Molecular and genetic characterization of a biotin biosynthetic gene in Arabidopsis encoding both 7,8-diaminopelargonic acid aminotransferase and dethiobiotin synthetase

Elve Chen

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Biochemistry Commons

Recommended Citation

Chen, Elve, "Molecular and genetic characterization of a biotin biosynthetic gene in Arabidopsis encoding both 7,8-diaminopelargonic acid aminotransferase and dethiobiotin synthetase " (2006). Retrospective Theses and Dissertations. 902.

https://lib.dr.iastate.edu/rtd/902
Molecular and genetic characterization of a biotin biosynthetic gene in
Arabidopsis encoding both 7,8-diaminopelargonic acid aminotransferase
and dethiobiotin synthetase

by

Elve Chen

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Biochemistry

Program of Study Committee:
Basil J. Nikolau, Major Professor
Thomas A. Bobik
Eve S. Wurtele

Iowa State University
Ames, Iowa
2006
Copyright © Elve Chen, 2006. All rights reserved.
# TABLE OF CONTENTS

CHAPTER 1. GENERAL INTRODUCTION ................................................................. 1
  Biotin Biosynthesis ................................................................................. 2
  Rationale and Significance ................................................................. 4
  Thesis Organization ............................................................................. 5
  References ............................................................................................. 5

CHAPTER 2. MOLECULAR AND GENETIC CHARACTERIZATION INDICATES
  THE ARABIDOPSIS BIO1 LOCUS IS CHIMERIC THAT CODES FOR BOTH 7,8-
  DIAMINOPELARGONIC ACID AMINOTRANSFERASE AND DETHIOBIOTIN
  SYNTHETASE IN BIOTIN BIOSYNTHESIS .............................................. 10
  ABSTRACT ............................................................................................. 10
  INTRODUCTION ..................................................................................... 10
  RESULTS ................................................................................................ 12
    Molecular Characterization of bio1 Locus of Arabidopsis .................. 12
    Isolation and Cloning of Full Length cDNA of BIO1 ......................... 13
    Accumulation of the BIO3/BIO1 mRNA ............................................ 15
    Molecular Characterization of the bio1-1 Mutant Lesion ................. 15
    Characterization of SALK_023399 Mutant of Arabidopsis ............... 16
    SALK_023399 and bio1-1 are Allelic ................................................ 17
    Functional Complementation of *E. coli* bioD* and bioA* Mutants .... 18
    Overexpression and Purification of the BIO3 Protein ....................... 20
    Western Blot Analysis of Arabidopsis BIO3/BIO1 Protein ............... 20
  DISCUSSION ........................................................................................... 21
  MATERIALS AND METHODS ................................................................. 25
    Plant Materials and Growth Conditions .......................................... 25
CHAPTER 1. GENERAL INTRODUCTION

Biotin is a water-soluble vitamin known as vitamin $B_8$ or vitamin H. It was first discovered in 1901 as a growth factor in yeast (Wildiers, 1901), and then isolated from egg yolk by Kögl and Tönnis in 1936. The structure of biotin was later established in early 1940s by du Vigneaud et al. (1942).

It is an essential enzyme cofactor that facilitates the transfer of carboxyl group during carboxylation, decarboxylation, and transcarboxylation reactions. Biotin-containing enzymes are therefore involved in many important metabolic processes, such as lipogenesis, gluconeogenesis, and amino acid metabolism. Biotin-dependent carboxylases include heteromeric and homomeric acetyl-CoA carboxylase (ACCase), methylcrotonoyl-CoA carboxylase (MCCase), propionyl-CoA carboxylase (PCCase), pyruvate carboxylase (PC), and geranoyl-CoA carboxylase (GCCase). Although biotin is essential for life in all organisms, not all organisms can synthesize biotin. Most bacteria, all plants, and some fungi can synthesize biotin. Humans and animals must acquire biotin from their diets.

Biotin is also present on a specific seed biotin protein (SBP) in plants (Duval 1994). SBP is free of any known biotin-dependent carboxylase activity. It appears to serve as a sink for biotin during late stages of embryo development and is rapidly degraded during germination.

Biotin may also act as a signaling molecule. It has been shown to regulate the expression of MCCase genes (Che et al., 2003). Both subunits of MCCase hyperaccumulate in response to $bio1$-associated depletion of biotin (biotin biosynthesis is blocked in the $bio1$ mutant). The accumulation is the result of increased mRNAs translation and/or decreased protein turnover. Biotin is also
shown to be required for inducing MCCase subunits transcription under CO₂ or light deprivation.

**Biotin Biosynthesis**

Mutations in biotin biosynthesis are lethal in bacteria (Eisenberg, 1987) and plants (Shellhammer and Meinke, 1990), which demonstrates that biotin is essential for viability. The genetics and biochemistry of biotin biosynthesis and its regulation have been extensively studied in bacteria such as *E. coli* and *B. subtilis* (Pai and Lichstein, 1965; Rolfe and Eisenberg, 1968; Eisenberg, 1973; Izumi et al., 1981; Gloeckler et al., 1990; Bowel et al., 1996). The biotin biosynthesis is a four reaction pathway. The first reaction involves the decarboxylative condensation of pimeloyl-CoA and L-alanine catalyzed by 7-keto-9-aminopelargonic acid (KAPA) synthase (encoded by *bioF* gene) to yield KAPA. The next reaction is catalyzed by 7,8-diaminopelargonic acid (DAPA) aminotransferase (encoded by *bioA* gene) where an amino group from S-adenosyl-L-methionine (SAM) is transferred to KAPA. The resulting product DAPA is then carboxylated with the use of bicarbonate to form dethiobiotin. This carboxylation step is catalyzed by dethiobiotin synthetase (encoded by *bioD* gene). The last reaction is catalyzed by biotin synthase (encoded by *bioB* gene) where sulfur is donated from SAM to dethiobiotin to generate biotin (Figure 1). Plant homologs to *E. coli* *bioA* and *bioB* were identified later through the analysis of biotin auxotrophic mutants, *bio1* and *bio2*, in Arabidopsis (Shellhammer and Meinke, 1990; Patton et al., 1998).

The initial reactions of biotin biosynthesis, the generation of the precursor pimeloyl-CoA, are still not well understood. In gram-positive bacteria such as *B. subtilis*, pimelic acid is activated by coenzyme A (CoA-SH) to form pimeloyl-CoA
with the use of pimeloyl-CoA synthase (encoded by \textit{bioW}). The source of pimelic acid is not clear. Stok et al. (2000) has proposed the formation of pimelic acid from $\omega$-oxidation of odd-chain fatty acid substrate with the use of cytochrome P450 (encoded by \textit{biol}). Gram-negative bacteria such as \textit{E. coli} lacks pimeloyl-CoA synthase and cannot use pimelic acid as precursor for the biotin biosynthesis (Gloeckler et al., 1990; Ifuku et al., 1994). \textit{E. coli} contains two genes, \textit{bioC} and \textit{bioH} which appear to be involved in early steps prior to pimeloyl-CoA synthesis, but their exact roles are unknown (Eisenberg, 1987). The metabolic pathway to synthesize odd chain fatty acid from propionyl-CoA or the $\beta$-oxidation of odd-chain fatty acid were also shown not to be the precursor of pimeloyl-CoA in $^{13}$C-labeling experiments. L-alanine and acetyl-CoA were proposed to be the possible precursors for pimeloyl-CoA synthesis. Other proposed precursor of pimeloyl-CoA includes the condensation of three malonyl-CoA. (Lezius et al., 1963).

