYEDV QSGSEIRSVV LAGRRFYKE GLPAFFMGNI DQTRMWKVE	446
<i>S. cerevisiae</i>	EILYECYEDV QSGSEIRSVV LAGRRFYKE GLPAFFMGNI DQTRMWKVE	270
<i>N. crassa</i>	EILYECYEDV QSGSEIRSVV LAGRRFYKE GLPAFFMGNI DQTRMWKVE	276
<i>A. thaliana</i>	RVRKSRPAGD LGPLYPTAG VYVALMAQI EILRKNGHSK SEAFNEIVE	490
<i>S. oleracea</i>	RVRKSRPAGD LGPLYPTAG VYVALMAQI EILRKNGHSK SEAFNEIVE	496
<i>S. cerevisiae</i>	RVRKSRPAGD LGPLYPTAG VYVALMAQI EILRKNGHSK SEAFNEIVE	295
<i>N. crassa</i>	RVRKSRPAGD LGPLYPTAG VYVALMAQI EILRKNGHSK SEAFNEIVE	301
<i>A. thaliana</i>	SVSLMEFMH ARGVSEMDN CSTTARLGR KMPHFYDIL --TQALVA	537
<i>S. oleracea</i>	SVSLMEFMH ARGVSEMDN CSTTARLGR KMPHFYDIL --TQALVA	543
<i>S. cerevisiae</i>	ATCSLMELIG KVGMDMDA CSTTARLGR KMPHFYDIL --TQALVA	345
<i>N. crassa</i>	ATCSLMELIG KVGMDMDA CSTTARLGR KMPHFYDIL --TQALVA	351
<i>A. thaliana</i>	VDSQAINRD LISNFSFDP HGAIEVCAQL RPTVDISVPA DVDFVPELR	587
<i>S. oleracea</i>	VDSQAINRD LISNFSFDP HGAIEVCAQL RPTVDISVPA DVDFVPELR	593
<i>S. cerevisiae</i>	TKNGTETKRS LEYSNFSFDP HGAIEVCAQL RPTVDISVPA DVDFVPELR	395
<i>N. crassa</i>	VDSQAINRD LISNFSFDP HGAIEVCAQL RPTVDISVPA DVDFVPELR	399
<i>A. thaliana</i>	QSSN	591
<i>S. oleracea</i>	QA	595
<i>S. cerevisiae</i>	K	395
<i>N. crassa</i>	K	400

Figure 6 Comparison of the deduced aa sequences for four eukaryotic genes encoding acetohydroxy acid isomeroreductase

Boxes indicate sequence similarity between the acetohydroxy acid isomeroreductases from *A. thaliana* (the present study), *S. oleracea* (Dumas et al., 1991), *S. cerevisiae* (Petersen and Holmberg, 1986) and *N. crassa* (Sista and Bowman, 1992). Putative NADPH- and Mg²⁺-binding sites are indicated. Arrows indicate the end of the presumptive chloroplastic or mitochondrial peptide transits.

merase), *leu2* (encoding isopropylmalate dehydrogenase) and *leu4* (encoding isopropylmalate synthase), and its DNA-binding site 'CCGGN/NCCGG' can be found in the promoter region of *ilv2* (encoding acetohydroxy acid synthase) and the acetohydroxy acid isomeroreductase gene *ilv5* (Friden and Schimmel, 1988; Sze et al., 1992). In higher plants, the regulation of this pathway has not yet been described. We note, however, that the acetohydroxy acid isomeroreductase gene from *A. thaliana* contains a palindromic region GGCCT/AGGCC located just upstream of the second putative TATA box (Figure 2). This sequence is reminis-

cent of the core sequence CCGGN/NCCGG recognized by the *leu3* transcriptional factor in *S. cerevisiae*, and therefore it is tempting to postulate that this sequence may act also as a potential binding site for a transcriptional regulatory protein in *A. thaliana*. The identification and the role of transcriptional regulatory sequences for acetohydroxy acid isomeroreductase are currently underway in our laboratory.

Comparison of predicted peptide sequences between *S. cerevisiae*, *N. crassa* and *S. oleracea* and *A. thaliana* shows that (a) the higher-plant proteins were 30% longer than the fungal proteins and (b) this difference corresponds essentially to an insertion of 141 amino acid residues into the plant protein. Interestingly, all but 11 amino acids of this inserted peptide are between introns 6 and 9, meaning that exons 7, 8 and 9 almost exclusively encode these 141 amino acids. Alignment of the higher-plant and bacterial peptides did not result in any insight towards the function of these inserted 141 amino acids. These differences raise questions concerning the function of these 141 amino acids in plants. To help determine the function(s) of this insertion, we are currently characterizing an overexpressed plant acetohydroxy acid isomeroreductase in which the 141 amino acids have been deleted.