In \textit{E. coli}, biotin biosynthesis is regulated by the \textit{birA} gene product. BirA is a bifunctional protein that possesses both regulatory and enzymatic activities (Campbell et al., 1972; Pai, 1972; Campbell et al., 1980; Barker and Campbell, 1981a; Barker and Campbell, 1981b; Eisenberg, 1982). BirA can activate biotin in the presence of ATP to form biotinyl-5'AMP, a co-repressor of biotin regulation, which remains tightly bound to BirA. The complex can either bind to the operator, bioO, to repress the transcriptional initiation of the biotin biosynthetic operon, or can function as a ligase to mediate the biotinylation of the biotin-dependent carboxylase, ACCase. The BirA-mediated regulation of biotin biosynthesis is conserved in \textit{B. subtilis} (Bower et al., 1995), and is also predicted to be conserved in eubacteria and archea by bioinformatics analysis (Rodionov et al., 2002).
Rationale and Significance

The first information about biotin biosynthetic pathway in plants came from the analysis of *bio1*, a biotin auxotrophic mutant of Arabidopsis (Schneider et al., 1989). Plants homozygous for this allele stop growing after 7 to 10 days, unless exogenous biotin, dethiobiotin or DAPA are supplied to the plant (Shellhammer and Meinke, 1990). It was later shown that the intermediates in biotin biosynthetic pathway between plants and bacteria are conserved (Baldet et al., 1993). Biotin plays crucial roles in various metabolic processes. Plants are the major source of biotin in the biosphere but its biosynthetic pathway is not completely characterized. The genes encoding for the first and last enzyme in this pathway have been identified. The latter enzyme, biotin synthase has been shown to be a mitochondrial enzyme, establishing that biotin biosynthesis is a mitochondrial process (Baldet et al., 1997). Biotin synthase is encoded by *BIO2*. The *BIO2* cDNA encoding product was shown to be capable of complementing the *E. coli bioB* mutant, a Arabidopsis homolog (Weaver et al., 1996). *BIO2* was the first plant biotin biosynthetic gene to be characterized. A second biotin auxotroph, *bio2* was isolated more recently (Patton et al., 1998). Arrested *bio2* embryos were defective in the final step of biotin biosynthesis. The first enzyme in the pathway KAPA synthase, *AtbioF*, was recently identified as a cytosolic enzyme in Arabidopsis (Pinon et al., 2005). The *AtbioF* cDNA is capable of complementing *E. coli bioF* mutant. Therefore, plant biotin biosynthesis separates between cytosol and mitochondria, revealing its complicated compartmentation. Identification and cloning of the additional genes in the biotin biosynthetic pathway will be required for understanding the regulation of biotin biosynthesis in plants.
Thesis Organization

This thesis consists of 3 chapters. The first chapter provides a general introduction of biotin biosynthesis in bacteria and plants. The second chapter is a manuscript to be submitted to the journal The Plant Cell. The manuscript describes the genetic and molecular analysis of the BIO1 locus in Arabidopsis. The last chapter summarizes the important findings of this work and identifies potential future research.

References


Wildiers, E. (1901). Nouvelle substance indispensable au developpement de la levure. La Cellule 18, 313-331.
CHAPTER 2. MOLECULAR AND GENETIC CHARACTERIZATION

INDICATES THE ARABIDOPSIS BIO1 LOCUS IS CHIMERIC THAT CODES FOR BOTH 7,8-DIAMINOPELARGONIC ACID AMINOTRANSFERASE AND DETHIOBIOTIN SYNTHETASE IN BIOTIN BIOSYNTHESIS

A manuscript to be submitted to The Plant Cell

Elve Chen, Libuse Brachova and Basil J. Nikolau

ABSTRACT

Biotin is an essential enzyme cofactor required for different metabolic processes. Plants are a major source of biotin, however, the genes and the protein products of biotin biosynthetic pathway are not fully characterized. We have molecularly characterized the genetically defined Arabidopsis bio1 locus as encoding 7,8-diaminopelargonic acid (DAPA) aminotransferase. Molecular and genetic analysis revealed that the full length BIO1 cDNA is capable of producing a fusion protein that catalyzes not only DAPA aminotransferase, but also dethiobiotin synthetase (BIOS) which is another enzymatic step in the biotin biosynthetic pathway. We refer to this fusion-gene as BIO3/BIO1. This bifunctional enzyme was demonstrated by functional complementation of E. coli bioA' and bioD' mutants deficient in DAPA aminotransferase and dethiobiotin synthetase reactions, respectively. This gene organization is unique in biology and establishes that the BIO3/BIO1 gene has dual functions.

INTRODUCTION

The pathway of biotin biosynthesis is conserved between plants and bacteria (Baldet et al., 1993). Initial decarboxylative condensation reaction between pimeloyl-CoA and L-alanine is catalyzed by 7-keto-9-aminopelargonic acid
(KAPA) synthase. The KAPA is transaminated by 7,8-diaminopelargonic acid (DAPA) aminotransferase. The next carboxylation reaction is catalyzed by dethiobiotin synthetase. The final reaction is catalyzed by biotin synthase to produce biotin (Figure 1).

The genes encoding the enzymes in this pathway are well characterized in bacteria, such as *E. coli* and *B. sphearicus* (Pai and Lichstein, 1965; Rolfe and Eisenberg, 1968; Izumi et al., 1981; Gloeckler et al., 1990; Bowel et al., 1996). KAPA synthase is encoded by *bioF*, DAPA aminotransferase is encoded by *bioA*, dethiobiotin synthetase is encoded by *bioD*, and biotin synthase is encoded by *bioB*.

The genes encoding the enzymes in the biotin biosynthetic pathway in plants, however, are not completely characterized. The first enzyme in this pathway is encoded by *AtbioF* which is localized in the cytosol (Pinon et al., 2005) and the last enzyme of the pathway is encoded by *BIO2* which is localized in the mitochondria (Weaver et al., 1996, Patton et al., 1998). The genes encoding the enzymes for the intermediate steps of the pathway are not identified.

Prior to identification of *AtbioF* and *BIO2* genes, the analysis of *bio1* (a biotin auxotrophic mutant of Arabidopsis) provided the initial information about biotin biosynthetic pathway in plant (Schneider et al., 1989). Plants homozygous for the *bio1-1* allele stop growing after 7 to 10 days, unless exogenous biotin, dethiobiotin or DAPA are supplied to these plants (Shellhammer and Meinke, 1990). *E. coli bioA* gene was able to rescue the growth of *bio1-1*.

Here, we report the identification and characterization of genes for the rest of the biotin biosynthetic pathway. We molecularly investigated the *bio1-1* allele and identified a full length cDNA encoding DAPA aminotransferase (also known as BIO1) using an expressed sequence tag. The recombinant protein was
overexpressed in E. coli and purified. We established that the full length cDNA is
a fusion between a gene coding for DAPA aminotransferase and an upstream
gene coding for dethiobiotin synthetase (BIO3) that we named BIO3/BIO1. This
gene organization is unique in biology and indicates that the BIO3/BIO1 gene
codes for a bifunctional protein.

RESULTS

Molecular Characterization of bio1 Locus of Arabidopsis

The bio1 auxotroph was originally isolated as an embryo-lethal mutant following
EMS seed mutagenesis (Meinke 1985). This locus was mapped to position 74
cM on chromosome 5 (Patton et al., 1991; Franzmann et al., 1995),
corresponding to coordinates between 23335734 bp and 23338765 bp of
chromosome 5 (TAIR). An Arabidopsis homolog to E. coli bioA has been
identified by searching the Arabidopsis genome database. BLASTP analysis of
the Arabidopsis genome sequence identified a series of homologs of which the
third and forth best matches to Arabidopsis genes from chromosome 5. The third
best match corresponds to gene At5g46180 which has been shown to encode
ornithine aminotransferase (Roosens et al., 1998). The forth best match
corresponds to gene At5g57590 which is predicted to encode an
aminotransferase class III family protein. At5g57590 was tentatively identified as
the BIO1 locus which shares 24% identity (37% similarity) to the E. coli BioA
protein sequence (TAIR BLASTP).

Genomic DNA was isolated from homozygous bio1-1 mutant plants enabling
PCR amplification and sequencing of At5g57590, which revealed the presence of
a point mutation “A” relative to the Arabidopsis genome database “G” (Figures 2A
and 2B). To ensure the point mutation is of the bio1-1 mutant plants, genomic
DNA was also isolated from Arabidopsis wild-type (WT) Columbia plants. PCR amplification and sequencing of this locus isolated from the WT plants agreed with the sequence in the Arabidopsis genome database.

**Isolation and Cloning of Full Length cDNA of BIO1**

The above described analyses established that the bio1-1 lesion, a point mutation, lies in the gene identified as At5g57590. To further characterize this gene, a cDNA corresponding to At5g57590 was identified in the Arabidopsis EST collection, clone RZ128g09R (Cunxi Wang, Iowa State University, unpublished data). 5'-RACE and 3'-RACE PCR products were sequenced to define the full length cDNA of At5g57590. The 5'-untranslated region (5'-UTR) contains 122 nucleotides, the predicted coding region contains 2,502 nucleotides, and the 3'-UTR contains 121 nucleotides. The cDNA was aligned with the Arabidopsis genome sequence to define exons and introns. These analyses revealed that the 5'-end of BIO1 was chimeric with the annotated gene At5g57600, which is immediately upstream of At5g57590 (Figure 3A). Sequence analysis using BLASTP (NCBI) showed that Arabidopsis At5g57600 shares 26% sequence identity (41% similarity) with the *E. coli* bioD gene. Subsequent RT-PCR analysis using a primer in At5g57600 and a primer in At5g57590 confirmed the fusion of both of these genes (Figure 3B). This RT-PCR product was sequenced and its result agreed with that of full length cDNA of At5g57590. Next, RT-PCR using primers located at the annotated 3'-UTR of At5g57600 were carried out. As expected no RT-PCR products were detected (Figure 3C). We define this At5g57600-At5g57590 cDNA as BIO3/BIO1.

In these analyses we found two different BIO3/BIO1 transcripts which appear to be produced through differential splicing. We named these transcripts splice
variant (-10) and splice variant (+10) (Figures 4A and 4B). Splice variant (-10) contains a predicted coding region of 2,502 nucleotides. This sequence is predicted to encode a 833-amino acid protein for a predicted molecular mass of 91.9 kD. The possible open reading frames (ORFs) of splice variant (-10) are shown in Table 1. Amino acid sequence analysis using BLASTP (NCBI) showed that the first 385 amino acids of the 833-amino acid sequence share 26% sequence identity (41% similarity) with the *E. coli* BioD protein (GenBank accession no. AAA23518) and the last 448 amino acids share 26% sequence identity (41% similarity) with the *E. coli* BioA protein (GenBank accession no. AAA23514) (Figures 4C and 4D). This ORF is therefore capable of producing a fusion protein that may catalyze two different reactions in the biotin biosynthetic pathway (BIO3 would catalyze the dethiobiotin synthetase reaction, and BIO1 would catalyze the DAPA aminotransferase reaction. Splice variant (+10) contains a predicted coding region of 2,512 nucleotides that does not have a contiguous ORF. The 10 extra nucleotides locate between 1,208 – 1,217-nucleotide, which disrupts the ORF and introduces a premature stop codon (Figure 4E). This sequence is predicted to encode a 405-amino acid protein with a predicted molecular mass of 44.5 kD. The possible ORFs of splice variant (+10) are shown in Table 2. The 405-amino acid sequence shares 26% sequence identity (41% similarity) with the *E. coli* BioD protein, which is the same as the N-terminal portion of splice variant (-10) (Figure 4C). Sequence analysis with TargetP (Nielsen et al., 1997; Emanuelsson et al., 2000) predicts that both proteins from splice variant (-10) and splice variant (+10) are mitochondrial proteins with predicted 23-amino acid N-terminal targeting peptide.
Accumulation of the *BIO3/BIO1* mRNA

To investigate the accumulation of mRNA of splice variant (+10) and splice variant (-10), semi-quantitative RT-PCR was carried out using isolated WT Arabidopsis (Col-0) organs (Figures 5A and 5B). The tissues examined include leaves and roots from 21-day old plants and buds, flowers, siliques, and rosette leaves from 37-day old plants, and a pool of fluorescent stem and leaves from 37-day old plants. In these analyses, RT-PCR was conducted with two primers that spanned the 10-nucleotides polymorphism, which distinguish the (+10) and (-10) splice variants. Hence, the two respective RT-PCR product derived from these variants are 221 bp and 211 bp in length, and these were separated by electrophoresis and the band intensities of RT-PCR products were quantified (Figure 5C and Table 3). The splice variant (+10) transcript is more abundant than the splice variant (-10) transcript in all tissues. The (+10) splice variant was most abundant in flowers, while the (-10) splice variant was most abundant in roots.

Molecular Characterization of the *bio1-1* Mutant Lesion

After defining the exons and introns of the *BIO3/BIO1* gene, we realized that the *bio1-1* point mutation occurred at the last intron of this gene (Figures 6A and 6B). RT-PCR was conducted using primers in exons 13 and 14. The resulting RT-PCR products were analyzed by electrophoresis and were sequenced. Sequencing result revealed the mutation at the 3' splice site caused the terminal intron of *BIO3/BIO1* to be included in the mature mRNA (Figures 6C and 6D).
Characterization of SALK_023399 Mutant of Arabidopsis

SALK_023399 is a T-DNA insertion allele that disrupts the second exon of At5g57600 (Figure 7A). A genetic stock carrying this allele was obtained from the Arabidopsis Biological Resource Center. Seeds were germinated on MS media plates with biotin and the resulting plants were watered with 1 mM biotin throughout their lifetime. PCR was used to genotype individual progeny from this stock. Twenty-three out of 33 plants were WT and did not carry the T-DNA insertion. Ten out of 33 plants were heterozygous for the T-DNA mutant allele. Plants homozygous for the SALK_023399 allele were not recovered in this T3 generation. These results indicate that the SALK_023399 allele is recessive lethal allele, and that exogenous biotin does not rescue these plants. Although biotin-rescue was attempted in all subsequent experiments, we were not able to demonstrate that plants homozygous for this mutant allele could be rescued with exogenous supply of biotin.

Seeds from a single heterozygous mutant in the T3 generation were collected. The seeds were kept at 4°C for 3 days to break dormancy, then planted on MS media plates and then transferred to soil. Thirty one seedlings germinated and were genotyped. Sixteen out of 31 plants were WT and 15 out of 31 plants were heterozygous for the T-DNA mutant allele. No homozygous mutants were recovered in this T4 generation (Table 4).

We noticed that not all seeds from a heterozygous parent germinated, and the numbers above appear not to follow the normal Mendelian ratio of 1:2:1 (WT:heterozygous:homozygous). Therefore, more progeny seeds from a selfed heterozygous parent were analyzed. These plants were sown on MS media plates and their growth phenotypes were observed after 18 days and subsequently each plant was genotyped. In the plates without the addition of
biotin, 51 out of 85 sown seeds had a healthy normal phenotype, another 19 had germinated but the seedling demonstrated an altered, dying phenotype (the seedlings were small, with only cotyledons having emerged, and these were becoming chlorotic), and finally 15 of the sown seeds did not germinate. In the plates where biotin was added, 52 out of 94 sown seeds germinated and had a healthy phenotype, 19 had the dying phenotype, and 23 seeds did not germinate. The germinated seedlings demonstrating the normal and dying phenotypes from plates with biotin were genotyped. Seventeen out of 52 healthy seedlings were WT, and 35 were heterozygous, and no homozygous mutant were found. Eight out of 19 dying seedlings were WT, 11 were heterozygous, and again, no homozygous mutants were found (Table 5). Chi-squared test was conducted to analyze the segregation ratio from all heterozygous plants that were grown with the supply of biotin. The observed deviation from the 1:2:1 segregation ratio is statistically significant ($\chi^2 = 32.5$, p value<0.0001) and the null hypothesis can be rejected. This indicates that the homozygous mutant of SALK_023399 T-DNA allele is recessive lethal. We therefore could only use heterozygous SALK_023399 for subsequent analyses.