We are very grateful to Viviane Brozek for the synthesis of oligonucleotides. We thank Marie-Line Ruffet, Bernard Dumas, Alain Sailland, Michel Lebrun and Dominique Job for helpful discussions. This study has been conducted under the BIOAVENIR programme financed by Rhône-Poulenc with the contribution of the Ministère de la Recherche et de l'Espace and the Ministère de l'Industrie et du Commerce Extérieur.

REFERENCES

- Aguilar, O. M. and Grasso, D. H. (1991) J. Bacteriol. **173**, 7756-7764
- Arfin, S. M. and Umbarger, H. E. (1969) J. Biol. Chem. **244**, 1118-1127
- Aulabaugh, A. and Schloss, J. V. (1990) Biochemistry **29**, 2824-2830
- Birnboim, H. C. and Doly, J. (1979) Nucleic Acids Res. **7**, 1513-1523
- Blazey, D. L. and Burns, R. O. (1984) J. Bacteriol. **159**, 951-957
- Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349-383
- Cavener, D. R. and Ray, S. C. (1991) Nucleic Acids Res. **19**, 3185-3192
- Chunduru, S. K., Mrachko, G. T. and Calvo, K. C. (1989) Biochemistry **28**, 486-493
- Curien, G., Dumas, R. and Douce, R. (1993) Plant Mol. Biol. **21**, 717-722
- Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988) Nucleic Acids Res. **16**, 6127-6145
- Dumas, R., Joyard, J. and Douce, R. (1989) Biochem. J. **262**, 971-976
- Dumas, R., Lebrun, M. and Douce, R. (1991) Biochem. J. **277**, 469-475
- Dumas, R., Job, D. and Douce, R. (1992a) In Biosynthesis and Molecular Regulation of Amino Acids in Plants (Flores, H., Shannon, J. and Singh, B., eds.), pp. 315-317, American Society of Plant Physiologists, Rockville, MD
- Dumas, R., Job, D., Ortholand, J.-Y., Emeric, G., Greiner, A. and Douce, R. (1992b) Biochem. J. **288**, 865-874
- Friden, P. and Schimmel, P. (1988) Mol. Cell. Biol. **8**, 2690-2697
- Godon, J. J., Chopin, M.-C. and Ehrlich, S. D. (1992) J. Bacteriol. **174**, 6580-6589
- Hoffer, J. G., Decedue, C. J., Luginbuhl, G. H., Reynolds, J. A. and Burns, R. O. (1975) J. Biol. Chem. **250**, 877-882
- Kakar, S. N. and Wagner, R. P. (1964) Genetics **49**, 213-222
- Lawther, R. P., Wek, R. C., Lopes, J. M., Pereira, R., Taillon, B. E. and Hatfield, G. W. (1987) Nucleic Acids Res. **15**, 2137-2155
- Mackey, C. J., Warburg, R. J., Halvorson, H. O. and Zahler, S. A. (1984) Gene **32**, 49-56
- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Mortimer, R. K., Contopoulos, C. R. and King, J. S. (1991) In The Molecular and Cellular Biology of the yeast *Saccharomyces* (Broach, J. R., Pringle, J. R. and Jones, E. W., eds.), pp. 737-812, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Pattee, P. A. (1976) J. Bacteriol. **127**, 1167-1172
- Petersen, J. G. L. and Holmberg, S. (1986) Nucleic Acids Res. **14**, 9631-9651
- Rieble, S. and Beale, S. I. (1992) J. Bacteriol. **174**, 7910-7918
- Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467
- Schulz, A., Spönmann, P., Köcher, H. and Wengenmayer, F. (1988) FEBS Lett. **238**, 375-378
- Shematek, E. M., Arfin, S. M. and Diven, W. F. (1973) Arch. Biochem. Biophys. **158**, 132-138

- Sista, H. and Bowman, B. (1992) *Gene* **120**, 115–118
- Somers, J. M., Amzallag, A. and Middleton, R. B. (1973) *J. Bacteriol.* **113**, 1268–1272
- Squires, C. H., DeFelice, M., Devereux, J. and Calvo, J. M. (1983) *Nucleic Acids Res.* **11**, 5299–5313
- Sze, J.-Y., Woontner, M., Jaehning, J.-A. and Kohlhaw, G. B. (1992) *Science* **258**, 1143–1145
- Umbarger, H. E. (1983) in *Amino acids: Biosynthesis and Genetic Regulation* (Herrman, K. M. and Somerville, R. L., eds.), pp. 245–266, Addison–Wesley, Reading, MA
- Wagner, R. P., Berquist, A., Barbee, T. and Kiritani, K. (1964) *Genetics* **49**, 865–882
- Wek, R. C. and Hatfield, G. W. (1986) *J. Biol. Chem.* **261**, 2441–2450
- Wek, R. C. and Hatfield, G. W. (1988) *J. Mol. Biol.* **203**, 643–663
- Wek, R. C., Hauser, C. A. and Hatfield, G. W. (1985) *Nucleic Acids Res.* **13**, 3995–4010

Received 9 March 1993/13 May 1993; accepted 18 May 1993