**SALK_023399 and bio1-1 are Allelic**

Molecular evidence showed that At5g57600 and At5g57590 are fused genes. We therefore test if SALK_023399 and bio1-1 are allelic. Plants heterozygous for the SALK_023399 allele (male) were crossed with bio1-1 homozygous mutant plants (female) (Figure 7B). The plants were supplied with 1 mM biotin twice a week throughout their life time. Eighty progeny seeds were collected from the cross and planted on MS media containing biotin. Thirty-eight out of 80 seeds germinated and were genotyped.
To determine if the *bio1-1* point mutation was present in the progeny, PCR was carried out to amplify the genomic region carrying the *bio1-1* lesion. The difference between homozygous and heterozygous *bio1-1* cannot be distinguished from gel electrophoresis of the PCR product. We therefore sequenced 23 out of the 38 PCR products. The sequence obtained from heterozygous *bio1-1* gave a “half peak” at the position of the point mutation “A” and a half peak of its corresponding WT base “G” (Figure 7C). All of the sequenced PCR products were heterozygous *bio1-1*. The result showed that the recovered plants were not selfs, which would have *bio1-1/bio1-1* genotype. To determine if there was SALK_023399/bio1-1 progeny present, PCR with primers specific to the T-DNA insertion were used. PCR products were visualized using gel electrophoresis. Only the WT allele was recovered from the SALK_023399 stock. None of the 38 germinated seedlings from the allelic cross had the T-DNA insertion allele. The overall result showed that all germinated offsprings from the allelic cross have the genotype of +/-bio1-1, and because SALK_023399/bio1-1 genotype was not recovered, it must be lethal. Because SALK_023399 does not complement *bio1-1*, SALK_023399 and *bio1-1* are allelic. Therefore, genetically At5g57600 and At5g57590 are one and the same locus.

**Functional Complementation of *E. coli* bioD' and bioA' Mutants**

The cDNA sequence of *BIO3/BIO1* (splice variant (-10)) and the allelic cross mentioned above indicates that the *BIO3/BIO1* encoded protein may be chimeric. To test the functionality of the *BIO3/BIO1* chimeric protein, two different Arabidopsis *BIO3/BIO1* cDNA (splice variant (-10) and splice variant (+10)) were cloned into the *E. coli* expression vector pDEST17 that has a N-terminal 6xHis tag and the proteins were expressed in BL21-AI. Splice variant (-10) was
expected to produce a 91.9 kD protein with the addition of 2.6 kD from the 6xHis tag. Spice variant (+10) was expected to produce a 44.5 kD protein with the addition of 2.6 kD from the 6xHis tag.

The constructs were used to transform *E. coli* bioD' and bioA' mutants (strain ID: JW0761, JW0757), which lack dethiobiotin synthetase activity and DAPA aminotransferase activity, respectively. The mutants were disrupted in *bioD* and *bioA* genes respectively by replacement with a kanamycin resistant gene. PCR amplification and sequencing confirmed the replacement of the corresponding genes with the kanamycin resistant gene. The mutants did not carry T7 RNA polymerase gene, therefore they were made into λ(DE3) lysogens prior to transformation. The mutants could grow in M9 minimal medium with or without the addition of kanamycin (Figures 8A and 8B). To absolutely deplete the biotin from the medium, avidin was added to the M9 minimal medium. The mutants could not grow in biotin free medium (Figure 8C). The mutants could grow when 0.1 mM biotin was added onto biotin free medium (Figure 8D). A search in NCBI database of *E. coli* showed a putative dethiobiotin synthetase, ynfK (Gene ID: 944927, JW5264), which is not located in the biotin biosynthetic operon. Because the ynfK' mutant did not require biotin for growth, we did not use this mutant for subsequent analysis.

We transformed 3 constructs (pDEST17-splice variant (-10), pDEST17-splice variant (+10), and WT pDEST17) into *E. coli* bioD', bioA' mutants and WT strain BW25113, respectively. The pDEST17-splice variant (-10) construct contained the full length Arabidopsis BIO3/BIO1 cDNA. The pDEST17-splice variant (+10) construct expressed a truncated version of Arabidopsis BIO3/BIO1 because of its premature stop codon. We also included a construct that contain an empty pDEST17 vector as a negative control. We maintained the transformed *E. coli*
strains on LB plates. The transformed *E. coli* were then streaked onto M9 with avidin plates to determine their biotin dependency. The pDEST17-splice variant (-10) construct could rescue the *E. coli* bioD' and bioA' mutants in biotin-free medium, but they grew slightly slower than the WT strain. The pDEST17-splice variant (+10) construct could only rescue the *E. coli* bioD' mutant slightly in biotin-free medium (Figure 9). The pDEST17-splice variant (+10) construct could not rescue the *E. coli* bioA' mutant in biotin-free medium (Figure 10). The mutants transformed with the empty pDEST17 vector could not grow (Figures 9 and 10).

**Overexpression and Purification of the BIO3 Protein**

To further characterize the plant dethiobiotin synthetase and DAPA aminotransferase reactions, antibody specific to BIO3 was made. The BIO3-specific antibody contains an N-terminal 6xHis tag from pDEST17 vector. The recombinant BIO3 protein was overexpressed in *E. coli* strain BL21-AI and purified by nickel affinity column. Its size is as expected, 53 kD (Figures 11A and 11B). The BIO3-specific antibody recognized the proteins expressed from both the pDEST17-splice variant (+10) construct and the pDEST17-splice variant (-10) construct with sizes of 48 kD and 95 kD, respectively (Figures 12B and 12C).

**Western Blot Analysis of Arabidopsis BIO3/BIO1 Protein**

The BIO3-specific antibody was used to detect the protein extracts of WT Arabidopsis and bio1-1 homozygous mutant. Initial western blot analyses were done with protein extracts prepared from rosette leaves because large amount of leaves tissues can be easily obtained. The BIO3-specific antibody identified a 47.8 ± 2.4 kD (average of 3 determinations ± standard deviation) from 47-d leaves. This observation agreed with predicted MW of 44.5 kD from translation of the splice variant (+10). The observed band intensity was at least 2-fold stronger
in WT extracts than that from bio1-1 mutant extracts. The point mutation of bio1-1 mutant probably affected its protein translation (Figure 12A). The BIO3-specific antibody also recognized the protein made from splice variant (-10) with a calculated MW of 91.9 kD. However, no such band was observed in rosette leaves extracts. Analysis of mRNA accumulation of splice variants, described earlier, showed that in rosette leaves the splice variant (+10) transcript was more than 4-fold higher than that of splice variant (-10). Such difference might hinder the detection of splice variant (-10) protein. The pooled influorescent stem and leaves sample expressed the highest relative amounts of splice variant (-10) to (+10) (Table 3). To investigate if BIO3-specific antibody was able to recognize the BIO3/BIO1 protein translated from splice variant (-10), protein extracts from pooled influorescent stem and leaves were used for western blot analyses.

The BIO3-specific antibody was able to recognize both proteins from splice variant (-10) and splice variant (+10) in the pooled influorescent stem and leaves from WT, with MW of 92 and 45 kD respectively. The 92 kD band is very weak comparing to the 45 kD band. However, in this organ sample from bio1-1 homozygous mutant, the BIO3-specific antibody could only detect the 45 kD band but not the 92 kD band. This indicates the full length protein from splice variant (-10) in bio1-1 is protein null. The observed 45 kD band from WT had stronger signal than that of bio1-1 mutant in these extracts (Figures 12B and 12C). This reduction in abundance was also observed in extracts from rosette leaves (Figure 12A).

**DISCUSSION**

In this study, we have cloned the BIO3/BIO1 cDNA coding for both dethiobiotin synthetase (BI03) and DAPA aminotransferase (BI01) which is involved in 2
consecutive enzymatic steps in the biotin biosynthetic pathway of Arabidopsis. Our experimental results indicate the gene coding for dethiobiotin synthetase function consists not only of the annotated At5g57600 locus, but is also part of a bifunctional protein that has both dethiobiotin synthetase and DAPA aminotransferase functions. This bifunctional protein is produced by a locus that is a chimera of annotated genes At5g57600 and At5g57590. To complicate this story further, we identified 2 BIO3/BIO1 transcripts which appear to be produced through alternative splicing. The splice variant (-10) is chimeric and has the capacity to produce a bifunctional protein encoded by genes annotated as At5g57600 and At5g57590. Splice variant (+10) encodes a protein that primarily originates from At5g57600 and is monofunctional (supporting dethiobiotin synthetase activity only).

This conclusion was reached both from amino acid sequence deductions and from the functional complementation of *E. coli* mutants. We were able to show that the plant protein from splice variant (-10) could genetically complement both *E. coli* bioD<sup>+</sup> mutant which is defective in dethiobiotin synthetase and bioA<sup>-</sup> mutant which is defective in DAPA aminotransferase. The splice variant (+10), on the other hand, could only genetically complement the bioD<sup>+</sup> mutant but not the bioA<sup>-</sup> mutant.

Our findings demonstrated the BIO3/BIO1 cDNA from splice variant (-10) is genetically bifunctional that codes for a bifunctional enzyme that involves in 2 distinct reactions: carboxylation and transamination. Such gene organization is rare in biology. For example, Fujimori and Ohta (1998) isolated an Arabidopsis cDNA, At-IE, that encodes for a bifunctional protein that catalyze 2 enzymatic steps of histidine biosynthetic pathway (phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase). Genetic and
biochemical evidences supported their hypothesis of the bifunctional protein. However, the bifunctional enzyme performs similar reactions in catalyzing the addition of water. Stover et al. (2005) discovered a fusion gene, *FSF1*, that encodes a putative fusion protein which is involved in 2 sequential enzymatic reactions in the formaldehyde detoxification pathway in ciliate *Oxytricha trifallax* (of protist family). Experimental evidence is yet required to confirm the bifunction of the putative fusion protein. Gene fusion event was also suggested in a bifunctional glutamyl- and prolyl-tRNA synthetases (GluProRS) in several organisms from the eukaryotic phylum of coelomate metazoans (Berthonneau and Mirande, 2000). A cDNA was isolated for the linker domain of GluProRS and examined by phylogenetic analysis. The analysis suggested that the bifunctional protein was a result of the fusion of GluRS and ProRS genes which are distinct in bacteria and archaea. However, no direct experimental evidence has yet shown the enzymatic functions of the cDNA isolated.

The *BOI3/BOI1* cDNA isolated here is genetically bifunctional and may have the ability to catalyze distinct reactions. The dethiobiotin synthetase and DAPA aminotransferase genes are adjacent in the genomes of diverse bacterial species (Gloeckler et al., 1990; Perkins et al., 1996; Rodionov 2002) or separated by only a few genes (Eisenberg, 1987). Physical linkage of these related proteins represent the evolutionarily linkage of prokaryotes and eukaryotes. Our lab is currently using biochemical approach to analyze the putative bifunctional enzyme.

The mRNA accumulation of splice variant (-10) in all the tissues sampled is lower than that of splice variant (+10). Because biotin biosynthesis is a linear pathway, it will be interesting to ascertain if the accumulation of both transcripts are coordinated with accumulation of mRNAs coding for other enzymes of the pathway. Even though the expression ratio of splice variant (-10) to splice variant
(+10) was close to the ratio of 1 to 1 in the pooled influorescent stem and leaves from 37-day old plants, the western blot result showed that the 92 kD BIO3/BIO1 protein from splice variant (-10) accumulated to lower levels than the 45 kD BIO3 protein from splice variant (+10). Protein regulation may occur and result in less protein expression from the splice variant (-10).

Plants that are homozygous for the bio1-1 allele carrying a point mutation, accumulate an intron to be included in the BIO3/BIO1 mRNA transcript. The BIO3/BIO1 protein from the mutant has a predicted molecular mass of 85 kD (6.6 kD smaller than that of WT because of a premature stop codon). We were not able to detect such protein. Therefore, the bio1-1 homozygous mutant plants are protein null for the splice variant (-10). The mutation also had effect on the protein from splice variant (+10) because its protein band intensity from the western blot was at least 2-fold lower than that of the WT.

The biotin biosynthetic pathway in plants occur in at least 2 compartments. KAPA synthase was shown to be localized in the cytosol (Pinon et al., 2005) and biotin synthase was shown to be localize in mitochondria (Baldet et al., 1997; Picciocchi et al., 2001). The TargetP program predicts that the protein product from BIO3/BIO1 is localized in the mitochondria. Further experiments are required to detect the subcellular localization of the BIO3/BIO1 protein. Even though the biotin biosynthetic pathway is linear, the complex spatial organization of the enzymes, and the finding that BIO3/BIO1 is genetically bifunctional, reveal the complexity of the biotin biosynthesis in plants.
MATERIALS AND METHODS

Plant Materials and Growth Conditions

The wild-type *Arabidopsis thaliana* and *bio1-1* mutant (ecotype Columbia) seed stocks were obtained from ABRC (Arabidopsis Biological Resource Center, Columbus, OH). SALK_023399 T-DNA insertion mutant (ecotype Columbia) seed stock was obtained from Salk Institute (Salk Institute Genomic Analysis Laboratory, La Jolla, CA). Seeds were sterilized by washing them in 50% ethanol for 1-min, then a 10-min incubation in 50% bleach solution containing 0.1% Tween-20.

Seeds were grown on Murashige and Skoog (Murashige, 1973) (Invitrogen, Carlsbad, CA) agar solid media with 1X multivitamin (Sigma-Aldrich, St-Louis, MO), 0.1% sucrose and 1 mM biotin (USB Corporation, Cleveland, OH), pH 7.0, under continuous illumination (170μmol m⁻² s⁻¹) at 22°C. Seedlings were transferred to LC1 Sunshine Mix soil (Sun Gro Horticulture, Bellevue, WA). Plants were watered with 1 mM biotin twice a week unless otherwise stated.

PCR Analysis

Taq polymerase (Invitrogen, Carlsbad, CA) were used for PCR amplification. PCR conditions: 1. Incubate at 96°C for 10 min. 2. Incubate at 94°C for 15 sec. 3. Incubate at 65°C for 30 sec. 4. Incubate at 72°C for 2 min. 5. Cycle to step 2 for 35 more times. 6. Incubate at 72°C for 4 min for final extension. The resulting PCR products were analyzed by electrophoresis in 1% agarose gels. Gel slices containing DNA were excised and DNA purified (QIAquick® Gel Extraction Kit, QIAGEN, Valencia, CA) following manufacturer protocol. DNA was sequenced (DNA Facility, Iowa State University) with automated Applied Biosystems sequencing equipment (Foster City, CA). For SALK_023399 characterization,
forward primer sequence (LP) 5’-ACCATCGGATTGACAAACCA-3’ and reverse primer sequence (RP) 5’-GCCTCTTTTCCTCCCTCCAGCAA-3’ were used to amplify wild-type allele; forward primer sequence (Lba1) 5’-TGGTTCACGTAGTGGCCATCG-3’ and reverse primer sequence (RP) were used to amplify T-DNA mutant allele.

**Cloning, Overexpression and Purification**

Total RNA was extracted from young leaves using RNeasy kit as recommended by the manufacturer (QIAGEN). RNA was then digested with DNasel (Invitrogen) to minimize the presence of genomic DNA. Step 2 of the manufacturer protocol is modified to incubate the reaction at 37°C for 30 min instead of at room temperature for 15 min. First strand cDNA is then synthesized using Oligo dT (Superscript™ First-strand synthesis system for RT-PCR, Invitrogen). BIO3/BIO1 cDNA and BIO3 cDNA (for making BIO3-specific antibody) were amplified using PrimeStar HS DNA Polymerase as recommended (Takara Mirus Bio, Madison, WI) with forward primer sequence 5’-CACCATGATACCCGTAACCGC-3’ and reverse primer sequence 5’-AGCTGGAGAGAGTTTTGGGT-3’. The final PCR products were cloned in pENTR vector (Invitrogen). Splice variant (-10) and splice variant (+10) were identified by sequencing. The constructs from pENTR were then moved to pDEST17 vector using Gateway Technology as recommended by the manufacturer (Invitrogen). The constructs were then expressed in BL21-AI (Invitrogen) at 37°C and purified using His-Bind kits inclusion body purification as recommended by the manufacturer (Novagen, San Diego, CA).
Semi-quantitative RT-PCR Analysis

To identify the mRNA accumulation of splice variant (-10) and (+10), BIO3/BIO1 cDNA was amplified with forward primer sequence 5'-CCAGCAATTGGATGGTTGCAA-3' and reverse primer sequence 5'-AAGCCCGACTTTTCCACTC-3'. Ubiquitin 10 were used as an internal control with forward primer sequence 5'-'GGTCTGCTTTGGCGGATTACA-3' and reverse primer sequence 5'-'CCTTGACGTTGTAATGTC-3'. The RT-PCR products were subjected to electrophoresis in 20% polyacrylamide gels and the band intensities were analyzed using ImageJ (http://rsb.info.nih.gov/ij/).

RT-PCR Analysis

For bio1-1 mutant characterization, forward primer sequence 5'-CATCTCAGCTCGGTTAAACG-3' was designed from exon region before the bio1 point mutation, reverse primer sequence 5'-TGAATTCTCAGCTCTTTTGTAG-3' was designed from exon region after the bio1 point mutation. To show At5g57590 is fused to an upstream gene At5g57600, forward primer sequence 5'-TGTCTTCTCCAAGCTCGATT-3' was designed from At5g57600, reverse primer sequence 5'-AAGTAGGATCTGCCCCTGTG-3' was designed from At5g57590.

RT-PCR using primers located at the annotated 3'UTR of At5g57600 were carried out to confirm the fusion of the 2 genes: 5'-primer 1: 5'-TCGTCTTCTCCAAGCTCGATT-3', 5'-primer 2: 5'-AATTTTTCAGTGGCTTTATCAGTTTA-3', 3'-primer 3: 5'-TAAACTGATAAAAGCCACTGAAAATT-3', 3'-primer 4: 5'-AAGCCCGCTTTTCCACTC-3'. The RT-PCR products were visualized with 1% agarose gel electrophoresis.
Bacteria Materials and Growth Conditions

The EST clone RZ128g09R was obtained from the First Laboratory for Plant Gene Research (Chiba, Japan). The *Escherichia coli* wild-type strain (BW25113) and mutant strains (ID: JW0757, JW0761, JW5264) were gifts from Dr. Tomas A. Bobik (Iowa State University, Ames, IA) which were originally obtained from Keio collection (Japan) (Datsenko and Wanner, 2000; Baba et al., 2006). *E. coli* was cultured in Luria broth medium at 37°C supplemented when appropriate with 50 μg/mL kanamycin and/or 100 μg/mL ampicillin. Biotin auxotrophy of *E. coli* strains was tested on M9 minimal medium containing 0.2% (w/v) glucose (Sambrook et al., 1989) with the addition of avidin (0.5 μg/mL for solid medium, 10 μg/mL for liquid medium).

Functional Complementation of the *E. coli* mutants

The *E. coli* mutants were made into λ.DE3 lysogen as recommended by manufacturer (λ.DE3 lysogenization kit, Novagen) prior to transformation. The pDEST17-splice variant (-10), pDEST17-splice variant (+10), and an empty pDEST17 vector were transformed into the *E. coli* mutants by electrophoresion using MicroPulser (Bio-Rad Laboratories (Hercules, CA)).

Protein Extraction, Quantification and Western Blot Analysis

Plant proteins were extracted with the following buffer: 0.1M Hepes-KOH pH 7.0, 20 mM 2-mercaptoethanol, 2% SDS, 20% glycerol, 1 mM EDTA. Bacterial proteins were extracted with 1X SDS buffer as recommended by manufacturer (Gateway Technology, Invitrogen).

Proteins were measured by the Bradford method (Bradford, 1976) using Bio-Rad protein assay reagent. The BIO3-specific antibody was produced from mice by
Hybridoma Facility (Iowa State University). Protein extracts were subjected to SDS-PAGE and/or western blotting with BIO3-specific antibody (dilution 1:1,000) following mouse-HRP secondary antibody (dilution 1:3,000) (Bio-Rad). The detection was achieved by chemiluminescence (ECL western blotting reagents, GE Healthcare Bio-Sciences Corporation, Piscataway, NJ).

ACKNOWLEDGEMENTS

We thank Dr. Thomas Bobik for guidance during the use of his microplate reader to generate growth curves. We are grateful to the National Science Foundation for funding this project.

REFERENCES


(1996). Cloning, sequencing, and characterization of the *Bacillus subtilis*


Rasband, W., pp. ImageJ.


Figure 1. Biotin biosynthesis pathway (modified from Kiyasu et al, 2001).

There are four reactions involved from the synthesis of biotin from pimeloyl-CoA. \textit{bioF}, \textit{bioA}, \textit{bioD}, and \textit{bioB} are bacteria genes for these enzymes. \textit{AtbioF} (\textit{BIO4}), \textit{BIO1}, \textit{BIO3}, and \textit{BIO2} represent the homologous plant genes.
Figure 2. Genetic characterization of bio1-1.

(A) Chromatogram of sequencing result. Arrow points to the site of mutation. WT: wild-type Columbia, hm: homozygous.

(B) Genomic sequence around the mutation site that is denoted with an asterisk.
Figure 3. Gene structure of the full length BIO1 cDNA.

(A) RZ128g09R is an EST clone corresponding to At5g57590. 5'-RACE and 3'-RACE were used to obtain the full length sequence of this cDNA. At5g57590 is fused to an upstream gene, At5g57600. The arrow heads represent the location of primers used for RT-PCR.

(B) Image of gel electrophoresis of RT-PCR. Lane 1: DNA ladder; 2: RT-PCR product from WT Columbia first strand cDNA.

(C) PCR and RT-PCR using corresponding primers. The arrow heads represent the location of primers. Primer 2 and Primer 3 are located at the annotated 3'-UTR of At5g57600. RT-PCR template: WT Columbia first strand cDNA. PCR template: WT Columbia genomic DNA.
A

<table>
<thead>
<tr>
<th>1171</th>
<th>1180</th>
<th>1190</th>
<th>1200</th>
<th>1210</th>
<th>1220</th>
<th>1230</th>
<th>1240</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpliceVariant(-10)</td>
<td>TGGTGGACACGGGCGGCACTCTTTTTTCGAGGCTG---------AGCTTTGCAAGAAGATGGTTAAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpliceVariant(+10)</td>
<td>TGGTGGACACGGGCGGCACTCTTTTTTCGAGGCTG---------AGCTTTGCAAGAAGATGGTTAAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td>TGGTGGACACGGGCGGCACTCTTTTTTCGAGGCTG---------AGCTTTGCAAGAAGATGGTTAAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

BIO3/BIO1

Splice variant (-10)

Splice variant (+10)

C

>gi|145428|gb|AAA23518.1| dethiobiotin synthetase
Length=219

Score = 30.8 bits (68), Expect = 4.2
Identities = 30/115 (26%), Positives = 48/115 (41%), Gaps = 13/115 (11%)

Query 156 YAUEAISPHLAERENATVEDSVVLQMIKLEEMGCVKS-ESGDLLCLVETAGGVA 216
  Y + SPH+ +E +E V M G++ E+ LVE AGG

Sbjct 75 YTFALPTSPIISAQEGRPIESLVS----------MSAGLRALEGADVLVEAGGUF 122

Query 217 SPGPSGTLCDDLSPFRLPGILVGDGLGGISGTLAAYESLKLGLYDIAAVVFED 271
  +P D +LP ILV +LG I+ + ++ G +A V D

Sbjct 123 TPLSDTFTPADWVTQEQLPVILVQVVLGCINMLTAQIQHAGLTLAGWAND 177
7,8-diamino-pelargonic acid aminotransferase

Score = 129 bits (325), Expect = 7e-30
Identities = 114/434 (26%), Positives = 176/434 (40%), Gaps = 66/434 (15%)

Query: DACASUUTQGPDTPFQAEALAREMGYTAARFGHVFPPENVYFALKCAELLDGDVGKGVWAS

Sbjct: DGMSWAAHGYN-HPQLNAARMAQDHSVYFGTIIAELCRKLVARNSGRNALE

Query: RVYFSDNGSTAIEALKMFRKFCVBNFCEATEEEKHIVKVIALRGYESMGDGAMEA

Sbjct: CVFLADSGSVAVEVANKMALQ-------YVQAKGERQ-----RFLFKNYGMGDTGAMS

Query: QAPSYTGFLQQPWTGRGFLDPPVFLSNGSWHISLPSFSSETSAPIPEYGFSTSRDEIFD

Sbjct: CDPDNSHMHL----------WKGVLPENLFAAPQSRMDGEWD---

Query: KSRDASTLARIYSAYLSKHLQHSGVRQASAHGVATLIFIEPVIMAGGMMHVDFLPGVRVLW

Sbjct: RDAR+ +A+ +H +A+IEF++GAGMNMP+P+++R

Query: ECRNRIKPVIFDEFVGFPURLVGETTTELGCKPDIAFCAKLTGMVCLAVTLATDAVF

Sbjct: IDCREGILIAEDIATOFORTGKLFACEMHAEIAPPDILGALKLTGMHTLBTATLTM

Query: DFSGSQKALLNHGSHSAYAMCATAKAIQWFKDFPETNMTSQGKTRELWDEELV

Sbjct: ETIS-NGEACGCFMHPFTKGMPLACAAANASLAILESGDWQQQACAFVOLR------EQL

Query: QQISSHSAVQRVTVGGTLFALELKDASNSGYASEYAKSLLILMLREDGIFTRPLGQIVYL

Sbjct: APARDQHVAQCDVRGAIGSVEETTHFCVNNAAALQKFFV----------IQGVWIRPFQKLTYL

Query: MCGPCTSPEICRRI

Sbjct: MPPYILFQQLQR
Figure 4. Two BI03/BI01 differentially spliced transcripts.

(A) Comparison of BI03/BI01 cDNA. Splice variant (-10) cDNA contains 2,502 nucleotides. Splice variant (+10) cDNA contains 2,512 nucleotides; the difference is the 10-nucleotides inserted at position 1208 of the (-10) variant.

(B) Comparison of BI03/BI01 gene structure and its two differentially spliced transcripts. Partial repeat of the 10 nucleotides are boxed.

(C) Amino acid alignment between predicted amino acid from splice variant (-10) and *E. coli* BioD protein (GenBank accession AAA23518). Query: predicted amino acid sequence from splice variant (-10), Sbjct: *E. coli* BioD amino acid sequence.

(D) Amino acid alignment between predicted amino acid from splice variant (-10) and *E. coli* BioA protein (GenBank accession AAA23514). Query: predicted amino acid sequence from splice variant (-10), Sbjct: *E. coli* BioA amino acid sequence.

(E) Comparison of translation of BI03/BI01. Splice variant (-10) transcript gives one contiguous open reading frame. The first ORF in splice variant (+10) transcript is identical to the ORF from splice variant (-10) but it stops prematurely at position 405.
Leaves
Roots
Pooled influorescent stem and leaves
Buds
Flowers
Siliques
Rosette leaves
Figure 5. Semi-quantitative RT-PCR analysis of the differentially spliced transcripts.

(A) Arrow heads represent location of primers for RT-PCR. 5'-primer is located in exon 4, 3'-primer spanned between coding region from exon 5 and exon 6.

(B) Polyacrylamide gel electrophoresis of RT-PCR products. Template: RNA was isolated from the indicated organs of WT Arabidopsis (Columbia). Ubiquitin (UBQ) served as internal standard.

(C) Band intensities from panel B were quantified using ImageJ program (http://rsb.info.nih.gov/ij/). The age and type of organs sampled is indicated.
Figure 6. Molecular characterization of the bio1-1 mutant lesion.

(A) Gene structure of BIO3/BIO1 locus of bio1-1 mutant. The asterisk represents the site of the point mutation. Closed box represents exons and solid line represents introns. All exons and introns are drawn to scale. The arrow heads represent the location of primers used for RT-PCR analysis.

(B) Genomic sequence around the mutation site that is denoted with an asterisk. Small letters represent intron and capital letters represent exon.

(C) Agarose gel electrophoresis analysis of RT-PCR products. Lane 1: DNA ladder; 2: RT-PCR product from WT; 3: RT-PCR product from bio1-1 homozygous mutant; 4: WT no RT control; 5: bio1-1 no RT control.
**Figure 7.** SALK_023399 and bio1-1 are allelic.

(A) Gene structure of SALK_023399. The T-DNA insertion disrupts the second exon of At5g57600.

(B) Expected genotype structure of crossing a heterozygous SALK_023399 line with a homozygous bio1-1 line.

(C) Chromatogram of heterozygous bio1-1. Arrow points to the site of mutation: half peak of base "A" corresponds to the point mutation and a half peak of base "G" corresponds to WT.
Figure 8. Biotin dependency of *E. coli* bioD' and bioA' mutants.

The indicated strains were plated on the following medium:
(A) M9 minimal medium
(B) M9 minimal medium with kanamycin
(C) M9 minimal medium with 0.5 μg/mL avidin
(D) M9 minimal medium with 0.5 μg/mL avidin and 0.1 mM biotin
Figure 9. Complementation of *E. coli* bioD' (λ.D3) mutant.

(A) The strains indicated were plated on M9 minimal medium with 0.5 μg/mL avidin.

(B) The strains indicated were cultured in liquid M9 minimal medium with 10 μg/mL avidin. Growth of each culture was monitored by the increase in A<sub>600nm</sub>.
Figure 10. Complementation of *E. coli* bioA<sup>+</sup> (JDE3) mutant.

(A) The strains indicated were plated on M9 minimal medium with 0.5 µg/mL avidin.
(B) The strains indicated were cultured in liquid M9 minimal medium with 10 µg/mL avidin. Growth of each culture was monitored by the increase in A<sub>600nm</sub>. 
Figure 11. Overexpression and purification of BIO3-specific antibody.

(A) SDS-PAGE analysis stained with Coomassie blue. Lane 1: MW standard; 2: uninduced BIO3 in pDEST17 in BL21-AI; 3: induced BIO3 in pDEST17 in BL21-AI.

(B) SDS-PAGE and western blot analysis using BIO3-specific antibody. Lane 1: 1 µg crude protein from BL21-AI; 2: 0.01 µg purified BIO3 in pDEST17 in BL21-AI.
Figure 12. SDS-PAGE and western blot analysis using BIO3-specific antibody.

(A) Lane 1: protein extract from rosette leaves of 47-d old WT plants; Lane 2: protein extract from rosette leaves of 47-d bio1-1 mutant plants. Each lane contained 23 μg of protein.

(B) Lane 1: purified protein from splice variant (-10) expressed from pDEST17 in BL21-AI; Lane 2: protein extract from pooled aerial organs of 37-d old WT plants; Lane 3: protein extract from pooled aerial organs of 37-d old bio1-1 mutant plants. Each lane contained 30 μg of protein.

(C) Lane 1: purified protein from splice variant (+10) expressed from pDEST17 in BL21-AI; Lane 2: protein extract from pooled fluorescent stem and leaves of 37-d old WT plants; Lane 3: protein extract from pooled fluorescent stem and leaves of 37-d old bio1-1 mutant plants. Each lane contained 30 μg of protein.
<table>
<thead>
<tr>
<th>Length of amino acid</th>
<th>Predicted MW (kD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>833</td>
<td>91.935</td>
</tr>
<tr>
<td>699</td>
<td>77.206</td>
</tr>
<tr>
<td>649</td>
<td>71.702</td>
</tr>
<tr>
<td>640</td>
<td>70.598</td>
</tr>
<tr>
<td>512</td>
<td>56.926</td>
</tr>
<tr>
<td>498</td>
<td>55.316</td>
</tr>
<tr>
<td>426</td>
<td>47.110</td>
</tr>
<tr>
<td>415</td>
<td>45.919</td>
</tr>
<tr>
<td>371</td>
<td>41.132</td>
</tr>
<tr>
<td>331</td>
<td>36.627</td>
</tr>
<tr>
<td>222</td>
<td>24.673</td>
</tr>
<tr>
<td>220</td>
<td>34.404</td>
</tr>
<tr>
<td>162</td>
<td>17.864</td>
</tr>
<tr>
<td>126</td>
<td>14.123</td>
</tr>
<tr>
<td>46</td>
<td>5.376</td>
</tr>
<tr>
<td>28</td>
<td>3.287</td>
</tr>
</tbody>
</table>
Table 2. Potential ORFs of splice variant (+10)

<table>
<thead>
<tr>
<th>Length of amino acid</th>
<th>Predicted MW (kD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>405</td>
<td>44.488</td>
</tr>
<tr>
<td>271</td>
<td>29.759</td>
</tr>
<tr>
<td>221</td>
<td>24.255</td>
</tr>
<tr>
<td>212</td>
<td>23.150</td>
</tr>
<tr>
<td>84</td>
<td>9.479</td>
</tr>
<tr>
<td>70</td>
<td>7.869</td>
</tr>
</tbody>
</table>
Table 3. Quantification of semi-quantitative RT-PCR products from splice variants (+10) and (-10)

<table>
<thead>
<tr>
<th>WT tissues</th>
<th>Splice variant (+10)</th>
<th>Splice variant (-10)</th>
<th>Ratio Splice variant (-10)/Splice variant (+10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-d old plants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaves</td>
<td>7.14</td>
<td>1.60</td>
<td>0.224</td>
</tr>
<tr>
<td>Roots</td>
<td>12.06</td>
<td>3.73</td>
<td>0.309</td>
</tr>
<tr>
<td>Pooled fluorescent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stem and leaves</td>
<td>3.51</td>
<td>3.28</td>
<td>0.934</td>
</tr>
<tr>
<td>37-d old plants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buds</td>
<td>5.65</td>
<td>3.59</td>
<td>0.635</td>
</tr>
<tr>
<td>Flowers</td>
<td>18.22</td>
<td>3.05</td>
<td>0.167</td>
</tr>
<tr>
<td>Siliques</td>
<td>11.76</td>
<td>2.70</td>
<td>0.230</td>
</tr>
<tr>
<td>Rosette leaves</td>
<td>2.45</td>
<td>0.56</td>
<td>0.229</td>
</tr>
</tbody>
</table>
Table 4. Segregation of SALK_023399 mutant allele in progeny from a heterozygous parent

<table>
<thead>
<tr>
<th></th>
<th>Biotin</th>
<th>WT</th>
<th>HT</th>
<th>HM</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3</td>
<td>+</td>
<td>23</td>
<td>10</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>T4</td>
<td>+</td>
<td>11</td>
<td>10</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 5. Growth phenotype distribution of germinated seedlings of progeny derived from a heterozygous parent carrying the SALK_023399 mutant allele (phenotype was scored 18-days after sowing of seeds)

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>- Biotin</th>
<th>+ Biotin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy, normal</td>
<td>51</td>
<td>52</td>
</tr>
<tr>
<td>Dying seedlings</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Not germinated</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>85</strong></td>
<td><strong>94</strong></td>
</tr>
</tbody>
</table>
Table 6. Segregation of phenotype and SALK_023399 mutant allele in progeny derived from a heterozygous parent

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>+ biotin</th>
<th>WT</th>
<th>HT</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy, normal</td>
<td>52</td>
<td>17</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>Dying seedlings</td>
<td>19</td>
<td>8</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>25</td>
<td>46</td>
<td>0</td>
</tr>
</tbody>
</table>
CHAPTER 3. GENERAL CONCLUSIONS

The biotin biosynthetic pathway is well characterized in bacteria, with the exception of the reactions prior to the synthesis of pimeloyl-CoA. The synthesis of biotin from pimeloyl-CoA consists of 4 reaction steps, which are conserved between bacteria and plants. The second step of the pathway is catalyzed by DAPA aminotransferase and the third step of the pathway is catalyzed by dethiobiotin synthetase. The genes encoding both of these enzymes were not characterized in Arabidopsis.

We have cloned the \textit{BIO3/BIO1} cDNA. During the molecular analysis of this gene, we found 2 transcripts, splice variant (-10) and splice variant (+10), that appear to be produced through differential splicing.

Splice variant (-10) transcript originates from the annotated At5g57590 locus, and At5g57600 locus, which are immediately adjacent on chromosome 5 of Arabidopsis. We have shown that the BIO3/BIO1 protein, encoded from the splice variant (-10), has the ability to rescue both \textit{E. coli bioD} (defective in dethiobiotin synthetase) and \textit{bioA} (defective in DAPA aminotransferase) mutants. We have genetically shown that the BIO3/BIO1 bifunctional protein is able to carry out 2 distinct reactions, carboxylation and transamination. This finding is unique in biology because the bifunctional protein is encoded by 1 gene as opposed to 1 gene encoding one functional protein.

Splice variant (+10) transcript originates primarily from the annotated At5g57600 locus. Its encoded protein was able to carry the dethiobiotin synthetase reaction, because it could rescue the \textit{E. coli bioD} mutant.

The spatial and temporal expression patterns of the \textit{BIO3/BIO1} mRNA were identified. Among the organs we examined, the splice variant (+10) has the
highest accumulation in 37-day flowers while the splice variant (-10) has the highest accumulation in 21-day roots. The abundance of splice variant (+10) was always higher than that of splice variant (-10). The ratio of splice variant (-10) and splice variant (+10) was almost 1:1 in the pooled influorescent stem and leaves from 37-day old plants. However, we could not detect the protein from splice variant (-10) as much as that of splice variant (+10) in the western blot analysis.

Our major finding described the discovery of the unique gene-protein organization. The gene and protein responsible for coding dethiobiotin synthetase were also mentioned for the first time which were not found elsewhere. To prove the bifunction of the BIO3/BIO1 protein biochemically, enzymatic assay may be used in the future.